



(24) 등록일자 2023년 07월 05일

- (73) 특허관자
**템플 유니버시티-오브 더 커먼웰쓰 시스템 오브
 하이어 에듀케이션**
 미합중국 19122 펜실바니아, 필라델피아 브로드
 스트리트 앤드 몽고메리 애비뉴
- (72) 발명자
칼리리 카멜
 미국, 펜실바니아 19004, 발라 신위드, 프리지덴
 설 보울레바드 190, 애플. 718
휴 웬후이
 미국, 뉴저지 08003, 체리 힐, 유로파 코트 224
장 영강
 미국, 뉴저지 08052, 메이플 세이드, 빌리지 오브
 스톤리 런 48이
- (74) 대리인
강명구, 이경민

심사관 : 최홍석

(54) 발명의 명칭 HIV 감염의 RNA-가이드된 치료를 위한 방법 및 조성물

레트로바이러스 내의 보다 큰 서열을 특이적으로 절단하기 위한 조성물은 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 엔도뉴클레아제 및 레트로바이러스 게놈 내 하나 이상의 표적 핵산 서열에 상보적인 가이드 RNA 서열을 코딩하는 핵산을 포함한다.

- 1 -

(52) CPC특허분류

A61P 31/18 (2018.01)
C12N 15/1132 (2013.01)
 C12N 2310/10 (2013.01)
 C12N 2310/20 (2017.05)
 C12N 2320/11 (2013.01)

(30) 우선권주장

62/169,633 2015년06월02일 미국(US)
 62/308,320 2016년03월15일 미국(US)

명세서

청구범위

청구항 1

다음을 포함하는, 조성물:

- (a) 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 Cas9 엔도뉴클레아제를 코딩하는 제 1 핵산 서열;
- (b) 인간 면역결핍 바이러스 (HIV) 서열의 긴 말단 반복(LTR) 서열 내부에 상보적인 제 1 가이드 RNA (gRNA)를 코딩하는 제 2 핵산 서열; 및
- (c) 인간 면역결핍 바이러스 (HIV) 서열의 핵산 서열 GGATAGATGTAAAAGACACCA 를 포함하는 표적 서열에 상보적인 제 2 gRNA를 코딩하는 제 3 핵산 서열, 및
- (d) 부형제.

청구항 2

청구항 1에 있어서, 인간 면역결핍 바이러스 (HIV)은 포유동물 게놈 내 통합됨을 특징으로 하는, 조성물.

청구항 3

삭제

청구항 4

삭제

청구항 5

청구항 1에 있어서, 제 1 gRNA는 서열 번호: 1-8 중 어느 하나에 100% 서열 동일성을 가지는 핵산 서열을 포함함을 특징으로 하는, 조성물.

청구항 6

삭제

청구항 7

청구항 1에 있어서, 제 1 핵산 서열, 제 2 핵산 서열, 및 제 3 핵산 서열은 동일한 발현 벡터에 존재함을 특징으로 하는, 조성물.

청구항 8

청구항 1에 있어서, 제 1 핵산 서열, 제 2 핵산 서열, 및 제 3 핵산 서열은 상이한 발현 벡터에 존재함을 특징으로 하는, 조성물.

청구항 9

다음에 포함하는 제약 조성물:

(a) 다음을 포함하는 발현 벡터:

(i) 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 Cas9 엔도뉴클레아제를 코딩하는 제 1 핵산 서열;

(ii) 인간 면역결핍 바이러스 (HIV) 서열의 긴 말단 반복서열 (LTR)에 상보적인 제 1 가이드 RNA (gRNA)를 코딩하는 제 2 핵산 서열; 및

(iii) 인간 면역결핍 바이러스 (HIV) 서열의 핵산 서열 GGATAGATGTAAAAGACACCA를 포함하는 표적 서열에 상보적인 제 2 gRNA를 코딩하는 제 3 핵산 서열; 및

(b) 제약상 허용되는 부형제.

청구항 10

청구항 9에 있어서, 인간 면역결핍 바이러스 (HIV) 서열은 포유동물 계통에 통합됨을 특징으로 하는, 제약 조성물.

청구항 11

삭제

청구항 12

삭제

청구항 13

청구항 9에 있어서, 제 1 gRNA는 서열 번호: 1-8 중 어느 하나에 100% 서열 동일성을 가지는 핵산 서열을 포함함을 특징으로 하는, 제약 조성물.

청구항 14

삭제

청구항 15

청구항 9에 있어서, 필요로 하는 대상체의 인간 면역결핍 바이러스 감염을 치료하는데 사용하기 위한 것임을 특징으로 하는, 제약 조성물.

청구항 16

다음을 포함하는 발현 벡터:

(a) 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 Cas9 엔도뉴클레아제를 코딩하는 제 1 핵산 서열;

(b) 인간 면역결핍 바이러스(HIV) 서열의 긴 말단 반복(LTR) 서열에 상보적인 제 1 가이드 RNA (gRNA)를 코딩하는 제 2 핵산 서열;

(c) 인간 면역결핍 바이러스 (HIV) 서열의 핵산 서열 GGATAGATGTAAAAGACACCA 를 포함하는 표적 서열에 상보적인 제 2 gRNA를 코딩하는 제 3 핵산 서열, 및

(d) 조절 영역.

청구항 17

청구항 16에 있어서, 인간 면역결핍 바이러스(HIV) 서열은 포유동물 계통에 통합됨을 특징으로 하는 발현 벡터.

청구항 18

청구항 16에 있어서, 제 1 gRNA는 서열 번호: 1-8 중 어느 하나에 100% 서열 동일성을 가지는 핵산 서열을 포함함을 특징으로 하는 발현 벡터.

청구항 19

삭제

청구항 20

다음을 코딩하는 폴리뉴클레오타이드:

- (a) 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 Cas9 엔도뉴클레아제를 코딩하는 제 1 핵산 서열;
- (b) 인간 면역결핍 바이러스 (HIV) 서열의 긴 말단 반복(LTR) 서열에 상보적인 제 1 가이드 RNA (gRNA)를 코딩하는 제 2 핵산 서열; 및
- (c) 인간 면역결핍 바이러스 (HIV) 서열의 핵산 서열 GGATAGATGTAAAAGACACCA 를 포함하는 표적 서열에 상보적인 제 2 gRNA를 코딩하는 제 3 핵산 서열.

청구항 21

청구항 20에 있어서, 인간 면역결핍 바이러스 (HIV) 서열은 포유동물 게놈에 통합됨을 특징으로 하는, 폴리뉴클레오타이드.

청구항 22

청구항 20에 있어서, 제 1 gRNA는 서열 번호: 1-8 중 어느 하나에 100% 서열 동일성을 가지는 핵산 서열을 포함함을 특징으로 하는 폴리뉴클레오타이드.

청구항 23

삭제

청구항 24

삭제

청구항 25

삭제

청구항 26

삭제

청구항 27

삭제

청구항 28

삭제

청구항 29

삭제

청구항 30

삭제

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

삭제

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

발명의 설명

기술 분야

[0001] <연방정부 후원 연구에 관한 진술>

[0002] 본 발명은 미국 국립보건원(NIH)에 의해 카멜 칼라일리 (Kamel Khalili; P30MH092177) 및 베누이 후 (Wenhui Hu; R01NS087971)에 허여된 미국 정부 지원하에 이루어진 것이다. 미국 정부는 본 발명에 대해 특정의 권리를 가질 수 있다.

[0003] 본 발명은 인간 면역결핍 바이러스 (HIV)와 같은 레트로바이러스 내 표적 서열을 특이적으로 절단하는 조성물 및 방법에 관한 것이다. 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (Clustered Regularly Interspace Short Palindromic Repeat: CRISPR)-관련 엔도뉴클레아제 및 인간 면역결핍 바이러스 내 표적 서열에 상보적인 가이드 RNA 서열을 코딩하는 핵산을 포함할 수 있는 조성물이 HIV 감염증에 걸렸거나 걸릴 위험이 있는 대상에 투여될 수 있다.

배경 기술

[0004] HIV-1이 발견된 이래 30년이 넘는 시간 동안, AIDS는 세계적으로 3,530만 명에 영향을 줄 수 있는 주된 공중 건강 문제점으로 남아있다. AIDS는 HIV-1이 숙주 게놈 내로 영구적으로 통합됨으로써 치료가 불가능한 것으로 되어 있다. HIV-1 감염을 억제하고 AIDS 발현을 지연시키는 현재의 치료법 (고활성 항레트로바이러스 요법 또는 HAART)은 HIV-1 감염을 지원하는 세포에서 바이러스 복제를 현격하게 감소시키고, 바이러스혈증을 최소 수준으로 저하시킨다. 그러나, HAART는 조직에서 낮은 수준의 바이러스 게놈 발현 및 복제를 억제하지 못하고, HIV-1 저장소로 작용하는, 예컨대, 휴지기 기억 T 세포, 뇌 매크로파지, 소신경교세포 및 성상 교세포, 장관-관련 림프계 세포와 같은 잠복-감염된 세포를 표적으로 하지 못한다. 지속적 HIV-1 감염은 또한 심장 및 신장 질환, 골감소증, 및 신경계 질환을 포함하는 합병증과 관련이 있다. 지속적 바이러스 저장소를 표적으로 하는 근본적 치료 전략에 대한 지속적인 요구가 있다.

[0005] HIV-1 감염을 억제하고 AIDS 진행을 예방하는 현재의 요법은 HIV-1 감염에 감수성인 세포에서 바이러스 복제를 획기적으로 감소시켰으나, 통합된 HIV-1 프로바이러스 DNA 카피를 함유하는 잠복 감염된 세포에서 낮은 수준의 바이러스 복제를 제거하지 못한다. 숙주 세포 게놈으로부터 프로바이러스 DNA를 퇴치하는 전략을 포함하여, 지속적 바이러스 저장소를 표적으로 하는 근본적 치료 전략의 개발에 대한 시급한 필요가 있다.

[0006] 최근에, 내생 유전자를 뿌리뽑는 몇몇 신규 시스템이 개발되었으며, 이들은 호밍 엔도뉴클레아제 (HE), 징크 핑거 뉴클레아제 (ZFN), 전사 액티베이터-유사 이펙터 뉴클레아제 (TALEN) 및 CRISPR-관련 시스템 9 (Cas9) 단백질을 포함한다.

[0007] CRISPR (주기적 간격으로 분포하는 짧은 회문구조 반복서열) 방법에서, Cas9 뉴클레아제 및 바이러스 DNA 서열 내 표적에 상보적인 가이드 RNA (gRNA)를 포함하는 유전자 편집 복합체가 구성된다. gRNA는 Cas9 뉴클레아제를 유도하여 표적 서열을 함유하는 바이러스 DNA 가닥에 맞물려 그를 절단시키도록 한다. Cas9/gRNA 유전자 편집 복합체는 바이러스 DNA 내로 하나 이상의 변이를 도입시킨다.

[0008] 보존 바이러스 단백질 서열을 표적으로 하는 HE를 사용하여 통합된 HIV-1 프로바이러스를 유전적으로 파괴시키는 것의 실행가능성이 보고된 바 있다. HIV-1 숙주 코-리셉터 CCR5 유전자를 표적으로 하는 ZFN는 HIV/AIDS 치료를 위한 임상 실험 2상에 들어갔다. TALEN은 예상된 위치에서 CCR5를 유효하게 절단하는 것으로 실험적으로 밝혀졌다. Cas9/gRNA 편집 복합체는 또한 HIV-1 엔트리 코-리셉터 (CCR5, CXCR4) 및 프로바이러스 구조 단백질을 파괴하는데 사용되어 왔다 (Manjunath *et al.*, *Viruses*, 14;5(11):2748-2766 (2013); Stone *et al.*, *Curr. Opin. HIV AIDS*, 8(3):217-223 (2013); Wang *et al.*, *PLoS One*, 26;9(12):e115987 (2014)). 그러나, CCR5는 HIV-1 감염에 독특한 리셉터는 아니며, 많은 다른 세포 기능을 갖는다.

발명의 내용

해결하려는 과제

과제의 해결 수단

- [0009] <요약>
- [0010] 본 발명은 레트로바이러스 감염, 특히 인간 면역결핍 바이러스 HIV 감염의 치료 및 예방과 관련된 조성물 및 방법을 제공한다. 조성물 및 방법은 숙주 세포의 게놈 내로 통합된 프로바이러스 HIV를 공격한다.
- [0011] 구체적으로, 본 발명은 CRISPR-관련 엔도뉴클레아제를 코딩하는 핵산 서열, 및 각각의 gRNA가 레트로바이러스 게놈 내의 표적 서열에 상보적인 것인, gRNA를 코딩하는 하나 이상의 단리된 핵산 서열을 포함하는 조성물을 제공한다. 바람직한 실시양태에서, 두 개의 gRNA가 조성물 내에 포함되며, 각 gRNA는 Cas 엔도뉴클레아제를 통합된 레트로바이러스 DNA, 특히 HIV DNA 내의 상이한 표적 부위로 유도한다. 절단 부위 사이의 DNA 연장부가 제거되어, HIV 게놈의 일부 또는 전부가 제거된다. 가장 효과적인 gRNA 조합은 하나의 gRNA가 LTR 영역 내의 부위를 표적으로 하고, 다른 것은 gag 또는 pol과 같은 구조 유전자 내의 부위를 표적으로 하는 쌍; 및 두 개의 gRNA가 LTR 영역 내의 부위를 표적으로 하는 쌍을 포함한다.
- [0012] 본 발명은 또한 포유동물 세포를 유전자 편집 복합체를 코딩하는 하나 이상의 단리된 핵산을 포함하는 조성물에 노출시켜, 포유동물 세포에서 레트로바이러스를 불활성화시키는 방법을 제공한다. 유전자 편집 복합체는 CRISPR-관련 엔도뉴클레아제 및 하나 이상의 gRNA를 포함하며, 각각의 gRNA는 레트로바이러스 내 표적 서열에 상보적이다.
- [0013] 본 발명은 또한 포유동물 대상체 내에서 통합된 레트로바이러스 프로바이러스 DNA를 불활성화시키기 위한 제약 조성물을 제공한다. 그러한 조성물은 Cas 엔도뉴클레아제를 코딩하는 단리된 핵산 서열, 및 HIV DNA와 같은 레트로바이러스 프로바이러스 DNA 내의 표적 서열에 상보적인 하나 이상의 gRNA를 코딩하는 하나 이상의 단리된 핵산 서열을 포함한다. 레트로바이러스 게놈 내의 상이한 부위를 표적으로 하는 gRNA의 쌍이 바람직하다. 단리된 핵산 서열은 하나 이상의 발현 벡터 내에 포함된다.
- [0014] 본 발명은 또한 레트로바이러스, 예를 들어, HIV에 감염된 포유동물 대상체를 치료하는 방법을 제공한다. 그러한 방법은 포유동물 대상체가 HIV에 감염되었음을 판정하는 단계, 유효량의 상기 제약 조성물을 투여하는 단계, 및 포유동물 대상체의 HIV 감염을 치료하는 단계를 포함한다.
- [0015] 본 발명은 또한 레트로바이러스 (예를 들어, HIV) 감염의 위험이 있는 포유동물 대상체에서 그 위험을 감소시키는 치료 방법을 제공한다. 그러한 방법은 포유동물 대상체에 HIV 감염의 위험이 있음을 판정하는 단계, 유효량의 상기 제약 조성물을 투여하는 단계, 및 포유동물 대상체의 HIV 감염 위험을 감소시키는 단계를 포함한다.
- [0016] 본 발명은 또한 HIV 감염의 치료 또는 예방을 위한 키트를 제공한다. 그러한 키트는 CRISPR-관련 엔도뉴클레아제를 코딩하는 하나 이상의 단리된 핵산 서열, 및 HIV 내의 표적 부위에 상보적인 하나 이상의 gRNA를 코딩하는 하나 이상의 단리된 핵산 서열을 포함하는 조성물을 측정된 양으로 포함한다. 키트는 또한 핵산을 코딩하는 하나 이상의 벡터를 포함할 수 있다. 키트는 또한 포장재, 사용 설명이 있는 포장 삽입물, 멸균 액체, 시린지 및/또는 멸균 용기를 포함할 수 있다.

도면의 간단한 설명

- [0017] 도 1A는 스트렙토코커스 파이오제네스 (*Streptococcus pyogenes*) Cas9 (SpCas9), 싱글 가이드 RNA (sgRNA) 및 프로토스페이서 인접 모티프 (PAM)의 다이어그램을 도시한 것으로, 양 DNA 가닥에서 PAM인 NGG 또는 NAG로부터 세번째 뉴클레오타이드에 있는 절단 부위 및 sgRNA 서열을 보여준다. sgRNA는 20 bp 스페이서 (시드 또는 표적 서열) 및 12 bp 스템 루프 (GAAA)를 함유하는 CRISPR RNA (crRNA) 및 트랜스활성화 cRNA (tracRNA)로 이루어져 있다.
- 도 1B는 spCas9 발현 렌티바이러스 벡터 (상측) 및 sgRNA 발현 렌티바이러스 벡터 (하측)의 맵 다이어그램이다. 렌티바이러스 리포터 벡터 (상측)는 바이오세티아 (Biosettia)로부터 구입하였으며, 면역검출용으로 3xFlag를, 간편한 타이타링과 FACS 분석용으로 리포터 적색 형광 단백질 (RFP)을, RFP의 Cas9 기능에 대한 잠재적 효과를 방지하는 자가-절단용 T2A 펩타이드를 발현한다. sgRNA 발현 렌티바이러스 벡터는 애드진 (Addgene) 벡터 (#50946)로부터 변형된 것으로, BbsI 클로닝 부위 및 항생제 선택 마커 푸로마이신, 및 간편한 타이타링과 FACS 분석용으로 리포터 청색 형광 단백질 (BFP)을 포함한다.
- 도 1C는 HIV-1 게놈의 다이어그램으로, 선택된 gRNA 표적 부위인 HIV-1 LTR, Gag (A 내지 D) 및 Pol (A 및 B) 영역이 나타나 있다.

도 1D는 HIV-1 LTR의 U3 영역 내의 sgRNA의 표적이 되는 400 bp의 상세한 다이어그램이다. 시드 서열은 센스 가닥에서 녹색의 PAM (밑줄친 볼드체), 안티센스 가닥에서 적색 PAM (밑줄친 볼드체)로 라벨링되었다. 이들 대부분은 또한 Cas9 Nickase 및 RNA-가이드된 FokI 뉴클레아제 (RFN)와 쌍을 이룰 수 있으며, 이는 잠재적 표적-이탈 효과를 1500배까지 감소시킬 수 있다. 400 bp 영역의 선택은 유전자 및 sgRNA 전달을 위해 현재 사용되고 있는 모든 렌티바이러스 벡터에 그것이 존재하지 않는다는 것에 기초한 것이다. 그와 같은 선택은 Cas9/gRNA에 의한 렌티바이러스 벡터의 자가-절단을 방지할 것이다.

도 1E는 인간 HIV-1_{NL4-3} 벡터로부터 유도된, 증강된 반딧불이 루시페라제 (eLuc)를 함유하는 EcoHIV 리포터 벡터의 다이어그램을 도시한 것이다. eLuc 유전자는 Env와 Nef 사이에 삽입되고, Nef 앞에 자가-절단 2A 펩타이드가 있는 한편, HIV-1의 gp120은 동종지향성 마우스 백혈병 바이러스로부터의 gp80으로 대체되었다.

도 2A 및 2D는 EcoHIV-루시페라제 리포터 어세이에 의한 sgRNA 스크리닝을 도시한 것이다. HEK293T 세포를 EcoHIV-eLuc 리포터, pLV-EF1a-spCas9-T2A-RFP 및 언급된 sgRNA 발현 렌티바이러스 벡터로 동시형질감염시켰다. 2일 후, 세포 용해물 내 루시페라제 활성을 ONE-Glo™ 루시페라제 어세이 시스템으로 측정하였다. 데이터는 4회 독립적인 형질감염의 평균 ± 표준오차를 나타낸다.

도 2B는 임의의 LTR-sgRNA와 쌍을 이룬 Gag-D sgRNA가 루시페라제 활성을 64 내지 96% 감소시킨 것을 보여준다. HEK293T 세포를 EcoHIV-eLuc 리포터, pLV-EF1a-spCas9-T2A-RFP 및 언급된 sgRNA 발현 렌티바이러스 벡터로 동시형질감염시켰다. 2일 후, 세포 용해물 내 루시페라제 활성을 ONE-Glo™ 루시페라제 어세이 시스템으로 측정하였다. 데이터는 4회 독립적인 형질감염의 평균 ± 표준오차를 나타낸다.

도 2C는 임의의 Gag-sgRNA 또는 Pol-sgRNA와 쌍을 이룬 LTR-3 sgRNA가 루시페라제 활성을 73 내지 93% 감소시킨 것을 보여준다. HEK293T 세포를 EcoHIV-eLuc 리포터, pLV-EF1a-spCas9-T2A-RFP 및 언급된 sgRNA 발현 렌티바이러스 벡터로 동시형질감염시켰다. 2일 후, 세포 용해물 내 루시페라제 활성을 ONE-Glo™ 루시페라제 어세이 시스템으로 측정하였다. 데이터는 4회 독립적인 형질감염의 평균 ± 표준오차를 나타낸다.

도 3A는 PCR 제노타이핑이 5'-LTR 표적 부위와 Gag-D 절단 부위 사이에서 HIV-1 DNA를 제거해낸 것을 입증하는 것을 보여준다. 상측 패널: PCR 프라이머 위치. 하측 패널: LTR을 표적으로 하는 각종 sgRNA와 쌍을 이룬 GagD sgRNA. HEK293T 세포를 EcoHIV-eLuc 리포터, pLV-EF1a-spCas9-T2A-RFP 및 언급된 gRNA 발현 벡터로 동시형질감염시켰다. 2일 후, 세포를 50 mM NaOH로 95°C에서 10분 동안 용해시키고, 1M Tris-HCl로 중화시켰다. 조직의 추출물을 직접 PCR에 사용하였으며, 이때 테라 (Terra) PCR 다이렉트 폴리머라제 믹스 (Clontech) 및 5'-LTR 및 5'-부분 Gag를 커버링하는 PCR 프라이머 T361/T458를 사용하였으며(1364 bp), 이는 속이 빈 sgRNA 발현 벡터로 형질감염시킨 대조 샘플에서 1.35 kb 단편을 생산한다.

도 3B는 PCR 제노타이핑이 3'-LTR 표적 부위와 Gag-D 절단 부위 사이에서 HIV-1 DNA를 제거해낸 것을 입증하는 것을 보여준다. 상측 패널: PCR 프라이머 위치. 하측 패널: LTR을 표적으로 하는 각종 sgRNA와 쌍을 이룬 GagD sgRNA. HEK293T 세포를 EcoHIV-eLuc 리포터, pLV-EF1a-spCas9-T2A-RFP 및 언급된 gRNA 발현 벡터로 동시형질감염시켰다. 2일 후, 세포를 50 mM NaOH로 95°C에서 10분 동안 용해시키고, 1M Tris-HCl로 중화시켰다. 조직의 추출물을 직접 PCR에 사용하였으며, 이때 테라 (Terra) PCR 다이렉트 폴리머라제 믹스 (Clontech) 및 3'-LTR 및 부분 5' Gag 서열을 제외한 전체 HIV-1 게놈을 커버링하는 PCR 프라이머 T758 (뉴클레오타이드 796-817)/T645 (3'-LTR 이후의 벡터 서열을 표적으로 함)를 사용하였으며, 이는 속이 빈 sgRNA 발현 벡터로 형질감염시킨 대조 샘플에서 보통의 PCR 조건 하에 검출될 수 없는, 예상된 9.5 kb 단편을 생산한다.

도 3C는 PCR 제노타이핑이 LTR-3 표적 부위와 Gag 또는 Pol 표적 부위 사이에서 HIV-1 DNA를 제거해낸 것을 입증하는 것을 보여준다. 상측 패널: PCR 프라이머 위치. 하측 패널: LTR-1, 2, 3 sgRNA와 쌍을 이룬 GagD sgRNA, 또는 Gag 또는 Pol을 표적으로 하는 각종 sgRNA와 쌍을 이룬 LTR-3 sgRNA. HEK293T 세포를 EcoHIV-eLuc 리포터, pLV-EF1a-spCas9-T2A-RFP 및 언급된 gRNA 발현 벡터로 동시형질감염시켰다. 2일 후, 세포를 50 mM NaOH로 95°C에서 10분 동안 용해시키고, 1M Tris-HCl로 중화시켰다. 조직의 추출물을 직접 PCR에 사용하였으며, 이때 테라 (Terra) PCR 다이렉트 폴리머라제 믹스 (Clontech) 및 5'-LTR과 5'-게놈 서열 사이의 서열 (좌측 패널) 또는 3'-게놈 서열과 3'-LTR 사이의 서열 (우측 패널)을 커버링하는 PCR 프라이머를 사용하였다.

도 3D는 PCR 제노타이핑에 의한 유효 gRNA의 검증에 사용된 PCR 프라이머 및 Gag/Pol gRNA 표적 부위 위치의 다이어그램이다.

도 3E 및 3F는 각종 LTR-gRNA와 쌍을 이룬 GagD의 결과를 보여주는 블롯이다.

도 3G 및 3H는 Gag 및 Pol을 표적으로 하는 각종 gRNA와 쌍을 이룬 LTR-R의 결과를 보여주는 블롯이다. 5'-LTR-Gag 또는 Gag-3'-LTR의 결실이 검출되었다. 야생형 (WT), 결실 및 삽입의 밴드 밀도가 NIH 이미지 J 프로그램으로 정상화되었다. 겔 아래의 숫자는 Cas9 특이적 프라이머 T477/T478를 사용한 Cas9 PCR 생성물로 표준화한 후의 속이 빈 gRNA 대조에 대한 WT 밴드 변화율 (%), 및 WT 밴드와 비교하여 삽입 및 결실 밴드에서의 배수 변화를 나타낸다. 상응하는 gRNA에 의해 유도된 획기적인 변화는 박스로 강조하였다. 박스는 TA-클로닝 및 생거 (Sanger) 시퀀싱에 선택된 샘플을 나타낸다.

도 3I 및 3J는 구조 영역과 3'-LTR을 커버링하는 추가의 PCR 프라이머를 사용한 다이렉트 PCR 제노타이핑에 의한 유효 gRNA의 또다른 검증을 보여준다. 도 3I는 gRNA GagD의 5'-상류의 정방향 프라이머 (T758)가 3'-말단 LTR 하류의 백터상의 역방향 프라이머 (T645)와 쌍을 이루는 것을 보여준다. 도 3J는 gRNA GagD의 5'-상류의 정방향 프라이머 (T758)가 3'-말단 LTR의 R 영역의 역방향 프라이머 (T422)와 쌍을 이루는 것을 보여준다. 화살표는 비-특이적 밴드 (ns)를 나타낸다.

도 4는 대표적인 TA-클로닝 및 생거 시퀀싱이 LTR-1과 LTR-3 사이에서 296 bp 결실 및 두 개의 절단 부위 사이에서 180 bp 추가 삽입을 확인하는 것을 보여준다. 샘플 제조 및 다이렉트 PCR은 도 3A에 기재된 바와 같이 수행되었다. 절단 후 PCR 단편을 TA 클로닝 및 생거 시퀀싱을 위해 추출하였다. 빨간 화살표는 PAM으로부터 세번째 뉴클레오타이드에서 예상된 절단 부위를 나타낸다. 빨간색 밑줄은 PAM 서열을 나타낸다.

도 5A는 HIV-1 게놈의 개략적인 표시로서, (통합된 프로바이러스를 플랭킹하는) 바이러스 LTR 서열을 Cas9/gRNA 복합체로 성공적으로 표적화하는 전략의 예상되는 결과를 포함한다.

도 5B는 LTR의 상세한 구조이다.

도 5C는 표적 부위의 서열 및 LTR 내 그의 위치를 보여준다.

도 6A는 저캣 (Jurkat) 2D10 리포터 세포주의 다이어그램으로서, 통합된 HIV-1 리포터 서열의 표시를 포함한다.

도 6B는 잠복된 프로바이러스 서열의 PMA/TSA 유도된 재활성화를 보여주는 형광현미경 사진이다.

도 6C는 잠복된 프로바이러스 서열의 PMA/TSA 유도된 재활성화를 나타내는 유세포분석 히스토그램이다.

도 7A는 단세포 클론 스크린 결과이다.

도 7B는 단세포 클론 스크린에서 얻어진 클론을 유세포 분석한 것에 따른, 최상의 클론의 확인을 나타낸다.

도 7C는 FLAG-Cas9에 대한 예시적인 웨스턴 블롯 (상측 패널) 및 gRNA 발현에 대한 RT-PCR 아가로즈 겔 전기영동 (하측 패널)이다.

도 8A는 숙주 DNA로부터 HIV-1 프로바이러스 게놈의 제거 확인을 위한 PCR 분석에 사용된 프라이머의 위치를 나타낸다. 프라이머는 프로바이러스 Env 유전자 서열 모티프 (RRE), 통합 리포터 프로바이러스를 플랭킹하는 게놈 서열 (염색체 16, MSRB1 유전자), LTR 및 대조 b-액틴 유전자에 특이적이었다.

도 8B는 PCR 반응의 아가로즈 겔 사진이며, 화살표는 HIV-1 서열의 절제로 인해 사라진 밴드의 위치를 표시한다.

