



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2025년07월10일
(11) 등록번호 10-2829712
(24) 등록일자 2025년07월01일

- (51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) A61P 31/18 (2006.01)
C12N 15/10 (2017.01) C12N 15/113 (2010.01)
- (52) CPC특허분류
C12N 9/22 (2025.05)
A61P 31/18 (2018.01)
- (21) 출원번호 10-2023-7002157(분할)
(22) 출원일자(국제) 2014년08월29일
심사청구일자 2023년02월16일
(85) 번역문제출일자 2023년01월18일
(65) 공개번호 10-2023-0017364
(43) 공개일자 2023년02월03일
(62) 원출원 특허 10-2016-7008253
원출원일자(국제) 2014년08월29일
심사청구일자 2019년08월26일
(86) 국제출원번호 PCT/US2014/053441
(87) 국제공개번호 WO 2015/031775
국제공개일자 2015년03월05일
(30) 우선권주장
61/871,626 2013년08월29일 미국(US)
(뒷면에 계속)
- (56) 선행기술조사문헌
SCIENTIFIC REPORTS, 2013.08.26, vol. 3, no. 2510, pages 1-7.
Science, 2013.03.15., vol. 339, pages 819-823.
GENOME BIOLOGY, 2013.02.26., vol. 14, no. 2, pages 1-4.

- (73) 특허권자
템플 유니버시티-오브 더 커먼웰스 시스템 오브
하이어 에듀케이션
미합중국 19122 펜실베이니아, 필라델피아 브로드
스트리트 앤드 몽고메리 애비뉴
- (72) 발명자
칼릴리, 카멜
미국, 펜실베이니아 19004, 발라 신위드, #718, 프
래지덴셜 블러바드 190
후, 웬후이
미국, 뉴저지 08003, 체리 힐, 유로파 코트 224
- (74) 대리인
강명구

전체 청구항 수 : 총 4 항

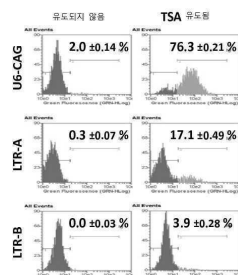
심사관 : 김정희

(54) 발명의 명칭 HIV 감염증의 RNA-유도 치료를 위한 방법 및 조성물

(57) 요약

본 발명은 면역결핍 바이러스 감염증의 치료를 위한 방법 및 조성물에 관한 것이다. 조성물은 CRISPR-연관 엔도 뉴클레아제 및 가이드 RNA를 포함하는 단리된 핵산 서열을 포함하고, 여기서 가이드 RNA는 인간 면역결핍 바이러스 내 표적 서열에 상보적이다.

대표도



(52) CPC특허분류

C12N 15/102 (2013.01)

C12N 15/113 (2013.01)

C12N 2310/20 (2025.05)

(30) 우선권주장

62/018,441 2014년06월27일 미국(US)

62/026,103 2014년07월18일 미국(US)

명세서

청구범위

청구항 1

포유류 세포의 인간 면역결핍 바이러스(HIV) 서열의 일부 또는 모두를 절단하기 위한 조성물로서, 이때 조성물은 다음을 인코딩하는 발현 벡터를 포함하고:

- (a) 클러스터링된 규칙적 간격의 짧은 회문식 반복부(CRISPR)-연관 엔도뉴클레아제;
- (b) 제 1 가이드 RNA(gRNA), 이 때 제 1 gRNA는 HIV 서열의 5' 긴 말단 반복부(LTR) 내부의 제 1 표적 핵산 서열에 상보적이고; 및
- (c) 제 2 gRNA, 이때 제 2 gRNA는 HIV 서열 내부의 제 2 표적 핵산 서열에 상보적이고;

이때 상기 제 1 표적 핵산 서열과 제 2 표적 핵산 서열의 뉴클레오티드 서열은 동일하지 않고,

이때 상기 gRNA는 crRNA 및 tracrRNA를 포함하고, 그리고

상기 CRISPR-연관 엔도뉴클레아제, 제 1 gRNA, 및 제 2 gRNA는 CRISPR-연관 엔도뉴클레아제에 의해 인식되는 제 1 표적 핵산 서열과 제 2 표적 핵산 서열 사이의 HIV 서열을 절단할 수 있는, 조성물.

청구항 2

제1항에 있어서, 상기 제 1 표적 핵산 서열은 상기 5' LTR의 U3, R, 또는 U5 부위 내에 존재하는, 조성물.

청구항 3

제1항에 있어서, 상기 제 1 표적 핵산 서열은 3' LTR의 U3, R, 또는 U5 부위 내에 존재하는, 조성물.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, CRISPR-연관 엔도뉴클레아제를 인코딩하는 핵산 서열, 제 1 gRNA를 인코딩하는 핵산 서열, 및 제 2 gRNA를 인코딩하는 핵산 서열은 하나의 발현 벡터 내부에 존재하는, 조성물.

발명의 설명

기술 분야

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2013년 8월 29일에 출원된 미국 가출원 특허 제61/871,626호; 2014년 6월 27일에 출원된 미국 가출원 특허 제62/018,441호; 및 2014년 7월 18일에 출원된 미국 가출원 특허 제62/026,103호의 우선일 이익을 주장한다. 미국 가출원 특허 제61/871,626호, 미국 가출원 특허 제62/018,441호, 및 미국 가출원 특허 제62/026,103호의 이익을 주장할 수 있는 임의 출원의 목적에 있어서, 이들 선행 출원된 출원의 내용은 본 명세서에 그 전체가 참고로서 포함된다.

[0003] 서열 목록

[0004] 본 출원은 ASCII 형태의 전자 파일로 제출되었고 그 전체가 본 명세서에 참고로서 포함된 서열 목록을 포함한다. 2014년 8월 26일에 생성된 상기 ASCII 파일의 제목은 F5129-00031_SL.txt이며 크기는 74,547 바이트이다.

[0005] 연방 후원 연구에 관한 선언

[0006] 본 발명은 미 국립보건원이 수여한 인가 번호 제R01MH093271호, 제R01NS087971호, 및 제P30MH092177호 하의 정부지원으로 이루어졌다. 미국 정부가 본 발명의 특정한 권리를 가질 수 있다.

[0007] 발명의 분야

[0008] 본 발명은 레트로바이러스, 예를 들면 인간 면역결핍 바이러스 (HIV)에서 표적 서열을 특이적으로 절단하는 조성물에 관한 것이다. CRISPR (클러스터링된 규칙적 간격의 짧은 회문식 반복부, Clustered Regularly Interspace Short Palindromic Repeat) 연관 엔도뉴클레아제 및 인간 면역결핍 바이러스 내 표적 서열에 상보적인 가이드 RNA 서열을 인코딩하는 핵산을 포함할 수 있는 그러한 조성물이 HIV 감염증에 걸렸거나 걸릴 위험이 있는 개체에 투여될 수 있다.

배경 기술

[0009] 발명의 배경

[0010] HIV-1가 발견된 지 삼십 년이 넘는 기간 동안, AIDS는 전세계 삼천오백삼십만명이 넘는 사람들에게 발생한 중요한 공공 보건 문제가 되었다. AIDS는 숙주 유전체로의 HIV-1의 계속되는 통합 때문에 불치로 남아있다. HIV-1 감염증을 제어하고 AIDS 발달을 지연시키려는 현재의 치료 (고도로 활성인 항레트로바이러스 요법 또는 HAART)는 HIV-1 감염을 유지시키는 세포 내 바이러스 복제를 크게 줄이고 혈장 바이러스혈증을 최소 수준까지 낮춘다. 하지만 HAART는 낮은 수준의 바이러스 유전체 발현 및 조직 내 복제를 억제하지 못하며 잠복성으로-감염된 세포, 예를 들면, 휴지기 기억 T 세포, 뇌 대식 세포, 미소아교세포, 및 성상아교세포, HIV-1의 저장소로서 기능하는 장관-연관 림프양 세포를 표적하지 못한다. 지속적인 HIV-1 감염은 또한 심장 및 신장 질환, 골감소증, 및 신경학적 장애를 비롯한 동반질환(comorbidity)으로 이어진다. 지속되는 바이러스 저장소를 표적하는 치료 요법 전략에 대한 끊임없는 수요가 존재한다.

발명의 내용

[0011] 요약

[0012] 본 명세서에 제공된 것은 레트로바이러스 감염의 치료 및 예방과 관련한 조성물 및 방법이다. 레트로바이러스는 렌티바이러스, 예를 들면, 인간 면역결핍 바이러스; 원숭이 면역결핍 바이러스; 고양이 면역결핍 바이러스; 및 소 면역결핍 바이러스일 수 있다. 인간 면역결핍 바이러스는 HIV-1 또는 HIV-2일 수 있다. 한 구체예에서, 조성물은 CRISPR-연관 엔도뉴클레아제 및 하나 이상의 가이드 RNA를 인코딩하는 서열을 포함하는 핵산 서열을 포함하고, 여기서 가이드 RNA는 인간 면역결핍 바이러스 내 표적 서열에 상보적이다. 일부 구체예에서 핵산은 발현 벡터 내에 내포된다. 한 구체예에서, 조성물은 CRISPR-연관 엔도뉴클레아제 폴리펩티드 및 하나 이상의 가이드 RNA를 포함하고, 여기서 가이드 RNA는 인간 면역결핍 바이러스 내 표적 서열에 상보적이다. 또한 본 명세서에 개시된 핵산, 발현 벡터, 또는 폴리펩티드를 포함하는 약제학적 조성물이 제공된다. 또한 본 명세서에 제공된 것은 인간 면역결핍 바이러스 감염을 가지거나 가질 위험이 있는 개체의 치료 방법이며, 여기서 치료 방법은 치료적으로 효과적인 양의 CRISPR-연관 엔도뉴클레아제 및 하나 이상의 가이드 RNA를 인코딩하는 벡터를 포함하는 조성물을 개체에 투여하는 단계를 포함하며, 여기서 가이드 RNA는 인간 면역결핍 바이러스 내 표적 서열에 상보적이다. 또한 CRISPR-연관 엔도뉴클레아제 및 하나 이상의 가이드 RNA를 포함하는 유전자 편집 복합체(gene editing complex)를 인코딩하는 단리된 핵산을 포함하는 조성물에 세포를 노출시킴으로써 인간 세포에서 레트로바이러스를 비활성화시키는 방법이 제공되며, 여기서 가이드 RNA는 레트로바이러스 내 표적 핵산 서열에 상보적이다. 유전자 편집 복합체는 하나 이상의 돌연변이를 프로바이러스 DNA에 도입시킨다. 일부 구체예에서 돌연변이는 프로바이러스 DNA 전체 또는 실질적으로 전체를 포함할 수 있는 결실을 포함할 수 있다. 또다른 양태에서, 또다른 양태에서본 명세서에 개시된 소정량의 조성물을 포함하는 키트가 또한 제공된다.

[0013] 본 발명의 하나 이상의 구체예의 상세한 사항이 첨부된 도면 및 하기 상세한 설명에 제시된다. 본 발명의 다른 특징, 목적, 및 장점이 명세서 및 도면, 및 청구범위로부터 명백할 것이다.

도면의 간단한 설명

[0014] 도면의 간단한 설명

도 1은 Cas9/LTR-gRNA가 HIV-1로 잠복성으로 감염된 CHME5 미소아교 세포에서 HIV-1 리포터 바이러스 생산을 억제함을 나타낸다. (A) EGFP 유세포 분석의 대표적인 게이팅(gating) 다이어그램은 안정하게 발현되는 Cas9 더하기 LTR-A 또는 -B, vs. 텅빈 U6-유발 gRNA발현 벡터 (U6-CAG)에 의해 잠복성 pNL4-3-ΔGag-d2EGFP 리포터 바이러스의 TSA-유도 재활성화가 극적으로 감소했음을 나타낸다. (B) 선택된 LTR-A- 또는 -B-을 발현하는 안정한 (stable) 클론으로부터의 PCR 산물 (LTR 내 -453 내지 +43)의 SURVEYOR *Cel-I* 뉴클레아제 어세이는 극적인 삽입결실 돌연변이 패턴 (화살표)을 나타낸다. (C, D) PCR 단편 분석은 정확히 LTR A 및 B 절단 부위 (D 내 화살

표머리 및 화살표) 사이의 190-bp 부위의 결실을 나타내며, TA-클로닝 및 서열분석 결과에 의해 입증된 306-bp 단편 (C 내 화살표)을 남긴다. 도 1D는 보이는 순서에 따라 각각 SEQ ID NO 1-3을 보여준다. (E-G) LTR-A/B 안정한 클론의 서브클로닝은 EGFP 유세포 분석기로 측정하여 리포터 재활성화의 완전한 손실 (E) 및 EGFP 및 HIV-1 Rev 반응 요소 (RRE)에 대한 유전체 DNA의 표준 (F) 및 실-시간 (G) PCR 증폭에 의해 검출하여 pNL4-3-ΔGag-d2EGFP 프로바이러스 유전체의 제거를 나타내고; β-액틴은 DNA 정제 및 로딩 대조이다. (H) HIV-1 LTR U3/R/U5 부위 (-411 내지 +129)를 포괄하는 DNA 단편을 증폭하기 위해 프라이머를 이용한 LTR-A/B 서브클론 (#8, 13)의 PCR 유전형검증은 삽입결실(indel) (a, 결실; c, 삽입) 및 "온전한" 또는 혼합 LTR (b)을 나타낸다.

도 2는 Cas9/LTR-gRNA가 U1 단세포성 세포에서 잠복성 HIV-1 바이러스를 효율적으로 제거함을 나타낸다. (A) 우측, 염색체 Xp11.4. HIV-1 통합 부위에서 HIV-1 전체 유전체의 절제를 나타내는 다이어그램은 Genome-Walker 연결 PCR 키트를 이용하여 확인하였다. 좌측, 염색체 X 통합 부위-측면부착 서열을 표적하는 프라이머 쌍 (P1/P2)을 이용한 PCR 증폭절(amplicon) 길이의 분석은 전체 HIV-1 유전체 (9709-bp)가 제거되고, 두 개의 단편 (833- 및 670-bp)이 잔여함을 나타낸다. (B) 숙주 유전체 서열을 나타내는 LTR 단편 (833-bp)의 TA 클로닝 및 서열분석 (소문자, 226-bp) 및 5'-LTR의 부분 서열 (634-27=607 bp) (파선 밑줄) 및 LTR-A 표적화 부위 (두 번째 밑줄친 부분) 주변에 27-bp 결실을 갖는 3'-LTR (첫 번째 밑줄친 부분). 하단, 15개 서열 클론 증폭절로부터 규명된 두 가지 삽입결실 알렐(allele). 670-bp 단편은 숙주 서열 (226-bp) 및 LTR-A 및 B 표적 부위를 동시에 절단하여 190-bp를 절제한 후에 잔여한 LTR 서열 (634-190=444 bp)로 이루어진다. 밑줄치고 강조된 서열은 gRNA LTR-A 표적 부위와 PAM을 표시한다. 도 2B는 보이는 순서에 따라 각각 SEQ ID NO 4-13을 보여준다. (C) TSA/PMA 재활성화-유도성 p24 바이러스입자 방출이 실질적으로 차단됨을 나타내는, HIV-1 유전체의 LTR-A/B-유도 제거의 기능 분석. U1 세포를 pX260-LTR-A, -B, 또는 -A/B로 형질주입시켰다. 2-주의 프로마이신 선별 후에, 세포를 2일 동안 TSA (250 nM)/PMA로 처리한 후에 p24 Gag ELISA를 수행하였다.

도 3은 Cas9 더하기 LTR-A/B의 안정적인 발현이 새로운 HIV-1 감염으로부터 T2M-bI 세포를 면역화함을 나타낸다. (A) 항-Flag 항체를 이용한 면역세포화학 (ICC) 및 웨스턴 블롯 (WB) 분석은 2주 동안 프로마이신 (2 μg/ml)-선별된 T2M-bI 안정한 클론에서 Flag-Cas9의 발현을 확인해 준다. (B) Cas9/LTR-A/B 안정한 클론 (c1-c7)의 PCR 유전형검증은 LTR 루시페라아제 리포터 활성화의 억제와 LTR 절제의 밀접한 상관관계를 보여준다. 배수 변화는 상응하는 비-유도 수준을 넘는 TSA/PMA-유도 수준을 나타낸다. (C) 안정한 Cas9/LTR-A/B-발현 세포 (c4)를 표시된 다중감염도 (MOI)의 위형(pseudotyped)-pNL4-3-Nef-EGFP 렌티바이러스로 감염시키고 감염 2일 후, 감염 효율을 EGFP 유세포 분석기로 측정하였다. (D) 대표적인 위상차/형광 현미경은 LTR-A/B 안정한 세포가 pNL4-3-ΔE-EGFP HIV-1 리포터 바이러스 (화색)에 의한 신규한 감염 (우측 패널)에 저항성이 있지만 대조 (U6-CAG; 흑색) 세포는 없음을 나타낸다.

도 4는 인간 유전체에 대한 Cas9/LTR-A/B의 표적-이탈 효과를 나타낸다. (A) SURVEYOR 어세이는 인간 T2M-bI 및 U1 세포 내 예상/잠재 표적-이탈 부위에 삽입결실 돌연변이가 없음을 보여준다. LTR-A 적중-표적 부위 (A)를 양성 대조로서 및 텅빈 U6-CAG 벡터 (U6)를 음성 대조로서 사용하였다. (B-D) U6-CAG 대조 및 LTR-A/B 샘플에서 지칭된 삽입결실의 수를 나타내는 LTR-A/B 안정한 T2M-bI 서브클론의 전체-유전체 서열분석 (B), 양쪽 샘플에서 gRNA 표적 부위 부근 10개의 지칭된 삽입결실에 대한 상세한 정보 (C), 및 표적-이탈 지칭된 삽입결실의 분포 (D). 도 4C는 보이는 순서에 따라 각각 SEQ ID NO 14-15를 보여준다.

도 5는 인간 T2M-bI 세포의 유전체 DNA로부터 TA-클로닝 및 PCR 산물 (-411 내지 -10) 서열분석에 의해 규명된 통합된 렌티바이러스 LTR-반딧불 루시페라아제 리포터의 LTR U3 서열을 보여준다. 4가지 gRNA (LTR-A 내지 D)의 프로토스페이스(protospacer) 및 PAM (NGG) 서열 및 표시된 전사 인자의 예정 결합 부위가 강조된다. 정확한 절단 부위가 가위로 표시된다. +1은 전사 시작 부위를 표시한다. 도 5는 SEQ ID NO: 16를 보여준다.

도 6은 LTR-C 및 LTR-D가 CHME5 미소아교세포 세포에서 잠복성 pNL4-3-ΔGag-d2EGFP 바이러스의 TSA-유도 비활성화를 현저하게 억제함을 나타낸다. (A) Tat, Rev, Env, Vpu, 및 Nef를 리포터 유전자 d2EGFP와 함께 함유하는 pNL4-3-ΔGag-d2EGFP 벡터를 도식적으로 나타내는 다이어그램. (B) Cas9/LTR-D의 표적-적중 LTR 유전체에서 삽입결실 돌연변이를 나타내는 SURVEYOR 어세이, Cas9/LTR-C 형질주입된 세포는 나타내지 않음. (C) 텅빈 U6-유발 gRNA 발현 벡터 (U6-CAG)에 비교하여 Cas9/LTR-C 또는 LTR-D의 안정적인 발현에 의해 잠복성 pNL4-3-ΔGag-d2EGFP 리포터 바이러스의 TSA-유도 비활성화가 극적으로 감소됨을 나타내는 EGFP 유세포 분석의 대표적인 게이팅 다이어그램.

도 7은 양쪽 LTR-C 및 LTR-D이 HIV-1 LTR-반딧불 루시페라아제 리포터 유전자와 안정하게 통합된 T2M-bI 세포에

서 삽입결실 돌연변이를 유도했고 항시성 및 TSA/PMA-유도 루시페라아제 활성을 상당히 감소시켰음을 보여준다. (A) LTR-C, LTR-D 또는 둘다에 의한 LTR 비활성화의 상당한 감소를 나타내는 기능적 루시페라아제 리포터 어세이. (B) LTR-C 및 LTR-D (상단 화살표)에 의해 유도된 LTR DNA (-453 내지 +43) 내 삽입결실 돌연변이를 나타내는 SURVEYOR 어세이. LTR-C 및 LTR-D의 조합은 LTR-C 및 LTR-D 사이의 302 bp 부위의 결실로부터 야기된 194 bp 단편 (하단 화살표)를 생성한다. (C, D) LTR-C에 있어서 23% 및 LTR-D에 있어서 13%로 삽입결실 효율을 나타내는 30개 클론의 Sanger 서열분석 및 삽입/결실을 나타내는 예시 크로마토그램. 도 7C는 보이는 순서에 따라 각각 SEQ ID NO 17-25를 보여준다. 도 7D는 보이는 순서에 따라 각각 SEQ ID NO 26-30를 보여준다. (E) U6-CAG 대조 샘플에서 두 개의 주요 밴드 (96 bp 및 270 bp)를 나타내지만, 96/102 부위의 LTR-C-유도 삽입결실 돌연변이 후 추가적인 372 bp 밴드 (상단 화살표), 372 부위의 LTR-D-유도 돌연변이 후 290 bp 밴드 (중간 화살표) 및 LTR-C/D-유도 절제 후 180 bp 단편 (하단 화살표)를 나타내는 LTR의 -453 내지 +43를 포괄하는 PCR 산물의 5 부위 (96, 102, 372, 386, 482)를 절단하기 위해 *Bsa*II를 이용하는 PCR-제한 단편 길이 다형성 (RFLP) 분석. (F) LTR-C 및 LTR-D 사이의 302 bp 단편의 결실(상단) 및 추가적인 17 bp 결실(하단)을 나타내는 예시 크로마토그램. 붉은 화살표는 연결 부위를 표시한다. *P<0.05는 U6-CAG 대조에 비하여 LTR-C 또는 LTR-D-매개 루시페라아제 활성화의 상당한 감소를 나타낸다. 도 7F는 보이는 순서에 따라 각각 SEQ ID NO 31-32를 보여준다.

도 8은 HIV-1 LTR U3/R/U5 부위 (-411 내지 +129)를 포괄하는 프라이머를 이용한 LTR-A/B 및 텅빈 U6-CAG 대조의 CHME5 서브클론으로부터의 PCR 산물의 TA 클로닝 및 Sanger 서열분석을 나타낸다. (A) 표시된 바와 같은 잠재 단편 a-c를 생성하는 5'- 및 3'- LTR 양쪽에 대한 LTR-A 및 LTR-B 절단의 가능한 조합. (B) LTR-A 및 LTR-B 절단 부위 사이에 190 bp 결실을 나타내는 절편 (351 bp)의 블라스트(Blast) 처리 결과. (C) LTR-A 절단 부위의 175 bp 삽입 및 LTR-B 절단 부위의 27 bp 결실을 나타내는 단편 c (682 bp)의 블라스트 결과. 도 8C는 보이는 순서에 따라 각각 SEQ ID NO 33-34를 보여준다.

도 9는 Cas9/LTR-gRNA가 U1 단세포성 세포에서 잠복성 HIV-1 바이러스를 효율적으로 제거함을 보여준다. (A) 염색체 2 통합 부위-측면부착 서열 (소문자, 467-bp)을 표적하는 프라이머 쌍 (T492/T493)을 이용한 긴-범위 PCR로부터의 1.1 kb 단편의 Sanger 서열분석은 전체 HIV-1 유전체 (9709-bp)이 제거되고, PAM (*TGG*) LTR-A 표적화 부위 (밑줄) 및 4-bp 결실 (*nnnn*)로부터 조합된 5'-LTR (파선 밑줄) 및 정확히 세 번째 뉴클레오타이드에 6-bp 삽입(박스)을 갖는 3'-LTR 이 잔여함을 보여준다. 도 9A는 SEQ ID NO: 35를 보여준다. (B) 대표적인 DNA 겔 사진은 HIV-1 유전체의 특이적인 소거를 보여준다. NS, 비-특이적 밴드. (C, D) Gag 유전자 (T457/T458)를 표적하는 프라이머 쌍을 이용한 정량적 PCR 분석은 Cas9/LTR-A/B-발현 U1 세포에서 85% 효율로 전체 HIV-1 유전체 소거를 나타낸다. U1 세포를 pX260 텅빈 벡터 (U6-CAG) 또는 LTR-A/B-인코딩 벡터로 형질주입시켰다. 2-주 푸로마이신 선별 후에, 세포 유전체 DNA를 기준화된(spiked) pNL4-3-ΔE-EGFP 인간 유전체 DNA를 표준으로서 이용하여 절대 정량 qPCR 분석을 위해 사용하였다. **P<0.01은 U6-CAG 대조에 비하여 상당한 감소를 나타낸다.

도 10은 Cas9/LTR gRNA이 J-Lat 잠복성으로 감염된 T 세포에서 HIV-1 촉진 바이러스를 효과적으로 소거함을 나타낸다. (A) EGFP 유세포 분석에 의한 기능적 분석은 PMA의 대략 50% 감소 및 EGFP 리포터 바이러스의 TNF α-유도 비활성화를 보여준다. (B) SURVEYOR 어세이는 Cas9/LTR-A/B 형질주입된 세포의 표적-적중 LTR 유전체에서 삽입결실 돌연변이(화살표)를 나타낸다. J-Lat 세포를 pX260 텅빈 벡터 또는 LTR-A 및 -B로 형질주입시켰다. 2-주의 푸로마이신 선별 후에, 세포를 24시간 동안 PMA 또는 TNF α로 처리하였다. 유전체 DNA를 HIV-1 LTR U3/R/U5 부위 (-411 내지 +129)를 포괄하는 프라이머를 이용하여 PCR 처리하였고 따라서 SURVEYOR 어세이를 수행하였다. **P<0.01은 U6-CAG 대조에 비하여 상당한 감소를 나타낸다. (C) HIV-1 LTR (-374 내지 +43)를 포괄하는 프라이머를 이용한 PCR 단편 분석은 LTR A 및 B 절단 부위 사이의 190-bp 부위가 정확히 결실되어, 227-bp 단편 (화살표)이 잔여함을 나타낸다. 항존(House-keeping) 유전자 β-액틴은 DNA 정제 및 로딩 대조로서 기능한다.

도 11은 유전체 편집 효율이 Cas9 및 gRNA의 존재에 의존함을 나타낸다, (A, B) PCR 유전형검증은 푸로마이신-선별된 TZM-bI 서브클론에서 유전체 편집의 어떠한 흔적도 없이 U6-유발 LTR-A 또는 LTR-B 발현 카세트 (A)의 부재 및 CMV-유발 Cas9 DNA (B)의 부재/감소를 보인다. 지정된 서브클론으로부터의 유전체 DNA를 U6 프로모터 (T351) 및 LTR-A (T354) 또는 -B (T356)를 포괄하고, Cas9 (T477/T491)를 표적하는 프라이머 쌍을 이용하여 통상적인 (A) 또는 실시간 (B) PCR 분석으로 처리하였다. (C, D) Cas9 단백질 발현은 비효과적인 TZM-bI 서브클론에서 부재한다. Flag-태그된 Cas9 융합 단백질은 항-Flag 단클론 항체를 이용한 웨스턴 블롯 (WB) 및 면역세포 화학 (ICC)으로 검출하였다. 안정적으로 Flag-Cas9를 발현하는 HEK293T 세포주를 WB를 위한 양성 대조로 사용하였다 (C). GAPDH는 단백질 로딩 대조로서 기능한다. 클론 c6는 Cas9 DNA는 보유하나 Cas9 단백질 발현이 없기 때문에, 푸로마이신 선별 후 후성적 억제의 가능한 메커니즘을 시사한다. 클론 c5 및 c3은 말단이 잘린 Flag-

Cas9 (tCas9)를 나타낼 수 있다. 핵을 Hoechst 33258 (D)로 염색하였다.

도 12는 TZM-b1 세포에서의 Cas9/LTR-A/B gRNA의 안정한 발현이 위형 또는 자연적인 HIV-1 바이러스에 대한 면역성을 일으켰음을 보여준다. (A) 유세포 분석은 TZM-b1 서브클론을 발현하는 Cas9/LTR-A/B에서 자연적인 pNL4-3-ΔE-EGFP 리포터 바이러스 감염 효율이 상당히 감소했음을 나타낸다. (B, C) 실시간 PCR 분석은 Cas9/LTR-A/B gRNA에 의한 바이러스 RNA (B) 및 DNA (C)의 억제 또는 제거를 나타낸다. (D) 반딧불-루시페라아제 발광 어세이는 Cas9/LTR-A/B gRNA에 의한 바이러스 감염-촉진된 LTR 프로모터 활성의 극적인 저해를 보여준다. 안정한 Cas9/LTR-A/B gRNA-발현 TZM-b1 세포를 지정된 자연적인 HIV-1 바이러스로 2 시간 동안 감염시키고, PBS로 두 차례 세척하였다. 감염 2일 후에, 세포를 수집하고, 고정하고 유세포 분석에 의해 EGFP 발현 (A)을 측정하고, 또는 총 RNA 추출 및 RT-qPCR (B), qPCR을 위한 유전체 DNA 정제 (C) 및 형광 측정 (D)을 위해 용리하였다. * $P < 0.05$ 및 ** $P < 0.01$ 은 U6-CAG 대조에 비하여 상당한 감소를 나타낸다.

도 13은 예측된 LTR gRNA 및 이들의 표적-이탈 수(100% 일치)를 나타낸다. pHR'-CMV-LacZ 렌티바이러스 벡터 (AF 105229)의 5'-LTR 센스 및 안티센스 서열 (각각 SEQ ID NO 79-111 및 112-141) (634 bp)를 활용하여 20-bp 가이드 서열 (프로토스페이서)를 내포하는 Cas9/gRNA 표적 부위 더하기 프로토스페이서 인접 모티프 서열 (NGG)을 검색하기 위해 Jack Lin CRISPR/Cas9 gRNA 검색 도구 (<http://spot.colorado.edu/~slln/cas9.html>)를 사용하였다. 각각의 gRNA 더하기 NGG (AGG, TGG, GGG, CGG)를 이용가능한 인간 유전체 및 전사 서열에 대해 블라스트 처리하였고 1000개의 나열된 서열이 보여진다. Control + F 키를 누른 후, 표적 서열 (1-23부터 9-23까지 뉴클레오타이드)를 복사/붙여넣기하고 100% 일치하는 유전체 표적의 수를 찾는다. 반복된 유전체 라이브러리 때문에 각 검색에서 표적-이탈의 수를 3으로 나눴다. 표시된 수는 4회 검색의 합 (NGG)를 나타낸다. 높은 수 (예를 들면, gRNA 서열 (센스)에 있어서: 20, 19, 19, 17, 16, 15, 14, 13, 12)는 NGG로부터 가장 먼 gRNA 표적 서열을 나타낸다. 선택된 LTR-A/B 및 LTR-C/D에 대한 서열 및 표적-이탈 수는 각각 적색과 녹색으로 강조된다.

도 14는 PCR 및 서열분석을 위해 사용된 gRNA 표적화 부위 및 프라이머 (보이는 순서에 따라 각각 SEQ ID NOS 36-78)에 해당하는 올리고뉴클레오타이드를 도시한다.

도 15는 LTR-A 및 LTR-B의 예측된 gRNA 표적화 부위의 위치를 나타내고 모두 각각, 보이는 순서에 따라 "검색 (query) Seq" 서열을 SEQ ID NO 142-252로서, 및 "응답(ref) Seq" 서열을 SEQ ID NO 253-363로서 나타낸다.

도 16은 양쪽 LTR-C 및 LTR-D이 HIV-1 LTR반딧불 루시페라아제 리포터 유전자와 안정하게 통합된 TZMBI 세포에서 항시성 및 TSA/PMA-유도 루시페라아제 활성을 감소시켰고 조합이 정확한 유전체 절제를 유도했음을 나타낸다. 여섯 개의 gRNA 표적을 HIV-LTR의 프로모터 부위를 위해 설계하였다 (도 16A). 도 16A는 SEQ ID NO: 16를 보여준다. TZMBI 세포를 리포펙타민 2000에 의해 Cas9-EGFP 및 키메라 gRNA 발현 카세트 (PCR 산물)로 동시형질감염시켰다. 3일 후에, EGFP-양성 세포를 FACS를 통해 분류하고 군당 2000 세포를 루시페라아제 어세이를 위해 수집하였다 (도 16B). 도 16B는 SEQ ID: 31를 보여준다. 집단 분류된 세포를 2일 동안 배양하고 루시페라아제 어세이 하루 전에 TSA/PMA로 처리하였다 (도 16C). 단일 세포를 96-웰 플레이트에 분류하고 하루 동안 TSA/PMA의 부재 (도 16D) 또는 존재 (도 16E)에서 루시페라아제 어세이를 위해 합류할 때까지 배양하였다. 집단 분류된 세포로부터의 PCR 산물을 Surveyor Cel-I 뉴클레아제 어세이 (도 16F) 및 BsaI을 이용한 제한 단편 길이 다형성 (도 16G)으로 분석하여 돌연변이 (도 16F) 또는 미절단 (도 16G) 밴드 (적색 화살표)를 나타내었다. 예측한 바와 같은 LTR-C 및 LTR-D 사이의 321 bp 부위의 결실(도 16A, 적색 화살표머리)로부터 생성된 200 bp 단편 (도 16F, 16G, 흑색 화살표)을 TA-클로닝 및 서열분석으로 확인하여 정확한 유전체 절제를 관찰했다 (도 16H). 개별적인 LTR-C 및 -D 규명된 % 및 % 삽입결실 돌연변이 효율 각각으로부터 PCR 산물의 Sanger 서열분석 (도 16). * $p < 0.05$ 는 상응하는 U6-CAG 대조에 비교하여 스튜던트(student)의 t 시험을 이용할 때 통계학적으로 유의한 감소를 표시한다. 프로토스페이스 (E), 프로토스페이스 (C), 프로토스페이스 (A), 프로토스페이스 (B), 프로토스페이스 (D), 및 프로토스페이스 (F)는 보이는 순서에 따라 각각 SEQ ID NO 365, 367, 369, 371, 373, 및 375에 상응한다.

도 17은 Cas9/LTR-gRNA가 HIV-1 잠복성으로 감염된 CHME5 미소아교세포 세포주에서 EGFP 유세포 분석에 의해 측정된 HIV-1 바이러스의 항시성 및 유도성 생산을 저해했음을 나타낸다. 리포트된 유전자 d2EGFP와 함께 Tat, Rev, Env, Vpu, 및 Nef를 함유하는 pHR' 렌티바이러스 벡터를 인간 태아 미소아교세포 세포주 CHME5에 형질도입 시켰고 3'-LTR의 U3 부위 내 400 bp 결실이 예시된다(도 17A). Cas9/gRNA의 일시적인 형질감염 후에, 인간 HIV-1 LTR-A, B, C, D 단독 또는 조합은 EGFP의 강도를 감소시켰지만 LTR 프로모터 활성의 억제로 인해 비율은 감소시키지 못했다 (도 17B, 17C). 1-2주 동안의 항생제 선별 후에, EGFP 세포의 비율이 또한 감소했다 (도 17D, 17E). 안정한 선택된 클론으로부터의 PCR 산물을 Surveyor Cel-I 뉴클레아제 어세이로 분석하였고 (도

17F) 이 LTR-A 및 LTR-B에서 극적이지만 LTR-A/B의 조합(적색 화살표)에서 약한 삽입결실 돌연변이를 보여준다. 예측한 바와 같은 LTR-A 및 LTR-B 사이의 190 bp 부위의 결실(도 17H, 적색 화살표머리)로부터 생성된 331 bp 단편 (도 17F, 17G, 흑색 화살표)을 TA-클로닝 및 서열분석으로 확인하여 정확한 유전체 절제를 관찰했다 (도 17H). 도 17H는 보이는 순서에 따라 각각 SEQ ID NO 1-3을 보여준다.

도 18은 대표적인 HIV-1 서열의 LTR을 보여준다 (SEQ ID NO: 376). U3 부위는 뉴클레오타이드 1 내지 뉴클레오타이드 432에 이르고 (SEQ ID NO: 377), R 부위는 뉴클레오타이드 432 내지 뉴클레오타이드 559에 이르며 (SEQ ID NO: 378), U5 부위는 560 내지 뉴클레오타이드 644에 이른다 (SEQ ID NO: 379).

도 19는 대표적인 SIV 서열의 LTR을 보여준다 (SEQ ID NO: 380). U3 부위는 뉴클레오타이드 1 내지 뉴클레오타이드 517에 이르고 (SEQ ID NO: 381), R 부위는 뉴클레오타이드 518 내지 뉴클레오타이드 693에 이르며 (SEQ ID NO: 382), U5 부위는 694 내지 뉴클레오타이드 818에 이른다 (SEQ ID NO: 383).

발명을 실시하기 위한 구체적인 내용

상세한 설명

본 발명은, 부분적으로, RNA-유도된 클러스터링된 규칙적 간격의 짧은 회문식 반복부(CRISPR)-Cas 9 뉴클레아제 시스템 (Cas9/gRNA)을 단일 및 다중 배열로 이용함으로써 통합된 HIV-1 유전체를 HIV-1 감염된 세포로부터 제거할 수 있었다는 본 발명자의 발견을 기초로 한다. 본 발명자는 HIV-1 LTR U3 부위 내부에서 고도로 특이적인 표적을 규명하였고 이는 잠복성으로-감염된 미소아교, 전단구 및 T 세포에서 Cas9/gRNA에 의해 효율적으로 편집되어, 바이러스 유전자 발현 및 복제가 비활성화되었다. Cas9/gRNA는 숙주 세포에게 유전독성이나 표적-이탈 편집을 일으키지 않았고, 5'- 내지 3'-LTR에 걸쳐 통합된 프로바이러스 DNA의 9709-bp 단편을 완전히 절제했다. 게다가, Cas9-발현 세포 내 다중 gRNA의 존재는 HIV-1 감염을 예방하였다. 본 발견의 결과는 Cas9/gRNA가 AIDS에 대한 특이적이고, 효력이 있는 예방 및 치료적 접근법을 제공하기 위해 유전조작될 수 있음을 시사한다.

따라서, 본 발명은 CRISPR-연관 엔도뉴클레아제 및 레트로바이러스, 예컨대, HIV 내 표적 서열에 상보적인 가이드 RNA를 인코딩하는 핵산을 포함하는 조성물, 뿐만 아니라 CRISPR-연관 엔도뉴클레아제 및 HIV 내 표적 서열에 상보적인 가이드 RNA를 인코딩하는 핵산을 포함하는 약제학적 제형을 특징으로 한다. 또한 CRISPR-연관 엔도뉴클레아제 폴리펩티드 및 HIV 내 표적 서열에 상보적인 가이드 RNA를 포함하는 조성물, 뿐만 아니라 CRISPR-연관 엔도뉴클레아제 폴리펩티드 및 HIV 내 표적 서열에 상보적인 가이드 RNA를 포함하는 약제학적 제형을 특징으로 한다.

또한 레트로바이러스 감염, 예컨대, HIV 감염증을 치료하기 위해 조성물을 투여하는 방법, 바이러스 복제를 제거하는 방법, 및 HIV 감염을 예방하는 방법을 특징으로 한다. 본 명세서에 기술된 치료적 방법은 다른 항레트로바이러스 요법 (예컨대, HAART)과 함께 수행될 수 있다.

HIV 감염의 임상 경로는 다양한 요인, 가령 개체의 유전적 배경, 연령, 일반 건강, 영양, 받은 치료, 및 HIV 아형에 따라 달라질 수 있다. 일반적으로, 대부분의 개체는 감염 후 수 주 또는 수 개월 이내에 감기-유사 증상을 겪는다. 증상은 발열, 두통, 근육통, 발진, 오한, 후두염, 구강 또는 성기 궤양, 림프절 부종, 관절 통증, 도한, 및 설사를 포함할 수 있다. 증상의 강도는 개체에 따라 온화한 정도부터 심각한 정도까지 달라질 수 있다. 급성 단계 도중, HIV 바이러스 입자는 적절한 CD4 수용체 분자를 발현하는 세포에 부착하여 침입한다. 바이러스가 숙주 세포에 침입하면, HIV가 인코딩한 역전사효소가 HIV RNA의 프로바이러스 DNA를 생성시키고 프로바이러스 DNA는 숙주의 유전체 DNA에 통합되게 된다. 숙주 세포에 의해 복제되는 것은 바로 이 HIV 프로바이러스이며, 다른 세포를 이후 감염시킬 수 있는 새로운 HIV 바이러스입자를 방출시킨다. 본 발명의 방법 및 조성물은 통합된 HIV 프로바이러스 DNA의 절제에 있어서 일반적으로 및 다양하게 유용하지만, 이에 본 발명이 제한되는 것은 아니며, 조성물은 임의의 감염 단계에 있는 개체 또는 HIV 감염의 위험을 가지는 미감염 개체에게 투여될 수 있다.

최초의 HIV 감염은 수 주 내지 수 개월 내에 진정되고, 이후 전형적으로 최대 10년까지 유지될 수 있는 긴 임상적 "잠복" 기간에 들어간다. 잠복기는 또한 무증상 HIV 감염 또는 만성 HIV 감염으로 지칭된다. 개체의 CD4 림프구 수는 회복되지만, 감염-전 수준에 미치지 못하며 대부분의 개체에서 혈전전환이 일어나고, 즉, 이들은 감염 후 2주 내지 4주 이내에 혈액 내에 검출가능한 수준의 항-HIV 항체를 갖는다. 이러한 잠복 기간 도중에, 말초혈 단핵세포에는 검출가능한 바이러스 복제가 존재하지 않고 말초혈에는 배양가능한 바이러스가 없거나 거의 없을 수 있다. 잠복 기간, 또한 임상적 잠복기로 지칭되는 기간 도중에, HIV에 감염된 사람들은 HIV-연관된 증상을 전혀 경험하지 않거나, 온화한 증상만 겪을 수 있다. 하지만, HIV 바이러스는 매우 낮은 수준으로 계속 복

제된다. 항-레트로바이러스 요법으로 치료받은 개체에서, 이러한 잠복기는 수십년 또는 그 이후까지 연장될 수 있다. 그러나, 비록 항레트로바이러스 요법이 전염의 위험을 낮추기는 해도 이러한 단계의 개체는 이들이 항레트로바이러스 요법을 받아도 여전히 HIV를 다른 이들에게 전달할 수 있다. 위에서 언급한 바와 같이, 항-레트로바이러스 요법은 낮은 수준의 바이러스 유전체 발현을 억제하거나 잠복성으로 감염된 세포 가령 휴지기 기억 T 세포, 뇌 대식 세포, 미소아교세포, 성상아교세포 및 장관 연관 림프양 세포를 효율적으로 표적하지 못한다.

[0021] AIDS(후천성 면역결핍 증후군)의 임상적인 징후 및 증상은 CD4 림프구 수가 감소하여, 면역 시스템에 비가역적인 손상이 일어나면 나타난다. 많은 환자들이 또한 AIDS-연관 합병증, 가령, 예를 들면, 기회 감염 가령 결핵, 살모넬라증, 거대세포바이러스, 칸디다증, 크립토코쿠스 뇌막염, 톡소플라즈마증, 및 크립토스포리디움증; 뿐만 아니라 특정 종류의 암, 가령 예를 들면, 카포시 육종, 및 림프종; 뿐만 아니라 고갈 증후군, 신경학적 합병증, 및 HIV-연관 신장병을 보인다.

[0022] 조성물

[0023] 본 발명의 조성물은 CRISPR-연관 엔도뉴클레아제, 예컨대, Cas9, 및 레트로바이러스, 예컨대, HIV 내 표적 서열에 상보적인 가이드 RNA를 인코딩하는 핵산을 포함한다. 박테리아에서 CRISPR/Cas 자리(loci)는 이동하는 유전요소(바이러스, 전이성 원소 및 접합성 플라스미드)에 대항하는 RNA-유도된 후천성 면역 시스템을 인코딩한다. 세 가지 유형(I-III)의 CRISPR 시스템이 규명된 바 있다. CRISPR 집합은 선행 이동하는 원소에 상보적인 서열인 스페이서를 내포한다. CRISPR 집합은 전사되고 프로세싱되어 성숙한 CRISPR(클러스터링된 규칙적 간격의 짧은 회문식 반복부) RNA(crRNA)가 된다. CRISPR-연관 엔도뉴클레아제인, Cas9는 유형 II CRISPR/Cas 시스템에 속하며 표적 DNA를 절단하는 강력한 엔도뉴클레아제 활성을 가진다. Cas9는 약 20 염기 쌍(bp)의 고유한 표적 서열(스페이서로 지칭) 및 전구-crRNA의 리보뉴클레아제 III-보조 과정을 위한 가이드로서 기능하는 전사-활성화된 소형 RNA(tracrRNA)를 보유하는 성숙 crRNA에 의해 유도된다. crRNA:tracrRNA 이중복합체는 crRNA 상의 스페이서와 표적 DNA 상의 상보적 서열(프로토스페이서로 지칭) 사이의 상보적인 염기 결합을 통해 Cas9를 표적 DNA로 유도한다. Cas9는 절단 부위(PAM으로부터 3번째 뉴클레오타이드)를 지정하기 위해 트리뉴클레오타이드(NGG) 프로토스페이서 인접 모티프(PAM)를 인식한다. crRNA 및 tracrRNA는 개별적으로 발현되거나 합성 줄기고리(AGAAAU)를 통해 천연 crRNA/tracrRNA 이중복합체를 모방하도록 인공적인 융합 소형 가이드 RNA(sgRNA) 내로 조작될 수 있다. 그러한 sgRNA, 가령 shRNA는 합성되거나 시험관 내에서 직접 RNA 형질감염을 위해 전사되거나 U6 또는 H1-촉진된 RNA 발현 벡터로부터 발현될 수 있지만, 인공적인 sgRNA의 절단 효율은 개별적으로 발현된 crRNA 및 tracrRNA를 갖는 시스템보다 더 낮다.

[0024] 본 발명의 조성물은 CRISPR-연관 엔도뉴클레아제를 인코딩하는 핵산을 포함할 수 있다. 일부 구체예에서, CRISPR-연관 엔도뉴클레아제는 Cas9 뉴클레아제일 수 있다. Cas9 뉴클레아제는 야생형 스트렙토코쿠스 피로게네스(*Streptococcus pyrogenes*) 서열과 동일한 뉴클레오타이드 서열을 가질 수 있다. 일부 구체예에서, CRISPR-연관 엔도뉴클레아제는 다른 종, 예를 들면 다른 스트렙토코쿠스 종(*Streptococcus species*), 가령 써모필러스(*thermophilus*); 슈도모나 에루기노사(*Psuedomona aeruginosa*), 에세리아 콜라이(*Escherichia coli*), 또는 다른 서열을 가진 박테리아 유전체 및 고세균, 또는 다른 원핵 미생물로부터의 서열일 수 있다. 대안적으로, 야생형 스트렙토코쿠스 피로게네스(*Streptococcus pyrogenes*) Cas9 서열은 변형될 수 있다. 핵산 서열은 포유류 세포에서 효율적인 발현을 위해 최적화된, 즉, "인간화된" 코돈일 수 있다. 인간화된 Cas9 뉴클레아제 서열은 예를 들면, 유전자은행(Genbank) 접근 번호 KM099231.1 GI:669193757; KM099232.1 GI:669193761; 또는 KM099233.1 GI:669193765에 나열된 발현 벡터 중 어느 하나에 의해 인코딩된 Cas9 뉴클레아제 서열일 수 있다. 대안적으로, Cas9 뉴클레아제 서열은 예를 들면, 시판되는 벡터 가령 Addgene(Cambridge, MA)사로부터의 PX330 또는 PX260 내에 내포된 서열일 수 있다. 일부 구체예에서, Cas9 엔도뉴클레아제는 유전자은행 접근 번호 KM099231.1 GI:669193757; KM099232.1 GI:669193761; 또는 KM099233.1 GI:669193765의 Cas9 엔도뉴클레아제 서열 중 어느 하나의 변이체 또는 단편인 아미노산 서열 또는 PX330 또는 PX260(Addgene, Cambridge, MA)사의 Cas9 아미노산 서열을 가질 수 있다. Cas9 뉴클레오타이드 서열은 생물학적으로 활성인 Cas9의 변이체를 인코딩하도록 변형될 수 있고, 이들 변이체는, 예를 들면, 하나 이상의 돌연변이(예컨대, 부가, 결실, 또는 치환 돌연변이 또는 그러한 돌연변이의 조합)를 내포하는 점에서 야생형 Cas9로부터 차이나는 아미노산 서열을 가질 수 있거나 포함할 수 있다. 하나 이상의 치환 돌연변이가 치환(예컨대, 보존적 아미노산 치환)일 수 있다. 예를 들면, 생물학적으로 활성인 Cas9 폴리펩티드의 변이체는 야생형 Cas9 폴리펩티드에 대해 적어도 또는 약 50% 서열 동일성(예컨대, 적어도 또는 약 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 서열 동일성)을 가지는 아미노산 서열을 가질 수 있다. 보존적 아미노산 치환은 전형적으로 다음 군에 속하는 치환을 포함한다: 글리신 및 알라닌; 발린, 이소류신, 및 류신; 아스파르트산 및 글루탐산; 아스파라긴, 글루타

민, 세린 및 트레오닌; 리신, 히스티딘 및 아르기닌; 및 페닐알라닌 및 티로신. Cas9 아미노산 서열 내 아미노산 잔기는 비-자연발생적 아미노산 잔기일 수 있다. 자연발생적 아미노산 잔기는 유전자 암호에 의해 인코딩되는 자연적인 것들 뿐만 아니라 비-표준 아미노산 (예컨대, L-배열 대신에 D-배열을 가지는 아미노산)을 포함한다. 본 발명의 펩티드는 또한 표준 잔기의 변형된 버전인 아미노산 잔기를 포함할 수 있다 (예컨대 피롤리신은 리신 대신에 사용될 수 있고 셀레노시스테인은 시스테인 대신에 사용될 수 있다). 비-자연발생적 아미노산 잔기는 자연에서 발견되지 않지만, 아미노산의 기본 형태와 일치하며 펩티드 내에 포함될 수 있는 것들이다. 이들은 D-알로이소류신(2R,3S)-2-아미노-3-메틸펜타노산 및 L-사이클로펜틸 글리신 (S)-2-아미노-2-사이클로펜틸 아세트산을 포함한다. 다른 예시에 있어서, 교과서나 인터넷 정보를 참조할 수 있다 (현재 캘리포니아 공대가 사이트를 운영중이며 기능적 단백질에 성공적으로 통합된 비-천연 아미노산의 구조를 공개한다).

[0025] Cas9 뉴클레아제 서열은 돌연변이를 일으킨 서열일 수 있다. 예를 들면 Cas9 뉴클레아제는 가닥 특이적 절단에 관여하는 보존된 HNH 및 RuvC 도메인에 돌연변이가 일어날 수 있다. 예를 들면, RuvC 촉매 도메인 내 아스파르트레이트-에서-알라닌으로 (D10A) 돌연변이는 Cas9 틱내기효소(nickase) 돌연변이 (Cas9n)가 DNA를 자르는 대신 틱을 내게 하여 수율 단일-가닥 단절을 얻게 하며, 및 이어지는 HDR을 통한 선별적 복구는 잠재적으로 표적을-이탈한 이중-가닥 단절로 인한 원치않는 삽입결실 돌연변이의 빈도를 증가시킬 수 있다.

[0026] 일부 구체예에서, 본 발명의 조성물은 상기 기술된 핵산 서열 중 어느 하나에 의해 인코딩된 CRISPR-연관 엔도뉴클레아제 폴리펩티드를 포함할 수 있다. 용어 "펩티드", "폴리펩티드", 및 "단백질"은, 비록 이들이 전형적으로 다양한 크기의 펩티드 서열을 지칭함에도, 본 명세서에서 상호교환적으로 사용된다. 본 발명의 아미노산-기반 조성물을 이들이 아미노산 잔기의 선형 중합체임을 암시하고, 전장 단백질과 차별됨을 돕기 위해 "폴리펩티드"로 지칭할 수 있다. 본 발명의 폴리펩티드는 CRISPR-연관 엔도뉴클레아제의 단편을 "구성"하거나 "포함"할 수 있고, 본 발명은 CRISPR-연관 엔도뉴클레아제의 생물학적으로 활성인 변이체를 구성하거나 포함하는 폴리펩티드를 포괄한다. 그러므로 폴리펩티드가 CRISPR-연관 엔도뉴클레아제 (또는 이들의 생물학적으로 활성인 변이체)의 단편만을 포함할 수도 있지만 추가적인 잔기를 역시 포함할 수 있음이 이해될 것이다. 생물학적으로 활성인 변이체는 표적 DNA를 절단하기에 충분한 활성을 보유할 것이다.