도 8C는 짧은 길이 생성물에 최적화된 조건하의 롱 레인지 PCR 데이터를 나타내며, 이는 통합 부위의 프로바이러스 래리어트 (lariat) 서열의 검출을 가능하게 한다.

도 8D는 프로바이러스 래리어트의 시퀀싱 결과이다.

도 9A는 저캣 2D10 단세포 클론의 형광분석 결과로서, 클론을 HIV-1 NL-4-3-EGFP-P2A-Nef 리포터 바이러스로 감염시키고, 감염 진행을 18일 동안 형광 분석 형식으로 모니터링한 결과이다.

도 9B는 시험된 클론에서 Cas9-FLAG 발현을 나타내는 웨스턴 블롯을 보여준다.

도 9C는 선택된 클론에서 gRNA 발현을 위한 역전사 PCR의 아가로즈 겔 사진이다.

도 10A는 RFP-Cas9 및/또는 LTR A/B' gRNA를 발현하는 렌티바이러스로 형질도입된 저캣 2D10 세포의 형광현미경 사진이다.

도 10B는 RFP-Cas9 및/또는 LTR A/B' gRNA 발현 및 PMA/TSA 유도 후 바이러스 재활성화에 대한 유세포분석이다.

도 11A는 LTR A 및 LTR B'의 서열이다.

도 11B는 Cas9 및 LTR A/B'의 발현 후 표적-이탈 인델 (indel)의 서베이어 (Surveyor) 어레이 분석 결과이다.

도 11C는 표적-이탈 인델의 생거 시퀀싱 결과이다.

도 11D는 염색체 16 내의 MSRB1 유전자의 두번째 엑손 내의 HIV-1 리포터 통합 부위, 및 인접 유전자의 위치를 나타낸다.

도 11E는 대조 및 Cas9/LTR AB' 발현 세포에서 인접 유전자 발현에 대한 qRT-PCR 비교의 결과이다.

도 12A 내지 12C는 대표적인 샘플에 대한 TA-클로닝 및 생거 시퀀싱의 다이어그램이 상응하는 gRNA 표적 부위에서 예상된 단편의 결실 (도 12A 내지 12C) 및 다양한 추가 삽입 (도 12B 및 12C)을 확인하는 것을 보여준다. PAM 서열은 강조되고, 가위는 PAM으로부터 세번째 뉴클레오타이드를 나타낸다. 화살표는 절단 및 리게이션 후 연결 부위를 나타낸다.

도 13A 내지 13E는 EcoHIV-반딧불이-루시퍼라제 안정한 HEK293T 세포주의 수립을 보여준다. 6-웰 플레이트 중 하나의 웰에서 HEK293T 세포를 3 μ g의 pEcoHIV-eLuc 플라스미드 및 1 μ g의 VSV-G 플라스미드로 동시형질감염시키고, 48시간 후에 EcoHIV-eLuc 바이러스를 수확하였다. 같은 부피의 바이러스 상등액 (250 μ l)을 12-웰 플레이트 상의 HEK293T 세포 (2×10^4 /웰)에 가하였다. 48시간 후, eLuc 루시퍼라제 활성을 ONE-Glo™ 루시퍼라제 어레이 (프로메가; Promega)로 측정한 결과, 바이러스 처리가 없는 대조 HEK293T 세포보다 10^5 -배 높게 나타났다. 이어서, 단세포를 4개의 96-웰 플레이트 상에서 한계 희석으로 배양하였다. 2 내지 3주 후, 생존 세포 콜로니를 단리하여, ONE-Glo™ 루시퍼라제 어레이로 eLuc 리포터 활성에 대해 시험하고 (도 13A, 13B), 프라이머 T361 (5'-gatctgtggatctaccacacaca-3') 및 T458 (5'-ccactgtgttagcatgtatt-3')을 사용한 다이렉트-PCR 제노타이핑으로 EcoHIV-eLuc 도입유전자를 검증하였다 (도 13C 내지 13E). 1차 실험에서의 단세포-유도된 13개 클론 (도 13A, 1016으로 번호부여)의 반과, 2차 실험에서의 20개 클론 (도 13B, 1021로 번호부여)의 반은 상이한 정도의 구성적 eLuc 활성을 보였다. 이들 중 단지 두 개 (1016-12 및 1021-19)가 TNF α (10 ng/ml) 및 PMA (10 ng/ml)에 의한 자극에 상당히 반응성이었다. eLuc-발현 클론은 클론 1021-6 (녹색 사각형)을 제외하고 도입 유전자를 함유하였으며 (도 13C 내지 13E), 클론 1021-6의 경우 통합 중에 Gag 서열을 소실하였을 수 있으나, 다른 스플라이싱으로 eLuc를 전사한다. 두 개의 클론 (1016-6 및 1016-9)은 형질도입 유전자를 함유하였으나, TNF α /PMA 등과 같은 잠복성-역전제로 처리한 후에도 eLuc 활성을 나타내지 않았다 (도 13A 및 13D에서 적색 사각형). 따라서, eLuc 활성을 갖는 클론을 선택하여, 후속 연구를 위해 유지시켰다. 이들 클론의 일부는 추측하기로는 독성 바이러스 단백질의 계속적인 생산으로 인해 계대하지 못하였으며, 일부 클론은 리포펙타틴 형질감염 키트를 사용하는 경우에도 형질감염 저항성을 나타냈다.

도 14A 내지 14D는 EcoHIV-반딧불이-루시퍼라제 안정성 HEK293T 세포주를 사용한 gRNA 효율 스크리닝의 결과이다. 도 14A 및 14B: EcoHIV-eLuc 안정적 발현 클론을 2주 동안 pCW-Cas9-퓨로마이신 렌티바이러스로 10 MOI로 감염시키고, 퓨로마이신 (1 μ g/ml)으로 선택하였다. 세포를 언급된 gRNA-발현 벡터로 형질감염시켰다. 2일 후, ONE-Glo™ 루시퍼라제 어레이를 수행하였다. 도 14C 및 14D: EcoHIV-eLuc 안정적 발현 세포를 언급된 gRNA-발현 벡터 및 pLV-EF1 α -spCas9-T2A-RFP 벡터로 동시형질감염시켰다. 48시간 후, 루시퍼라제 활성을 ONE-Glo™ 루시퍼라제 어레이로 측정하였다. 데이터는 4회 독립적인 형질감염에서의 평균 \pm 표준 오차를 나타내며, 속이 빈 gRNA 제로(Zero) 그룹과 비교한 eLuc 활성의 퍼센트 변화율이다.

발명을 실시하기 위한 구체적인 내용

[0018] 본 발명은 부분적으로는 통합된 인간 면역결핍 바이러스 (HIV) 게놈이 싱글 및 멀티플렉스 배열의 RNA-가이드된 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-Cas9 뉴클레아제 시스템 (Cas9/gRNA)을 사용하여 HIV 감염된 세포로부터 제거될 수 있다는 발견에 기초한 것이다.

[0019] 정의

[0020] 달리 언급이 없는 한, 본 명세서에 사용된 모든 기술 및 과학 용어는 본 발명이 속하는 기술 분야의 통상의 전문가에 의해 통상적으로 이해되는 것과 같은 의미를 갖는다. 본 명세서에 기재된 것과 유사하거나 동등한 방법 및 재료는 어느 것이나 본 발명의 시험을 실시하는데 사용될 수 있지만, 바람직한 재료 및 방법이 본 명세서에 기재되어 있다. 본 발명을 기재하고 청구함에 있어서, 다음과 같은 용어가 사용될 것이다. 본 명세서에 사용된 용어는 단지 특정 실시양태를 기술하기 위한 것으로, 본 발명을 제한하려는 것이 아님이 또한 이해되어야 한다.

- [0021] 본 명세서에 개시된 모든 유전자, 유전자 명칭 및 유전자 생성물은 본 명세서에 개시된 조성물 및 방법이 적용될 수 있는 어떠한 종으로부터의 상동체나 해당하는 것으로 의도된다. 특정 종으로부터의 유전자 또는 유전자 생성물이 개시된 경우, 그러한 개시는 단지 예시를 위한 것일 뿐, 문맥에서 명확하게 언급하고 있지 않은 한, 그에 한정적인 것으로 해석되어서는 안된다는 것이 이해되어야 한다. 따라서, 예를 들어, 본 명세서에 개시된 유전자 또는 유전자 생성물은 상동 및/또는 다른 종으로부터의 이종상동 유전자 및 유전자 생성물을 포함하는 것으로 의도된다.
- [0022] 본 명세서에서 수사의 수식이 없는 명사는 하나 또는 하나를 초과하는 (즉, 적어도 하나의) 그러한 명사를 의미한다. 예를 들어, "요소"는 하나의 요소 또는 하나를 초과하는 요소를 의미한다. 따라서, 예를 들어, "세포"라 함은 같은 형태의 복수 개의 세포를 포함하는 것이다. 또한, "포함하는 (including)", "포함한다", "갖는 (having)", "갖는다" 또는 그의 변형체가 상세한 설명이나 청구범위에 사용되는 한, 그러한 용어는 "포함하는 (comprising)"이라는 용어와 유사한 방식으로 포괄적인 의미를 나타내려는 것이다.
- [0023] 본 명세서에서, 물건, 조성물, 장치, 방법, 프로세스, 시스템 등의 정의되거나 기재된 요소와 관련하여 사용된 "포함하는 (comprising)," "포함한다 (comprise)" 또는 "이루어진 (comprised)" 및 그의 변형체는 포괄적이거나 확장 가능한 것을 의미하는 것으로, 추가의 요소를 허용함으로써 상기 정의되거나 기재된 물건, 조성물, 장치, 방법, 프로세스, 시스템 등이 구체적으로 언급된 요소나 그의 등가물, 경우에 따라서는 다른 요소를 포함할 수 있으며, 그로써 여전히 정의된 물건, 조성물, 장치, 방법, 프로세스, 시스템 등의 범주/정의에 든다는 것을 나타낸다.
- [0024] 본 명세서에서 양, 시간적 경과 등과 같이 측정 가능한 값과 관련하여 사용되는 "약"이란 용어는 변화량의 범위가 개시된 방법을 수행하는데 적절한 한, 구체적으로 명시된 값으로부터 $\pm 20\%$, $\pm 10\%$, $\pm 5\%$, $\pm 1\%$ 또는 $\pm 0.1\%$ 의 변화량을 포함하는 것을 의미한다. 또한, 특히 생물학적 시스템이나 프로세스와 관련하여, 그러한 용어는 주어진 값의 5-배 이내의 크기, 또는 2-배 이내의 크기를 의미할 수 있다. 명세서 및 청구범위에 특정 값이 기재된 경우, 달리 언급이 없는 한, 특정 값에 대해 허용되는 오차 범위 내를 의미하는 "약"이란 의미가 전제되어야 한다.
- [0025] "코딩하는"이란 생물학적 프로세스에서 정의된 뉴클레오타이드 서열 (즉, rRNA, tRNA 및 mRNA) 또는 정의된 아미노산 서열 및 그들로부터 도출되는 생물학적 특성을 갖는 다른 중합체 및 매크로분자 합성시에 주형으로 기능하는, 유전자, cDNA 또는 mRNA와 같은 폴리뉴클레오타이드 내의 특이적 뉴클레오타이드 서열의 고유의 특성을 언급하는 것이다. 따라서, 하나의 유전자에 상응하는 mRNA의 전사 및 번역에 의해 세포 또는 다른 생물학적 시스템 내에서 단백질이 생산되면, 그러한 유전자는 그러한 단백질을 코딩하는 것이다. 뉴클레오타이드 서열이 mRNA 서열과 일치하며, 대개 서열 리스트에 제공되는 코딩 가닥, 및 유전자 또는 cDNA 전사에 주형으로 사용되는 비-코딩 가닥 모두가 그러한 유전자 또는 cDNA의 단백질 또는 다른 생성물을 코딩하는 것으로 될 수 있다.
- [0026] 화합물의 "유효량" 또는 "치료 유효량"이란 그 화합물이 투여되는 대상체에 유익한 효과를 제공하기에 충분한 화합물의 양을 의미한다. 전달 비히클의 "유효량"이란 화합물과 유효하게 결합하거나 그를 전달하기에 충분한 양이다.
- [0027] 본 명세서에서 바이러스, 예를 들어, HIV의 "퇴치"란 바이러스가 복제를 할 수 없거나, 게놈이 제거되거나, 단편화되거나, 분해되거나, 유전적으로 불활성화되거나, 바이러스가 전염되거나 다른 세포 또는 대상체를 감염시킬 수 없도록 기타 물리적, 생물학적, 화학적 또는 구조적으로 조작되어 생체내에서 바이러스가 사라진 것을 의미한다. 일부 경우에, 바이러스 게놈의 단편이 검출될 수 있으나, 바이러스는 복제 또는 감염 등이 불가능하다.
- [0028] "발현 벡터"란 발현될 뉴클레오타이드 서열에 작동가능하게 연결된 발현 조절 서열을 포함하는 재조합 폴리뉴클레오타이드를 포함하는 벡터를 이른다. 발현 벡터는 발현에 충분한 시스-액팅 요소를 포함하며, 발현에 필요한 다른 요소는 숙주 세포 또는 시험관내 발현 시스템에 의해 공급될 수 있다. 발현 벡터는 해당 기술 분야에 알려진 모든 것, 예를 들어, 재조합 폴리뉴클레오타이드를 포함하는 코스미드, 플라스미드 (예를 들어, 네이키드 또는 리포솜 내 함유된 형태) 및 바이러스 (예를 들어, 렌티바이러스, 레트로바이러스, 아데노바이러스, 아데노-연관 바이러스)를 포함한다.
- [0029] "상동성"은 두 개의 폴리펩타이드 또는 두 개의 핵산 분자 사이의 서열 유사성 또는 서열 동일성을 이른다. 두 개의 비교 서열 중의 어떤 위치가 동일한 염기 또는 아미노산 단량체 소단위에 의해 점유된 경우, 예를 들어, 두 개의 DNA 분자 각각에서 어느 한 위치가 아데닌에 의해 점유된 경우, 두 분자는 그 위치에서 상동이다. 두 서열 사이의 상동성 퍼센트는 두 서열에 의해 공유된 매칭 또는 상동 위치의 수를 비교된 위치의 수로 나누고,

이에 100을 곱하여 계산된다. 예를 들어, 두 서열 내 10개 위치 중 6개가 매칭되거나 상동인 경우, 두 서열은 60% 상동성이다. 예를 들어, DNA 서열 ATTGCC와 TATGGC는 50% 상동성이다. 일반적으로, 비교는 두 서열이 최대 상동성으로 정렬되었을 때 이루어진다.

- [0030] "단리된"이란 자연 상태에서부터 변화되었거나 회수된 것을 이른다. 예를 들어, 살아있는 동물체 내에 자연적으로 존재하는 핵산 또는 펩타이드는 "단리된" 것이 아니지만, 같은 핵산이나 펩타이드가 그의 자연 상태에서 공존하는 물질로부터 부분적으로 또는 완전히 분리되면 "단리된" 것이다. 단리된 핵산 또는 단백질은 실질적으로 정제된 형태로 존재할 수 있거나, 숙주 세포와 같이 본래의 것이 아닌 환경 내에 존재할 수 있다.
- [0031] 본 발명의 기재에서, 통상적으로 존재하는 핵산 염기에 대하여 다음과 같은 약어가 사용된다. "A"는 아데노신을, "C"는 시토신을, "G"는 구아노신을, "T"는 티미딘을, "U"는 우리딘을 이른다.
- [0032] 달리 언급이 없는 한, "아미노산 서열을 코딩하는 뉴클레오타이드 서열"은 서로에 대해 축퇴 (degenerate) 형태이고, 같은 아미노산 서열을 코딩하는 모든 뉴클레오타이드 서열을 포함한다. 단백질 또는 RNA를 코딩하는 뉴클레오타이드 서열은, 그것이 일부 형태에서 인트론(들)을 함유할 수 있는 한, 인트론을 또한 포함할 수 있다.
- [0033] "환자," "대상체," "개체" 등은 본 명세서에서 상호교환적으로 사용되며, 본 명세서에 기재된 방법에 사용될 수 있는 동물, 또는 그의 시험관내 또는 생체내 세포를 이른다. 일부 비-제한적 실시양태에서, 환자, 대상체 또는 개체는 인간이다.
- [0034] 조성물의 "비경구" 투여는, 예를 들어, 피하 (s.c.), 정맥내 (i.v.), 근육내 (i.m.) 투여, 또는 흉골내 주사 또는 주입 기술을 포함한다.
- [0035] 본 명세서에서, "폴리뉴클레오타이드," "핵산 서열" 및 "유전자"는 명세서 전체를 통하여 상호교환적으로 사용되며, 상보적 DNA (cDNA), 천연 및/또는 변형된 단량체 또는 결합의 선형 또는 환형 올리고머 또는 중합체를 포함하며, 단량체는, 예를 들어, 데옥시리보뉴클레오사이드, 리보뉴클레오사이드, 그의 치환 및 알파-아노머 형태, 펩타이드 핵산 (PNA), 록트 (locked) 핵산 (LNA), 포스포로티오에이트, 메틸포스포네이트 등을 포함한다. 본 명세서에서 사용되는 "폴리뉴클레오타이드"는 뉴클레오타이드의 사슬로서 정의된다. 또한, 핵산은 뉴클레오타이드의 중합체이다. 따라서, 본 명세서에서 핵산과 폴리뉴클레오타이드는 상호교환적으로 사용될 수 있다. 해당 기술 분야의 숙련가는 핵산은 폴리뉴클레오타이드로서 단량체인 "뉴클레오타이드"로 가수분해될 수 있다는 일반적인 지식을 가지고 있다. 단량체 뉴클레오타이드는 뉴클레오사이드로 가수분해될 수 있다. 폴리뉴클레오타이드는 해당 기술 분야에서 이용될 수 있는 것으로서, 비-제한적으로는 재조합 수단, 즉, 통상의 클로닝 기술 및 PCR™ 등을 사용하여 재조합 라이브러리 또는 세포 게놈으로부터 핵산 서열을 클로닝하는 수단, 및 합성 수단에 의해 얻어지는 모든 핵산 서열을 포함하나, 이에 한정되는 것은 아니다. 핵산 서열은 "키메라", 즉, 다른 영역들로 이루어질 수 있다. 본 발명의 기재에서, "키메라" 화합물은 둘 이상의 화학적 영역, 예를 들어, DNA 영역(들), RNA 영역(들), PNA 영역(들) 등을 함유하는 올리고뉴클레오타이드이다. 각 화학적 영역은 하나 이상의 단량체 단위, 즉, 뉴클레오타이드로 이루어진다. 이들 서열은 전형적으로는 하나 이상의 목적하는 특성을 나타내기 위하여 변형이 이루어진 하나 이상의 영역을 포함한다.
- [0036] 뉴클레오사이드와 관련하여 "유사체"는 변형된 염기 모이어티 및/또는 변형된 당 모이어티를 갖는 합성 뉴클레오사이드를 포함하며, 이와 같은 합성 뉴클레오사이드의 예는 문헌 (Scheit, Nucleotide Analogs, John Wiley, New York, 1980; Freier & Altmann, *Nucl. Acid. Res.*, 1997, 25(22), 4429-4443; Toulme, J.J., *Nature Biotechnology* 19:17-18 (2001); Manoharan M., *Biochemica et Biophysica Acta* 1489:117-139(1999); Freier S. M., *Nucleic Acid Research*, 25:4429-4443 (1997); Uhlman, E., *Drug Discovery & Development*, 3:203-213 (2000); Herdewin P., *Antisense & Nucleic Acid Drug Dev.*, 10:297-310 (2000))에 일반적으로 기재된 뉴클레오사이드; 및 2'-O,3'-C-결합[3.2.0]비시클로아라비도뉴클레오사이드 (예를 들어, N.K Christensen., *et al.*, *J. Am. Chem. Soc.*, 120:5458-5463 (1998))이다. 그러한 유사체는 듀플렉스 또는 트리플렉스 안정성, 특이성 등과 같은 결합 특성을 향상시키도록 설계된 합성 뉴클레오사이드를 포함한다.
- [0037] 폴리뉴클레오타이드 서열과 관련하여 사용되는 "변이체"는 야생형 유전자와 관련된 폴리뉴클레오타이드 서열을 포함할 수 있다. 이러한 정의는 또한, 예를 들어, "대립유전자," "스플라이스," "중," 또는 "다형성" 변이체를 포함할 수 있다. 스플라이스 변이체는 표준 분자와 상당한 일치도를 가질 수 있으나, 일반적으로는 mRNA 프로세싱 중 엑손을 번갈아 가며 스플라이싱하는 것에 기인하여 더 많거나 더 적은 수의 뉴클레오타이드를 갖는다. 상응하는 폴리펩타이드는 추가의 기능성 도메인을 갖거나 도메인 결실을 나타낼 수 있다. 중 변이체는 종에 따라 달라지는 폴리뉴클레오타이드 서열이다. 본 발명에 특히 유용한 것은 야생형 유전자 생성물의 변이체이다. 변이

체는 핵산 서열 내의 적어도 하나의 돌연변이에 의해 생겨나서, 변화된 mRNA, 또는 구조나 기능이 변화되거나 변화되지 않을 수 있는 폴리펩타이드를 생성할 수 있다. 주어진 천연 또는 재조합 유전자는 대립유전자 형을 가지지 않거나, 하나 또는 다수 가질 수 있다. 변이체를 생성시키는 통상의 돌연변이 변화는 일반적으로는 뉴클레오타이드의 자연적인 결실, 부가 또는 치환에 의한다. 이들 변화 형태 각각은 주어진 서열에서 한 번 이상, 단독으로 또는 다른 것과 병행하여 일어날 수 있다.

[0038] 달리 언급이 없는 한, "아미노산 서열을 코딩하는 뉴클레오타이드 서열"은 서로에 대해 동의(degenerate) 형태이며, 동일한 아미노산 서열을 코딩하는 모든 뉴클레오타이드 서열을 포함한다. 단백질 또는 RNA를 코딩하는 뉴클레오타이드 서열은, 그것이 일부 형태에서 인트론(들)을 함유할 수 있는 한, 인트론을 또한 포함할 수 있다.

[0039] 본 명세서에서 "펩타이드," "폴리펩타이드," 및 "단백질"은 상호교환적으로 사용되며, 펩타이드 결합에 의해 공유결합된 아미노산 잔기들로 이루어지는 화합물을 이른다. 단백질 또는 펩타이드는 적어도 두 개의 아미노산을 함유하여야 하며, 단백질 또는 펩타이드 서열을 구성할 수 있는 아미노산의 최대 갯수에는 제한이 없다. 따라서, 예를 들어, 올리고펩타이드, 단백질 및 효소는 그들이 재조합 기술, 화학적 또는 효소 합성에 의해 생산된 것이거나 천연에 존재하는 것이거나를 불문하고 폴리펩타이드 또는 펩타이드의 정의 내에 포함된다. 폴리펩타이드는 서로 펩타이드 결합으로 연결된 두 개 이상의 아미노산을 포함하는 펩타이드 또는 단백질은 어느 것이나 포함한다. 본 명세서에서 폴리펩타이드란 용어는 당 기술 분야에서 통상적으로 펩타이드, 올리고펩타이드 및 올리고머 등으로 지칭되는 단쇄 폴리펩타이드, 및 당 기술 분야에서 일반적으로 여러 가지 형태의 단백질로 지칭되는 장쇄 폴리펩타이드를 모두 이르는 것이다. "폴리펩타이드"는, 예를 들어, 생물학적 활성 단편, 실질적 상동성 폴리펩타이드, 올리고펩타이드, 호모이합체, 헤테로이합체, 폴리펩타이드 변이체, 변형 폴리펩타이드, 유도체, 유사체, 융합 단백질을 포함한다. 폴리펩타이드는 또한, 대표적으로는 글리코실화, 아세틸화, 포스포릴화 등에 의해 변형되거나 유도된 폴리펩타이드를 포함한다. 폴리펩타이드는 천연 펩타이드, 재조합 펩타이드, 합성 펩타이드 또는 그들의 조합을 포함한다.

[0040] 본 명세서에서, 폴리펩타이드의 "변이체"란 하나 이상의 아미노산 잔기에서 변화가 있는 아미노산 서열을 이른다. 변이체는 "보존적" 변화를 가질 수 있으며, 이때 치환된 아미노산은 유사한 구조적 또는 화학적 특성을 갖는다 (예를 들어, 류신이 이소류신으로 치환된 경우). 보다 드물게는, 변이체는 "비보존적" 변화를 가질 수 있다 (예를 들어, 글리신이 트립토판으로 치환된 경우). 유사한 경미한 변이는 또한 아미노산 결실 또는 삽입, 또는 둘 다를 포함할 수 있다. 생물학적 활성을 소실하지 않으면서, 어떠한 아미노산 잔기가 치환, 삽입 또는 결실될 수 있는지를 결정함에 있어서, 예를 들어, LASERGENE 소프트웨어 (DNASTAR)와 같이 당 기술 분야에 잘 알려진 컴퓨터 프로그램을 사용하여 안내받을 수 있다.

[0041] "치료적" 처치는 병의 증후를 나타내는 대상체에 그러한 증후를 감소시키거나 제거할 목적으로 부여되는 처치이다.

[0042] 본 명세서에서 사용되는 "치료 유효량"이란 질병 또는 증상을, 그의 증후를 경감시키는 것을 포함하여, 예방 또는 치료 (질병 또는 증상의 개시 지연 또는 방지, 질병 또는 증상의 진행 방지, 또는 질병 또는 증상의 억제, 감소, 역전)하는데 충분하거나 유효한 양을 이른다.

[0043] "치료"란 질병 또는 질병 증후의 진행을 방지하거나 변화시킬 목적으로 개입된 행위이다. 따라서, "치료"란 치료적 처치와, 예방 또는 방지 수단 둘 다를 이른다. "치료"는 또한 일시적 치료를 의미할 수도 있다. 치료를 필요로 하는 대상은 이미 질병이 있는 개체 뿐만 아니라 질병이 예방되어야 할 개체를 포함한다. 따라서, 어떠한 상태, 질병 또는 증상을 "치료하는" 또는 "치료"란 (1) 그러한 상태, 질병 또는 증상에 걸렸을 수 있으나 아직 임상 또는 잠복 증후를 경험하거나 나타내지 않은 인간 또는 기타 포유동물에서 그러한 상태, 질병 또는 증상의 임상 증후의 발현을 방지하거나 지연시키는 것; (2) 그러한 상태, 질병 또는 증상을 억제하는 것, 즉, 질병의 발현, (유지 치료의 경우) 그의 재발, 또는 그의 적어도 하나의 임상 또는 잠복 증후를 저지, 감소 또는 지연시키는 것; 또는 (3) 질병을 완화시키는 것, 즉, 그러한 상태, 질병 또는 증상, 또는 그의 적어도 하나의 임상 또는 잠복 증후의 퇴행을 유발시키는 것을 포함한다. 치료될 개체에 대한 잇점은 통계학적으로 유의하거나, 적어도 환자 또는 의사에 의해 감지될 수 있다.

[0044] "벡터"는 단리된 핵산을 포함하며, 그를 세포의 내부로 전달하는데 사용될 수 있는 조성물이다. 다수의 벡터가 당 기술 분야에 알려져 있으며, 선형 폴리뉴클레오타이드, 이온성 또는 양쪽성 화합물과 결합된 폴리뉴클레오타이드, 플라스미드 및 바이러스를 포함하나, 이에 한정되는 것은 아니다. 따라서, "벡터"는 자체적으로 복제하는 플라스미드 또는 바이러스를 포함한다. 벡터는 또한 핵산의 세포 내로의 전달을 촉진하는 비-플라스미드 및 비-바이러스 화합물, 예를 들어, 폴리리신 화합물, 리포솜 등을 포함하는 것으로 간주되어야 한다. 바이러스 벡터

의 예는 아데노바이러스 벡터, 아데노-연관 바이러스 벡터, 레트로바이러스 벡터 등을 포함하나, 이에 한정되는 것은 아니다.

[0045] 범위: 본 명세서 전체를 통하여 본 발명의 여러 가지 측면이 범위 형식으로 제시될 수 있다. 범위 형식의 기재는 단지 편리성 및 간결성을 목적으로 한 것이므로, 본 발명의 범위에 엄격한 제한을 가하는 것으로 간주되어서는 안된다. 따라서, 범위의 기재는 가능한 모든 하위 범위 뿐만 아니라 그러한 범위 내에 드는 개개의 수치가 구체적으로 개시된 것으로 간주되어야 한다. 예를 들어, 1 내지 6이라는 범위의 기재는, 1 내지 3, 1 내지 4, 1 내지 5, 2 내지 4, 2 내지 6, 3 내지 6 등의 하위 범위 뿐만 아니라, 예컨대, 1, 2, 2.7, 3, 4, 5, 5.3 및 6과 같은, 그러한 범위 내에 드는 개개의 수치가 구체적으로 개시된 것으로 간주되어야 한다. 이는 범위의 폭에 무관하게 적용된다.

[0046] 아미노산 서열이 스위스 프롯 (Swiss Prot.) 또는 진뱅크 (GENBANK) 수탁 번호로 구체적으로 언급된 경우, 그러한 서열은 본 명세서에 인용된다. 수탁번호와 관련된 정보, 예를 들어, 시그널 펩타이드, 세포외 도메인, 막관통 도메인, 프로모터 서열, 번역 개시부 등의 확인 등도 그 전체가 본 명세서에 인용된다.

[0047] CRISPR-뉴클레아제 조성물

[0048] HIV-1 저장소를 퇴치함에 있어서, 특히 LTR을 표적으로 하는 Cas9 기술의 적용은 AIDS를 치료하고, 가능하게는 완치하는 유망한 전략인 것으로 밝혀졌다. 문헌 (Hu, *et al.*, *PNAS* 2014, 111:114616)은 HIV-1 LTR 내의 부위를 표적으로 한 Cas9/gRNA를 발현하는 플라스미드를 사용하여 인간 세포 배양을 안정하게 감염시킨 결과 숙주 세포 기능을 상해하지 않으면서 HIV-1 게놈의 일부 및/또는 전부를 성공적으로 퇴치할 수 있었음을 개시하고 있다. 표적으로 된 부위는 LTR-A, LTR-B, LTR-C 및 LTR-D로 명명되었다. LTR 내의 두 개의 상이한 부위를 표적으로 하는 것은 프로바이러스 DNA 서열의 전부 또는 실질적으로 전부를 절제해 내기에 충분한 길이의 결실을 생성하는데 특히 유효하였다. 세포 내에 이미 존재하는 Cas9/gRNA가 또한 새로운 HIV-1 감염을 방지하였다.

[0049] HIV 및 다른 레트로바이러스는 변이가 매우 잘 일어나므로, 통합된 HIV 게놈을 표적으로 하기 위해 보다 넓은 스펙트럼의 Cas9/gRNA 반응제 및 방법이 필요하다. 특히 유용한 것은 gag 및 pol과 같은 HIV 구조 유전자를 유효하게 표적으로 하는 Cas9/gRNA 반응제일 것이다.

[0050] 따라서, 본 발명의 실시양태는 시험관내 또는 생체내 숙주 세포로부터 고도의 변이성 및/또는 잠복성 바이러스를 치료 및 퇴치하는 조성물 및 방법에 관한 것이다. 본 발명의 방법은 숙주 유기체로부터 그의 유전 물질의 완전성에 간섭하지 않으면서 바이러스 또는 기타 외래 유전물질을 제거하는데 사용될 수 있다. 뉴클레아제는 바이러스 핵산을 표적으로 하여 바이러스 복제 또는 전사에 간섭하거나 숙주 게놈으로부터 바이러스 유전 물질을 절단해 내는데 사용될 수 있다. 뉴클레아제는 특이적으로 특화되어 바이러스 핵산이 숙주 세포 내에 입자로서 존재할 때나 숙주 게놈 내로 통합되어 있을 때 숙주 물질에는 작용하지 않으면서 바이러스 핵산만을 제거할 수 있다. 바이러스 핵산을 표적으로 하는 것은 뉴클레아제에 의해 파괴되도록 바이러스 게놈 물질을 표적으로 하지만 숙주 세포 게놈은 표적으로 하지 않는, 가이드 RNA와 같은 서열-특이적 모이어티를 사용하여 수행될 수 있다. 일부 실시양태에서, CRISPR/Cas 뉴클레아제와 가이드 RNA (gRNA)는 바이러스 게놈 물질을 표적으로 하여, 그를 선택적으로 편집하거나 파괴시키도록 함께 사용된다. CRISPR (주기적 간격으로 분포하는 짧은 회문구조 반복서열)는 박테리아를 과거 감염으로부터 보호하는 박테리아 면역계의 천연 요소이다. 가이드 RNA는 CRISPR/Cas 결합체를 바이러스 표적 서열에 배치시킨다. 결합체의 결합은 Cas 엔도뉴클레아제를 바이러스 게놈 표적 서열에 배치시켜 바이러스 게놈 내에 절단을 일으킨다. 다른 뉴클레아제 시스템이 사용될 수 있으며, 그러한 시스템은, 예를 들어, 징크 핑거 뉴클레아제, 전사 액티베이터-유사 이펙터 뉴클레아제 (TALEN), 메가뉴클레아제, 또는 숙주 유전 물질의 정상적 기능과 간섭하지 않으면서 바이러스 핵산을 분해하거나 그에 간섭하는데 사용될 수 있는 다른 모든 시스템을 포함한다.

[0051] 본 발명에서 실시되는 조성물은 바이러스 핵산을 그가 어떤 형태로 있던지 또는 바이러스 생명 주기의 어떠한 단계에서나 표적으로 하는데 사용될 수 있다. 표적으로 된 바이러스 핵산은 숙주 세포 내에 독립적인 입자로서 존재할 수 있다. 바람직한 실시양태에서, 바이러스 감염은 잠복성이며, 바이러스 핵산은 숙주 게놈 내로 통합되어 있다. 적절한 바이러스 핵산은 어느 것이나 절단과 소화의 표적이 될 수 있다.

[0052] CRISPR/Cas 시스템: CRISPR-Cas 시스템은 CRISPR-관련 뉴클레아제, 예를 들어, Cas9, 및 DNA 가닥 상의 표적 서열, 예를 들어, 포유동물 게놈 내로 통합된 프로바이러스 DNA 내의 표적 서열에 상보적인 가이드 RNA를 포함하는 유전자 편집 복합체를 포함한다. 예시적인 유전자 편집 복합체가 도 1A에 도시되어 있다. 유전자 편집 복합체는 표적 서열 내에서 DNA를 절단할 수 있다. 이러한 절단은 다시 프로바이러스 DNA 내로 다양한 변이의 도

입을 유발시켜 HIV 프로바이러스를 불활성화시킨다. 그러한 변이가 프로바이러스를 불활성화시키는 메카니즘은 다양할 수 있다. 예를 들어, 변이는 프로바이러스 복제 및 바이러스 유전자 발현에 영향을 끼칠 수 있다. 변이는 조절 서열 또는 구조 유전자 서열 내에 존재할 수 있으며, HIV 생산 결함을 유발한다. 변이는 결실을 포함할 수 있다. 결실의 크기는 단일 뉴클레오타이드 염기쌍으로부터 약 10,000개 염기쌍까지 다양할 수 있다. 일부 실시양태에서, 결실은 통합된 레트로바이러스 핵산 서열의 전부 또는 실질적으로 전부를 포함할 수 있다. 일부 실시양태에서, 결실은 통합된 레트로바이러스 핵산 서열 전체를 포함할 수 있다. 변이는 삽입, 즉, 하나 이상의 뉴클레오타이드 염기쌍의 프로바이러스 서열로의 부가를 포함할 수 있다. 삽입된 서열의 크기 또한, 예컨대, 약 하나의 염기쌍으로부터 약 300개 뉴클레오타이드 염기쌍까지 다양할 수 있다. 변이는 점 변이, 즉, 하나의 뉴클레오타이드가 다른 뉴클레오타이드로 치환된 것을 포함할 수 있다. 유용한 점 변이는 기능적 결과를 갖는 것, 예컨대, 아미노산 코돈을 종결 코돈으로 전환시키거나 비기능적 단백질을 생산하게 하는 변이 이다.

[0053] 일반적으로, CRISPR/Cas 단백질은 적어도 하나의 RNA 인식 및/또는 RNA 결합 도메인을 포함한다. RNA 인식 및/또는 RNA 결합 도메인은 가이드 RNA와 상호작용한다. CRISPR/Cas 단백질은 또한 뉴클레아제 도메인 (즉, DN아제 또는 RN아제 도메인), DNA 결합 도메인, 헬리카제 도메인, RN아제 도메인, 단백질-단백질 상호작용 도메인, 이합체화 도메인, 및 기타 도메인을 포함할 수 있다.

[0054] 실시양태에서, CRISPR/Cas-유사 단백질은 야생형 CRISPR/Cas 단백질, 변형 CRISPR/Cas 단백질, 또는 야생형 또는 변형 CRISPR/Cas 단백질의 단편일 수 있다. CRISPR/Cas-유사 단백질은 핵산 결합 친화도 및/또는 특이성을 증가시키고/거나, 효소 활성을 변화시키고/거나 단백질의 다른 특성을 변화시키도록 변형될 수 있다. 예를 들어, CRISPR/Cas-유사 단백질의 뉴클레아제 (즉, DN아제, RN아제) 도메인은 변형되거나, 결실되거나 불활성화될 수 있다. 다른 방법으로는, CRISPR/Cas-유사 단백질은 융합 단백질의 기능에 필수적이지 않은 도메인을 제거하도록 절단(truncated)될 수 있다. CRISPR/Cas-유사 단백질은 또한 융합 단백질의 이펙터 도메인의 활성을 최적화하도록 절단되거나 변형될 수 있다.

[0055] 실시양태에서, CRISPR/Cas 시스템은 타입 I, 타입 II 또는 타입 III 시스템일 수 있다. 적절한 CRISPR/Cas 단백질의 비제한적인 예는 Cas3, Cas4, Cas5, Cas5e (또는 CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (또는 CasA), Cse2 (또는 CasB), Cse3 (또는 CasE), Cse4 (또는 CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4 및 Cu1966을 포함한다.

[0056] 하나의 실시양태에서, RNA-가이드된 엔도뉴클레아제는 타입 II CRISPR/Cas 시스템으로부터 유도된다. 다른 실시양태에서, RNA-가이드된 엔도뉴클레아제는 Cas9 단백질로부터 유도된다. Cas9 단백질은 스트렙토코커스 파이오제네스 (*Streptococcus pyogenes*), 스트렙토코커스 썬모필러스 (*Streptococcus thermophilus*), 스트렙토코커스 종 (*Streptococcus* sp.), 노카르디옵시스 다쎌빌라이 (*Nocardia d'assonvillei*), 스트렙토마이세스 프리스티니스피랄리스 (*Streptomyces pristinaespiralis*), 스트렙토마이세스 비리도크로모제네스 (*Streptomyces viridochromogenes*), 스트렙토마이세스 비리도크로모제네스 (*Streptomyces viridochromogenes*), 스트렙토스포란지움 로제움 (*Streptosporangium roseum*), 알리사이클로바실러스 애시도칼다리우스 (*Alicyclobacillus acidocaldarius*), 바실러스 슈도마이코이데스 (*Bacillus pseudomycoides*), 바실러스 셀레니티레듀센스 (*Bacillus selenitireducens*), 엑시구오박테리움 시비리쿰 (*Exiguobacterium sibiricum*), 락토바실러스 델브뤼키 (*Lactobacillus delbrueckii*), 락토바실러스 살리바리우스 (*Lactobacillus salivarius*), 마이크로실라 마리나 (*Microscilla marina*), 부르크홀데리아레스 박테리움 (*Burkholderiales bacterium*), 폴라로모나스 나프탈레니보란스 (*Polaromonas naphthalenivorans*), 폴라로모나스 종 (*Polaromonas* sp.), 크로코스피라 와트소니 (*Crocospira watsonii*), 사이아노테스 종 (*Cyanotheca* sp.), 마이크로시스티스 에어루지노사 (*Microcystis aeruginosa*), 시네초코커스 종 (*Synechococcus* sp.), 아세토할로비움 아라바티쿰 (*Acetohalobium arabaticum*), 암모니펙스 데젠지 (*Ammonifex degensii*), 칼디셀룰로시루프터 벅사이 (*Caldicelulosiruptor beccsii*), 캔디다투스 디설포루디스 (*Candidatus Desulforudis*), 클로스트리듐 보툴리눔 (*Clostridium botulinum*), 클로스트리듐 디피칠레 (*Clostridium difficile*), 피넨골디아 마그나 (*Finegoldia magna*), 나트란에어로비우스 썬모필러스 (*Natran aerobius thermophilus*), 펠로토마쿨럼 썬모프로피오니쿰 (*Pelotomaculum thermopropionicum*), 애시디티오바실러스 칼더스 (*Acidithiobacillus caldus*), 애시디티오바실러스 페록시단스 (*Acidithiobacillus ferrooxidans*), 알로크로마티움 비노숨 (*Allochromatium vinosum*), 마리노박터 종 (*Marinobacter* sp.), 니트로소코커스 할로필러스 (*Nitrosococcus halophilus*), 니트로소코커스 와트소니 (*Nitrosococcus watsoni*), 슈도알테로모나스 할로플랑크티스 (*Pseudoalteromonas haloplanktis*), 테도노박터 라세미페르 (*Ktedonobacter*

racemifer), 메타노할로비움 이베스티가툼 (*Methanohalobium evestigatum*), 아나배나 배리아빌리스 (*Anabaena variabilis*), 노듈라리아 스푸미제나 (*Nodularia spumigena*), 노스톡 종 (*Nostoc* sp.), 아쓰로피라 맥시마 (*Arthrospira maxima*), 아쓰로피라 플라텐시스 (*Arthrospira platensis*), 아쓰로피라 종 (*Arthrospira* sp.), 링비아 종 (*Lyngbya* sp.), 마이크로콜레우스 츠토노플라스테스 (*Microcoleus chthonoplastes*), 오실라토리아 종 (*Oscillatoria* sp.), 페트로토가 모빌리스 (*Petrotoxa mobilis*), 써모시포 아프리카누스 (*Thermosiphon africanus*) 또는 아카리오크로리스 마리나 (*Acaryochloris marina*)로부터 유래할 수 있다.

[0057] 일부 실시양태에서, CRISPR/Cas-유사 단백질은 야생형 Cas9 단백질 또는 그의 단편으로부터 유도될 수 있다. 다른 실시양태에서, CRISPR/Cas-유사 단백질은 변형된 Cas9 단백질로부터 유도될 수 있다. 예를 들어, Cas9 단백질의 아미노산 서열은 단백질의 하나 이상의 특성 (예를 들어, 뉴클레아제 활성, 친화도, 안정성 등)을 변화시키도록 변형될 수 있다. 다른 방식으로, RNA-가이드된 절단에 관여하지 않는 Cas9 단백질의 도메인이 변형된 Cas9 단백질이 야생형 Cas9 단백질 보다 작아지도록 단백질로부터 제거될 수 있다.

[0058] 예시적인 바람직한 CRISPR-관련 엔도뉴클레아제는 Cas9 뉴클레아제이다. Cas9 뉴클레아제는 야생형 스트렙토코커스 파이오제네스 서열과 동일한 뉴클레오타이드 서열을 가질 수 있다. 일부 실시양태에서, CRISPR-관련 엔도뉴클레아제는 다른 종, 예를 들어, 스트렙토코커스 써모필러스와 같은 다른 스트렙토코커스 종; 슈도모나스 에어루지노사 (*Pseudomonas aeruginosa*), 에스케리치아 콜라이 (*Escherichia coli*), 또는 다른 서열결정된 박테리아 계통 및 고세균, 또는 다른 원핵 미생물로부터의 서열일 수 있다. 또한, 야생형 스트렙토코커스 파이오제네스 Cas9 서열은 변형될 수 있다. 핵산 서열은 포유동물 세포 내에서 효율적 발현에 최적화된, 즉, "인간화"된 코돈일 수 있다. 인간화 Cas9 뉴클레아제 서열은, 예를 들어, 진뱅크 수탁번호 KM099231.1 GI:669193757; KM099232.1 GI:669193761; 또는 KM099233.1 GI:669193765에 기재된 발현 벡터 중 어느 하나에 의해서 코딩된 Cas9 뉴클레아제 서열일 수 있다. 또한, Cas9 뉴클레아제 서열은, 예를 들어, 애드진 (Addgene; Cambridge, MA)의 PX330 또는 PX260과 같은 시판 벡터 내에 함유된 서열일 수 있다. 일부 실시양태에서, Cas9 엔도뉴클레아제는 진뱅크 수탁번호 KM099231.1 GI:669193757; KM099232.1 GI:669193761; 또는 KM099233.1 GI:669193765 중 어느 하나의 Cas9 뉴클레아제 서열의 변이체 또는 단편인 아미노산 서열, 또는 PX330 또는 PX260 (Addgene, Cambridge, MA)의 Cas9 아미노산 서열을 가질 수 있다. Cas9 뉴클레오타이드 서열은 Cas9의 생물학적 활성 변이체를 코딩하도록 변형될 수 있으며, 이들 변이체는, 예를 들어, 하나 이상의 변이 (예를 들어, 부가, 결실 또는 치환 변이, 또는 그러한 변이의 조합)를 함유함으로써 야생형 Cas9와 상이한 아미노산 서열을 가지거나 포함할 수 있다. 치환 변이의 하나 이상은 치환 (예를 들어, 보존적 아미노산 치환)일 수 있다. 예를 들어, Cas9 폴리펩타이드의 생물학적 활성 변이체는 야생형 Cas9 폴리펩타이드와 적어도 또는 약 50%의 서열 동일성 (예를 들어, 적어도 또는 약 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98% 또는 99% 서열 동일성)을 갖는 아미노산 서열을 가질 수 있다. 보존적 아미노산 치환은 다음과 같은 그룹 내의 치환을 포함한다: 글리신 및 알라닌; 발린, 이소류신 및 류신; 아스파르트산 및 글루탐산; 아스파라진, 글루타민, 세린 및 트레오닌; 리신, 히스티딘 및 아르기닌; 및 페닐알라닌 및 타이로신. Cas9 아미노산 서열 중 아미노산 잔기는 비-천연 아미노산 잔기일 수 있다. 천연 아미노산 잔기는 유전자 코드에 의해 자연적으로 코딩되는 아미노산 및 비-표준 아미노산 (예를 들어, L-배치 대신 D-배치를 갖는 아미노산)을 포함한다. 이 펩타이드는 또한 표준 잔기의 변형체인 아미노산 잔기를 포함할 수 있다 (예를 들어, 리신 대신 피롤리신이 사용될 수 있으며, 시스테인 대신 셀레노시스테인이 사용될 수 있음). 비-천연 아미노산 잔기는 자연에서 발견되지 않았지만, 아미노산의 기본적인 구조식에 부합하고, 펩타이드 내에 함유될 수 있는 아미노산이다. 이들의 예는 D-알로이소류신, (2R,3S)-2-아미노-3-메틸펜타노산 및 L-시클로헥실 글리신, (S)-2-아미노-2-시클로헥실 아세트산을 포함한다. 다른 예를 위해서는 관련 서적이나 세계적인 웹 사이트 (캘리포니아 인스티튜트 오브 테크놀로지에 의해 현재 운용되고 있는 사이트는 기능적 단백질 내로 성공적으로 혼입된 비-천연 아미노산의 구조를 개시하고 있음)를 참조할 수 있다.

[0059] Cas9 뉴클레아제 서열은 변이된 서열일 수 있다. 예를 들어, Cas9 뉴클레아제는 가닥 특이적 절단에 관여하는 보존 HNH 및 RuvC 도메인 내에서 변이될 수 있다. 예를 들어, RuvC 촉매 도메인 내에서 아스파르트이트로부터 알라닌으로 (D10A)의 변이는 Cas9 니카제 변이체 (Cas9n)로 하여금 DNA를 절단하기 보다는 틈을 내어 단일 가닥 잘려진 부위를 생성함으로써, 후속되는 HDR을 통한 우선적 복구가 표적-이탈 이중 가닥 잘려진 부위로부터 원하지 않는 인텔 변이의 빈도를 상당히 감소시킬 수 있게 한다.

[0060] 본 발명은 문헌 (Hu, *et al*, *PNAS* 2014, 111:114616)에 개시된 Cas9/gRNA 시스템보다 몇 가지 진보된 점을 포함한다. 실시예에 개시된 실험에서, HIV-1 LTR 내 및 HIV-1 구조 유전자 내 둘 다에서 고도로 특이적인 표적 서열을 추가로 밝혀냈다. 이들 표적 서열 (표적 "부위"라고도 함)은 Cas9/gRNA에 의해 효율적으로 편집됨으로써, 잠복-감염된 포유동물 세포에서 바이러스의 유전자 발현 및 복제를 불활성화시켰다. 이들 추가의 Cas9/gRNA 구

조물 중 일부 및 그들의 조합은 숙주 세포 계놈으로부터 통합된 HIV 프로바이러스 DNA의 전부 또는 일부를 절제해 내는 것으로 밝혀졌다. 하나는 LTR 표적 부위를 향하는 것이고, 다른 하나는 구조 유전자 표적 부위를 향하는 것인 구조물의 쌍은 HIV 계놈의 절제 또는 퇴치를 이루어내는데 특히 효과적인 것으로 밝혀졌다. 이는 LTR 부위 및 구조 유전자 부위를 병행 공격함으로써 사이에 끼어 있는 통합된 HIV DNA의 절제를 이루어낼 수 있다는 최초의 주장이다. 본 발명은 따라서 숙주 세포 내 통합된 HIV DNA를 표적으로 하는데 사용될 수 있는 Cas9/gRNA 조성물의 스펙트럼을 대폭 확대하는 것이다.

[0061] 따라서, 본 발명은 숙주 세포 내로 통합된 프로바이러스 DNA를 불활성화시키는데 사용되는 것으로서, CRISPR-관련 엔도뉴클레아제를 코딩하는 단리된 핵산 서열, 및 HIV 또는 다른 레트로바이러스 내의 표적 서열에 상보적인 하나 이상의 gRNA를 코딩하는 하나 이상의 단리된 핵산 서열을 포함하는 조성물을 특징으로 한다.

[0062] gRNA는 약 20 염기쌍(bp)의 독특한 표적 서열 ("스페이서"라 함)을 함유하는 성숙한 crRNA, 및 프리-crRNA의 리보뉴클레아제 III-보조된 프로세싱에 가이드로서 작용하는 트랜스-활성화 작은 RNA (tracrRNA)를 포함한다. crRNA:tracrRNA 듀플렉스는 crRNA 상의 스페이서와 표적 DNA 상의 상보적인 서열 ("프로토스페이서"라 함) 사이의 상보적인 염기쌍 형성을 통해 Cas9을 표적 DNA로 유도한다. Cas9은 트리뉴클레오타이드 (NGG) 프로토스페이서 인접 모티프 (PAM)를 인식하여 절단 위치 (PAM으로부터 세번째 뉴클레오타이드)를 특정한다. 본 발명에서, crRNA와 tracrRNA는 독립적으로 발현되거나 합성 스템 루프 (AGAAAU)를 통해 천연 crRNA/tracrRNA 듀플렉스를 흉내낸 인공 융합 gRNA로 엔지니어링될 수 있다. 그러한 gRNA는 직접 RNA 형질감염을 위해 합성되거나, 시험관 내 전사되거나, 또는 U6 또는 H1-프로모팅된 RNA 발현 벡터로부터 발현될 수 있다.

[0063] 본 발명의 조성물에서 각 gRNA는 레트로바이러스 내 표적 서열에 상보적인 서열을 포함한다. 표적 레트로바이러스의 예는 HIV이지만, 본 발명의 조성물은 또한 다른 레트로바이러스, 예를 들어, HIV-2 및 원숭이 면역결핍 바이러스 (SIV)-1을 표적으로 하는 경우에도 유용하다.

[0064] 본 발명의 예시적 gRNA의 일부는 HIV의 긴 말단 반복서열 (LTR) 영역 내의 표적 서열에 상보적이다. LTR은 U3, R 및 U5 영역으로 세분된다. HIV-1의 U1, R 및 U5 영역의 배치는 도 3A에 도시되어 있다. LTR은 유전자 발현에 필요한 시그널을 모두 함유하며, 프로바이러스를 숙주 세포의 계놈 내로 통합시키는데 관여한다. 예를 들어, 기본 또는 코어 프로모터, 코어 인핸서 및 조절 영역이 U3 내에서 발견되는 한편, 트랜스활성화 응답 요소는 R 내에서 발견된다. HIV-1 내에서, U5 영역은 몇몇 세부 영역, 예를 들어, 전사 활성화에 관여하는 TAR 또는 트랜스-액팅 응답 요소; 이합체화 및 계놈 패키징에 관여하는 Poly A; PBS 또는 프라이머 결합 부위; Psi 또는 패키징 시그널; DIS 또는 이합체 개시 부위를 포함한다. 따라서, 일부 실시양태에서, gRNA 표적 서열은 HIV 프로바이러스 DNA의 LTR 영역 내의 하나 이상의 표적 서열, 및 HIV 프로바이러스 DNA의 구조 유전자 및/또는 비-구조 유전자 내의 하나 이상의 표적 서열을 포함한다. 다른 실시양태에서, gRNA 표적 서열은 HIV 프로바이러스 DNA의 LTR 영역 내의 하나 이상의 표적 서열, 및 구조 유전자 내의 하나 이상의 표적 서열을 포함한다. 또다른 실시양태에서, gRNA 표적 서열은 HIV 프로바이러스 DNA의 LTR 영역 내의 하나 이상의 표적 서열, 및 HIV 프로바이러스 DNA의 비-구조 유전자 내의 하나 이상의 표적 서열을 포함한다. 또다른 실시양태에서, gRNA 표적 서열은 HIV 프로바이러스 구조 유전자 내의 하나 이상의 표적 서열, 및 HIV 프로바이러스 DNA의 비-구조 유전자 내의 하나 이상의 표적 서열을 포함한다. 또다른 실시양태에서, gRNA 표적 서열은 HIV 프로바이러스 비-코딩 유전자 내의 하나 이상의 표적 서열, 및 HIV 프로바이러스 DNA의 코딩 유전자 내의 하나 이상의 표적 서열을 포함한다. 또다른 실시양태에서, gRNA 표적 핵산 서열은 제1 유전자 내의 하나 이상의 표적 핵산 서열, 및 제2 유전자 내의 하나 이상의 표적 핵산 서열; 또는 제1 유전자 내의 하나 이상의 표적 핵산 서열, 및 제3 유전자 내의 하나 이상의 표적 핵산 서열; 또는 제1 유전자 내의 하나 이상의 표적 핵산 서열, 제2 유전자 내의 하나 이상의 표적 핵산 서열, 및 제3 유전자 내의 하나 이상의 표적 핵산 서열; 또는 제2 유전자 내의 하나 이상의 표적 핵산 서열, 및 제3 또는 제4 유전자 내의 하나 이상의 표적 핵산 서열; 또는 그들의 임의의 조합을 포함한다. 알 수 있는 바와 같이, 표적 핵산 서열의 어떠한 조합이나 사용될 수 있으며, 단지 당 기술 분야 숙련가의 상상에 의해서만 제한될 뿐이다.

[0065] 실시예에 개시된 실험 결과에서, LTR의 U3, R 및 U5 영역 내의 몇몇 서열이 유용한 표적 서열인 것으로 밝혀졌다. 이들 표적 서열에 상보적인 gRNA는 실시예 2의 도 1D 및 3A, 실시예 3의 도 11A, 및 실시예 4의 표 1에 나타나 있다. 이들은 LTR 1, LTR 2, LTR 3, LTR A, LTR B, LTR B', LTR C, LTR D, LTR E, LTR F, LTR G, LTR H, LTR I, LTR J, LTR K, LTR L, LTR M, LTR N, LTR O, LTR P, LTR Q, LTR R, LTR S 및 LTR T를 포함한다. 이들 gRNA의 서열은 도 11A, 12A, 12B 및 12C에 나타나 있다. 본 발명의 조성물은 단지 비제한적으로 이들 예시적 gRNA를 포함하며, HIV LTR 내의 적절한 표적 부위에 상보적인 gRNA는 어느 것이나 포함할 수 있다.