[0027] 아미노산 잔기 사이의 결합은 통상적인 펩티드 결합 또는 또다른 공유 결합 (가령 에스테르 또는 에테르 결합)일 수 있고, 폴리펩티드는 아미드화, 인산화 또는 글리코실화에 의해 변형될 수 있다. 변형은 폴리펩티드 백본 및/또는 하나 이상의 곁사슬에 일어날 수 있다. 화학적 변형은 폴리펩티드를 인코딩하는 mRNA의 번역 이후에 생체 내에서 일어난 자연발생적 변형 (예컨대, 박테리아 숙주 내 글리코실화) 또는 시험관 내에서 일어난 합성적 변형일 수 있다. CRISPR-연관 엔도뉴클레아제의 생물학적으로 활성인 변이체는 하나 이상의 구조적 변형을 포함하여 자연발생적 (즉, 생체 내에서 자연적으로 일어난 것) 및 합성적 변형 (즉, 시험관 내에서 자연발생적 또는 비-자연발생적으로 일어난 변형)의 조합을 만들 수 있다. 변형의 예시는, 아미드화 (예컨대, C-말단에서 유리 카르복실기가 아미노기에 의해 교체됨); 비오틴화 (예컨대, 리신 또는 다른 반응성 아미노산 잔기가 비오틴 분자로 아실화됨); 글리코실화 (예컨대, 글리코실기가 아스파라긴, 하이드록실, 세린 또는 트레오닌 잔기 중 어느 하나에 부가되어 당단백질 또는 당펩티드가 생성됨); 아세틸화 (예컨대, 아세틸기가, 전형적으로 폴리펩티드의 N-말단에 부가됨); 알킬화 (예컨대, 알킬기의 부가); 이소프레닐화 (예컨대, 이소프레노이드기의 부가); 리포일화 (예컨대 리포에이트 모이어티의 부착); 및 인산화 (예컨대, 포스페이트기가 세린, 티로신, 트레오닌 또는 히스티딘에 부가됨)를 포함하지만 이에 제한되지 않는다.

[0028] 생물학적으로 활성인 변이체 내 하나 이상의 아미노산 잔기는 비-자연발생적 아미노산 잔기일 수 있다. 자연발생적 아미노산 잔기는 유전자 암호에 의해 인코딩되는 자연적인 것들 뿐만 아니라 비-표준 아미노산 (예컨대, L-배열 대신에 D-배열을 가지는 아미노산)을 포함한다. 본 발명의 펩티드는 또한 표준 잔기의 변형된 버전인 아미노산 잔기를 포함할 수 있다 (예컨대 피롤리신은 리신 대신에 사용될 수 있고 셀레노시스테인은 시스테인 대신에 사용될 수 있다). 비-자연발생적 아미노산 잔기는 자연에서 발견되지 않지만, 아미노산의 기본 형태와 일치하며 펩티드 내에 포함될 수 있는 것들이다. 이들은 D-알로이소류신(2R,3S)-2-아미노-3-메틸펜타노산 및 L-사이클로펜틸 글리신 (S)-2-아미노- 2-사이클로펜틸 아세트산을 포함한다. 다른 예시에 있어서, 교과서나 인터넷 정보를 참조할 수 있다 (현재 캘리포니아 공대가 사이트를 운영중이며 기능적 단백질에 성공적으로 통합된 비-천연 아미노산의 구조를 공개한다).

[0029] 대안적으로, 또는 부가적으로, 생물학적으로 활성인 변이체 내 하나 이상의 아미노산 잔기는 야생형 서열의 상응하는 위치에서 발견되는 자연발생적 잔기와 상이한 자연발생적 잔기일 수 있다. 달리 말하면, 생물학적으로 활성인 변이체는 하나 이상의 아미노산 치환을 포함할 수 있다. 아미노산 잔기의 치환, 부가, 또는 결실을 야생형 서열의 돌연변이로서 지칭할 수 있다. 언급한 바와 같이, 치환은 자연발생적 아미노산 잔기를 비-자연발생적

잔기 또는 그저 상이한 자연발생적 잔기로 교체할 수 있다. 추가로 치환은 보존적 또는 비-보존적 치환을 구성할 수 있다. 보존적 아미노산 치환은 전형적으로 다음 군에 속하는 치환을 포함한다: 글리신 및 알라닌; 발린, 이소류신, 및 류신; 아스파르트산 및 글루탐산; 아스파라긴, 글루타민, 세린 및 트레오닌; 리신, 히스티딘 및 아르기닌; 및 페닐알라닌 및 티로신.

[0030] CRISPR-연관 엔도뉴클레아제의 생물학적으로 활성인 변이체인 폴리펩티드는 이들의 서열이 상응하는 야생형 폴리펩티드와 어디까지 유사하거나 동일한 정도인지 특징지을 수 있다. 예를 들면, 생물학적으로 활성인 변이체의 서열은 야생형 폴리펩티드 내 상응하는 잔기와 적어도 또는 약 80% 동일할 수 있다. 예를 들면, CRISPR-연관 엔도뉴클레아제의 생물학적으로 활성인 변이체는 CRISPR-연관 엔도뉴클레아제 또는 이들의 동족체 또는 상동체와 적어도 또는 약 80% 서열 동일성 (예컨대, 적어도 또는 약 85%, 90%, 95%, 97%, 98%, 또는 99% 서열 동일성)을 가지는 아미노산 서열을 가질 수 있다.

[0031] CRISPR-연관 엔도뉴클레아제의 생물학적으로 활성인 변이체 폴리펩티드는 본 발명의 방법에서 유용하기에 충분한 생물학적 활성을 보유할 것이다. 생물학적으로 활성인 변이체는 표적화된 DNA의 절단에서 기능하기에 충분한 활성을 보유할 것이다. 생물학적 활성은 당해 분야의 숙련가에게 공지된 방식으로 평가될 수 있고 이는 시험관 내 절단 어세이 또는 기능적 어세이를 포함하지만 이에 제한되지 않는다.

[0032] 폴리펩티드는 예를 들면, 제조합 기술 또는 화학 합성을 비롯한 다양한 방법에 의해 생성될 수 있다. 일단 생성되면, 폴리펩티드는 당해 분야에 널리 공지된 수단에 의해 요망되는 임의의 정도까지 단리되고 정제될 수 있다. 예를 들면, 동결건조 이후, 예를 들면, 역상 (바람직하게는) 또는 정상 HPLC, 또는 다당류 겔 매체 가령 Sephadex G-25에서의 크기 배제 또는 분배 크로마토그래피를 이용할 수 있다. 최종 폴리펩티드의 조성물은 표준 수단, 아미노산 서열분석, 또는 FAB-MS 기술에 의해 펩티드를 분해한 후에 아미노산 분석에 의해 확인될 수 있다. 폴리펩티드의 아미노 기의 염, 가령 산염, 에스테르, 아마이드, 및 N-아실 유도체는 당해 분야에 공지된 방법을 이용하여 제조될 수 있고, 그러한 펩티드는 본 발명의 맥락에서 유용하다.

[0033] 본 발명의 조성물은 레트로바이러스 내 표적 서열에 상보적인 서열을 포함하는 가이드 RNA (gRNA)를 인코딩하는 서열을 포함한다. 레트로바이러스는 렌티바이러스, 예를 들면, 인간 면역결핍 바이러스; 원숭이 면역결핍 바이러스; 고양이 면역결핍 바이러스; 및 소 면역결핍 바이러스일 수 있다. 인간 면역결핍 바이러스는 HIV-1 또는 HIV-2일 수 있다. 표적 서열은 임의의 HIV, 예를 들면, HIV-1 및 HIV-2로부터의 서열, 및 이들의 임의의 순환하는 제조합 형태를 포함할 수 있다. HIV의 유전적 가변성은 기술된 복합적인 군 및 아형에 반영된다. HIV 서열의 목록은 로스 알라모드(Los Alamos) HIV 데이터베이스 및 전서에 따른 것이다 (즉, 서열 데이터베이스 웹 주소는 <http://www.hiv.lanl.gov/>이다). 본 발명의 방법 및 조성물은 이들의 다양한 군, 아형, 및 순환하는 제조합 형태의 임의의 HIV에 적용될 수 있다. 이들은 예를 들면, HIV-1 주요 군 (흔히 그룹 M으로 지칭) 및 비주요 군, 그룹 N, O, 및 P, 뿐만 아니라 이에 제한되지 않지만, 이하의 아형, A, B, C, D, F, G, H, J 및 K 중 어느 하나 또는 HIV의 군(예를 들면, 이에 제한되지 않지만 이하의 군, N, O 및 P 중 어느 하나)을 포함한다. 본 방법 및 조성물은 또한 HIV-2 및 임의의 A, B, C, F 또는 G 계통 (또한 "아형" 또는 "군"으로도 지칭됨), 뿐만 아니라 HIV-2의 임의의 순환하는 제조합 형태에 적용될 수 있다.

[0034] 가이드 RNA는 암호화 또는 비-암호화 서열에 상보적인 서열일 수 있다. 예를 들면, 가이드 RNA는 HIV 서열, 가령 긴 말단 반복 (LTR) 서열, 단백질 암호화 서열, 또는 조절 서열일 수 있다. 일부 구체예에서, 가이드 RNA는 HIV 긴 말단 반복 (LTR) 부위에 상보적인 서열을 포함한다. HIV-1 LTR은 대략 640 bp 길이다. 예시적인 HIV-1 LTR은 SEQ ID NO: 376의 서열이다. 예시적인 SIV LTR은 SEQ ID NO: 380의 서열이다. HIV-1 긴 말단 반복 (LTR)은 U3, R 및 U5 부위로 나뉜다. 예시적인 HIV-1 LTR U3, R 및 U5 부위는 각각 SEQ ID NO: 377, 378 및 379이다. 예시적인 SIV LTR U3, R 및 U5 부위는 각각 SEQ ID NO: 381, 382, 및 383이다. 예시적인 HIV-1 및 SIV 서열에 있어서 U1, R, U5 부위의 배열은 각각 도 18 및 19에 나타난다. LTR은 유전자 발현을 위해 필요한 모든 신호를 보유하며 프로바이러스가 숙주 세포의 유전체로 통합되는데 관여한다. 예를 들면, 기본 또는 핵심 프로모터, 핵심 인핸서 및 조절 부위는 U3 내부에서 발견되는 반면 전사활성화 반응 요소는 R 내부에서 발견된다. HIV-1에서, U5 부위는 전사 활성화에 관여하는 여러가지 하위-부위, 예를 들면, TAR 또는 전사-작용 반응성 요소; 이량체화 및 유전체 패키징에 관여하는 다중 A; PBS 또는 프라이머 결합 부위; Psi 또는 패키징 신호전달; DIS 또는 이량체 개시 부위를 포함한다.

[0035] 유용한 가이드 서열은 LTR의 U3, R, 또는 U5 부위에 상보적이다. HIV-1의 U3 부위를 표적하는 예시적인 가이드 RNA 서열이 도 13에 나타난다. 가이드 RNA 서열은, 예를 들면, 다음의 서열을 포함할 수 있다:

[0036] LTR A: ATCAGATATCCACTGACCTTGG (SEQ ID NO: 96),

- [0037] LTR B: CAGCAGTTCTTGAAGTACTCCGG (SEQ ID NO: 121),
- [0038] LTR C GATTGGCAGAACTACACACCAGG (SEQ ID NO: 87), 또는
- [0039] LTR D: GCGTGGCCTGGGCGGGACTGGGG (SEQ ID NO: 110).
- [0040] U3 (SEQ ID NO: 16) 부위 내부의 LTR A(SEQ ID NO: 96), LTR B (SEQ ID NO: 121), LTR C(SEQ ID NO: 87) 및 LTR D (SEQ ID NO: 110)의 위치가 도 5에 나타난다. U3 부위를 표적하는 추가적인 예시적인 가이드 RNA 서열은 도 13의 표에 나열되며 SEQ ID NO: 79-111 및 SEQ ID NO: 111-141 중 어느 하나의 서열을 가질 수 있다. 일부 구체예에서, 가이드 서열은 SEQ ID NO: 79-111 및 SEQ ID NO: 111-141 중 어느 하나와 95% 동일성을 가지는 서열을 포함할 수 있다. 따라서, 가이드 RNA 서열은, 예를 들면, 다음의 서열에 95% 동일성을 가지는 서열을 포함할 수 있다:
- [0041] LTR A: ATCAGATATCCACTGACCTTTGG (SEQ ID NO: 96),
- [0042] LTR B: CAGCAGTTCTTGAAGTACTCCGG (SEQ ID NO: 121),
- [0043] LTR C GATTGGCAGAACTACACACCAGG (SEQ ID NO: 87), 또는
- [0044] LTR D: GCGTGGCCTGGGCGGGACTGGGG (SEQ ID NO: 110).
- [0045] 또한 가이드 RNA 서열을 프로토스페이서, 예컨대, 프로토스페이스 (A), 프로토스페이스 (B), 프로토스페이스 (C), 및 프로토스페이스 (D)로 지칭할 수 있다.
- [0046] 가이드 RNA 서열은 HIV-1 U3, R, 또는 U5 부위 기준 서열 또는 공통 서열 내에서 발견되는 서열일 수 있다. 본 발명은 제한적이지 않으나, 가이드 RNA 서열은 임의의 변이체 또는 돌연변이 HIV 서열을 표적하기 위해 선택될 수 있다. 일부 구체예에서, 가이드 RNA는 변이체 서열 또는 유사(quasi)-종 서열을 포함할 수 있다. 일부 구체예에서, 가이드 RNA는 치료를 받는 개체에 잠복하는 바이러스의 유전체 내 서열에 상응하는 서열일 수 있다. 따라서 예를 들면, 개체에 잠복한 HIV 바이러스 내 특정 U3, R, 또는 U5 부위의 서열을 얻을 수 있고 환자의 특정 서열에 상보적인 가이드 RNA를 사용할 수 있다.
- [0047] 일부 구체예에서, 가이드 RNA는 단백질 암호화 서열, 예를 들면, 하나 이상의 바이러스 구조 단백질, (예컨대, gag, pol, env 및 tat)을 인코딩하는 서열에 상보적인 서열일 수 있다. 따라서, 서열은 gag 다단백질, 예컨대, MA (기질 단백질, p17); CA (캡시드 단백질, p24); SP1 (스페이서 펩티드 1, p2); NC (뉴클레오프로틴 단백질, p7); SP2 (스페이서 펩티드 2, p1) 및 P6 단백질; pol, 예컨대, 역전사효소 (RT) 및 리보뉴클레아제 H, 인테그라제 (IN), 및 HIV 프로테아제 (PR); env, 예컨대, gp160, 또는 gp160의 분해 산물, 예컨대, gp120 또는 SU, 및 gp41 또는 TM; 또는 tat, 예컨대, 72-아미노산 1-엑손 Tat 또는 86-101 아미노산 2-엑손 Tat 내부의 서열에 상보적일 수 있다. 일부 구체예에서, 가이드 RNA는 예를 들면, vif, n willef (음성 인자) vpu (바이러스 단백질 U) 및 tev를 비롯한 부속 단백질을 인코딩하는 서열에 상보적인 서열일 수 있다.
- [0048] 일부 구체예에서, 서열은 구조 또는 조절 요소, 예를 들면, 상기 기술된 바와 같은 LTR; TAR (바이러스 전사활성화를 위한 표적 서열), HIV-1 내 바이러스 mRNA의 대략 첫 번째 45 뉴클레오티드 (또는 HIV-2에서는 첫 번째 100 뉴클레오티드)로 이루어지고 헤어핀 자루-고리 구조를 이루는 Tat 단백질 및 세포 단백질을 위한 결합 부위; RRE (Rev 반응성 원소) 대략 200 뉴클레오티드로 이루어진 HIV-1의 env 부위 내에서 인코딩되는 RNA 원소 (HIV-1에서 전사 시작점부터 gp120 및 gp41의 경계를 이루는 위치 7710 내지 8061); PE (Psi 원소), Gag 출발 코돈에 선행하고 겹치는 4개의 자루-고리 구조의 집합; 자루-고리 구조가 이어지는 SLIP, TTTTTT "미끄러지는 부위"; CRS (Cis-액틴 억제 서열); INS 예를 들면, HIV-1의 gag 부위 내 뉴클레오티드 414 내지 631에서 발견되는 저해/불안정 RNA 서열)에 상보적인 서열일 수 있다.
- [0049] 가이드 RNA 서열은 센스 또는 항-센스 서열일 수 있다. 가이드 RNA 서열은 일반적으로 프로토-스페이서 인접 모티프 (proto-spacer adjacent motif, PAM)를 포함한다. PAM의 서열은 사용된 CRISPR 엔도뉴클레아제의 특이성 요건에 따라 달라질 수 있다. *S. pyogenes*에서 유도된 CRISPR-Cas 시스템에서, 표적 DNA는 전형적으로 5'-NGG 프로토-스페이서 인접 모티프 (PAM) 바로 앞에 선행한다. 따라서, *S. pyogenes* Cas9에 있어서, PAM 서열은 AGG, TGG, CGG 또는 GGG일 수 있다. 다른 Cas9 상동체는 상이한 PAM 특이성을 가질 수 있다. 예를 들면, *S. thermophilus*로부터의 Cas9는 CRISPR 1을 위해 5'-NNAGAA를 및 CRISPR3을 위해 5'-NGGNG를) 그리고 나이제리아 메니기디티스(*Neisseria meningitidis*)는 5'-NNNGATT를) 필요로 한다. 가이드 RNA의 특이적 서열은 달라질 수 있지만, 서열과 상관없이, 유용한 가이드 RNA 서열은 표적-이탈 효과를 최소화함과 동시에 유전체에 통합된 HIV-1 프로바이러스를 고효율로 및 완전하게 고갈시키는 서열일 것이다. 가이드 RNA 서열의 길이는 약 20 내지

약 60 또는 그 이상의 뉴클레오타이드, 예를 들면 약 20, 약 21, 약 22, 약 23, 약 24, 약 25, 약 26, 약 27, 약 28, 약 29, 약 30, 약 31, 약 32, 약 33, 약 34, 약 35, 약 36, 약 37, 약 38, 약 39, 약 40, 약 45, 약 50, 약 55, 약 60 또는 그 이상의 뉴클레오타이드로 다양할 수 있다. 외부 바이러스 유전체 및 내인성 레트로바이러스 DNA를 비롯한 숙주 세포 유전체 사이에 극도로 낮은 상동성을 가지는 부위를 규명하기에 유용한 선별 방법은 표적-이탈 인간 전사체 또는 (월씬 드물게는) 미변역된-유전체 부위를 배제하고; HIV-1 LTR 프로모터 (가능하게는 숙주 유전체 내에 보존된 것) 내부의 전사 인자 결합 부위를 피하기 위해 12-bp+NGG 표적-선별 기준을 이용하는 생물정보학적 선별; LTR-A- 및 -B-공략성, 30-bp gRNA 및 또한 특이성/효율을 증가시키기 위해 본래의 박테리아 면역 메커니즘을 반영하는 전구-crRNA 시스템 vs. 20-bp gRNA-, 키메라 crRNA-tracrRNA-기반 시스템 및 WGS의 선별, 가능한 표적-이탈 효과를 규명하고 배제하기 위한 Sanger 서열분석 및 SURVEYOR 어세이를 포함한다.

[0050] 가이드 RNA 서열은 단일 서열로서 또는 하나 이상의 상이한 서열의 조합, 예컨대, 다중 배열로서 구성될 수 있다. 다중 배열은 둘, 셋, 넷, 다섯, 여섯, 일곱, 여덟, 아홉, 열, 또는 그 이상의 상이한 가이드 RNA의 조합, 예를 들면 U3, R, 또는 U5 내 서열의 임의의 조합을 포함할 수 있다. 일부 구체예에서, LTR A, LTR B, LTR C 및 LTR D의 조합이 사용될 수 있다. 일부 구체예에서, 서열 LTR A (SEQ ID NO: 96), LTR B (SEQ ID NO: 121), LTR C (SEQ ID NO: 87), 및 LTR D (SEQ ID NO: 110) 중 임의의 조합이 사용될 수 있다. 일부 구체예에서, SEQ ID NO: 79-111 및 SEQ ID NO: 111-141의 서열을 가지는 임의의 서열 조합이 사용될 수 있다. 조성물이 발현 벡터에 투여되는 경우, 가이드 RNA는 단일 벡터에 의해 인코딩될 수 있다. 대안적으로, 다중 벡터가 각각 둘 이상의 상이한 가이드 RNA를 포함하도록 유전조작될 수 있다. 유용한 배열은 절단 부위 사이의 바이러스 서열을 절제하여 HIV 유전체 또는 HIV 단백질 발현의 제거를 야기할 것이다. 따라서, 둘 이상의 상이한 가이드 RNA의 사용은 CRISPR 엔도뉴클레아제에 의해 인식되는 절단 부위 사이의 바이러스 서열의 절제를 촉진한다. 절제되는 부위는 하나의 뉴클레오타이드부터 수 천개의 뉴클레오타이드까지 다양한 크기일 수 있다. 예시적인 절제 부위가 실시예에 기술된다.

[0051] 조성물이 핵산으로서 투여되거나 발현 벡터 내에 내포되는 경우, CRISPR 엔도뉴클레아제는 가이드 RNA 서열과 동일한 핵산 또는 벡터에 의해 인코딩될 수 있다. 대안적으로 또는 부가적으로, CRISPR 엔도뉴클레아제는 가이드 RNA 서열로부터 물리적으로 떨어진 핵산에서 또는 별도의 벡터에서 인코딩될 수 있다.

[0052] 일부 구체예에서, RNA 분자 예컨대 crRNA, tracrRNA, gRNA는 하나 이상의 변형된 핵염기를 포함하도록 유전조작된다. 예를 들면, RNA 분자의 공지된 변형은, 예를 들면, Genes VI, Chapter 9 ("Interpreting the Genetic Code"), Lewis, ed. (1997, Oxford University Press, New York), 및 Modification and Editing of RNA, Grosjean 및 Benne, eds. (1998, ASM Press, Washington DC)에서 찾을 수 있다. 변형된 RNA 성분은 다음을 포함한다: 2'-O-메틸시티딘; N⁴-메틸시티딘; N⁴-2'-O-디메틸시티딘; N⁴-아세틸시티딘; 5-메틸시티딘; 5,2'-O-디메틸시티딘; 5-하이드록시메틸시티딘; 5-포르밀시티딘; 2'-O-메틸-5-포르마일시티딘; 3-메틸시티딘; 2-티오시티딘; 리시딘; 2'-O-메틸유리딘; 2-티오유리딘; 2-티오-2'-O-메틸유리딘; 3,2'-O-디메틸유리딘; 3-(3-아미노-3-카르복시프로필)유리딘; 4-티오유리딘; 리보실티딘; 5,2'-O-디메틸유리딘; 5-메틸-2-티오유리딘; 5-하이드록시유리딘; 5-메톡시유리딘; 유리딘 5-옥시아세트산; 유리딘 5-옥시아세트산 메틸 에스테르; 5-카르복시메틸유리딘; 5-메톡시카르보닐메틸유리딘; 5-메톡시카르보닐메틸-2'-O-메틸유리딘; 5-메톡시카르보닐메틸-2'-티오유리딘; 5-카르바모일메틸유리딘; 5-카르바모일메틸-2'-O-메틸유리딘; 5-(카르복시하이드록시메틸)유리딘; 5-(카르복시하이드록시메틸) 유리딘메틸 에스테르; 5-아미노메틸-2-티오유리딘; 5-메틸아미노메틸유리딘; 5-메틸아미노메틸-2-티오유리딘; 5-메틸아미노메틸-2-셀레노유리딘; 5-카르복시메틸아미노메틸유리딘; 5-카르복시메틸아미노메틸-2'-O-메틸-유리딘; 5-카르복시메틸아미노메틸-2-티오유리딘; 디하이드로유리딘; 디하이드로리보실티민; 2'-메틸아데노신; 2-메틸아데노신; N^{sup}.6N-메틸아데노신; N⁶,N⁶-디메틸아데노신; N⁶,2'-O-트리메틸아데노신; 2-메틸티오-N⁶N-이소펜테닐아데노신; N⁶-(cis-하이드록시이소펜테닐)-아데노신; 2-메틸티오-N⁶-(cis--하이드록시이소펜테닐)-아데노신; N⁶-글리시닐카르바모일)아데노신; N⁶-트레오닐카르바모일 아데노신; N⁶-메틸-N⁶-트레오닐카르바모일 아데노신; 2-메틸티오-N⁶-메틸-N⁶-트레오닐카르바모일 아데노신; N⁶-하이드록시노르발릴카르바모일 아데노신; 2-메틸티오-N⁶-하이드록시노르발릴카르바모일 아데노신; 2'-O-리보실아데노신 (포스페이트); 이노신; 2'-O-메틸 이노신; 1-메틸 이노신; 1;2'-O-디메틸 이노신; 2'-O-메틸 구아노신; 1-메틸 구아노신; N²-메틸 구아노신; N²,N²-디메틸 구아노신; N²,2'-O-디메틸 구아노신; N²,N²,2'-O-트리메틸 구아노신; 2'-O-리보실 구아노신 (포스페이트); 7-메틸 구아노신; N²;7-디메틸 구아노신; N²; N²;7-트리메틸 구아노신; 와이오신; 메틸와이

오신; 저-변형된 하이드록시와이부토신; 와이부토신; 하이드록시와이부토신; 피옥시와이부토신; 케오신; 에폭시 케오신; 갈락토실-케오신; 만노실-케오신; 7-시아노-7-테아자구아노신; 아라케오신 [7-포름아미도-7-테아자구아노신으로도 지칭됨]; 및 7-아미노메틸-7-테아자구아노신.

[0053] 용어 "핵산" 및 "폴리뉴클레오타이드"는 RNA 및 DNA, 가령 cDNA, 유전체 DNA, 합성 DNA, 및 핵산 유사체를 내포한 DNA (또는 RNA)를 모두 지칭하기 위해 상호교환적으로 사용할 수 있고 이들 중 어느 것도 본 발명의 폴리펩티드를 인코딩할 수 있으며 상기 모두가 본 발명에 포함된다. 폴리뉴클레오타이드는 본질적으로 임의의 3-차원 구조를 가질 수 있다. 핵산은 이중-가닥 또는 단일-가닥 (즉, 센스 가닥 또는 안티센스 가닥)일 수 있다. 폴리뉴클레오타이드의 비-제한적인 예시는 유전자, 유전자 단편, 엑손, 인트론, 전령 RNA (mRNA) 및 이들의 부분, 운반 RNA, 리보솜 RNA, siRNA, 마이크로-RNA, 리보자임, cDNA, 재조합 폴리뉴클레오타이드, 분지형 폴리뉴클레오타이드, 플라스미드, 벡터, 임의 서열의 단리된 DNA, 임의 서열의 단리된 RNA, 핵산 탐침, 및 프라이머, 뿐만 아니라 핵산 유사체를 포함한다. 본 발명의 맥락에서, 핵산은 자연발생적 Cas9 또는 이의 생물학적으로 활성인 변이체 및 가이드 RNA의 단편을 인코딩할 수 있고 여기서 가이드 RNA는 HIV 내 서열에 상보적이다.

[0054] "단리된" 핵산은, 예를 들면, 자연-발생적 DNA 분자 또는 이의 단편일 수 있고, 적어도 하나의 핵산 서열이 일반적으로 자연-발생적 유전체 내 DNA 분자의 바로 옆에서 발견된다는 전제하에 제거되거나 부재한다. 따라서, 단리된 핵산은, 제한 없이, 다른 서열과 독립적으로 개별 분자로서 존재하는 DNA 분자 (예컨대, 화학적으로 합성된 핵산, 또는 중합효소 사슬 반응 (PCR) 또는 제한 엔도뉴클레아제 처리에 의해 제조된 cDNA 또는 유전체 DNA 단편)를 포함한다. 단리된 핵산은 또한 벡터, 독자적으로 복제하는 플라스미드, 바이러스에 통합되거나, 또는 원핵생물 또는 진핵생물의 유전체 DNA에 통합된 DNA 분자를 지칭한다. 또한, 단리된 핵산은 유전조작된 핵산 가령 하이브리드 또는 융합 핵산의 일부인 DNA 분자를 포함할 수 있다. 많은 (예컨대, 수십, 또는 수백 내지 수백 만) 다른 핵산 중에 존재하는, 예를 들면, cDNA 라이브러리 또는 유전체 라이브러리 내부의 핵산, 또는 유전체 DNA 제한 단편을 포함하는 겔 절편은 단리된 핵산이 아니다.

[0055] 단리된 핵산 분자는 표준 기술로 제조될 수 있다. 예를 들면, 중합효소 사슬 반응 (PCR) 기술을 사용하여 본 명세서에 기술된 뉴클레오타이드 서열, 가령 본 명세서에 기술된 폴리펩티드를 인코딩하는 뉴클레오타이드 서열을 내포하는 단리된 핵산을 얻을 수 있다. PCR은 총 유전체 DNA 또는 총 세포 RNA로부터의 서열을 포함하여, DNA 뿐만 아니라 RNA로부터의 특정한 서열을 증폭하기 위해 사용될 수 있다. 다양한 PCR 방법, 예를 들면, *PCR Primer: A Laboratory Manual*, Dieffenbach 및 Dveksler, eds., Cold Spring Harbor Laboratory Press, 1995에 기술되어 있다. 일반적으로, 관심 또는 그 이외의 부위의 말단으로부터의 서열 정보는 증폭시킬 주형의 반대 가닥에 대해 동일하거나 유사한 서열을 가진 올리고뉴클레오타이드 프라이머를 설계하기 위해 사용된다. 부위-특이적 뉴클레오타이드 서열 변형을 주형 핵산에 도입할 수 있는 다양한 PCR 전략이 또한 이용가능하다.

[0056] 단리된 핵산은 또한 단일 핵산 분자로서 (예컨대, 포스포라미디트 기술을 이용한 3'에서 5' 방향의 자동 DNA 합성을 이용하여) 또는 일련의 올리고뉴클레오타이드로서 화학적으로 합성될 수 있다. 예를 들면, 긴 올리고뉴클레오타이드 (예컨대, >50-100 뉴클레오타이드)의 하나 이상의 쌍은 요망되는 서열을 내포하도록 합성될 수 있고, 여기서 각각의 쌍은 올리고뉴클레오타이드 쌍이 합쳐질 때 이중복합체가 형성되도록 상보적인 짧은 조각 (예컨대, 약 15 뉴클레오타이드)를 내포한다. DNA 중합효소는 올리고뉴클레오타이드를 연장하기 위해 사용되어, 올리고뉴클레오타이드 쌍마다 나중에 벡터에 결합될 수 있는 단일, 이중-가닥 핵산 분자를 생성한다. 본 발명의 단리된 핵산은 또한 예컨대, Cas9-인코딩 DNA의 자연발생적 부분 (예를 들면, 상기 식에 따른 것)에 돌연변이를 유발시킴으로써 수득될 수 있다.

[0057] 이들이 인코딩하는 두 핵산 또는 폴리펩티드는 서로 특정 정도의 동일성을 가지도록 기술될 수 있다. 예를 들면, Cas9 단백질 및 이의 생물학적으로 활성인 변이체는 특정 정도의 동일성을 나타내도록 기술될 수 있다. 정렬은 PIR(Protein Information Research, 단백질 정보 검색) 웹사이트 (<http://pir.georgetown.edu>)에서 짧은 Cas9 서열을 배치하고, 이후 NCBI 웹사이트(<http://www.ncbi.nlm.nih.gov/blast>)에서 BLAST(Basic Local Alignment Search Tool, 유전자위치 정보 검색 도구) 알고리즘의 "짧은 거의 동일한 서열"을 이용하여 조립될 수 있다.

[0058] 본 명세서에서 사용된, 용어 "백분율 서열 동일성"은 주어진 임의의 검색 서열 및 대상 서열 사이의 동일성 정도를 가리킨다. 예를 들면, 자연발생적 Cas9은 검색 서열이 될 수 있고 Cas9 단백질의 단편이 대상 서열이 될 수 있다. 유사하게, Cas9 단백질의 단편이 검색 서열이 될 수 있고 이의 생물학적으로 활성인 변이체가 대상 서열이 될 수 있다.

[0059] 서열 동일성을 확인하기 위해, 검색 핵산 또는 아미노산 서열을 컴퓨터 프로그램 ClustalW (버전 1.83, 디폴트

변수)를 이용하여 각각 하나 이상의 개체 핵산 또는 아미노산 서열에 정렬시킬 수 있고, 상기 프로그램은 핵산 또는 단백질 서열을 이들의 전체 길이에 걸쳐 정렬할 수 있게 해준다(전반적 정렬). Chenna *et al*, *Nucleic Acids Res.* 31:3497-3500, 2003을 참조하라.

[0060] ClustalW는 검색 및 하나 이상의 개체 서열 사이의 최적의 대응을 산출하고 이를 정렬하여 동일성, 유사성 및 차이점이 확인될 수 있게 한다. 서열 정렬을 최대화하기 위해 하나 이상의 잔기의 간격이 검색 서열, 개체 서열, 또는 둘 다에 삽입될 수 있다. 핵산 서열의 빠른 쌍 방식의 정렬을 위해, 하기 디폴트 변수가 사용된다: 글자 크기: 2; 창 크기: 4; 점수 매김 방법: 백분율; 최대 항의 수: 4; 및 간격 페널티: 5. 핵산 서열의 다중 정렬을 위해, 하기 변수가 사용된다: 간격 개방 페널티: 10.0; 간격 연장 페널티: 5.0; 및 가중 전이: 있음. 단백질 서열의 빠른 쌍 방식의 정렬을 위해, 하기 변수가 사용된다: 글자 크기: 1; 창 크기: 5; 점수 매김 방법: 백분율; 최대 항의 수: 5; 및 간격 페널티: 3. 단백질 서열의 다중 정렬을 위해, 하기 변수가 사용된다: 가중 매트릭스: 블로섬(blosum); 간격 개방 페널티: 10.0; 간격 연장 페널티: 0.05; 친수성 간격: 사용; 친수성 잔기: Gly, Pro, Ser, Asn, Asp, Gin, Glu, Arg, 및 Lys; 잔기-특이적 간격 페널티: 사용. 출력 결과는 서열 사이의 관계를 반영하는 서열 정렬이다. ClustalW는 인터넷 예를 들면, Baylor 의대 검색 런처(Search Launcher) 사이트 (searchlauncher.bcm.tmc.edu/multi-align/multi-align.html) 및 유럽 생물정보연구소 사이트 (ebi.ac.uk/clustalw)에서 실행시킬 수 있다.

[0061] 검색 서열 및 개체 서열 사이의 백분율 동일성을 결정하기 위해, ClustalW는 최적의 정렬에서 동일성의 수를 비교된 잔기의 수로 나누고(간격 위치는 배제함), 결과에 100을 곱한다. 출력 결과는 검색 서열에 대한 개체 서열의 백분율 동일성이다. 백분율 동일성 수치는 소수 첫째 자리까지 반올림할 수 있음이 주의된다. 예를 들면, 78.11, 78.12, 78.13, 및 78.14는 78.1로 버림되는 반면, 78.15, 78.16, 78.17, 78.18, 및 78.19는 78.2로 올림된다.

[0062] 본 명세서에 기술된 핵산 및 폴리펩티드는 "외생성"으로 지칭될 수 있다. 용어 "외생성"은 핵산 또는 폴리펩티드가 재조합 핵산 구조체의 일부이거나 이에 의해 인코딩되고, 또는 이의 천연 환경이 아닌 것을 가리킨다. 예를 들면, 외생성 핵산은 또다른 종에 도입된 하나의 종으로부터의 서열, 즉, 이종 핵산일 수 있다. 전형적으로, 그러한 외생성 핵산은 재조합 핵산 구조체를 통해 다른 종으로 도입된다. 외생성 핵산은 또한 하나의 유기체에 게 자연적이며 유기체의 세포에 재도입된 서열일 수 있다. 자연 서열을 포함하는 외생성 핵산은 흔히 외생성 핵산에 연결된 비-자연적 서열의 존재, 예컨대, 재조합 핵산 구조체 내 자연 서열의 측면에 부착된 비-자연적 조절 서열에 의해 자연발생적 서열과 구분될 수 있다. 또한, 안정하게 형질전환된 외생성 핵산은 전형적으로 자연 서열이 발견되는 위치가 아닌 위치에 통합된다.

[0063] 재조합 구조체가 또한 본 명세서에 제공되며 세포가 Cas9 및/또는 HIV 내 표적 서열에 상보적인 가이드 RNA를 발현하도록 형질전환시키기 위해 사용될 수 있다. 재조합 핵산 구조체는 본 명세서에 기술된 바와 같이 세포에서 Cas9 및/또는 HIV 내 표적 서열에 상보적인 가이드 RNA를 발현하기에 적절한 조절 부위에 작동가능하도록 연결된, Cas9 및/또는 HIV 내 표적 서열에 상보적인 가이드 RNA를 인코딩하는 핵산을 포함한다. 수많은 핵산이 특정 아미노산 서열을 가지는 폴리펩티드를 인코딩할 수 있음이 이해될 것이다. 유전 암호의 퇴화는 당해 분야에 널리 공지되어 있다. 많은 아미노산에 있어서, 아미노산에 대한 코돈으로서 기능하는 뉴클레오타이드 삼중항은 둘 이상이 존재한다. 예를 들면, Cas9에 대한 암호화 서열 내 코돈은 특정 유기체에서 최적의 발현을 얻도록, 해당 유기체에게 적절한 코돈 바이어스(bias) 표를 이용하여 변형될 수 있다.

[0064] 본 명세서에 기술된 것과 같은, 핵산을 내포하는 벡터가 또한 제공된다. "벡터"는 또다른 DNA 조각이 삽입되어 삽입된 조각의 복제를 일으킬 수 있는 플라스미드, 파지, 또는 코스미드와 같은 복제단위이다. 일반적으로, 벡터는 적절한 조절 요소와 연합되었을 때 복제가 가능하다. 적절한 벡터 백본은, 예를 들면, 당해 분야에서 일상적으로 사용되는 것들 가령 플라스미드, 바이러스, 인공적인 염색체, BAC, YAC, 또는 PAC를 포함한다. 용어 "벡터"는 클로닝 및 발현 벡터, 뿐만 아니라 바이러스 벡터 및 통합 벡터를 포함한다. "발현 벡터"는 조절 부위를 포함하는 벡터이다. 다양한 숙주/발현 벡터 조합물이 본 명세서에 기술된 핵산 서열을 발현시키기 위해 사용될 수 있다. 적절한 발현 벡터는 제한없이, 예를 들면, 박테리오파지, 배큘로바이러스, 및 레트로바이러스에서 유래한 플라스미드 및 바이러스 벡터를 포함한다. 수많은 벡터 및 발현 시스템이 Novagen (Madison, WI), Clontech (Palo Alto, CA), Stratagene (La Jolla, CA), 및 Invitrogen/Life Technologies (Carlsbad, CA)와 같은 제조사로부터 구입가능하다.

[0065] 본 명세서에 제공된 벡터는 또한, 예를 들면, 복제원점, 스캐폴드 부착 부위 (SAR), 및/또는 마커를 포함할 수 있다. 마커 유전자는 숙주 세포에 선별가능한 표현형을 전달할 수 있다. 예를 들면, 마커는 살생물제 내성, 가

령 항생제 내성 (예컨대, 카나마이신, G418, 블레오마이신, 또는 히그로마이신)을 전달할 수 있다. 상기 언급된 바와 같이, 발현 벡터는 발현된 폴리펩티드의 조작 또는 검출(예컨대, 정제 또는 국제화)이 용이하도록 설계된 태그 서열을 포함할 수 있다. 태그 서열, 가령 녹색 형광 단백질 (GFP), 글루타티온 S-전이효소 (GST), 폴리히스티딘, c-myc, 헤마글루티닌, 또는 Flag™ 태그 (Kodak, New Haven, CT) 서열은 전형적으로 인코딩된 폴리펩티드와 융합되어 발현된다. 그러한 태그는 카복실 말단 또는 아미노 말단 중 어느 하나를 비롯하여, 폴리펩티드 내 어디든 삽입될 수 있다.

[0066] 추가적인 발현 벡터는 또한, 예를 들면, 염색체, 비-염색체 및 합성 DNA 서열의 조각을 포함할 수 있다. 적절한 벡터는 SV40 유도체 및 공지의 박테리아 플라스미드, 예컨대, *E. coli* 플라스미드 col E1, pCR1, pBR322, pMal-C2, pET, pGEX, pMB9 및 이들의 유도체, 플라스미드 가령 RP4; 파지 DNA, 예컨대, 파지 1의 수많은 유도체, 예컨대, NM989, 및 다른 파지 DNA, 예컨대, M13 및 섬유상 단일 가닥 파지 DNA; 효모 플라스미드 가령 2 μ 플라스미드 또는 이들의 유도체, 진핵 세포에서 유용한 벡터, 가령 곤충 또는 포유류 세포에서 유용한 벡터; 플라스미드 및 파지 DNA의 조합에서 유래된 벡터, 가령 파지 DNA 또는 다른 발현 조절 서열을 사용하도록 변형된 플라스미드를 포함한다.

[0067] 효모 발현 시스템이 또한 사용될 수 있다. 예를 들면, 두 가지만 언급하자면, 비-융합 pYES2 벡터 (XbaI, SphI, ShoII, NotI, GstXI, EcoRI, BstXI, BamHI, SacI, KpnI, 및 HindIII 클로닝 부위; Invitrogen) 또는 융합 pYESHisA, B, C (XbaI, SphI, ShoI, NotI, BstXI, EcoRI, BamHI, SacI, KpnI, 및 HindIII 클로닝 부위, ProBond 수지로 정제하고 엔테로키나제로 절단한 N-말단 펩티드; Invitrogen)가 본 발명에 따라 사용될 수 있다. 효모 두 가지-하이브리드 발현 시스템이 또한 본 발명에 따라 제조될 수 있다.

[0068] 벡터는 또한 조절 부위를 포함할 수 있다. 용어 "조절 부위"는 전사 또는 번역 개시 및 속도, 및 안정성 및/또는 전사물 또는 번역 산물의 이동성에 영향을 주는 뉴클레오티드 서열을 지칭한다. 조절 부위는, 제한없이, 프로모터 서열, 인핸서 서열, 반응 요소, 단백질 인식 부위, 유도 요소, 단백질 결합 서열, 5' 및 3' 미번역 부위 (UTR), 전사 시작 부위, 종말 서열, 폴리아데닐화 서열, 핵국제화신호, 및 인트론을 포함한다.

[0069] 본 명세서에서 사용된, 용어 "작동가능하도록 연결된"은 조절 부위 및 전사될 핵산 내 서열이 그러한 서열의 전사 또는 번역에 영향을 끼치도록 배치되는 것을 가리킨다. 예를 들면, 암호화 서열을 프로모터의 제어 아래 두기 위해, 폴리펩티드의 번역 관독 프레임의 번역 개시 부위는 전형적으로 프로모터의 일 및 약 오십 뉴클레오티드 하류에 위치한다. 프로모터는, 그러나, 번역 개시 부위의 약 5,000 뉴클레오티드 상류 또는 전사 시작 부위의 약 2,000 뉴클레오티드 상류만큼 위치할 수도 있다. 프로모터는 전형적으로 적어도 핵심 (기본) 프로모터를 포함한다. 프로모터는 또한 적어도 하나의 제어 요소, 가령 인핸서 서열, 상류 요소 또는 상류 활성화 부위 (UAR)를 포함할 수 있다. 포함될 프로모터의 선택은 효율성, 선택성, 유도성, 요망되는 발현 수준, 및 세포- 또는 조직-선별 발현을 포함하지만, 이에 제한되지 않는 여러가지 요인에 의존적이다. 적절히 선택 및 배치 프로모터 및 암호화 서열에 상대적인 다른 조절 부위를 적절히 선택하고 배치함으로써 암호화 서열의 발현을 조정하는 것은 당해 분야의 숙련자에게 일상적인 일이다.

[0070] 벡터는, 예를 들면, 바이러스 벡터 (가령 아데노바이러스 ("Ad"), 아데노-연관 바이러스 (AAV), 및 소포성 구내염 바이러스 (VSV) 및 레트로바이러스), 리포솜 및 다른 지질-함유 복합체, 및 숙주 세포에 폴리뉴클레오티드의 송달을 매개할 수 있는 다른 거대분자 복합체를 포함한다. 벡터는 또한 추가로 유전자 송달 및/또는 유전자 발현을 조정하거나, 그렇지 않으면 표적화 세포에게 유익한 특징을 제공하는 다른 성분 또는 관능기를 포함할 수 있다. 하기에 더 상세하게 기술되고 예시되는 바와 같이, 그러한 다른 성분은, 예를 들면, 세포에 대한 결합 또는 표적화에 영향을 주는 성분 (가령 세포-유형 또는 조직-특이적 결합을 매개하는 성분); 세포에 의한 벡터 핵산의 흡수에 영향을 주는 성분; 흡수 후 세포 내부에서 폴리뉴클레오티드의 국제화에 영향을 주는 성분 (가령 핵 국제화를 매개하는 물질); 및 폴리뉴클레오티드의 발현에 영향을 주는 성분을 포함한다. 그러한 성분은 또한 벡터를 흡수하고 벡터에 의해 송달되는 핵산을 발현하는 세포를 검출 또는 선별하기 위해 사용될 수 있는 검출 가능한 및/또는 선별가능한 마커와 같은 마커를 포함할 수 있다. 그러한 성분은 벡터의 자연적인 특징으로서 제공될 수 있고 (가령 결합 및 흡수를 매개하는 성분 또는 관능기를 가지는 특정한 바이러스 벡터의 이용), 또는 벡터는 그러한 관능기를 제공하도록 변형될 수 있다. 다른 벡터는 Chen *et al*; *BioTechniques*, 34: 167-171 (2003)에 기술된 것들을 포함한다. 아주 다양한 그러한 벡터가 당해 분야에 공지되어 있고 일반적으로 이용가능하다.

[0071] "재조합 바이러스 벡터"는 하나 이상의 이중 유전자 산물 또는 서열을 포함하는 바이러스 벡터를 가리킨다. 많은 바이러스 벡터가 패키징과 관련하여 크기-제한을 나타내기 때문에, 이중 유전자 산물 또는 서열은 전형적으

로 하나 이상의 분량의 바이러스 유전체를 대체하면서 도입된다. 그러한 바이러스는 복제-결손이 되어, 결실된 기능(들)을 바이러스 복제 및 단백질막 생성 도중에 교차 제공받아야 할 수 있다 (예컨대, 복제 및/또는 단백질막 생성을 위해 필요한 유전자 산물을 지니는 보조 바이러스 또는 패키징 세포주를 이용함). 전달해야 할 폴리뉴클레오티드를 바이러스 입자의 외부에 지니는 변형된 바이러스 벡터가 또한 기술된 바 있다(예컨대, Curriel, D T, *et al.* *PNAS* 88: 8850-8854, 1991를 참조).

[0072] 적절한 핵산 송달 시스템은 재조합 바이러스 벡터, 전형적으로는 아데노바이러스, 아데노바이러스-연관 바이러스 (AAV), 보조-의존성 아데노바이러스, 레트로바이러스, 또는 일본 혈구응집 바이러스-리포좀 (HVJ) 복합체 중 적어도 하나로부터의 서열을 포함한다. 그러한 경우, 바이러스 벡터는 폴리뉴클레오티드에 작동가능하도록 연결된 강력한 진핵세포 프로모터 예컨대, 거대세포바이러스 (CMV) 프로모터를 포함한다. 재조합 바이러스 벡터는 내부에 하나 이상의 폴리뉴클레오티드, 바람직하게는 약 하나의 폴리뉴클레오티드를 포함할 수 있다. 일부 구체예에서, 본 발명의 방법에서 사용되는 바이러스 벡터는 약 10^8 의 pfu (플라그 형성 단위) 내지 약 5×10^{10} pfu를 가진다. 폴리뉴클레오티드가 비-바이러스 벡터와 함께 투여되는 구체예에서, 약 0.1 나노그램 내지 약 4000 마이크로그램의 사용 예컨대, 약 1 나노그램 내지 약 100 마이크로그램의 사용이 흔히 유용할 것이다.

[0073] 추가적인 벡터는 바이러스 벡터, 융합 단백질 및 화학 접합체를 포함한다. 레트로바이러스 벡터는 몰로니 (Moloney) 쥐 백혈병 바이러스 및 HIV-계 바이러스를 포함한다. 하나의 HIV-계 바이러스 벡터는 적어도 두 가지 벡터를 포함하며 여기서 gag 및 pol 유전자는 HIV 유전체에서 유래하고 env 유전자는 또다른 바이러스에서 유래한다. DNA 바이러스 벡터는 포스(pox) 벡터 가령 오르소포스 또는 아비포스 벡터, 헤르페스바이러스 벡터 가령 I형 단순 헤르페스 바이러스 (HSV) 벡터 [Geller, A.I. *et al.*, *J. Neurochem*, 64: 487 (1995); Lim, F., *et al.*, *DNA Cloning: Mammalian Systems*, D. Glover, Ed. (Oxford Univ. Press, Oxford England) (1995) 수록; Geller, A.I. *et al.*, *Proc Natl. Acad. Sci.* U.S.A.:90 7603 (1993); Geller, A.I., *et al.*, *Proc Natl. Acad. Sci.* USA: 87: 1149 (1990)], 아데노바이러스 벡터 [LeGal LaSalle *et al.*, *Science*, 259:988 (1993); Davidson, *et al.*, *Nat. Genet.* 3: 219 (1993); Yang, *et al.*, *J. Virol.* 69: 2004 (1995)] 및 아데노-연관 바이러스 벡터 [Kaplitt, M.G., *et al.*, *Nat. Genet.* 8: 148 (1994)]를 포함한다.

[0074] 포스 바이러스 벡터는 유전자를 세포의 세포질에 도입한다. 아비포스 바이러스 벡터는 오로지 핵산의 단기 발현을 야기한다. 아데노바이러스 벡터, 아데노-연관 바이러스 벡터 및 단순 헤르페스 바이러스 (HSV) 벡터는 일부 발명의 구체예를 위해 지정될 수 있다. 아데노바이러스 벡터는 아데노-연관 바이러스보다 더 짧은 단기 발현 (예컨대, 약 한 달 미만)을 야기하며, 일부 구체예에서, 훨씬 더 긴 발현을 나타낼 수 있다. 선택되는 특정한 벡터는 표적 세포 및 치료되는 조건에 의존적일 것이다. 적절한 프로모터의 선택은 쉽게 달성될 수 있다. 적절한 프로모터의 예시는 763-염기-쌍 거대세포바이러스 (CMV) 프로모터이다. 유전자 발현을 위해 사용될 수 있는 다른 적절한 프로모터는, Rous 육종 바이러스 (RSV) (Davis, *et al.*, *Hum Gene Ther* 4: 151 (1993)), SV40 초기 프로모터 부위, 헤르페스 티미딘 키나아제 프로모터, 메탈로티오네인 (MMT) 유전자 조절 서열, 원핵세포 발현 벡터 가령 β -락타마제 프로모터, tac 프로모터, 효모 또는 다른 균류로부터의 프로모터 요소 가령 Gal 4 프로모터, ADC (알코올 탈수소화효소) 프로모터, PGK (포스포글리세롤 키나아제) 프로모터, 알칼리성 포스파타제 프로모터; 및 조직 특이성을 나타내고 형질전환 동물에서 이용되는 동물 전사 제어 부위: 췌장 소엽 세포에서 활성인 엘라스타제 I 유전자 제어 부위, 췌장 베타 세포에서 활성인 인슐린 유전자 제어 부위, 림프양 세포에서 활성인 면역글로불린 유전자 제어 부위, 고환, 유방, 림프양 및 비만세포에서 활성인 마우스 유방 종양 바이러스 제어 부위, 간에서 활성인 알부민 유전자 제어 부위, 간에서 활성인 알파-태아단백질 유전자 제어 부위, 간에서 활성인 알파 1-안티트립신 유전자 제어 부위, 골수 세포에서 활성인 베타-글로빈 유전자 제어 부위, 뇌의 회소돌기아교세포 세포에서 활성인 미엘린 염기성 단백질 유전자 제어 부위, 골격 근육에서 활성인 미오신 경사슬-2 유전자 제어 부위, 및 시상하부에서 활성인 생식선자극 방출 호르몬 유전자 제어 부위를 포함하지만 이에 제한되지 않는다. 특정 단백질은 이들의 자연적인 프로모터를 이용하여 발현될 수 있다. 발현을 촉진할 수 있는 다른 요소 가령 인핸서 또는 tat 유전자 및 tar 요소와 같은 높은 수준의 발현을 야기하는 시스템이 또한 포함될 수 있다. 이러한 카세트는 이후 벡터, 예컨대, 플라스미드 벡터 가령, pUC19, pUC118, pBR322, 또는 다른 공지의 플라스미드 벡터, 예를 들면, *E. coli*의 복제원점을 포함하는 벡터에 삽입될 수 있다. Sambrook *et al.*, *Molecular cloning: A laboratory manual*, Cold Spring Harbor Laboratory Press, (1989)를 참조하라. 플라스미드 벡터는 또한 마커 폴리펩티드가 처리되는 유기체의 대사에 유해하게 영향끼치지 않는다는 전제하에, 선별 가능한 마커 가령 암피실린 내성에 대한 β -락타마제 유전자를 포함할 수 있다. 카세트는 또한 합성 송달 시스템, 가령 WO 95/22618에 개시된 시스템 내 핵산 결합 모이어티에 결합될 수 있다.

[0075] 요망되는 경우, 본 발명의 폴리뉴클레오티드는 또한 미세수송 비히클 가령 양이온성 리포좀 및 아데노바이러스

백터와 함께 사용될 수 있다. 리포솜 제조, 표적화 및 내용물의 송달에 관한 절차를 알아보려면, Mannino 및 Gould-Fogerite, *BioTechniques*, 6:682 (1988)를 참조하라. 또한, Felgner 및 Holm, Bethesda Res. Lab. Focus, 11(2):21 (1989) 및 Maurer, R.A., Bethesda Res. Lab. Focus, 11(2):25 (1989)를 참조하라.

[0076] 복제-결손 재조합 아데노바이러스 백터는 공지된 기술에 따라 제조될 수 있다. Quantin, *et al*, *Proc. Natl. Acad. Sci. USA*, 89:2581-2584 (1992); Stratford-Perricadet, *et al.*, *J. Clin. Invest.*, 90:626-630 (1992); 및 Rosenfeld, *et al.*, *Cell*, 68: 143-155 (1992)를 참조하라.