- [0066] 본 발명의 예시적 gRNA의 일부는 HIV의 단백질 코딩 게놈 내의 서열을 표적으로 한다. 구조 단백질 gag를 코딩하는 유전자 내의 서열이 유용한 표적 서열인 것으로 밝혀졌다. 이들 표적 서열에 상보적인 gRNA는 Gag A, Gag B, Gag C 및 Gag D를 포함한다. 그들의 HIV-1 게놈 내의 표적 부위는 도 3A에, 그들의 핵산 서열은 표 1에 나타나 있다. 유용한 표적 서열은 또한 구조 단백질 pol을 코딩하는 유전자 내에서 밝혀졌다. 이들 표적 서열에 상보적인 gRNA는 Pol A 및 Pol B를 포함한다. 그들의 HIV-1 내의 표적 부위는 도 3A에, 그들의 핵산 서열은 표 1에 나타나 있다. HIV-1 외피 단백질 env 내의 표적 부위에 상보적인 gRNA 서열이 또한 표 1에 기재되어 있다.
- [0067] 따라서, 본 발명의 조성물은 비제한적으로 이들 예시적 gRNA를 포함하며, HIV의 단백질 코딩 유전자 내의 적절한 표적 부위 어느 것에나 상보적인 gRNA를 포함할 수 있으며, 이러한 HIV의 단백질 코딩 유전자는 구조 단백질을 코딩하는 tat, 및 부속 단백질을 코딩하는 vif, nef (음성 인자), vpu (바이러스 단백질 U), vpr 및 tcv를 포함하나, 이에 한정되는 것은 아니다.
- [0068] 본 발명에 따른 가이드 RNA 서열은 센스 또는 안티-센스 서열일 수 있다. 가이드 RNA 서열은 일반적으로 프로토-스페이서 인접 모티프 (PAM)를 포함한다. PAM의 서열은 사용되는 CRISPR 엔도뉴클레아제의 특이성 요건에 따라서 달라진다. 스트렙토코커스 파이오제네스로부터 유래하는 CRISPR-Cas 시스템에 있어서, 표적 DNA는 전형적으로 5'-NGG 프로토-스페이서 인접 모티프 (PAM) 바로 앞에 온다. 따라서, 스트렙토코커스 파이오제네스 Cas9에 있어서, PAM 서열은 AGG, TGG, CGG 또는 GGG일 수 있다. 다른 Cas9 이중상동체는 상이한 PAM 특이성을 가질 수 있다. 예를 들어, 스트렙토코커스 써모필러스로부터 유래하는 Cas9은 CRISPR 1에 대하여 5'-NNAGAA를, CRISPR3에 대하여 5'-NGGNG를 요구하고, 나이세리아 메니지디티스 (*Neisseria meningitidis*)는 5'-NNNGATT를 요구한다. 가이드 RNA의 특이적 서열은 서열과는 무관하게 달라질 수 있으나, 유용한 가이드 RNA 서열은 게놈상으로 통합된 레트로바이러스, 예를 들어, HIV의 고효율적이고 완전한 제거를 달성하는 한편 표적-이탈 효과는 최소화하는 서열일 것이다. 가이드 RNA 서열의 길이는 약 20 내지 약 60개 이상의 뉴클레오타이드로 다양할 수 있으며, 예를 들어, 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 60개 이상의 뉴클레오타이드일 수 있다. 유용한 선택 방법은 외래 바이러스 게놈과 내생 레트로바이러스 DNA를 포함하는 숙주 세포 게놈 사이에 극히 낮은 상동성을 갖는 영역을 찾아내는 것이며, 표적-이탈 인간 트랜스크립토 또는 (극히 드물게는) 비번역되는 게놈 위치를 배제하기 위해 12-bp+NGG 표적-선택 기준을 사용하는 생물정보 스크리닝; (숙주 게놈 내에 보존되어 있을 확률이 높은) HIV LTR 프로모터 내의 전사 인자 결합 부위를 피하는 것; 및 WGS, 생거 시퀀싱 및 서베이어 (SURVEYOR) 어세이를 사용하여 유력한 표적 이탈 효과를 찾아내고 배제시키는 것을 포함한다.
- [0069] 가이드 RNA 서열은 단일 서열로 또는 하나 이상의 상이한 서열들의 조합으로, 예컨대, 멀티플렉스 배열로 배치될 수 있다. 멀티플렉스 배열은 2, 3, 4, 5, 6, 7, 8, 9, 10종 또는 그 이상의 상이한 가이드 RNA의 조합을 포함할 수 있다.
- [0070] 실시예 2 및 3에 개시된 실험에서, gRNA의 조합은 멀티플렉스 방식으로, 즉, 동일한 세포 내에서 동시에 발현될 때 특히 효과적인 것으로 밝혀졌다. 많은 경우에, gRNA 조합은 표적 부위 사이에 개재된 HIV 프로바이러스를 절제해 내었다. 절제는 복수의 표적 부위 각각에서 Cas9에 의해 유도된 절단부 사이에 존재하는 서열의 결실로 인한 것이다. 이들 조합 중의 한 예는 하나는 레트로바이러스의 LTR 내의 표적 부위에 상보적이고, 다른 하나는 레트로바이러스의 구조 유전자 내의 표적 부위에 상보적인 gRNA 쌍이다. 효과적인 조합의 예는 Gag D와, LTR 1, LTR 2, LTR 3, LTR A, LTR B, LTR C, LTR D, LTR E, LTR F, LTR G; LTR H, LTR I, LTR J, LTR K, LTR L, LTR M, LTR N, LTR O, LTR P, LTR Q, LTR R, LTR S 및 LTR T 중 하나와의 조합을 포함한다. 효과적인 조합의 예는 또한 LTR 3와, LTR-1, Gag A, Gag B, Gag C, Gag D, Pol A 및 Pol B 중 하나와의 조합을 포함한다.
- [0071] LTR A와 LTR B'의 조합은 또한, 실시예 3에 기재된 바와 같이, HIV-1 게놈의 세그먼트를 절제하였다. 본 발명의 조성물은 이러한 조합에 한정되지 않으며, HIV-1 프로바이러스 내의 둘 이상의 상이한 표적 부위에 상보적인 gRNA의 적절한 조합은 어느 것이나 포함한다.
- [0072] 일부 실시양태에서, 표적 핵산 서열은 레트로바이러스 게놈의 코딩 및 비-코딩 핵산 서열 내의 하나 이상의 핵산 서열을 포함한다. 표적 핵산 서열은 구조 단백질, 비-구조 단백질 또는 그들의 조합을 코딩하는 서열 내에 위치할 수 있다. 구조 단백질을 코딩하는 서열은 Gag, Gag-Pol 전구체, 단백질불해효소 (Pro), 역전사효소 (RT), 삽입효소 (In), Env 또는 그들의 조합을 코딩하는 핵산 서열을 포함한다. 비-구조 단백질을 코딩하는 서열은, 예를 들어, Tat, Rev와 같은 조절 단백질, Nef, Vpr, Vpu, Vif와 같은 부속 단백질 또는 그들의 조합을 코딩하는 핵산 서열을 포함한다.
- [0073] 일부 실시양태에서, gRNA 서열은 Gag, Gag-Pol 전구체, Pro, 역전사효소 (RT), 삽입효소 (In), Env, Tat, Rev,

Nef, Vpr, Vpu, Vif 또는 그들의 조합을 코딩하는 상보적 표적 핵산 서열에 대해 75% 이상의 서열 동일성 갖는다.

- [0074] 일부 실시양태에서, gRNA 서열은 Gag, Gag-Pol 전구체, Pro, 역전사효소 (RT), 삽입효소 (In), Env, Tat, Rev, Nef, Vpr, Vpu, Vif 또는 그들의 조합을 코딩하는 표적 핵산 서열에 상보적이다.
- [0075] 다른 실시양태에서, gRNA 핵산 서열은 서열 번호 1 내지 57, 또는 그들의 조합을 포함하는 서열에 대해 75% 이상의 서열 동일성 갖는다. 다른 실시양태에서, gRNA 핵산 서열은 서열 번호 1 내지 57을 포함한다.
- [0076] 또다른 실시양태에서, 핵산 서열은 서열 번호 1 내지 57, 또는 그들의 조합을 포함하는 서열에 대해 75% 이상의 서열 동일성 갖는 서열을 포함한다. 다른 실시양태에서, 핵산 서열은 서열 번호 1 내지 57에 기재된 서열을 포함한다.
- [0077] 다른 실시양태에서, 레트로바이러스로 잠복 감염된 숙주 세포의 게놈 내로 통합된 레트로바이러스 DNA를 불활성화시키기 위한 조성물은 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 엔도뉴클레아제 및 하나 이상의 가이드 RNA (gRNA)를 코딩하는 단리된 핵산 서열을 포함하며, 이때 gRNA는 통합된 레트로바이러스 DNA 내의 표적 서열에 상보적이며, 레트로바이러스는 인간 면역결핍 바이러스 (HIV)이다. 하나 이상의 gRNA는 적어도 통합된 레트로바이러스 DNA 내의 표적 서열에 상보적인 제1 gRNA; 및 통합된 레트로바이러스 DNA 내의 또다른 표적 서열에 상보적인 제2 gRNA를 포함함으로써, 두 개의 gRNA 사이에 개재된 서열이 제거된다.
- [0078] 일부 실시양태에서, 표적 핵산 서열은 인간 면역결핍 바이러스 (HIV) 프로바이러스 DNA의 긴 말단 반복서열 (LTR) 영역 내의 하나 이상의 표적 및 HIV 통합된 DNA의 구조 및/또는 비-구조 유전자 내의 하나 이상의 표적; 또는 제2 유전자 내의 하나 이상의 표적; 또는, 제1 유전자 내의 하나 이상의 표적 및 제2 유전자 내의 하나 이상의 표적; 또는, 제1 유전자 내의 하나 이상의 표적, 제2 유전자 내의 하나 이상의 표적 및 제3 유전자 내의 하나 이상의 표적; 또는, 제2 유전자 내의 하나 이상의 표적 및 제3 또는 제4 유전자 내의 하나 이상의 표적; 또는 그들의 조합을 포함한다.
- [0079] 또다른 실시양태에서, 시험관내 또는 생체내 레트로바이러스 퇴치를 위한 조성물은 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 엔도뉴클레아제 및 하나 이상의 가이드 RNA (gRNA)를 코딩하는 단리된 핵산 서열을 포함하며, 이때 gRNA는 레트로바이러스 게놈 내의 표적 서열에 상보적이며, 레트로바이러스는 인간 면역결핍 바이러스 (HIV)이다. 실시양태에서, 하나 이상의 gRNA는 적어도 HIV 게놈 내의 표적 서열에 상보적인 제1 gRNA; 및 HIV 게놈 내의 또다른 표적 서열에 상보적인 제2 gRNA를 포함함으로써, 두 개의 gRNA 사이에 개재된 서열이 제거된다.
- [0080] 또다른 실시양태에서, 조성물은 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 엔도뉴클레아제 및 둘 이상의 가이드 RNA (gRNA)를 코딩하는 단리된 핵산 서열을 포함하며, 이때 gRNA는 각각 레트로바이러스 게놈 내의 상이한 표적 서열에 상보적이며, 레트로바이러스는 인간 면역결핍 바이러스 (HIV)이다. 실시양태에서, 둘 이상의 가이드 RNA (gRNA)는 적어도 HIV 게놈 내의 표적 서열에 상보적인 제1 gRNA; 및 HIV 게놈 내의 또다른 표적 서열에 상보적인 제2 gRNA를 포함함으로써, 두 개의 gRNA 사이에 개재된 서열이 제거된다.
- [0081] 일부 실시양태에서, 표적 핵산 서열은 레트로바이러스 게놈의 코딩 및 비-코딩 핵산 서열 내의 하나 이상의 핵산 서열을 포함한다. 표적 핵산 서열은 구조 단백질, 비-구조 단백질 또는 그들의 조합을 코딩하는 서열 내에 위치할 수 있다. 구조 단백질을 코딩하는 서열은 Gag, Gag-Pol 전구체, 단백질분해효소 (Pro), 역전사효소 (RT), 삽입효소 (In), Env 또는 그들의 조합을 코딩하는 핵산 서열을 포함한다. 비-구조 단백질을 코딩하는 서열은, 예를 들어, Tat, Rev와 같은 조절 단백질, Nef, Vpr, Vpu, Vif와 같은 부속 단백질 또는 그들의 조합을 코딩하는 핵산 서열을 포함한다.
- [0082] 일부 실시양태에서, gRNA 서열은 Gag, Gag-Pol 전구체, Pro, 역전사효소 (RT), 삽입효소 (In), Env, Tat, Rev, Nef, Vpr, Vpu, Vif 또는 그들의 조합을 코딩하는 상보적 표적 핵산 서열에 대해 75% 이상의 서열 동일성 갖는다.
- [0083] 일부 실시양태에서, gRNA 서열은 Gag, Gag-Pol 전구체, Pro, 역전사효소 (RT), 삽입효소 (In), Env, Tat, Rev, Nef, Vpr, Vpu, Vif 또는 그들의 조합을 코딩하는 표적 핵산 서열에 상보적이다.
- [0084] 다른 실시양태에서, gRNA 핵산 서열은 서열 번호 1 내지 57, 또는 그들의 조합을 포함하는 서열에 대해 75% 이상의 서열 동일성 갖는다. 다른 실시양태에서, gRNA 핵산 서열은 서열 번호 1 내지 57을 포함한다.
- [0085] 따라서, 본 발명은 레트로바이러스로 잠복 감염된 숙주 세포의 게놈 내로 통합된 레트로바이러스 DNA를 불활성

화시키기 위한 방법으로서, 숙주 세포를 CRISPR-관련 엔도뉴클레아제 및 프로바이러스 DNA 내의 표적 부위에 상보적인 하나 이상의 gRNA를 포함하는 조성물로 처리하는 단계; CRISPR-관련 엔도뉴클레아제 및 하나 이상의 gRNA를 포함하는 유전자 편집 복합체를 발현시키는 단계; 및 프로바이러스 DNA를 불활성화시키는 단계를 포함하는 방법을 포함한다. 상기 언급된 gRNA 및 Cas9 엔도뉴클레아제가 바람직하다. 또다른 바람직한 실시양태에서, 시험관내 또는 생체내 숙주 세포를 처리하는 단계는 둘 이상의 gRNA로 처리하는 것을 포함하며, 둘 이상의 gRNA 각각은 프로바이러스 DNA 내 상이한 표적 핵산 서열에 상보적이다. 특히 바람직한 것은 둘 이상의 gRNA의 조합으로, 적어도 하나의 gRNA는 레트로바이러스의 LTR 내의 표적 부위에 상보적이고, 적어도 하나의 gRNA는 레트로바이러스의 구조 유전자 내의 표적 부위에 상보적인 것인 조합을 포함한다. HIV는 바람직한 레트로바이러스이다.

[0086] 또다른 실시양태에서, 시험관내 또는 생체내 레트로바이러스를 퇴치하기 위한 조성물은, 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 엔도뉴클레아제 및 하나 이상의 가이드 RNA (gRNA)를 코딩하는 단리된 핵산 서열을 포함하고, 이때 gRNA는 레트로바이러스 게놈 내의 표적 서열에 상보적이며, 레트로바이러스는 인간 면역결핍 바이러스 (HIV)이고, gRNA는 적어도 HIV 게놈 내의 표적 서열에 상보적인 제1 gRNA; 및 HIV 게놈 내의 또다른 표적 서열에 상보적인 제2 gRNA를 포함함으로써, 두 개의 gRNA 사이에 개재된 서열이 제거된다. 표적 핵산 서열은 HIV 게놈의 코딩 및 비-코딩 핵산 서열 내의 하나 이상의 핵산 서열을 포함한다. 하나의 실시양태에서, 표적 서열은 긴 말단 반복서열 (LTR) 핵산 서열, 구조 단백질, 비-구조 단백질 또는 그들의 조합을 코딩하는 핵산 서열을 포함하는, HIV 게놈 내의 하나 이상의 핵산 서열을 포함한다. 일부 실시양태에서, 구조 단백질을 코딩하는 서열은 Gag, Gag-Pol 전구체, 단백질분해효소 (Pro), 역전사효소 (RT), 삽입효소 (In), Env 또는 그들의 조합을 코딩하는 핵산 서열을 포함한다. 실시양태에서, 비-구조 단백질을 코딩하는 서열은 조절 단백질, 부속 단백질 또는 그들의 조합을 코딩하는 핵산 서열을 포함한다. 조절 단백질의 예는 Tat, Rev 또는 그들의 조합을 포함한다. 부속 단백질의 예는 Nef, Vpr, Vpu, Vif 또는 그들의 조합을 포함한다. 일부 실시양태에서, gRNA 핵산 서열은 서열 번호 1 내지 57에 대해 75% 이상의 서열 동일성 갖는 핵산을 포함한다. 일부 실시양태에서, gRNA 핵산 서열은 서열 번호 1 내지 57을 포함하는 핵산 서열을 포함한다.

[0087] 일부 실시양태에서, 단리된 핵산 서열은 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 엔도뉴클레아제 및 하나 이상의 가이드 RNA (gRNA)를 코딩하는 단리된 핵산 서열을 포함하며, 이때 gRNA는 레트로바이러스 게놈, 예를 들어, HIV 게놈 내의 표적 핵산 서열에 상보적이다.

[0088] 조성물이 핵산으로서 투여되거나 발현 벡터 내에 함유될 때, CRISPR 엔도뉴클레아제는 가이드 RNA 서열과 같은 핵산 또는 벡터에 의해 코딩될 수 있다. 다른 방식으로, 또는 부가적으로, CRISPR 엔도뉴클레아제는 gRNA 서열과는 물리적으로 분리된 핵산 또는 분리된 개별적 벡터 내에 코딩될 수 있다.

[0089] 변형 또는 변이된 핵산 서열: 일부 실시양태에서, 본 발명에서 실시된 핵산 서열 중 어느 것이나 천연 핵산 서열로부터, 예컨대, 변이, 결실, 치환의 도입, 또는 핵염기, 골격의 변형 등에 의해 변형되거나 유도될 수 있다. 핵산 서열은 벡터, 유전자 편집제, gRNA, tracrRNA 등을 포함한다. 본 발명에 고려될 수 있는 일부 변형된 핵산 서열의 예는, 변형된 골격, 예를 들어, 포스포로티오에이트, 포스포트리에스테르, 메틸 포스포네이트, 단쇄 알킬 또는 시클로알킬 당간 결합 또는 단쇄 헤테로원자 또는 헤테로시클릭 당간 결합을 포함하는 서열을 포함한다. 일부 실시양태에서, 변형된 올리고뉴클레오타이드는 포스포로티오에이트 골격을 갖는 것, 헤테로원자 골격을 갖는 것, $\text{CH}_2\text{--NH--O--CH}_2$, $\text{CH}_2\text{--N(CH}_3\text{)--O--CH}_2$ [메틸렌(메틸이미노) 또는 MMI 골격으로 알려짐], $\text{CH}_2\text{--O--N(CH}_3\text{)--CH}_2$, $\text{CH}_2\text{--N(CH}_3\text{)--N(CH}_3\text{)--CH}_2$ 및 $\text{O--N(CH}_3\text{)--CH}_2\text{--CH}_2$ 골격을 포함하며, 여기서 천연 포스포디에스테르 골격은 O--P--O--CH_2 로 표시된다. 문헌 (De Mesmaeker *et al.*, *Acc. Chem. Res.* 1995, 28:366-374)에 개시된 아미드 골격 또한 본 발명에서 실시된다. 일부 실시양태에서, 핵산 서열은 모르폴리노 골격 구조를 갖거나 (Summerton and Weller, 미국 특허 제5,034,506호); 올리고뉴클레오타이드의 포스포디에스테르 골격이 폴리아미드 골격으로 치환되고, 핵염기가 폴리아미드 골격의 아자 질소 원자에 직접 또는 간접적으로 결합된 것인 펩타이드 핵산 (PNA) 골격을 포함한다 (Nielsen *et al.*, *Science* 1991, 254, 1497). 핵산 서열은 또한 하나 이상의 치환된 당 모이어티를 포함할 수 있다. 핵산 서열은 또한 펜토피라노실 기 대신에 사이클로부틸과 같은 당 미메틱스 (mimetics)를 포함할 수 있다.

[0090] 핵산 서열은, 부가적으로 또는 대체적으로, 핵염기 (종종 간단하게 "염기"로 지칭됨) 변형 또는 치환을 포함할 수 있다. 본 명세서에서, "비변형된" 또는 "천연" 핵염기는 아데닌 (A), 구아닌 (G), 티민 (T), 시토신 (C) 및 우라실 (U)을 포함한다. 변형된 핵염기는 천연 핵산에서는 단지 드물게 또는 순간적으로만 발견되는 핵염기, 예를 들어, 하이포잔틴, 6-메틸아데닌, 5-Me 피리미딘, 특히 5-메틸시토신 (5-메틸-2'-데옥시시토신으로도 칭해지

며, 당 기술 분야에서 종종 5-Me-C로 지칭됨), 5-히드록시메틸시토신 (HMC), 글리코실 HMC 및 젠토바이오실 HMC; 합성 핵염기, 예를 들어, 2-아미노아데닌, 2-(메틸아미노)아데닌, 2-(이미다졸릴알킬)아데닌, 2-(아미노알킬아미노)아데닌 또는 기타 헤테로치환된 알킬아데닌, 2-티오우라실, 2-티오티민, 5-브로모우라실, 5-히드록시메틸우라실, 8-아자구아닌, 7-테아자구아닌, N₆(6-아미노헥실)아데닌 및 2,6-디아미노퓨린을 포함한다. 문헌 (Kornberg, A., DNA Replication, W. H. Freeman & Co., San Francisco, 1980, pp75-77; Gebeyehu, G., *et al.*, *Nucl. Acids Res.* 1987, 15:4513)을 참조할 수 있다. 이노신과 같이 당 기술 분야에 알려진 "유니버설" 염기가 또한 포함될 수 있다. 5-Me-C 치환은 핵산 듀플렉스 안정성을 0.6 내지 1.2배 증가시키는 것으로 밝혀졌다 (Sanghvi, Y. S., in Crooke, S. T. and Lebleu, B., eds., *Antisense Research and Applications*, CRC Press, Boca Raton, 1993, pp. 276-278).

[0091] 본 발명의 핵산 서열의 또다른 변형은 핵산 서열에 올리고뉴클레오타이드의 활성 또는 세포 흡수를 향상시키는 하나 이상의 모이어티 또는 컨주게이트를 화학적으로 결합시키는 것을 포함한다. 그러한 모이어티는 지질 모이어티, 예컨대, 콜레스테롤 모이어티, 콜레스테릴 모이어티 (Letsinger *et al.*, *Proc. Natl. Acad. Sci. USA* 1989, 86, 6553); 콜산 (Manoharan *et al.*, *Bioorg. Med. Chem. Lett.* 1994, 4, 1053); 티오에테르, 예를 들어, 헥실-S-트리틸티올 (Manoharan *et al.*, *Ann. N.Y. Acad. Sci.* 1992, 660, 306; Manoharan *et al.*, *Bioorg. Med. Chem. Lett.* 1993, 3, 2765); 티오콜레스테롤 (Oberhauser *et al.*, *Nucl. Acids Res.* 1992, 20, 533); 지방족 설프, 예를 들어, 도데칸디올 또는 운테실 잔기 (Saison-Behmooaras *et al.*, *EMBO J.* 1991, 10, 111; Kabanov *et al.*, *FEBS Lett.* 1990, 259, 327; Svinarchuk *et al.*, *Biochimie* 1993, 75, 49); 인지질, 예를 들어, 디-헥사테실-rac-글리세롤 또는 트리에틸암모늄 1,2-디-O-헥사테실-rac-글리세로-3-H-포스포네이트 (Manoharan *et al.*, *Tetrahedron Lett.* 1995, 36, 3651; Shea *et al.*, *Nucl. Acids Res.* 1990, 18, 3777); 폴리아민 또는 폴리에틸렌 글리콜 설프 (Manoharan *et al.*, *Nucleosides & Nucleotides* 1995, 14, 969); 또는 아다만탄 아세트산 (Manoharan *et al.*, *Tetrahedron Lett.* 1995, 36, 3651)을 포함하나, 이에 한정되는 것은 아니다.

[0092] 주어진 핵산 서열에서 모든 위치가 균일하게 변형되어야 할 필요는 없으며, 사실상 상기 언급된 하나 이상의 변형이 단일 핵산 서열 또는 핵산 서열 내 단일 뉴클레오타이드에 혼입될 수 있다.

[0093] 일부 실시양태에서, RNA 분자, 예를 들어, crRNA, tracrRNA, gRNA는 하나 이상의 변형된 핵염기를 포함하도록 엔지니어링된다. 예컨대, RNA 분자의 공지된 변형 방법은, 예를 들어, 문헌 (Genes VI, Chapter 9, "Interpreting the Genetic Code", Lewis, ed., 1997, Oxford University Press, New York; 및 Modification and Editing of RNA, Grosjean and Benne, eds., 1998, ASM Press, Washington DC)에서 찾아볼 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-포르밀시티딘; 3-메틸시티딘; 2-티오시티딘; 라이시딘; 2'-O-메틸우리딘; 2-티오우리딘; 2-티오-2'-O-메틸우리딘; 3,2'-O-디메틸우리딘; 3-(3-아미노-3-카르복시프로필)우리딘; 4-티오우리딘; 리보실티민; 5,2'-O-디메틸우리딘; 5-메틸-2-티오우리딘; 5-하이드록시우리딘; 5-메톡시우리딘; 우리딘 5-옥시아세트산; 우리딘 5-옥시아세트산 메틸 에스테르; 5-카르복시메틸우리딘; 5-메톡시카르보닐메틸우리딘; 5-메톡시카르보닐메틸-2'-O-메틸우리딘; 5-메톡시카르보닐메틸-2'-티오우리딘; 5-카르바모일메틸우리딘; 5-카르바모일메틸-2'-O-메틸우리딘; 5-(카르복시하이드록시메틸)우리딘; 5-(카르복시하이드록시메틸)우리딘 메틸 에스테르; 5-아미노메틸-2-티오우리딘; 5-메틸아미노메틸우리딘; 5-메틸아미노메틸-2-티오우리딘; 5-메틸아미노메틸-2-셀레노우리딘; 5-카르복시메틸아미노메틸우리딘; 5-카르복시메틸아미노메틸-2'-O-메틸우리딘; 5-카르복시메틸아미노메틸-2-티오우리딘; 디하이드로우리딘; 디하이드로리보실티민; 2'-메틸아데노신; 2-메틸아데노신; N⁶메틸아데노신; N⁶,N⁶-디메틸아데노신; N⁶,2'-O-트리메틸아데노신; 2-메틸티오-N⁶이소펜테닐아데노신; N⁶-(시스-하이드록시이소펜테닐)-아데노신; 2-메틸티오-N⁶-(cis-하이드록시이소펜테닐)-아데노신; N⁶-글리시닐카르바모일아데노신; 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; 메

틸와이오신; 미변형 하이드록시와이부토신; 와이부토신; 하이드록시와이부토신; 퍼옥시와이부토신; 케오신; 에폭시케오신; 갈락토실-케오신; 만노실-케오신; 7-시아노-7-테아자구아노신; 아라키오신 [7-포름아미도-7-테아자구아노신으로도 불림]; 및 7-아미노메틸-7-테아자구아노신.

[0094] 본 발명의 단리된 핵산 분자는 표준 기술로 제조될 수 있다. 예를 들어, 폴리머라제 연쇄 반응 (PCR) 기술을 사용하여 본 명세서에 기재된 뉴클레오타이드 서열을 함유하는 단리된 핵산을 수득할 수 있다. 각종 PCR 방법, 예를 들어, 문헌 (*PCR Primer: A Laboratory Manual*, Dieffenbach and Dveksler, eds., Cold Spring Harbor Laboratory Press, 1995)에 기재되어 있다. 일반적으로, 관심을 갖고 있는 영역의 말단 또는 그 너머의 서열 정보는 증폭될 주형의 반대쪽 가닥과 서열이 일치하거나 유사한 올리고뉴클레오타이드 프라이머를 설계하는데 사용된다. 부위-특이적 뉴클레오타이드 서열 변형을 주형 핵산 내로 도입할 수 있는 각종 PCR 전략을 또한 이용할 수 있다.

[0095] 단리된 핵산은 또한 단일 핵산 분자 (예를 들어, 포스포르아미다이트 기술을 사용한 3'에서 5' 방향으로의 자동화 DNA 합성으로)로서 또는 일련의 올리고뉴클레오타이드로서 화학적으로 합성될 수 있다. 예를 들어, 목적하는 서열을 함유하는 한 쌍 이상의 긴 올리고뉴클레오타이드 (예를 들어, > 50 내지 100개 뉴클레오타이드)가 합성될 수 있으며, 각각의 쌍은 상보적인 짧은 세그먼트 (예를 들어, 약 15개 뉴클레오타이드)를 함유하므로, 올리고뉴클레오타이드 쌍이 어닐링될 때 듀플렉스가 형성된다. DNA 폴리머라제는 올리고뉴클레오타이드를 신장시켜, 올리고뉴클레오타이드 쌍마다 단일한 이중-가닥 핵산 분자를 생성시켜 벡터 내로 리게이팅될 수 있도록 하는데 사용된다.

[0096] 본 발명은 또한 포유동물 대상체 내에서 통합된 프로바이러스 HIV DNA를 불활성화시키기 위한 제약 조성물을 포함한다. 조성물은 Cas 엔도뉴클레아제를 코딩하는 단리된 핵산 서열, 및 프로바이러스 HIV DNA 내의 표적 서열에 상보적인 하나 이상의 gRNA를 코딩하는 하나 이상의 단리된 핵산 서열을 하나 이상의 발현 벡터 내에 포함된 상태로 포함한다. 바람직한 실시양태에서, 제약 조성물은 제1 gRNA 및 제2 gRNA를 포함하며, 상기 언급한 바와 같이, 제1 gRNA는 HIV LTR 내의 부위를 표적으로 하고, 제2 gRNA는 HIV 구조 유전자 내의 부위를 표적으로 한다.

[0097] 제약 조성물 내에 포함되는 발현 벡터의 예는 플라스미드 벡터 및 렌티바이러스 벡터를 포함하나, 이에 한정되는 것은 아니다. 광범위한 숙주/발현 벡터 조합을 사용하여 본 명세서에 기재된 핵산 서열을 발현시킬 수 있다. 적절한 발현 벡터는, 플라스미드 및 바이러스 벡터, 예컨대, 박테리오파이지, 배클로바이러스 및 레트로바이러스로부터 유래된 것을 포함하나, 이에 한정되는 것은 아니다. 다수의 벡터 및 발현 시스템이 여러 제조원, 예컨대, 노바젠 (Novagen; Madison, WI), 클론텍 (Clontech; Palo Alto, CA), 스트라타젠 (Stratagene; La Jolla, CA), 및 인비트로젠/라이프 테크놀로지스 (Invitrogen/Life Technologies; Carlsbad, CA)로부터 시판되고 있다. 마커 유전자는 숙주 세포에 선택할 수 있는 표현형을 부여할 수 있다. 예를 들어, 마커는 살생물제 내성, 예컨대, 항생제 (예를 들어, 가나마이신, G418, 블레오마이신 또는 하이그로마이신)에 대한 내성을 부여할 수 있다. 발현 벡터는 발현된 폴리펩타이드의 조작 또는 검출 (예를 들어, 정제 또는 위치결정)을 촉진하도록 설계된 태그 서열을 포함할 수 있다. 태그 서열, 예를 들어, 녹색 형광 단백질 (GFP), 글루타티온 S-트랜스퍼라제 (GST), 폴리히스티딘, c-myc, 헤마글루티닌, 또는 FLAGTM 태그 (Kodak, New Haven, CT) 서열은 전형적으로는 코딩된 폴리펩타이드와의 융합체로서 발현된다. 그러한 태그는 카르복실 또는 아미노 말단을 포함하여, 폴리펩타이드 내 어떠한 위치에나 삽입될 수 있다.