[0077] 또다른 송달 방법은 세포내에서 발현된 산물을 제조할 수 있는 백터를 생산하는 단일 가닥 DNA를 사용하는 것이다. 예를 들면, Chen *et al*, *BioTechniques*, 34: 167-171 (2003)를 참조, 상기 문헌은 본 명세서에서 그 전체가 참고로서 포함된다.

[0078] **약제학적 조성물**

[0079] 상기 기술된 바와 같이, 본 발명의 조성물은 당해 분야의 숙련자에게 공지된 다양한 방식으로 제조될 수 있다. 조성물이 수득되는 본래 출처 또는 방식과 관계없이, 본 발명의 조성물은 이들의 용도에 따라 제형화될 수 있다. 예를 들면, 상기 기술된 핵산 및 백터는 조직 배양물 내 세포에 적용하기 위해 또는 환자 또는 개체에게 투여하기 위해 조성물 내에 제형화될 수 있다. 본 발명의 임의의 약제학적 조성물은 의학의 제조에서 사용하기 위해 제형화될 수 있고, 특정한 용도가 하기에서 치료, 예컨대, HIV 감염증을 가지거나 또는 HIV 감염증에 걸릴 위험이 있는 개체의 치료의 맥락에서 나타난다. 약물로 사용되는 경우, 임의의 핵산 및 백터는 약제학적 조성물의 형태로 투여될 수 있다. 이들 조성물은 약제학 분야에 널리 공지된 방식으로 제조될 수 있고, 국부 또는 전신 치료의 요망 여부 및 처리되는 부분에 따라, 다양한 경로에 의해 투여될 수 있다. 투여는 국소 (가령 눈 및 비강내, 질 및 직장 송달을 비롯한 점액성 막), 폐 (예컨대, 분무기에 의한 것을 비롯한 분말 또는 에어로졸의 흡입 또는 통기에 의해; 기관내, 비강내, 상피 및 경피), 안내, 경구 또는 비경구일 수 있다. 안내 송달을 위한 방법은 국소 투여 (점안액), 결막하, 안주위 또는 유리체내 주입 또는 풍선 카테터 또는 결막하에 수술로 배치한 안내 삽입물에 의한 도입을 포함할 수 있다. 비경구 투여는 정맥내, 동맥내, 피하, 복강내 또는 근육내 주사 또는 주입; 또는 두개내, 예컨대, 척추강내 또는 심실내 투여를 포함한다. 비경구 투여는 단일 볼루스(bolus) 투여의 형태일 수 있거나, 또는 예를 들면, 연속 관류 펌프에 의한 것일 수 있다. 국소 투여를 위한 약제학적 조성물 및 제형은 경피 패치, 연고, 로션, 크림, 젤, 점적액, 좌제, 분사제, 액체, 분말, 등을 포함할 수 있다. 통상적인 약제학적 담체인, 수용성, 분말 또는 유성 베이스, 증점제 등이 필요하거나 요망될 수 있다.

[0080] 본 발명은 또한 활성 성분으로서, 본 명세서에 기술된 핵산 및 백터를 하나 이상의 약제학적으로 허용되는 담체와 조합하여 함유하는 약제학적 조성물을 포함한다. 용어 "약제학적으로 허용되는" (또는 "약리학적으로 허용되는")은 동물 또는 인간에게, 적절히 투여될 때 부작용, 알러지, 또는 다른 원치않는 반응을 일으키지 않는 분자 독립체 및 조성물을 가리키기 위해 사용된다. 본 명세서에서 사용된 용어 "약제학적으로 허용되는 담체"는 약제학적으로 허용되는 물질을 위한 매체로서 사용될 수 있는 임의의 및 모든 용매, 분산 매체, 코팅제, 항생제, 등장화제 및 흡수지연제, 완충제, 부형제, 결합제, 활택제, 젤, 계면활성제 등을 포함한다. 본 발명의 조성물의 제조에서, 활성 성분은 전형적으로 부형제와 혼합되거나, 부형제로 희석되거나 또는 예를 들면, 캡셀, 정제, 사세(sachet), 종이, 또는 다른 저장소의 형태로 된 담체 내에 봉입된다. 부형제가 희석제로서 기능하는 경우, 부형제는 고체, 반고체, 또는 액체 물질 (예컨대, 일반적인 식염수)일 수 있고, 활성 성분을 위한 비히클, 담체 또는 매체로서 작용한다. 따라서, 조성물은 정제, 알약, 분말, 로젠지(lozenge), 사세, 카세(cachet), 엘릭서, 현탁액, 에멀전, 용액, 시럽, 에어로졸 (고체로서 또는 액체 매체 내), 로션, 크림, 연고, 젤, 연질 및 경질 젤라틴 캡셀, 좌제, 무균 주사용 용액, 및 무균 포장된 분말의 형태일 수 있다. 당해 분야에 공지된 바와 같이, 희석제의 유형은 의도되는 투여의 경로에 따라 달라질 수 있다. 생성된 조성물은 추가적인 물질, 가령 보존제를 포함할 수 있다. 일부 구체예에서, 담체는 지질-기반 또는 중합체-기반 콜로이드일 수 있거나 이들을 포함할 수 있다. 일부 구체예에서, 담체 물질은 리포솜, 하이드로겔, 미소입자, 나노입자, 또는 블록 공중합체 미셀(micelle)로서 제형화된 콜로이드일 수 있다. 언급한 바와 같이, 담체 물질은 캡셀을 형성할 수 있고, 해당 물질은 중합체-기반 콜로이드일 수 있다.

[0081] 본 발명의 핵산 서열은 개체의 적절한 세포에 송달될 수 있다. 이는, 예를 들면, 식세포 가령 대식 세포에 의한 식세포작용에 최적화된 크기의 중합체, 생분해가능한 미소입자 또는 마이크로캡슐 송달 비히클의 사용에 의해 성취될 수 있다. 예를 들면, 대략 1-10 μm 직경의 PLGA (폴리-락토-코-글리콜리드) 미소입자가 사용될 수 있다. 폴리뉴클레오티드는 이들 미소입자에 캡슐화되고, 이는 대식 세포에 의해 흡수되고 세포 내에서 점차 생분해되어, 이를 통해 폴리뉴클레오티드를 방출한다. 일단 방출되면, DNA는 세포 내에 발현된다. 두 번째 유형의

미소입자는 세포에 의해 직접 흡수되는 대신, 처음에는 생분해를 통해 미소-입자로부터 방출되었을 때에만 세포에 의해 흡수되는 핵산의 서방출 저장소로서 기능하도록 의도된다. 이들 중합체 입자는 그러므로 식세포작용이 불가능하도록 충분히 더 커야 한다(즉, 5 μm 초과 및 바람직하게는 20 μm 초과). 핵산의 흡수를 성취하기 위한 또다른 방법은 표준 방법에 의해 제조된, 리포솜을 이용하는 것이다. 핵산은 이들 송달 비히클에 단독으로 포함되거나 조직-특이적 항체, 예를 들면 HIV 감염에서 흔히 잠복성으로 감염되는 저장소인 세포 유형, 예를 들면, 뇌 대식 세포, 미소아교세포, 성상아교세포, 및 장관-연관 림프양 세포를 표적하는 항체와 동시-포함될 수 있다. 대안적으로, 정전 또는 공유 힘에 의해 폴리-L-리신에 부착된 플라스미드 또는 다른 벡터로 이루어진 분자 복합체를 제조할 수 있다. 폴리-L-리신은 표적 세포 상의 수용체에 결합할 수 있는 리간드에 결합한다. 근육 내, 피내, 또는 피하 부위로 "노출된 DNA" (즉, 송달 비히클이 없는 것)의 송달은 생체 내 발현을 얻기 위한 또다른 송달 수단이다. 관련된 폴리뉴클레오타이드 (예컨대, 발현 벡터)에서 CRISPR-연관 엔도뉴클레아제 및 가이드 RNA를 인코딩하는 서열을 포함하는 단리된 핵산 서열을 인코딩하는 핵산 서열은 프로모터 또는 인핸서-프로모터 조합에 작동가능하게 연결된다. 프로모터 및 인핸서는 위에서 기술되었다.

[0082] 일부 구체예에서, 본 발명의 조성물은 나노 입자, 예를 들면, DNA를 가지고 폴리에틸렌글리콜-변형된 (폐길화) 저분자량 LPEI의 겹질에 둘러싸인 고분자량 선형 폴리에틸렌이민 (LPEI) 복합체의 본체로 이루어진 나노입자로서 제형화될 수 있다.

[0083] 핵산 및 벡터는 또한 장치 (예컨대, 카테터)의 표면에 도포되거나 펌프, 패치, 또는 다른 약물 송달 장치 내에 내포될 수 있다. 본 발명의 핵산 및 벡터는 약제학적으로 허용되는 부형제 또는 담체 (예컨대, 생리학적 식염수)의 존재에서 단독으로, 또는 조합으로 투여될 수 있다. 부형제 또는 담체는 투여의 방식 및 경로를 기준으로 선택된다. 약제학적 제형에서 사용하기 위한 적절한 약제학적 담체, 뿐만 아니라 약제학적 필수품이 본 발명의 분야에 널리-공지된 참고 서적인 Remington's Pharmaceutical Sciences (E. W. Martin), 및 USP/NF (미국 약전 및 처방집, United States Pharmacopeia and the National Formulary)에 기술되어 있다.

[0084] 일부 구체예에서, 조성물은 HIV의 성적 전염을 방지하기 위한 국소 겔로서 제형화될 수 있다. 국소 겔은 성적 행위 전 남성 또는 여성의 생식기 부위의 피부 또는 점액성 막에 직접 도포될 수 있다. 대안적으로 또는 부가적으로 국소 겔은 남성용 또는 여성용 콘돔 또는 페서리(diaphragm)의 표면에 도포되거나 내포될 수 있다.

[0085] 일부 구체예에서, 조성물은 Cas9 또는 변이체 Cas9 및 표적 HIV에 상보적인 가이드 RNA 서열을 인코딩하는 핵산 또는 Cas9 및 표적 HIV에 상보적인 가이드 RNA 서열을 인코딩하는 핵산을 포함하는 벡터에 피막을 형성하는 나노입자로서 제형화될 수 있다. 대안적으로, 조성물은 CRISPR-연관 엔도뉴클레아제 폴리펩티드, 예컨대, Cas9 또는 변이체 Cas9 및 표적에 상보적인 가이드 RNA 서열에 피막을 형성하는 나노입자로서 제형화될 수 있다.

[0086] 본 발명의 제형은 Cas9 및 표적 HIV에 상보적인 가이드 RNA 서열을 인코딩하는 벡터를 둘러쌀 수 있다. 가이드 RNA 서열은 단일 부위, 예컨대 LTR A, B, C, 또는 D에 상보적인 서열을 포함할 수 있거나 LTR A, B, C, 및 D에 상보적인 서열의 임의의 조합을 포함할 수 있다. 대안적으로 Cas9를 인코딩하는 서열 및 가이드 RNA 서열을 인코딩하는 서열은 별도의 벡터에 존재할 수 있다.

[0087] 치료 방법

[0088] 본 명세서에 개시된 조성물은 레트로바이러스 감염, 예컨대, HIV 감염증을 가지는 개체의 치료에 있어서 일반적으로 및 다양하게 유용하다. 개체, 환자, 또는 개인은 상호교환적으로 지칭될 수 있다. 본 방법은 임의의 HIV, 예를 들면, HIV-1, HIV-2 및 이들의 임의의 순환하는 재조합 형태를 표적하는데 유용하다. 임상적으로 유의한 결과가 이어지는 경우 개체는 효과적으로 치료된다. 이는 예를 들면, 질환 증상의 완전한 해소, 질환 증상의 중증도 감소, 또는 질환 진행의 지연을 의미할 수 있다. 이들 방법은 a) HIV 감염증을 가진 개체(예컨대, 환자 및, 더 특정하게는, 인간 환자)를 확인하는 단계; 및 b) 개체에게 CRISPR-연관 뉴클레아제, 예컨대, Cas9, 및 HIV 표적 서열, 예컨대 HIV LTR에 상보적인 가이드 RNA를 인코딩하는 핵산을 포함하는 조성물을 제공하는 단계를 추가로 포함할 수 있다. 개체는 개체의 혈청에서 HIV 항체 또는 HIV 폴리펩티드 p24의 존재를 검출하기 위한 표준 임상 시험, 예를 들면, 면역어세이를 이용하거나, HIV 핵산 증폭 어세이를 통해 확인될 수 있다. 개체에게 제공되어 감염 증상의 완전한 해소, 감염 증상의 중증도 감소, 또는 감염 진행의 지연을 야기하는 그러한 조성물의 양은 치료적으로 효과적인 양으로 간주된다. 본 발명의 방법은 또한 투여량 및 투여 스케줄을 최적화하는 것을 돕고 결과를 예측하기 위한 관찰 단계를 포함할 수 있다. 본 발명의 일부 방법에 있어서, 먼저 환자가 잠복성 HIV-1 감염을 가졌는지 확인하고, 이후 환자를 본 명세서에 기술된 하나 이상의 조성물로 치료할 것인지 여부를 결정할 수 있다. 관찰하는 단계는 또한 약물 저항성의 발생을 검출하고 응답하는 환자를 응답하지 않는 환자와 빠르게 구분하기 위해 사용될 수 있다. 일부 구체예에서, 본 방법은 추가로 환자에 잠복하는 특정 HIV의

핵산 서열을 검사하는 단계 및 이후 이러한 특정 서열에 상보적인 가이드 RNA를 설계하는 단계를 포함할 수 있다. 예를 들면, 개체의 LTR U3, R 또는 U5 부위의 핵산 서열을 확인하고 이후 환자의 서열에 정확히 상보적이라도 록 하나 이상의 가이드 RNA를 설계할 수 있다.

[0089] 조성물은 또한 레트로바이러스 감염, 예컨대, HIV 감염증을 가질 위험이 있는 개체의 치료, 예를 들면, 예방적 치료에 유용하다. 이들 방법은 a) HIV 감염증을 가질 위험이 있는 개체를 확인하는 단계; b) 개체에게 CRISPR-연관 뉴클레아제, 예컨대, Cas9, 및 HIV 표적 서열, 예컨대 HIV LTR에 상보적인 가이드 RNA를 인코딩하는 핵산을 포함하는 조성물을 제공하는 단계를 추가로 포함할 수 있다. HIV 감염증을 가질 위험이 있는 개체는, 예를 들면, 복수의 파트너와 보호되지 않은 즉, 콘돔을 사용하지 않고 성행위를 하는 임의의 개체; 또다른 성적으로 전염되는 감염증을 가진 복수의 파트너와 성행위를 하는 개체; 정맥 주사 마약 사용자; 또는 포경수술을 받지 않은 남성일 수 있다. HIV 감염증을 가질 위험이 있는 개체는, 예를 들면, 직업상 HIV-감염된 집단과 접촉할 수 있는 개체, 예컨대, 보건 의료 종사자 또는 최초응급대응자일 수 있다. HIV 감염증을 가질 위험이 있는 개체는, 예를 들면, 범죄 교정 시설의 수감자 또는 성 노동자, 즉 노동 대가 또는 음식, 약품, 또는 숙소와 같은 비화폐적인 대가를 위해 성적인 행위를 이용하는 개체일 수 있다.

[0090] 조성물은 또한 모친에게서 자녀로 HIV가 전염될 가능성을 낮추기 위해 HIV 감염증을 가지는 임산부 또는 수유부에게 투여될 수 있다. HIV에 감염된 임산부는 자궁에서 태반을 통해, 출산시 산도를 통해 또는 출산후 모유 수유를 통해 자녀에게 바이러스를 옮길 수 있다. 본 명세서에 개시된 조성물은 산전기, 주산기 또는 산후 수유기 중 어느 한 기간에 HIV에 감염된 모친에게 투여될 수 있고, 또는 산전기, 주산기, 및 산후기의 복합적인 투여가 이루어질 수 있다. 조성물은 하기 기술된 바와 같은 표준 항레트로바이러스 요법에 따라 모친에게 투여될 수 있다. 일부 구체예에서, 본 발명의 조성물은 또한 출산 직후 즉시 및, 일부 구체예에서, 출산 후 기간에 신생아에게 투여된다. 신생아는 또한 표준 항레트로바이러스 요법을 받을 수 있다.

[0091] 본 명세서에 개시된 방법 및 조성물은 레트로바이러스 감염증의 치료에 유용하다. 예시적인 레트로바이러스는 인간 면역결핍 바이러스, 예컨대 HIV-1, HIV-2; 원숭이 면역결핍 바이러스 (SIV); 고양이 면역결핍 바이러스 (FIV); 소의 면역결핍 바이러스 (BIV); 말 감염성 빈혈 바이러스 (EIAV); 및 염소 관절염/뇌염 바이러스 (CAEV)를 포함한다. 본 명세서에 개시된 방법은 광범위한 종, 예컨대, 인간, 인간이-아닌 영장류 (예컨대, 원숭이), 말 또는 다른 가축, 개, 고양이, 페럿 또는 애완용으로 키워지는 다른 포유동물, 래트, 마우스, 또는 다른 실험실용 동물에 투여될 수 있다.

[0092] 본 발명의 방법은 의약의 제조 측면에서 표현될 수 있다. 따라서, 본 발명은 의약의 제조에서 본 명세서에 기술된 물질 및 조성물의 사용을 포괄한다. 본 명세서에 기술된 화합물은 치료 조성물 및 계획에서 또는 본 명세서에 기술된 바와 같은 질환 또는 용태의 치료에 사용하기 위한 의약의 제조에 있어서 유용하다.

[0093] 본 명세서에 기술된 임의의 조성물은 표적 세포로의 추후 송달을 위해 숙주의 신체 중 임의의 부위에 투여될 수 있다. 조성물은 제한없이, 포유동물의 뇌, 뇌척수액, 관절, 비강점막, 혈액, 폐, 장, 근육 조직, 피부, 또는 복막강에 전달될 수 있다. 송달 경로의 관점에서, 조성물은 정맥내, 두개내, 복강내, 근육내, 피하, 근내, 직장내, 질내, 척추강내, 기관내, 피내, 또는 경피 주입에 의해, 경구 또는 비강 투여에 의해, 또는 시간에 따른 점진적인 관류에 의해 투여될 수 있다. 추가의 예시에서, 조성물의 에어로졸 제제가 흡입에 의해 숙주에게 제공될 수 있다.

[0094] 요구되는 투여량은 투여 경로, 제형의 성질, 환자 질병의 성질, 환자의 체형, 체중, 체표면적, 연령, 및 성별, 투여되는 다른 약물, 및 진료하는 임상치의 판단에 따라 달라질 것이다. 세포 표적의 다양성 및 다양한 투여 경로의 상이한 효율의 측면에서 필요한 투여량의 다양한 변조가 예상된다. 이러한 투여량 수준의 변조는 당해 분야에서 숙지되는 바와 같이, 최적화를 위한 표준 실증 루틴을 이용하여 조정될 수 있다. 투여는 단일 분량 또는 다회 분량일 수 있다 (예컨대, 2- 또는 3-, 4-, 6-, 8-, 10-, 20-, 50-, 100-, 150-배, 또는 그 이상). 적절한 송달 비히클 (예컨대, 중합체 미소입자 또는 이식가능한 장치) 내 화합물의 봉입은 송달 효율을 증가시킬 수 있다.

[0095] 본 명세서에 제공된 임의의 조성물을 이용한 치료 기간은 하루 만큼 짧은 시간부터 숙주의 평생 만큼 긴 시간까지 (예컨대, 수 년) 임의의 길이일 수 있다. 예를 들면, 화합물은 주당 한 회(예를 들면, 4 주 내지 수 개월 또는 년간); 개월당 한 회(예를 들면, 3 내지 12개월간 또는 수년 간); 또는 5년, 10년, 또는 그 이상의 기간 동안 연간 한 회 투여될 수 있다. 치료 빈도가 변할 수 있음이 또한 주의된다. 예를 들면, 본 발명의 화합물은 매일, 매주, 매달, 또는 매년 한 회(또는 두 차례, 세 차례 등) 투여될 수 있다.

- [0096] 본 명세서에 제공된 임의의 조성물의 효과적인 양이 치료가 필요한 개체에게 투여될 수 있다. 본 명세서에서 사용된 용어 "효과적인"은 환자에서 요망되는 반응을 유도하는 반면 상당한 독성을 유발하지 않는 임의의 양을 지칭한다. 그러한 양은 확인된 양의 특정 조성물의 투여 후에 환자의 반응을 평가함으로써 결정될 수 있다. 또한, 독성이 존재하는 경우, 독성의 수준은 확인된 양의 특정 조성물의 투여 전후에 환자의 임상적인 증상을 평가함으로써 결정될 수 있다. 환자에게 투여되는 특정 조성물의 효과적인 양은 요망되는 결과 뿐만 아니라 환자의 반응 및 독성 수준에 따라 조정될 수 있다. 상당한 독성은 각각의 특정 환자에서 다를 수 있고 제한없이, 환자의 질환 상태, 연령, 및 부작용에 대한 저항성을 비롯한 복수의 요인에 의존적이다.
- [0097] 특정한 반응이 유발되는지 여부를 확인하기 위해 당해 분야의 숙련가에게 공지된 임의의 방법이 사용될 수 있다. 반응이 유발되는지 확인하기 위해 특정한 질환 상태의 정도를 평가할 수 있는 임상적인 방법이 사용될 수 있다. 반응을 평가하기 위해 사용되는 특정 방법은 환자의 장애의 성질, 환자의 연령, 및 성별, 투여되는 다른 약물, 및 진료하는 임상의의 판단에 따라 달라질 것이다.
- [0098] 조성물은 또한 또다른 치료제, 예를 들면, HAART에서 사용되는 항-레트로바이러스제와 함께 투여될 수 있다. 예시적인 항레트로바이러스제는 역전사효소 저해제 (예컨대, 뉴클레오사이드/뉴클레오타이드 역전사효소 저해제, 지도부딘(zidovudine), 엠트리시타빈(emtricitabine), 라미부딘(lamivudine) 및 테노피비어(tenofivir); 및 비-뉴클레오사이드 역전사효소 저해제 가령 에파바렌즈(efavarens), 네비라핀(nevirapine), 릴피비린(rilpivirine)); 프로테아제 저해제, 예컨대, 티피라비어(tipiravir), 다루나비어(darunavir), 인디나비어(indinavir); 침입 저해제, 예컨대, 마라비록(maraviroc); 융합 저해제, 예컨대, 엔푸비리티드(enfuviritide); 또는 인테그라제 저해제 예컨대, 랄테그리비어(raltegravir), 도루테그라비어(dolutegravir)를 포함한다. 예시적인 항레트로바이러스제는 또한 다중-단계 복합제 예를 들면, 엠트리시타빈(emtricitabine), 에파바렌즈, 및 테노피비어의 조합; 엠트리시타빈; 릴피비린, 및 테노피비어의 조합; 또는 엘비테그라비어(elvitegravir), 코비시스탯(cobicistat), 엠트리시타빈 및 테노피비어의 조합을 포함할 수 있다.
- [0099] 둘 이상의 치료제의 동시적인 투여는 치료제가 치료 효과를 발생시키는 시간이 겹치는 동안은, 치료제를 동일한 시간 또는 동일한 경로로 투여할 필요가 없다. 투여가 서로다른 날짜 또는 주에 이루어지는 것처럼, 동시적 또는 연속적인 투여가 고려된다. 치료제는 규칙적인 계획, 예컨대, 치료제의 연속적인 낮은-투여 하에 투여될 수 있다.
- [0100] 그러한 조성물의 투여량, 독성 및 치료적 효능은 세포 배양 또는 실험 동물의 표준 약학 절차, 예컨대, LD₅₀ (집단 중 50%의 치사량) 및 ED₅₀ (집단 중 50%의 치료적 으로 효과적인 투여량)을 측정하기 위한 절차에 의해 결정될 수 있다. 독성 및 치료 효과 사이의 용량 비율이 치료 지수이며 이는 비율 LD₅₀/ED₅₀로 표현될 수 있다.
- [0101] 세포 배양 어세이 및 동물 연구로부터 얻은 데이터가 인간에서 사용을 위한 투여량 범위를 구성하는데 사용될 수 있다. 그러한 조성물의 투여는 바람직하게는 독성이 없거나 거의 없는 ED₅₀를 포함하는 순환 농도의 범위 내에 존재한다. 투여량은 활용되는 투여 형태 및 사용되는 투여 경로에 따라 상기 범위 내에서 달라질 수 있다. 본 발명의 방법에서 사용되는 임의의 조성물에 있어서, 치료적으로 효과적인 투여량은 처음에는 세포 배양 어세이로부터 추정될 수 있다. 투여량은 세포 배양에서 측정된 바와 같은 IC₅₀ (즉, 증상의 최대저해-절반을 달성하는 시험 화합물의 농도)을 포함하는 순환하는 혈장 농도 범위를 동물 모델에서 얻도록 형성될 수 있다. 그러한 정보는 인간에서 유용한 투여량을 더욱 정확하게 결정하기 위해 사용될 수 있다. 혈장 내 수준은, 예를 들면, 고성능 액체 크로마토그래피에 의해 측정될 수 있다.
- [0102] 기술된 바와 같이, 치료적으로 효과적인 양의 조성물 (즉, 효과적인 투여량)은 치료적으로 (예컨대, 임상적으로) 요망되는 결과를 생성하기에 충분한 양을 의미한다. 조성물은 하루 한 회부터 격일에 한 회를 포함하여 한주에 한 회까지 투여될 수 있다. 당해 분야의 숙련가는 질환 또는 장애의 중증도, 기존 치료, 개체의 일반적인 건강 및/또는 연령, 및 존재하는 다른 질환을 포함하지만, 이에 제한되지 않는 특정 요인이 개체를 효과적으로 치료하기 위해 요구되는 투여 및 투여 시간에 영향을 줄 수 있음을 이해할 것이다. 게다가, 치료적으로 효과적인 양의 본 발명의 조성물로 개체를 치료하는 것은 단일 치료 또는 일련의 치료를 포함할 수 있다.
- [0103] 본 명세서에 기술된 조성물은 상기 기술된 다양한 약물 송달 시스템에서 사용하기에 적절하다. 추가적으로, 투여된 화합물의 생체 내 혈청 반-감기를 증가시키기 위해, 조성물은 캡슐화되거나, 리포솜의 내장 내부에 도입되거나, 콜로이드로서 제조될 수 있고, 또는 조성물의 연장된 혈청 반-감기를 제공하는 다른 통상적인 기술이 사용될 수 있다. 예컨대, 각각 본 명세서에 참고로서 포함된 Szoka, et al, 미국 특허 제4,235,871호, 제

4,501,728호 및 제4,837,028호에 기술된 리포솜을 제조하는 다양한 방법이 이용가능하다. 추가로, 표적화 약물 전달 시스템, 예를 들면, 조직-특이적 항체로 코팅된 리포솜에 내포된 약물을 투여할 수 있다. 리포솜은 조직을 표적으로 하며 조직에 의해 선별적으로 흡수될 것이다.

[0104] 또한 제공된 것은 포유류 세포에서 레트로바이러스, 예를 들면 렌티바이러스 가령 인간 면역결핍 바이러스, 원숭이 면역결핍 바이러스, 고양이 면역결핍 바이러스, 또는 소의 면역결핍 바이러스를 비활성화하는 방법이다. 인간 면역결핍 바이러스는 HIV-1 또는 HIV-2일 수 있다. 인간 면역결핍 바이러스는 염색체 내에 통합된 프로바이러스일 수 있다. 포유류 세포는 CD4+ 림프구, 대식 세포, 섬유아세포, 단핵백혈구, T 림프구, B 림프구, 자연 살해 세포, 수지상 세포 가령 랑게르한스(Langerhans) 세포 및 소포성 수지상 세포, 조혈 줄기세포, 내피 세포, 뇌 미소아교 세포, 및 위장 상피 세포를 포함하지만, 이에 제한되지 않는 HIV에 감염된 임의의 세포 유형일 수 있다. 그러한 세포 유형은 전형적으로 원발성 감염 도중에 감염되는 세포 유형, 예를 들면, CD4+ 림프구, 대식 세포, 또는 랑게르한스 세포, 뿐만 아니라 잠복성 HIV 저장소를 이루는 세포 유형, 즉, 잠복성으로 감염된 세포를 포함한다.

[0105] 본 방법은 CRISPR-연관 엔도뉴클레아제 및 하나 이상의 가이드 RNA를 포함하는 유전자 편집 복합체(gene editing complex)를 인코딩하는 단리된 핵산을 포함하는 조성물에 세포를 노출시키는 단계를 포함하며, 여기서 가이드 RNA는 레트로바이러스 내 표적 핵산 서열에 상보적이다. 접촉하는 단계는 생체 내에서 일어날 수 있고, 즉 조성물은 HIV 감염증을 가지는 개체에 직접 투여될 수 있다. 본 방법은 제한되지 않지만, 접촉하는 단계는 생체외에서 일어날 수 있다. 예를 들면, 하나의 세포 또는 다수의 세포, 또는 조직 외식편이 HIV 감염증을 가지는 개체에서 제거되어 배양되고, 이후 CRISPR-연관 엔도뉴클레아제 및 가이드 RNA를 포함하는 조성물과 접촉할 수 있고 여기서 가이드 RNA는 인간 면역결핍 바이러스 내 핵산 서열에 상보적이다. 상기 기술된 바와 같이, 조성물은 CRISPR-연관 엔도뉴클레아제 및 가이드 RNA를 인코딩하는 핵산, 여기서 가이드 RNA는 인간 면역결핍 바이러스 내 핵산 서열에 상보적임; 상기 핵산 서열을 포함하는 발현 벡터; 또는 CRISPR-연관 엔도뉴클레아제 및 가이드 RNA를 인코딩하는 핵산, 여기서 가이드 RNA는 인간 면역결핍 바이러스 내 핵산 서열에 상보적임; 또는 상기 핵산 서열을 포함하는 발현 벡터를 포함하는 약제학적 조성물일 수 있다. 일부 구체예에서, 유전자 편집 복합체는 CRISPR-연관 엔도뉴클레아제 폴리펩티드 및 가이드 RNA를 포함할 수 있고 여기서 가이드 RNA는 인간 면역결핍 바이러스 내 핵산 서열에 상보적이다.

[0106] 조성물이 핵산 또는 폴리펩티드로서 투여되는지와 상관없이, 조성물은 포유류 세포에 의한 흡수를 촉진하는 방식으로 제형화된다. 유용한 벡터 시스템 및 제형은 위에서 기술되었다. 일부 구체예에서 벡터는 조성물을 특정한 세포 유형으로 전달할 수 있다. 본 발명은 제한적이지 않지만, 다른 DNA 전달 방법 가령 예를 들면 칼슘 포스페이트, DEAE 텍스트란, 리포솜, 지질접합체(lipoplex), 계면활성제, 및 퍼플루오로 화학적 액체를 이용하는 화학적 형질감염, 마찬가지로 물리적 전달 방법, 가령 전기천공, 미세 주입, 탄도 입자, 및 "유전자 총(gun)" 시스템이 또한 고려된다.

[0107] 표준 방법, 예를 들면, CRISPR-연관 엔도뉴클레아제를 검출하기 위한 면역어세이, 또는 핵산-기반 어세이 가령 gRNA를 검출하기 위한 PCR이 복합체가 도입된 세포에 의해 흡수되고 발현되는지 확인하기 위해 사용될 수 있다. 조작된 세포는 이후 하기 기술된 바와 같이 이들이 유래한 개체로 재도입될 수 있다.

[0108] 유전자 편집 복합체는 CRISPR-연관 뉴클레아제, 예컨대, Cas9, 및 레트로바이러스 표적 서열, 예를 들면, HIV 표적 서열에 상보적인 가이드 RNA를 포함한다. 유전자 편집 복합체는 다양한 돌연변이를 프로바이러스 DNA에 도입시킬 수 있다. 그러한 돌연변이가 바이러스를 비활성화하는 메커니즘은 다양할 수 있고, 예를 들면 돌연변이는 프로바이러스 복제, 바이러스 유전자 발현 또는 프로바이러스 절제에 영향을 줄 수 있다. 돌연변이는 조절 서열 또는 구조적 유전자 서열에 배치되어 HIV 생산 결함을 야기할 수 있다. 돌연변이는 결실을 포함할 수 있다. 결실의 크기는 단일 뉴클레오티드 염기 쌍 내지 약 10,000 염기 쌍까지 달라질 수 있다. 일부 구체예에서 결실은 프로바이러스 서열 전체 또는 실질적으로 전체를 포함할 수 있다. 일부 구체예에서 결실은 프로바이러스 서열 전체를 포함할 수 있다. 돌연변이는 삽입을 포함할 수 있고, 즉 하나 이상의 뉴클레오티드 염기 쌍이 프로-바이러스 서열에 부가될 수 있다. 삽입되는 서열의 크기 또한 다양할 수 있고, 예를 들면 약 하나의 염기 쌍 내지 약 300 염기 쌍까지 달라질 수 있다. 돌연변이는 점 돌연변이를 포함할 수 있고, 즉, 단일 뉴클레오티드가 또다른 뉴클레오티드로 교체될 수 있다. 유용한 점 돌연변이는 기능적 결과를 내는 것들, 예를 들면, 아미노산 코돈을 종말 코돈으로 전환하거나 비기능적 단백질의 생산을 야기하는 돌연변이다.

[0109] 다른 구체예에서, 조성물은 하나 이상의 Cas/gRNA 벡터로 형질전환 또는 형질주입된 세포를 포함한다. 일부 구체예에서, 본 발명의 방법은 생체외에서 적용될 수 있다. 즉, 개체의 세포는 신체에서 제거되어 HIV 서열을 절

제하기 위해 배양물 내에서 조성물로 처리되고 처리된 세포가 개체의 신체로 되돌아갈 수 있다. 세포는 개체의 세포일 수 있거나 세포는 단상형(haplotype) 일치된 것 또는 세포주일 수 있다. 세포는 복제를 방지하기 위해 방사선 조사될 수 있다. 일부 구체예에서, 세포는 인간 백혈구 항원 (HLA)-일치된 것, 자가, 세포주, 또는 이들의 조합이다. 다른 구체예에서, 세포는 줄기 세포일 수 있다. 예를 들면, 배아 줄기세포 또는 인공적인 만능 줄기세포 (유도된 만능 줄기세포 (iPS 세포)). 배아 줄기세포 (ES 세포) 및 인공적인 만능 줄기세포 (유도된 만능 줄기세포, iPS 세포)는 인간을 비롯한 많은 동물종으로부터 만들어졌다. 이러한 유형의 만능 줄기세포는 적절한 분화 유도에 의해 다능성을 유지하면서 활발히 분열하는 능력을 보유하면서도 거의 모든 장기로 분화할 수 있기 때문에 재생 의학을 위한 세포의 가장 유용한 공급원일 것이다. 특히 iPS 세포는 자가-유래 체세포로부터 만들어질 수 있고, 따라서 배아를 파괴하여 생성되는 ES 세포에 비해 윤리적이고 사회적인 논란을 덜 야기한다. 게다가, 자가-유래 세포인 iPS 세포는 재생 의학 또는 이식 요법에서 가장 큰 장애물인 거부 반응을 피할 수 있게 해준다.

[0110] gRNA 발현 카세트는 당해 분야에 공지된 방법, 예를 들면, siRNA를 송달하는 방법으로 개체에 용이하게 송달될 수 있다. 일부 양태에서, Cas는 Cas 분자의 활성 도메인이 포함된 단편으로, 이를 통해 분자 크기로 절단될 수 있다. 따라서, Cas9/gRNA분자는 최신 유전자 요법에서 취하는 접근법과 유사하게 임상적으로 사용될 수 있다. 특히, 세포 이식 요법 뿐만 아니라 HIV-1 면역화를 위한 Cas9/다중 gRNA 안정한 발현 줄기세포 또는 iPS 세포가 개체에서 사용되기 위해 개발될 것이다.

[0111] 형질도입된 세포가 공지의 방법에 따른 재융합을 위해 제조된다. 약 2-4 주의 배양 기간 후에, 세포는 1×10^6 및 1×10^{10} 의 수일 수 있다. 이와 관련하여, 세포의 성장 특징은 환자마다 및 세포 유형마다 달라질 수 있다. 형질도입된 세포를 재융합하기 약 72시간 전에, 치료제를 발현하는 세포의 표현형, 및 백분율을 분석하기 위해 일부를 분취한다. 투여를 위해, 본 발명의 세포는 환자의 체중 및 전반적인 건강에 적용하는 것처럼, 세포 유형의 LD₅₀, 및 다양한 농도에서 세포 유형의 부작용에 의해 결정된 비율로 투여될 수 있다. 투여는 단일 또는 분할 용량으로 이루어질 수 있다. 성인 줄기세포가 또한 이들의 생산을 촉진하고 골수 또는 지방질 조직을 포함할 수 있지만, 이에 제한되지 않는 조직 또는 장소로부터 떠나게 하는 외생적으로 투여된 요인을 이용하여 동원될 수 있다.

[0112] 제조 물품

[0113] 본 명세서에 기술된 조성물은 예를 들면, 레트로바이러스 감염, 예를 들면, HIV 감염증을 가지는 개체 또는 레트로바이러스 감염, 예를 들면, HIV 감염증에 걸릴 위험이 있는 개체를 치료하는 요법으로서 사용하기 위해 라벨이 달린 적절한 용기에 포장될 수 있다. 용기는 CRISPR-연관 엔도뉴클레아제, 예를 들면, Cas9 엔도뉴클레아제, 및 인간 면역결핍 바이러스 내 표적 서열에 상보적인 가이드 RNA를 인코딩하는 핵산 서열, 또는 상기 핵산을 인코딩하는 벡터, 및 하나 이상의 적절한 안정화제, 담체 분자, 착향제, 및/또는 의도하는 용도를 위해 적절한 것 등을 포함하는 조성물을 포함할 수 있다. 따라서, 적어도 하나의 본 발명의 조성물, 예컨대, CRISPR-연관 엔도뉴클레아제, 예를 들면, Cas9 엔도뉴클레아제, 및 인간 면역결핍 바이러스 내 표적 서열에 상보적인 가이드 RNA를 인코딩하는 핵산 서열, 또는 상기 핵산을 인코딩하는 벡터 및 사용 설명서를 포함하는, 포장된 제품 (예컨대, 농축된 또는 바로 쓸 수 있는(ready-to-use) 농도로 본 명세서에 기술된 하나 이상의 조성물을 내포하고 저장, 배송, 또는 판매를 위해 포장된 무균 용기) 및 키트가 본 발명의 범위 내에 포함된다. 제품은 하나 이상의 본 발명의 조성물을 내포하는 용기(예컨대, 바이알, 자(jar), 병, 봉투, 등)를 포함할 수 있다. 또한, 제조 물품은, 예를 들면 예방 또는 치료가 요구되는 용태를 치료하거나 관찰하기 위한 포장 재료, 사용 설명서, 주사기, 송달 장치, 완충액 또는 다른 대조 시약을 추가로 포함할 수 있다.

[0114] 일부 구체예에서, 키트는 하나 이상의 추가적인 항레트로바이러스 물질, 예를 들면, 역전사효소 저해제, 프로테아제 저해제 또는 침입 저해제를 포함할 수 있다. 추가적인 물질이 CRISPR-연관 엔도뉴클레아제, 예를 들면, Cas9 엔도뉴클레아제, 및 인간 면역결핍 바이러스 내 표적 서열에 상보적인 가이드 RNA를 인코딩하는 핵산 서열, 또는 상기 핵산을 인코딩하는 벡터와 동일한 용기에 함께 포장될 수 있거나 별도로 포장될 수 있다. CRISPR-연관 엔도뉴클레아제, 예를 들면, Cas9 엔도뉴클레아제, 및 인간 면역결핍 바이러스 내 표적 서열에 상보적인 가이드 RNA를 인코딩하는 핵산 서열, 또는 상기 핵산을 인코딩하는 벡터 및 추가적인 물질은 사용 직전에 조합되거나 별도로 투여될 수 있다.

[0115] 제품은 또한 설명서(예컨대, 인쇄된 라벨 또는 삽입물 또는 제품 사용을 설명하는 다른 매체 (예컨대, 오디오- 또는 비디오테이프))를 포함할 수 있다. 설명서는 용기와 연계될 수 있고 (예컨대, 용기에 부착) 포함된 조성물

이 투여되어야 하는 방식(예컨대, 투여의 빈도 및 경로), 이를 위한 적응증, 및 다른 용도를 설명할 수 있다. 조성물은 바로 투여할 수 있고 (예컨대, 투여-적합 단위로 존재), 하나 이상의 추가적인 약제학적으로 허용되는 어쥬번트, 담체 또는 다른 희석제 및/또는 추가적인 치료제를 포함할 수 있다. 대안적으로, 조성물은 희석제 및 희석을 위한 설명과 함께 농축 형태로 제공될 수 있다.

[0116] 실시예

[0117] 실시예 1: 재료 및 방법

[0118] 플라스미드 제조: 인간 Cas9 및 gRNA 발현 카세트, pX260, 및 pX330 (Addgene)을 내포하는 벡터를 사용하여 다양한 구조체, LTR-A, B, C, 및 D를 생성하였다.

[0119] 세포 배양물 및 안정한 세포주: TZM-bI 리포터 및 U1 세포주를 NIH AIDS Reagent Program으로부터 얻었고 CHME5 미소아교 세포는 당해 분야에 공지되어 있다.

[0120] 면역조직화학 및 웨스턴 블롯: 세포의 조직화학적 관찰을 위한 표준 방법 및 웨스턴 블롯에 의한 단백질 발현 평가를 사용하였다.

[0121] 반딧불-루시페라아제 어세이: 세포를 부동 용해 완충액(Passive Lysis Buffer, Promega)을 이용하여 24시간 처리-후 용리하였고 루시페라아제 리포터 유전자 어세이 키트(Promega)를 이용하여 제조사의 프로토콜에 따라 분석하였다. 루시페라아제 활성은 평행 MTT 어세이 (Vybrant, Invitrogen)에 의해 측정된 세포수로 정상화하였다.

[0122] p24 ELISA: 감염 또는 비활성화 이후, p24 Gag ELISA (Advanced Bioscience Laboratories, Inc)에 의해 제조사의 프로토콜에 따라 상청액 내 HIV-1 바이러스 로드 수준을 정량화하였다. 처리에 따른 세포 생존능을 평가하기 위해, MTT 어세이를 제조사의 프로토콜에 따라 평행하게 수행하였다 (Vybrant, Invitrogen).

[0123] EGFP 유세포 분석: 세포를 트립신화하고, PBS로 세척하고 실온의 2% 파라포름알데히드에서 10분간 고정시킨 후, PBS로 두 차례 세척하고 Guava EasyCyte Mini 유세포 분석기(Guava Technologies)를 이용하여 분석하였다.

[0124] HIV-1 리포터 바이러스 제조 및 감염: HEK293T 세포를 pNL4-3-ΔE-EGFP (NIH AIDS Research and Reference Reagent Program)와 함께 리포펙타민 2000 시약 (Invitrogen)을 이용하여 형질주입시켰다. 48 시간 후에, 상청액을 수집하고, 0.45 μm 여과하고 HeLa 세포에서 감염 마커로서 EGFP를 이용하여 적정하였다. 바이러스 감염을 위해, 안정한 Cas9/gRNA TZM-bI 세포를 2 시간 동안 희석된 바이러스 스톡으로 배양하고, 이후 PBS로 두 차례 세척하였다. 감염-후 2일 및 4일에, 세포를 수집하고, 고정하고 EGFP 발현에 대해 유세포 분석으로 분석하거나, 또는 PCR 및 전체 유전체 서열분석을 위해 유전체 DNA 정제를 수행하였다.

[0125] 유전체 DNA 증폭, PCR, TA-클로닝, 및 Sanger 서열분석, Genome Walker 연결 PCR: 클로닝 및 서열분석을 위해 DNA 조각을 위한 표준 방법을 사용하였다. HIV-1 통합 부위를 식별하기 위해, 본 발명자는 Lenti-X™ 통합 부위 분석 키트를 사용하였다.

[0126] Surveyor 어세이: PCR 산물 내 돌연변이의 존재를 SURVEYOR 돌연변이 검출 키트 (Transgenomic)를 이용하여 제조사의 프로토콜에 따라 검사하였다. 간단히 뒤섞인 PCR 산물을 95℃에서 10분간 변성시키고 유전자증폭기 (thermocycler)를 이용하여 점차 냉각시키며 혼성화시켰다. 이후, 300 ng의 혼성화된 DNA (9 μl)을 0.25 μl SURVEYOR 인핸서 S 및 15 mM MgCl₂의 존재에서 4시간 동안 42℃에서 0.25 μl의 SURVEYOR 뉴클레아제로 절단 처리하였다. 중지 용액을 부가하고 샘플을 2% 아가로스 겔에서 동일한 양의 미절단된 PCR 산물 대조군과 함께 분석하였다.

[0127] 일부 PCR 산물을 제한 단편 길이 다형성 분석을 위해 사용하였다. 동일한 양의 PCR 산물을 BsaJI로 절단시켰다. 절단된 DNA를 에티디움 브로마이드-함유 아가로스 겔 (2%)에 분배하였다. 서열분석을 위해, PCR 산물을 TA 클로닝® 키트 이중 프로모터를 이용하여 pCR™II 벡터와 함께 (TA Cloning® Kit Dual Promoter with pCR™II vector, Invitrogen) 클로닝하였다. 삽입물을 EcoRI로 분해하여 확인하고 양성 클론을 Sanger 서열 분석을 위해 Genewiz로 보냈다.

[0128] LTR 표적 부위 선별, 전체 유전체 서열분석 및 생물정보학 및 통계 분석. 본 발명자는 처음에 LTR 내부의 가능한 표적 부위를 식별하기 위해 Jack Lin CRISPR/Cas9 gRNA 검색 도구를 이용하였다.

[0129] 플라스미드 제조. 전구-crRNA를 위한 LTR-A 또는 LTR-B를 발현하는 DNA 조각을 퓨로마이신 선별 유전자가 내포된 pX260 벡터 (Addgene, 플라스미드 #42229)에 클로닝하였다. 키메라 crRNA-tracrRNA를 위한 LTR-C 또는 LTR-D를 발현하는 DNA 조각을 pX330 벡터 (Addgene, 플라스미드 #42230)에 클로닝하였다. 두 벡터는 모두 CAG 프로

모터에 의해 유도되는 인간화된 Cas9 암호화 서열 및 인간 U6 프로모터에 의해 유도되는 gRNA 발현 카세트를 내포한다. 벡터를 *Bbs*I로 절단하고 안탈틱 포스파타제(Antarctic Phosphatase)로 처리하고, 선형화된 벡터를 Quick 뉴클레오티드 제거 키트 (Qiagen)로 정제하였다. 각각의 표적화 부위에 대한 올리고뉴클레오티드 쌍 (도 14, AlphaDNA)을 다시 결합시키고, 인산화하고, 선형 벡터에 결합시켰다. gRNA 발현 카세트를 GENEWIZ에서 U6 서열분석 프라이머 (도 14)로 서열분석하였다. pX330 벡터에 있어서, 본 발명자는 즉시 형질감염 또는 다른 벡터로 서브클로닝하기 위해 gRNA 발현 카세트 (U6-gRNA-crRNA-줄기-tracrRNA)를 가지런하게 만들 수 있는 오버행 (overhang) 절단 부위 (도 14)를 가지는 보편 PCR 프라이머의 쌍을 설계하였다.

[0130] 세포 배양. Dr John C. Kappes, Dr Xiaoyun Wu 및 Tranzyme Inc로부터의 TZM-bI 리포터 세포주, Dr. Thomas Folks로부터의 U1/Hiv-1 세포주 및 Dr. Eric Verdin으로부터의 J-Lat 전장 클론을 NIH AIDS Reagent Program의 미국립보건원 에이즈 부서(Division of AIDS, NIAID, NIH)를 통해 얻었다. CHME5/HIV 태아 미소아교세포 세포주를 앞서 기술한 바와 같이 생성하였다. TZM-bI 및 CHME5 세포를 고글루코스 보충된 돌베코(Dulbecco) 최소 영양 배지에서 10% 열-비활성화된 태아 소 혈청 (FBS) 및 1% 페니실린/스트렙토마이신과 함께 배양하였다. U1 및 J-Lat 세포를 2.0 mM L-글루타민, 10% FBS 및 1% 페니실린/스트렙토마이신을 함유하는 RPMI 1640에서 배양하였다.

[0131] 안정한 세포주 및 서브클로닝. TZM-bI 또는 CHME5/HIV 세포를 6-웰 플레이트에 1.5×10^5 세포/웰로 시딩하고 리포팩틴 2000 시약 (Invitrogen)을 이용하여 1 μ g의 pX260 (LTR-A 및 B를 위해) 또는 1 μ g/0.1 μ g의 pX330/pX260 (LTR-C 및 D를 위해) 플라스미드로 형질주입하였다. 다음날, 세포를 100-mm 접시에 옮기고 1 μ g/ml의 푸로마이신 (Sigma)을 함유하는 성장 배지에서 배양하였다. 두 주 후에, 생존하는 세포 콜로니를 클로닝 실린더 (Corning)를 이용하여 단리하였다. U1 세포(1.5×10^5)를 Neon™ 형질감염 시스템 (Invitrogen)에서 10 μ l 팁, 3 x 10 ms 1400 V 자극을 이용하여 1 μ g의 DNA로 전기천공시켰다. 세포를 0.5 μ g/ml의 푸로마이신으로 두 주간 선별하였다. 안정한 클론을 96-웰 플레이트에서 제한 희석 방법을 이용하여 계대배양하고 단일 세포-유래 서브클론을 추가적인 연구를 위해 유지하였다.

[0132] 면역세포화학 및 웨스턴 블롯. Cas9/gRNA 안정한 발현 TZM-bI 세포를 8-웰 챔버 슬라이드에서 2일 동안 배양하면 4% 파라포름알데히드/PBS에서 10분 동안 고정하였다. 세 차례 행균 후에, 세포를 0.5% Triton X-100/PBS로 20분간 처리하고 10% 당나귀 혈청으로 1시간 동안 차단하였다. 세포를 밤새 4°C에서 마우스 항-Flag M2 일차 항체 (1:500, Sigma)와 함께 배양하였다. 세 차례 행균 후에, 세포를 당나귀 항-마우스 Alexa-Fluor-594 이차 항체와 함께 1시간 동안 배양하고, Hoechst 33258과 함께 5분간 배양하였다. PBS로 세 차례 행균 후에, 세포에 페이딩-방지 수성 고정 매체 (Biomeda)를 얇게 덮고 Leica DMI6000B 형광 현미경 하에서 분석하였다.

[0133] 6-웰 플레이트에서 배양한 TZM-bI 세포를 20 mM Tris-HCl (pH 7.4), 1% Triton X-100, 5 mM 에틸렌디아민테트라아세트산, 5 mM 디티오프레이트, 150 mM NaCl, 1 mM 페닐메틸설포닐 플루오라이드, 1x 핵 추출 프로테이나제 저해제 카테일 (Cayman Chemical, Ann Arbor, MI), 1 mM 소듐 오르소바나테이트 및 30 mM NaF를 함유하는 200 μ l의 Triton X-100-기반 용해 완충액에서 용해하였다. 세포 용해물을 4°C에서 30분간 원심분리시켰다. 핵 및 세포 파편을 4°C에서 20,000 g로 20분간 원심분리하여 제거하였다. 동일한 양의 용해물 단백질 (20 μ g)을 5분간 나트륨 도데실 설페이트 (SDS) 샘플 완충액 내에서 끓임으로써 변성시키고, 트리-글리신 완충액에서 SDS-폴리아크릴아미드 겔 전기영동에 의해 분별하고, 니트로셀룰로스 막 (BioRad)으로 이동시켰다. SeeBlue로 선염색한 표준 (Invitrogen)을 분자량 기준으로서 이용하였다. 블롯을 5% BSA/트리스-완충 식염수 (pH 7.6) 더하기 0.1% Tween-20 (TBS-T)에서 1시간 동안 차단하고 이후 밤새 4°C에서 마우스 항-Flag M2 단클론 항체 (1:1000, Sigma) 또는 마우스 항-GAPDH 단클론 항체 (1:3000, Santa Cruz Biotechnology)와 함께 배양하였다. TBS-T로 행균 후에, 블롯을 IRDye 680LT-접합된 항-마우스 항체와 함께 1시간 동안 실온에서 배양하였다. 막을 스캔하고 Odyssey Infrared Imaging System (LI-COR Biosciences)를 이용하여 분석하였다.

[0134] 반딧불-루시페라아제 어세이. 세포를 부동 용해 완충액(Promega)을 이용하여 24시간 처리-후 용리하였고 루시페라아제 리포터 유전자 어세이 키트(Promega)를 이용하여 제조사의 프로토콜에 따라 분석하였다. 루시페라아제 활성은 평행 MTT 어세이 (Vybrant, Invitrogen)에 의해 측정된 세포수로 정상화하였다.

[0135] p24 ELISA 감염 또는 비활성화 이후, p24 Gag ELISA (Advanced Bioscience Laboratories, Inc)에 의해 제조사의 프로토콜에 따라 상청액 내 HIV-1 바이러스 로드 수준을 정량화하였다. 처리에 따른 세포 생존능을 평가하기 위해, MTT 어세이를 제조사의 프로토콜에 따라 평행하게 수행하였다 (Vybrant, Invitrogen).

[0136] EGFP 유세포 분석. 세포를 트립신화하고, PBS로 세척하고 실온의 2% 파라포름알데히드에서 10분간 고정시킨 후,

PBS로 두 차례 세척하고 Guava EasyCyte Mini 유세포 분석기(Guava Technologies)를 이용하여 분석하였다.

- [0137] Hiv-1 리포터 바이러스 제조 및 감염. HEK293T 세포를 pNL4-3-ΔE-EGFP, SF162 및 JRFL (NIH AIDS Research and Reference Reagent Program)과 함께 리포펙타민 2000 시약 (Invitrogen)을 이용하여 형질주입시켰다. 위험 pNL4-3-ΔE-EGFP에 있어서, VSVG 벡터를 동시형질감염시켰다. 48 시간 후에, 상청액을 수집하고, 0.45 μm 여과하고 HeLa 세포에서 감염 마커로서 발현된 EGFP를 이용하여 적정하였다. 바이러스 감염을 위해, 안정한 Cas9/gRNA T2M-b1 세포를 2 시간 동안 희석된 바이러스 스톡으로 배양하고, PBS로 두 차례 세척하였다. 감염-후 2일 및 4일에, 세포를 수집하고, 고정하고 EGFP 발현에 대해 유세포 분석으로 분석하거나, 또는 PCR 및 전체 유전체 서열분석을 위해 유전체 DNA 정제를 수행하였다.
- [0138] 유전체 DNA 정제, PCR, TA-클로닝, 및 Sanger 서열분석. 유전체 DNA를 ArchivePure DNA 세포/조직 정제 키트 (5PRIME)를 이용하여 제조자가 권고하는 프로토콜에 따라 세포로부터 단리하였다. 100 ng의 추출된 DNA를 도 14에 나열된 프라이머를 이용하여 높은-정확성의 FailSafe PCR 키트 (Epicentre)를 이용하여 PCR 처리하였다. 세 단계의 표준 PCR을 55°C 어닐링 및 72°C 연장으로 30 사이클 수행하였다. 산물을 2% 아가로스 겔에서 분석하였다. 관심의 밴드를 겔-정제하고 pCRII T-A 벡터 (Invitrogen)로 클로닝시키고, 개별적인 클론의 뉴클레오티드 서열을 Genewiz에서 보편 T7 및/또는 SP6 프라이머를 이용하여 서열분석함으로써 결정하였다.
- [0139] 통상적인 및 실시간 역전사 (RT)-PCR. 총 RNA 추출을 위해, 세포를 RNeasy Mini 키트 (Qiagen)를 이용하여 제조자의 설명에 따라 처리하였다. 잠재적으로 잔여하는 유전체 DNA를 RNase-Free DNase Set (Qiagen)과 함께 컬럼-상 DNA 가수분해효소 절단을 통해 제거하였다. 각 샘플의 1 μg을 랜덤 헥사뉴클레오티드 프라이머를 이용하여 고성능 cDNA 역전사 키트 (Invitrogen, Grand Island, NY)로 cDNA로 역전사하였다. 통상적인 PCR을 표준 프로토콜을 이용하여 수행하였다.
- [0140] 정량적 PCR (qPCR) 분석을 SYBR® Green PCR Master 혼합 키트 (Applied Biosystems)를 이용하여 LightCycler480 (Roche)에서 수행하였다. RT 반응물을 반응물 마이크로-리터당 5 ng의 총 RNA까지 희석하였고 20-μl PCR 반응물에서 2 μl를 사용하였다. HIV-1 프로바이러스의 qPCR 분석을 위해, 50 ng의 유전체 DNA를 사용하였다. 프라이머를 AlphaDNA에서 합성하였고 도 14에 나타난다. 인간 항존유전자 GAPDH 및 RPL13A를 위한 프라이머를 RealTimePrimers (Elkins Park, PA)에서 수득하였다. 각각의 샘플을 삼중복으로 시험하였다. 표적 유전자 및 항존-유전자에 대한 사이클 역치 (Ct) 값을 그래프로부터 얻었다. 항존유전자 및 표적 유전자 사이의 Ct 값 차이는 ΔCt 값으로 표현된다. 실험 샘플에서 대조 샘플의 ΔCt 값을 뺀으로써 ΔΔCt 값을 얻었다. 비교적인 배수 또는 백분율 변화를 2-ΔΔCt로서 산출하였다. 일부 경우에, 인간 유전체 DNA에서 표준으로서 기준화된 pNL4-3-ΔE-EGFP 플라스미드를 이용하여 절대 정량화를 수행하였다. HIV-1 바이러스 복제물의 수를 항존유전자로 정규화한 후에 표준 곡선을 기준으로 산출하였다.
- [0141] Genome Walker 연결 PCR 및 긴-범위 PCR. 숙주 세포에서 HIV-1의 통합 부위를 Lenti-X™ 통합 부위 분석 키트 (Clontech)를 이용하여 제조자의 설명에 따라 식별하였다. 간략하게, 고품질 유전체 DNA를 NucleoSpin 조직 키트 (Clontech)를 이용하여 U1 세포로부터 추출하였다. 바이러스 통합 라이브러리를 구성하기 위해, 각각의 유전체 DNA 샘플을 평활-말단(blunt-end)-생성절단효소 *Dra* I, *Ssp* I 또는 *Hpa* I를 이용하여 개별적으로 밤새 37°C에서 절단하였다. 절단 효율을 0.6% 아가로스 상에서 전기영동에 의해 확인하였다. 절단된 DNA를 NucleoSpin 겔 및 PCR 제거 키트를 이용하여 정제하고 이후 절단된 유전체 DNA 단편을 Genome Walker™ Adaptor로 16°C에서 밤새 결합시켰다. 결합 반응을 5분간 70°C에서 배양함으로써 중지시키고 TE 완충액으로 5배 희석하였다. 일차 PCR을 DNA 조각에 대해 어댑터(adaptor) 프라이머 1 (AP1) 및 LTR-특이적 프라이머 1 (LSP1)와 함께 Advantage 2 Polymerase Mix를 이용하여 수행하고 이후 이차 (내포된) PCR을 AP2 및 LSP2 프라이머를 이용하여 수행하였다 (도 14). 이차 PCR 산물을 1.5% 에티디움 브로마이드-함유 아가로스 겔 상에 분배하였다. 주요 밴드를 겔-정제하고 pCRII T-A 벡터 (Invitrogen)로 클로닝시키고, 개별적인 클론의 뉴클레오티드 서열을 Genewiz에서 보편 T7 및 SP6 프라이머를 이용하여 서열분석함으로써 결정하였다. 서열 판독을 NCBI BLAST 검색에 의해 분석하였다. U1 세포 내 HIV-1의 두 가지 통합 부위를 염색체 X 및 2에서 확인하였다. 각각의 통합 부위를 포괄하는 프라이머의 쌍(도 14)을 AlphaDNA에서 합성하였다. U1 유전체 DNA를 이용한 긴-범위 PCR을 Phusion 고-정확도 PCR 키트 (New England Biolabs)를 이용하여 제조사의 프로토콜에 따라 수행하였다. PCR 산물을 1% 아가로스 겔 상에 가시화하고 Sanger 서열분석으로 확인하였다.
- [0142] Surveyor 어레이. PCR 산물 내 돌연변이의 존재를 SURVEYOR 돌연변이 검출 키트 (Transgenomic)를 이용하여 제조자의 프로토콜에 따라 시험하였다. 간단히 뒤섞인 PCR 산물을 95°C에서 10분간 변성시키고 유전자증폭기 (thermocycler)를 이용하여 점차 냉각시키며 혼성화시켰다. 이후, 300 ng의 혼성화된 DNA (9 μl)을 0.25 μl

SURVEYOR 인헨서 S 및 15 mM MgCl₂의 존재에서 4시간 동안 42℃에서 0.25 μl의 SURVEYOR 뉴클레아제로 절단 처리하였다. 증지 용액을 부가하고 샘플을 2% 아가로스 겔에서 동일한 양의 미절단된 PCR 산물과 함께 분석하였다.

[0143] 일부 PCR 산물을 제한 단편 길이 다형성 분석을 위해 사용하였다. 동일한 양의 PCR 산물을 *Bsa*II로 절단시켰다. 절단된 DNA를 에티디움 브로마이드-함유 아가로스 겔 (2%)에 분배하였다. 서열분석을 위해, PCR 산물을 TA 클로닝® 키트 이중 프로모터를 이용하여 pCR™II 벡터와 함께 (TA Cloning® Kit Dual Promoter with pCR™II vector, Invitrogen) 클로닝하였다. 삽입물을 *Eco*RI로 분해하여 확인하고 양성 클론을 Sanger 서열 분석을 위해 Genewiz로 보냈다.

[0144] LTR 표적 부위의 선별 및 가능한 표적-이탈 부위의 예측. 최초 연구를 위해, 본 발명자는 계대배양 도중의 LTR의 돌연변이 가능성 때문에, 통합된 렌티바이러스 LTR-루시페라아제 리포터의 LTR 프로모터 서열 (-411 내지 -10)을 인간 TZM-b1 세포의 유전체로부터의 PCR 산물의 TA-클로닝 서열분석에 의해 얻었다. 이러한 프로모터 서열은 pHR'-CMV-LacZ 렌티바이러스 벡터 (AF105229)의 5'-LTR과 100% 일치한다. 따라서, 전장 pHR' 5'-LTR (634 bp)의 센스 및 안티센스 서열을 활용하여 20 bp gRNA 표적화 서열 더하기 PAM 서열 (NRG)을 내포하는 Cas9/gRNA 표적 부위를 검색하기 위해 Jack Lin CRISPR/Cas9 gRNA 검색 도구를 사용하였다 (<http://spot.colorado.edu/~slln/cas9.html>). 정확한 일치치를 보이는 가능한 표적-이탈의 수는 NCBI/blastn 스위트를 이용하고 E-값 한계 1,000 및 글자 크기 7로 하여 각각의 gRNA 표적화 서열 더하기 NRG (AGG, TGG, GGG 및 CGG; AAG, TAG, GAG, CAG)를 모든 이용가능한 인간 유전체 및 전사 서열에 대해 블라스트 처리하여 예측하였다. Control + F 키를 누른 후, 표적 서열 (1-23부터 9-23까지 뉴클레오타이드)를 복사/붙여넣기하고 표적 서열에 100% 일치하는 유전체 표적의 수를 찾는다. 반복된 유전체 라이브러리 때문에 각 검색에서 표적-이탈의 수를 3으로 나눴다.

[0145] 전체 유전체 서열분석 및 생물정보 분석. TZM-b1 세포의 대조 서브클론 C1 및 실험 서브클론 AB7을 LTR-루시페라아제 리포터의 표적 조각 효율 및 기능적 억제에 대하여 검증하였다. 유전체 DNA를 NucleoSpin 조직 키트 (Clontech)로 단리하였다. DNA 샘플을 탭플 대학교 Fox Chase 암 센터의 NextGen 서열분석 설비에 제출하였다. 복제된 유전체 DNA 라이브러리를 Illumina를 위한 NEBNext Ultra DNA 라이브러리 제조 키트 (New England Biolab)을 이용하여 제조자의 설명에 따라 각각의 서브클론으로부터 제조하였다. 모든 라이브러리를 HiSeq 2500 장비 (Illumina)에서 두 대의 Illumina Rapid Run 플로우셀(flowcells)에서 대응-말단(paired-end) 141-bp 판독을 이용하여 서열분석하였다. 서열분석된 라이브러리로부터 탈복합화된 판독 데이터는 전문적인 생물정보 분석을 위해 AccuraScience, LLC (<http://www accurascience.com>)로 보냈다. 간단히, 원본 판독물은 Bowtie2를 이용하여 인간 유전체 (hg19) 및 HIV-1 유전체에 대해 유전자지도로 그려졌다. 중복된 판독물 제거, 국부 정렬, 염기 품질 재조정 및 삽입결실 호출을 위해 유전체 분석 킷 (GATK, 버전 2.8.1)를 사용하였다. 신뢰성 점수 10 및 30은 저품질 (LowQual) 및 고신뢰성 호출 (PASS)에 대한 역치였다. 다양한 미스매치를 갖는 LTR-A 및 LTR-B의 가능한 표적-이탈 부위를 상기 기술된 바와 같이 NCBI/blastn 스위트에 의해 및 CRISPR Design Tool (<http://crispr.mit.edu/>)에 의해 예측하였다. 모든 가능한 gRNA 표적 부위 (도 15)를 사용하여 GATK에 의해 확인된 각각의 삽입결실 주변으로 ± 300 bp 부위를 지도로 그렸다. 인간 유전체 및 HIV-1 유전체에서 중복된 부위의 위치를 대조 C1 및 실험 AB7 사이에 비교하였다.

[0146] 통계 분석. 정량 데이터는 3-5개의 독립적인 실험으로부터 평균 ± 표준 편차를 나타내었고, 스튜던트 *t*-시험 또는 ANOVA 및 Newman-Keuls 다중 비교 시험으로 평가하였다. <0.05 또는 0.01인 *p* 값은 통계학적으로 유의한 차이로 간주하였다.

[0147] 실시예 2: Cas9/LTR-gRNA는 HIV-1로 잠복성으로 감염된 CHME5 미소아교 세포에서 HIV-1 리포터 바이러스 생산을 억제한다

[0148] 본 발명자는 HIV-1-공략성 가이드 RNA (gRNA)가 특히 처리하기 어려운 표적 집단인 뇌내 HIV-1 저장소로서 기능하는, 잠복성으로-감염된 골수 세포의 유전체로부터 LTR 전사 활성을 없애고 프로바이러스 DNA를 제거하는 능력을 평가하였다. 본 발명자의 전략은 HIV-1 LTR 프로모터 U3 부위를 표적화하는데 집중되었다. 생물정보학적 선별 및 효율/표적-이탈 예측에 의해, 본 발명자는 보존된 전사 인자 결합 부위를 피하여, 숙주 유전자 발현이 변화할 가능성을 최소화하는 네 가지 gRNA 표적 (프로토스페이서; LTR A-D)을 규명하였다(도 5 및 13). 본 발명자는 gRNA A-D에 상보적인 DNA 단편을 인간화된 Cas9 발현 벡터 (A/B는 pX260 내; C/D는 pX330 내)에 삽입하였고 통합된 HIV-1 유전체 활성을 변화시키는 이들의 개별적 및 조합적 능력을 시험하였다. 본 발명자는 먼저 5' 및 3' LTR, 및 증강된 녹색 형광 단백질 (EGFP) 리포터 대체 Gag (pNL4-3-ΔGag-d2EGFP)를 인코딩하는 유전자를

포함하는 싱글 라운드 HIV-1 벡터의 통합된 복제물이 잠복하는 미소아교 세포주 CHME5를 활용하였다. CHME5 세포를 트리코스타틴 A (TSA), 히스톤 탈아세틸효소 저해제로 처리하면, 대부분의 통합된 프로바이러스로부터 전사가 재활성화되며 EGFP 및 잔여 HIV-1 단백질체(proteome)의 발현이 유도된다. gRNA 더하기 Cas9의 발현은 TSA-유도된 EGFP-양성 CHME5 세포의 비율을 확연하게 감소시켰다 (도 1A 및 6). 본 발명자는 *Ce1* I 뉴클레아제-기반 이형이중가닥-특이적 SURVEYOR 어세이를 이용하여 LTR A-D에 대한 삽입/결실 유전자 돌연변이 (삽입결실)를 검출하였다(도 1B 및 6B). 유사하게, 안정하게 포함된 HIV-1 LTR 복제물을 내포하여 *반딧불*-루시페라아제 리포터 유전자를 유도하는, HeLa-유래 TZM-b1 세포에서 LTR C 및 D를 표적하는 gRNA의 발현은, 바이러스 프로모터 활성을 억제하고 (도 7A), SURVEYOR 및 Sanger 서열분석에 의해 입증된 LTR U3 부위 내부의 삽입결실을 유발시켰다(도 7B-D). 게다가, 이들 세포에서 LTR C/D-표적화 gRNA의 조합된 발현은 예측된 302-bp 바이러스 DNA 서열의 절제, 및 잔여 194-bp 단편의 발생을 야기하였다(도 7E-F).

[0149] 혼합된 클론 CHME5 세포에서 LTR-A/B gRNA의 복합적인 발현은 A 및 B 표적 부위 사이에서 190-bp 단편의 결실을 야기하고 다양한 정도로 삽입결실을 유발하였다(도 1C-D). >20 퓨로마이신-선별된 안정한 서브클론 중에서, 본 발명자는 EGFP에 대한 유세포 분석에 의해 측정된 TSA-유도 HIV-1 프로바이러스 비활성화의 완전한 차단을 가지는 세포 집단을 발견하였다 (도 1E). 프로바이러스 유전체에서 EGFP 및 HIV-1 Rev 반응 요소 (RRE)에 대한 PCR-기반 분석은 HIV-1 유전체의 소거를 확인해주었다 (도 1F, G). 게다가, PCR 산물의 서열분석 결과, 전체 5'-3' LTR에 이르는 바이러스 유전체 (5'-3' LTR-spanning viral genome)가 결실되어, 절단 부위 A 및 B 사이의 190-bp 절제를 통해 351-bp 단편이 얻어졌고(도 1G 및 8), 및 LTR-A 및 -B 부위에 각각 175-bp 삽입 및 27-bp 결실로 682-bp 단편이 얻어졌음을 나타내었다(도 8C). 잔여 HIV-1 유전체 (도 1F-H)는 미량의 Cas9/gRNA-음성 세포의 존재를 반영할 수 있다. 이들 결과는 LTR-표적화 Cas9/gRNA A/B가 HIV-1 유전체를 소거하고 잠복성으로 감염된 미소아교 세포에서 이의 비활성화를 차단함을 나타낸다.

[0150] **실시예 3: Cas9/LTR-gRNA는 U1 단세포성 세포에서 잠복성 HIV-1 바이러스를 효율적으로 제거한다**

[0151] 감염된 혈관주위 대식 세포 및 단핵백혈구에 대한 HIV-1 잠복 모델인 전단구 U-937 세포 서브클론 U1은 만성적으로 HIV-1-감염되며 낮은 수준의 항시성 바이러스 유전자 발현 및 복제를 나타낸다. GenomeWalker 지도는 두 개의 통합된 프로바이러스 DNA 복제물을 U1 세포 내 염색체 Xp11-4 (도 2A) 및 2p21 (도 9A)에서 검출하였다. 전체 9709-bp 프로바이러스 HIV-1 DNA 더하기 측면에 부착된 226-bp X-염색체-유래 서열을 나타내는 9935-bp DNA 단편 (도 2A), 및 9709-bp HIV-1 유전체 더하기 이의 측면에 부착된 2-염색체-유래 467-bp를 내포하는 10176-bp 단편 (도 9A, B)을 부모 대조군 또는 텅빈-벡터 (U6-CAG) U1 세포의 긴-범위 PCR 분석에 의해 규명하였다. 226-bp 및 467-bp 단편은 염색체 X 및 2의 다른 복제물로부터의 예측된 조각을 각각 나타내며, 여기에는 통합된 프로바이러스 DNA가 없다. LTR-A/B gRNA 및 Cas9를 발현하는 U1 세포에서, 본 발명자는 염색체 X 내 833 및 670 bp의 두 개의 추가적인 DNA 단편 및 염색체 2 내 하나의 추가적인 1102-bp 단편을 발견하였다. 따라서, gRNA A/B는 Cas9가 두 염색체 모두에서 HIV-1 5'-3' LTR에 이르는 바이러스 유전체 (5'-3' LTR-spanning viral genome) 조각을 절제하도록 만든다. 833-bp 단편은 숙주 유전체로부터의 예측된 226-bp 및 LTR-A 부위 주변에 27-bp 결실을 갖는 607-bp 바이러스 LTR 서열을 포함한다(도 2A-B). 670-bp 단편은 양쪽 LTR의 gRNA-A/B-유도 절단으로 야기된(도 2A) 190-bp 단편 절제 후에 226-bp 숙주 서열 및 잔여 444-bp 바이러스 LTR 서열을 포함했다(도 1D). 환형 LTR 통합을 통해서이 이것이 부모 U1 세포에 부재했기 때문에 추가적인 단편이 발생하지 않았고, 그러한 환형 LTR 바이러스 유전체 배열은 HIV-1 감염 후 즉시 발생하지만 수명이 짧으며 반복되는 계대배양에 취약하다. 이들 세포는 실질적으로 감소된 HIV-1 바이러스 로드를 나타내었고, 기능적 p24 ELISA 복제 어레이 (도 2C) 및 실시간 PCR 분석 (도 9C, D)에서 보여진다. 검출가능하지만 낮은 잔여 바이러스 로드 및 비활성화는 세포 집단 이종성 및/또는 불완전한 유전체 편집에서 야기될 수 있다. 본 발명자는 또한 통합된 HIV-R7/E-EGFP가 숨어있는 잠복성으로 감염된 J-Lat T 세포에서 Cas9/LTR-A/B gRNA에 의한 HIV-1 유전체의 고갈을 유세포 분석기 분석, SURVEYOR 어레이 및 PCR 유전형검증 (도 10)을 통하여 확인하였고, 이는 Cas9/gRNA 및 ZFN에 의한 Jurkat T 세포 내 HIV-1 프로바이러스 결실에 관한 기존 보고 결과를 지지한다. 모두 고려하면, 본 발명자들의 결과는 복합 LTR-gRNA/Cas9 시스템이 인간 잠복성 HIV-1 감염증에서 전형적인 잠복성으로 HIV-1-감염된 "저장소" (미소아교, 단세포성 및 T) 세포에서, 및 HIV-1 전사 및 비활성화를 검출하는데 있어서 고도로 민감한 TZM-b1 세포에서 효율적으로 HIV-1 복제 및 비활성화를 억제함을 시사한다. 5'- 및 3'-LTR을 표적하는 단일 또는 복합 gRNA는 전체 HIV-1 유전체를 효과적으로 소거하였다.

[0152] **실시예 4: Cas9 더하기 LTR-A/B의 안정적인 발현은 새로운 HIV-1 감염으로부터 TZM-b1 세포를 면역화한다**

[0153] 다음으로 본 발명자는 조합된 Cas9/LTR gRNA가 세포를 HIV-1 감염에 대해 면역화할 수 있는지 안정한 Cas9/gRNA-A 및 -B-발현 TZM-b1-기반 클론을 이용하여 시험하였다 (도 3A). 7가지의 퓨로마이신-선택된 서브클

론 중 두 가지가 190-bp LTR-A/B 부위에 이르는 DNA 단편 (LTR-A/B site-spanning DNA fragment)의 효율적인 절제를 나타내었다 (도 3B). 그러나, 나머지 5가지 서브클론은 Sanger 서열분석으로 검증하여 절제가 없고 (도 3B) 삽입결실 돌연변이가 없음을 나타내었다. Cas9 및 U6-LTR을 표적하는 프라이머를 이용한 PCR 유전형검증은 이들 비효과적인 서브클론 중 어느 것도 Cas9/LTR-A/B gRNA 발현 카세트의 통합된 복제물을 보유하지 않았음을 나타내었다. (도 11A, B). 그 결과, 전장 Cas9의 발현은 검출되지 않았다(도 11C, D). Cas9/LTR-A/B gRNA의 장기 발현은 세포 성장 또는 생존능에 유해한 영향을 주지 않았으며, 이는 상기 모델에서 숙주 유전체 또는 Cas9-유도 독성을 가지는 표적-이탈 방해가 낮게 발생함을 암시한다. 본 발명자는 세포를 VSVG-위형 pNL4-3-ΔE-EGFP 리포터 바이러스로 감염시킴으로써 신규(*de novo*) HIV-1 복제를 평가하였고, 여기서 유세포 분석에 의한 EGFP-양성은 HIV-1 복제를 나타낸다. 대조 U6-CAG 세포와 달리, Cas9/gRNA LTR-A/B를 안정적으로 발현하는 세포는 감염 후 2일에 HIV-1 복제를 지원하지 못했고, 이는 이들이 신규한 HIV-1 감염에 대해 효과적으로 면역화되었음을 나타낸다 (도 3C-D). 자연적인 T-국한(T-tropic) X4 계통 pNL4-3-ΔE-EGFP 리포터 바이러스 (도 12A) 또는 자연적인 M-국한(M-tropic) R5 계통 가령 SF162 및 JRFL (도 12B-D)로 감염된 Cas/LTR-A/B gRNA 발현 세포에서 HIV-1에 대한 유사한 면역성이 관찰되었다.

[0154] 실시예 5: 인간 유전체에 대한 Cas9/LTR-A/B의 표적-이탈 효과

[0155] 개입 접근법으로서의 Cas9/gRNA의 장점은 이의 고도로 특이적인 표적-적중 삽입결실-발생 절단에 있지만, 복합적인 gRNA는 잠재적으로 숙주 유전체 돌연변이유발 및 염색체 장애, 세포독성, 유전독성, 또는 발암을 야기할 수 있다. 상당히 낮은 바이러스-인간 유전체 상동성은 이러한 위험을 감소시키지만, 인간 유전체는 수많은 내인성 레트로바이러스 유전체를 내포하므로 잠재적으로 HIV-1-공략성 gRNA에 취약하다. 그러므로, 본 발명자는 인간 유전체에 대한 선택된 HIV-1 LTR gRNA의 표적-이탈 효과를 평가하였다. 프로토스페이서-인접 모티프 (PAM) 부위 (NGG)에 가장 가까운 12-14-bp 시드 서열이 절단 특이성을 위해 중요하기 때문에, 본 발명자는 >14-bp 시드+NGG를 검색하였고, LTR gRNA A-D에 의한 표적-이탈 후보 부위를 찾지 못했다 (도 13). 점점 더 짧은 gRNA 조각이 상응하는 표적-적중 서열에 대해 100% 일치하는 표적-이탈 절단 부위를 더 많이 얻었다는 사실은 놀랍지 않다 (즉, NGG+13bp는 6, 0, 2 및 9개 표적-이탈 부위를 각각 얻은 반면, NGG+12bp는 16, 5, 16 및 29개를 얻었다; 도 13). 인간 유전체 DNA로부터 본 발명자는 고성능 PCR을 이용하여 예측된 표적-이탈 부위 중 하나를 포괄하는 500-800-bp 서열을 얻었고, 가능한 돌연변이를 SURVEYOR 및 Sanger 서열분석으로 분석하였다. 본 발명자는 어떠한 돌연변이도 찾지 못했다 (TZM-bI 및 U1 세포에서 대표적인 표적-이탈 부위 #1, 5 및 6를 참조하라; 도 4A).

[0156] 표적-이탈 효과의 위험을 철저히 측정하기 위해, 본 발명자는 안정한 Cas9/gRNA A/B-발현 및 대조 U6-CAG TZM-bI 세포를 이용하여 전체 유전체 서열분석 (WGS)을 수행하였다 (도 4B-D). 본 발명자는 인간 (hg19) 및 HIV-1 유전체를 기준 서열로 가지는 유전체 분석 킷 (GATK, v.2.8.1)를 이용하여 676,105개 삽입결실을 규명하였다. 삽입결실 중에서, 24%는 U6-CAG 대조에서 발생했고, 26%는 LTR-A/B 서브클론에서 발생했고, 50%가 양쪽 모두에서 발생했다 (도 4B). 그러한 실질적인 샘플-내 삽입결실 호출 차이는 가능한 표적-이탈 효과를 암시하지만, 아마도 이의 제한된 신뢰성, 제한된 WGS 적용범위 (15-30X), 및 세포 이종성 때문일 것이다. GATK는 오직 신뢰성-규명된 삽입결실만 보고하였다: 일부는 U6-CAG 대조에서 발견되었지만 LTR-A/B 서브클론에는 없었고, 다른 일부는 LTR-A/B에 있었으나 U6-CAG에 없었다. 본 발명자는 양쪽 샘플에서 없어진 수많은 삽입결실 호출을 제한된 WGS 적용범위 때문이라고 예상했다. 그러한 제한된 삽입결실 호출 신뢰도는 또한 잘못된 부정의 가능성을 의미한다: 없어진 삽입결실은 LTR-A/B에 발생하나 U6-CAG 대조에는 없었다. 세포 이종성은 Cas9/gRNA 편집 효율의 가변성 및 계대배양의 영향을 반영할 수 있다. 그러므로, 본 발명자는 HIV-1 유전체의 LTR-A/-B-표적화 부위 및 숙주 유전체의 예상/가능한 gRNA 표적-이탈 부위에 대한 각각의 삽입결실의 측면에 부착된 ±300 bp를 분석함으로써, 각각의 삽입결실이 LTR-A/B gRNA-유도된 것인지 시험하였다 (도 15). 시드 (12-bp) 더하기 NRG를 내포하는 서열에 100% 일치하는 서열을 위해, 본 발명자는 676,105개 삽입결실에 대한 92개의 가능한 표적-이탈 부위 중 단지 8개의 중복된 부위를 규명하였다: 6개의 삽입결실은 두 샘플 모두에서 발생했고, 2개는 U6-CAG 대조에서만 발생했다 (도 4C, D). 본 발명자는 또한 LTR-A/B 서브클론에서만 발생하고, 예상한 대로, U6-CAG에서는 발생하지 않는 HIV-1 LTR 상의 2 개 삽입결실을 규명하였다. 상기 결과는 LTR-A/B gRNA가 언급된 표적-적중 삽입결실을 유도하지만, 표적-이탈 삽입결실은 유도하지 않음을 시사하며, 이는 예상/가능한 표적-이탈 부위를 포괄하는 PCR 산물의 세밀한 서열분석을 이용한 이전 발견과 일치한다.

[0157] 본 발명자의 복합적 접근법은 표적-이탈 효과를 최소화함과 동시에 유전체에 통합된 HIV-1 프로바이러스를 고효율로 및 완전하게 소거하였다. 외부 바이러스 유전체 및 내인성 레트로바이러스 DNA를 비롯한 숙주 세포 유전체 사이에 극도로 낮은 상동성 외에, 본 발명의 연구에서 특징적인 핵심 설계는: 표적-이탈 인간 전사체 또는 (헬

원 드물게는) 미번역된-유전체 부위를 배제하고; HIV-1 LTR 프로모터 (가능하게는 숙주 유전체 내에 보존된 것) 내부의 전사 인자 결합 부위를 피하기 위해 가장 짧은 12-bp+NGG 표적-선별 기준을 이용하는 생물정보학적 선별; LTR-A- 및 -B-공략성, 30-bp gRNA 및 또한 특이성/효율을 증가시키기 위해 본래의 박테리아 면역 메커니즘을 반영하는 전구-crRNA 시스템 vs. 20-bp gRNA-, 키메라 crRNA-tracrRNA-기반 시스템 및 WGS의 선별, 가능한 표적-이탈 효과를 규명하고 배제하기 위한 Sanger 서열분석 및 SURVEYOR 어세이를 포함한다. 실제로, 신규하게 개발된 Cas9 이중-틈생성 및 RNA-유도된 *FokI* 뉴클레아제의 사용은 감소된 표적-이탈 효과를 가지며 HIV-1의 다양한 보존된 부위 내부에서 신규한 표적의 식별을 더욱 도울 수 있다.

[0158] 본 결과는 HIV-1 Cas9/gRNA 시스템이 상이한 염색체에 위치한 LTR의 하나 초과의 복제물을 표적하는 능력을 가짐을 나타내며, 이는 이러한 유전체 편집 시스템이 다중 프로바이러스 DNA가 숨어있는 잠복성으로 감염된 환자의 세포에서 HIV-1의 DNA 서열을 바꿀 수 있음을 시사한다. 본 기술의 높은 편집 효율 및 일관성을 더욱 보장하기 위해, 환자 샘플 내 HIV-1을 소거하기 위한 표적으로서 HIV-1 유전체의 가장 안전한 부위를 고려할 수 있고, 여기에는 단 하나의 HIV-1 계통만 숨어있지 않을 수 있다. 대안적으로, 치료적 Cas9/gRNA 분자를 조작하기 이전에 환자-유래 바이러스 유전체의 세밀한 서열분석으로부터의 데이터를 기반으로 하여 맞춤형 치료 요법을 개발할 수 있다.

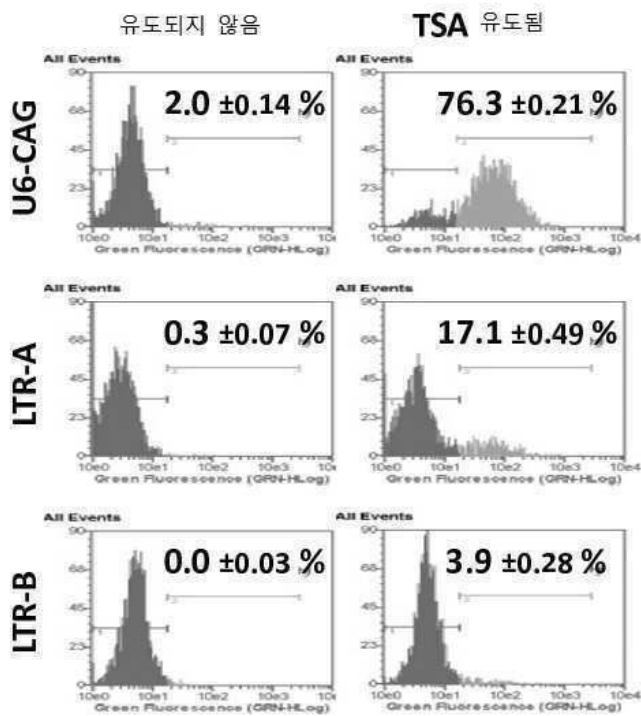
[0159] 본 결과는 또한 HIV-1 감염에 대해 세포를 면역화하기 위해 Cas9/gRNA 유전체 편집이 사용될 수 있음을 보여준다. 예방적 면역화는 어떻게 바이러스가 감염된 세포에 침입하는지 상관없이 시스템이 유전체 서열을 표적하기 때문에 HIV-1 계통의 다양성과 관계없다. 세포 내 Cas9/gRNA 시스템의 선존재는 새로운 HIV-1이 숙주 유전체로 통합되기 전에 이의 빠른 제거를 야기한다. 고위험 개체를 면역화하기 위해 Cas9/LTR-gRNA를 송달하기 위한 다양한 시스템, 예컨대, HIV-1 감염을 제거하기 위한 유전자 요법(바이러스 벡터 및 나노입자) 및 자가 Cas9/gRNA-변형된 골수 줄기/전구세포 세포 또는 유발성 만능 줄기세포의 이식이 연구될 수 있다.

[0160] 여기서, 본 발명자는 HIV-1 표적 유전체를 편집하는데 있어서 Cas9/gRNA의 높은 특이성을 증명하였다. 서브클론 데이터로부터의 결과는 Cas9 및 gRNA 둘다의 존재에 대한 유전체 편집의 엄격한 의존성을 드러내었다. 게다가, 설계된 gRNA 표적 내 단 하나의 뉴클레오타이드 미스매치라도 편집 효율을 저해할 것이다. 또한, 본 발명의 4 개의 설계된 LTR gRNA는 모두 상이한 세포주에 잘 작용했고, 이는 편집이 숙주 세포 유전체보다 HIV-1 유전체에서 더 효율적임을 나타내며, 여기서 설계된 gRNA가 기능적인 것은 아닌데, 이는 상이한 후생적 조절, 가변 유전체 접근성, 또는 다른 이유로 인한 것일 수 있다. Cas9/gRNA 개발의 용이성 및 빠른 속도를 감안하면, 만약 HIV-1 돌연변이가 하나의 Cas9/gRNA-기반 요법에 대한 내성을 전달한다고 해도, 위에서 기술한 바와 같이, 개별적인 환자를 위한 또다른 맞춤형 요법을 가능하게 하기 위한 HIV-1 변이체의 유전형이 검증될 수 있다.

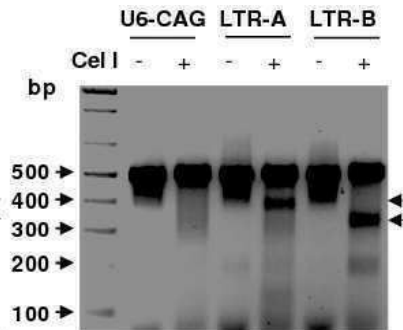
[0161] 본 발명의 수많은 구체예가 기술되었다. 그렇지만, 다양한 변형이 본 발명의 사상 및 범위로부터 벗어나지 않고 만들어질 수 있음이 이해될 것이다. 따라서, 다른 구체예가 하기 청구범위의 범위 내에 있다.

도면

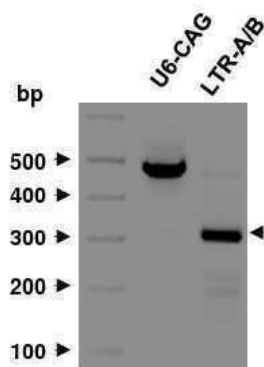
도면1a



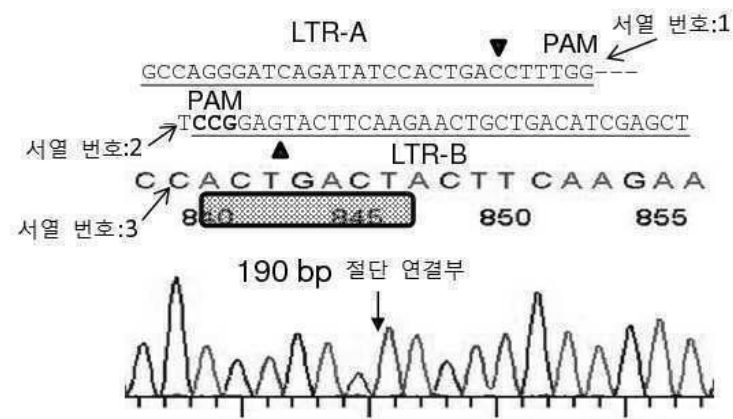
도면1b



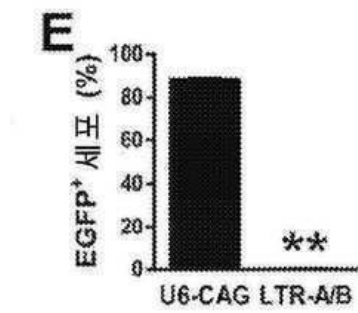
도면1c



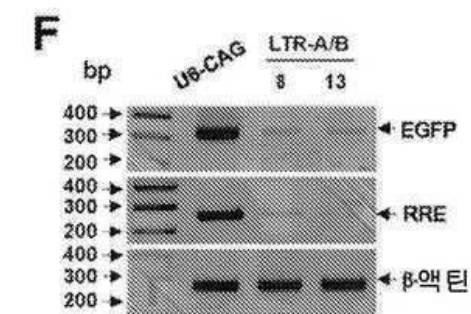
도면1d



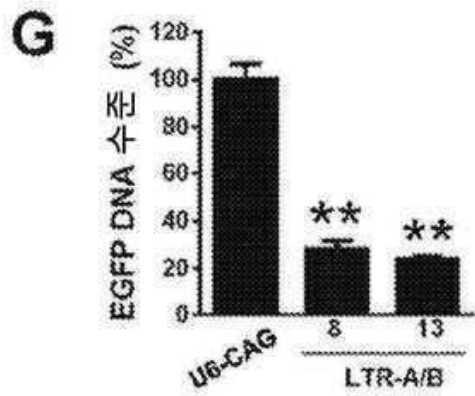
도면1e



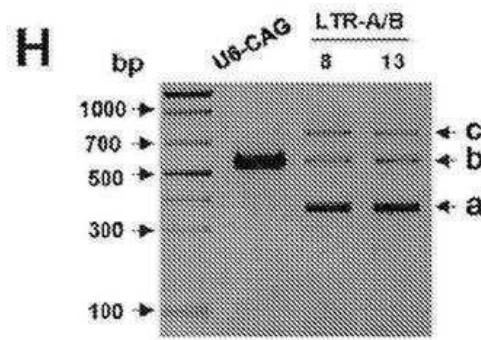
도면1f



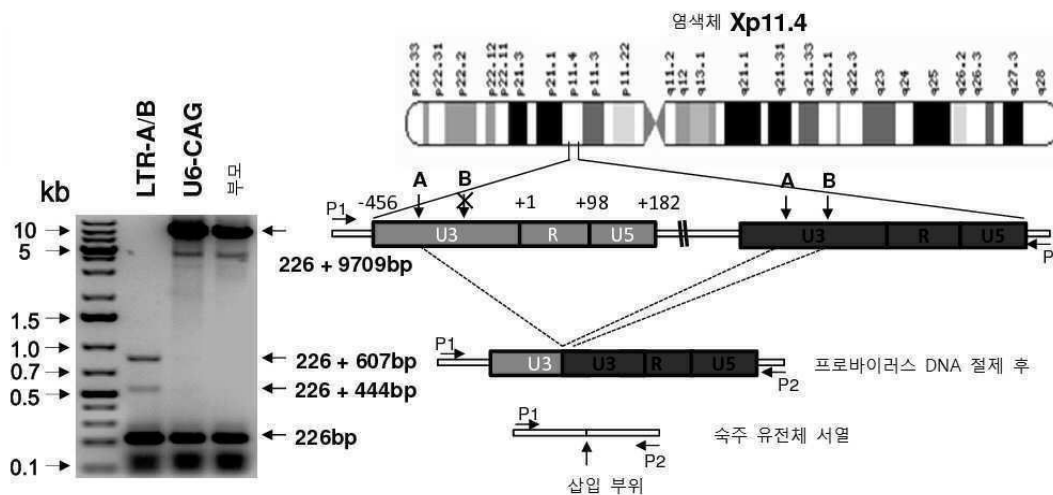
도면1g



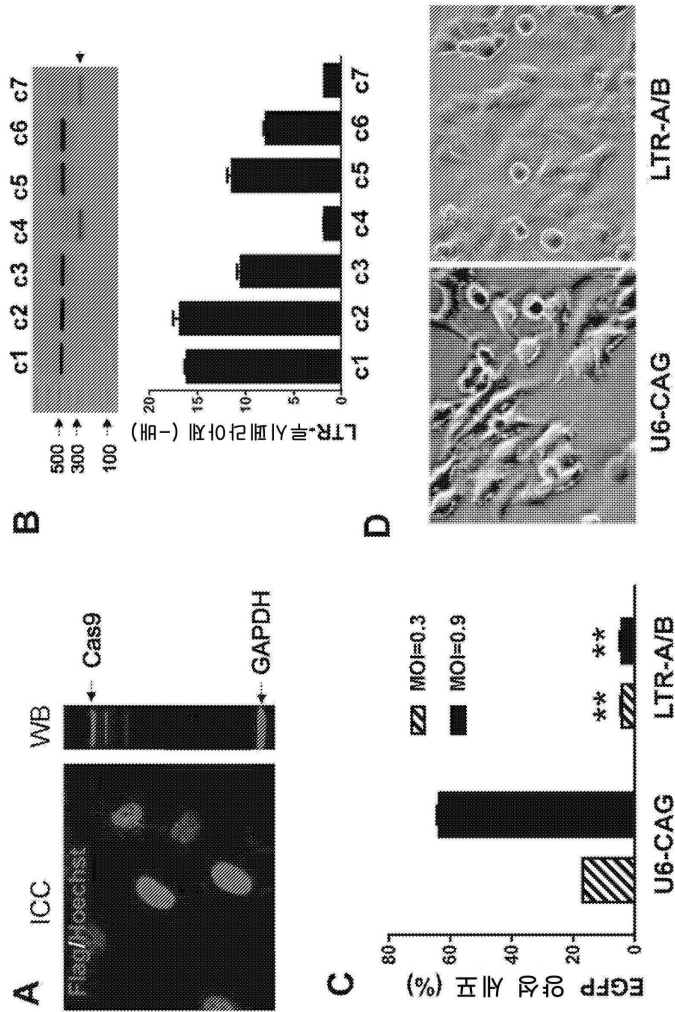
도면1h



도면2a

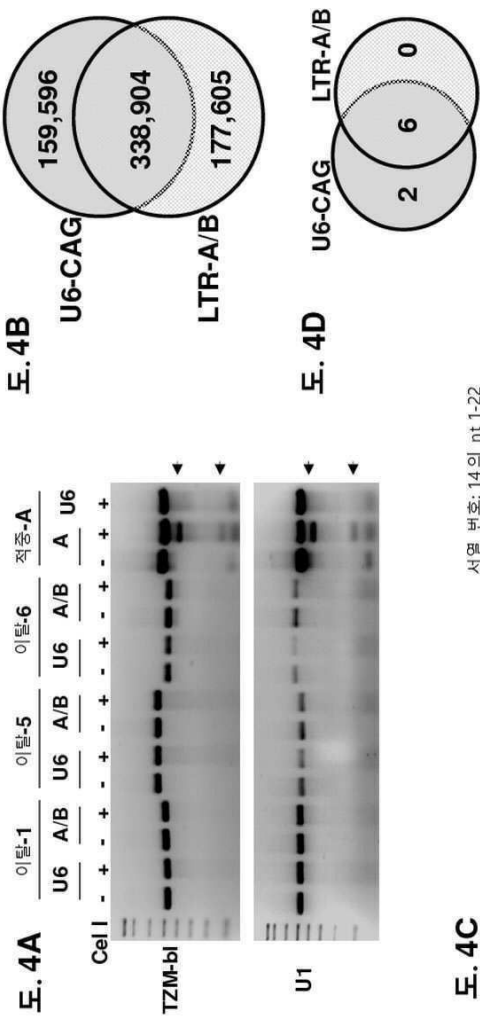


도면3



도 3

도면4



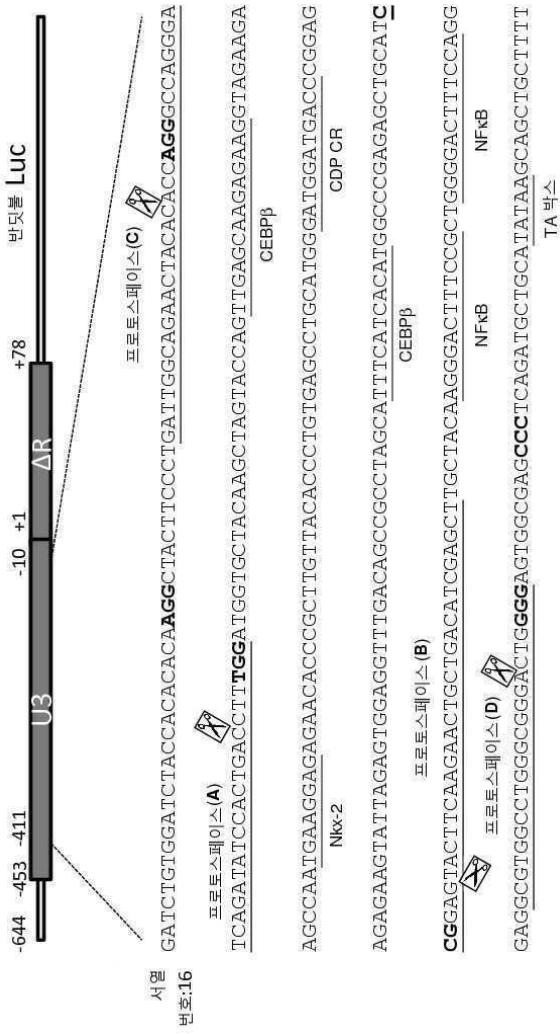
서열 번호: 14의 nt 1-22

도. 4C

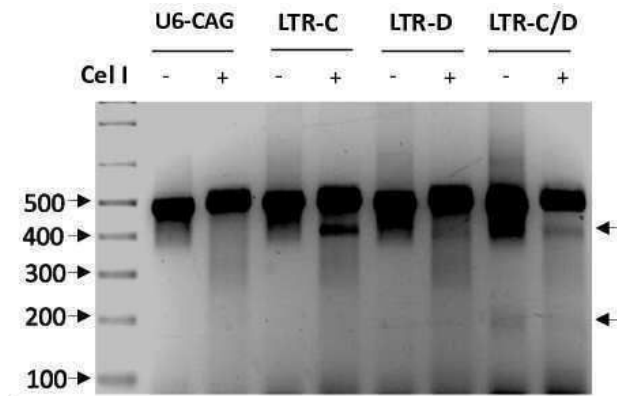
명칭	gRNA 표적 부위				상인결실				유전형	
	유전체	시작	종결	가닥	유전체	시작	종결	특성	돌연변이	U6-CAG LTR-A/B
LTR-A	chr11	103532094	103532108	+	chr11	1,048,408	103532208	ATC	A	1/1 1/1
LTR-A	chr14	24193180	24193194	-	chr14	24192852	24192875	GACATCTGTC TCTATTTTATG		0/1 0/0
LTR-A	chr6	98901053	98901068	+	chr6	98901264	98901287	CACTA	C	1/1 1/1
LTR-A	chr6	98901053	98901068	+	chr6	98901280	98901282	CTA	G	1/1 1/1
LTR-A	chr7	72210628	72210642	-	chr7	72210627	72210628	GT	G	0/1 0/1
LTR-A	chr7	72210628	72210642	-	chr7	72210629	72210632	GTTT	G	0/1 0/0
LTR-A	chr9	95871690	95871704	+	chr9	95871594	95871596	CTA	C	0/1 0/1
LTR-A	HIV	162	184	+	HIV	371	375	CTACA	C	0/0 0/1
LTR-B	chr15	82132445	82132459	+	chr15	82132704	82132704	T	TTTGG, TTTTGTG	1/1 1/1
LTR-B	HIV	9291	9313	-	HIV	9616	9616	A	TTTCTTTTCTTCTCTCT	0/0 0/1

서열 번호:15

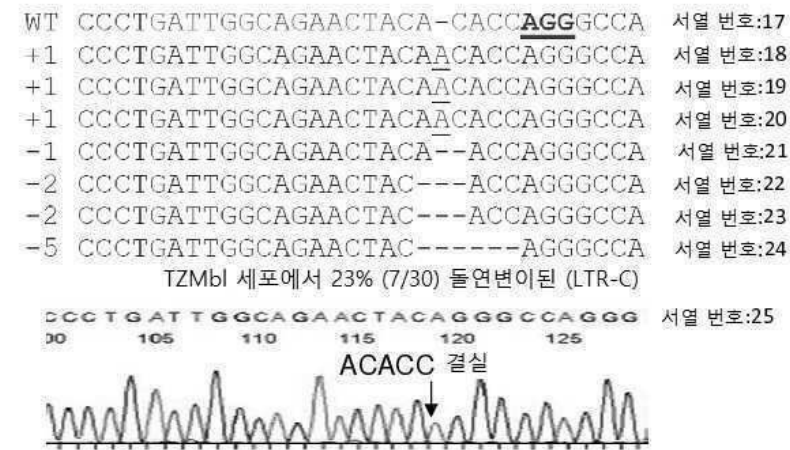
도면5



도면7b



도면7c

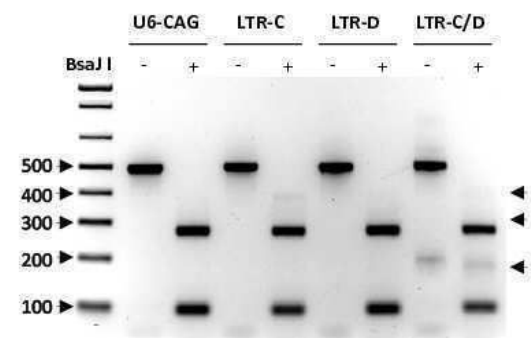


도면7d

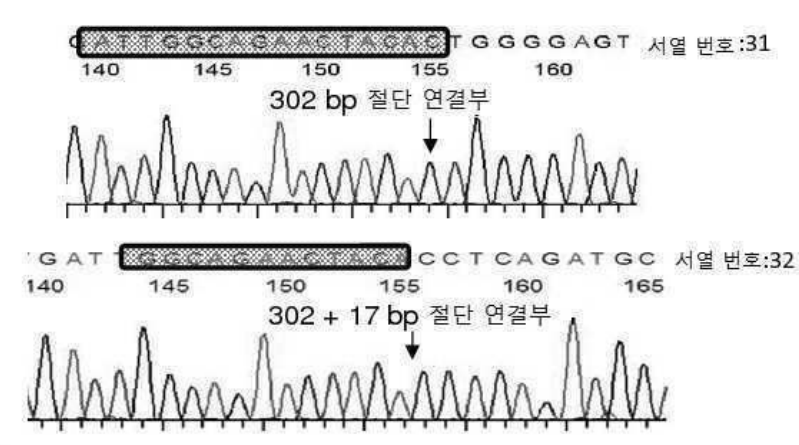
서열 번호:26 WT GACTTTCAGGGAGGCGTGGCCTGGGCGGG-A-CTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCGGTGAAGCCGAATTC
 서열 번호:27 A (g) GACTTTCAGGGAGGCGTGGCCTGGGCGGG-A-CTGGGGGGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCGGTGAAGCCGAATTC
 서열 번호:28 +2 GACTTTCAGGGAGGCGTGGCCTGGGCGGGTATCTGGGGAGTGGCGAGCCCTCAGATGCTGCATATAAGCAGCGGTGAAGCCGAATTC
 서열 번호:29 -1 GACTTTCAGGGGGGCGTGGCCTGGGCGGG-A-CTGGGGAGTGGCGAGCCCTCAGATGCTGCATA-AAGCAGCGGTGAAGCCGAATTC
 서열 번호:30 -63 GACTTTCAGG-----GAAGCCGAATTC

TZMbl 세포에서 13% (4/30) 돌연변이된 (LTR-D)