[0098] 벡터는 또한 조절 영역을 포함할 수 있다. "조절 영역"이란 전사 또는 번역 개시 및 속도, 전사 또는 번역 생성물의 안정성 및/또는 이동에 영향을 주는 뉴클레오타이드 서열을 이른다. 조절 영역은 프로모터 서열, 인핸서 서열, 응답 요소, 단백질 인식 부위, 가유도성 요소, 단백질 결합 서열, 5' 및 3' 비번역 영역 (UTR), 전사 개시 부위, 종결 서열, 폴리아데닐화 서열, 핵 위치확인 시그널, 인트론 등을 포함하나, 이에 한정되는 것은 아니다.

[0099] 바람직한 경우, 본 발명의 폴리뉴클레오타이드는 또한 양이온성 리포솜 및 아데노바이러스 벡터와 같은 마이크로전달 비히클과 함께 사용될 수 있다. 리포솜 제작, 표적화 및 내용물 전달 과정에 대해서는 문헌 (Mannino and Gould-Fogerite, *BioTechniques*, 6:682 (1988))을 참조할 수 있다. 또한, 문헌 (Felgner and Holm, *Bethesda Res. Lab. Focus*, 11(2):21 (1989); Maurer, R.A., *Bethesda Res. Lab. Focus*, 11(2):25 (1989))을 참조할 수 있다.

[0100] 본 발명의 조성물은 프로바이러스 HIV-1 활성화의 억제, 및 통합된 HIV-1의 부분적 또는 완전한 절제를 일으키고 (실시예 2 및 3), 본 발명은 레트로바이러스, 예를 들어, HIV에 감염된 포유동물 대상체를 치료하는 방법을

제공한다. 본 발명의 방법은 포유동물 대상체가 레트로바이러스에 감염된 것을 판정하는 단계, 유효량의 상기 제약 조성물을 투여하는 단계, 및 포유동물 대상체의 레트로바이러스 감염증을 치료하는 단계를 포함한다.

[0101] 그러한 방법은 통합된 프로바이러스의 문제에 대한 해결, AIDS 및 기타 레트로바이러스 질환의 치료 및 예방에 필수적인 해결책을 제공한다. HIV 감염증 급성기에, HIV 바이러스 입자는 적절한 CD4 수용체 분자를 발현하는 세포에 침투한다. 일단 바이러스가 숙주 세포 내에 진입하면, HIV-코딩된 역전사효소가 HIV RNA의 프로바이러스 DNA 카피를 생성하고, 프로바이러스 DNA는 숙주 세포 게놈 DNA 내로 통합된다. 바로 이런 HIV 프로바이러스가 숙주 세포에 의해 복제되어, 새로운 HIV 비리온을 방출시켜, 이것이 다른 세포를 감염시키도록 할 수 있다.

[0102] 1차 HIV 감염은 수 주 내지 수 개월 동안 잠잠하다가, 전형적으로는 장기의 임상적 "잠복"기로 이어지며, 잠복기는 길게는 10년까지 지속될 수 있다. 이러한 잠복기 동안, 임상 증상이나 말초 혈액 단핵구 중 검출가능한 바이러스 복제는 없을 수 있으며, 말초 혈액에 배양가능한 바이러스 또한 없거나 거의 없다. 그러나, HIV 바이러스는 아주 낮은 수준으로 계속 증식하고 있다. 항-레트로바이러스 요법으로 치료된 대상체에 있어서, 이러한 잠복기는 수 십년 또는 그 이상 지속될 수 있다. 항-레트로바이러스 요법은 낮은 수준의 바이러스 게놈 발현을 억제하지 않을 뿐더러, 휴지기 기억 T 세포, 뇌 매크로파지, 소신경교세포 및 성상 교세포, 장관-관련 림프계 세포와 같은 잠복-감염된 세포를 효율적으로 표적으로 하지 못한다. 본 발명의 조성물은 HIV 프로바이러스를 불활성화시키거나 절제해 낼 수 있으므로, 그러한 조성물을 사용한 치료 방법은 HIV 감염에 대한 새로운 공격로를 개척하는 것이다.

[0103] 본 발명의 조성물은 잠재적 숙주 세포 내에서 안정하게 발현될 때 레트로바이러스, 예컨대, HIV-1에 의한 새로운 감염을 감소시키거나 예방한다 (실시예 3). 따라서, 본 발명은 또한 레트로바이러스 감염, 예를 들어, HIV 감염 위험이 있는 포유동물 대상체 내에서 감염 위험을 감소시키는 치료 방법을 제공한다. 그러한 방법은 포유동물 대상체가 HIV 감염의 위험이 있음을 판정하는 단계, 유효량의 상기 제약 조성물을 투여하는 단계, 및 포유동물 대상체의 HIV 감염 위험을 감소시키는 단계를 포함한다. 바람직하게는, 제약 조성물은 상기 언급한 적어도 하나의 서열의 안정적이고/거나 유도 가능한 발현을 제공하는 벡터를 포함한다.

[0104] 본 발명에 따른 제약 조성물은 당 기술 분야의 숙련가에 알려진 다양한 방법으로 제조될 수 있다. 예를 들어, 상기 핵산과 벡터는 조직 배양 내 세포에 도포하거나 환자 또는 대상체에 투여하기 위한 조성물 내에 배합될 수 있다. 이들 조성물은 제약 기술 분야에 공지된 방법으로 제조될 수 있으며, 부분적 또는 전신적 처치가 바람직한지 여부 및 처치될 부위에 따라서 다양한 경로로 투여될 수 있다. 투여는 국소적 경로 (안 투여, 및 비내, 질 내 및 직장 전달을 포함하는 점막 투여를 포함), 폐 경로 (예를 들어, 네블라이저 등을 사용한 분말 또는 에어로졸의 흡입 또는 취입; 기관지내, 비내, 표피 및 경피 경로), 안구, 경구 또는 비경구 경로를 통할 수 있다. 안구 전달을 하기 위한 방법은 국소 투여 (점안), 결막하, 안주위 또는 유리체내 주사, 별론 카테터에 의한 도입 또는 결막하낭에 외과적으로 심은 안내 삽입물을 포함한다. 비경구 투여는 정맥내, 동맥내, 피하, 복강내 또는 근육내 주사 또는 주입; 또는 두개내, 예를 들어, 경막내 또는 뇌실내 투여를 포함한다. 비경구 투여는 단일 볼러스 투여의 형태일 수 있거나, 또는 연속 주입 펌프에 의할 수도 있다. 국소 투여를 위한 제약 조성물과 제형은 경피 패치, 연고, 로션, 크림, 젤, 액적, 좌약, 스프레이, 액체, 분말 등을 포함할 수 있다. 통상의 제약용 담체, 액상, 분말상 또는 오일상 기재, 증점제 등이 필요하거나 바람직할 수 있다.

[0105] 본 발명은 또한 활성 성분으로서 본 명세서에 기재된 핵산 및 벡터와 함께 1종 이상의 제약상 허용되는 담체를 포함한다. 본 명세서에서 사용되는 "제약상 허용되는" (또는 "약물학상 허용되는")이란 동물 또는 적절하게는 인간에 투여될 때, 역효과, 알리지 또는 기타 부적합 반응을 나타내지 않는 분자체 및 조성물을 이른다. 본 명세서에 사용된 "제약상 허용되는 담체"는 제약상 허용되는 물질에 매질로서 사용될 수 있는 모든 종류의 용매, 분산 매질, 코팅제, 항생제, 삼투압 조절제, 흡수지연제, 완충제, 부형제, 결합제, 윤활제, 젤, 계면활성제 등을 포함한다. 본 발명의 조성물 제조시, 활성 성분은 전형적으로는 부형제와 혼합되거나, 부형제에 의해 희석되거나, 그러한 담체 내에 함유되어, 예컨대, 캡슐, 정제, 사세, 페이퍼 또는 기타 용기의 형태로 된다. 부형제가 희석제로 작용하는 경우, 고체, 반고체 또는 액상 물질 (예를 들어, 생리식염수)일 수 있으며, 활성 성분을 위한 비히클, 담체 또는 매질로서 작용한다. 따라서, 조성물은 정제, 필, 분말, 로젠지, 사세, 카세, 엘릭서, 현탁액, 유액, 용액, 시럽, 에어로졸 (고체로서 또는 액상 매질 중에), 로션, 크림, 연고, 젤, 연질 및 경질 젤라틴 캡슐, 좌제, 멸균 주사 용액, 및 멸균 포장 분말의 형태일 수 있다. 당 기술 분야에 알려져 있는 바와 같이, 희석제의 형태는 의도하는 투여 경로에 따라 달라질 수 있다. 생성된 조성물은 보존제와 같은 추가의 기능제를 함유할 수 있다. 일부 실시양태에서, 담체는 지질-기재 또는 중합체-기재 콜로이드이거나 그를 포함할 수 있다. 일부 실시양태에서, 담체 물질은 리포솜, 하이드로젤, 마이크로입자, 나노입자, 또는 블록 공중합체 미셀의 형태로 제조된 콜로이드일 수 있다. 언급한 바와 같이, 담체 물질은 캡슐을 형성할 수 있으며, 그 물질은 중합체-

기재 콜로이드일 수 있다.

- [0106] 본 발명의 핵산 서열은 대상체의 적절한 세포에 전달될 수 있다. 이는, 예를 들어, 매크로파아지와 같은 식세포의 식작용에 적절한 크기로 된, 중합체성 생분해성 마이크로입자 또는 마이크로캡슐 전달 비히클을 사용하여 달성될 수 있다. 예를 들어, 직경이 대략 1 내지 10 μm 인 PLGA (폴리-락토-코-글리콜라이드) 마이크로입자가 사용될 수 있다. 폴리뉴클레오타이드가 이들 마이크로입자 내에 캡슐화되며, 입자가 식세포에 흡수되어 세포 내에서 점차적으로 분해됨으로써 폴리뉴클레오타이드를 방출시킨다. 일단 방출되면, DNA가 세포 내에서 발현된다. 두번째 형태의 마이크로입자는 세포에 의해 직접 흡수되는 것이 아니라, 생분해를 통하여 마이크로입자로부터 방출될 때에 이르러 핵산이 세포에 의해 흡수되는 핵산의 서방성 저장소로서 주로 작용하도록 한 것이다. 따라서, 이들 중합체성 입자는 식작용을 배제할 수 있도록 충분히 커야한다 (즉, 5 μm 초과, 바람직하게는 20 μm 를 초과하는 크기). 핵산의 흡수를 달성하는 또다른 방법은 표준 방법으로 제조된 리포솜을 사용하는 것이다. 핵산은 이들 전달 비히클 내로 단독으로 함입되거나, 조직-특이적 항체, 예를 들어, HIV 감염에 통상적인 잠복 감염 저장소인 세포 유형, 예를 들어, 뇌 매크로파지, 소신경교세포 및 성상 교세포, 장관-관련 림프계 세포를 표적으로 하는 항체와 함께 함입될 수 있다. 다른 방법으로는, 폴리-L-리신에 정전기력 또는 공유결합력에 의해 부착된 플라즈미드 또는 기타 벡터로 이루어진 분자 결합체를 제조할 수 있다. 폴리-L-리신은 표적 세포 상의 수용체에 결합할 수 있는 리간드에 결합한다. "네이키드 (즉, 전달 비히클이 없는) DNA"의 근육내, 피내 또는 피하 부위로의 전달은 생체내 발현을 달성하기 위한 또다른 수단이다. 관련 폴리뉴클레오타이드 (예를 들어, 발현 벡터) 내에, CRISPR-관련 엔도뉴클레아제 및 가이드 RNA를 코딩하는 서열을 포함하는 단리된 핵산 서열을 코딩하는 핵산 서열은 프로모터 또는 인핸서-프로모터 조합에 작동가능하게 연결되어 있다. 프로모터와 인핸서는 상기 기재되어 있다.
- [0107] 일부 실시양태에서, 본 발명의 조성물은 나노파티클, 예를 들어, DNA와 결합된 고분자량 선형 폴리에틸렌이민 (LPEI)의 코어와, 그를 둘러싸고 있는 폴리에틸렌글리콜-변형 (PEG화) 저분자량 LPEI 외피로 이루어진 나노입자의 형태로 제조될 수 있다.
- [0108] 핵산 및 벡터는 장치 (예를 들어, 카테터)의 표면에 도포되거나, 펌프, 패치 또는 다른 약물 전달 장치 내에 함유될 수 있다. 본 발명의 핵산 및 벡터는 단독으로 또는 제약상 허용되는 부형제 또는 담체 (예를 들어, 생리식염수)의 존재하에 혼합물로 투여될 수 있다. 부형제 또는 담체는 투여 방식 및 경로에 기초하여 선택된다. 제형에 사용되는 적절한 제약용 담체 및 제약용 필요 성분은 관련 기술 분야에 주지된 문헌 (*Remington's Pharmaceutical Sciences*; E. W. Martin), 및 USP/NF (미국 약전 및 국립 규격서)에 기재되어 있다.
- [0109] 일부 실시양태에서, 조성물은 Cas9 또는 변이체 Cas9 및 표적 HIV에 상보적인 하나 이상의 gRNA 서열을 코딩하는 핵산을 캡슐화한 나노입자로 제형화될 수 있거나, 이들 성분을 코딩하는 벡터를 포함할 수 있다. 또한, 조성물은 CRISPR-관련 엔도뉴클레아제 등, 본 발명의 하나 이상의 핵산 조성물에 의해 코딩되는 폴리펩타이드를 캡슐화한 나노입자로 제형화될 수 있다.
- [0110] 조성물은 핵산으로 투여되는지 또는 폴리펩타이드로 투여되는지와 무관하게 포유동물 세포에 의한 흡수를 촉진하는 방식으로 제형화된다. 유용한 벡터 시스템 및 제형은 상기 기재되어 있다. 일부 실시양태에서, 벡터는 조성물을 특정 세포 유형에 전달할 수 있다. 본 발명에 따라서 다른 DNA 전달 방법, 예컨대, 인산칼슘, DEAE 텍스트란, 리포솜, 리포플렉스, 계면활성제, 퍼플루오로 화학약제를 사용한 화학적 형질감염; 전기천공, 마이크로주사, 탄도 입자, 및 "유전자 총" 시스템을 사용한 물리적 전달 방법이 고려될 수 있으나, 이에 한정되는 것은 아니다.
- [0111] 다른 실시양태에서, 조성물은 하나 이상의 Cas/gRNA 벡터로 형질전환 또는 형질감염된 세포를 포함한다. 일부 실시양태에서, 본 발명의 방법은 생체외에서 적용될 수 있다. 즉, 대상체의 세포를 신체로부터 제거하여, 배양 중 조성물로 처리하여, 예컨대, HIV 바이러스 서열을 절제해내고, 처리된 세포를 대상체의 신체로 되돌릴 수 있다. 세포는 대상체의 세포일 수 있거나, 매칭된 하플로타입 또는 세포주일 수 있다. 세포는 복제를 방지하도록 방사선 조사될 수 있다. 일부 실시양태에서, 세포는 인간 백혈병 항원 (HLA)-매칭된 자생 세포주이거나 그 조합일 수 있다. 다른 실시양태에서, 세포는 줄기 세포일 수 있다. 예를 들어, 배아 줄기세포 또는 인공 만능 줄기세포 (유도 만능 줄기세포 (iPS 세포))일 수 있다. 배아 줄기세포 (ES 세포) 및 인공 만능 줄기세포 (유도 만능 줄기세포, iPS 세포)는 인간을 포함한 여러 동물 종에서 확립되었다. 이러한 형태의 만능 줄기세포는 그의 분화를 적절하게 유도함으로써, 만능성을 유지하며 활발하게 분할하는 능력을 보유하면서 거의 모든 기관으로 분화될 수 있으므로, 재생 의학에 가장 유용한 세포 공급원이다. iPS 세포는 특히 자가-유도된 체세포로부터 확립될 수 있으므로, 배아 파괴에 의해 생산되는 ES 세포에 비하여 윤리적, 사회적 문제가 될 여지는 없다. 또한, 자가

-유도된 세포인 iPS 세포는 재생 의학 또는 이식 요법에서 가장 큰 걸림돌이 되는 거부 반응을 피해갈 수 있다.

[0112] 단리된 핵산은 당 기술 분야에 공지된 방법, 예를 들어, siRNA를 전달하는 방법으로 대상체에 쉽게 전달될 수 있다. 일부 측면에서, Cas는 Cas 분자의 활성 도메인이 포함된 단편으로서, 분자 크기를 감소시킨 것일 수 있다. 따라서, Cas9/gRNA 분자는 현재 유전자 치료에 사용되고 있는 방법과 유사하게 임상적으로 사용될 수 있다. 특히, 세포 이식 요법 및 백신 접종을 위한 Cas9/멀티플렉스 gRNA 안정한 발현 줄기세포 또는 iPS 세포가 대상체에 사용될 수 있도록 개발될 수 있다.

[0113] 형질도입된 세포가 재주입을 위해 확립된 방법에 따라 제조된다. 배양 중 약 2 내지 4주 후에, 세포수는 1×10^6 내지 1×10^{10} 일 수 있다. 이와 관련하여, 세포의 성장 특성은 환자마다 또한 세포 유형마다 상이하다. 형질도입된 세포를 재주입하기 약 72시간 전에, 표현형의 분석 및 치료제를 발현하는 세포의 퍼센트를 알기 위하여 분량을 취한다. 투여에 있어서, 본 발명의 세포는 환자의 체중과 전체적인 건강 상태를 고려하여, 세포 유형의 LD₅₀ 및 여러 농도에서 세포 유형의 부작용에 의해 결정된 속도로 투여될 수 있다. 투여는 단일 또는 분할 투여로 달성될 수 있다. 성체 줄기세포는 또한, 예컨대, 골수, 지방 조직 등을 포함할 수 있는 조직이나 공간에서 그의 생산과 배출을 자극하는 외래 투여된 인자를 사용하여 동원될 수 있다.

[0114] 치료 방법

[0115] 일부 실시양태에서, 세포 또는 대상체 중 레트로바이러스 게놈을 퇴치하는 방법은 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 엔도뉴클레아제 및 레트로바이러스 게놈 내 표적 핵산 서열에 상보적인 하나 이상의 가이드 RNA (gRNA)를 코딩하는 단리된 핵산 서열을 치료 유효량으로 포함하는 제약 조성물을 세포와 접촉시키거나 대상체에 투여하는 것을 포함한다.

[0116] 다른 실시양태에서, 세포 또는 대상체 내에서 레트로바이러스의 복제를 억제하는 방법은 주기적 간격으로 분포하는 짧은 회문구조 반복서열 (CRISPR)-관련 엔도뉴클레아제 및 레트로바이러스 게놈 내 표적 핵산 서열에 상보적인 하나 이상의 가이드 RNA (gRNA)를 코딩하는 단리된 핵산 서열을 치료 유효량으로 포함하는 제약 조성물을 세포와 접촉시키거나 대상체에 투여하는 것을 포함한다.

[0117] 레트로바이러스 감염, 예를 들어, HIV 감염의 치료 방법에서, 대상체는 표준 임상 시험, 예컨대, 대상체 혈청 내의 HIV 항체 또는 HIV 폴리펩타이드 p24를 검출하는 면역어세이 또는 HIV 핵산 증폭 어세이를 통해 감염여부가 판정될 수 있다. 대상체에 감염 증후의 완전한 소멸, 감염 증후의 중증도 저하, 또는 감염 진행의 서행의 효과를 제공하는 조성물의 양은 치료 유효량으로 간주된다. 본 발명의 방법은 또한 투여량과 투여 스케줄을 최적화하고 결과를 예측하는데 도움을 주는 모니터링 단계를 포함할 수 있다. 본 발명의 일부 방법에서, 우선 환자에 잠복성 HIV 감염이 있는지 판정하고, 환자를 상기 하나 이상의 조성물로 치료할 지 여부를 결정할 수 있다. 일부 실시양태에서, 그러한 방법은 환자에 보유된 특정 HIV의 핵산 서열을 결정한 다음, 그러한 특정 서열에 상보적인 가이드 RNA를 설계하는 단계를 포함한다. 예를 들어, 대상체의 LTR U3, R 또는 U5 영역, 또는 pol, gag 또는 env 유전자 영역을 결정한 다음, 대상체의 서열에 정확히 상보적인 하나 이상의 gRNA를 설계하거나 선택할 수 있다. 본 발명에 의해 제공되는 신규 gRNA는 효과적인 치료를 수립할 확률을 크게 향상시킨다.

[0118] HIV 감염의 위험을 감소시키는 방법에 있어서, HIV 감염이 생길 위험이 있는 대상은, 예를 들어, 보호되지 않은 성관계를 하는, 즉, 콘돔을 사용하지 않고 성관계를 하는 성적으로 활동적인 개인; 다른 성관계 전달 감염증이 있는 성적으로 활동적인 개인; 정맥 주사 약물 사용자; 또는 포경수술을 하지 않은 남자일 수 있다. HIV 감염이 생길 위험이 있는 대상은, 예를 들어, 직업상 HIV-감염 집단과 접촉할 수 있는 개인, 예컨대, 건강관리 종사자 또는 최초 대처자일 수 있다. HIV 감염이 생길 위험이 있는 대상은, 예를 들어, 교도소 수감자 또는 성매매 종사자, 즉, 수입, 또는 음식, 약물 또는 주거와 같은 돈이 아닌 물품을 위해 성적 활동을 하는 개인일 수 있다.

[0119] 본 발명은 또한 CRISPR-관련 엔도뉴클레아제, 예를 들어, Cas9 엔도뉴클레아제를 코딩하는 단리된 핵산 서열, 및 HIV 프로바이러스 내의 표적 서열에 상보적인 gRNA를 코딩하는 하나 이상의 단리된 핵산 서열을 포함하는 키트를 포함한다. 또한, 단리된 핵산 서열 중 적어도 하나는 벡터, 예컨대, 발현 벡터 내에 코딩될 수 있다. 키트는 HIV 감염의 치료 또는 예방에 사용될 수 있다. 바람직하게는, 키트는 사용 설명, 시린지, 전달 장치, 완충액 용기 및 희석제, 또는 치료 또는 예방에 필요한 기타 기능제를 포함한다. 키트는 또한 목적하는 용도에 적절한 경우 적절한 안정화제, 담체 분자, 향미제 등을 포함할 수 있다.

[0120] 본 발명의 각종 실시양태가 상기 기재되었지만, 이들은 단지 예시적인 것으로 본 발명을 제한하려는 것이 아니며 이해되어야 한다. 본 명세서에 개시된 바에 따라 본 발명의 요지 및 범주를 벗어남이 없이, 기재된 실시양태

에 여러 가지 변형이 이루어질 수 있다. 따라서, 본 발명의 범위 및 범주는 상기 실시양태 중 어느 것에 의해서도 제한되어서는 안된다.

[0121] 본 명세서에 인용된 모든 간행물 및 특허 문헌은 각각의 간행물 또는 특허 문헌이 본 명세서에 개별적으로 기재된 것으로 간주되는 효과를 갖는다. 본 명세서에 여러 참고 문헌이 인용되었다하여, 그로써 어떤 특정 문헌이 본 발명의 "선 행 기술"에 해당함을 시인하는 것은 아니다.

[0122] <실시예>

[0123] 실시예 1: 재료 및 방법

[0124] 렌티바이러스 벡터 내 sgRNA의 클로닝: 가장 좋은 점수의 고효율 및 고품질을 얻기 위해 바이오인포매틱스 spCas9-sgRNA 디자인 툴을 사용하여, HIV-1 LTR-U3 영역 내의 표적 부위에 대해 20개 sgRNA, Gag에 대해 4개 sgRNA, Pol에 대해 2개 sgRNA, Env에 대해 2개 sgRNA를 설계하였다 (표 1). 이들 모든 sgRNA 시드 서열을, sgRNA 렌티바이러스 벡터 (애드진 #50946)로부터 변형된, 변형 sgRNA 발현 pKLV-Wg 렌티바이러스 벡터 내로 클로닝하였다 (도 1B). 간단히 설명해서, pKLV-Wg 벡터를 BbsI로 소화시키고, 안타크틱 (Antarctic) 포스포타제 처리한 다음, 선형화 벡터를 퀵 (Quick) 뉴클레오타이드 제거 키트 (Qiagen)로 정제하였다. 등량의 센스 및 안티센스 가이드 (1 μ l 중 100 μ M)를 폴리뉴클레오타이드 키나제 (PNK, 1 μ l), 1x PNK 완충제 및 1 mM ATP와 혼합하고, 37 $^{\circ}$ C에서 30분 동안 인큐베이션한 후, PCR 기계 (95 $^{\circ}$ C에서 5분, -1 $^{\circ}$ C/사이클로 15초씩 70 사이클)에서 어닐링하였다. 포스포릴화 올리고 듀플렉스 (10 μ M)를 1:100으로 희석하여 작업 용액 (100 nM)을 얻었다. 이어서, 1 μ l 올리고 듀플렉스 (0.1 pmol)를 3.5 μ l BbsI-소화된 pKLV-Wg 벡터 (0.015 pmol), 5 μ l 2x T7 리가제 반응 완충액 및 0.5 μ l T7 DNA 리가제 (NEB)와 혼합하였다. 혼합물을 실온에서 15 내지 30분 동안 인큐베이션하고, 얼음 상에서 냉각시킨 다음, Stab13 적격 세포 내로 형질전환시켰다. 2 내지 4개의 콜로니를 취하여 T351 (U6/5') 및 sgRNA 역방향 프라이머로 PCR시켰다. 2개의 PCR 포지티브 클론을 LB/Amp 배지 중 37 $^{\circ}$ C에서 밤새 성장시켰다. 다음 날, 미니프렙 플라스미드 DNA를 진위즈 인코포레이티드 (Genewiz Inc.)로 보내 T428 (hU6-서열/5'/F)로 시퀀싱하였다. 진리너 (GeneRunner) 프로그램을 사용한 서열분석에 의해, 전체 sgRNA 발현 카세트는 U6 프로모터, sgRNA 및 폴리 T 터미네이터를 포함하는 것으로 입증되었다.

[0125] EcoHIV-루시퍼라제 리포터 어세이: HEK293T 세포 (5 x 10⁴/웰)를 96-웰 플레이트에서, 10% FBS 및 항생제 (100 U/ml 페니실린 및 100 μ g/ml 스트렙토마이신)를 함유하는 고농도 포도당 DMEM 중, 5% CO₂의 습한 대기 하에 37 $^{\circ}$ C에서 배양하였다. 다음 날, 세포를 표준 인산칼슘 침전법으로 EcoHIV-eLuc 리포터 벡터, pLV-Cas9-RFP 벡터, 및 상기 sgRNA 발현 pKLV-Wg 벡터로 동시형질감염시켰다. 형질감염 2일 후, 세포 용해물을 ONE-GloTM 루시퍼라제 어세이 시스템 (프로메가; Promega)을 사용하여 제조하고, 2104 인비전(EnVision)[®] 멀티라벨 리더 (퍼킨엘머; PerkinElmer)로 발광을 측정하였다. 데이터는 4회 독립적인 형질감염의 평균 \pm 표준 오차를 나타낸다. 단일 또는 쌍을 이룬 sgRNA에서의 상대적인 변화를 속이 빈 sgRNA 벡터 대조와 비교하여 계산하였다.

[0126] PCR 제노타이핑, TA 클로닝 및 생거 시퀀싱: 96-웰 플레이트 중의 HEK293T 세포를 EcoHIV-eLuc 리포터, pLV-EF1a-spCas9-T2A-RFP 및 상기 gRNA 발현 벡터로 동시형질감염시켰다. 2일 후, 세포를 90 μ l 50 mM NaOH로 95 $^{\circ}$ C에서 10분 동안 용해시키고, 10 μ l 1M Tris-HCl로 중화시켰다. 조절의 추출물을 PCR에 직접 사용하였으며, 테라 (Terra) PCR 다이렉트 폴리머라제 믹스 (클론텍; Clontech) 및 기재된 PCR 프라이머를 사용하였다. 두 단계의 표준 PCR을 35 사이클 수행하였으며, 68 $^{\circ}$ C에서 어닐링/신장 1분, 98 $^{\circ}$ C에서 15초로 하였다. 생성물을 1.5% 아가로스 겔 중에 분할하였다. 관심있는 밴드를 겔-정제하여, pCRII T-A 벡터 (인비트로젠; Invitrogen) 내로 클로닝하고, 개개 클론의 뉴클레오타이드 서열을 진위즈에서 유니버설 T7 및/또는 SP6 프라이머를 사용하여 시퀀싱하여 결정하였다.

[0127] 실시예 2: 시험관내 HIV-1 퇴치를 위한 HIV-1 gRNA 스크리닝 및 기능 특성화

[0128] HIV 프로바이러스 게놈의 CRISPR-매개된 편집에 유효한 gRNA의 스펙트럼을 넓히기 위하여, 후보 gRNA를 찾아내고, HIV 발현을 억제하는 효과 및 숙주 세포 내 HIV-1 프로바이러스 게놈의 제거 또는 퇴치를 유도하는 능력에 대하여 스크리닝하였다.