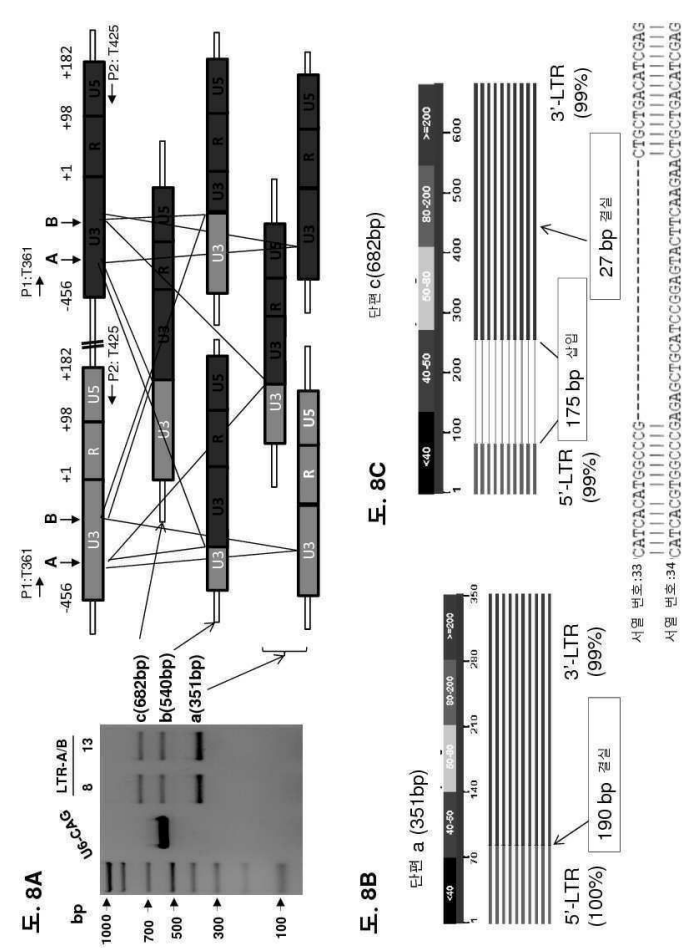
도면7e



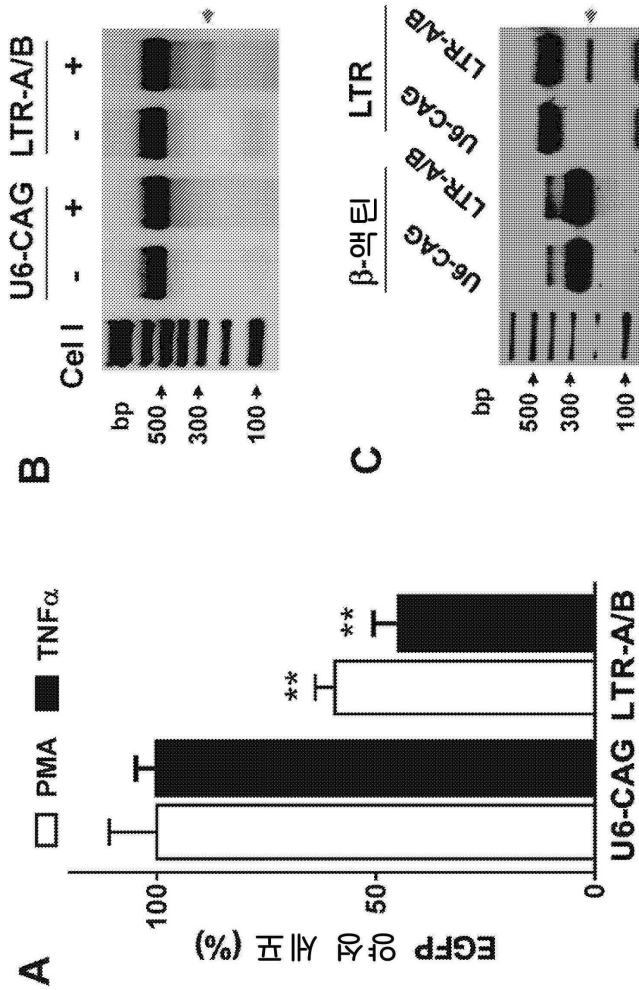
도면7f



도면8

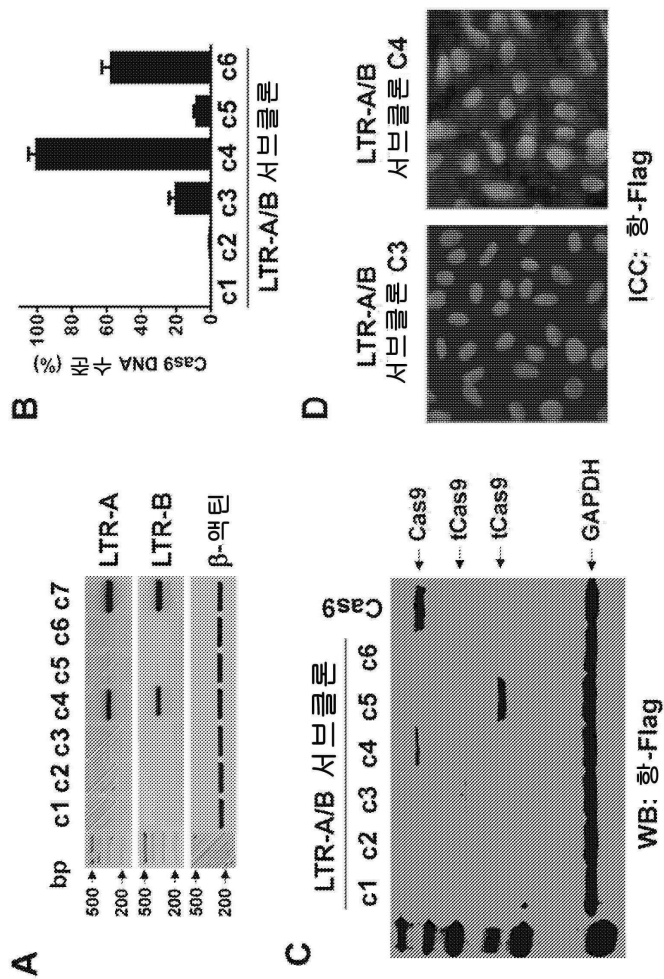


도면10



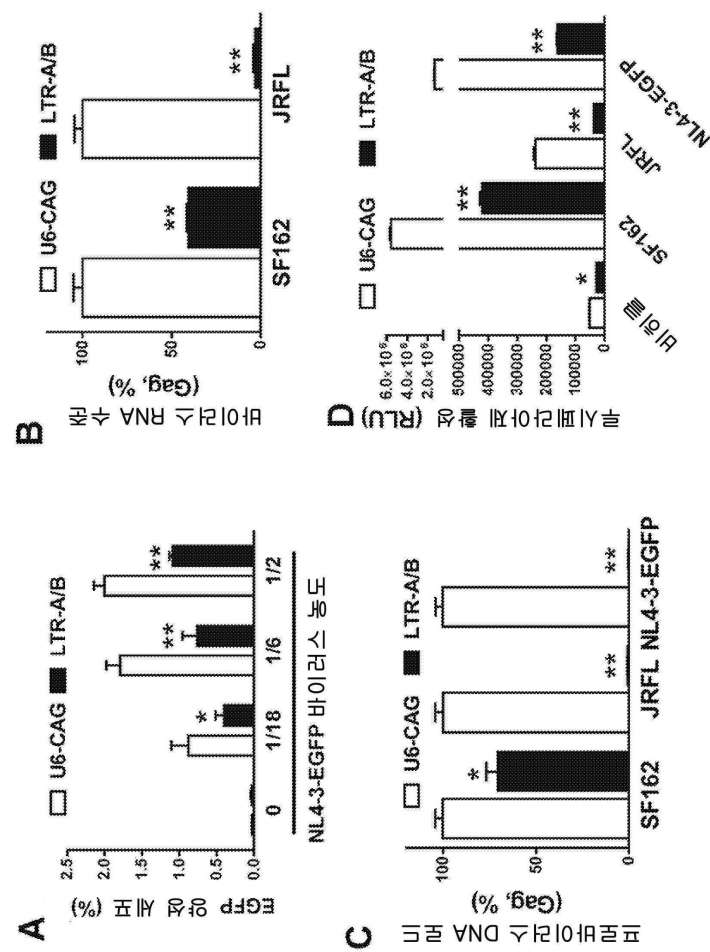
도 10

도면11



도 11

도면12



도 12

도면13a

예측된 LTR gRNA 및 이들의 표적-이탈 수 (100% 일치)											
명칭	번호	gRNA 서열	20	19	18	17	16	15	14	13	12
센스											
서열 번호: 79	1	TCAGACCCCTTTTAGTCAGTGTGG	0	0	0	0	0	0	0	2	25
서열 번호: 80	2	TTGCTTGACTGGGTCTCTCTGG	0	0	0	0	0	0	0	3	10
서열 번호: 81	3	CAGCTGCTTTTGTGCTGTACTGG	0	0	0	0	0	0	2	4	11
서열 번호: 82	4	CTGACATCGAGCTTGCTACAAGG	0	0	0	0	0	0	0	0	17
서열 번호: 83	5	CCGCCTAGCATTTTCATCACATGG	0	0	0	0	0	1	2	9	55
서열 번호: 84	6	CGGAGAGAGAAGTATTAGAGTGG	0	0	0	0	0	0	0	3	18
서열 번호: 85	7	AGTACCAGTTGAGCAAGAGAAGG	0	0	0	0	0	0	0	1	18
서열 번호: 86	8	GATATCCACTGACCTTTGGATGG	0	0	0	0	0	1	3	22	211
서열 번호: 87	LTR-C 9	GATTGGCAGAACTACACACCAGG	0	0	0	0	0	0	0	2	12
서열 번호: 88	10	CACAAGGCTACTTCCCTGATTGG	0	0	0	0	0	0	0	2	11
서열 번호: 89	11	CTGTGGATCTACCACACACAAGG	0	0	0	0	0	2	5	12	37
서열 번호: 90	12	TGGGAGCTCTCTGGCTAACTAGG	0	0	0	0	0	0	0	4	14
서열 번호: 91	13	GGTTAGACCAGATCTGAGCCTGG	0	0	0	0	0	0	0	9	33
서열 번호: 92	14	TGCTACAAGGGACTTTCGCTGG	0	0	0	0	0	0	0	2	5
서열 번호: 93	15	AGAGAGAAGTATTAGAGTGGAGG	0	0	0	0	0	0	5	7	16
서열 번호: 94	16	TTACACCCTGTGAGCCTGCATGG	0	0	0	0	0	0	3	14	35
서열 번호: 95	17	AAGGTAGAAGAAGCCAATGAAGG	0	0	0	0	0	0	4	36	75
서열 번호: 96	LTR-A 18	TCAGATATCCAGTGAAGTTGG	0	0	0	0	0	0	0	0	16
서열 번호: 97	19	GACAAGATATCCTTGATCTGTGG	0	0	0	0	0	0	0	0	10
서열 번호: 98	20	GCCCGTCTGTTGTGTGACTCTGG	0	0	0	0	0	0	3	7	35
서열 번호: 99	21	ATCTGAGCCTGGGAGCTCTCTGG	0	0	0	0	1	2	9	32	78
서열 번호: 100	22	CTTTCCGCTGGGGACTTTCAGG	0	0	0	0	0	0	0	0	4
서열 번호: 101	23	CAGAACTACACACCAGGGCCAGG	0	0	0	0	0	0	0	3	20
서열 번호: 102	24	CCTGCATGGGATGGATGACCCGG	0	0	0	0	2	2	2	5	21

도면13b

도 13 계속											
명칭	번호	gRNA 서열	20	19	18	17	16	15	14	13	12
센스											
서열 번호: 103	25	CCCTGTGAGCCTGCATGGGATGG	0	0	0	0	1	1	2	9	30
서열 번호: 104	26	CTTCCAGGGAGCGTGGCCTGG	0	0	0	0	0	8	15	32	75
서열 번호: 105	27	GGGACTTTCAGGGAGGCGTGG	0	0	0	0	0	0	0	5	24
서열 번호: 106	28	CCGCTGGGGACTTTCAGGGAGG	0	0	0	0	0	1	3	9	25
서열 번호: 107	29	CATGGCCCCGAGAGCTGCATCCGG	0	0	0	0	0	0	0	0	16
서열 번호: 108	30	GCCTGGGCGGGACTGGGAGTGG	0	0	0	0	1	2	7	18	250
서열 번호: 109	31	AGGCGTGGCCTGGGCGGGACTGG	0	0	0	0	0	0	0	2	6
서열 번호: 110	LTR-D 32	CCGTGGCCTGGGCGGGACTGGGG	0	0	0	0	0	0	0	0	29
서열 번호: 111	33	CCAGGGAGGCGTGGCCTGGGCGG	0	0	0	0	2	2	2	11	22
안티센스											
서열 번호: 112	1	TGTGGTAGATCCACAGATCAAGG	0	0	0	0	0	0	0	3	13
서열 번호: 113	2	GGTGTGTAGTTCTGCCAATCAGG	0	0	0	0	0	0	0	8	11
서열 번호: 114	3	GTCAGTGGATATCTGATCCCTGG	0	0	0	0	0	0	0	2	10
서열 번호: 115	4	TAGCACCATCCAAAGGTCAGTGG	0	0	0	0	0	0	2	3	10
서열 번호: 116	5	TAGCTTGTAGCACCATCCAAAGG	0	0	0	0	0	1	1	6	12
서열 번호: 117	6	TCTACCTTCTCTTGTCTCAACTGG	0	0	0	0	0	0	1	1	6
서열 번호: 118	7	CACTCTAATACTTCTCTCTCCGG	0	0	0	0	0	0	0	1	26
서열 번호: 119	8	CCATGTGATGAAATGCTAGGCGG	0	0	0	0	0	1	1	5	17
서열 번호: 120	9	GGGCCATGTGATGAAATGCTAGG	0	0	0	0	0	0	0	3	48
서열 번호: 121	LTR-B 10	TAGCAGTCTTGAAGTACTCCGG	0	0	0	0	0	0	0	0	0
서열 번호: 122	11	CTGCTTATATGCAGCATCTGAGG	0	0	0	0	0	0	0	4	39
서열 번호: 123	12	CACACTACTTGAAGCACTCAAGG	0	0	0	0	0	0	0	3	19
서열 번호: 124	13	TACCAGAGTCACACAACAGACGG	0	0	0	0	0	0	0	2	12

도면13c

도 13 계속

명칭	번호	gRNA 서열	20	19	18	17	16	15	14	13	12
안티센스											
서열 번호: 125	14	ACACTGACTAAAAGGGTCTGAGG	0	0	0	0	1	2	3	4	15
서열 번호: 126	15	CAAGGATATCTTGTCTTCGTTGG	0	0	0	0	0	0	0	1	5
서열 번호: 127	16	CAGGGAAGTAGCCTTGTGTGTGG	0	0	0	0	0	1	2	4	17
서열 번호: 128	17	GCGGGTGTCTCTCCTTCATTGG	0	0	0	0	0	2	2	15	49
서열 번호: 129	18	TAGTTAGCCAGAGAGCTCCCAGG	0	0	0	0	0	2	4	24	93
서열 번호: 130	19	CTTTATTGAGGCTTAAGCAGTGG	0	0	0	0	0	0	0	4	25
서열 번호: 131	20	ACTCAAGGCAAGCTTTATTGAGG	0	0	0	0	0	0	0	1	28
서열 번호: 132	21	GGATATCTGATCCCTGGCCCTGG	0	0	0	0	0	1	2	8	43
서열 번호: 133	22	GGCTCACAGGGTGTAACAAGCGG	0	0	0	0	0	0	0	2	5
서열 번호: 134	23	TCCATCCCATGCAGGCTCACAGG	0	0	0	0	0	3	3	8	20
서열 번호: 135	24	AGTACTCCGGATGCAGCTCTCGG	0	0	0	0	0	0	1	5	48
서열 번호: 136	25	AGAGCTCCCAGGCTCAGATCTGG	0	0	0	0	0	0	4	15	38
서열 번호: 137	26	GATTTTCCCACTGACTAAAAGG	0	0	0	0	0	0	0	8	21
서열 번호: 138	27	CCGGTTCATCCATCCCATGCAGG	0	0	0	0	0	0	0	4	36
서열 번호: 139	28	CCTCCCTGGAAAGTCCCCAGCGG	0	0	0	0	0	0	3	14	37
서열 번호: 140	29	GCCACTCCCCAGTCCCGCCAGG	0	0	0	0	0	0	1	4	14
서열 번호: 141	30	CCGCCAGGCCACGCCCTCCCTGG	0	0	0	0	0	1	1	6	19

도면14a

표적 명칭	방향	서열 (5'에서 3'으로)	서열 번호
LTR-A	T353: 센스	aaacAGGGCCAGGGATCAGATATCCACTGACCTTgt	서열 번호: 36
	T354: 안티센스	taaacAAGGTCAGTGGATATCTGATCCCTGGCCCT	서열 번호: 37
LTR-B	T355: 안티센스	aaacAGCTCGATGTCAGCAGTTCTTGAAGTACTCgt	서열 번호: 38
	T356: 센스	taaacGAGTACTTCAAGAACTGCTGACATCGAGCT	서열 번호: 39
LTR-C	T357: 센스	caccGATTGGCAGAACTACACACC	서열 번호: 40
	T358: 안티센스	aaacGGTGTGTAGTTCTGCCAATC	서열 번호: 41
LTR-D	T359: 센스	caccGCGTGGCCTGGGCGGGACTG	서열 번호: 42
	T360: 안티센스	aaacCAGTCCC GCCCAGGCCACGC	서열 번호: 43
PCR 프라이머			
LTR -453/F	센스	TGGAAGGGCTAATTCACCTCCCAAC	서열 번호: 44
LTR +43/R	안티센스	CCGAGAGCTCCCAGGCTCAGATCT	서열 번호: 45
LTR -411/F	T361: 센스	caccGATCTGTGGATCTACCACACACA	서열 번호: 46
LTR +129/R	T363: 안티센스	aaacGAGTCACACAACAGACGGGC	서열 번호: 47
Cas-hU6/5'/XhoBm	T351: 센스	cgcctcgaggatccGAGGGCCTATTTCCCATGATTCC	서열 번호: 48
Cas-CAG/3'/EcoR	T352: 안티센스	tgtgaattcAGGCGGGCCATTACCGTAAGTTATG	서열 번호: 49
U1-염색체 X	T485: 센스	ACGACTATCTTATCAATCCTTCCTG	서열 번호: 50
	T486: 안티센스	CTAGGTGATTAGGATATTCTACAATC	서열 번호: 51
	T492: 센스	GCTATTGTATCTGATCACAAGCTG	서열 번호: 52
U1-염색체 2	T493: 안티센스	TTGATTGTGTGTCCAGGTCCTAGG	서열 번호: 53
	T494: 센스	GCAAGGGCGAGGAGCTGTTCCACC	서열 번호: 54
d2EGFP	T495: 안티센스	TTGTAGTTCGGTCGTCCTTGAAG	서열 번호: 55
	T457: 센스	AATGGTACATCAGGCCATATCAC	서열 번호: 56
Gag	T458: 안티센스	CCCACTGTGTTTAGCATGGTATT	서열 번호: 57
	T477: 센스	CACAGCATCAAGAAGAACCCTGAT	서열 번호: 58
Cas9	T491: 안티센스	TCTTCCGCTCGGIGTATCTTCTTC	서열 번호: 59
	센스	CGCCAAGCTTGAATAGGAGCTTTGTTC	서열 번호: 60
RRE	안티센스	CTAGGATCCAGGAGCTGTTGATCCTTTAGG	서열 번호: 61

도면14b

도 14 계속

표적 명칭	방향	서열 (5'에서 3'으로)	서열 번호
표적-이탈 (OT)			
LTR-A-OT-1	T465: 센스	GTGGACTTTGGATGGTGAGATAG	서열 번호: 62
	T466: 안티센스	GCCTGGCAAGAGTGAAGTGAATC	서열 번호: 63
LTR-A-OT-2	T467: 센스	AAGATAATGAGTTGTGGCAGAGC	서열 번호: 64
	T468: 안티센스	TCTACCTGGTAATCCAGCATCTGG	서열 번호: 65
LTR-A-OT-3	T469: 센스	ATAGGAGGAAGGCACCAAGAGGG	서열 번호: 66
	T470: 안티센스	AATGATGCTTTGGTCCTACTCCT	서열 번호: 67
LTR-A-OT-4	T471: 센스	TGCTCTTGCTACTCTGGCATGTAC	서열 번호: 68
	T472: 안티센스	AATCTACCTCTGAGAGCTGCAGG	서열 번호: 69
LTR-A-OT-5	T473: 센스	TCAGACACAGCTGAAGCAGAGGC	서열 번호: 70
	T474: 안티센스	ATGCCAGTGTGAGTATGTCAG	서열 번호: 71
LTR-A-OT-6	T475: 센스	TCAAGATCAGCCAGAGTGCACATG	서열 번호: 72
	T476: 안티센스	TGCTCTTCCGAGCCTCTCTGGAG	서열 번호: 73
기타			
hU6-서열	T428: 센스	ATGGACTATCATATGCTTACCG	서열 번호: 74
LSP1	센스	GCTTCAGCAAGCCGAGTCTGCGTCGAG	서열 번호: 75
LSP2	안티센스	GCTCCTCTGGTTTCCCTTTTCGCTTICAA	서열 번호: 76
AP1	센스	GTAATACGACTCACTATAGGGC	서열 번호: 77
AP2	안티센스	ACTATAGGGCACGCGTGGT	서열 번호: 78

도면15a

명칭	검색 서열	서열 번호:	대상 ID	동일성 (%)	E-값	시작	종결	ORF	기준 서열	서열 번호: (12bp 시드)	미스매치
LTRA	ATCAGATATCCACTGACCTTTGG	142	HIV	100	7.00E-04	162	184	+	ATCAGATATCCACTGACCTTTGG	253	0
LTRA	TCAGATATCCACTGACCTTTGG	143	HIV	100	0.003	163	184	+	TCAGATATCCACTGACCTTTGG	254	0
LTRA	TCAGATATCCACTGACCTTTGG	144	HIV	100	0.003	9091	9112	+	TCAGATATCCACTGACCTTTGG	255	0
LTRA	CAGATATCCACTGACCTTTGG	145	HIV	100	0.009	164	184	+	CAGATATCCACTGACCTTTGG	256	0
LTRA	CAGATATCCACTGACCTTTGG	146	HIV	100	0.009	9092	9112	+	CAGATATCCACTGACCTTTGG	257	0
LTRA	AGATATCCACTGACCTTTGG	147	HIV	100	0.033	165	184	+	AGATATCCACTGACCTTTGG	258	0
LTRA	AGATATCCACTGACCTTTGG	148	HIV	100	0.033	9093	9112	+	AGATATCCACTGACCTTTGG	259	0
LTRA	GATATCCACTGACCTTTGG	149	HIV	100	0.12	166	184	+	GATATCCACTGACCTTTGG	260	0
LTRA	GATATCCACTGACCTTTGG	150	HIV	100	0.12	9094	9112	+	GATATCCACTGACCTTTGG	261	0
LTRA	ATATCCACTGACCTTTGG	151	HIV	100	0.42	167	184	+	ATATCCACTGACCTTTGG	262	0
LTRA	ATATCCACTGACCTTTGG	152	HIV	100	0.42	9095	9112	+	ATATCCACTGACCTTTGG	263	0
LTRA	TATCCACTGACCTTTGG	153	chr5	100	1.5	21926317	21926333	+	TATCCACTGACCTTTGG	264	0
LTRA	TATCCACTGACCTTTGG	154	HIV	100	1.5	168	184	+	TATCCACTGACCTTTGG	265	0
LTRA	TATCCACTGACCTTTGG	155	HIV	100	1.5	9096	9112	+	TATCCACTGACCTTTGG	266	0
LTRA	TATCCACTGACCTTAAG	156	chr3	100	1.5	116712577	116712593	+	TATCCACTGACCTTAAG	267	0
LTRA	TATCCACTGACCTTGAG	157	chr6	100	1.5	32460607	32460623	+	TATCCACTGACCTTGAG	268	0
LTRA	ATCCACTGACCTTAGG	158	chr3	100	5.4	2669092	2669107	+	ATCCACTGACCTTAGG	269	0
LTRA	ATCCACTGACCTTAGG	159	chr3	100	5.4	158293369	158293384	+	ATCCACTGACCTTAGG	270	0
LTRA	ATCCACTGACCTTGGG	160	chr20	100	5.4	46918344	46918359	+	ATCCACTGACCTTGGG	271	0
LTRA	ATCCACTGACCTTGGG	161	chr14	100	5.4	86310067	86310082	-	ATCCACTGACCTTGGG	272	0
LTRA	ATCCACTGACCTTGGG	162	chr5	100	5.4	21926318	21926333	+	ATCCACTGACCTTGGG	273	0
LTRA	ATCCACTGACCTTGGG	163	chr4	100	5.4	95491921	95491936	-	ATCCACTGACCTTGGG	274	0
LTRA	ATCCACTGACCTTTGG	164	HIV	100	5.4	169	184	+	ATCCACTGACCTTTGG	275	0
LTRA	ATCCACTGACCTTTGG	165	HIV	100	5.4	9097	9112	+	ATCCACTGACCTTTGG	276	0
LTRA	ATCCACTGACCTTTGG	166	chr6	100	5.4	98901053	98901068	+	ATCCACTGACCTTTGG	277	0
LTRA	ATCCACTGACCTTAAG	167	chr7	100	5.4	155511293	155511308	-	ATCCACTGACCTTAAG	278	0
LTRA	ATCCACTGACCTTAAG	168	chr3	100	5.4	116712578	116712593	+	ATCCACTGACCTTAAG	279	0
LTRA	ATCCACTGACCTTCAG	169	chr5	100	5.4	152371289	152371304	+	ATCCACTGACCTTCAG	280	0

도면15b

도 15 계속

명칭	검색 서열	서열 번호:	대상 ID	동일성 (%)	E-값	시작	종결	방향	기준 서열	서열 번호:	미스매치 (12bp 시드)
LTRA	ATCCACTGACCTTCAG	170	chr4	100	5.4	110823169	110823184	-	ATCCACTGACCTTCAG	281	0
LTRA	ATCCACTGACCTTGAG	171	chrX	100	5.4	74260260	74260275	+	ATCCACTGACCTTGAG	282	0
LTRA	ATCCACTGACCTTGAG	172	chr6	100	5.4	32460608	32460623	+	ATCCACTGACCTTGAG	283	0
LTRA	TCCACTGACCTTAGG	173	chr12	100	20	14485012	14485026	-	TCCACTGACCTTAGG	284	0
LTRA	TCCACTGACCTTAGG	174	chr7	100	20	72210628	72210642	-	TCCACTGACCTTAGG	285	0
LTRA	TCCACTGACCTTAGG	175	chr6	100	20	160845640	160845654	+	TCCACTGACCTTAGG	286	0
LTRA	TCCACTGACCTTAGG	176	chr3	100	20	2669093	2669107	+	TCCACTGACCTTAGG	287	0
LTRA	TCCACTGACCTTAGG	177	chr3	100	20	158293370	158293384	+	TCCACTGACCTTAGG	288	0
LTRA	TCCACTGACCTTAGG	178	chr2	100	20	237551230	237551244	-	TCCACTGACCTTAGG	289	0
LTRA	TCCACTGACCTTGGG	179	chr20	100	20	46918345	46918359	+	TCCACTGACCTTGGG	290	0
LTRA	TCCACTGACCTTGGG	180	chr14	100	20	86310067	86310081	-	TCCACTGACCTTGGG	291	0
LTRA	TCCACTGACCTTGGG	181	chr12	100	20	116054688	116054702	+	TCCACTGACCTTGGG	292	0
LTRA	TCCACTGACCTTGGG	182	chr11	100	20	103532094	103532108	+	TCCACTGACCTTGGG	293	0
LTRA	TCCACTGACCTTGGG	183	chr10	100	20	132186431	132186445	-	TCCACTGACCTTGGG	294	0
LTRA	TCCACTGACCTTGGG	184	chr8	100	20	144600475	144600489	-	TCCACTGACCTTGGG	295	0
LTRA	TCCACTGACCTTGGG	185	chr5	100	20	21926319	21926333	+	TCCACTGACCTTGGG	296	0
LTRA	TCCACTGACCTTGGG	186	chr4	100	20	95491921	95491935	-	TCCACTGACCTTGGG	297	0
LTRA	TCCACTGACCTTTGG	187	HIV	100	20	170	184	+	TCCACTGACCTTTGG	298	0
LTRA	TCCACTGACCTTTGG	188	HIV	100	20	9098	9112	+	TCCACTGACCTTTGG	299	0
LTRA	TCCACTGACCTTTGG	189	chr16	100	20	86962569	86962583	+	TCCACTGACCTTTGG	300	0
LTRA	TCCACTGACCTTTGG	190	chr11	100	20	68156214	68156228	+	TCCACTGACCTTTGG	301	0
LTRA	TCCACTGACCTTTGG	191	chr6	100	20	98901054	98901068	+	TCCACTGACCTTTGG	302	0
LTRA	TCCACTGACCTTTGG	192	chr5	100	20	72600080	72600094	+	TCCACTGACCTTTGG	303	0
LTRA	TCCACTGACCTTTGG	193	chr5	100	20	136458169	136458183	+	TCCACTGACCTTTGG	304	0
LTRA	TCCACTGACCTTTGG	194	chr4	100	20	25353030	25353044	+	TCCACTGACCTTTGG	305	0
LTRA	TCCACTGACCTTTGG	195	chr2	100	20	207833373	207833387	+	TCCACTGACCTTTGG	306	0
LTRA	TCCACTGACCTTAAG	196	chr15	100	20	67850506	67850520	-	TCCACTGACCTTAAG	307	0
LTRA	TCCACTGACCTTAAG	197	chr7	100	20	155511293	155511307	-	TCCACTGACCTTAAG	308	0
LTRA	TCCACTGACCTTAAG	198	chr5	100	20	25142541	25142555	-	TCCACTGACCTTAAG	309	0

도면15c

도 15 계속

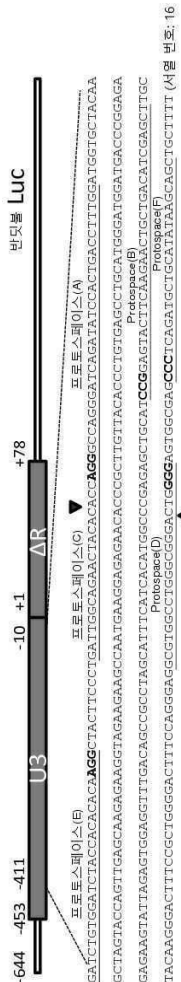
명칭	검색 서열	서열 번호:	대상 ID	동일성 (%)	E-값	시작	종결	방향	기준 서열	서열 번호:	미스매치 (12bp 시드)
LTRA	TCCACTGACCTTAAG	199	chr3	100	20	116712579	116712593	+	TCCACTGACCTTAAG	310	0
LTRA	TCCACTGACCTTAAG	200	chr1	100	20	163298514	163298528	+	TCCACTGACCTTAAG	311	0
LTRA	TCCACTGACCTTCAG	201	chr20	100	20	22136764	22136778	-	TCCACTGACCTTCAG	312	0
LTRA	TCCACTGACCTTCAG	202	chr19	100	20	50519462	50519476	-	TCCACTGACCTTCAG	313	0
LTRA	TCCACTGACCTTCAG	203	chr18	100	20	74623621	74623635	-	TCCACTGACCTTCAG	314	0
LTRA	TCCACTGACCTTCAG	204	chr16	100	20	71402733	71402747	+	TCCACTGACCTTCAG	315	0
LTRA	TCCACTGACCTTCAG	205	chr14	100	20	24193180	24193194	-	TCCACTGACCTTCAG	316	0
LTRA	TCCACTGACCTTCAG	206	chr11	100	20	133664063	133664077	+	TCCACTGACCTTCAG	317	0
LTRA	TCCACTGACCTTCAG	207	chr9	100	20	140394271	140394285	+	TCCACTGACCTTCAG	318	0
LTRA	TCCACTGACCTTCAG	208	chr6	100	20	47685256	47685270	-	TCCACTGACCTTCAG	319	0
LTRA	TCCACTGACCTTCAG	209	chr5	100	20	152371290	152371304	+	TCCACTGACCTTCAG	320	0
LTRA	TCCACTGACCTTCAG	210	chr4	100	20	110823169	110823183	-	TCCACTGACCTTCAG	321	0
LTRA	TCCACTGACCTTCAG	211	chr3	100	20	46255327	46255341	+	TCCACTGACCTTCAG	322	0
LTRA	TCCACTGACCTTCAG	212	chr3	100	20	186757301	186757315	+	TCCACTGACCTTCAG	323	0
LTRA	TCCACTGACCTTGAG	213	chrX	100	20	74260261	74260275	+	TCCACTGACCTTGAG	324	0
LTRA	TCCACTGACCTTGAG	214	chr11	100	20	76052171	76052185	-	TCCACTGACCTTGAG	325	0
LTRA	TCCACTGACCTTGAG	215	chr9	100	20	33927660	33927674	-	TCCACTGACCTTGAG	326	0
LTRA	TCCACTGACCTTGAG	216	chr9	100	20	71035331	71035345	+	TCCACTGACCTTGAG	327	0
LTRA	TCCACTGACCTTGAG	217	chr9	100	20	95871690	95871704	+	TCCACTGACCTTGAG	328	0
LTRA	TCCACTGACCTTGAG	218	chr7	100	20	137681847	137681861	+	TCCACTGACCTTGAG	329	0
LTRA	TCCACTGACCTTGAG	219	chr6	100	20	32460609	32460623	+	TCCACTGACCTTGAG	330	0
LTRA	TCCACTGACCTTGAG	220	chr3	100	20	42344237	42344251	+	TCCACTGACCTTGAG	331	0
LTRA	TCCACTGACCTTGAG	221	chr2	100	20	64643586	64643600	+	TCCACTGACCTTGAG	332	0
LTRA	TCCACTGACCTTTAG	222	chr16	100	20	55133552	55133566	-	TCCACTGACCTTTAG	333	0
LTRA	TCCACTGACCTTTAG	223	chr15	100	20	90072212	90072226	-	TCCACTGACCTTTAG	334	0
LTRA	TCCACTGACCTTTAG	224	chr12	100	20	69006300	69006314	+	TCCACTGACCTTTAG	335	0
LTRA	TCCACTGACCTTTAG	225	chr3	100	20	170680338	170680352	-	TCCACTGACCTTTAG	336	0
LTRA	TCCACTGACCTTTAG	226	chr2	100	20	215414950	215414964	-	TCCACTGACCTTTAG	337	0

도면 15d

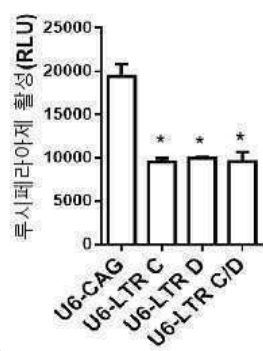
도 15 계속

명칭	검색 서열	서열 번호	대상 ID	동일성 (%)	E-값	시작	종결	가중치	기준 서열	서열 번호	미스매치 (12bp 시드)
LTRB	CAGCAGTCTTCTGAAGTACTCCGG	227	HIV	100	7.00E-04	9291	9313	-	CAGCAGTCTTCTGAAGTACTCCGG	338	0
LTRB	AGCAGTCTTCTGAAGTACTCCGG	228	HIV	100	0.003	9291	9312	-	AGCAGTCTTCTGAAGTACTCCGG	339	0
LTRB	GCAGTCTTCTGAAGTACTCCGG	229	HIV	100	0.009	9291	9311	-	GCAGTCTTCTGAAGTACTCCGG	340	0
LTRB	CAGTCTTCTGAAGTACTCCGG	230	HIV	100	0.033	9291	9310	-	CAGTCTTCTGAAGTACTCCGG	341	0
LTRB	AGTCTTCTGAAGTACTCCGG	231	HIV	100	0.12	9291	9309	-	AGTCTTCTGAAGTACTCCGG	342	0
LTRB	GTCTCTGAAGTACTCCGG	232	HIV	100	0.42	9291	9308	-	GTCTCTGAAGTACTCCGG	343	0
LTRB	TTCTTGAAGTACTCCGG	233	HIV	100	1.5	9291	9307	-	TTCTTGAAGTACTCCGG	344	0
LTRB	TCTTGAAGTACTCCGG	234	HIV	100	5.4	9291	9306	-	TCTTGAAGTACTCCGG	345	0
LTRB	TCTTGAAGTACTCTAG	235	chr11	100	5.4	91845834	91845849	-	TCTTGAAGTACTCTAG	346	0
LTRB	CTTGAAGTACTCAGG	236	chr19	100	20	45672789	45672803	-	CTTGAAGTACTCAGG	347	0
LTRB	CTTGAAGTACTCAGG	237	chr15	100	20	82132445	82132459	+	CTTGAAGTACTCAGG	348	0
LTRB	CTTGAAGTACTCAGG	238	chr11	100	20	94282411	94282425	+	CTTGAAGTACTCAGG	349	0
LTRB	CTTGAAGTACTCAGG	239	chr2	100	20	193312744	193312758	-	CTTGAAGTACTCAGG	350	0
LTRB	CTTGAAGTACTCCGG	240	HIV	100	20	9291	9305	-	CTTGAAGTACTCCGG	351	0
LTRB	CTTGAAGTACTCTGG	241	chr15	100	20	61274973	61274987	-	CTTGAAGTACTCTGG	352	0
LTRB	CTTGAAGTACTCAAG	242	chrX	100	20	36051764	36051778	-	CTTGAAGTACTCAAG	353	0
LTRB	CTTGAAGTACTCAAG	243	chr15	100	20	31316465	31316479	-	CTTGAAGTACTCAAG	354	0
LTRB	CTTGAAGTACTCAAG	244	chr13	100	20	23054474	23054488	-	CTTGAAGTACTCAAG	355	0
LTRB	CTTGAAGTACTCAAG	245	chr9	100	20	83208046	83208060	+	CTTGAAGTACTCAAG	356	0
LTRB	CTTGAAGTACTCAAG	246	chr8	100	20	13956368	13956382	+	CTTGAAGTACTCAAG	357	0
LTRB	CTTGAAGTACTCCAG	247	chr16	100	20	57449025	57449039	-	CTTGAAGTACTCCAG	358	0
LTRB	CTTGAAGTACTCCAG	248	chr15	100	20	41397831	41397845	-	CTTGAAGTACTCCAG	359	0
LTRB	CTTGAAGTACTCCAG	249	chr11	100	20	70255488	70255502	-	CTTGAAGTACTCCAG	360	0
LTRB	CTTGAAGTACTCCAG	250	chr3	100	20	134149643	134149657	+	CTTGAAGTACTCCAG	361	0
LTRB	CTTGAAGTACTCTAG	251	chr11	100	20	91845834	91845848	-	CTTGAAGTACTCTAG	362	0
LTRB	CTTGAAGTACTCTAG	252	chr1	100	20	224520600	224520614	+	CTTGAAGTACTCTAG	363	0

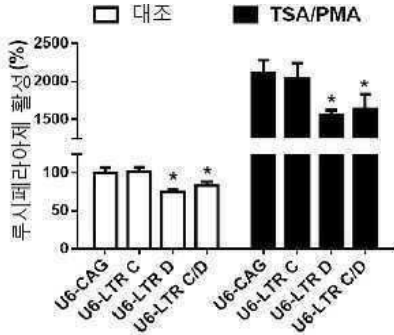
도면 16a



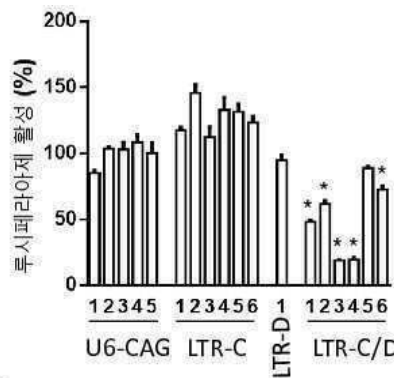
도면16b



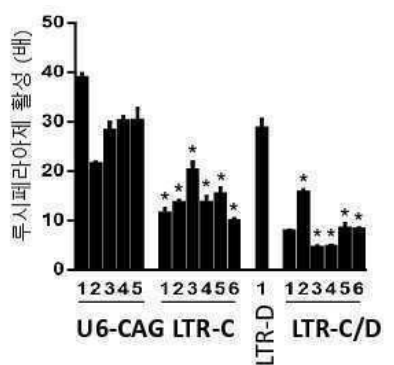
도면16c



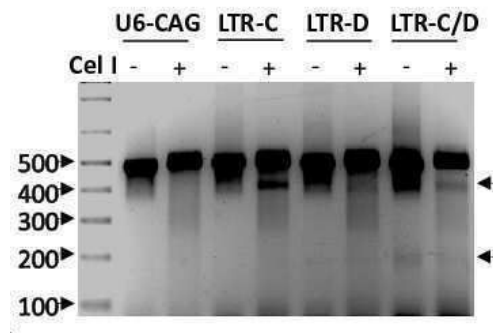
도면16d



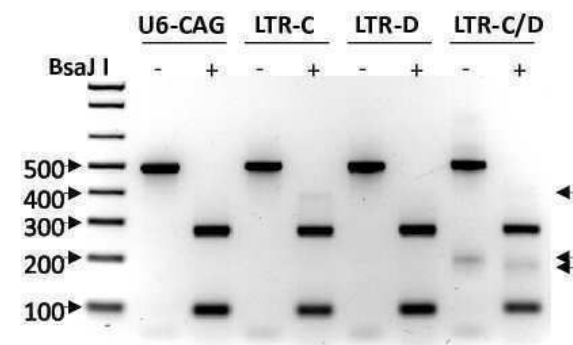
도면16e



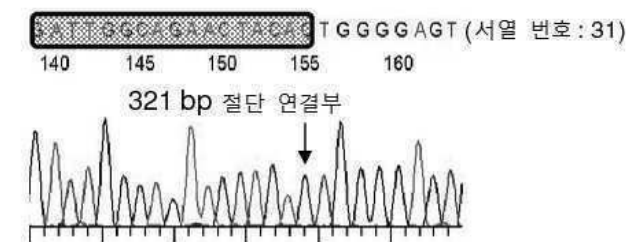
도면16f



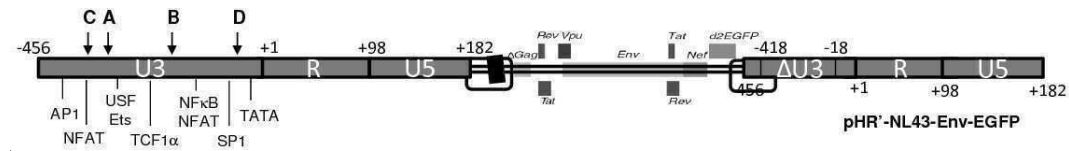
도면16g



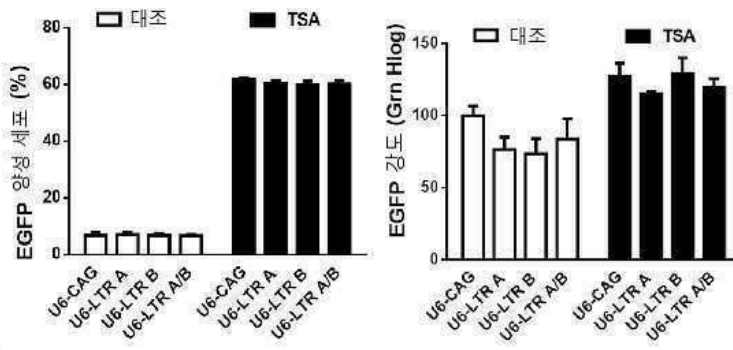
도면16h



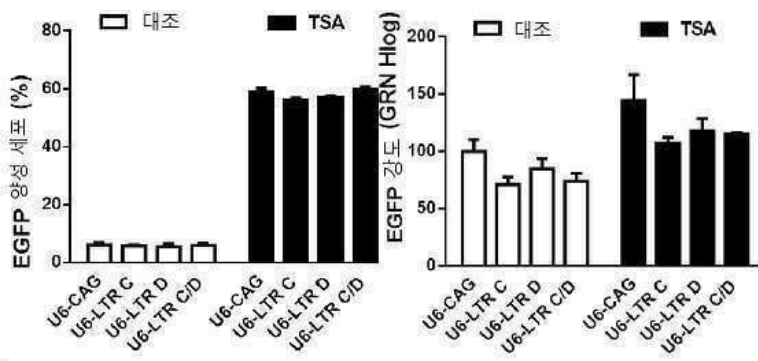
도면17a



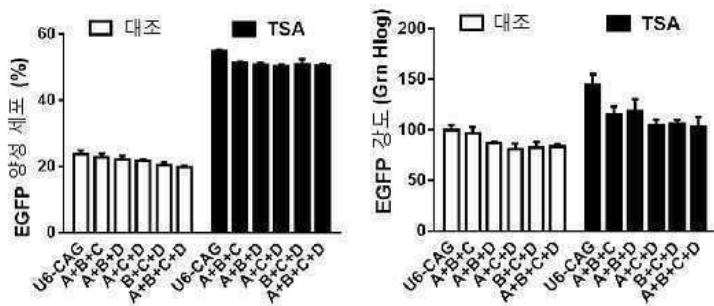
도면17b



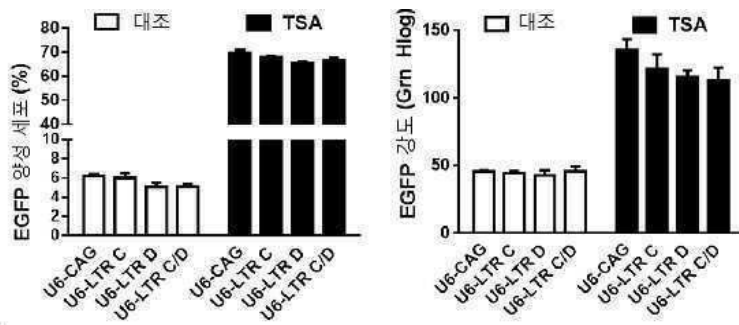
도면17c



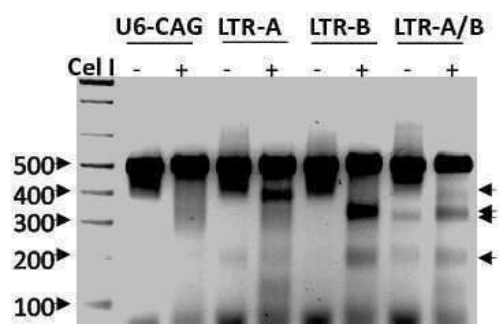
도면17d



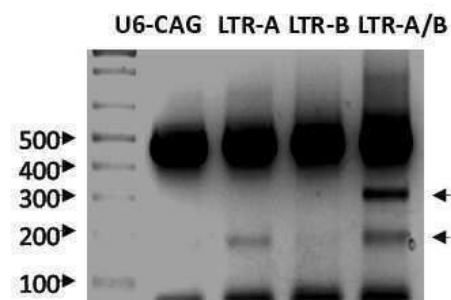
도면17e



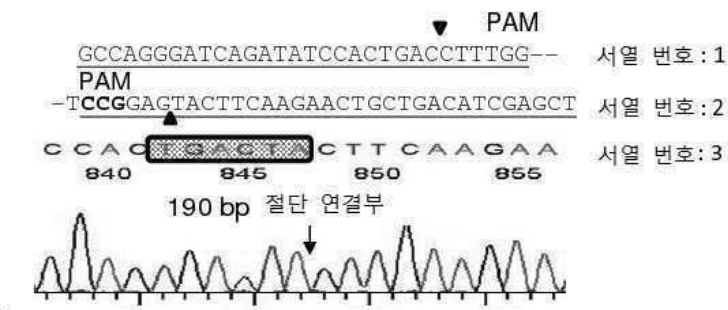
도면17f



도면17g



도면17h



도면 18

HIV-1 LTR (644 bp)

인간 면역결핍 바이러스 제1형, NY5/BRU (LAV-1) 재조합 클론 pNL4-3.
접근 번호: M19921

TGGAAGGGCTAATTCACCTCCCAACGAAGACAAAGATATCCTTGTATCTGTGGATCTACCAACAAGGCTACT
TCCCTGATTGGCAGAACTACACACCGAGGCCAGGATCAGATATCCACTGACCTTTGGATGGTGTCTACAAG
CTAGTACCAGTTGAGCAAGAGAAAGGTAGAAGAAGCCCAATGAAGGAGAGAACAACCCGCTTGTTACACCCCTG
TGAGCCTGCATGGGATGGATGACCCGGAGAGAGAAGTATTAGAGTGGAGGTTTGACAGCCGCTTAGCATT
TCATCACATGGCCCGAGAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGAC
TTTCCGCTGGGACTTTCAGGGAGGCGTGGCCTGGCGGACTGGGGAGTGGCGAGCCCTCAGATGCT
GCATATAAGCAGCTGCTTTTGTCTTGTACTGGGTCTCTCTGTGTAGACCAGATCTGAGCCTGGGAGCTCTCT
GGCTAACTAGGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCCTTGAGTGTCTCAAGTAGTGTGTGGCCG
TCTGTGTGTGACCTGTGTAAGATAGAGATCCCTCAGACCCCTTTAGTCAGTGTGGAAAATCTTAGCA

184

도면19

SIVmm239-LTR(818bp)

원숭이 (마카크) 면역결핍 바이러스, 단리체 239
유전자은행 번호: M33262

TGGAAGGATTTATTACAGTCGAAGAAGACATAGAAATCTTAGACATATACCTTAGAAAAGGAAGGATCAT
ACCAGATTGGCAGGATTACACCTCAGGACCAGGAATTAGATACCCAAAGACATTTGGCTGGCTATGGAAATTA
GTCCCTGTAAATGTATCAGATGAGGCACAGGAGGATGAGGAGCATTTATTAATGCATCCAGCTCAAACTTCCC
AGTGGGATGACCCCTTGGGGAGAGGTTCTAGCATGGAAAGTTGATCCAACTCTGGCCTACACTTATGAGGCATA
TGTTAGATACCCAGAAAGAGTTTGGAAAGCAAGTCAGGCTGTGAGGAAGAGGTTAGAAAGGCTAACCGCA
AGAGGCTTCTTAACATGGCTCACAAAGAGAAACCTCGCTGAACACAGCAGGGACTTCCACAAGGGGATGTTA
CGGGGAGGTACTGGGGAGGAGCGGTCGGGAACGCCACTTCTTGATGTATAAATATCACTGCATTTGCGCTC
TGTAATCAGTCGCTCTGCGGAGAGGCTGGCAGATTGAGCCCTGGGAGGTTCTCTCCAGCACTAGCAGGTAGAG
CCTGGGTGTTCCCTGCTAGACTCTCACAGCACCTTGGCGGGTGCTGGGCAGAGTCACTCCACGCTTGCTTGCT
TAAAGCCCTCTTCAATAAAGCTGCCATTTTAGAAGTAAGTGTGTGTTCCTCCTAGCCGCGGCC
TGGTCAACTCGGTACTCAATAAATAAGACACCCCTGGTCTCTTAGGACCCCTTCTGCTTTGGGAACCCGAGCA
GGAAATCCCTAGCA

U3 부위

R 부위

U5 부위

도 19

서열목록

SEQUENCE LISTING

<110> TEMPLE UNIVERSITY OF THE COMMONWEALTH SYSTEM OF HIGHER
EDUCATION

<120> METHODS AND COMPOSITIONS FOR RNA-GUIDED TREATMENT OF HIV
INFECTION

<130> F5129-00031

<140><141><150> 62/026,103

<151> 2014-07-18

<150> 62/018,441

<151> 2014-06-27

<150> 61/871,626
 <151> 2013-08-29
 <160> 389
 <170> PatentIn version 3.5
 <210> 1
 <211> 30
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 1
 gccagggatc agatatccac tgacctttgg 30

 <210> 2
 <211> 34
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 2
 tccggagtac ttcaagaact gctgacatcg agct 34

 <210> 3
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 3
 ccactgacta cttcaagaa 19

 <210> 4
 <211> 859
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 polynucleotide

 <220><221> modified_base
 <222> (289)..(313)
 <223> a, c, t, g, unknown or other

<400> 4

ctaggtgatt aggatattct acaatccaaa ttcttaccag tttgggatta ttcaaattgg	60
gcaccttggc agatatgttt tgaaaactgc taggcaaagc attctggaag aatagacaaa	120
gaagtaataa aatataacaa aaagcagtgg aagttacaaa aaaaaatgtt tctcttttgg	180
aagggtctaat ttggtcccaa agaagacaag atatccttga tctgtggatc taccacacac	240
aaggctactt cctgattgg cagaactaca acaccagggc cagggatcnn nnnnnnnnnn	300
nnnnnnnnnn nnnntcaagt tagtaccagt tgagccaggg caggtagaag aggccaatga	360

aggagagaac aacaccttgt tacaccctat gagcctgcat gggatggagg acccgagggg	420
agaagtatta gtgtggaagt ttgacagcct cctagcattt cgtcacatgg cccgagagct	480
gcatccggag tactacaaag actgctgaca tcgagttttc tacaaggagc tttccgctgg	540
ggactttcca gggagggtgtg gcctgggagg gactggggag tggcgagccc tcagatgctg	600
catataagca gctgcttttt gcctgtactg ggtctctctg gtttagaccag atctgagcct	660
gggagctctc tggctagcta gggaaccac tgcttaagcc tcaataaagc ttgccttgag	720
tgctacaagt agtgtgtgcc cgtctgttgt gtgactctgg taactagaga tccctcagac	780

ccttttagtc agtgtggaat atctctagca tctttaaagt acagaatgcc aaaacaggaa	840
ggattgataa gatagtcgt	859

<210> 5

<211> 10

<212> DNA

<213> Human immunodeficiency virus 1

<400> 5

tcttttgga	10
-----------	----

<210> 6

<211> 76

<212> DNA

<213> Human immunodeficiency virus 1

<400> 6

gattggcaga actacacacc agggccaggg atcagatata cactgacatt tggatgggtc	60
ttcaagttag taccag	76

<210> 7

<211> 10

<212> DNA

<213> Human immunodeficiency virus 1

<400> 7

tcttttaaagt 10

<210> 8

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 8

tcttttgga 10

<210> 9

<211> 63

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 9

gattggcaga actacaacac cagggccagg gatcagatgg atggtgcttc aagttagtagtac 60

cag 63

<210> 10

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 10

tcttttaaagt 10

<210> 11

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 11

tcttttggaa 10

<210> 12

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 12

gattggcaga actacaacac cagggccagg gatcttcaag ttagtaccag 50

<210> 13

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 13

tctttaagt 10

<210> 14

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 14

gagatcctgt ctcaaaaaa agtt 24

<210> 15

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 15

atctatccat gagggcg 17

<210> 16

<211> 402

<212> DNA

<213> Human immunodeficiency virus 1

<400> 16

gatctgtgga tctaccacac acaaggctac ttccctgatt ggcagaacta cacaccaggg 60

ccagggatca gatatccact gacctttgga tgggtgctaca agctagtacc agttgagcaa 120

gagaaggtag aagaagccaa tgaaggagag aacacccgct tgttacaccc tgtgagcctg 180

catgggatgg atgacccgga gagagaagta ttagagtgga ggtttgacag cgcctagca 240

tttcatcaca tggcccgaga gctgcatccg gagtacttca agaactgctg acatcgagct 300

tgctacaagg gactttccgc tggggacttt ccagggaggc gtggcctggg cgggactggg 360

gagtggcgag ccctcagatg ctgcatataa gcagctgctt tt 402

<210> 17

<211> 31

<212> DNA

<213> Human immunodeficiency virus 1

<400> 17

ccctgattgg cagaactaca caccagggcc a 31

<210> 18

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 18

ccctgattgg cagaactaca acaccagggc ca 32

<210> 19

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 19

ccctgattgg cagaactaca acaccagggc ca 32

<210> 20

<211> 32

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 20

ccctgattgg cagaactaca acaccagggc ca 32

<210> 21

<211> 30

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 21

ccctgattgg cagaactaca accagggccca 30

<210> 22

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 22

ccctgattgg cagaactaca ccagggccca 29

<210> 23

<211> 29

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 23

ccctgattgg cagaactaca ccagggccca	29
<210> 24	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 24	
ccctgattgg cagaactaca gggcca	26
<210> 25	
<211> 29	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 25	
ccctgattgg cagaactaca gggccagg	29
<210> 26	
<211> 86	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 26	
gactttccag ggaggcgtgg cctgggcggg actggggagt ggcgagccct cagatgctgc	60
atataagcag cgtgaagcc gaattc	86
<210> 27	
<211> 86	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 27	
gactttccag ggaggcgtgg cctgggcggg actgggggt ggcgagccct cagatgctgc	60
atataagcag cgtgaagcc gaattc	86

<210> 28

<211> 88

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 28

gactttccag ggaggcgtgg cctgggcggg tatctgggga gtggcgagcc ctcagatgct 60
gcatataagc agcgggtgaag ccgaattc 88

<210> 29

<211> 85

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 29

gactttccag gggggcgtgg cctgggcggg actggggagt ggcgagccct cagatgctgc 60
ataaagcagc ggtgaagccg aattc 85

<210> 30

<211> 23

<212> DNA

<213>

> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 30

gactttccag ggaagccgaa ttc 23

<210> 31

<211> 25

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 31

gattggcaga actacactgg ggagt 25

<210> 32

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 32

gattggcaga actacacctc agatgc 26

<210> 33

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 33

catcacatgg cccgctgctg acatcgag 28

<210> 34

<211> 55

<212> DNA

<213> Human immunodeficiency virus 1

<400> 34

catcacatgg cccgagagct gcatccggag tacttcaaga actgctgaca tcgag 55

<210> 35

<211> 1106

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<220><221> modified_base

<222> (152)..(155)

<223> a, c, t, g, unknown or other

<400> 35

gctattgtat ctgatacaaa gctgttaaaa gcggtcatgc cacttcttga atgctttgca 60
gctggaaggg ctaatttggc cccaagaag acaagatata cttgatctgt ggatctacca 120
cacacaaggc tacttccttg attggcagaa cnnnccacca gggccaggga tcagatatcc 180
actgaccatc cactttggat ggtgcttcaa gttagtacca gttgagccag ggcaggtaga 240

agaggccaat gaaggagaga acaacacctt gttacaccct atgagcctgc atgggatgga 300
ggaccgggag ggagaagtat tagtgtgga gtttgacagc ctctagcat ttcgtcacat 360
ggcccagagag ctgcatccgg agtactacaa agactgctga catcgagttt tctacaaggg 420
actttccgct ggggacttcc caggagagtg tggcctgggc gggactgggg agtggcgagc 480
ctcagatgc tgcataaag cagctgcttt ttgctgtac tgggtctctc tggtagacc 540
agatctgagc ctgggagctc tctggctagc tagggaacc actgcttaag cctcaataaa 600
gcttgccctg agtgctacaa gtagtgtgtg cccgtctgtt gtgtgactct ggtaactaga 660

gatccctcag acccttttag tcagtgtgga aaatctctag cagcagctta gaaatttttt 720
ccaccagagg ccgggcgtgg tggctcacgc ctgtaatccc agcacttttg gaggccgagg 780
tgggcggatc acctgaagtc aggagtcca gaccagctc aacatggaga aaccccatct 840
ctactaaaaa taaaaatta gctgggcgtg gtggtgcatg cctgtaatcc cagctacttg 900
ggaggctgag acaggataat tgcttgaacc tggaaggcag aggttgcggt gagccgagat 960
tgcgccattg cattccagcc tgggcaacag gagcgaaact tcgtctcaaa aaaaaaaaaa 1020
aaagacattt ttccaccag atacctaga tcatgactgt taagtctggc cttccacgaa 1080

gccctaggac ctggacacac aatcaa 1106

<210> 36
<211> 36
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 36
aaacagggcc agggatcaga tatccactga ccttgt 36

<210> 37
<211> 35
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 37

taaacaaggt cagtggatat ctgatccctg gccct 35

<210> 38

<211> 36

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 38

aaacagctcg atgtcagcag ttcttgaagt actcgt 36

<210> 39

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 39

taaacgagta cttcaagaac tgctgacatc gagct 35

<210> 40

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 40

caccgattgg cagaactaca cacc 24

<210> 41

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 41
aaacggtgtg tagttctgcc aatc 24

<210> 42
<211> 24
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 42
caccgcgtgg cctgggcggg actg 24

<210> 43
<211> 24
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 43
aaaccagtcc cgcccaggcc acgc 24

<210> 44
<211> 24
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 44
tggaagggt aattcactcc caac 24

<210> 45
<211> 24
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 45

ccgagagctc ccaggctcag atct 24

<210> 46

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 46

caccgatctg tggatctacc acacaca 27

<210> 47

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

><223> Description of Artificial Sequence: Synthetic

primer

<400> 47

aaacgagtca cacaacagac gggc 24

<210> 48

<211> 37

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 48

cgctcgagg atccgagggc ctatttccca tgattcc 37

<210> 49

<211> 35

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 49

tgtgaattca ggcgggcat ttaccgtaag ttatg 35

<210> 50
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 50
 acgactatct tatcaatcct tcttg 25
 <210> 51
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 51
 ctaggtgatt aggatattct acaatc 26
 <210> 52
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 52
 gctattgtat ctgacacaa gctg 24
 <210> 53
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 53
 ttgattgtgt gtccaggtcc tagg 24
 <210> 54
 <211> 23

<212> DNA

<213> Artificial Sequence

<220

><223> Description of Artificial Sequence: Synthetic

primer

<400> 54

gcaagggcga ggagctgttc acc 23

<210> 55

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 55

ttgtagttgc cgtcgtcctt gaag 24

<210> 56

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 56

aatggtacat caggccatat cac 23

<210> 57

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 57

cccactgtgt ttagcatggt att 23

<210> 58

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
primer

<400> 58

cacagcatca agaagaacct gat 23

<210> 59

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
primer

<400> 59

tcttccgtct ggtgtatctt ctte 24

<210> 60

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
primer

<400> 60

cgccaagctt gaataggagc ttgtgtcc 28

<210> 61

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

><223> Description of Artificial Sequence: Synthetic
primer

<400> 61

ctaggatcca ggagctgttg atcctttagg 30

<210> 62

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 62

gtggactttg gatggtgaga tag 23

<210> 63

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 63

gcctggcaag agtgaactga gtc 23

<210> 64

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 64

aagataatga gttgtggcag agc 23

<210> 65

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 65

tctacctggt aatccagcat ctgg 24

<210> 66

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 66

ataggaggaa ggcaccaaga ggg 23

<210> 67

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 67

aatgatgctt tggctctact cct 23

<210> 68

<211> 24

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 68

tgctcttgct actctggcat gtac 24

<210> 69

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 69

aatctacctc tgagagctgc agg 23

<210> 70

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 70	
tcagacacag ctgaagcaga ggc	23
<210> 71	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 71	
atgccagtgt cagtagatgt cag	23
<210> 72	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 72	
tcaagatcag ccagagtga catg	24
<210> 73	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 73	
tgctcttccg agcctctctg gag	23
<210> 74	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 74	
atggactatc atatgcttac cg	22

<210> 75

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 75

gcttcagcaa gccgagtcct gcgtcgag 28

<210> 76

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 76

gctcctctgg tttccctttc gctttcaa 28

<210> 77

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 77

gtaatacgac tcactatagg gc 22

<210> 78

<211> 19

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 78

actatagggc acgcgtggt 19

<210> 79

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 79

tcagaccctt ttagtcagtg tgg 23

<210> 80

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 80

ttgcttgtac tgggtctctc tgg 23

<210> 81

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 81

cagctgcttt ttgcttgtac tgg 23

<210> 82

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 82

ctgacatcga gcttgctaca agg 23

<210> 83

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 83

ccgcctagca ttcatcaca tgg 23

<210> 84

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 84

cggagagaga agtattagag tgg 23

<210> 85

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 85

agtaccagtt gagcaagaga agg 23

<210> 86

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 86

gatatccact gacctttgga tgg 23

<210> 87

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 87

gattggcaga actacacacc agg 23

<210> 88

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 88

cacaaggcta cttccctgat tgg 23

<210> 89

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 89

ctgtggatct accacacaca agg 23

<210> 90

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 90

tgggagctct ctggctaact agg 23

<210> 91

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 91

ggtttagacca gatctgagcc tgg

23

<210> 92

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 92

tgctacaagg gactttccgc tgg

23

<210> 93

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 93

agagagaagt attagagtgg agg

23

<210> 94

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 94

ttacaccctg tgaccctgca tgg

23

<210> 95

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 95

aaggtagaag aagccaatga agg 23

<210> 96

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 96

atcagatatc cactgacctt tgg 23

<210> 97

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 97

gacaagatat ccttgatctg tgg 23

<210> 98

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 98

gcccgtctgt tgtgtgactc tgg 23

<210> 99

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 99
atctgagcct gggagctctc tgg 23
<210> 100
<211> 23
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 100
ctttccgctg gggactttcc agg 23
<210> 101
<211> 23
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 101
cagaactaca caccagggcc agg 23
<210> 102
<211> 23
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 102
cctgcatggg atggatgacc cgg 23
<210> 103
<211> 23
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 103

ccctgtgagc ctgcatggga tgg	23
<210> 104	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 104	
ctttccaggg aggcgtggcc tgg	23
<210> 105	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 105	
ggggactttc caggagggcg tgg	23
<210> 106	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 106	
ccgctgggga ctttccaggg agg	23
<210> 107	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic	
oligonucleotide	
<400> 107	
catgcccga gagctgcac cgg	23

<210> 108
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 108
 gcctgggcgg gactggggag tgg 23
 <210> 109
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 109
 aggcgtggcc tgggcgggac tgg 23
 <210> 110
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 110
 gcgtggcctg ggcgggactg ggg 23
 <210> 111
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide
 <400> 111
 ccagggaggc gtggcctggg cgg 23
 <210> 112
 <211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 112

tgtggtagat ccacagatca agg 23

<210> 113

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 113

ggtgtgtagt tctgccaatc agg 23

<210> 114

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 114

gtcagtggat atctgatccc tgg 23

<210> 115

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 115

tagcaccatc caaaggtcag tgg 23

<210> 116

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 116

tagcttgtag caccatcaa agg 23

<210> 117

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 117

tctacattct ctgtctaac tgg 23

<210> 118

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 118

cactctaata cttctctctc cgg 23

<210> 119

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 119

ccatgtgatg aaatgctagg cgg 23

<210> 120

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 120

gggccatgtg atgaaatgct agg

23

<210> 121

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 121

cagcagttct tgaagtactc cgg

23

<210> 122

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 122

ctgcttatat gcagcatctg agg

23

<210> 123

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 123

cacactactt gaagcactca agg

23

<210> 124

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 124

taccagagtc acacaacaga cgg 23

<210> 125

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 125

acactgacta aaagggtctg agg 23

<210> 126

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 126

caaggatatac ttgtcttcgt tgg 23

<210> 127

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 127

cagggaagta gccttggtg tgg 23

<210> 128

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 128

gcgggtgttc tctccttcat tgg	23
<210> 129	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 129	
tagttagcca gagagctccc agg	23
<210> 130	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 130	
ctttattgag gcttaagcag tgg	23
<210> 131	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 131	
actcaaggca agctttattg agg	23
<210> 132	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220><223> Description of Artificial Sequence: Synthetic oligonucleotide	
<400> 132	
ggatatctga tccctggccc tgg	23

<210> 133

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 133

ggctcacagg gtgtaacaag cgg 23

<210> 134

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 134

tccatcccat gcaggctcac agg 23

<210> 135

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 135

agtactccgg atgcagctct cgg 23

<210> 136

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 136

agagctccca ggctcagatc tgg 23

<210> 137

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 137

gattttccac actgactaaa agg 23

<210> 138

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 138

ccgggtcatc catcccatgc agg 23

<210> 139

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 139

cctccctgga aagtccccag cgg 23

<210> 140

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 140

gccactcccc agtcccgccc agg 23

<210> 141

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 141

ccgcccaggc cacgcctccc tgg 23

<210> 142

<211> 23

<212> DNA

<213> Human immunodeficiency virus 1

<400> 142

atcagatatc cactgacctt tgg 23

<210>

> 143

<211> 22

<212> DNA

<213> Human immunodeficiency virus 1

<400> 143

tcagatatcc actgaccttt gg 22

<210> 144

<211> 22

<212> DNA

<213> Human immunodeficiency virus 1

<400> 144

tcagatatcc actgaccttt gg 22

<210> 145

<211> 21

<212> DNA

<213> Human immunodeficiency virus 1

<400> 145

cagatatcca ctgacctttg g 21

<210> 146

<211> 21

<212>

DNA

<213> Human immunodeficiency virus 1
 <400> 146
 cagatatcca ctgacctttg g 21
 <210> 147
 <211> 20
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 147
 agatatccac tgacctttgg 20
 <210> 148
 <211> 20
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 148
 agatatccac tgacctttgg 20
 <210> 149
 <211> 19
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 149
 gatataccact gacctttgg 19
 <210> 150
 <211> 19
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 150
 gatataccact gacctttgg 19
 <210> 151
 <211> 18
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 151
 atataccactg acctttgg 18
 <210> 152

<211> 18
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 152
 atatccactg acctttgg 18

<210> 153
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 153
 tatccactga ccttggg 17

<210> 154
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 154
 tatccactga cctttgg 17

<210> 155
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 155
 tatccactga cctttgg 17

<210> 156
 <211> 17
 <

212> DNA
 <213> Human immunodeficiency virus 1
 <400> 156
 tatccactga ccttaag 17

<210> 157
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1

<400> 157	
tatccactga ccttgag	17
<210> 158	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 158	
atccactgac cttagg	16
<210> 159	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 159	
atccactgac cttagg	16
<210> 160	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 160	
atccactgac cttggg	16
<210> 161	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 161	
atccactgac cttggg	16
<210> 162	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 162	
atccactgac cttggg	16
<210> 163	

<211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 163
 atccactgac cttggg 16
 <210> 164
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 164
 atccactgac ctttgg 16
 <210> 165
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 165
 atccactgac ctttgg 16
 <210> 166
 <211> 16
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 166
 atccactgac ctttgg 16
 <210> 167
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 167
 atccactgac ctttgg 16
 <210> 168
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 168

atccactgac ctttaag	16
<210> 169	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 169	
atccactgac cttcag	16
<210> 170	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 170	
atccactgac cttcag	16
<210> 171	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 171	
atccactgac cttgag	16
<210> 172	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 172	
atccactgac cttgag	16
<210> 173	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 173	
tccactgacc ttagg	15
<210> 174	
<211> 15	

<212> DNA
 <213> Human immunodeficiency virus 1
 <400> 174
 tccactgacc ttagg 15
 <210> 175
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 175
 tccactgacc ttagg 15
 <210> 176
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 176
 tccactgacc ttagg 15
 <210> 177
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 177
 tccactgacc ttagg 15
 <210> 178
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 178
 tccactgacc ttagg 15
 <210> 179
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 179

tccactgacc ttggg	15
<210> 180	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 180	
tccactgacc ttggg	15
<210> 181	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 181	
tccactgacc ttggg	15
<210> 182	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 182	
tccactgacc ttggg	15
<210> 183	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 183	
tccactgacc ttggg	15
<210> 184	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 184	
tccactgacc ttggg	15
<210> 185	
<211> 15	
<212> DNA	

<213> Human immunodeficiency virus 1
 <400> 185
 tccactgacc ttggg 15
 <210> 186
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 186
 tccactgacc ttggg 15
 <210> 187
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 187
 tccactgacc tttgg 15
 <210> 188
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 188
 tccactgacc tttgg 15
 <210> 189
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 189
 tccactgacc tttgg 15
 <210> 190
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 190
 tccactgacc tttgg 15

<210> 191
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 191
 tccactgacc ttggtg 15
 <210> 192
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 192
 tccactgacc ttggtg 15
 <210> 193
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 193
 tccactgacc ttggtg 15
 <210> 194
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 194
 tccactgacc ttggtg 15
 <210> 195
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 195
 tccactgacc ttggtg 15
 <210> 196
 <211> 15
 <
 212> DNA

<213> Human immunodeficiency virus 1
 <400> 196
 tccactgacc ttaag 15
 <210> 197
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 197
 tccactgacc ttaag 15
 <210> 198
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 198
 tccactgacc ttaag 15
 <210> 199
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 199
 tccactgacc ttaag 15
 <210> 200
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 200
 tccactgacc ttaag 15
 <210> 201
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 201
 tccactgacc ttcag 15
 <210> 202

<211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 202
 tccactgacc ttcag 15

 <210> 203
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 203
 tccactgacc ttcag 15

 <210> 204
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 204
 tccactgacc ttcag 15

 <210> 205
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 205
 tccactgacc ttcag 15

 <210> 206
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 206
 tccactgacc ttcag 15

 <210> 207
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1

<400> 207	
tccactgacc ttcag	15
<210> 208	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 208	
tccactgacc ttcag	15
<210> 209	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 209	
tccactgacc ttcag	15
<210> 210	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 210	
tccactgacc ttcag	15
<210> 211	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 211	
tccactgacc ttcag	15
<210> 212	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 212	
tccactgacc ttcag	15
<210> 213	

<211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 213
 tccactgacc ttgag 15
 <210> 214
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 214
 tccactgacc ttgag 15
 <210> 215
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 215
 tccactgacc ttgag 15
 <210> 216
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 216
 tccactgacc ttgag 15
 <210> 217
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 217
 tccactgacc ttgag 15
 <210> 218
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 218

tccactgacc ttgag	15
<210> 219	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 219	
tccactgacc ttgag	15
<210> 220	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 220	
tccactgacc ttgag	15
<210> 221	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 221	
tccactgacc ttgag	15
<210> 222	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 222	
tccactgacc tttag	15
<210> 223	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 223	
tccactgacc tttag	15
<210> 224	
<211> 15	

<212> DNA
 <213> Human immunodeficiency virus 1
 <400> 224
 tccactgacc ttttag 15
 <210> 225
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 225
 tccactgacc ttttag 15
 <210> 226
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 226
 tccactgacc ttttag 15
 <210> 227
 <211> 23
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 227
 cagcagttct tgaagtactc cgg 23
 <210> 228
 <211> 22
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 228
 agcagttctt gaagtactcc gg 22
 <210> 229
 <211> 21
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 229

gcagttcttg aagtactccg g	21
<210> 230	
<211> 20	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 230	
cagttcttga agtactccgg	20
<210> 231	
<211> 19	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 231	
agttcttgaa gtactccgg	19
<210> 232	
<211> 18	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 232	
gttcttgaag tactccgg	18
<210> 233	
<211> 17	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 233	
ttcttgaagt actccgg	17
<210> 234	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 234	
tcttgaagta ctccgg	16
<210> 235	
<211> 16	
<212> DNA	

<213> Human immunodeficiency virus 1
 <400> 235
 tcttgaagta ctctag 16
 <210> 236
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 236
 cttgaagtac tcagg 15
 <210> 237
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 237
 cttgaagtac tcagg 15
 <210> 238
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 238
 cttgaagtac tcagg 15
 <210> 239
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 239
 cttgaagtac tcagg 15
 <210> 240
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 240
 cttgaagtac tccgg 15

<210> 241
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 241
 cttgaagtac tctgg 15
 <210> 242
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 242
 cttgaagtac tcaag 15
 <210> 243
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 243
 cttgaagtac tcaag 15
 <210> 244
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 244
 cttgaagtac tcaag 15
 <210> 245
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 245
 cttgaagtac tcaag 15
 <210> 246
 <211> 15
 <
 212> DNA

<213> Human immunodeficiency virus 1
 <400> 246
 cttgaagtac tcaag 15
 <210> 247
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 247
 cttgaagtac tccag 15
 <210> 248
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 248
 cttgaagtac tccag 15
 <210> 249
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 249
 cttgaagtac tccag 15
 <210> 250
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 250
 cttgaagtac tccag 15
 <210> 251
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 251
 cttgaagtac tctag 15
 <210> 252

<211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 252
 cttgaagtac tctag 15

<210> 253
 <211> 23
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 253
 atcagatatc cactgacctt tgg 23

<210> 254
 <211> 22
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 254
 tcagatatcc actgaccttt gg 22

<210> 255
 <211> 22
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 255
 tcagatatcc actgaccttt gg 22

<210> 256
 <211> 21
 <

212> DNA
 <213> Human immunodeficiency virus 1
 <400> 256
 cagatatcca ctgacctttg g 21

<210> 257
 <211> 21
 <212> DNA
 <213> Human immunodeficiency virus 1

<400> 257	
cagatatcca ctgacctttg g	21
<210> 258	
<211> 20	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 258	
agatatccac tgacctttgg	20
<210> 259	
<211> 20	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 259	
agatatccac tgacctttgg	20
<210> 260	
<211> 19	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 260	
gatataccact gacctttgg	19
<210> 261	
<211> 19	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 261	
gatataccact gacctttgg	19
<210> 262	
<211> 18	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 262	
atatccactg acctttgg	18
<210> 263	

<211> 18
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 263
 atatccactg acctttgg 18
 <210> 264
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 264
 tatccactga cctttggg 17
 <210> 265
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 265
 tatccactga cctttgg 17
 <210> 266
 <211> 17
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 266
 tatccactga cctttgg 17
 <210> 267
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 267
 tatccactga ccttaag 17
 <210> 268
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 268

tatccactga ccttgag	17
<210> 269	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 269	
atccactgac cttagg	16
<210> 270	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 270	
atccactgac cttagg	16
<210> 271	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 271	
atccactgac cttggg	16
<210> 272	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 272	
atccactgac cttggg	16
<210> 273	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 273	
atccactgac cttggg	16
<210> 274	
<211> 16	

<212> DNA
 <213> Human immunodeficiency virus 1
 <400> 274
 atccactgac cttggg 16
 <210> 275
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 275
 atccactgac ctttgg 16
 <210> 276
 <211> 16
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 276
 atccactgac ctttgg 16
 <210> 277
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 277
 atccactgac ctttgg 16
 <210> 278
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 278
 atccactgac ctttaag 16
 <210> 279
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 279

atccactgac ctttaag	16
<210> 280	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 280	
atccactgac cttcag	16
<210> 281	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 281	
atccactgac cttcag	16
<210> 282	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 282	
atccactgac cttgag	16
<210> 283	
<211> 16	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 283	
atccactgac cttgag	16
<210> 284	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 284	
tccactgacc ttagg	15
<210> 285	
<211> 15	
<212> DNA	

<213> Human immunodeficiency virus 1
 <400> 285
 tccactgacc ttagg 15
 <210> 286
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 286
 tccactgacc ttagg 15
 <210> 287
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 287
 tccactgacc ttagg 15
 <210> 288
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 288
 tccactgacc ttagg 15
 <210> 289
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 289
 tccactgacc ttagg 15
 <210> 290
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 290
 tccactgacc ttggg 15

<210> 291
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 291
 tccactgacc ttggg 15
 <210> 292
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 292
 tccactgacc ttggg 15
 <210> 293
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 293
 tccactgacc ttggg 15
 <210> 294
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 294
 tccactgacc ttggg 15
 <210> 295
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 295
 tccactgacc ttggg 15
 <210> 296
 <211> 15
 <
 212> DNA

<213> Human immunodeficiency virus 1
 <400> 296
 tccactgacc ttggg 15
 <210> 297
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 297
 tccactgacc ttggg 15
 <210> 298
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 298
 tccactgacc tttgg 15
 <210> 299
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 299
 tccactgacc tttgg 15
 <210> 300
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 300
 tccactgacc tttgg 15
 <210> 301
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 301
 tccactgacc tttgg 15
 <210> 302

<211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 302
 tccactgacc ttg 15

<210> 303
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 303
 tccactgacc ttg 15

<210> 304
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 304
 tccactgacc ttg 15

<210> 305
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 305
 tccactgacc ttg 15

<210> 306
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 306
 tccactgacc ttg 15

<210> 307
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1

<400> 307	
tccactgacc ttaag	15
<210> 308	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 308	
tccactgacc ttaag	15
<210> 309	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 309	
tccactgacc ttaag	15
<210> 310	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 310	
tccactgacc ttaag	15
<210> 311	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 311	
tccactgacc ttaag	15
<210> 312	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 312	
tccactgacc ttcag	15
<210> 313	

<211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 313
 tccactgacc ttcag 15
 <210> 314
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 314
 tccactgacc ttcag 15
 <210> 315
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 315
 tccactgacc ttcag 15
 <210> 316
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 316
 tccactgacc ttcag 15
 <210> 317
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 317
 tccactgacc ttcag 15
 <210> 318
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 318

tccactgacc ttcag	15
<210> 319	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 319	
tccactgacc ttcag	15
<210> 320	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 320	
tccactgacc ttcag	15
<210> 321	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 321	
tccactgacc ttcag	15
<210> 322	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 322	
tccactgacc ttcag	15
<210> 323	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 323	
tccactgacc ttcag	15
<210> 324	
<211> 15	

<212> DNA
 <213> Human immunodeficiency virus 1
 <400> 324
 tccactgacc ttgag 15
 <210> 325
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 325
 tccactgacc ttgag 15
 <210> 326
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 326
 tccactgacc ttgag 15
 <210> 327
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 327
 tccactgacc ttgag 15
 <210> 328
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 328
 tccactgacc ttgag 15
 <210> 329
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 329

tccactgacc ttgag	15
<210> 330	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 330	
tccactgacc ttgag	15
<210> 331	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 331	
tccactgacc ttgag	15
<210> 332	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 332	
tccactgacc ttgag	15
<210> 333	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 333	
tccactgacc tttag	15
<210> 334	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 334	
tccactgacc tttag	15
<210> 335	
<211> 15	
<212> DNA	

<213> Human immunodeficiency virus 1
 <400> 335
 tccactgacc tttag 15
 <210> 336
 <211> 15
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 336
 tccactgacc tttag 15
 <210> 337
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 337
 tccactgacc tttag 15
 <210> 338
 <211> 23
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 338
 cagcagttct tgaagtactc cgg 23
 <210> 339
 <211> 22
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 339
 agcagttctt gaagtactcc gg 22
 <210> 340
 <211> 21
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 340
 gcagttcttg aagtactccg g 21

<210> 341
 <211> 20
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 341
 cagttcttga agtactccgg 20
 <210> 342
 <211> 19
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 342
 agttcttgaa gtactccgg 19
 <210> 343
 <211> 18
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 343
 gttcttgaag tactccgg 18
 <210> 344
 <211> 17
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 344
 ttcttgaagt actccgg 17
 <210> 345
 <211> 16
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 345
 tcttgaagta ctccgg 16
 <210> 346
 <211> 16
 <
 212> DNA

<213> Human immunodeficiency virus 1
 <400> 346
 tcttgaagta ctctag 16
 <210> 347
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 347
 cttgaagtac tcagg 15
 <210> 348
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 348
 cttgaagtac tcagg 15
 <210> 349
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 349
 cttgaagtac tcagg 15
 <210> 350
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 350
 cttgaagtac tcagg 15
 <210> 351
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 351
 cttgaagtac tccgg 15
 <210> 352

<211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 352
 cttgaagtac tctgg 15

<210> 353
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 353
 cttgaagtac tcaag 15

<210> 354
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 354
 cttgaagtac tcaag 15

<210> 355
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 355
 cttgaagtac tcaag 15

<210> 356
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 356
 cttgaagtac tcaag 15

<210> 357
 <211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1

<400> 357	
cttgaagtac tcaag	15
<210> 358	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 358	
cttgaagtac tccag	15
<210> 359	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 359	
cttgaagtac tccag	15
<210> 360	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 360	
cttgaagtac tccag	15
<210> 361	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 361	
cttgaagtac tccag	15
<210> 362	
<211> 15	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 362	
cttgaagtac tctag	15
<210> 363	

<211> 15
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 363
 cttgaagtac tctag 15
 <210> 364
 <211> 23
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 364
 gatctgtgga tctaccacac aca 23
 <210> 365
 <211> 26
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 365
 gatctgtgga tctaccacac acaagg 26
 <210> 366
 <211> 20
 <
 212> DNA
 <213> Human immunodeficiency virus 1
 <400> 366
 gattggcaga actacacacc 20
 <210> 367
 <211> 23
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 367
 gattggcaga actacacacc agg 23
 <210> 368
 <211> 27
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 368

gccagggatc agatatccac tgacctt	27
<210> 369	
<211> 30	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 369	
gccagggatc agatatccac tgacctttgg	30
<210> 370	
<211> 30	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 370	
gagtacttca agaactgctg acatcgagct	30
<210> 371	
<211> 33	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 371	
ccggagtact tcaagaactg ctgacatcga gct	33
<210> 372	
<211> 20	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 372	
gcgtggcctg ggcgggactg	20
<210> 373	
<211> 23	
<212> DNA	
<213> Human immunodeficiency virus 1	
<400> 373	
gcgtggcctg ggcgggactg ggg	23
<210> 374	
<211> 22	

<212> DNA
 <213> Human immunodeficiency virus 1
 <400> 374
 tcagatgctg catataagca gc 22
 <210> 375
 <211> 25
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 375
 ccctcagatg ctgcatataa gcagc 25
 <210> 376
 <211> 634
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 polynucleotide
 <400> 376
 tggaagggt aattcactcc caacgaagac aagatatacct tgatctgtgg atctaccaca 60
 cacaaggcta cttccctgat tggcagaact acacaccagg gccagggatc agatatccac 120
 tgacctttgg atggtgctac aagctagtac cagttgagca agagaaggta gaagaagcca 180
 atgaaggaga gaacaccgc ttgttacacc ctgtgagcct gcatgggatg gatgaccgg 240
 agagagaagt attagagtgg aggtttgaca gccgcctagc atttcacac atggcccag 300
 agctgcatcc ggagtacttc aagaactgct gacatcgagc ttgctacaag ggactttccg 360
 ctggggactt tccaggagg cgtggcctgg gcgggactgg ggagtggcga gccctcagat 420
 gctgcatata agcagctgct ttttgcttgt actgggtctc tctggttaga ccagatctga 480
 gcctgggagc tctctggcta actaggggaac cactgctta agcctcaata aagcttgcc 540
 tgagtgttc aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc 600
 agaccctttt agtcagtgtg gaaaatctct agca 634
 <210> 377
 <211> 453
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 377

tggaagggct aattcactcc caacgaagac aagataatcct tgatctgtgg atctaccaca	60
cacaaggcta cttccctgat tggcagaact acacaccagg gccagggatc agatatccac	120
tgacctttgg atggtgctac aagctagtac cagttgagca agagaaggta gaagaagcca	180
atgaaggaga gaacacccgc ttgttacacc ctgtgagcct gcatgggatg gatgaccg	240
agagagaagt attagagtgg aggtttgaca gccgcctagc attcatcac atggcccag	300
agctgcatcc ggagtaactc aagaactgct gacatcgagc ttgtacaag ggactttccg	360

ctggggactt tccaggagg cgtggcctgg gcgggactgg ggagtggcga gccctcagat	420
gctgcatata agcagctgct ttttcttgt act	453

<210> 378

<211> 97

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 378

gggtctctct ggtagacca gatctgagcc tgggagctct ctggctaact agggaaccca	60
ctgcttaagc ctcaataaag cttgccttga gtgcttc	97

<210> 379

<211> 84

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 379

aagtagtgtg tgcccgctctg ttgtgtgact ctggtaacta gagatccctc agaccctttt	60
agtcagtgtg gaaaatctct agca	84

<210> 380

<211> 818

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 380

tggaagggat ttattacagt gcaagaagac atagaatctt agacatatac ttagaaaagg	60
aagaaggcat cataccagat tggcaggatt acacctcagg accaggaatt agatacccaa	120
agacatttgg ctggctatgg aaattagtcc ctgtaaatgt atcagatgag gcacaggagg	180
atgaggagca ttatttaatg catccagctc aaacttccca gtgggatgac cttggggag	240
aggttctagc atggaagttt gatccaactc tggcctacac ttatgaggca tatgttagat	300
accagaaga gtttgaagc aagtcaggcc tgcagagga agaggttaga agaaggctaa	360
ccgcaagagg ctttcttaac atggctgaca agaaggaaac tcgctgaaac agcagggact	420
ttccacaagg ggatgttacg gggagggtact ggggaggagc cggtcgggaa cgcccacttt	480
cttgatgtat aaatatcact gcatttcgct ctgtattcag tcgctctgcg gagaggctgg	540
cagattgagc cctgggaggt tctctccagc actagcaggt agagcctggg tgttcctgc	600
tagactctca ccagcacttg gccggtgctg ggcagagtga ctccacgctt gcttgcttaa	660
agccctcttc aataaagctg ccatittaga agtaagctag tgtgtgttcc catctctct	720
agccgccgcc tggtaactc ggtactcaat aataagaaga ccctggctctg ttaggacct	780
ttctgctttg ggaaaccgaa gcaggaaaat ccctagca	818

<210> 381

<211> 517

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 381

tggaagggat ttattacagt gcaagaagac atagaatctt agacatatac ttagaaaagg	60
aagaaggcat cataccagat tggcaggatt acacctcagg accaggaatt agatacccaa	120
agacatttgg ctggctatgg aaattagtcc ctgtaaatgt atcagatgag gcacaggagg	180
atgaggagca ttatttaatg catccagctc aaacttccca gtgggatgac cttggggag	240
aggttctagc atggaagttt gatccaactc tggcctacac ttatgaggca tatgttagat	300
accagaaga gtttgaagc aagtcaggcc tgcagagga agaggttaga agaaggctaa	360
ccgcaagagg ctttcttaac atggctgaca agaaggaaac tcgctgaaac agcagggact	420

ttccacaagg ggatgttacg gggagggtact ggggaggagc cggtcgggaa cgcccacttt 480

cttgatgtat aaatatcact gcatttcgct ctgtatt 517

<210> 382

<211> 176

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 382

cagtcgctct gcggagaggc tggcagattg agccctggga ggttctctcc agcactagca 60

ggttagagcct ggggtgttccc tgctagactc tcaccagcac ttggccgggtg ctgggcagag 120

tgactccacg cttgcttgct taaagccctc ttcaataaag ctgccatttt agaagt 176

<210> 383

<211> 125

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 383

aagctagtgt gtgttcccat ctctcctagc cgccgctgg tcaactcgtt actcaataat 60

aagaagacct tggctctgta ggaccctttc tgctttggga aaccgaagca ggaaaatccc 120

tagca 125

<210> 384

<211> 14825

<212> DNA

<213> Human immunodeficiency virus 1

<400> 384

tggaagggt aatttggctc caaaaaagac aagagatcct tgatctgtgg atctaccaca 60

cacaaggcta cticcctgat tggcagaact acacaccagg gccagggatc agatatccac 120

tgacctttgg atggtgcttc aagttagtac cagttgaacc agagcaagta gaagaggcca 180

atgaaggaga gaacaacagc ttgttacacc ctatgagcca gcatgggatg gaggacccgg 240

agggagaagt attagtgtgg aagtttgaca gcctcctagc atttcgtcac atggcccag 300

agctgcattc ggagtactac aaagactgct gacatcgagc tttctacaag ggactttccg 360

ctggggactt tccagggagg tgtggcctgg gcgggactgg ggagtggcga gccctcagat 420
gctacatata agcagctgct ttttgcctgt actgggtctc tctggttaga ccagatctga 480

gcctgggagc tctctggcta actaggaac ccactgetta agcctcaata aagcttgcc 540
tgagtgtca aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc 600
agaccctttt agtcagtgtg gaaaatctct agcagtggcg cccgaacagg gacttgaaag 660
cgaaagtaaa gccagaggag atctctcgac gcaggactcg gcttgctgaa gcgcgcacgg 720
caagaggcga ggggcggcga ctgggtgagta cgccaaaaat tttgactagc ggaggctaga 780
aggagagaga tgggtgcgag agcgtcggta ttaagcgggg gagaattaga taaatgggaa 840
aaaattcggg taaggccagg gggaaagaaa caatataaac taaaacatat agtatgggca 900

agcagggagc tagaacgatt cgcagttaat cctggccttt tagagacatc agaagctgt 960
agacaaatac tgggacagct acaaccatcc cttcagacag gatcagaaga acttagatca 1020
ttatataata caatagcagt cctctattgt gtgcatcaaa ggatagatgt aaaagacacc 1080
aaggaagcct tagataagat agaggaagag caaaacaaaa gtaagaaaaa ggcacagcaa 1140
gcagcagctg acacaggaag caacagccag gtcagccaaa attaccctat agtgcagaac 1200
ctccaggggc aaatggtaca tcaggccata tcacctagaa ctttaaatgc atgggtaaaa 1260
gtagtagaag agaaggtttt cagcccagaa gtaataccca tgttttcagc attatcagaa 1320

ggagccaccc cacaagattt aaataccatg ctaaacacag tggggggaca tcaagcagcc 1380
atgcaaatgt taaaagagac catcaatgag gaagctgcag aatgggatag attgcatcca 1440
gtgcatgcag ggcctattgc accaggccag atgagagaac caaggggaag tgacatagca 1500
ggaactacta gtacccttca ggaacaaata ggatggatga cacataatcc acctatccca 1560
gtaggagaaa tctataaaag atggataatc ctgggattaa ataaaatagt aagaatgtat 1620
agccctacca gcattctgga cataagacaa ggaccaaagg aaccctttag agactatgta 1680
gaccgattct ataaaactct aagagccgag caagcttcac aagaggtaaa aaattggatg 1740

acagaaacct tgttgggtcca aaatgcgaac ccagattgta agactatttt aaaagcattg 1800
ggaccaggag cgacactaga agaatgatg acagcatgtc agggagtggg gggacccggc 1860
cataaagcaa gagtthttggc tgaagcaatg agccaagtaa caaatccagc taccataatg 1920
atacagaaag gcaattttgag gaaccaaaaga aagactgtta agtgtttcaa ttgtggcaaa 1980
gaagggcaca tagccaaaaa ttgcagggcc cctaggaaaa agggctgttg gaaatgtgga 2040
aaggaaggac accaaatgaa agattgtact gagagacagg ctaatthttt agggaagatc 2100

tgcccttccc acaagggaag gccaggggaat tttcttcaga gcagaccaga gccaacagcc	2160
ccaccagaag agagcttcag gtttggggaa gagacaacaa ctccctctca gaagcaggag	2220
ccgatagaca aggaactgta tccttttagct tccttcagat cactctttgg cagcgacccc	2280
tcgtcacaat aaagataggg gggcaattaa aggaagctct attagataca ggagcagatg	2340
atacagtatt agaagaaatg aatttgccag gaagatggaa accaaaaatg atagggggaa	2400
ttggagggtt tatcaaagta agacagtatg atcagatact catagaaatc tgcggacata	2460
aagctatagg tacagtatta gtaggacctt cacctgtcaa cataattgga agaaatctgt	2520
tgactcagat tggtgcact ttaaattttc ccattagtcc tattgagact gtaccagtaa	2580
aattaaagcc aggaatggat ggcccaaaag ttaaacaatg gccattgaca gaagaaaaaa	2640
taaaagcatt agtagaaatt tgtacagaaa tggaaaagga aggaaaaatt tcaaaaattg	2700
ggcctgaaaa tcatacaat actccagtat ttgccataaa gaaaaaagac agtactaaat	2760
ggagaaaaatt agtagatttc agagaactta ataagagaac tcaagatttc tgggaagttc	2820
aattaggaat accacatcct gcagggttaa aacagaaaaa atcagtaaca gtactggatg	2880
tgggcgatgc atatitttca gtcccttag ataaagactt caggaagtat actgcattta	2940
ccatacctag tataaacaat gagacaccag ggattagata tcagtaaat gtgcttcac	3000
agggatggaa aggatcacca gcaatattcc agtgtagcat gacaaaaatc ttagagcctt	3060
ttagaaaaca aaatccagac atagtcact atcaatacat ggatgatttg tatgtaggat	3120
ctgacttaga aatagggcag catagaacaa aaatagagga actgagacaa catctgttga	3180
ggtggggatt taccacacca gacaaaaaac atcagaaaga acctccattc ctttggatgg	3240
gttatgaact ccatcctgat aaatggacag tacagcctat agtgctgcca gaaaaggaca	3300
gctggactgt caatgacata cagaaattag tgggaaaatt gaattgggca agtcagattt	3360
atgcagggat taaagtaagg caattatgta aacttcttag gggaaccaa gcactaacag	3420
aagtagtacc actaacagaa gaagcagagc tagaactggc agaaaacagg gagattctaa	3480
aagaaccggt acatggagtg tattatgacc catcaaaaga cttaatatga gaaatacaga	3540
agcaggggca aggccaatgg acatatcaaa ttatcaaga gccatttaaa aatctgaaaa	3600
caggaaagta tgcaagaatg aagggtgccc aactaatga tgtgaaacaa ttaacagagg	3660
cagtacaaaa aatagccaca gaaagcatag taatatgggg aaagactcct aaattttaat	3720
taccataca aaaggaaaca tgggaagcat ggtggacaga gtattggcaa gccacctgga	3780
ttcctgagtg ggagtttgc aataccctc ccttagtgaa gttatggtac cagttagaga	3840

aagaacccat aataggagca gaaactttct atgtagatgg ggcagccaat agggaaacta 3900
aattaggaag agcaggatat gtaactgaca gaggaagaca aaaagttgtc cccctaacgg 3960
acacaacaaa tcagaagact gagttacaag caattcatct agctttgcag gattcgggat 4020
tagaagtaaa catagtaca gactcacaat atgcattggg aatcattcaa gcacaaccag 4080
ataagagtga atcagagtta gtcagtcaaa taatagagca gttataaaaa aaggaaaaag 4140
tctacctggc atgggtacca gcacacaaaag gaattggagg aaatgaacaa gtagataaat 4200
tggtcagtgc tggaatcagg aaagtactat ttttagatgg aatagataag gcccaagaag 4260

aacatgagaa atatcacagt aattggagag caatggctag tgattttaac ctaccacctg 4320
tagtagcaaa agaaatagta gccagctgtg ataatgtca gctaaaaggg gaagccatgc 4380
atggacaagt agactgtagc ccaggaatat ggcagctaga ttgtacacat ttagaaggaa 4440
aagttatctt ggtagcagtt catgtagcca gtggatatat agaagcagaa gtaattccag 4500
cagagacagg gcaagaaaca gcatacttcc tcttaaaatt agcaggaaga tggccagtaa 4560
aaacagtaca tacagacaat ggcagcaatt tcaccagtac tacagttaag gccgcctgtt 4620
ggtggcgagg gatcaagcag gaatttggca ttcctacaa tcccaaaagt caaggagtaa 4680

tagaatctat gaataaagaa ttaaagaaaa ttataggaca ggtaagagat caggctgaac 4740
atcttaagac agcagtacaa atggcagtat tcattccaaa ttttaaaaga aaagggggga 4800
ttggggggta cagtgcaggg gaaagaatag tagacataat agcaacagac atacaaacta 4860
aagaattaca aaaaacaaatt acaaaaattc aaaattttcg gggtttattac agggacagca 4920
gagatccagt ttggaaagga ccagcaaagc tcctctggaa aggtgaaggg gcagtagtaa 4980
tacaagataa tagtgacata aaagttagtc caagaagaaa agcaaagatc atcagggatt 5040
atggaaaaca gatggcaggt gatgattgtg tggcaagtag acaggatgag gattaacaca 5100

tggaagagat tagtaaaaca ccatatgtat atttcaagga aagctaagga ctggttttat 5160
agacatcact atgaaagtac taatccaaaa ataagttcag aagtacacat cccactaggg 5220
gatgctaaat tagtaataac aacatattgg ggtctgcata caggagaaag agactggcat 5280
ttgggtcagg gagtctccat agaattggagg aaaaagagat atagcacaca agtagaccct 5340
gacctagcag accaactaat tcattctcac tattttgatt gttttcaga atctgctata 5400
agaaatacca tattaggacg tatagttagt cctaggtgtg aatatcaagc aggacataac 5460
aaggtaggat ctctacagta cttggcacta gcagcattaa taaaacaaa acagataaag 5520

ccacctttgc ctagtgtag gaaactgaca gaggacagat ggaacaagcc ccagaagacc 5580
aagggccaca gagggagcca tacaatgaat ggacactaga gcttttagag gaacttaaga 5640
gtgaagctgt tagacatttt cctaggatat ggctccataa cttaggacaa catatctatg 5700

aaacttacgg ggatacttgg gcaggagtgg aagccataat aagaattctg caacaactgc 5760
 tgtttatcca tttcagaatt ggggtgtcgac atagcagaat aggcgttact cgacagagga 5820
 gagcaagaaa tggagccagt agatcctaga ctagagccct ggaagcatcc aggaagtcag 5880
 cctaaaactg ctigtaccaa ttgctattgt aaaaagtgtt gctttcattg ccaagtttgt 5940

 ttcattgacaa aagccttagg catctcctat ggcaggaaga agcggagaca gcgacgaaga 6000
 gctcatcaga acagtcagac tcatcaagct tctctatcaa agcagtaagt agtacatgta 6060
 atgcaaccta taatagtagc aatagtagca ttagtagtag caataataat agcaatagtt 6120
 gtgtgtgtcca tagtaatcat agaatatagg aaaatatata gacaaagaaa aatagacagg 6180
 ttaattgata gactaataga aagagcagaa gacagtggca atgagagtga aggagaagta 6240
 tcagcacttg tggagatggg ggtggaaatg gggcaccatg ctccctggga tattgatgat 6300
 ctgtagtgtc acagaaaaat tgtgggtcac agtctattat ggggtacctg tgtggaagga 6360

 agcaaccacc acitctatfff gtgcatcaga tgctaaagca tatgatacag aggtacataa 6420
 tgtttgggcc acacatgcct gtgtaccac agaccccaac ccacaagaag tagtattggt 6480
 aaatgtgaca gaaaatttta acatgtggaa aaatgacatg gtagaacaga tgcattgagga 6540
 tataatcagt ttatgggac aaagcctaaa gccatgtgta aaattaaccc cactctgtgt 6600
 tagtttaag tgcactgatt tgaagaatga tactaatacc aatagtagta gcgggagaat 6660
 gataatggag aaaggagaga taaaaaactg ctctttcaat atcagcacia gcataagaga 6720
 taagggtcag aaagaatatg cattctttta taaacttgat ataglaccaa tagataatac 6780

 cagctatagg ttgataagtt gtaacacctc agtcattaca caggcctgtc caaaggtatc 6840
 ctttgagcca attcccatc attattgtgc cccggctggt tttgcgattc taaaatgtaa 6900
 taataagacg ttcaatggaa caggaccatg tacaaatgtc agcacagtac aatgtacaca 6960
 tggaaatcagg ccagtagtat caactcaact gctgttaaat ggcagtctag cagaagaaga 7020
 ttagtaatt agatctgcca atttcacaga caatgctaaa accataatag tacagctgaa 7080
 cacatctgta gaaattaatt gtacaagacc caacaacaat acaagaaaaa gtatccgtat 7140
 ccagagggga ccaggagag catttgttac aataggaaaa ataggaaata tgagacaagc 7200

 acattgtaac attagtagag caaaatggaa tgccacttta aaacagatag ctagcaaatt 7260
 aagagaacaa tttgaaata ataaaacaat aatctttaag caatcctcag gaggggaccc 7320
 agaaattgta acgcacagtt ttaattgtgg aggggaattt ttctactgta attcaacaca 7380
 actgtttaat agtacttgg ttaatagtag ttggagtact gaagggtcaa ataactga 7440
 aggaagtgtc acaatcacac tcccatgcag aataaaacaa ttataaaca tgtggcagga 7500
 agtaggaaaa gcaatgtatg cccctcccat cagtggacaa attagatgtt catcaatat 7560

tactgggctg ctattaacaa gagatgggtg taataacaac aatgggtccg agatcttcag	7620
acctggagga ggcgatatga gggacaattg gagaagtga ttatataaat ataaagtagt	7680
aaaaattgaa ccattaggag tagcacccac caaggcaaag agaagagtgg tgcagagaga	7740
aaaaagagca gtgggaatag gagctttgtt ccttgggttc ttgggagcag caggaagcac	7800
tatgggcgca gcgtcaatga cgctgacggt acaggccaga caattattgt ctgatatagt	7860
gcagcagcag aacaatttgc tgagggtctat tgaggcgcaa cagcatctgt tgcaactcac	7920
agtctggggc atcaaacagc tccaggcaag aatcctggct gtggaaagat acctaaagga	7980
tcaacagctc ctggggattt ggggttgctc tggaaaactc atttgcacca ctgctgtgcc	8040
ttggaatgct agttggagta ataaatctct ggaacagatt tggaataaca tgacctggat	8100
ggagtgggac agagaaatta acaattacac aagcttaata cactccttaa ttgaagaatc	8160
gcaaaaccag caagaaaaga atgaacaaga attattggaa ttagataaat gggcaagttt	8220
gtggaattgg tttaacataa caaattggct gtggtatata aaattattca taatgatagt	8280
aggaggcttg gtaggtttaa gaatagtttt tgctgtactt tctatagtga atagagttag	8340
gcagggatat tcaccattat cgtttcagac ccacctcca atcccgaggg gacccgacag	8400
gcccgaagga atagaagaag aaggaggaga gagagacaga gacagatcca ttcgattagt	8460
gaacggatcc ttagcactta tctgggacga tctgcggagc ctgtgcctct tcagctacca	8520
ccgcttgaga gacttactct tgattgtaac gaggattgtg gaacttctgg gacgcagggg	8580
gtgggaagcc ctcaaatatt ggtggaatct cctacagtat tggagtcagg aactaaagaa	8640
tagtgcgtgt aacttgcctc atgccacagc catagcagta gctgagggga cagatagggt	8700
tatagaagta ttacaagcag cttatagagc tattcgccac atacctagaa gaataagaca	8760
gggcttggaa aggatattgc tataagatgg gtggcaagtg gtcaaaaagt agtgtgattg	8820
gatggcctgc tgtaaggga agaattgagc gagctgagcc agcagcagat ggggtgggag	8880
cagtatctcg agacctagaa aaacatggag caatcacaag tagcaataca gcagctaaca	8940
atgctgcttg tgcttggtta gaagcacaag aggaggaaga ggtgggtttt ccagtcacac	9000
ctcaggtacc tttaagacca atgacttaca aggcagctgt agatcttagc cactttttaa	9060
aagaaaagg gggactggaa gggctaattc actcccaaag aagacaagat atccttgatc	9120
tgtggatcta ccacacaca ggctacttcc ctgattggca gaactacaca ccagggccag	9180
gggtcagata tccactgacc tttaggatgt gctacaagct agtaccagtt gagccagata	9240
aggtagaaga ggccaataaa ggagagaaca ccagcttggt acaccctgtg agcctgcatg	9300