[0129] HIV-1 게놈 내의 표적 부위에 특이적인 후보 gRNA를 생물정보학적 과정으로 찾아냈다. 후보 gRNA를 최소의 표적-이탈 가능성, 즉, 숙주 게놈 내의 부위에 손상을 유발할 수 있는 최소의 가능성으로, 효과적인 유전자 편집을 제공할 가능성이 가장 높은 것으로 선택하였다. LTR의 U3 조절 영역 내의 후보 gRNA에 대한 표적 부위 시드 서열은 도 1D에 나타나 있다. gRNA는 LTR 1, LTR 2, LTR 3, LTR A, LTR B, LTR C, LTR D, LTR E, LTR F, LTR G,

LTR H, LTR I, LTR J, LTR K, LTR L, LTR M, LTR N, LTR O, LTR P, LTR Q, LTR R, LTR S 및 LTR T를 포함한다. 이들 gRNA의 서열은 표 1에 나타나 있다. 이들 후보 gRNA의 대부분은 Cas9 니카제 및 RNA 가이드된 FokI 뉴클레아제와 함께 사용하기 위해 쌍을 이룰 수 있으며, 이는 가능한 표적-이탈 효과를 1,500배 감소시킬 수 있다. 구조 유전자 gag 및 pol 내의 후보 gRNA 표적 부위는 도 1C 및 3A에 화살표로 표시되어 있다. 이들 gRNA는 gag A, gag B, gag C, gag D; pol A 및 pol B를 포함한다. 그들의 서열은 표 1에 나타나 있다. 후보 gRNA를 렌티바이러스 리포터 벡터 내로 클로닝하였으며, 도 1B의 하측 패널에 나타난 바와 같다. spCas9를 또한 렌티바이러스 리포터 벡터 내로 클로닝하였으며, 도 1B의 상측 패널에 나타난 바와 같다.

[0130] HEK293T 숙주 세포를 HIV 리포터 구조물인 EcoHIV-eLuc, spCas9에 대한 리포터 발현 구조물, gRNA에 대한 하나 이상의 리포터 발현 구조물로 동시형질감염시켰다. 대조 세포는 대조 구조물 ("LTR 0")을 수용하였다. 2일 후, 세포 용해물 중의 루시페라제 활성을 ONE-GLOTM 루시페라제 어세이 시스템으로 측정하였다.

[0131] 단독으로 투여된 후보 LTR gRNA의 대부분은 루시페라제 발현의 감소에 의해 알 수 있는 바와 같이 (도 2A), 숙주 세포에서 HIV-1의 발현을 억제하는데 유효하였다. 이어서, 이들 gRNA를 gag 또는 pol 유전자 내의 부위를 표적으로 하는 gRNA와의 여러 가지 다양한 조합으로 동시형질감염시켰다. 표 2B 및 2C의 가로축에 나타난 조합들은 Gag D와 LTR 1, LTR 2, LTR 3, LTR A, LTR B, LTR C, LTR D, LTR E, LTR F, LTR G; LTR H, LTR I, LTR J, LTR K, LTR L, LTR M; LTR N, LTR O, LTR P, LTR Q, LTR R, LTR S 및 LTR T 중 하나와의 조합; LTR 3와 Gag A, Gag B, Gag C, Gag D, Pol A 및 Pol B 중 하나와의 조합을 포함한다.

[0132] 단일 gRNA의 대부분은 구성적 루시페라제 활성을 억제하였지만, 일부는 리포터 활성을 증진시키거나 아무런 효과가 없었다 (도 2A). 이 데이터는 LTR 내의 단일 부위 편집이 인텔 변이를 유도하여 HIV-1 LTR 프로모터 활성화에 대한 전사 액티베이터 또는 리프레서의 결합에 영향을 줄 수 있다는 것을 암시한다.

[0133] 그러나, LTR-gRNA 중 어느 하나와 쌍을 이룬 Gag-D의 gRNA는 루시페라제 활성을 64 내지 96% 감소시켰다 (도 2B). 이들 데이터는 설계된 gRNA 모두가 플라스미드 수준에서 표적 부위의 유전자 편집을 유효하게 유도하였다는 것을 암시한다. Gag와 LTR-sgRNA 어느 하나와의 조합은 LTR의 코어 프로모터를 제거함으로써 프로모터 활성의 급격한 감소를 유도할 것이다. 남아 있는 리포터 루시페라제 활성은 spCas9이나 두 sgRNA를 함유하지 않는 EcoHIV 리포터-발현 세포의 소집단을 반영하는 것일 수 있다. 감소 효율은 동일 세포 내 4종의 플라스미드, 즉, EcoHIV-eLuc, spCas9, 두 개의 sgRNA 모두가 존재할 것을 요한다는 점에 유의해야 할 것이다.

[0134] 설계된 Gag 또는 Pol gRNA 중 어느 하나와 쌍을 이룬 LTR-3 gRNA도 또한 루시페라제 리포터 활성을 73 내지 93% 급격하게 감소시켰다 (도 2C). 이들 데이터는 구조 유전자 Gag 또는 Pol 영역을 표적으로 하여 설계된 모든 gRNA가 플라스미드 수준에서 표적 부위의 유전자 편집을 유효하게 유도하였다는 것을 암시한다. LTR-3와 구조 영역 내의 어느 하나의 sgRNA와의 조합은 LTR의 코어 프로모터와 절단 부위 사이의 구조 계층을 제거함으로써 프로모터 활성의 현격한 감소를 유도할 것이다. 그러한 개념 검증 과정은 LTR-sgRNA 뿐만 아니라 Gag, Pol, Env와 같은 구조 영역 또는 양쪽 말단 LTR 사이의 다른 비-구조 영역을 표적으로 하는 다른 sgRNA 어느 것에나 적용될 수 있다. 남아 있는 리포터 루시페라제 활성은 spCas9이나 두 sgRNA 벡터를 함유하지 않는 EcoHIV 리포터-발현 세포의 소집단을 반영하는 것일 수 있다. 감소 효율은 동일 세포 내 4종의 플라스미드, 즉, EcoHIV-eLuc, spCas9, 두 개의 sgRNA 모두가 존재할 것을 요한다는 점에 유의해야 할 것이다.

[0135] 다음으로, HIV-1 발현의 억제가 HIV-1 프로바이러스 게놈 세그먼트의 제거를 반영하는 것인지가 밝혀졌다. HEK293T 세포를 EcoHIV-eLuc 리포터, pLV-EF1a-spCas9-T2A-RFP 및 gRNA 발현 벡터로 동시형질감염시켰다. 2일 후, 세포를 50 mM NaOH로 95 °C에서 10분 동안 용해시키고, 1M Tris-HCl로 중화시켰다. 조직의 추출물을 PCR에 직접 사용하였으며, 테라 (Terra) PCR 다이렉트 폴리머라제 믹스 (클론텍) 및 언급된 PCR 프라이머를 사용하였다.

[0136] 첫번째 세트의 프라이머를 사용하였을 때 (도 3A), 설계된 표적 부위를 절단한 후의 PCR 단편은 20개의 LTR sgRNA 중 17개가 예상된 크기를 나타냈다. 단지 LTR-F, LTR-G 및 LTR-K만 절제를 나타내지 않았으며, 이는 이들이 5'-LTR을 절단할 수 없거나 덜 효율적으로 절단한다는 것을 암시한다. 이는 도 2A에 나타난 LTR-F, LTR-G 및 LTR-K 단일 sgRNA 형질감염에서의 유효하지 않거나 덜 효율적이라는 결과와 일치한다. 유효 sgRNA 중, 상응하는 비절단 단편에 대한 절단 단편의 비율로 결정된 절단 효율은 sgRNA LTR에서 가장 높았다. 흥미롭게도, 예상된 절단 단편보다 큰 크기의 추가의 밴드가 대부분의 쌍에서 발견되었다.

[0137] 두번째 세트의 프라이머를 사용하였을 때 (도 3B), 모든 샘플에서 두 개의 약한 밴드가 관찰되었으며, 이는 비-특이적 PCR 생성물을 의미한다. 그러나, 설계된 표적 부위를 절단한 후의 특징적 PCR 단편은 20개의 LTR sgRNA

중 16개가 예상된 크기를 나타냈다. 유효 sgRNA 중에서, 예상된 PCR 단편의 강도로 결정된 절단 효율은 sgRNA LTR-Q, L, B, S, O, C, I에서 가장 높았다.

- [0138] LTR과 각종 구조 유전자를 표적으로 하는 추가의 sgRNA 조합을 검사하였다. 이 조합들은 도 3C에 나타나 있다. LTR-1, 2, 3 sgRNA 중에서, LTR-1이 최고의 효율을 나타냈다. Gag-A, C, D 및 Pol-A, B는 강한 퇴치 효율을 나타냈다. Gag-B는 절단을 나타내지 않았다 (PCR 반복으로 확인, 우측 패널). 또다시, 예상된 절단 단편보다 큰 크기의 추가의 밴드가 5'LTR 절단을 위한 쌍들에서 발견되었다.
- [0139] LTR U3 영역 내의 표적 부위에 상보적인 gRNA 쌍들이 또한 HIV-1 프로바이러스 게놈 내에 결실을 생성한 것으로 또한 밝혀졌다. 샘플 제조 및 다이렉트 PCR을 전술한 바와 같이 수행하였다. 절단 후의 PCR 단편을 TA 클로닝 및 생거 시퀀싱을 위해 추출하였다 (도 4). 대표적인 TA-클로닝 및 생거 시퀀싱은 LTR-1과 LTR-3 사이에서 296 bp 결실, 및 두 절단 부위 사이에서 180 bp 추가의 삽입을 확인해주었다. 15개 클론의 시퀀싱은 두 절단 부위 사이에서 인텔 변이없이 완벽한 리게이션을 하는 유사한 패턴을 보여주었다. 추가의 180 bp 삽입 서열은 NCBI 블라스트 (Blast)에 의하면 벡터 서열과 매칭된다.
- [0140] 모든 것을 종합한 결과는 후보 gRNA의 대부분이, 루시페라제 리포터 활성화에서 나타난 바와 같이, 선택된 두 표적 부위 사이의 예상된 HIV-1 게놈 서열을 퇴출시키고, 프로바이러스 발현을 억제하는데 유효하다는 것을 보여 준다. 특히, 바이러스 구조 gRNA와 하나 또는 두 개의 LTR gRNA의 조합은 보다 높은 게놈 퇴치 효율을 제공하였다. 이들 실험의 결과는 HIV-1 게놈을 유효하게 절단시키기 위하여 CRISPR 시스템에 사용될 수 있는 gRNA의 스펙트럼을 확대한다.
- [0141] 실시예 3: 신규 gRNA 조합, 및 CRISPR/Cas9 유전자 편집 시스템을 사용하여 잠복 감염된 T 세포로부터 통합된 HIV-1 게놈을 퇴출시키기 위한 벡터 시스템
- [0142] HIV-1 퇴치를 위한 유효 gRNA의 스펙트럼을 확대하고, 숙주 세포로의 gRNA 및 Cas9 전달의 유연성을 증가시키기 위해 실험을 수행하였다.
- [0143] 우선 병합 처리 전략을 시험하였다. 처리는 gRNA LTR B'과 LTR A (문헌 (Hu, *et al.*, 2014)에 개시)의 신규 조합을 사용하였다. LTR A 및 LTR B'의 서열은 도 11A에 나타나 있으며, HIV-1 LTR 내 그들의 위치는 도 5A 내지 5D에 도시되어 있다.
- [0144] Cas9, LTR A 및 LTR B' gRNA의 병합 발현은 잠복 HIV 프로바이러스의 활성화를 폐쇄하고, 프로바이러스 서열을 절제한다: 실험은 도 6A에 도식화된 저칼2D10 리포터 세포주 내에서 실시하였다. 이 세포주는 통합된, 전사단계 잠복 HIV-1 프로바이러스와, 프로바이러스 활성화 리포터로서 Nef 대신에 eGFP를 함유한다. 활성화는 포르볼 12-미리스테이트 13-아세테이트 (PMA) 또는 트리코스타틴 A (TSA)로 처리하여 유도된다. 통합된 HIV-1 리포터 서열은 도 6A에 나타나 있다. 유도된 (우측 패널) 및 비유도된 (좌측 패널) 2D10 세포의 형광현미경 사진이 도 6B에 있다. 유도된 (우측 패널) 및 비유도된 (좌측 패널) 2D10 세포의 유세포 분석이 도 6C에 나타나 있다.
- [0145] 2D10 리포터 세포 (2×10^6 /조건)에 10 μ g의 대조 pX260 플라스미드 또는 각각 5 μ g의 pX260 LTR-A 및 pX260 LTR-B' 플라스미드를 전기천공법으로 도입하였다 (네온 시스템, 인비트로젠, 3회 10ms/1350V 임펄스). 48시간 후, 배지를 푸로마이신 0.5 μ g/ml를 함유하는 배지로 교체하였다. 1주 동안 선택한 후, 푸로마이신을 제거하고, 세포가 다시 1주일 동안 성장하도록 두었다. 세포는 또한 플래그-태깅된 Cas9을 발현하였다.
- [0146] 이어서, 세포를 10 세포/ml 농도로 희석하고, 96웰 플레이트에 50 μ l/웰로 플레이팅하였다. 2주 후, 단세포 클론을 GFP 태깅된 HIV-1 리포터 재활성화에 대해 구아바 이지사이트 미니 (Guava EasyCyte Mini) 유세포 분석기를 사용하여 스크리닝하였다 (12h PMA 25nM/TSA 250nM 처리).
- [0147] Cas9, LTR A 및 LTR B'을 발현하는 클론을 단지 Cas9만을 발현하는 클론과 비교하였다 (도 7A 내지 7C). Cas9, LTR A 및 LTR B'을 발현하는 클론만이 PMA 유도 후 HIV-1 리포터 재활성화를 나타내지 못하는 것으로 밝혀졌다 (도 7B, 하측 패널). 대조적으로, 단지 Cas9만을 발현하는 클론은 EGFP 발현에 의해 입증되는 바와 같이 HIV-1 리포터 재활성화를 보였다 (도 7B, 상측 패널).
- [0148] 다음으로, 잠복 리포터 HIV-1 프로바이러스 재활성화의 폐쇄가 숙주 게놈으로부터 프로바이러스 서열을 성공적으로 절제해낸 것에 기인하는 것인지 여부를 판단하였다. 이전의 실험에서 분석된 클론으로부터 유도된 DNA에 PCR을 수행하여 프로바이러스 *env* 유전자 서열 모티프 RRE, 또는 통합 리포터 프로바이러스를 플랭킹하는 게놈 서열 (MSRB1 유전자)을 증폭시켰다. 프라이머의 T 위치는 도 8A에 나타나 있다.

- [0149] PCR 분석에 따르면 Cas9과 LTR A/B' 조합을 발현하는 클론이 RRE 및 MSRB1을 포함하는 PCR 생성물을 내지 못하는 것으로 나타났으며, 이는 그러한 서열을 포함하는 DNA의 절제를 암시한다. 대조적으로, 단지 Cas9 만을 발현하는 클론에서 RRE 및 MSRB1은 증폭되고 쉽게 검출되었다 (도 8B). 롱 레인지 PCR 제노타이핑은 LTR A/B 조합의 발현이 5' U3 영역과 3' U3 영역 사이에 존재하는 652 bp 서열을 절제하였음을 확인시켰다 (도 8C). 절제는 염색체 16 내 통합 위치로부터 절단 래리어트를 시퀀싱함으로써 추가로 확인되었다 (도 8D). 종합된 결과는 LTR A/B' 및 Cas9 발현에 의한, 잠복 HIV-1 프로바이러스 재활성화의 폐쇄가 숙주 게놈으로부터 프로바이러스 서열의 절제에 기인한 것임을 보여주었다.
- [0150] LTR A/B'과 Cas9의 안정한 발현은 2D10 클론이 HIV-1으로 새로이 감염되는 것을 방지한다는 것이 또한 밝혀졌다. 클론을 Cas9 발현에 대해서는 웨스턴 블롯팅으로 (도 9B), LTR B' 발현에 대해서는 리버스 PCR 분석으로 (도 9C) 특성화하였다. 클론을 HIV-1 NL4-3-EGFP-P2A-NEF 리포터 바이러스로 감염시키고, 감염의 진행에 대해 FACS 분석으로 모니터링하였다. LTR A/B'과 Cas9을 둘 다 발현하는 클론만이 HIV-1 감염에 내성을 나타냈다. 이는 LTR A/B' (ctr17) 또는 Cas9 (AB8)이 결여된 클론과 비교하여, 현저히 낮은 EGFP 발현 수준에 의해 입증된다 (도 9A).
- [0151] Cas9/gRNA의 렌티바이러스를 통한 전달은 프로바이러스 서열을 효율적이고 시간-조절된 방식으로 표적화한다: 다음으로, 렌티바이러스 벡터가 숙주 세포 내에서 Cas9/gRNA 성분을 발현하는데 사용될 수 있는지를 밝혀냈다. 렌티바이러스 벡터는 다양하고 유연한 발현 수단을 제공하며, 다양한 약물로 유도가능한 렌티바이러스 벡터를 이용할 수 있다. 저각 2D10을 RFP-Cas9 (적색 형광) 및/또는 LTR A/B' gRNA (BFP 마커, 청색 형광)를 발현하는 렌티바이러스로 MOI 5로 형질도입시켰다 (도 10A). 72 시간 후, GFP 리포터 바이러스를 PMA/TSA 처리로 재활성화시켰으며, 유세포 분석으로 정량화하였다. 점 플롯 분석이 도 10B에 나타나 있다. HIV 활성화는 유도 시의 점 플롯에서 상향 (녹색) 이동으로 검출할 수 있다 (예를 들어, 도 10B의 좌상측 패널 대 좌하측 패널). Cas9 및 LTR A/B' 둘다로 형질도입된 세포만이 녹색 상향이동을 나타내지 않는 상당량의 세포 분획을 보인다는 것을 알 수 있다 (도 10B, 우하측 패널). 결과는 Cas9/gRNA 성분이 렌티바이러스 벡터에 의해 유효하게 전달될 수 있음을 확인시킨다.
- [0152] Cas9/LTR A/B' 발현은 검출가능한 표적-이탈 효과를 나타내지 않으며, 인접 유전자 발현에 최소의 변화를 나타낸다: CRISPR 편집에 의한 HIV-1 바이러스의 유효한 절제는 표적 서열에 유사한 서열을 함유하는 정상 숙주 유전자 내에서의 유도된 변이를 동반하는 경우라면 거의 쓸모가 없다. LTR A/B'에 대하여 6개의 예상/가능한 표적-이탈 부위를, HIV-1 게놈이 성공적으로 퇴치된 저각 클론 내에서 검사하였다. LTR A 및 LTR B'의 서열은 도 11A에 나타나 있다. 서베이어 (Surveyor) 어레이 반응 (도 11B)에 의해서나 생거 시퀀싱 (도 11C)에 의해서나 인텔 변이는 나타나지 않았다. 염색체 16 내 MSRB1 유전자의 두번째 엑손 내의 HIV-1 리포터 통합 부위 및 이웃 유전자의 위치가 도 11D에 나타나 있다. HIV-1 서열 절제 후 통합 부위에 인접한 유전자의 발현 수준을 qRT-PCR로 측정하고, 대조 세포에서의 발현 수준과 비교하였다. 도 11E에 나타난 결과는 Cas9 및 LTR A/B'으로의 유효한 처리가 절제된 HIV 서열에 이웃한 유전자의 발현에 의미있는 영향을 끼치지 않는다는 것을 보여준다.
- [0153] 실시예 4: AIDS의 치료를 위한, 조절 및 구조 HIV-1 바이러스 게놈을 표적으로 하는 가이드 RNA의 기능적 스크리닝
- [0154] 이 연구에서, HIV-1 LTR 및 바이러스 구조 영역을 표적으로 하는 최상의 gRNA를 찾아내었으며, HIV-1 게놈을 효율적으로 퇴치할 수 있는 gRNA 쌍 형성을 최적화하였다.
- [0155] 생물정보학적 도구를 사용하여 고도의 특이적 gRNA를 설계하였으며, Cas9를 가이딩하여 HIV-1 프로바이러스 DNA를 절단하는 그들의 능력을 고용량 HIV-1 루시퍼라제 리포터 어레이 및 신속 다이렉트-PCR 제노타이핑으로 평가하였다.

표 1

- [0156] <표 1> HIV-1 LTR, Gag 및 Pol을 표적으로 하는 gRNA 올리고뉴클레오타이드 및 PCR 프라이머

| 표적 명칭 | 방향 | 서열 (5' → 3') |
|-------|-----------|--|
| LTR-A | T353: 포워드 | aaacAGGGCCAGGGATCAGATATCCACTGACCTTgt (서열 번호: 1) |
| | T354: 리버스 | taaacAAGGTCAGTGGATATCTGATCCCTGGCCCT (서열 번호: 2) |

| | | |
|-------|-----------|--|
| LTR-B | T355: 포워드 | aaacAGCTCGATGTCAGCAGTTCTTGAAGTACTCgt (서열 번호: 3) |
| | T356: 리버스 | t aaacGAGTACTTCAAGAACTGCTGACATCGAGCT (서열 번호: 4) |
| LTR-C | T357: 포워드 | caccGATTGGCAGAACTACACACC (서열 번호: 5) |
| | T358: 리버스 | aaacGGTGTGTAGTTCTGCCAATC (서열 번호: 6) |
| LTR-D | T359: 포워드 | caccGCGTGGCCTGGGCGGGACTG (서열 번호: 7) |
| | T360: 리버스 | aaacCAGTCCCGCCCAGGCCACGC (서열 번호: 8) |
| LTR-E | T361: 포워드 | caccGATCTGTGGATCTACCACACACA (서열 번호: 9) |
| | T362: 리버스 | aaacTGTGTGTGGTAGATCCACAGATC (서열 번호: 10) |
| LTR-F | T363: 포워드 | caccGCTGCTTATATGCAGCATCTGAG (서열 번호: 11) |
| | T364: 리버스 | aaacCTCAGATGCTGCATATAAGCAGC (서열 번호: 12) |
| LTR-G | T530: 포워드 | caccGTGTGGTAGATCCACAGATCA (서열 번호: 13) |
| | T531: 리버스 | aaacTGATCTGTGGATCTACCACAC (서열 번호: 14) |
| LTR-H | T532: 포워드 | caccGCAGGGAAGTAGCCTTGTGTG (서열 번호: 15) |
| | T533: 리버스 | aaacCACACAAGGCTACTTCCCTGC (서열 번호: 16) |
| LTR-I | T534: 포워드 | caccGATCAGATATCCACTGACCTT (서열 번호: 17) |
| | T535: 리버스 | aaacAAGGTCAGTGGATATCTGATC (서열 번호: 18) |
| LTR-J | T536: 포워드 | caccGCACACTAATACTTCTCCCTC (서열 번호: 19) |
| | T537: 리버스 | aaacGAGGGAGAAGTATTAGTGTGC (서열 번호: 20) |
| LTR-K | T538: 포워드 | caccGCCTCCTAGCATTTCTGTCACA (서열 번호: 21) |
| | T539: 리버스 | aaacTGTGACGAAATGCTAGGAGGC (서열 번호: 22) |
| LTR-L | T540: 포워드 | caccGCATGGCCCAGAGCTGCATC (서열 번호: 23) |
| | T541: 리버스 | aaacGATGCAGCTCTCGGGCCATGC (서열 번호: 24) |
| LTR-M | T542: 포워드 | caccGCAGCAGTCTTTGTAGTACTC (서열 번호: 25) |
| | T543: 리버스 | aaacGAGTACTACAAAGACTGCTGC (서열 번호: 26) |
| LTR-N | T544: 포워드 | caccGCTGACATCGAGCTTTCTACA (서열 번호: 27) |
| | T545: 리버스 | aaacTGTAAGAAAGCTCGATGTCAGC (서열 번호: 28) |
| LTR-O | T546: 포워드 | caccGTCTACAAGGGACTTTCGCT (서열 번호: 29) |
| | T547: 리버스 | aaacAGCGGAAAGTCCCTTGTAGAC (서열 번호: 30) |
| LTR-P | T548: 포워드 | caccGCTTTCCGCTGGGGACTTTCC (서열 번호: 31) |
| | T549: 리버스 | aaacGGAAGTCCCCAGCGGAAAGC (서열 번호: 32) |
| LTR-Q | T687: 포워드 | caccGCCTCCCTGGAAAGTCCCCAG (서열 번호: 33) |
| | T688: 리버스 | aaacCTGGGGACTTTCCAGGGAGGC (서열 번호: 34) |
| LTR-R | T689: 포워드 | caccGCCTGGGCGGGACTGGGGAG (서열 번호: 35) |
| | T690: 리버스 | aaacCTCCCCAGTCCCGCCAGGC (서열 번호: 36) |
| LTR-S | T691: 포워드 | caccGTCCATCCCATGCAGGCTCAC (서열 번호: 37) |
| | T692: 리버스 | aaacGTGAGCCTGCATGGGATGGAC (서열 번호: 38) |
| LTR-T | T548: 포워드 | caccGCGGAGAGAGAAGTATTAGAG (서열 번호: 39) |
| | T549: 리버스 | aaacCTCTAATACTTCTCTCTCCGC (서열 번호: 40) |
| Gag-A | T687: 포워드 | caccGGCCAGATGAGAGAACCAAG (서열 번호: 41) |
| | T688: 리버스 | aaacCTTGGTTCTCTCATCTGGCC (서열 번호: 42) |
| Gag-B | T714: 포워드 | caccGCCTTCCCACAAGGGAAGGCCA (서열 번호: 43) |
| | T715: 리버스 | aaacTGGCCTTCCCTTGTGGGAAGGC (서열 번호: 44) |
| Gag-C | T758: 포워드 | caccGCGAGAGCGTCCGTATTAAGCG (서열 번호: 45) |
| | T759: 리버스 | aaacCGCTTAATACCGACGCTCTCGC (서열 번호: 46) |
| Gag-D | T760: 포워드 | caccGGATAGATGTAAAGACACCA (서열 번호: 47) |
| | T761: 리버스 | aaacTGGTGTCTTTTACATCTATCC (서열 번호: 48) |
| Pol-A | T689: 포워드 | caccGCAGGATATGTAAGTACAGC (서열 번호: 49) |
| | T690: 리버스 | aaacCTGTCAGTTACATATCTCTGC (서열 번호: 50) |
| Pol-B | T716: 포워드 | caccGCATGGGTACCAGCACAAA (서열 번호: 51) |
| | T717: 리버스 | aaacTTGTGTGCTGGTACCCATGC (서열 번호: 52) |
| PCR | T422 | caccGCTTTATTGAGGCTTAAGCAG (서열 번호: 53) |
| | T425 | aaacGAGTCACACAACAGACGGGC (서열 번호: 54) |
| | T645 | TGGAATGCAGTGGCGCGATCTTGGC (서열 번호: 55) |
| | T477 | CACAGCATCAAGAAGAACCTGAT (서열 번호: 56) |
| | T478 | TGAAGATCTCTTGCAGATAGCAG (서열 번호: 57) |

[0157] <결과>

[0158] 고효율 및 낮은 표적-이탈을 나타내는 sgRNA의 생물정보학적 스크리닝: 표적 gRNA의 효율과 특이성은 감염성 질환에 Cas9/gRNA를 응용함에 있어서 매우 중요한 사항이다. spCas9-gRNA 시스템을 위한 gRNA의 설계 및 선택을 위해 몇몇 컴퓨팅 프로그램이 개발되었으며, 20 bp 시드 서열과 NRG PAM이 사용되었다. 대부분의 gRNA 설계 프로그램이 표적-이탈 효과를 예측하기 위하여 개발되었던 반면, 절단 효율을 예측할 수 있는 프로그램은 거의 없었다. 인간 게놈에 대하여 높은 점수의 절단 효율 및 특이성을 갖는, HIV-1 LTR을 표적으로 하는 20개의 gRNA를 (표 1) 다음의 기준을 사용하여 설계하였다: (1) LTR-U3 프로모터의 -18에서 -418 bp 영역을 표적으로 하여 HIV-1 초기 전사를 파괴함 (바이러스 생산을 억제함), 이러한 400 bp 영역은 대부분의 LV에서 배제되어 있으므로 LV 자가-절단을 피함; (2) 높은 상동성으로 인해 숙주 세포 유전자의 발현에 영향을 끼칠 수 있는 전사 인자 결합 부위를 피함 (도 1C); (3) 양쪽 말단의 LTR을 매칭하여 LTR 사이의 프로바이러스 DNA 전체를 제거할 수 있도록 함; (4) 표적-이탈 점수 50% 초과; 및 (5) 이중 spCas9 Nickase 또는 이합체성 RNA-가이드된 FokI 뉴클레아제에의 적용성. HIV-1 전체 게놈을 퇴치하는 최상의 gRNA 조합을 얻는다는 기대에 구조 영역 Gag 및 Pol을 표적으로 하는 몇몇 gRNA를 선택하였다 (도 1C). Env 구조 영역은 다른 균주 사이에서 이러한 구조 서열의 보존성이 보다 낮았으므로 선택하지 않았다. LV 유전자 전달 시스템을 spCas9 및 gRNA 발현을 위해 개별적으로 선택하였으며 (도 1B), 그 이유는 다음과 같다: (1) LV 자체는 형질감염시키기 어려운 HIV-잠복 세포주, 동물 연구 및 잠재적 임상 응용에서 고효율 유전자 요법을 위한 많은 잇점을 제공한다 (Integration-free LV; Hu P, *et al.*, *Mol Ther Methods Clin Dev* 2015, 2:15025; Liu KC, *et al.*, *Curr Gene Ther* 2014, 14:352-364); (2) 별도의 spCas9 LV는 보다 큰 크기의 spCas9 유전자에 대한 양호한 패키징 효율을 보장한다; (3) 별도의 gRNA 발현 LV는 양호한 패키징 효율을 위해 멀티플렉스 gRNA 발현 카세트의 하나의 벡터 내로 클로닝하는데 사용될 수 있다.

[0159] 유효 gRNA를 찾아내기 위한 HEK293T 세포의 기능적 스크리닝: 최상의 표적을 신속히 기능적 스크리닝하기 위하여, 고용량 인비전 멀티플 플레이트 리더를 사용하여 EcoHIV-eLuc 리포터 어세이를 수행하였다. EcoHIV-eLuc 리포터를 선택한 이유는 (1) 그것이 HIV Env를 제외하고 HIV-1 복제에 필요한 모든 성분을 함유하고, (2) Env 결실로 인해 생물안전성 레벨 II 용기에서 취급하기에 편리하며, (3) 생물발광이 형광보다 감수성이 크고, eLuc 리포터는 10개 미만의 단세포를 검출하는데 사용될 수 있기 때문이다 (Song J, *et al.*, *J Gen Virol* 2015, 96:3131-3142). HEK293T 세포주는 높은 형질감염 효율과 함께 비용경제적 인산칼슘 침전법을 사용할 수 있기 때문에 선택하였다. 단일 gRNA 형질감염에 있어서, LTR 프로모터 및 구조 영역을 표적으로 하는 대부분의 gRNA는 EcoHIV 프로바이러스 리포터 생산을 단지 근소하게 감소시킬 수 있었으나, 일부는 프로모터 활성을 증가시키거나 아무런 효과도 나타내지 않은 것으로 밝혀졌다 (도 2A, 2B). 프로모터 활성의 증가는 뉴우런 초기 응답 유전자 프로모터 내의 spCas9/gRNA-유도된 DSB가 그들의 발현을 자극한다는 최근의 보고와도 일치한다 (Madabhushi R, *et al.*, *Cell* 2015, 161:1592-1605). 단일 gRNA는 표적 부위에서 단일 절단을 유도함으로써 표적으로 된 영역 내에서의 인텔 변이 및/또는 양쪽 말단 LTR 사이에서 전체 프로바이러스 DNA의 결실을 생성하는 것으로 가설을 세웠다. 프로모터 내에서의 변이는 전사 액티베이터 및/또는 리프레서의 기능적 활성화에 영향을 줄 수 있으며, 이는 전사 활성의 증가 또는 감소를 일으킬 수 있다. 구조 영역에서의 변이는 HIV-1 구조 단백질의 오픈-리딩 프레임의 이동을 가져올 수 있으며, 따라서 eLuc 리포터의 발현을 감소시킨다.

[0160] 기능적 절단에 유효한 gRNA를 보다 신뢰성 있고 큰 감도로 스크리닝하기 위하여, 쌍을 이룬 gRNA로 동시형질감염시켰다: 각각의 LTR gRNA와 구조 영역을 표적으로 하는 gRNA 중 하나와의 쌍. 이 전략에서, 5' 또는 3' LTR과 구조 영역 사이에서 보다 큰 단편의 결실로 인해 리포터 바이러스의 보다 현격한 감소가 관찰되었다. 도 2B에 나타난 것을 예로 들면, GagD와 LTR-gRNA 중 어느 하나 사이의 모든 조합이 eLuc 발현을 상당히 감소시켰으며 (64 내지 96%), 이는 단일 gRNA를 사용한 경우보다 강력한 것이다. LTR-gRNA의 반 (10/20)은 eLuc 활성화에서 90%를 넘는 감소를 나타냈다. 마찬가지로, 도 2C에 나타난 또다른 예로서, GagA-D 또는 PolA-B 중 어느 하나와 쌍을 이룬 LTR-R gRNA는 루시퍼라제 리포터 활성을 7 내지 23%까지 상당히 감소시켰다. 쌍 형성을 위한 GagD 또는 LTR-R의 선택은 또한 스태필로코커스 오레우스 (*Staphylococcus aureus*) Cas9 시스템에 적용될 수 있는 그들의 PAM 부위, 및 HIV-1 잠복 세포주 (Jadlowsky JK, *et al.*, *Mol Cell Biol* 2014, 34:1911-192) 및 부분적 Gag 및 전체 Pol 서열이 결실된 Tg26 트랜스제닉 마우스 (Kopp JB, *et al.*, *Proc Natl Acad Sci USA* 1992, 89:1577-1581)에 적용될 수 있는 그들의 표적 부위에 기초한 것이다. 이들 데이터는 LTR-gRNA와 구조 gRNA의 조합이 고용량 HIV-1 eLuc 리포터 어세이를 사용하여 유효 gRNA를 스크리닝하는 보다 좋고 간편한 전략이었다는 증거를 제공하였다. 설계된 모든 gRNA는 EcoHIV-eLuc 리포터의 발현을 기능적으로 감소시켰으며, 이는 생물정보학 분석에 의한 높은 점수의 효율 예측과도 일치한다.

[0161] 다이렉트-PCR 제노타이핑을 사용한 유효 gRNA의 확인: 이들 후보 gRNA가 설계된 바와 같은 적절한 표적을 기능적으로 절단하는지를 검증하기 위하여, 쌍을 이룬 gRNA를 함유하는 DNA 샘플 및 상응하는 PCR 프라이머 (도 3D)를 사용하여 다이렉트-PCR 제노타이핑 분석을 수행하였다. 다이렉트-PCR 제노타이핑 방법은 DNA 추출 및 정제를 필요로 하지 않기 때문에 제노타이핑 스크리닝에 보다 편리하다. 구조 영역을 표적으로 하는 gRNA 중 하나를 사용하여 LTR 표적 부위와 쌍을 이루었을 때, PCR 제노타이핑은 5' LTR에서 Gag 사이의 단편의 결실 후 남아 있는 (나머지) 바이러스 LTR 및 Gag 서열로부터 유래한 새로운 단편 (편의상 "결실"로 기재)을 실제 생성하였다 (도 3E, 프라이머 T361/T458). eLuc 리포터 어세이와 일치되게, 거의 모든 gRNA가 크기 (1.3 kb)로 인해 5'-LTR/Gag에 대한 표준 PCR 조건으로 쉽게 증폭될 수 있는 야생형 밴드에서 다양한 정도의 감소를 유도하였다. 절단 후, 설계된 바에 따른 여러 정도의 "결실"이 대부분의 조합에서 관찰되었다 (도 3E). 흥미롭게도, 예상된 "결실"보다 큰 추가의 단편 ("삽입"으로 칭함)이 5'-LTR-Gag 절단시 대부분의 조합에서 관찰되었다 (도 3E). 야생형 밴드 강도의 정량에 의해 LTR-0가 최고 효율, 이어서 I, C, A 순으로 높은 효율을 갖는 것으로 나타났다 (도 3E의 막스). 이 야생형 밴드 절단 효율 패턴은 eLuc 리포터 활성의 감소 패턴과 완전히 연관되어 있는 것은 아니었으며(도 2B), 추측하기로는 믹스 집단 중 PCR 생성물의 증폭이 일반적으로 크기가 작은 생성물을 선호하기 때문이다. 다른 한편, 일부 쌍에서 야생형 밴드의 약한 감소는 단편의 "결실" 또는 "삽입" 없이 gRNA 표적 부위 내에서 다양한 정도의 (수 개의 뉴클레오타이드 이내의) 작은 인델 변이로부터 비롯된 것일 수 있다. 가능한 PCR 우선적 증폭의 영향을 피하기 위해서, PCR 제노타이핑을 사용된 PCR 셋팅에서 증폭되지 않을 7 kb 야생형 PCR 생성물을 생성할 것으로 예상되는 3'-LTR 및 Gag를 커버링하는 프라이머 (도 3E, 3I, 3J)를 사용하여 수행하였다. 단일 PCR 생성물에 의해 검출된 gRNA 사이의 단편 "결실" 패턴 (도 3E, 3I, 3J)은 상대적 비율 변화에 의해 밝혀진 것 (도 3E)과 일치하였다. LTR-K와 GagD의 쌍은 4 세트의 PCR 제노타이핑 반응 모두에서 "결실" 또는 "삽입" 단편 밴드를 나타내지 않았으며 (도 3E, 3F, 3I, 3J), 이는 야생형 밴드에서 단지 7% 감소를 나타내는 것과 관련이 있다 (도 3E). LTR-F와 GagD의 쌍은 1 세트의 PCR 반응에서 약한 "결실" 밴드를 보여주었으며 (3I), 이는 야생형 밴드에서 17% 감소를 나타내는 것과 관련이 있다 (도 3E). LTR-G, P와 GagD의 쌍은 야생형 밴드에서 대략 50% 감소를 나타내었으며, 이는 5'-LTR-Gag (도 3E) 또는 Gag-3'-LTR 절단 (도 3F, 3I, 3J)에서의 "결실"에 기인한다. LTR-R이 사용되어 GagA 내지 D 및 PolA, B 중 어느 하나와 쌍을 이루었을 때, 기재된 상응하는 프라이머로의 PCR 제노타이핑은 또한 매우 약한 편집 능력을 나타낸 Gag-B gRNA (도 3G)를 제외한 시험된 모든 gRNA에서 예상된 새로운 단편 ("결실") (도 3G, 3H) 및 5'-LTR/Gag상의 추가의 삽입 (도 3G)을 다양한 정도로 생성하였다. 그러나, 약하거나 전무한 결실 제노타이핑을 갖는 이들 모든 조합은 여전히 EcoHIV-eLuc 리포터 활성의 현저한 감소를 나타냈다 (도 2B). 이는 단일 gRNA가 작은 인델 (InDel) 변이를 유도하는데 매우 유효한 상태로 남아있는 세포 내로 두 개의 gRNA 플라스미드 중 어느 하나만이 형질감염되거나 둘다 형질감염되지 않았기 때문일 수 있다. 종합된 데이터는 다이렉트-PCR 제노타이핑이 단편 "결실" 및/또는 "삽입"의 존재를 검증하는데 신뢰성있고 신속한 도구를 제공한다는 것을 증명한다. 그러나, 여러 가지 gRNA에 의한 효과적인 HIV-1 퇴치의 평가에는 바이러스 리포터 어세이에 의한 기능 감소 및 5'-LTR 또는 3'-LTR-관련 PCR 제노타이핑에 의한 프로바이러스 DNA 단편 절제의 조합이 필요하다.

[0162] TA-클로닝 및 생거 시퀀싱에 의한 단편 "삽입"/"결실" 변이의 증명: spCas9/gRNA의 절단 효율을 더욱 입증하고, 절단 후 "결실"/"삽입" 변이의 패턴을 검사하기 위하여, TA-클로닝 및 생거 시퀀싱을 위한 3개의 대표적인 PCR 제노타이핑 샘플을 선택하였다. LTR-R/GagA의 쌍을 이룬 발현은 LTR-R과 GagA 표적 부위 사이에서 519-bp 단편의 "결실"을 유발하였다 (도 12A). LTR-L/GagD (도 12B) 및 LTR-M/GagD (도 12C)의 동시 발현은 각 표적 부위 쌍 사이에서 각각 772-bp 또는 763-bp 단편의 "결실"로 이어졌다. 또한, 그러한 동시 발현은 여러가지 크기 및 형태의 작은 인델 변이를 유발하였다. 일부 경우에, 추가 서열 (예를 들어, 159 내지 359 bp)의 큰 "삽입"이 밝혀졌다 (도 3E, 12B, 12C). NCBI 블라스트 분석은 이들 추가의 서열이 내생 숙주 세포 유전자가 아닌 외래 벡터로부터 유래하는 것임을 나타냈다. 이들 결과는 이들 후보 gRNA의 대부분이 절제 또는 삽입/결실에 의한 통합된 HIV 게놈의 표적화된 파괴를 효율적으로 매개할 수 있다는 것을 나타낸다.

[0163] <결론>

[0164] 멀티플렉스 gRNA가 표적 부위 사이에서 큰 단편의 결실을 유도할 수 있다는 개념 (Hu W, *et al.*, *Proc Natl Acad Sci USA* 2014, 111:11461-11466)은 Cas9/gRNA의 DNA 절단 효율을 평가하는 신뢰가능한 방법을 제공한다. 본 연구에서, 이러한 개념 검증 과정은 26종의 gRNA의 각종 멀티플렉스를 스크리닝함으로써 더욱 증명되었다. 설계된 gRNA의 대부분은 두 개의 선택된 표적 부위 사이에서 예상된 HIV-1 게놈 서열을 퇴치하는데 매우 효과적 이어서 HIV-1 리포터 바이러스를 상당 수준 절제해 낸다는 것이 입증되었다. 특히, 바이러스 구조 gRNA와 하나 또는 두 개의 LTR gRNA와의 조합은 훨씬 높은 효율의 게놈 퇴치 및 다이렉트-PCR 제노타이핑 및 고용량 리포터

스크리닝을 사용한 보다 쉬운 접근법을 제공하였다. 본 연구에서 HIV-1 프로바이러스 DNA를 절제하기 위해 선택된 gRNA의 유효성 및 특이성은 Cas9/gRNA 기술을 사용한 임상전 동물 및 임상 환자 연구에서의 성공을 보장하며, 이는 (1) 이들 gRNA가 바이러스 및 비-바이러스 유전자 요법을 개발하기 위해 즉시 사용할 수 있는 선택 소스로서 작용할 수 있고; (2) 개개 HIV-1 환자에 있어서, 이들 gRNA가 HIV-1의 높은 변이율에도 불구하고, 어떠한 HIV-1 단리물에 대해서나 특이적으로 설계된 새로운 gRNA를 스크리닝하는데 마스터로서 사용될 수 있으며; (3) 간편한 gRNA 클로닝, 신속한 리포터 스크리닝 및 신뢰가능한 다이렉트-PCR 제노타이핑이 Cas9/gRNA를 맞춤형 의학에 실제 적용하는 것의 실용성을 제공하기 때문이다.

[0165] 설계된 gRNA 모두가 예상된 표적 부위를 절단하는데 필요한 활성을 나타내지는 않는다. 이제까지 Cas9/gRNA 기술에 의해 유도된 게놈 편집의 효능을 평가하는 몇 가지 접근법이 개발되어 왔다. 지속적으로 발전하고 있는 효율을 예상을 위한 컴퓨터 프로그램은 설계 표적으로서 숙주 세포 게놈을 사용하여 시험되어 왔으나 (Doench JG, *et al.*, *Nat Biotechnol* 2014, 32:1262-1267; Gagnon JA, *et al.*, *PLoS One* 2014, 9:e98186; Liu H, *et al.*, *Bioinformatics* 2015), 감염성 바이러스와 같은 외래 게놈에 적용하기에는 신뢰성이 없을 수 있다. PCR 클로닝을 통한 표적 영역의 생거 시퀀싱은 게놈 편집 효율을 측정하는데 높은 감도와 특이성을 제공하지만 (Sander JD, *et al.*, *Nat Biotechnol* 2011, 29:697-698), 고용량 스크리닝은 노동집약적이다. 미스매치-기반 서베이어 어레이 (Qiu P, *et al.*, *Biotechniques* 2004, 36:702-707; Kim JM, *et al.*, *Nat Commun* 2014, 5:3157; Dahlem TJ, *et al.*, *PLoS Genet* 2012, 8:e1002861) 및 고분해능 용융 분석 (Bassett AR, Liu JL. *J Genet Genomics* 2014, 41:7-19)은 작은 인텔 변이를 검출하는데 감도가 높지만, 불량한 특이성은 그릇된 양성 결과를 생산하기 쉽게 만든다. 제한 단편 길이 다형성 (RFLP) 어레이에는 표적 영역과 함께 제한 효소 부위가 존재하여야 하나, 이는 대부분의 경우에 제한적이다 (Kim JM, *et al.*, *Nat Commun* 2014, 5:3157). 차세대 시퀀싱은 신뢰성 있고 특이적 수단을 제공하나, 고가이고 시간이 많이 든다 (Guell M, *et al.*, *Bioinformatics* 2014, 30:2968-2970). 최근의 몇몇 PCR-기반 어레이는 편집 효능을 정량하는 간편하고 신뢰성 있는 방법을 제공하나, 강력한 프라이머 설계, 미량 분해 또는 모세관 시퀀서를 필요로 한다 (Brinkman EK, *et al.*, *Nucleic Acids Res* 2014, 42:e168; Carrington B, *et al.*, *Nucleic Acids Res* 2015; Yu C, *et al.*, *PLoS One* 2014, 9:e98282). 본 연구에서, 고 감도의 고용량 생물발광 리포터 어레이와 함께 신속한 다이렉트-PCR 제노타이핑을 사용하여 유효 gRNA를 찾아내기 위한 신속하고, 비용-효율적이며, 신뢰성 있는 스크리닝 플랫폼을 확립하였다. 리포터 어레이는 두 gRNA 표적 부위 사이의 큰 단편의 절제, 및 각 gRNA 부위에서의 작은 인텔 변이에 기초한다. 단편의 절제는 프로모터 활성 또는 리포터 발현을 폐쇄하는 한편, 인텔 변이는 프로모터 조절을 변화시키거나 바이러스 단백질의 오픈 리딩 프레임 쉬프트를 유도할 수 있다. 이들 모든 사건은 후속적으로 리포터의 활성에 영향을 미칠 것이다. PCR 제노타이핑은 단편 절단 및 남아있는 말단 DNA 사이의 효율적인 재-연결을 기초로 한다. 재-연결된 PCR 단편의 존재는 양쪽 gRNA의 효율에 대한 긍정적인 증거를 제공한다. 재-연결 효율은 세포 분할에 따라 달라지므로, PCR 제노타이핑은 비-분할 세포의 경우에 한정될 수 있다. 또한, 일부 프라이머에 대한 PCR 조건은 최고의 제노타이핑 효율을 달성하기 위해 최적화할 필요가 있다.

[0166] 본 연구의 목적은 신뢰성 있고 감도가 좋은 고용량 어레이를 수립함으로써 유효한 gRNA를 스크리닝하고 찾아내는 것이었다. EcoHIV-eLuc 리포터를 HEK293T 세포 내에 일시적 형질감염시키는 것을 테스트 플랫폼으로 선택하였으며, 이는 spCas9/gRNA 성분에 비하여 소량의 리포터 플라스미드 (1:20)로도 모든 리포터-발현 세포에서 표적 절단을 보장할 수 있으므로 루시퍼라제 리포터 어레이 및 PCR 제노타이핑의 검출 효율을 최대화할 수 있기 때문이다. 대조적으로, HIV-1 잠복의 실제 상황에 보다 유사할 수 있는, HEK293T 세포 기반의 EcoHIV-eLuc 안정한 세포주 (도 13A 내지 13E, 14A 내지 14D)는 루시퍼라제 리포터 어레이 및 PCR 제노타이핑 모두에서 불량한 검출 감도를 나타냈다. 이는 형질감염 효율이 80 내지 90%로 높은 경우라도, gRNA 플라스미드가 없는 EcoHIV-eLuc-발현 세포는 형질감염 후에 항상 존재하므로 eLuc 리포터는 계속적으로 발현되기 때문이다. 일시적 리포터 형질감염의 또다른 잇점은 간편한 셋업, 비용-효율적 형질감염 및 고용량 발광 측정을 포함한다. 중요한 것은, 밝혀진 gRNA가 실제로 HIV 잠복-감염된 세포 또는 세포주 내에서도 여전히 유효하여, 동물 연구 및 임상 응용에 사용될 수 있다는 것이다. 일시적 형질감염된 EcoHIV-eLuc 리포터 (에피솜 DNA)가 숙주 게놈 (핵) 내의 잠복 HIV 프로바이러스 DNA를 반영하지는 않으나, spCas9/gRNA-매개 유전자 편집은 HIV 프로바이러스 (Hu W, *et al.*, *Proc Natl Acad Sci USA* 2014, 111:11461-11466) 및 다른 바이러스 (Ramanan V, *et al.*, *Sci Rep* 2015, 5:10833; Yuen KS, *et al.*, *J Gen Virol* 2015, 96:626-636)의 에피솜 및 핵 DNA 사이에 유사한 효율로 작동한다. 또한, 통합된 HIV-1 프로바이러스 DNA에 더하여 에피솜 DNA의 유효한 절단은 HIV (Hu W, *et al.*, *Proc Natl Acad Sci USA* 2014, 111:11461-11466) 및 다른 감염성 바이러스 (Peng C, Lu M, Yang D., *Virol Sin* 2015, 30:317-325)의 새로운 감염에 대한 신규 예방적 치료법이 될 수 있을 것이다.

[0167] 몇몇 혼재하는 요인이 상이한 gRNA의 비교 분석에서 일시적 형질감염 효율과 전환유전자 발현에 영향을 끼칠 수

있다. 이를 최소화하기 위하여, 몇 가지 주의를 하였다: 1) 리포터 및 spCas9 플라스미드의 마스터 혼합물을 제조하여 각 군의 gRNA에 동일한 양의 이들 공유 플라스미드가 혼합되도록 하였고; 2) 레닐라 (*Renilla*) 루시페라제 리포터 (1:100)를 형질감염 효율의 표준화에 사용하였으며; 3) 대규모 형질감염을 96-웰 플레이트 중에서 모든 gRNA에 대해 동시에 4 내지 6회 중복으로 수행하였고; 4) 모든 데이터는 각 실험에서 속이 빈 gRNA 대조와 비교한 상대적 변화로서 표현하였다.

[0168] LTR 영역을 표적으로 하는 gRNA가 양쪽 말단 LTR의 절단에 의해 전체 프로바이러스 DNA를 제거할 수 있으나, EcoHIV-eLuc 리포터 어세이에 의해 나타난 바와 같이, 퇴치 효율은 나타나지 않았다. 또한, LTR을 커버링하는 프라이머를 사용한 표준 PCR은 두 LTR 표적 부위 사이에서 단편을 제거한 후에는 5'-LTR과 3'-LTR을 구별할 수 없으므로, 전체 HIV-1 프로바이러스 DNA의 퇴치를 입증하기 위해서는 롱-레인지 PCR이 필요하다 (Hu W, *et al.*, *Proc Natl Acad Sci USA* 2014, 111:11461-11466). LTR 영역을 표적으로 하는 두 개의 gRNA는 각 LTR 영역 내에 단편 절단을 유도하여 LTR 프로모터 활성을 억제하고 HIV-1 RNA 안정성을 감소시킬 것이므로, 이전에 입증되었던 바와 같이 전체 퇴치 효율을 향상시킨다 (Hu W, *et al.*, 2014). 본 연구에서, LTR과 구조 영역 사이의 한 쌍의 gRNA는 어느 것이나 HIV-1 퇴치 효율을 평가하는 보다 효과적인 방법을 제공한다는 원리의 새로운 검증을 시험하였다. 이 방법에 의해, HIV-1 리포터 바이러스 생산에 있어서의 현격한 기능 감소가 5'LTR+Gag, Gag+3'LTR 및 양쪽 말단 LTR의 세 개의 가능한 절단으로부터 비롯되었으며, 민감하고 고용량인 생물발광 리포터 어세이로 쉽게 모니터링될 수 있다. 이들 절단은 LTR 및 구조 영역을 커버링하는 프라이머를 사용한 표준 및 신속 다이렉트 PCR-제노타이핑에 의해 효율적이고 신뢰도 있게 검출될 수 있다. 마찬가지로, 두 개의 LTR gRNA와 하나 또는 두 개의 구조 gRNA의 각테일은 임상전 및 임상 셋팅에서 HIV-1 게놈을 퇴치하는 적절하고 경제적인 치료법을 제공한다.

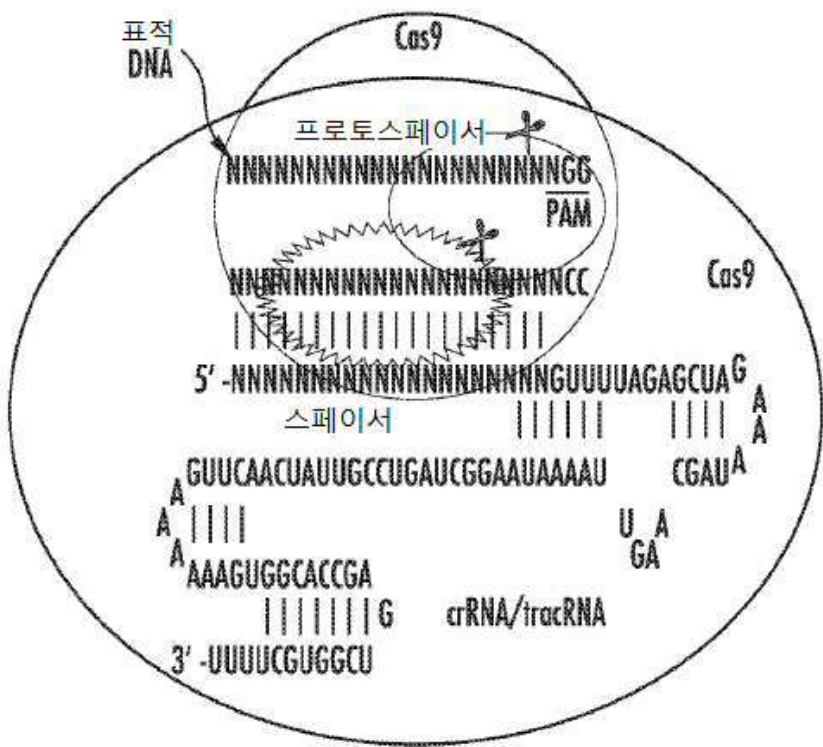
[0169] Cas9/gRNA 기술과 관련하여 표적-이탈 효과의 가능성은 게놈 편집 분야에 있어서 큰 문제점이 되어 왔다. 엄격한 gRNA 설계, 기능적 스크리닝 및 Cas9 기술 수정은 게놈 편집의 특이성을 증가시키기 위해 개발되어 왔다. 배양된 세포에서 spCas9/gRNA 관련된 표적-이탈 효과는 전체 게놈 시퀀싱 (WGS)에 의해 아주 드물게 입증되어 왔다 (Hu W, *et al.*, 2014; Zuckermann M, *et al.*, *Nat Commun* 2015, 6:7391; Smith C, *et al.*, *Cell Stem Cell* 2014, 15:12-13; Veres A, *et al.*, *Cell Stem Cell* 2014, 15:27-30; Yang L, *et al.*, *Nat Commun* 2014, 5:5507). 새로 개발된 비편향 프로파일링 기술은 이러한 spCas9-gRNA 시스템의 고도의 특이성을 더욱 입증하였다 (Ran FA, *et al.*, *Nature* 2015, 520:186-191; Tsai SQ, *et al.*, *Nat Biotechnol* 2015, 33:187-197; Frock RL, *et al.*, *Nat Biotechnol* 2015, 33:179-186). 생체내 표적 이탈은 후성적 보호에 의해 훨씬 낮을 것으로 예상된다. 본 연구에서, 높은 점수의 효율 및 특이성에 대해 숙주 게놈에 대비하여 외래 바이러스 DNA를 분석하였다. gRNA 스크리닝 중에 세포 독성은 관찰되지 않았다. 이중의 spCas9 Nickase 및 RNA-가이드된 FokI 뉴클레아제가 잠재적인 표적-이탈 효과를 최대 1500-배 감소시키는 것으로 밝혀진 바 있다 (Ran FA, *et al.*, *Cell* 2013, 154:1380-1389; Mali P, *et al.*, *Nat Biotechnol* 2013, 31:833-838; Wyvekens N, *et al.*, *Hum Gene Ther* 2015, 26:425-431; Tsai SQ, *et al.*, *Nat Biotechnol* 2014, 32:569-576).

[0170] 결론적으로, 설계된 gRNA의 대부분은 선택된 두 표적 부위 사이에서 예상된 HIV-1 게놈 서열을 퇴치하고, eLuc 리포터 활성화에 영향을 미치는데 매우 효과적이다. 특히, 바이러스 구조 gRNA와 하나 또는 두 개의 LTR gRNA의 조합은 보다 높은 게놈 퇴치 효율 및 PCR 제노타이핑을 위한 보다 쉬운 방법을 제공한다. HIV-1 eLuc 리포터 어세이 및 다이렉트-PCR 제노타이핑을 사용한 스크리닝은 유효 HIV-1 gRNA를 스크리닝하는데 신뢰성 있고 신속하며 편리한 방법을 제공한다. 이 방법은 새로운 개인맞춤형/정밀 의료 시대에 새로운 HIV-1 단리물 및 기타 감염성 바이러스 어느 것에 대해서나 고용량 gRNA 라이브러리 스크린을 셋업하는데 이용될 수 있다.