gaatggatga ccctgagaga gaagtgttag agtggagggt tgacagccgc ctagcatttc	9360
atcacgtggc ccgagagctg catccggagt acttcaagaa ctgctgacat cgagcttgct	9420
acaagggact ttccgctggg gactttccag ggaggcgtgg cctgggcggg actggggagt	9480
ggcgagccct cagatgtctg atataagcag ctgctttttg cctgtactgg gtctctctgg	9540
ttagaccaga tctgagcctg ggagctctct ggctaactag ggaaccact gcttaagcct	9600
caataaagct tgccttgagt gcttcaagta gtgtgtgccc gtctgtttgt tgactctggt	9660
aactagagat ccctcagacc cttttagtca gtgtggaaaa tctctagcac ccaggaggta	9720
gaggttgtag tgagccaaga tcgcgccact gcattccagc ctgggcaaga aaacaagact	9780
gtctaaaata ataataataa gtttaagggt ttaaataatat ttatacatgg aggtcataaa	9840
aatatatata ttgggctgg gcgcagtggt tcacacctgc gcccgccct ttgggaggcc	9900
gaggcaggtg gatcacctga gtttgggagt tccagaccag cctgaccaac atggagaaac	9960
cccttctctg tgiattttta gtagatttta ttttatgtgt attttattca caggtatttc	10020
tggaaaactg aaactgtttt tcctctactc tgataccaca agaatcatca gcacagagga	10080
agacttctgt gatcaaatgt ggtgggagag ggaggttttc accagcacat gagcagtcag	10140
ttctgccga gactcggcgg gtgtccttcg gttcagttcc aacaccgcct gcctggagag	10200
aggtcagacc acagggtgag ggctcagtc ccaagacata aacaccaag acataaacac	10260
ccaacaggtc caccgccct gctgccagg cagagccgat tcaccaagac gggaattagg	10320
atagagaaag agtaagtcac acagagccgg ctgtgcggga gaacggagt ctattatgac	10380
tcaaatcagt ctcccaagc attcggggat cagagttttt aaggataact tagtgtgtag	10440
ggggccagtg agttggagat gaaagcgtag ggagtcgaag gtgtcctttt gcgccgagtc	10500
agttcctggg tgggggccac aagatcggat gagccagttt atcaatccgg gggtgccagc	10560
tgatccatgg agtgcagggt ctgcaaaata tctcaagcac tgattgatct taggttttac	10620
aatagtgatg ttacccagg aacaatttgg ggaaggtcag aatctttag cctgtagctg	10680
catgactcct aaaccataat ttcttttttg ttttttttt ttttttttg agacagggtc	10740
tcactctgtc acctagctg gagtgcagtg gtgcaatcac agctcactgc agcctcaacg	10800
tcgtaagctc aagcgatcct ccacctcag cctgcctggg agctgagact acaagcgacg	10860
ccccagttaa tttttgtatt ttgtgtagag gcagcgtttt gccgtgtggc cctggctggt	10920
ctcgaactcc tgggctcaag tgatccagcc tcagcctccc aaagtgtgg gacaaccggg	10980
gccagtcact gcacctggcc ctaaaccata atttctaate ttttggttaa tttgttagtc	11040
ctacaaaggc agtctagtc ccaggcaaaa agggggtttg tttcgggaaa gggctgttac	11100
tgtctttgtt tcaaaactata aactaagttc ctctaaact tagttcggcc tacaccagg	11160

aatgaacaag gagagcttgg aggttagaag cacgatggaa ttggttaggt cagatctctt	11220
tcactgtctg agttataatt ttgcaatggt ggttcaaaga ctgcccgtt ctgacaccag	11280
tcgctgcatt aatgaatcgg ccaacgcgcg gggagaggcg gtttgcgtat tgggcgctct	11340
tccgcttcct cgctcactga ctgctgcgc tcggtcgttc ggctgcggcg agcgggtatca	11400
gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc aggaaagaac	11460
atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggcgcggtt gctggcggtt	11520
ttccataggc tccgcccc tgacgagcat cacaaaaatc gacgctcaag tcagaggtagg	11580
cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc cctcgtgcgc	11640
tctcctgttc cgacctgcc gcttaccgga tacctgtccg cctttctccc ttcgggaagc	11700
gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cgggttaggt cgttcgctcc	11760
aagctgggct gtgtgcacga accccccgtt cagcccgacc gctgcgcctt atccggtaac	11820
tatcgtcttg agtccaacc ggtaagacac gacttatcgc cactggcagc agccactggt	11880
aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa gtggtggcct	11940
aactacggct acactagaag aacagtatit ggtatctgcg ctctgctgaa gccagttacc	12000
ttcggaaaaa gagttagtag ctcttgatcc ggcaaaaaa ccaccgctgg tagcggtggt	12060
ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga agatcctttg	12120
atcttttcta cggggtctga cgctcagtgg aacgaaaact cacgttaagg gattttggtc	12180
atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg aagttttaaa	12240
tcaatctaaa gtatatatga gtaaaccttg tctgacagtt accaatgctt aatcagttag	12300
gcacctatct cagcgatctg tctatttcgt tcattcatag ttgcctgact ccccgctgtg	12360
tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat gataccgca	12420
gaccacgct caccggtcc agatttatca gcaataaacc agccagccgg aaggccgag	12480
cgcagaagtg gtctgcaac ttatccgcc tccatccagt ctattaattg ttgccgggaa	12540
gctagagtaa gtagttgcc agttaatagt ttgcgcaacg ttgttgccat tgctacaggc	12600
atcgtgggtg cagctcgtc gtttggtatg gcttattca gctccggttc ccaacgatca	12660
aggcgagtta catgatcccc catgttgtgc aaaaaagcgg ttagctcctt cggtcctccg	12720
atcgttgta gaagtaagtt ggccgcagtg ttatcactca tggttatggc agcactgcat	12780
aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactggtga gtactcaacc	12840
aagtcattct gagaatagt tatgcggcga ccgagttgct cttgcccgcg gtcaatacgg	12900
gataataccg cgccacatag cagaacttta aaagtgtca tcattggaaa acgtttctcg	12960
ggcgaaaaac tcicaaggat cttaccgtg ttgagatcca gttcgatga acccactcgt	13020

gcaccaact gatcttcagc atcttttact ttcaccagcg tttctgggtg agcaaaaaca	13080
ggaaggcaaa atgccgcaaa aaaggaata agggcgacac ggaaatgttg aatactcata	13140
ctcttccttt ttcaatatta ttgaagcatt tatcagggtt attgtctcat gagcggatac	13200
atatttgaat gtatttagaa aaataaaca ataggggttc cgcgcacatt tccccgaaaa	13260
gtgccacctg agctctaaga aaccattatt atcatgacat taacctataa aaataggcgt	13320
atcacgagge ctttctgtc cgcgcggttc ggtgatgacg gtgaaaacct ctgacacatg	13380
cagctcccg agacggtcac agcttgtctg taagcggatg cggggagcag acaagcccgt	13440
caggcgcggt cagcgggtgt tggcgggtgt cggggctggc ttaactatgc ggcatcagag	13500
cagattgtac tgagagtga ccatatgcgg tgtgaaatac cgcacagatg cgtaaggaga	13560
aaataccga tcaggcgcca ttccgcatc aggtctgcga actgttggga agggcgatcg	13620
gtcggggcct cttcgtatt acgccagggg aggcagagat tgcagtaagc tgagatcgca	13680
gcactgcact ccagcctggg cgacagagta agactctgtc tcaaaaataa aataaataa	13740
tcaatcagat attccaatct tttcctttat ttatttattt attttctatt ttggaaacac	13800
agtccttctt tattccagaa ttacacatat attctatttt tctttatatg ctccagtttt	13860
ttttagacct tcacctgaaa tgttgtata caaatctag gccagtccag cagagcctaa	13920
aggtaaaaa taaataata aaaaataaat aaaatctagc tcactccttc acatcaaat	13980
ggagatacag ctgttagcat taaataccaa ataaccatc ttgtcctcaa taattttaag	14040
cgctctctc caccacatct aactctgtc aaaggcatgt gcccttccg ggcgctctgc	14100
tgtgtgcca accaactggc atgtggactc tgcagggtcc ctaactgcca agccccacag	14160
tgtgccctga ggctgccct tcttctagc ggctgcccc actcggttt gctttcccta	14220
gtttcagtta ctgtcgctta gccaaagtct gaaactaggt gcgcacagag cggttaagact	14280
gcgagagaaa gagaccagct ttacaggggg tttatcacag tgcacctga cagtcgtcag	14340
cctcacaggg ggittatcac attgcacct gacagtcgtc agcctcacag ggggtttatc	14400
acagtgcacc cttacaatca ttccatttga ttcacaattt ttttagtctc tactgtgcct	14460
aacttgaag ttaaatttga tcagagggtgt gttcccagag gggaaaacag tatatacagg	14520
gttcagtact atcgcatctc aggcctccac ctgggtcttg gaatgtgtcc cccgagggt	14580
gatgactacc tcagttggat ctccacaggt cacagtgaca caagataacc aagacacctc	14640
ccaaggctac cacaatgggc cgccctccac gtgcacatgg ccggagggaac tgccatgtcg	14700
gaggtgcaag cacacctgcg catcagagtc cttggtgtgg agggagggaac cagcgcagct	14760

tccagccatc cacctgatga acagaacctt gggaaagccc cagttctact tacaccagga 14820
aaggc 14825
<210> 385
<211> 10535
<212> DNA
<213> Simian immunodeficiency virus
<400> 385
gcattgcacat tttaaaggct ttgtctaaat atagccaaaa gtccttctac aaattttcta 60
agagttctga ttcaaagcag taacaggcct tgtctcatca tgaactttgg catctcatct 120
acagctaagt ttatatcata aatagttctt tacaggcagc accaacttat acccttatag 180
catactttac tgtgtgaaaa ttgcatcttt cattaagctt actgtaaatt tactggctgt 240
cttccttgca ggtttctgga agggatttat tacagtgcaa gaagacatag aatcttagac 300
atatacttag aaaaggaaga aggcatacata ccagattggc aggattacac ctcaggacca 360
ggaattagat acccaaagac atttggtgga ctatggaaat tagtcctgt aaatgtatca 420
gatgaggcac aggaggatga ggagcattat ttaatgcac cagctcaaac tccccagtgg 480
gatgaccctt ggggagaggt tctagcatgg aagtttgatc caactctggc ctacacttat 540
gaggcatatg ttagataccc agaagagttt ggaagcaagt caggcctgtc agaggaagag 600
gttagaagaa ggctaaccgc aagaggcctt cttaacatgg ctgacaagaa ggaaactcgc 660
tgaaacagca gggaacttcc acaaggggat gttacgggga ggtactggg aggagccggt 720
cggaacgcc cactttcttg atgtataaat atcactgcat ttcgctctgt attcagtcgc 780
tctgcggaga ggctggcaga ttgagccctg ggaggttctc tccagcacta gcaggtagag 840
cctgggtgtt cctgctaga ctctcaccag cacttggccg gtgctgggca gagtgactcc 900
acgcttgctt gcttaaagcc ctcttcaata aagctgcat tttagaagta agctagtgtg 960
tgttcccatc tctcctagcc gccgcctggt caactcggtc ctcaataata agaagaccct 1020
ggtctgttag gaccctttct gctttgggaa accgaagcag gaaaatccct agcagattgg 1080
cgcctgaaca gggaacttgaa ggagagttag agactcctga gtacggctga gtgaaggcag 1140
taaggcggc aggaaccaac cagcagggag tgcctctata aaggcgcggg tcggtaccag 1200
acggcgttag gaggcggaga ggaagaggcc tccggttgca ggtaagtga acacaaaaaa 1260
gaaatagctg tcttttatcc aggaaggggt aataagatag agtgggagat gggcgtgaga 1320
aactccgtct tgtcagggaa gaaagcagat gaattagaaa aaattagct acgacccaac 1380
ggaaagaaaa agtacatgtt gaagcatgta gtatgggcag caaatgaatt agatagattt 1440

ggattagcag aaagcctgtt ggagaacaaa gaaggatgtc aaaaaatact ttcggtctta 1500

gctccattag tgccaacagg ctcaaaaaat ttaaaaagcc ttataatac tgtctgcgtc 1560

atctggtgca ttcacgcaga agagaaagtg aaacacactg aggaagcaaa acagatagtg 1620

cagagacacc tagtgggtga aacaggaaca acagaaacta tgccaaaaac aagtagacca 1680

acagcaccat ctacgggcag aggaggaaat tacccagtac aacaatagg tgtaactat 1740

gtccacctgc cattaagccc gagaacatta aatgcctggg taaaattgat agaggaaaag 1800

aaatttggag cagaagtagt gccaggattt caggcactgt cagaaggtt cacccttat 1860

gacattaatc agatgttaaa ttgtgtggga gaccatcaag cggctatgca gattatcaga 1920

gatattataa acgaggaggc tgcagattgg gacttgcagc acccacaacc agctccacaa 1980

caaggacaac ttagggagcc gtcaggatca gatattgcag gaacaactag ttcagtagat 2040

gaacaaatcc agtggatgta cagacaacag aacccatac cagtaggcaa catttacagg 2100

agatggatcc aactggggtt gcaaaaatgt gtcagaatgt ataaccaac aaacattcta 2160

gatgtaaaac aagggccaaa agagccattt cagagctatg tagacaggtt ctacaaaagt 2220

ttaagagcag aacagacaga tgcagcagta aagaattgga tgactcaaac actgctgatt 2280

caaatgcta acccagattg caagctagtg ctgaaggggc tgggtgtgaa tcccacccta 2340

gaagaaatgc tgacggcttg tcaaggagta ggggggccgg gacagaaggc tagattaatg 2400

gcagaagccc tgaaagagcc cctgcacca gtgccaatcc cttttgcagc agcccaacag 2460

aggggaccaa gaaagccaat taagtgttgg aattgtggga aagagggaca ctctgcaagg 2520

caatgcagag cccaagaag acagggatgc tggaatgtg gaaaaatgga ccatgttatg 2580

gccaaatgcc cagacagaca ggcgggtttt ttaggccttg gtccatgggg aaagaagccc 2640

cgcaatttcc ccatggctca agtgcacag ggctgatgc caactgctcc cccagaggac 2700

ccagctgttg atctgctaaa gaactacatg cagttgggca agcagcagag agaaaagcag 2760

agagaaagca gagagaagcc ttacaaggag gtgacagagg atttctgca cctcaattct 2820

ctctttggag gagaccagta gtcactgtc atattgaagg acagcctgta gaagtattac 2880

tggatacagg ggctgatgat tctattgtaa caggaataga gttaggtcca cattataccc 2940

caaaaatagt aggaggaata ggaggtttta ttaatactaa agaatacaaa aatgtagaaa 3000

tagaagtttt aggcaaaagg attaaaggga caatcatgac aggggacacc ccgattaaca 3060

tttttggtag aaatttgcta acagctctgg ggatgtctct aaattttccc atagctaaag 3120

tagagcctgt aaaagtcgcc ttaaagccag gaaaggatgg accaaaattg aagcagtggc 3180

cattatcaaa agaaaagata gttgcattaa gagaaatctg tgaaaagatg gaaaaggatg 3240
 gtcagtgtga ggaagctccc ccgaccaatc catacaacac cccacattt gctataaaga 3300
 aaaaggataa gaacaaatgg agaagtctga tagatttttag ggaactaaat agggtcactc 3360
 aggactttac ggaagtccaa ttaggaatac cacaccctgc aggactagca aaaaggaaaa 3420
 gaattacagt actggatata ggtgatgcat atttctccat acctctagat gaagaattta 3480
 ggcagtacac tgcctttact ttaccatcag taaataatgc agagccagga aaacgataca 3540
 ttataaggt tctgcctcag ggatggaagg ggtcaccagc catcttcaa tacactatga 3600

gacatgtgct agaacccttc aggaaggcaa atccagatgt gaccttagtc cagtatatgg 3660
 atgacatctt aatagctagt gacaggacag acctggaaca tgacagggtg gttttacagt 3720
 caaaggaact ctigaatagc atagggtttt ctacccaga agagaaattc caaaaagatc 3780
 cccatttca atggatgggg tacgaattgt ggccaacaaa atggaagtgt caaaagatag 3840
 agttgccaca aagagagacc tggacagtga atgatatata gaagttagta ggagtattaa 3900
 attgggcagc tcaaatttat ccaggtataa aaaccaacaa tctctgtagg ttaattagag 3960
 gaaaaatgac tctaacagag gaagttcagt ggactgagat ggcagaagca gaatatgagg 4020

aaaataaaat aattctcagt caggaacaag aaggatgtta ttaccaagaa ggcaagccat 4080
 tagaagccac ggtaataaag agtcaggaca atcagtggtc ttataaaatt caccaagaag 4140
 acaaaatact gaaagtagga aaatttgcaa agataaagaa tacacatacc aatggagtga 4200
 gactattagc acatgtaata cagaaaatag gaaaggaagc aatagtgatc tggggacagg 4260
 tcccaaaatt ccacttacca gttgagaagg atgtatggga acagtgggtg acagactatt 4320
 ggcaggtaac ctggataccg gaatgggatt ttatctcaac accaccgcta gtaagattag 4380
 tcttcaatct agtgaaggac cctatagagg gagaagaac ctattataca gatggatcat 4440

gtaataaaca gtcaaaagaa gggaaagcag gatatatcac agataggggc aaagacaaag 4500
 taaaagtgtt agaacagact actaatcaac aagcagaatt ggaagcattt ctcattggcat 4560
 tgacagactc agggccaaag gcaaatatta tagtagattc acaatatgtt atgggaataa 4620
 taacaggatg ccttacagaa tcagagagca ggctagttaa tcaaataata gaagaaatga 4680
 ttaaaaagtc agaaatttat gtagcatggg taccagcaca caaaggata ggaggaaacc 4740
 aagaaataga ccacctagt agtcaaggga ttagacaagt tctcttcttg gaaaagatag 4800
 agccagcaca agaagaacat gataaatacc atagtaatgt aaaagaattg gtattcaaat 4860

ttggattacc cagaatagt gccagacaga tagtagacac ctgtgataaa tgtcatcaga 4920
 aaggagaggc tatacatggg caggcaaat cagatctagg gacttggcaa atggattgta 4980
 cccatctaga gggaaaaata atcatagttg cagtacatgt agctagtgga ttcatagaag 5040

cagaggtaat tccacaagag acaggaagac agacagcact atttctgtta aaattggcag 5100

gcagatggcc tattacacat ctacacacag ataatggtgc taactttgct tcgcaagaag 5160

taaagatggt tgcattggtg gcagggatag agcacacctt tggggtacca tacaatccac 5220

agagtcaggg agtagtgga gcaatgaatc accacctgaa aaatcaaata gatagaatca 5280

gggaacaagc aaattcagta gaaaccatag tattaatggc agttcattgc atgaatttta 5340

aaagaagggg aggaataggg gatatgactc cagcagaaag attaattaac atgatcacta 5400

cagaacaaga gatacaattt caacaatcaa aaaactcaaa atttaaaaat tttcgggtct 5460

attacagaga aggcagagat caactgtgga agggacccgg tgagctattg tggaaagggg 5520

aaggagcagt catcttaag gtagggacag acattaaggt agtaccaga agaaaggcta 5580

aaattatcaa agattatgga ggaggaaaag aggtggatag cagttccac atggaggata 5640

ccggagaggc tagagaggtg gcatagcctc ataaaatc tgaaatataa aactaaagat 5700

ctacaaaagg ttgctatgt gccccatttt aaggctggat gggcatggtg gacctgcagc 5760

agagtaatct tcccactaca ggaaggaagc catttagaag tacaagggtg ttggcatttg 5820

acaccagaaa aagggtggct cagtacttat gcagttagga taacctggta ctcaaagaac 5880

ttttggacag atgtaacacc aaactatgca gacattttac tgcatagcac ttatttcct 5940

tgctttacag cgggagaagt gagaagggcc atcaggggag aacaactgct gtcttgctgc 6000

aggttcccga gagtcataa gtaccaggta ccaagcctac agtacttagc actgaaagta 6060

glaagcgatg tcagatccca gggagagaat cccacctgga aacagtggag aagagacaat 6120

aggagaggcc ttgcaatggc taaacagaac agtagaggag ataaacagag aggcggtaaa 6180

ccacctacca agggagctaa ttttccaggt ttggcaagg tcttgggaat actggcatga 6240

tgaacaaggg atgtaccaa gctatgtaaa atacagatac ttgtgtttta taaaaaggc 6300

tttatttatg cattgcaaga aaggctgtag atgtctaggg gaaggacatg gggcaggggg 6360

atggagacca ggacctctc ctctcccc tccaggacta gcataaatgg aagaaagacc 6420

tccagaaaat gaaggaccac aaaggaacc atgggatgaa tgggtagtgg aggttctgga 6480

agaactgaaa gaagaagctt taaaacattt tgatcctcgc ttgctaactg cacttggtaa 6540

tcatactat aatagacatg gagacacct tgaggagca ggagaactca ttagaatcct 6600

ccaacgagcg ctcttcagc atttcagagg cggatgcac cactccagaa tcggccaacc 6660

tgggggagga aatcctctct cagctatacc gcctctaga agcatgctat aacacatgct 6720

attgtaaaa gtgttgctac cattgccagt ttgttttct taaaaaggc ttggggatat 6780

gttatgagca atcacgaaag agaagaagaa ctccgaaaaa ggctaaggct aatacatctt 6840

ctgcatcaaa caagtaagta tgggatgtct tgggaatcag ctgcttatcg ccatcttgct 6900

ttaaagtgtc taigggatct attgtactct atatgtcaca gtcitttatg gtgtaccagc	6960
ttggaggaat gcgacaattc cccctcttttg tgcaaccaag aataggata cttggggaac	7020
aactcagtgc ctaccagata atgggtgatta ttcagaagtg gcccttaatg ttacagaaag	7080
ctttgatgcc tggaaataata cagtcacaga acaggcaata gaggatgtat ggcaactctt	7140
tgagacctca ataaagcctt gtgtaaaatt atccccatta tgcattacta tgagatgcaa	7200
taaaagttag acagatagat ggggattgac aaaatcaata acaacaacag catcaacaac	7260
atcaacgaca gcatcagcaa aagtagacat ggtcaatgag actagtctct gtatagccca	7320
ggataattgc acaggcttgg aacaagagca aatgataagc tgtaaattca acatgacagg	7380
gttaaaaaga gacaagaaaa aagagtacaa tgaaacttgg tactctgcag atttggtatg	7440
tgaacaaggg aataaactg gtaatgaaag tagatgttac atgaaccact gtaacacttc	7500
tgttatccaa gagtcttctg acaaacatta ttgggatgct attagattta ggtattgtgc	7560
acctccaggt tatgctttgc ttagatgtaa tgacacaaat tattcaggct ttatgcctaa	7620
atgttctaag gtggtggtct ctcatgcac aaggatgatg gagacacaga cttctacttg	7680
gtttggcttt aatggaacta gagcagaaaa tagaacttat atttactggc atggtaggga	7740
taataggact ataattagtt taaataagta ttataatcta acaatgaaat gtagaagacc	7800
aggaaataag acagttttac cagtcacat tatgtctgga ttggttttcc actcacaacc	7860
aatcaatgat aggccaaagc aggcattggtg ttggttttga ggaaaatgga aggatgcaat	7920
aaaagaggtg aagcagacca ttgtcaaaca tcccaggtat actggaacta acaatactga	7980
taaaatcaat ttgacggctc ctggaggagg agatccgga gttaccttca tgtggacaaa	8040
ttgcagagga gagtctctct actgtaaaat gaattggttt ctaaatggg tagaagatag	8100
gaatacagct aaccagaagc caaaggaaca gcataaaagg aattacgtgc catgtcatat	8160
tagacaaata atcaacactt ggcataaagt aggcaaaaat gtttatttgc ctccaagaga	8220
gggagacctc acgtgtaact ccacagtgc cagtctcata gcaaacatag attggattga	8280
tggaaaccaa actaatatca ccatgagtgc agagggtgca gaactgtatc gattggaatt	8340
gggagattat aaattagtag agatcactcc aattggcttg gccccacag atgtgaagag	8400
gtacactact ggtggcacct caagaaataa aagaggggtc tttgtgctag ggttcttggg	8460
ttttctcgca acggcaggtt ctgcaatggg cgcggcgtcg ttgacgtga ccgtcagtc	8520
ccgaacttta ttggctggga tagtgcagca acagcaacag ctgttgacg tggtaagag	8580
acaacaagaa ttgttgcgac tgaccgtctg gggaacaaag aacctccaga ctaggtcac	8640

tgccatcgag aagtacttaa aggaccaggc gcagctgaat gcttggggat gtgcgttttag	8700
acaagtctgc cacactactg taccatggcc aaatgcaagt ctaacaccaa agtggaaaca	8760
tgagacttgg caagagtggg agcgaaaggt tgacttcttg gaagaaaata taacagccct	8820
cctagaggag gcacaaattc aacaagagaa gaacatgtat gaattacaaa agttgaatag	8880
ctgggatgtg tttggcaatt ggtttgacct tgcttcttgg ataaagtata tacaatatgg	8940
agtttatata gttgtaggag taatactgtt aagaatagtg atctatatag taaaaatgct	9000
agctaagtta aggcaggggt ataggccagt gtctctctcc ccacctctt atttcagca	9060
gacccatatc caacaggacc cggcactgcc aaccagagaa ggcaaagaaa gagacggtgg	9120
agaaggcggg ggcaacagct cctggccttg gcagatagaa tatattcatt tcctgatccg	9180
ccaactgata cgctcttga cttggctatt cagcaactgc agaaccttgc tatcgagagt	9240
ataccagatc ctccaaccaa tactccagag gctctctgcg accctacaga ggattcgaga	9300
agtcctcagg actgaactga cctacctaca atatgggtgg agctatttcc atgaggcggg	9360
ccaggccgtc tggagatctg cgacagagac tcttgcgggc gcgtggggag acttatggga	9420
gactcttagg agaggtggaa gatggatact cgcaatcccc aggaggatta gacaagggct	9480
tgagctcact ctcttgtgag ggacagaaat acaatcaggg acagtatatg aatactccat	9540
ggagaaaccc agctgaagag agagaaaaat tagcatacag aaaacaaaat atggatgata	9600
tagatgagta agatgatgac ttggtagggg tatcagttag gccaaaagtt cccctaagaa	9660
caatgagtta caaattggca atagacatgt ctcatTTTTat aaaagaaaag gggggactgg	9720
aagggatTTa ttacagtgca agaagacata gaatcttaga catatactta gaaaaggaag	9780
aaggcatcat accagattgg caggattaca cctcaggacc aggaattaga tacccaaaga	9840
catttggctg gctatggaaa ttagtccctg taaatgtatc agatgaggca caggaggatg	9900
aggagcatta tttaatgcat ccagctcaaa ctccccagtg ggatgaccct tggggagagg	9960
ttctagcatg gaagtttgat ccaactctgg cctacactta tgaggcatat gttagatacc	10020
cagaagagtt tggaagcaag tcaggcctgt cagaggaaga ggtagaaga aggctaaccg	10080
caagaggcct tcttaacatg gctgacaaga aggaaactcg ctgaaacagc agggactttc	10140
cacaagggga tgttacgggg aggtactggg gaggagccgg tcgggaacgc ccactttctt	10200
gatgtataaa tatcactgca tttcgctctg tattcagtgc ctctcggag aggctggcag	10260
attgagccct gggaggttct ctccagcact agcaggtaga gcctgggtgt tccctgctag	10320
actctacca gcacttgccc ggtgctgggc agagtgactc cacgcttgct tgcttaaagc	10380
cctcttcaat aaagctgcca ttttagaagt aagctagtgt gtgttcccat ctctcctagc	10440
cgccgcctgg tcaactcggg actcaataat aagaagaccc tggctctgta ggaccctttc	10500

tgctttggga aaccgaagca ggaaaatccc tagca 10535

<210> 386

<211> 9713

<212> DNA

<213> Human immunodeficiency virus 2

<400> 386

agtcgctctg cggagaggct ggcagattga gccctgggag gttctctcca gcactagcag 60

gtagagcctg ggtgttcctt gctagactct caccggtgct tggccggcac tgggcagacg 120

gtccacgct tgettgtta aaagacctct taataaagct gccagttaga agcaagttaa 180

gtgtgtgttc ccatctctcc tagtcgccgc ctggtcattc ggtgttcac tgaataacaa 240

gacctggtc tgttaggacc ctttctgctt tgggaaacca aagcaggaaa atccctagca 300

ggttggcgcc cgaacaggga cttagagaag actgaaaagc cttggaacac ggctgagtga 360

aggcagtaag ggcggcagga acaaacaccg acggagtgtt cctagaaagg cgcaggccaa 420

ggtacaaaag gcggcgtgtg gagcgggagt aaagaggcct ccgggtgaag gtaagtacct 480

acacaaaaaa attgtagcca ggaagggtt gttatcctac ctttagacag gtagaagatt 540

gtgggagatg ggcgcgagaa actccgtctt gaaaggga aaagcagacg aattagaaac 600

aattaggtta cgccccggcg gaaagaaaaa atacaggcta aagcatattg tgtgggcagc 660

gaatgaattg gacagattcg gattagcaga gacctgttg gagtcaaaag aaggttgcca 720

aagaattctt acagttttag gtccattagt accgacaggt tcagaaaatt taaaaagcct 780

ttttaatact gtctgcgtca ttgtgtgcat acacgcagaa gagaaagtga aagatactga 840

aggagcaaaa caaataglac agagacatct agcggcagaa acaggaactg cagagaaaat 900

gccaaataca agtagaccaa cagcaccacc tagcgggaag ggaggaaact tccccgtaca 960

acaagtaggc ggcaattata cccatgtgcc gctgagtcct cgaaccctaa atgcttgggt 1020

aaaattagta gaggaaga agttcggggc agaggtagtg ccaggatttc aggcactctc 1080

agaaggctgc acgccctatg atatcaacca aatgcttaat tgtgtgggcg accatcaagc 1140

agctatgcaa ataacaggg agatcgtaa tgaagaagca gcagattggg atgtgcaaca 1200

tccaatacca ggtcccttac cagcggggca gcttagagaa ccaagagggt ctgacatagc 1260

agggacaaca agcacagtag atgaacagat ccagtggatg tttaggccac aaaatcccgt 1320

accagtggga aacatctata ggagatggat ccagatagga ctgcagaagt gcgtcaggat 1380

gtacaacccg accaacaatc tagacataaa acaaggacca aaggaacat tccaaagtta 1440

tgtagataga ttctacaaaa gcttgagggc agaacaaca gatccagcag tgaagaattg 1500

gatgaccag acactactag tacagaatgc caaccagac tgtaaattag tactaaaagg	1560
actagggatg aatcctacct tagaagagat gctaaccgcc tgccaagggg taggtgggcc	1620
agccagaaa gctagactaa tggcagaagc cttaaaagag gccttgacac cagcccctat	1680
cccatttgca gcagcccagc agaaaaggac aattaaatgc tggaattgtg gaaaggaagg	1740
acactcggca agacaatgcc gagcacctag aagacagggc tgctggaagt gtggtaaacc	1800
aggacatgtc atagcaaatt gcccagatag acaggtgggt ttttaggga tgggcccccg	1860
gggaaagaag ccccgaact tcccgtggc ccaagtcccg caggggctaa caccaacagc	1920
acccccagta gatccagcag tggacctact ggagaattat atgcagcaag gaaaaagaca	1980
aagagaacag agagagagac catacaaaga agtgacagag gacttactgc acctcgagca	2040
gggagaggca ccatgcagag agacgacaga ggacttgctg cacctcaatt ctctctttg	2100
aaaagaccag tagtcacggc atacgtcgag ggccagccag tagaagttct gctagacagc	2160
ggggctgacg actcaatagt agcagggata gagttaggga gcaattatag tccaaagata	2220
gtaggaggaa tagggggatt cataaatacc aaggaatata aaaatgtaa aatagaagtt	2280
ttaggtaaaa aggtaaaggc caccataatg acaggtgaca cccaatcaa catttttggc	2340
agaaatattc tgacagcctt aggcattgtc ttaaatttac cagtcgcaa aatagaacca	2400
ataaaaaata tgtaaagcc aggaaaagat ggacaaaaac tgaggcaatg gcccttaaca	2460
aaagaaaaaa tagaggcact aaaagaaatc tgtgaaaaaa tggaaagaga aggccagcta	2520
gaggaagcgc ctccaactaa tccttataac acccccacat ttgcaatcaa gaaaaaggac	2580
aaaaataaat ggaggatgct aatagatttt agagaactaa acaaggtaac tcaagatttc	2640
acagaaattc agttaggaat tccacacca gcaggattgg ccaagaaaaa aagaattact	2700
gtactagata taggggatgc ttacttttcc ataccactac atgaagactt tagacagtat	2760
actgcattta cttaccatc aataaacaat gcagaaccag gaaaaagata tatatataag	2820
gtcctgcctc agggatggaa ggggtcacca gcaatttttc aatacacaat gaggcaggtc	2880
ttagaaccat tcagaaaagc aaacctagat gtcattatca ttcagtacat ggatgatatc	2940
ctaatagcta gtgacaggac agatctagaa catgacaagg tggctcctgca gctaaaggaa	3000
cttctaata acctaggatt ttctaccca gatgagaagt tccaaaagga cctccatac	3060
cactggatgg gctatgaact gtggccaact aagtggaagc tgcagaagat acagttgccc	3120
caaaaagatg taigcacagt aaatgacatc caaaagttag tgggtgtctt aaactgggca	3180
gcacaaatct acccagggat aaaaaccaga cacttatgta agctaattag aggaaaaatg	3240
acactcacag aagaagtaca gtggacagaa ctacgagagg cggagttaga agagaacaag	3300
attatcttaa gccaggagca agagggacac tattaccaag aagaaaaaga gttagaagca	3360

acagtccaaa aggatcaaga caatcagtgg acatataaag tacaccaggg agagaaaatt	3420
ctaaaagtag ggaaatatgc aaagataaaa aataccata ccaatggggt cagattgtta	3480
gcacaagtag ttcaaaagat aggaaaagaa gcactaatca tttggggacg aatacaaaaa	3540
tttcacctac cagtagaaag agagacatgg gaacagtggg gggatgacta ctggcaggtg	3600
acatggatcc ctgactggga cttcgtatct accccgccgc tggtcagact agcatttaac	3660
ctggtaaaag atcctatacc aagaacagag actttctaca cagatggatc ctgcaatagg	3720
caatcaaagg aagggaaaagc aggatatgta acagatagag ggagagacaa ggtaaggatg	3780
ctagaacaaa ctaccaatca gcaagcagaa ttagaagcct ttgcaatggc actaacagac	3840
tcagggtccaa aagccaatat tatagtagac tcacagtatg taatggggat agtagcaggc	3900
cagccaacag aatcagagag tagaatagta aatcaaatca tagaggagat gataaaaaag	3960
gaagcaatct atgttgcagt ggtcccagcc cataaaggca taggaggga ttaggaggtg	4020
gatcagttag taagtcaggg catcagacaa gtgttgttcc tggaaaaaat agagcccgct	4080
caggaagaac atgagaaata ccatagcaat gtaaaagaac tatccataa atttgattg	4140
cccaaatag tagcaagaca aatagtaaac acatgtgccc aatgtcaaca gaaaggggag	4200
gctatacatg ggcaagtaga tgcagaatta ggcacttggc aaatggactg cacacactta	4260
gaaggaaaga tcattatagt agcagtacat gttgcaagtg gattcataga agcagaagtc	4320
atcccacagg aatcaggaag gcagacagca ctcttctat taaaactggc cagtaggtgg	4380
ccaataacac acttgcacac agataatggg gccacttca cttcacagga agtaaaaatg	4440
gtagcatggt gggtaggtat agaacaatct ttcggagtac cttacaatcc acaaagccaa	4500
ggagtagtag aagcaatgaa tcaccacctt aaaaatcaga taagtagaat tagagaacag	4560
gcaaatacag tagaaacaat agtactgatg gcaacacact gcatgaattt taaaagaagg	4620
ggaggaatag gggatatgac cccagcagaa agactaatca atatgatcac cacagaacaa	4680
gaaatacaat tcctccacgc caaaaattca aaattaataa attttcgggt ctatttcaga	4740
gaaggcagag atcagctgtg gaaaggaccc gggaactac tgtggaaggg agacggagca	4800
gtcatagtca aggtagggac agacataaaa gtagtaccaa ggaggaaagc caagatcatc	4860
aaagactatg gaggaaggca agaactggat agtggttccc acttggaggg tgccagggag	4920
gatggagaaa tggcatagcc ttgtcaata tctaaaatac agaacaaaag atctagaaga	4980
cgtgtgctat gtccccacc ataaagtagg atgggcatgg tggacttgca gcagggtaat	5040
attccatta aagggaacaa gtcactaga aatacaggca tattggaacc taacgccaga	5100

aaaaggatgg ctctcctctt attcagtaag aatgacttgg tatacggaaa ggttctggac 5160

 agatgttacc ccagactgtg cagactccct aatacatagc acttatttct ctgtctttac 5220
 agcaggtgaa gtaagaagag ccatcagagg ggaaaagtta ttgtcctgct gcaattatcc 5280
 ccaagcccat agagcccagg taccgtcact ccaatTTTTG gccttagtgg tagtgcagca 5340
 aatatgacaga cccagagaa acggtacccc caggaaacag tggcgaagag actatcgaag 5400
 aggccttcaa ttggctagac aggacggtag aagccataaa cagagaggca gtgaatcacc 5460
 tgccccgaga gcttatttct cagggtgtggc agaggtcctg gagatactgg catgatgaac 5520
 aagggatgtc acaaagttac acaaagtata gatatttgtg cttataacag aaggctatgt 5580

 tcacacattg taagagaggg tgcacttgcc tggggggagg acatgggcca ggagggtgga 5640
 gaccaggacc tccccctct cccctccag gtctagtcta atgactgaag caccaacaga 5700
 gtttccccg gaggatggga cccaccgag ggaaccaggg gatgagtgga taatagaaat 5760
 cctgagaaaa ataaagaaag aagctttaa gcattttgac cctcgcttgc taactgctct 5820
 tggcaactat atccatacta gacatggaga cacccttgaa ggccgagag agctcattaa 5880
 tgtcctacaa cgagccctct tcatgcactt cagagcggga tgtaggctct caagaattgg 5940
 ccaaacaggg ggaagaactc ctttcccagc tacatcgacc cctagaacca tgcaataaca 6000

 aatgctattg taaaggatgc tgcctccact gccagctgtg ttttttaaac aaggggctcg 6060
 ggatatgtta tgaccggaag ggcagacgaa gaagaactcc gaagaaaact aaggctcatt 6120
 catcttctgc atcagacaag tgagtatgat ggggtgtaga aatcagctgc ttgttgccat 6180
 ttgtctaact agtacttgc tgcataattg caccaattat gtgactgttt tctatggcat 6240
 acccgctgg agaaatgcat ccattcccct cttttgtgca accaagaata gggatacttg 6300
 gggaaccata cagtgttgc cagacaatga tgattatcag gagataactt tgaatgtgac 6360
 agaggtttc gatgcatggg ataatacagt aacagaacaa gcaatagaag atgtctggaa 6420

 tctatttgag acatcaataa aaccatgtgt caaattaacg ctttatgtg tagcaatgag 6480
 atgtaacaac acagatgcaa ggaacacaac cacaccaca acagcatccc cgcgtacaat 6540
 aaaaccgtg acagagataa gtgagaattc ctcatgcata cgcgcaaaca actgctcagg 6600
 attgggagaa gaagagggtg tcaattgtca attcaatatg acaggattag agagagataa 6660
 gaaaaagcaa tatagtgaga catggtactc gaaggatgta gtttgtgaag gaaatggcac 6720
 cacagataca tgttacatga accattgcaa cacatcggtc atcacagagt catgtgacaa 6780
 gcactattgg gatgctatga ggtttagata ctgtgcacca ccaggttttg ccctactaag 6840

atgcaatgat accaattatt caggctttgc gcccaattgc tctaaggtag tagctgctac	6900
atgcaccaga atgatggaaa cgcaaaacttc tacatggttt ggctttaatg gcactagagc	6960
agaaaataga acatttatct attggcatgg tagggataac agaactatca tcagcttaaa	7020
caaatattat aatctcacta tacattgtaa gaggccagga aataagacag tggtagcaat	7080
aacacttatg tcagggttaa ggtttcactc ccagccggtc atcaataaaa gaccagaca	7140
agcatggtgt tggttcaaag gtgaatggaa gggagccatg caggaggtga aggaaacct	7200
tgcaaacat cccaggata aaggaaccaa tgaacaaaag aatattaact ttacagcacc	7260
aggaaagggc tcagaccag aggtggcata catgtggact aactgcagag gagaatttct	7320
ctactgcaac atgacttggc tcctcaattg gatagaaaat aagacacacc gcaattatgt	7380
accgtgccat ataagacaaa taattaacac ctggcataag gtagggaaaa atgtatattt	7440
gcctcccagg gaaggggagt tgacctgcaa ctcaacagta actagcataa ttgctaaca	7500
tgatgcaaat ggaataata caaatattac ctttagtgca gaggtggcag aactataccg	7560
attagattg ggagattata aattggtaga aataacacca attggcttcg cacctacagc	7620
agaaaaaga tactctcta ctccaatgag gaacaagaga ggtgtgttcg tgctagggtt	7680
cttgggtttt ctgcaacag caggctctgc aatgggcgcg gcgtccttaa cgctgtcggc	7740
tcagtctcgg actttactgg ccgggatatg gcagcaacag caacagctgt tggacgtgg	7800
caagagacaa caggaaatgt tgcgactgac cgtctgggga acaaaaaatc tccaggcaag	7860
agtcactgct atcgagaagt acttaagga ccaggcgcaa ctaaattcat ggggatgtgc	7920
atttagacaa gtctgccaca ctactgtacc atgggtaa atgatacctaa cgcctgagt	7980
gaacaatatg acgtggcaag aatgggaagg caaaatccgc gacctggagg caaatatcag	8040
tcaacaatta gaacaagcac aaattcagca agagaagaat atgtatgaac tacaaaagtt	8100
aaatagctgg gatgtttttg gtaactggtt tgacttaacc tcctggatca agtatattca	8160
atatggagtt tatataataa taggaatagt agttcttaga atagtaatat atatagtaca	8220
gatgttaagt agacttagaa agggctatag gcctgttttc tcttcccccc ccggttacct	8280
ccaacagatc catatccaca aggactggga acagccagcc agagaagaaa cagaagaaga	8340
cggttgaaac aacgttggag acagctcgtg gccttgccg ataagatata tacatttcct	8400
gateccag ctgattcgcc tcttgccgg actatacaac atctcgagga acttactatc	8460
caggatctcc ctgacctcc gaccgtttt ccagagtctt cagagggcac tgacagcaat	8520
cagagactgg ctaagaactg acgcagccta ctgacagtat gggtgcgagt ggatccaagg	8580
agcgttccag gccttcgcaa gggctacgag agagactctt gcgggcacgt ggagagactt	8640
gtggggggca ctgcagcgga tcgggagggg aatacttgca gtcccaagaa gaatcaggca	8700

gggagcagag atcgccctcc tatgaggac agcggatca gcaggagac tttatgaata 8760
 ccccatggag aaccccgca aaagaagggg agaaagaatt gtacaagca caaatagag 8820
 atgatgtaga ttcgatgat gatgacctag taggggtctc tgtcacacca agagtaccac 8880
 taagagaatt gacacataga ttagcaatag atgtgtcaca tttataaaa gaaaaagggg 8940

gactggaagg gatgtattac agtgagagaa gacatagaat cttagacata taccttgaaa 9000
 aggaagaagg gataattgca gattggcaga actatactca tgggccagga ataagatacc 9060
 caatgttctt tgggtggcta tggaagctag taccagtaga tgtcacacca caggaggagg 9120
 acgatgggac tcaactgtta ctacaccag cacaacaag caggtttgat gaccgcgatg 9180
 gggaaacact gatatggaag tttagcccca cgttggtca tgattacaag gcttttatcc 9240
 tgcaccaga ggaatttggg cataagtcag gcctgccaga agaagactgg aaggcaagac 9300
 tgaaagcaag agggatacca tttagttaga gacaggaaca gctatattg gccagggcag 9360

gaaataacta ctgaaaacag ctgagactgc agggactttc cgaaggggct gtaaccaggg 9420
 gagggacatg ggaggagccg gtggggaacg ccctcact tctgtataa agataccgc 9480
 tgcttgcatg gtacttcagt cgctctgcgg agaggctggc agattgagcc ctgggaggtt 9540
 ctctccagca ctacaggtg gagcctgggt gtccctgct agactctcac cgttgcttgg 9600
 ccggcactgg gcagacggct ccacgttgc ttgcttaaaa gacctctaa taaagctgcc 9660
 agttagaagc aagttaagtg tgtgttccca tctctctag tcgcgcctg gtc 9713

<210> 387

<211> 11878

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 387

gcctcactga ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt 60
 gattttaaac ttcatTTTTA atttaaaagg atctaggtga agatcctttt tgataatctc 120
 atgacaaaaa tcccttaacg tgagttttcg ttccactgag cgtcagacc cgtagaaaag 180
 atcaaaggat ctcttgtaga tccttttttt ctgcgcgtaa tctgtgtctt gcaaacaaaa 240
 aaaccaccgc taccagcggg ggtttgtttg ccggatcaag agctaccaac tctttttccg 300
 aaggtaactg gcttcagcag agcgcagata ccaatactg ttcttctagt gtagccgtag 360
 ttaggccacc acttcaagaa ctctgtagca ccgcctacat acctcgctct gctaactctg 420

ttaccagtgg ctgctgccag tggcgataag tcgtgtctta ccgggttgga ctcaagacga	480
tagttaccgg ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc	540
ttggagcgaa cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc	600
acgcttcccc aaggagagaaa ggcggacagg tatccggtaa gcggcagggt cggaacagga	660
gagcgcacga gggagcttcc aggggggaaac gcctggtatc tttatagtcc tgcgggttt	720
cgccacctct gacttgagcg tcgatttttg tgatgctcgt caggggggcg gaggcctatgg	780
aaaaacgcca gcaacgcggc ctttttacgg ttcttggcct tttgctggcc ttttctcac	840
atgttctttc ctgcgttate ccttgattct gtgataacc gtattaccgc ctttgagtga	900
gtgataccg ctgcgcgag ccgaacgacc gagcgacgag agtcagttag cgaggaagcg	960
gaagagcgcc caatacgcaa accgcctctc cccgcgcgtt ggccgattca ttaatgcagc	1020
tggcacgaca ggtttccga ctggaagcg ggcagttagc gcaacgcaat taatgtgagt	1080
tagctcactc attaggcacc ccaggcttta cactttatgc ttccggctcg tatgttgtgt	1140
ggaattgtga gcggataaca atttcacaca ggaaacagct atgaccatga ttacccaag	1200
ctatttaggt gacactatag aatactcaag cttgggggga tcctctagag tcgacctgca	1260
ggcatgctat ttgatgaatt aactacactt aaaataatac aattattatt aaatttttt	1320
ttgatttatt tattaatttt taaacttaat catttgtatt tgggaggaat tatatatatc	1380
tttataatta ttttattttt ttttattttt ttattttttt attattatta ttttttttta	1440
ttttttttt ttactgtatc aaagaaaaac ctttaaaaaa aaaattataa tttcccatc	1500
ttactatatt ttaatacat acgttttaag gaattaaatt agacaaaagc tatattatgc	1560
ttacatata attagaattt ataaacgttt ggttattaga tatttcatgt ctgagtaaag	1620
tctttcaata catatgtaaa aaaatatata tgaatacaca taagtgtta atatatatta	1680
tatgcataaa tgtataaata tatatatata tatatatata tgtatgtatg tatatgtgtg	1740
tatatgaaat tatttcaatg ttttaatttt taaattttta tttttttttt tttttttttt	1800
tttattatgt atattgatct ttattattta aatattactt ttttcgtttt ttcttctttt	1860
tattattttt tttttttttt atattttata caaatggtaa ttcaaataaa aggtataaat	1920
ttatatatta ttttctttta tggataaata aaagaaaaat ataaatatat aaaaatataa	1980
aaatatatat atgtatatgg gggatgatgat aaaatgaaag ataatatata tatatatata	2040
tctttatttt tttttttttg tagaccccat tgtgagtaca taaatatatt atataactcg	2100
ggagcatcag tcatggaatt cttatttctt tttctttttt gcctggccgg cctttttcgt	2160
ggccgccggc cttttgtcgc ctcccagctg agacaggtcg atccgtgtct cgtacaggcc	2220
ggtgatgctc tggatgatca gggatggcgtc cagcacctct ttggtgctgg tgtacctctt	2280

ccggtcgatg gtggtgtcaa agtacttgaa ggccggcaggg gctcccagat tggtcagggt 2340
aaacaggtgg atgatattct cggcctgtct tctgatgggc ttatcccggg gcttggtgta 2400
ggcggacagc actttgtcca gattagcgtc ggccaggatc actctcttgg agaactcgt 2460

gatctgctcg atgatctcgt ccaggtatgt cttgtgctgt tccacaaaca gctgtttctg 2520
ctcattatcc tcgggggagc ccttcagctt ctcatagtgg ctggccaggt acaggaagtt 2580
cacatatattg gagggcaggg ccagttcgtt tcccttctgc agttcccg cagaggccag 2640
cattctcttc cggccgtttt ccagctcgaa caggagtagt ttaggcagct tgatgatcag 2700
gtcctttttc acttctttgt agcccttggc ttccagaaag tcgatgggat tcttctcgaa 2760
gtctcttctt tccatgatgg tgatccccag cagctcttcc acactcttca gtttcttgga 2820
cttgcccttt tccactttgg ccaccaccag cacagaatag gccacggtgg ggctgtcgaa 2880

gccgccgtac ttcttagggg cccagtcctt ctttctggcg atcagcttat cgctgttctt 2940
cttgggcagg atagactctt tgctgaagcc gcctgtctgc acctcggctt ttttcacgat 3000
attcacttgg ggcattgtca gcactttccg cacggtggca aaatcccggc ctttatccca 3060
cacgatctcc ccggtttcgc cgtttgtctc gatcagagge cgcttccgga tctcgccgtt 3120
ggccagggtg atctcggctt tgaaaaagtt catgatgttg ctgtagaaga agtacttggc 3180
ggtagccttg ccgatttctt gctcgtctt ggccgacatc ttccgcacgt cgtacacctt 3240
gtagtgcgg tacacgaact cgctttccag cttaggttac ttttgatca gggcggttcc 3300

cacgacggcg ttcaggtagg cgctgtgggc gtggtggtag ttgttgatct cgcgcacttt 3360
gtaaaactgg aaatccttcc ggaaatcgga caccagcttg gacttcaggg tgatcacttt 3420
cacttcccgg atcagcttgt cattctctgc gtaacttagt ttcatccggg agtccaggat 3480
ctgtgccacg tgcttttgta tctgcccggg ttccaccagc tgtctcttga tgaagccggc 3540
cttatccagt tcgtcagge cgctctctc ggcccttggc agattgtcga actttctctg 3600
ggtaatcagc ttggcgttca gcagctgccg ccagtagttc ttcatcttct tcacgacctc 3660
ttcggagggc acgttgtcgc tcttgccccg gttcttctcg cttctggtca gcacttgtt 3720

gtcgatggag tcgtccttca gaaagctctg aggcacgata tggccacat cgtagtcgga 3780
cagccggttg atgtccagtt cctggtccac gtacatatcc cgccattct gcaggtagta 3840
caggtacagc ttctcgttct gcagctgggt gttttccagc ggggttctt tcaggatctg 3900
gtgccccagc tctttgatgc cctcttcgat ccgcttcatt ctctcgccgc tgttctctg 3960
tcccttctgg gtggtctggt tctctctggc catttcgac acgatgttct cgggcttgtg 4020
ccggcccatc actttcacga gctcgtccac cacttcact gtctgcagga tgcccttctt 4080

aatggcgggg ctgccggcca gattggcaat gtgctcgtgc aggctatcgc cctggccgga	4140
cacctgggct tctcgatgt cctctttaaa ggtcaggctg tcgtcgtgga tcagctgcat	4200
gaagtttctg ttggcgaagc cgtcggactt caggaaatcc aggattgtct tgccggactg	4260
cttgtcccgg atgccgttga tcagcttccg gctcagcctg cccagccgg tgatctccg	4320
ccgttcagc tgcttcatca ctttgcgtc gaacaggtag gcataggttt tcagccgttc	4380
ctcgatcacc tctctgtcct caaacagtgt cagggtcagc acgatatctt ccagaatgtc	4440
ctcgttttcc tcattgtcca ggaagtcctt gtccttgata attttcagca gatcgtgga	4500
tgtgcccagg gaggcgttga accgatcttc cagccggag atttccacgg agtcgaagca	4560
ctcgattttc ttgaagtagt cctctttcag ctgcttcacg gtcactttcc ggttggtctt	4620
gaacagcagg tccacgatgg cttttttctg ctgccgctc aggaaggcgg gctttctcat	4680
tccctcggtc acgtatttca ctttggtag ctggttatac acggtgaagt actcgtacag	4740
caggctgtgc ttgggcagca ctttctgtt gggcaggttc ttatcgaagt tggatcatccg	4800
ctcgatgaag ctctggcgg aagcgcctt gtccaccact tcctcgaagt tccagggggt	4860
gatggtttcc tcgtcttttc tggatcacc ggcgaatctg ctgtttcccc tggccagagg	4920
gcccacgtag taggggatgc ggaaggtag gatcttctcg atcttttccc ggttgctctt	4980
caggaatggg taaaaatctt cctgccgccg cagaatggcg tgcagctctc ccaggtagat	5040
ctggtggggg atgctgccgt tgtcgaaggt ccgtgcttc cgcagcaggt cctctctgtt	5100
cagcttcacg agcagttcct cgggtccgtc catcttttcc aggatgggct tgatgaactt	5160
gtagaactct tcttggtcgg ctccgccgtc aatgtagccg gcgtagccgt tcttgcctg	5220
gtcgaagaaa atctctttgt acttctcagg cagctgctgc cgcacgagag ctttcagcag	5280
ggtcaggtec tggtagtct ctgctatct cttgatcata gaggcgtca ggggggcctt	5340
ggtgatctcg gtgttactc tcaggatgtc gtcagcagg atggcgtcgg acaggttctt	5400
ggcggccaga aacaggtagc cgtactggtc gccgatctgg gccagcaggt tgtccaggtc	5460
gtcgtcgtag gtgtccttc tcagctgcag ttggcatcc tcggccagg cgaagtgtct	5520
cttgaagtgg ggggtcaggc ccaggctcag ggcaatcagg ttccgaaca ggccattctt	5580
cttctgcgg ggcagctggg cgatcagatt ttccagccgt ctgctcttgc tcagtctggc	5640
agacaggatg gccttggcgt ccacgccgtt ggcgttgatg gggttttcct cgaacagctg	5700
gttgtaggtc tgcaccagct ggatgaacag cttgtccacg tcgctgttgt cggggttcag	5760
gtcgcctcgt atcaggaagt ggccccgaa cttgatcatg tgggccaggg ccagatagat	5820

cagccgcagg tcggccttgt cgggtgctgtc caccagtttc tttctcaggt ggtagatggt 5880

gggggtacttc tcgtggttagg ccacctcgtc cacgatgttg ccgaagatgg ggtgccgctc 5940

gtgctttetta tctctttcca ccaggaagga ctcttccagt ctgtggaaga agctgtcgtc 6000

caccttggcc atctcgttgc tgaagatctc ttgcagatag cagatccggt tcttccgtct 6060

ggtgtatctt ctcttgccgg ttctcttcag ccgggtggcc tcggctgttt cgcccgctgc 6120

gaacagcagg gctccgatca ggttcttctt gatgctgtgc cggtcggtgt tgcccagcac 6180

cttgaatttc ttgctgggca ccttgtactc gtcggtgatc acggcccagc ccacagagtt 6240

ggtgccgatg tccaggccga tgcgtactt ctgtcggt gctgggactc cgtggatacc 6300

gaccttccgc ttcttcttgg gggccatctt atcgatcatg tctttgtaac caatatcatg 6360

atcctttagt tctccgtcgt ggtccttata gtccattttt ctgagggat cctgatatat 6420

ttctattagg tatttattat tataaaatat aaatcttgaa tgataataaa taaaatatta 6480

gttattcctt ttctagttaa aaatatacat attataaata tatatatata tatatatatt 6540

tttattgtga caagaatata taattataaa ttatattatt tatitttgta tttttttttt 6600

tttttttttt ttttctttt ttgtttttat ttttctttt tttataaat attatttttt 6660

tcttttatca tgcacattgg aataatacat taatatatat atatataatta tattatacat 6720

atattgaata atgtttataa aaaatgcata acttatatga atataatttt ttttaaatat 6780

gacaaaaaga aaaaaaaaaa aaaccaaaaa aaattaaaat tgaaatgaaa tatataaata 6840

tattatttat atatattata cattgtttta tactactaca tgtatatata tatattatat 6900

atatatatat atatcaattt tttcaaaaat aaattaatat aaaaagagg gaaaaaaaaa 6960

aaaaaaaaaa aaaaagata attaagtaag catttaaaaa tatataaatt gataatatat 7020

aaaattaatc acatataaaa gcttataaac actaggttag ctaattcgct tgtaagaggt 7080

actctcgttt atgcaaaact atttgatata gcattttaac aagtacacat atatatatgt 7140

aatatatata ctatatatat ctattgcatg tgtactaagc atgtgcatgg catccccttt 7200

ttctcgtgtt taaaacagtt tgtatgataa aatataaagg atttgaaaaa gagaaaaaaaa 7260

tatatgatct catcctatat agcgccataa tttttatttg ggttgaataa aattttctac 7320

taaatttagg tgtaagtaaa ataatggaat atatataagt acaataaaaa agtgcataaa 7380

ttaaaaaatt tttataataa atattttttt taaaaaagtc aataataata ttaaatatat 7440

ataacacagg attatatatg ttactacaaa ttttttatat tataatataa attcttttca 7500

attttcattt tattttacat acactttcct tttttgtcac tatattttta tattcacata 7560

tttagtttaa atactggeta tttctttcta catttgctag taacaattgt gtagtgctta 7620

aatatatata cacacctaaa acttacaaag taccctagga ccatggccaa gcctttgtct 7680

caagaagaat ccacccatcat tgaagagca acggctacaa tcaacagcat ccccatctct	7740
gaagactaca gcgtcgccag cgcagctctc tctagcgacg gccgcatctt cactggtgtc	7800
aatgtatata attttactgg gggaccttgt gcagaactcg tgggtctggg cactgctgct	7860
gctgcggcag ctggcaacct gacttgtatc gtcgcgatcg gaaatgagaa caggggcatc	7920
ttgagccct gcggacggtg ccgacaggtg cttctcgatc tgcacctctg gatcaaagcc	7980
atagtgaagg acagtgatgg acagccgacg gcagttggga ttcgtgaatt gctgccctct	8040
ggttatgtgt gggagggcta accgcgggta cccattaaa tttatttaat aatagattaa	8100
aaatattata aaaaataaaa cataaacaca gaaattacaa aaaaaataca tatgaatatt	8160
ttttttgtaa tcttccttat aaatatagaa taatgaatca tataaaacat atcattattc	8220
atttatttac atttaaaatt attgtttcag tatctttaat ttattatgta tatataaaaa	8280
taacttacaa ttttattaat aaacaatata tgtttattaa ttcattgttt gtaatttatg	8340
ggatagcgat ttttttact gtctgtattt tcttttttaa ttatgtttta attgtattta	8400
ttttattttt attattgttc tttttatagt attattttta aacaaaatgt attttctaag	8460
aacttataat aataataata taaattttta taaaaattat atttatcttt tacaatatga	8520
acataaagta caacattaat atatagcttt taatatTTTT attcctaac atgtaaatct	8580
taaatttttc tttttaaaca tatgttaaat atttatttct cattatatat aagaacatat	8640
ttattacatc tagaggtagc gagctcgttt tcgacactgg atggcggcgt tagtatcgaa	8700
tcgacagcag tatagcgacc agcattcaca tacgattgac gcattgatatt actttctgcg	8760
cacttaactt cgcatctggg cagatgatgt cgaggcgaaa aaaaataata atcacgctaa	8820
catttgatta aaatagaaca actacaatat aaaaaaacta tacaatgac aagtcttga	8880
aaacaagaat ctttttattg tcagtactga ttagaaaaac tcatcgagca tcaaatgaaa	8940
ctgcaattta ttcatatcag gattatcaat accatatatt tgaaaaagcc gtttctgtaa	9000
tgaaggagaa aactcacga ggcagttcca taggatggca agatcctggt atcggctctgc	9060
gattccgact cgtccaacat caatacaacc tattaatttc cctcgtcaa aaataagggt	9120
atcaagttag aaatcacat gagtgacgac tgaatccggt gagaatggca aaagcttatg	9180
catttctttc cagacttggt caacaggcca gccattacgc tcgtcatcaa aatcactcgc	9240
atcaacaaaa ccgttatcca ttcgtgattg cgctgagcg agacgaaata cgcgatcgct	9300
gttaaaagga caattacaaa caggaatcga atgcaaccgg cgcaggaaca ctgccagcgc	9360
atcaacaata tttcacctg aatcaggata ttcttctaata acctggaatg ctgttttgcc	9420
ggggatcgca gtggtgagta accatgcac atcaggagta cggataaaat gcttgatggt	9480
cggaagaggc ataaattccg tcagccagtt tagtctgacc atctcatctg taacatcatt	9540

ggcaacgcta cctttgccat gtttcagaaa caactctggc gcatcgggct tcccatacaa	9600
tcgatagatt gtcgcacctg attgcccgcac attatcgcga gcccatTTat acccatataa	9660
atcagcatcc atgttggaaT ttaatcgcgT cctcgaaacg tgagtctttt ccttaccat	9720
ggttgtttat gttcggatgt gatgtgagaa ctgtatccta gcaagatttt aaaaggaagt	9780
atatgaaaga agaacctcag tggcaaatcc taacctttta ttttctcta caggggcgcg	9840
gcgtggggac aattcaacgc gtctgtgagg ggagcgtttc cctgctcgca ggtctgcagc	9900
gaggagccgt aatttttgcT tcgcgccgtg cggccatcaa aatglatgga tgcaaatgat	9960
tatacatggg gatgtatggg ctaaattgtac gggcgacagt cacatcatgc ccctgagctg	10020
cgcacgtcaa gactgtcaag gaggtatttc tgggcctcca tgtcgtggc ctaacattag	10080
taatgtaggt ctgactttca ctcatataag tcttatggta actaaactaa ggtcttacct	10140
ttactgatat atgtcttact ttactaaact taggtattac ttttactaac ttaggtctta	10200
aattcagtaa ctaaggTcat acttcgacta actaaggTct tacattcact gatataggTc	10260
ttatgattac taacttaggt cctaatttga ctaacataag tcctaacatt agtaatgtag	10320
gtcttaactt aactaaacta ggtcttacct tcactaatat aggtcttaat attactgact	10380
taagtaatta aggtactaac ttaggtcgta aggtactaa tataataggTc ttaaggtaac	10440
taatttaggt cttgacttaa taaatatagg tcctaacata aatagtatag gtcctaatat	10500
aagtactata ggccTTaact taaccaacat aggtcctaac ataagttata taggtcttaa	10560
cgtaactaac ataagTcatt aaggTactaa gtttggTctt aatttaacaa taacatgtcg	10620
ctggcctaac attagtaatg taggtctgac tttcactcat ataagTctta tggtaactaa	10680
actaaggTct tacctttact gatatatgtc ttactttcac taacttaggt attactttta	10740
ctaacttagg tcttaaatTc agtaactaag gtcatacttc gactaactaa ggtcttacat	10800
tcactgatat aggtcttatg attactaact taggtcctaa tttgactaac ataagTccta	10860
acattagtaa ttaggtcttT aacttaacta acttaggtct taccttact aatataggTc	10920
ttaatattac tgacttaagt aattaaggta ctaacttagg tcgtaaggta actaatatat	10980
aggtcttaag gtaactaatt taggtcttga cttaataaat ataggTccta acataaatag	11040
tataggTcct aatataagta ctataggcct taacttaacc aacataggTc ctaacataag	11100
ttatataggt cttaacgtaa ctaacataag tcattaaggT actaagtttg gtcttaattt	11160
aacaataacc atgtcgctgg ccgggtggTc ttaatttaac aaatatagac catgtcgctg	11220
gccgggtgac ccggcgggga cgaggcaagc taaacagatc ctcgtgatac gcctattttt	11280

ataggttaat gtcatgataa taatggtttc ttaggacgga tcgcttgctt gtaacttaca 11340
 cgcgctcgt atcttttaat gatggaataa tttgggaatt tactctgtgt ttattttatt 11400
 ttatgttttg tatttggatt ttagaaagta aataaagaag gtagaagagt tacggaatga 11460
 agaaaaaaaa ataaacaaag gtttaaaaaa tttcaacaaa aagcgtactt tacatatata 11520
 tttattagac aagaaaagca gattaaatag atatacattc gattaacgat aagtaaaatg 11580
 taaaatcaca ggattttcgt gtgtgggtctt ctacacagac aagatgaaac aattcggcat 11640
 taatactga gagcaggaag agcaagataa aaggtagtat ttgttggcga tccccctaga 11700

gtcttttaca tcttcggaag acaaaaacta ttttttcttt aatttctttt tttactttct 11760
 atttttaatt tatatattta tattaaaaaa tttaaattat aattattttt atagcacgtg 11820
 atgaaaagga cccagggtggc acttttcggg gaaatctcga cctgcagcgt acgaagct 11878

<210> 388

<211> 12044

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 388

gcctcactga ttaagcattg gtaactgtca gaccaagttt actcataat acttttagatt 60
 gattttaaag ttcattttta attttaaagg atctaggtga agatcctttt tgataatctc 120

atgaccaaaa tcccttaacg tgagttttcg ttcactgag cgtcagaccc cgtagaaaag 180
 atcaaaggat ctctttgaga tctttttttt ctgcgcgtaa tctgctgctt gcaaacaaaa 240
 aaaccaccgc taccagcggg ggtttgtttg ccggatcaag agctaccaac tctttttccg 300
 aaggttaactg gcttcagcag agcgcagata ccaataactg tttcttctagt gtagccgtag 360
 ttaggccacc acttcaagaa ctctgtagca ccgcctacat acctcgtctt gctaactctg 420
 ttaccagtgg ctgctgccag tggcgataag tcgtgtctta ccgggttgga ctcaagacga 480
 tagttaccgg ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc 540

ttggagcgaa cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc 600
 acgcttcccg aaggagagaa ggcggacagg tatccggtaa gcggcagggt cggaacagga 660
 gagcgcacga gggagcttcc aggggggaaac gcctggatc tttatagtcc tgtcgggttt 720
 cgccacctct gacttgagcg tcgatttttg tgatgctcgt caggggggcg gacgctatgg 780
 aaaaacgcca gcaacgcggc ctttttacgg ttcctggcct tttgctggcc ttttctcac 840
 atgttctttc ctgcgttacc ccctgattct gtggataacc gtattaccgc ctttgagtga 900

gctgataccg ctcgccgcag ccgaacgacc gagcgcagcg agtcagtgag cgaggaagcg 960

gaagagcgcc caatacgcga accgcctctc cccgcgcgtt ggccgattca ttaatgcagc 1020

tggcacgaca ggtttccgga ctggaaagcg ggccagtgagc gcaacgcaat taatgtgagt 1080

tagctcactc attaggcacc ccaggcttta cactttatgc ttccggctcg tatgttgtgt 1140

ggaattgtga gcggataaca atttcacaca ggaaacagct atgaccatga ttacgccaag 1200

ctatttaggt gacactatag aatactcaag ctggggggga tcctctagag tcgactaata 1260

cgactcacta taggaacata atctatagcg gcgtttttaga gctagaaata gcaagttaaa 1320

ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgcta gcataacccc 1380

ttggggcctc taaacgggtc ttgaggggtt ttttggtcga cctgcaggca tgctatttga 1440

tgaattaact acacttaaaa taatacaatt attattaaat ttttttttga tttatttatt 1500

aattttttaa cttaatcatt tgtatttggg aggaattata tatactttta taattatitt 1560

atTTTTTTT atTTTTTat tTTTTatta ttattatttt tttttatttt ttttttttac 1620

tgtatcaaag aaaaaccttt aaaaaaaaaa ttataatttc cccatcttac tatattttta 1680

atacatagct ttttaaggaat taaattagac aaaagctata ttatgcttta catataatta 1740

gaatttataa acgttttggt attagatatt tcatgtctca gtaaagtcct tcaatacata 1800

tgtaaaaaaa tatatatgaa tacacataag ttgttaatat attttatatg cataaatgta 1860

taaatatata tatatatata tatatatgia tgtatgtata tgtgtgtata tgaaattatt 1920

tcaatgttta atttttttaa ttttaatttt tttttttttt ttttttttta ttatgtatat 1980

tgatctttat tatttaaaata ttactttttt cgttttttct tctttttatt attttttttt 2040

ttttttatat tttatacaaa tggtaattca aataaaaggt ataaatttat atttaatttt 2100

cttttatgga taaataaaa aaaaatataa atataaaaa atataaaaat atatatatgt 2160

atattggggt gatgataaaa tgaaagataa tatatatata tatatatctt tatttttttt 2220

ttttttaga cccatttgtg agtacataaa tatattatat aactcgggag catcagtcac 2280

ggaattctta tttctttttc ttttttgctt ggccggcctt tttcgtggcc gccggccttt 2340

tgtcgctcc cagctgagac aggtcgatcc gtgtctcgta caggccggtg atgctctggt 2400

ggatcagggt ggcgtccagc acctctttgg tgctggtgta cctcttccgg tcgatgggtg 2460

tgtcaaagta ctgaaggcg gcaggggctc ccagattggt cagggtaaac aggtggatga 2520

tattctcggc ctgctctctg atgggcttat cccggtgctt gttgtaggcg gacagcactt 2580

gtccagatt agcgtcgcc aggatcactc tcttgagaa ctcgctgac tgctcgatga 2640

tctcgtccag gtagtgcttg tgctgttcca caaacagctg tttctgtca ttatcctcgg	2700
gggagccctt cagcttctca tagtggctgg ccaggtacag gaagttcaca tatttggagg	2760
gcagggccag ttctgttccc ttctgcagtt cgcgggcaga ggccagcatt ctcttccggc	2820
cgttttccag ctccaacagg gagtacttag gcagcttgat gatcaggtcc tttttcactt	2880
cttttagacc cttggcttcc agaaagtcga tgggattctt ctccaagctg cttctttcca	2940
tgatggtgat cccagcagc tctttcacac tcttcagttt cttggacttg cctttttcca	3000
ctttggccac caccagcaca gaataggcca cggtaggggt gtcgaagccg ccgtacttct	3060
tagggtecca gtccttcttt ctggcgatca gcttatcgtt gttcctcttg ggcaggatag	3120
actctttgct gaagccgctt gtctgcacct cggctttttt cacgatattc acttggggca	3180
tgctcagcac ttccgcacg gtggcaaaat cccggccctt atccacacg atctccccgg	3240
tttcgccgtt tgtctcgatc agaggccgct tccggatctc gccgttggcc agggtaatct	3300
cggctctgaa aaagtctatg atgttgctgt agaagaagta cttggcggta gccttgccga	3360
tttctgtctc gctcttggcg atcatcttcc gcacgtcgta cacctttagt tcgccgtaca	3420
cgaactcgtt ttccagctta gggtagcttt tgatcagggc ggttcccacg acggcgttca	3480
ggtaggcgtc gtgggcgttg tggtagttgt tgatctcgcg cactttgtaa aactggaaat	3540
ccttccggaa atcggacacc agcttggact tcagggtgat cactttcact tcccggatca	3600
gcttgcatt ctctcgatc ttagtgttca tccgggagtc caggatctgt gccacgtgct	3660
ttgtgatctg ccgggtttcc accagctgtc tcttgatgaa gccggcctta tccagttcgc	3720
tcaggccgcc tctctcgcc ttggtcagat tgcgaactt tctctgggta atcagcttgg	3780
cgttcagcag ctgccgccag tagttcttca tcttcttcac gacctcttcg gagggcacgt	3840
tgtcgtctt gccccggttc ttgtcgcttc tggtcagcac cttgtgtcg atggagtcgt	3900
ccttcagaaa gctctgaggc acgatatggt ccacatcgta gtcggacagc cggttgatgt	3960
ccagttcctg gtccacgtac atatccgcc cattctgcag gtagtacagg tacagcttct	4020
cgttctgcag ctgggtgttt tccacgggtt gttctttcag gatctggctg cccagctctt	4080
tgatgccctc ttgatccgc ttcatctct cgcggctgtt cttctgtccc ttctgggtgg	4140
tctggttctc tctggccatt tcgatcacga tgttctcggg cttgtgccgg cccatcactt	4200
tcacgagctc gtccaccacc ttactgtct gcaggatgcc cttcttaatg gcggggctgc	4260
cggccagatt ggcaatgtgc tcgtgcaggc tatcgccctg gccggacacc tgggctttct	4320
ggatgtctc tttaaaggtc aggtgtctgt cgtggatcag ctgcatgaag tttctgttgg	4380
cgaagccgtc ggacttcagg aaatccagga ttgtcttgcc ggactgcttg tcccggatgc	4440
cgttgatcag ctccggctc agcctgcccc agccggtgta tctccgccgc ttcagctgct	4500

tcatacattt gtcgtcgaac aggtgggcat aggttttcag ccgttcctcg atcatctctc	4560
tgtcctcaaa cagtgtcagg gtcagcacga tatcttcag aatgtcctcg ttttcctcat	4620
tgtccaggaa gtccttgcc ttgataattt tcagcagatc gtggtatgtg cccagggagg	4680
cgttgaaccg atcttcacg ccggagattt ccacggagtc gaagcactcg atttcttga	4740
agtagtctc tttcagctgc ttcacggta ctttcgggtt ggtcttgaac agcaggtcca	4800
cgatggcctt tttctgctcg ccgtcagga aggcgggctt tctcattccc tcggtcacgt	4860
atttcacttt ggtcagctcg ttatacacgg tgaagtactc gtacagcagg ctgtgcttgg	4920
gcagcacctt ctggttgggc aggttcttat cgaagtgggt catccgctcg atgaagctct	4980
gggcggaagc gcccttgcc accacttctt cgaagtcca ggggtgatg gtttcctcgc	5040
tctttctggt catccaggcg aatctgctgt tttccctggc cagagggccc acgtagtagg	5100
ggatgcggaa ggtcaggatc ttctcgatct tttcccggtt gtccttcagg aatgggtaaa	5160
aatcttctg ccgccgcaga atggcgtgca gctctccag gtggatctgg tgggggatgc	5220
tgccgttgtc gaaggtccgc tgcttcgca gcaggtctc tctgttcagc ttcacgagca	5280
gttctcgggt gccgtccatc ttttcagga tgggttgat gaactttag aactcttct	5340
ggctggctcc gccgtcaatg tagccggcgt agccgttctt gctctggtcg aagaaaatct	5400
ctttgtactt ctcaggcagc tgctccgca cgagagcttt cagcagggtc aggtcctggt	5460
ggtgctcgtc gtatctcttg atcatagagg cgtcagggg gcccttggtg atctcggtgt	5520
tcactctcag gatgtcgtc agcaggatgg cgtcggacag gttcttggcg gccagaaaca	5580
ggtcggcgta ctggtcgcc atctgggcca gcaggttgtc caggtcgtcg tcgtaggtgt	5640
ccttgctcag ctgcagtttg gcatcctcgg ccaggtcgaa gttgctcttg aagttggggg	5700
tcaggccag gtcaggcca atcaggtttc cgaacaggcc attcttcttc tcgccgggca	5760
gctggcgcat cagattttcc agccgtctgc tcttgctcag tctggcagac aggatggcct	5820
tggcgtccac gccgtggcg ttgatgggtt tttcctcgaa cagctggttg taggtctgca	5880
ccagctggat gaacagcttg tccacgtcgc tgtgtcggg gttcaggtcg ccctcgatca	5940
ggaagtggcc ccggaacttg atcatgtggg ccagggccag atagatcagc cgcaggtcgg	6000
ccttgctggt gctgtccacc agtttctttc tcaggtggta gatggtgggg tacttctcgt	6060
ggtaggccac ctggtccagc atgttgccga agatggggtg ccgctcgtgc ttcttactct	6120
cttcaccag gaaggactct tccagtctgt ggaagaagct gtcgtccacc ttggccatct	6180
cgttgctgaa gatctcttgc agatagcaga tccggttctt ccgtctggtg tatcttcttc	6240
tggcggttct cttcagccgg tggcctcgg ctgtttcgcc gctgtcgaac agcagggctc	6300
cgatcagggt cttcttgatg ctgtgccggg cgggtgtgcc cagcaccttg aatttcttgc	6360

tgggcacctt gtactcgtcg gtgacacagg cccagccac agagtgggtg ccgatgtcca	6420
ggccgatgct gtacttcttg tcggctgctg ggactccgtg gataccgacc ttccgcttct	6480
tctttggggc catcttatcg tcatcgtctt tgtaatcaat atcatgatcc ttgtagtctc	6540
cgtcgtggtc cttatagtcg atttttctcg agggatcctg atatatctt attaggtatt	6600
tattattata aaatataaat cttgaatgat aataaataaa atattagtta ttccttttct	6660
agtttaaaat atacatatta taaatatata tatatatata tatattttta ttgtgacaag	6720
aatatataat tataaattat attatttatt ttgtattttt tttttttttt tttttttttt	6780
tctttttttg ttttattttt cttttttttt ataaatatta tttttttctt ttatcatgca	6840
cattggaata atacattaat atatatatat atattatatt atacatatat tgaataatgt	6900
ttataaaaaa tgcataactt atatgaatat aatttttttt aaatatgaca aaaagaaaaa	6960
aaaaaaaaac caaaaaaaat taaaattgaa atgaaatata taaatatatt atttatatat	7020
attatacatt gttaataact actacatgta tatatatata ttatatatat atatatatat	7080
caattttttc aaaaataaat taatatataa agaggggaaa aaaaaaaaaa aaaaaaaaaa	7140
aagataatta agtaagcatt taaaaatata taaattgata atatatataa ttaatcacat	7200
ataaaagctt ataaacacta ggtagctaa ttcgcttgta agaggctactc tcgtttatgc	7260
aaaactatth gatatagcat ttaacaagt acacatatat atatgtaata tatatactat	7320
atatacttat tgcattgtga ctaagcatgt gcatggcatc ccttttttct cgtgtttaaa	7380
acagtttgta tgataaaata taaaggattt gaaaaagaga aaaaaatata tgatctcatc	7440
ctatatagcg ccataattht tatttgggtt gaataaaatt ttctactaaa tttaggtgta	7500
agtaaaataa tggaatatat ataagtacaa taaaaaagtg cataaattaa aaaattttta	7560
taataaatat tttttttaa aaagtaata ataatttaa atatatataa cacaggatta	7620
tatatgttca ctacaattht ttatatata atataaatc ttttcaatth tcattttatt	7680
ttacatacac tttccttttt tgtcactata ttttaattat cacatattta gtttaatac	7740
tggtattht tttctacatt tgctagtaac aattgtgtag tgcttaata tatacacaca	7800
cctaaaactt acaaagtatc ctaggacat ggccaagcct ttgtctcaag aagaatccac	7860
cctcattgaa agagcaacgg ctacaatcaa cagcatcccc atctctgaag actacagcgt	7920
cgccagcgca gctctctcta gcgacggcgg catcttcact ggtgtcaatg tatatcattt	7980
tactggggga ccttgtgcag aactcgtggt gctgggcact gctgctgctg cggcagctgg	8040
caacctgact tgtatcgtcg cgatcggaat tgagaacagg ggcatcttga gccctgcgg	8100

acggtgccga cagggtgcttc tcgatctgca tcctgggatc aaagccatag tgaaggacag 8160
 tgatggacag ccgacggcag ttgggattcg tgaattgctg ccctctggtt atgtgtggga 8220
 gggtcaaccg cgggtacccc attaaattta ttttaataata gattaaaaat attataaaaa 8280
 taaaaacata aacacagaaa ttacaaaaaa aatacatatg aatTTTTTTT ttgtaatctt 8340
 ccttataaat atagaataat gaatcatata aaacatatca ttattcattt atttacattt 8400
 aaaattattg tttcagtatc ttttaattat tatgtatata taaaaataac ttacaatttt 8460
 attaataaac aatatagttt tattaattca tgttttgtaa tttatgggat agcgattttt 8520

 tttactgtct gtattttctt ttttaattat gttttaattg tatttatTTT atttttatta 8580
 ttgttctttt tatagtatta ttttaaaaca aaatgtattt tctaagaact tataataata 8640
 ataataaaa ttttaataaa aattatattt atcttttaca atatgaacat aaagtacaac 8700
 attaatatat agcttttaat atttttattc ctaatcatgt aaatcttaaa tttttctttt 8760
 taaacatatg ttaaataattt atttctcatt atatataaga acatatttat tacatctaga 8820
 ggtaccgagc tcgttttcga cactggatgg cggcgttagt atcgaatcga cagcagtata 8880
 gcgaccagca ttcacatacg attgacgcat gatattactt tctgcgcact taacttcgca 8940

 tctgggcaga tgatgtcgag gcgaaaaaaa atataaatca cgctaacatt tgattaaaat 9000
 agaacaacta caatataaaa aaactataca aatgacaagt tcttgaaaac aagaatcttt 9060
 ttattgtcag tactgattag aaaaactcat cgagcatcaa atgaaactgc aatttatcca 9120
 tatcaggatt atcaatacca tatttttgaa aaagccgttt ctgtaatgaa ggagaaaact 9180
 caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt ccgactcgtc 9240
 caacatcaat acaacctatt aatttcccct cgtcaaaaat aaggttatca agtgagaaat 9300
 caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt tctttccaga 9360

 ctgtttcaac aggccagcca ttacgtcgt catcaaaatc actcgcatca accaaaccgt 9420
 tattcattcg tgattgcgcc tgagcgagac gaaatacgcg atcgtgttta aaaggacaat 9480
 tacaacagg aatcgaatgc aaccggcgca ggaacactgc cagcgcatca acaatatTTT 9540
 cacctgaatc aggatattct tctaatacct ggaatgctgt tttgccgggg atcgcagtgg 9600
 tgagtaacca tgcatcatca ggagtacgga taaaatgctt gatggtcgga agaggcataa 9660
 attccgtcag ccagtttagt ctgaccatct catctgtaac atcattggca acgtacctt 9720
 tgccatgttt cagaaacaac tctggcgcat cgggcttccc atacaatcga tagattgtcg 9780

 cacctgattg cccgacatta tcgcgagccc atttatacc atataaatca gcatccatgt 9840
 tggaaattta tcgcggcctc gaaacgtgag tcttttctt acccatggtt gtttatgttc 9900
 ggatgtgatg tgagaactgt atcctagcaa gatTTTaaa ggaagtatat gaaagaagaa 9960

cctcagtggc aaatcctaac cttttatatt tctctacagg ggcgcggcgt ggggacaatt	10020
caacgcgtct gtgaggggag cgtttccctg ctgcaggtc tgcagcgagg agccgtaatt	10080
tttgcttcgc gccgtgcggc catcaaatg tatggatgca aatgattata catggggatg	10140
tatgggctaa atgtacgggc gacagtcaca tcatgccctt gagctgcgca cgtcaagact	10200
gtcaaggagg gtattctggg cctccatgtc gctggcctaa cattagtaat gtaggtctga	10260
ctttcactca tataagtctt atggtaacta aactaaggtc ttacctttac tgatatatgt	10320
cttactttca ctaacttagg tattactttt actaacttag gtcttaaatt cagtaactaa	10380
ggtcatactt cgactaacta aggtcttaca ttcactgata taggtcttat gattactaac	10440
ttaggtccta atttgactaa cataagtcct aacattagta atgtaggtct taacttaact	10500
aacttaggtc ttaccttcac taatataggi cttaatatata ctgacttaag taattaaggt	10560
actaacttag gtcgtaaggt aactaatata taggtcttaa ggtaactaat ttaggtcttg	10620
acttaataaa tataggctct aacataaata gtataggctc taatataagt actataggcc	10680
ttaacttaac caacataggt cctaacataa gttatatagg tcttaacgta actaacataa	10740
gtcattaagg tactaagttt ggtcttaatt taacaataac atgtcgtctgg ctaacatta	10800
glaatgtagg tctgactttc actcatataa gtcttatggg aactaaacta aggtcttacc	10860
tttactgata tatgtcttac tttcactaac ttaggtatta cttttactaa cttaggtctt	10920
aaattcagta actaaggta tacttcgact aactaaggtc ttacattcac tgatataggt	10980
cttatgatta ctaacttagg tcctaatttg actaacataa gtcctaacat tagtaatgta	11040
ggtcttaact taactaactt aggtcttacc ttcactaata taggtcttaa tattactgac	11100
ttaagtaatt aaggtaacta cttaggtcgt aaggtaacta atatataggt cttaggtaa	11160
ctaatttagg tcttgactta ataaatatag gtccctaact aaatagtata ggtcctaata	11220
taagtactat aggccttaac ttaaccaaca taggtcctaa cataagttat ataggcttta	11280
acgtaactaa cataagtcac taaggtaacta agtttggtct taatttaaca ataaccatgt	11340
cgctggccgg gtggtcttaa ttaacaaat atagaccatg tcgctggccg ggtgacccgg	11400
cggggacgag gcaagctaaa cagatcctcg tgatacgctt atttttatag gttaatgtca	11460
tgataataat ggtttcttag gacggatcgc ttgcctgtaa cttacacgcg cctcgtatct	11520
tttaatgatg gaataatttg ggaatttact ctgtgtttat ttatttttat gttttgtatt	11580
tggatttttag aaagtaata aagaaggtag aagagttacg gaatgaagaa aaaaaataa	11640
acaaaggttt aaaaaatttc aacaaaaagc gtactttaca tataatttta ttagacaaga	11700
aaagcagatt aaatagatat acattcgatt aacgataagt aaaatgtaaa atcacaggat	11760
tttcgtgtgt ggtcttctac acagacaaga tgaacaatt cggcattaat acctgagagc	11820

aggaagagca agataaaagg tagtatttgt tggcgatccc cctagagtct ttacatctt 11880

cggaaaacaa aaactatttt ttctttaatt tcttttttta ctttctattt ttaatttata 11940

tatttatatt aaaaaattta aattataatt atttttatag cacgtgatga aaaggacceca 12000

ggtggcactt ttcggggaaa tctcgacctg cagcgtagca agct 12044

<210> 389

<211> 12044

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 389

gcctcactga ttaagcattg gtaactgtca gaccaagttt actcatatat actttagatt 60

gatttaaaac ttcatTTTTA atttaaaagg atctaggtga agatcctttt tgataatctc 120

atgaccaaaa tcccttaacg tgagttttcg ttccactgag cgtcagaccc cgtagaaaag 180

atcaaaggat ctctttgaga tctttttttt ctgcgcgtaa tctgctgctt gcaaacaaaa 240

aaaccaccgc taccagcggg ggtttgtttg ccggatcaag agctaccaac tctttttccg 300

aaggtaactg gcttcagcag agcgcagata ccaataactg ttcttctagt gtagccgtag 360

ttaggccacc acttcaagaa ctctgtagca ccgcctacat acctcgctct gctaactctg 420

ttaccagtgg ctgctgccag tggcgataag tctgtcttta ccgggttgga ctcaagacga 480

tagttaccgg ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc 540

ttggagcgaa cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc 600

acgcttcccg aaggagagaaa ggccgacagg tatccggtaa gcggcagggt cggaacagga 660

gagcgcacga gggagcttcc agggggaaac gcctggtatc tttatagtcc tgtcgggttt 720

cgccacctct gacttgagcg tcatTTTTG tgatgctcgt caggggggcg gacccatgg 780

aaaaacgcca gcaacgcggc ctttttacgg ttccctggcct tttgctggcc ttttctcac 840

atgttctttc ctgcgttacc ccctgattct gtggataacc gtattaccgc ctttgagtga 900

gctgataccg ctgcgccgag ccgaacgacc gagcgacgag agtcagttag cgaggaagcg 960

gaagagcgcc caatacgcaa accgcctctc cccgcgcgtt ggccgattca ttaatgcagc 1020

tggcacgaca ggtttccga ctggaaagcg ggacgtgagc gcaacgcaat taatgtgagt 1080

tagctcactc attaggcacc ccaggcttta cactttatgc ttccggctcg tatgttgtgt 1140

ggaattgtga gcggataaca atttcacaca ggaaacagct atgacatga ttacccaag 1200

ctatttaggt gacactatag aatactcaag cttgggggga tcctctagag tcgactaata	1260
cgactcacta taggaaatga tatggatttt gggttttaga gctagaaata gcaagttaaa	1320
ataaggctag tccgttatca acttgaaaaa gtggcaccga gtcggtgcta gcataacccc	1380
ttggggcctc taaacgggtc ttgaggggtt ttttggtcga cctgcaggca tgctatttga	1440
tgaattaact acacttaaaa taatacaatt attattaaat ttttttttga tttatttatt	1500
aatttttaaa cttaatcatt tgtatttggg aggaattata tatacttta taattatfff	1560
atfttttttt atftttttat tfttttatta ttattatfff tftttatfff tfttttttac	1620
tgtatcaaag aaaaacftt aaaaaaaaaa ttataatttc cccatcttac tatatfttta	1680
atacatacgt tftaaggaat taaattagac aaaagctata ttatgcttta catataatta	1740
gaattftataa acgtttgggt attagataft tcatgtctca gtaaagctft tcaatacata	1800
tgtaaaaaaa tatatatgaa tacacataag ttgttaatat atfttatatg cataaatgta	1860
taaatatata tatatatata tatatatgta tgtatgtata tgttgtata tgaaattatt	1920
tcaatgttta atfttttaaa tfttaattft tftttttttt tfttttttta ttatgtatat	1980
tgatctftat tftttaataa ttactftttt cgtfttttct tctftttatt atfttttttt	2040
tfttttatat tftatacaaa tggtaattca aataaaaagg ataaatttat atftaattft	2100
ctfttatgga taaataaaag aaaaatataa atataaaaa atataaaaat atatatatgt	2160
atattggggt gatgataaaa tgaaagataa tatatatata tatatatctt tftttttttt	2220
tftttttaga cccatttggt agtacataaa tatattatat aactcgggag catcagtcatt	2280
ggaattctta tftctftttc tfttttgcct ggccggcctt tftcgtggcc gccggcctft	2340
tgtcgcctcc cagctgagac aggtcgatcc gtgtctcgta caggccggtg atgctctggt	2400
ggatcagggt ggctccagc acctctttgg tgcgtgtgta cctcttccgg tcgatggtgg	2460
tgtcaaagta cttgaaggcg gcaggggctc ccagattggt cagggtaaac aggtggatga	2520
tattctcggc ctgctctctg atgggcttat cccggtgctt gttglaggcg gacagcattt	2580
tgtccagatt agcgtcggcc aggatcactc tcttgagaa ctcgtgatc tgctcgatga	2640
tctcgtccag gtagtgcttg tgctgttcca caaacagctg tftctgctca ttatcctcgg	2700
gggagccctt cagcttctca tagtggctgg ccaggtacag gaagttcaca tatttggagg	2760
gcagggccag ttcgtttccc tctgcagtt cgcgggcaga ggccagcatt ctcttccggc	2820
cgftttccag ctcgaacagg gagtacttag gcagcttgat gatcaggctc tftttcattt	2880
ctttttagcc cttggcttcc agaaagtcga tgggattctt ctcgaagctg ctcttttcca	2940
tgatggtgat cccagcagc tctttcacac tcttcagttt cttggacttg cctttttcca	3000

ctttggccac caccagcaca gaataggcca cggtagggct gtcgaagccg ccgtacttct 3060

taggggccca gtccttcttt ctggcgatca gcttatcgct gttcctcttg ggcaggatag 3120

actctttgct gaagccgcct gtctgcacct cggctctttt cagcatattc acttggggca 3180

tgctcagcac tticcgcacg gtggcaaaat cccggccctt atcccacacg atctccccgg 3240

tttcgccgtt tgtctcgatc agaggccgct tccggatctc gccgttggcc agggtaatct 3300

cggctcttgaa aaagtcatg atgttctgtt agaagaagta cttggcggta gccttgccga 3360

tttcctgctc gctcttggcg atcatcttcc gcacgtcgta cacctttagt tcgccgtaca 3420

cgaactcgct ttccagctta gggctacttt tgatcagggc gggtcccacg acggcgttca 3480

ggtaggcgtc gtggcgctgg tggtagttgt tgatctcgcg cactttgtaa aactggaaat 3540

ccttcggaa atcggacacc agcttggact tcagggtgat cactttcact tcccgatca 3600

gcttgcatt ctctcgatc ttagtgttca tccgggagtc caggatctgt gccacgtgct 3660

ttgtgatctg ccgggtttcc accagctgtc tcttgatgaa gccggcctta tccagttcgc 3720

tcaggccgcc tctctcgcc ttggtcagat tgcgaactt tctctgggta atcagcttgg 3780

cgttcagcag ctgccccag tagttcttca tcttcttcac gacctcttcg gagggcacgt 3840

tgtcgctctt gccccggtc ttgtcgcttc tggtcagcac cttgttctcg atggagtcgt 3900

ccttcagaaa gctctgagcg acgatatggt ccacatcgta gtcggacagc cggttgatgt 3960

ccagttcctg gtccacgtac atatccgcc cattctgcag gtagtacagg tacagcttct 4020

cgttctgcag ctgggtgttt tccacggggt gttctttcag gatctggctg cccagctctt 4080

tgatgccctc ttgatccgc ttacttctct cgcggctgtt cttctgtccc ttctgggtgg 4140

tctggttctc tctggccatt tcgatcacga tgttctcggg cttgtgccgg cccatcactt 4200

tcacgagctc gtccaccacc ttactgtct gcaggatgcc cttcttaatg gcggggctgc 4260

cggccagatt ggcaatgtgc tcgtgcaggc tatcgccctg gccggacacc tgggctttct 4320

ggatgtctc tttaaaggtc aggtgtctgt cgtggatcag ctgcatgaag tttctgttgg 4380

cgaagccgtc ggacttcagg aaatccagga ttgtcttgcc ggactgcttg tcccggatgc 4440

cgttgatcag ctccggctc agcctgcccc agccggtgta tctccgccgc ttcagctgct 4500

tcatcacttt gtcgtgaac aggtgggcat aggttttcag ccgttctctg atcatctctc 4560

tgtctcaaa cagtgtcagg gtcagcacga tatcttcag aatgtctctg ttttctcat 4620

tgtccaggaa gtccttctc ttgataattt tcagcagatc gtggtatgtg cccaggagg 4680

cgttgaaccg atcttcacg ccggagattt ccacggagtc gaagcactcg attttcttga 4740

agtagtcctc tticagctgc ttcacgggtca ctttccgggtt ggtcttgaac agcaggtcca 4800
 cgatggcctt tttctgctcg ccgctcagga aggcgggctt tctcattccc tcggtcacgt 4860
 atttcacttt ggtcagctcg ttatacacgg tgaagtactc gtacagcagg ctgtgcttgg 4920
 gcagcacctt ctcgttgggc aggttcttat cgaagtgggt catccgctcg atgaagctct 4980
 gggcggaagc gcccttgtcc accacttcct cgaagttcca gggggtgatg gtttcctcgc 5040
 tctttctggt catccaggcg aatctgctgt tttccctggc cagagggccc acgtagtagg 5100
 ggatgcggaa ggtcaggatc ttctcgatct tttcccggtt gtccttcagg aatgggtaaa 5160

 aatcttctg ccgccgcaga atggcgtgca gctctcccag gtggatctgg tgggggatgc 5220
 tgccgttgtc gaaggtccgc tgcttccgca gcaggtcctc tctgttcagc ttcacgagca 5280
 gttcctcggg gccgtccatc ttttccagga tgggcttgat gaactttag aactcttct 5340
 ggctggctcc gccgtcaatg tagccggcgt agccgttctt gctctggtcg aagaaaatct 5400
 ctttgtactt ctcaggcagc tgctccgca cgagagcttt cagcagggtc aggtcctggt 5460
 ggtgctcgtc gtatctcttg atcatagagg cgctcagggg gcccttggtg atctcgggtg 5520
 tcactctcag gatgtcgtc agcaggatgg cgtcggacag gttcttggcg gccagaaaca 5580

 ggtcggcgta ctggtcgccg atctgggcca gcaggttgtc caggctcgtc tcgtaggtgt 5640
 ccttgctcag ctgcagtttg gcacctcgg ccaggtcgaa gttgctcttg aagttggggg 5700
 tcaggcccag gtcagggca atcaggtttc cgaacaggcc attcttcttc tcgcccggca 5760
 gctgggcgat cagattttcc agccgtctgc tcttgctcag tctggcagac aggatggcct 5820
 tggcgtccac gccgtggcg ttgatggggt tttcctcgaa cagctggttg taggtctgca 5880
 ccagctggat gaacagcttg tccacgtcgc tgttgtcggg gttcaggtcg cctcgtatca 5940
 ggaagtggcc ccggaacttg atcatgtggg ccagggccag atagatcagc cgcaggtcgg 6000

 ccttgctggg gctgtccacc agtttctttc tcaggtggta gatggtgggg tacttctcgt 6060
 ggtaggccac ctggtccacg atgttgccga agatggggtg ccgctcgtgc ttcttatcct 6120
 cttccaccag gaaggactct tccagtctgt ggaagaagct gtcgtccacc ttggccatct 6180
 cgttgctgaa gatctcttgc agatagcaga tccggttctt ccgtctggtg tatcttcttc 6240
 tggcggttct cttcagccgg gtggcctcgg ctgtttcgcc gctgtcgaac agcagggtc 6300
 cgatcaggtt cttcttgatg ctgtgccggt cgggtgtgcc cagcaccttg aatttcttgc 6360
 tgggcacctt gtactcgtcg gtgatcacgg cccagccac agagtgggtg ccgatgtcca 6420

 ggccgatgct gtacttcttg tcggctgctg ggactccgtg gataccgacc ttccgttct 6480
 tctttggggc catcttatcg tcatctctt tgaatcaat atcatgatcc ttgtagtctc 6540
 cgtcgtggtc cttatagtc atttttctcg agggatcctg atatattct attaggtatt 6600

tattattata aaatataaat cttgaatgat aataaataaa atattagtta ttccttttct	6660
agttttaa atacatatata taaatatata tatatatata tatattttta ttgtgacaag	6720
aatatataat tataaattat attattttatt ttgtatttt tttttttttt tttttttttt	6780
tctttttttg ttttattttt cttttttttt ataaatatata tttttttctt ttatcatgca	6840
cattggaata atacattaat atatatatat atatttatatt atacatatat tgaataatgt	6900
ttataaaaaa tgcataactt atatgaatat aatttttttt aaatatgaca aaaagaaaaa	6960
aaaaaaaaac caaaaaaat taaaattgaa atgaaatata taaatatatt atttatatat	7020
attatacatt gttaataact actacatgta tatatatata ttatatatat atatatatat	7080
caattttttc aaaaataaat taatatataa agaggggaaa aaaaaaaaaa aaaaaaaaaa	7140
aagataatta agtaagcatt taaaaatata taaattgata atatatataa ttaatcacat	7200
ataaaagcct ataaacacta ggtagctaa ttcgcttgta agaggctact tcgtttatgc	7260
aaaactattt gatatagcat ttaacaagt acacatatat atatgtaata tatatactat	7320
atatacttat tgcattgta ctaagcatgt gcatggcatc ccttttttct cgtgtttaaa	7380
acagtttgta tgataaaata taaaggattt gaaaaagaga aaaaaatata tgatctcatc	7440
ctatatagcg ccataatttt tatttgggtt gaataaaatt ttctactaaa tttagggtga	7500
agtaaaataa tggaatatat ataagtacaa taaaaaagt cataaattaa aaaattttta	7560
taataaatat tttttttaaa aaagtcaata ataattata atatatataa cacaggatta	7620
tatatgttca ctacaatttt ttatatata atataaattc tttcaattt tcattttatt	7680
ttacatacac tttccttttt tgtcactata ttttaattt cacaattta gttaaatac	7740
tggctatttc ttctacatt tgctagtac aattgtgtag tgcttaata tatacacaca	7800
cctaaaactt acaaagtatc ctaggacat ggccaagcct ttgtctcaag aagaatccac	7860
cctcattgaa agagcaacgg ctacaatcaa cagcatcccc atctctgaag actacagcgt	7920
cgccagcgca gctctctcta gcgacggcgg catcttcact ggtgtcaatg tatatcattt	7980
tactggggga cttgtgtag aactcgtgtt gctgggcact gctgctgctg cggcagctgg	8040
caacctgact tgtatcgtcg cgatcggaat tgagaacagg ggcatcttga gccctgcgg	8100
acggtgccga caggtgcttc tcgactgca tcctgggac aaagccatag tgaaggacag	8160
tgatggacag ccgacggcag ttgggattcg tgaattgctg ccctctggtt atgtgtggga	8220
gggctaaccg cgggtacccc attaaattta ttaataata gattaaaaat attataaaaa	8280
taaaaacata aacacagaaa ttacaaaaaa aatacatatg aatttttttt ttgtaatctt	8340
ccttataaat atagaataat gaatcatata aaacatatca ttattcattt atttacattt	8400
aaaattattg tticagtatc ttaattttat tatgtatata taaaaataac ttacaatttt	8460

attaataaac aatatatgtt tattaattca tgttttgtaa tttatgggat agcgattttt	8520
tttactgtct gtattttctt ttttaattat gttttaattg tttttatttt atttttatta	8580
ttgttctttt tatagtatta ttttaaaaca aaatgtatit tctaagaact tataataata	8640
ataatataaa ttttaataaa aattatattt atcttttaca atatgaacat aaagtacaac	8700
attaatatat agcttttaat atttttattc ctaatcatgt aaatcttaaa tttttctttt	8760
taaacatatg ttaaatattt atttctcatt atatataaga acatatttat tacatctaga	8820
ggtaccgagc tcgttttcga cactggatgg cggcgttagt atcgaatcga cagcagtata	8880
gcgaccagca ttcacatacg attgacgcat gatattactt tctgcgcact taacttcgca	8940
tctgggcaga tgatgtcgag gcgaaaaaaa atataaatca cgctaacatt tgattaaaat	9000
agaacaacta caatataaaa aaactataca aatgacaagt tcttgaaaac aagaatcttt	9060
ttattgtcag tactgattag aaaaactcat cgagcatcaa atgaaactgc aatttatcca	9120
tatcaggatt atcaatacca ttttttgaa aaagccgttt ctgtaatgaa ggagaaaact	9180
caccgaggca gttccatagg atggcaagat cctggtatcg gtctgcgatt ccgactcgtc	9240
caacatcaat acaacctatt aatttccctt cgtcaaaaat aaggttatca agtgagaaat	9300
caccatgagt gacgactgaa tccggtgaga atggcaaaag cttatgcatt tctttccaga	9360
cttgttcaac aggccagcca ttacgtctgt catcaaaatc actcgcatca accaaaccgt	9420
tattcattcg tgattgcgcc tgagcgagac gaaatacgcg atcgtgttta aaaggacaat	9480
tacaaacagg aatcgaatgc aaccggcgca ggaacactgc cagcgcatca acaatatit	9540
cacctgaatc aggatattct tctaatacct ggaatgtctt tttgccgggg atcgcagtgg	9600
tgagtaacca tgcatcatca ggagtacgga taaaatgctt gatggtcgga agaggcataa	9660
attccgtcag ccagtttagt ctgaccatct catctgtaac atcattggca acgtacctt	9720
tgccatgttt cagaaacaac tctggcgcgt cgggcttccc atacaatcga tagattgtcg	9780
cacctgattg cccgacatta tcgcgagccc atttatacce atataaatca gcatccatgt	9840
tggaatttaa tcgcggcctc gaaacgtgag tcttttcctt acccatggtt gtttatgttc	9900
ggatgtgatg tgagaactgt atcctagcaa gatitaaaaa ggaagtatat gaaagaagaa	9960
cctcagtggc aaatcctaac cttttatatt tctctacagg ggcgcggcgt ggggacaatt	10020
caacgcgtct gtgaggggag cgtttccctg ctgcaggtc tgcagcgagg agccgtaatt	10080
tttgcttcgc gccgtgcggc catcaaaatg tatggatgca aatgattata catggggatg	10140
tatgggctaa atgtacgggc gacagtcaca tcatgccctt gagctgcgca cgtcaagact	10200

gtcaaggagg gtattctggg cctccatgic gctggcctaa cattagtaat gtaggtctga	10260
ctttcactca tataagtcct atggtaacta aactaaggtc ttacctttac tgatatatgt	10320
cttactttca ctaacttagg tattactttt actaacttag gtcttaaatt cagtaactaa	10380
ggtcatactt cgactaacta aggtcttaca ttcactgata taggtcttat gattactaac	10440
ttaggtccta atttgactaa cataagtcct aacattagta atgtaggtct taacttaact	10500
aacttaggtc ttaccttcac taatataggt ctaaatatta ctgacttaag taattaaggt	10560
actaacttag gtcgtaagg aactaataa taggtcttaa ggtaactaat ttaggtcttg	10620
acttaataaa tataggtcct aacataaata gtatagggcc taatataagt actataggcc	10680
ttaacttaac caacataggt cctaacataa gttatatagg tcttaacgta actaacataa	10740
gtcattaagg tactaagttt ggtcttaatt taacaataac atgtcgctgg ctaacatta	10800
gtaatgtagg tctgacttc actcatataa gtcttatggg aactaaacta aggtcttacc	10860
tttactgata tatgtcttac tttcactaac ttaggtatta cttttactaa cttaggtctt	10920
aaattcagta actaaggcca tacttcgact aactaaggtc ttacattcac tgatataggt	10980
cttatgatta ctaacttagg tcctaatttg actaacataa gtccatacat tagtaatgta	11040
ggtcttaact taactaactt aggtcttacc ttcactaata taggtcttaa tattactgac	11100
ttaagtaatt aaggtaacta cttaggtcgt aaggtaacta atatataggt ctttaaggtaa	11160
ctaatttagg tcttgactta ataatatag gtccatacat aaatagtata ggtcctaata	11220
taagtactat aggccttaac ttaaccaaca taggtcctaa cataagttat ataggcttta	11280
acgtaactaa cataagtcac taaggtaacta agtttggtct taatttaaca ataaccatgt	11340
cgctggccgg gtggtcttaa ttaacaaat atagaccatg tcgctggccg ggtgacccgg	11400
cggggacgag gcaagctaaa cagatcctcg tgatacgctt atttttatag gttaatgtca	11460
tgataataat ggtttcttag gacggatcgc ttgcctgtaa cttacacgcg cctcgtatct	11520
tttaatgatg gaataatttg ggaatttact ctgtgtttat ttatttttat gttttgtatt	11580
tggatttttag aaagtaataa aagaaggtag aagagttacg gaatgaagaa aaaaaataa	11640
acaaaggttt aaaaaatttc aacaaaaagc gtactttaca tataatttta ttagacaaga	11700
aaagcagatt aaatagatat acattcgatt aacgataagt aaaatgtaaa atcacaggat	11760
tttcgtgtgt ggtcttctac acagacaaga tgaacaatt cggcattaat acctgagagc	11820
aggaagagca agataaaagg tagtatattt tggcgatccc cctagagtct ttacatctt	11880
cggaaaacaa aaactatttt ttctttaatt tcttttttta ctttctattt ttaatttata	11940
tattttatatt aaaaaattta aattataatt atttttatag cacgtgatga aaaggacce	12000
ggtggcactt ttcggggaaa tctcgacctg cagcgtagca agct	12044