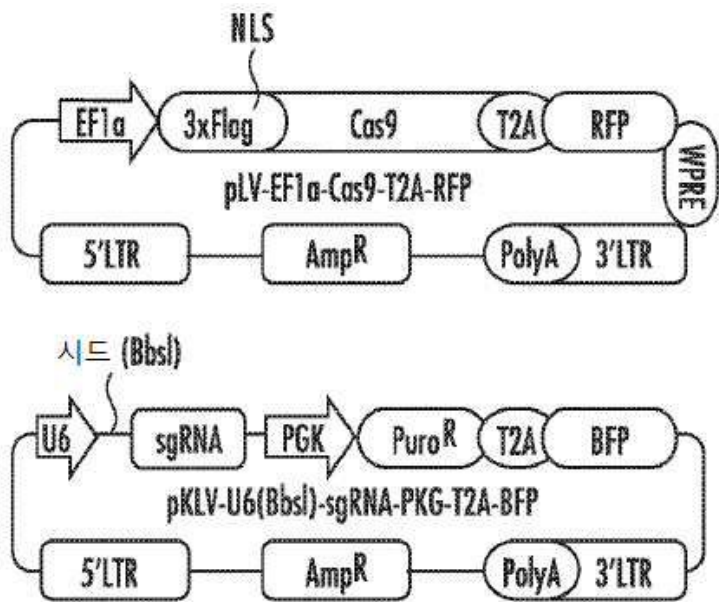
[0171] 본 발명은 구체적으로 설명하는 방식으로 기재되었으며, 사용된 용어는 한정을 위한 것이 아니라 기재를 위한 것임이 이해되어야 한다. 명백한 것은 본 발명의 여러 가지 수정 및 변경이 상기 교시 내용에 비추어 가능하다는 것이다. 따라서, 첨부된 특허청구 범위의 범주 내에서 본 발명이 구체적으로 기재된 것과는 다른 방식으로 실시될 수 있음이 이해되어야 한다.

도면

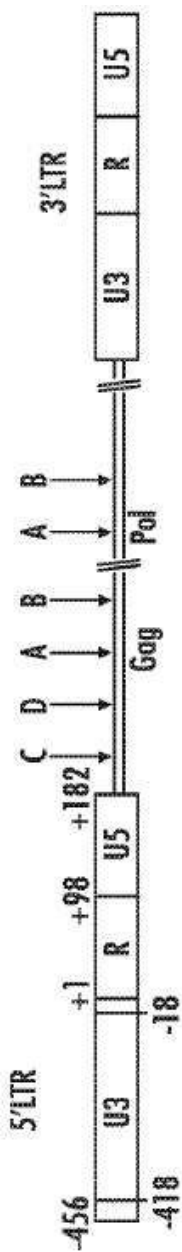
도면1a



도면1b



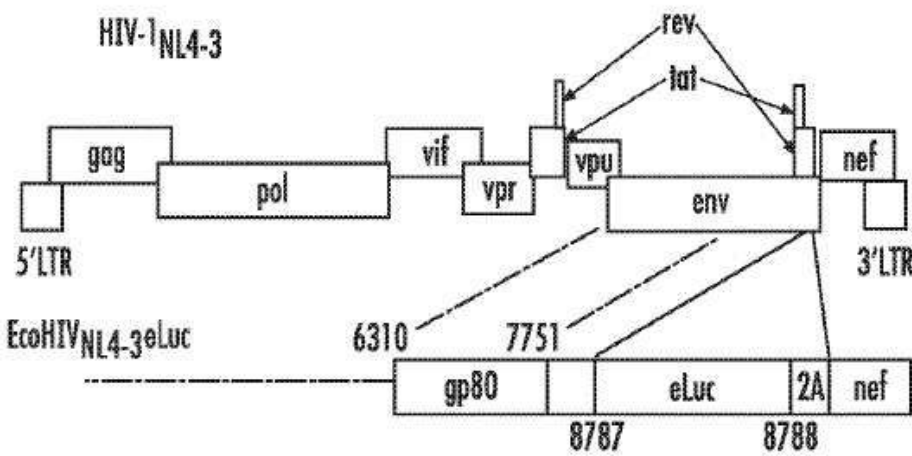
도면1c



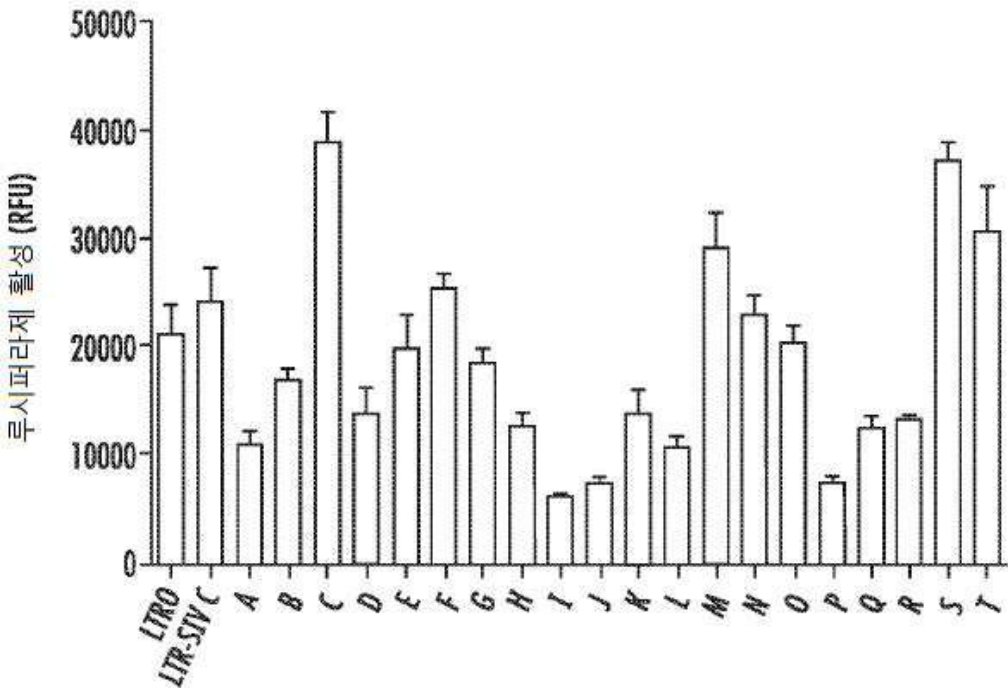
도면1d

^G
TCCTTGATCTGTGGATCTAC^HCCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATAT
^E
CCACTGACCTTTGGATGGTCTTCAAGTTAGTACCAGTTGAAC^SCCAGAGCAAGTAGAAGAGGCCCAATGAAGGAGAGAACAA
^{A,I}
CAGCTTGTTACACCCATGAGCCAGCAIGGGGATGGAGGACCCGGAGGGAGAAGTATTAGTGTGGAAAGTTGACAGCCCTCC
^{2,I}
TAGCATTTGTCACATGGCCCGAGAGCTGCA^{B,M}TCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCIACAAGGGACTT
^K
TCCGCTGGGGACTTTCCAGGGAGGGTGTGGCTTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAGC
⁰
^P
^{D,3,R}

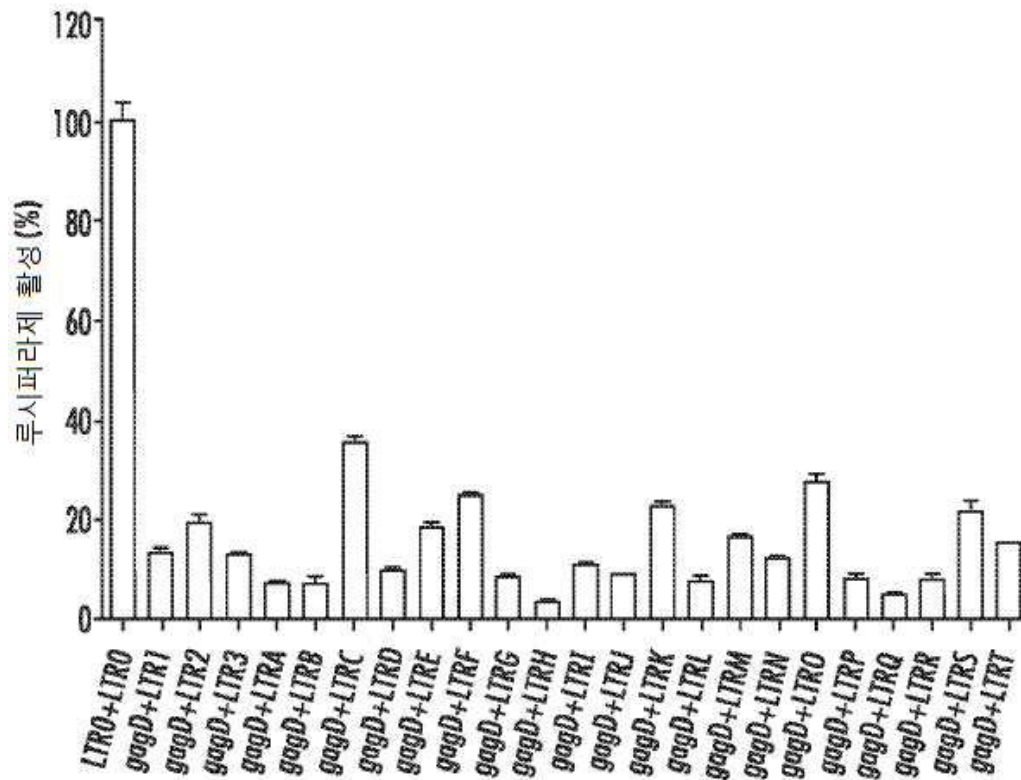
도면1e



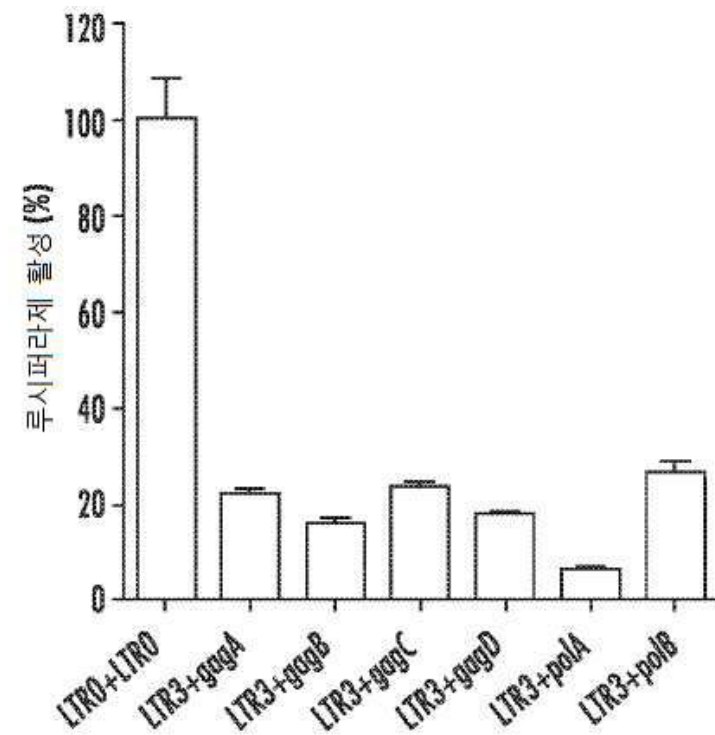
도면2a



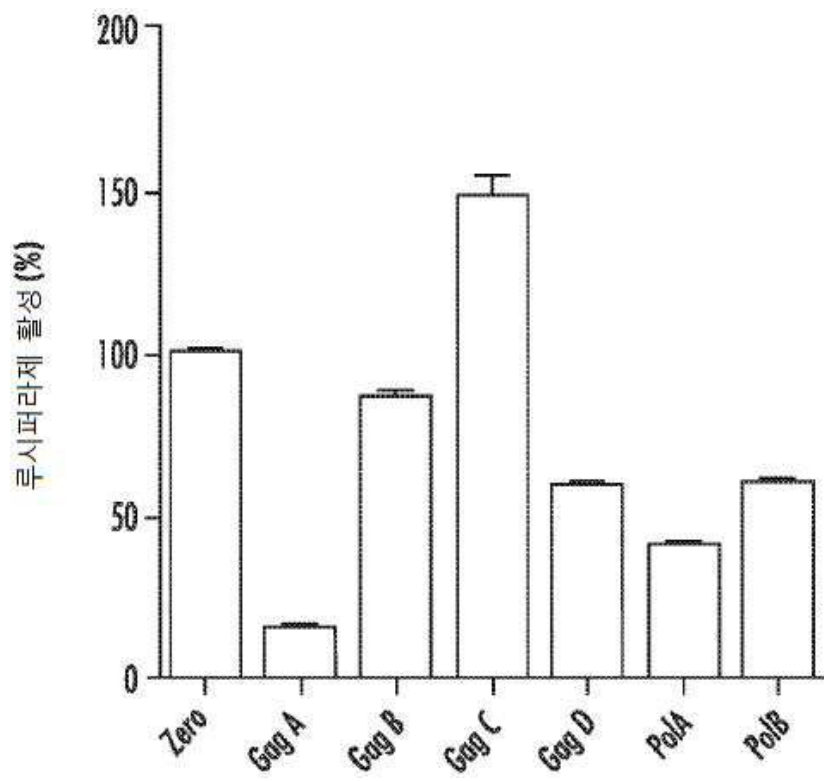
도면2b



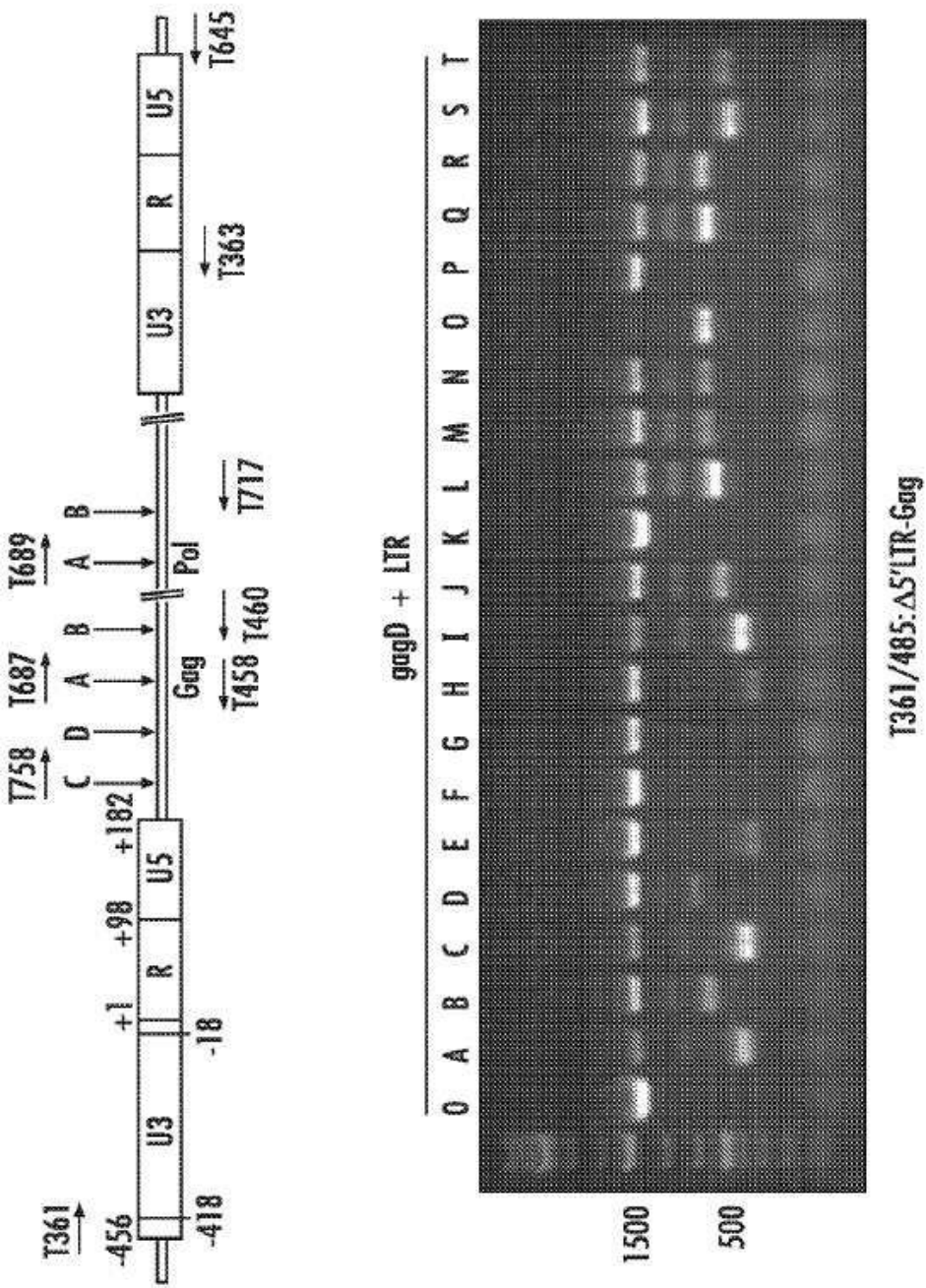
도면2c



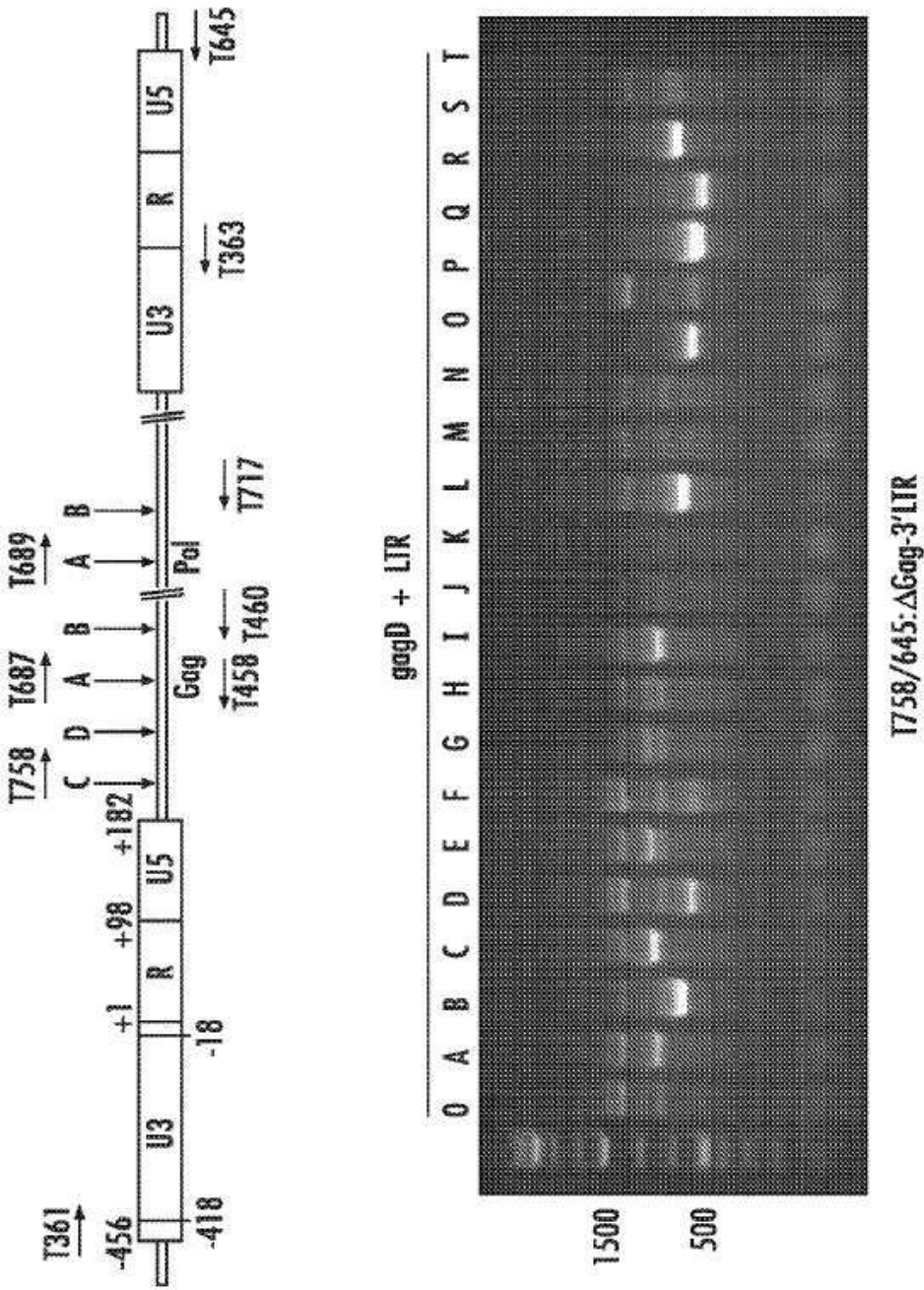
도면2d



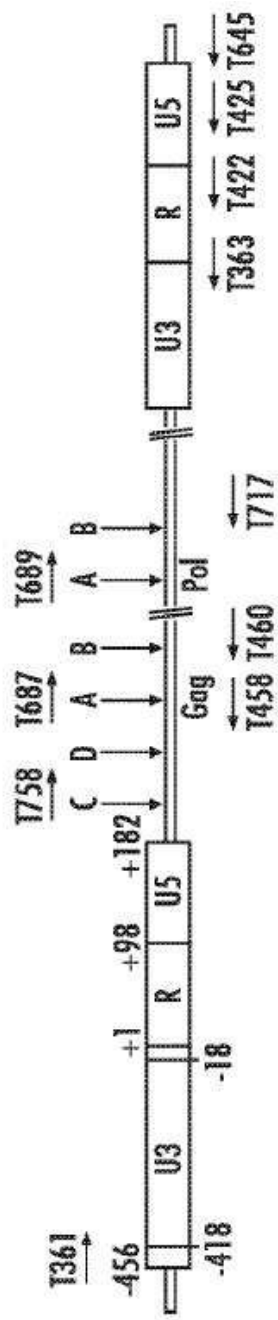
도면3a



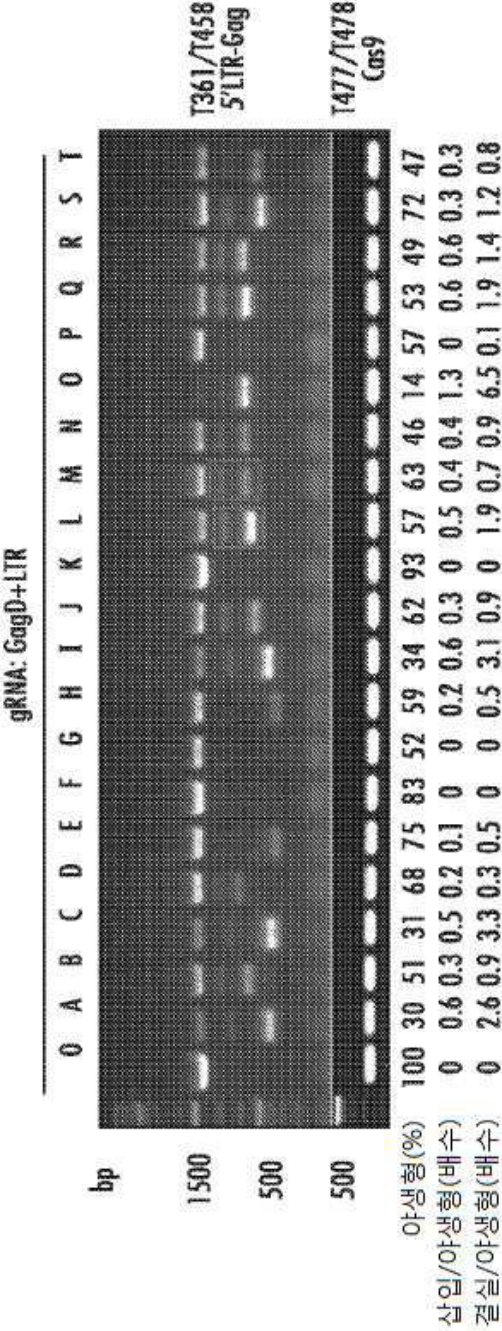
도면3b



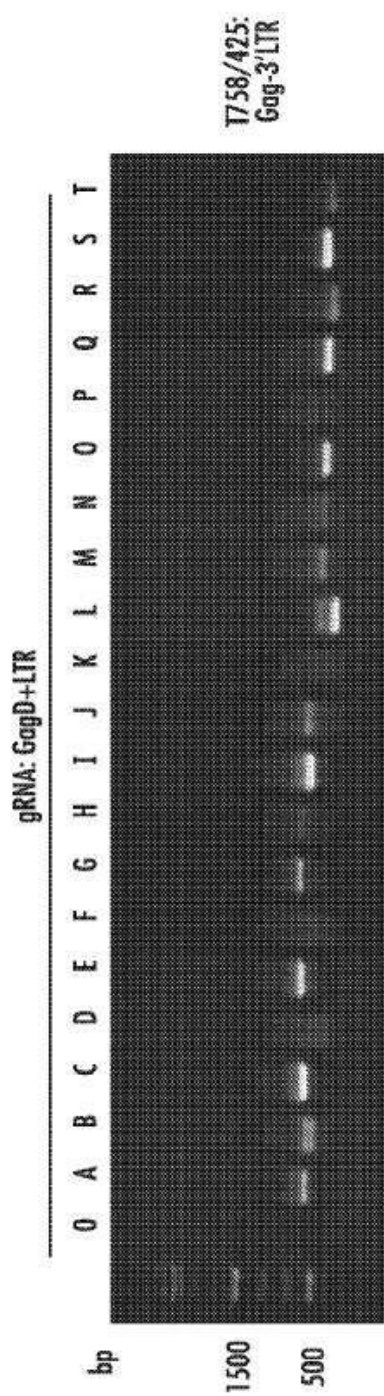
도면3d



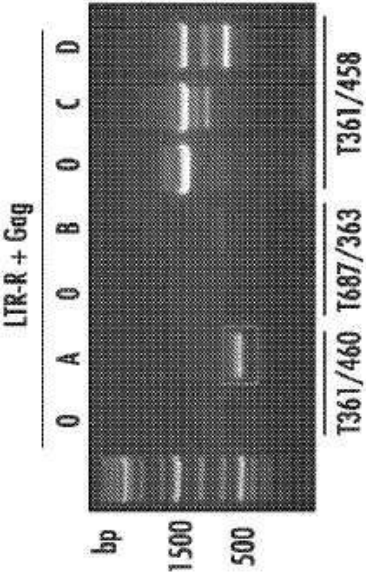
도면3e



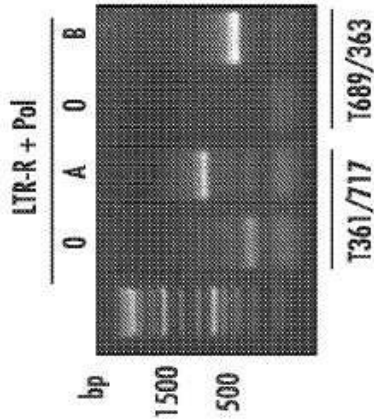
도면3f



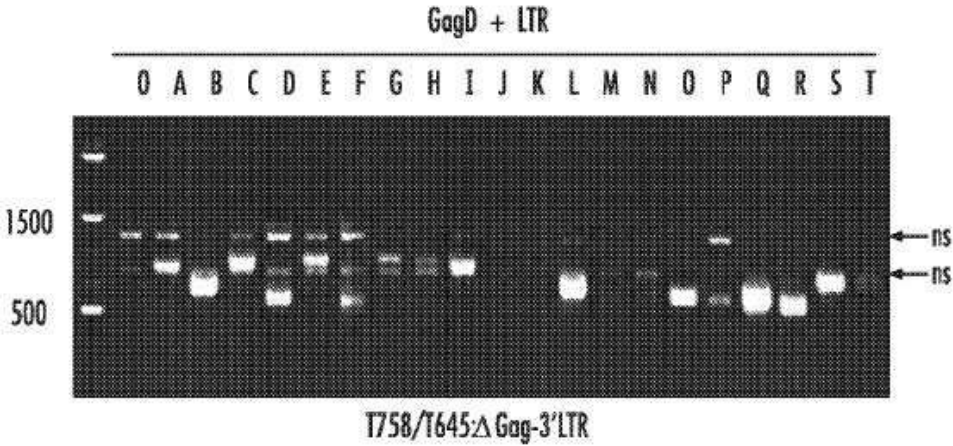
도면3g



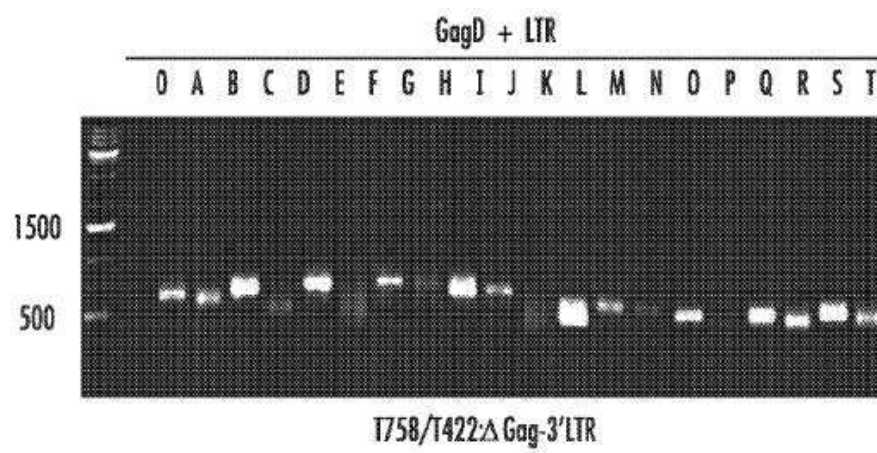
도면3h



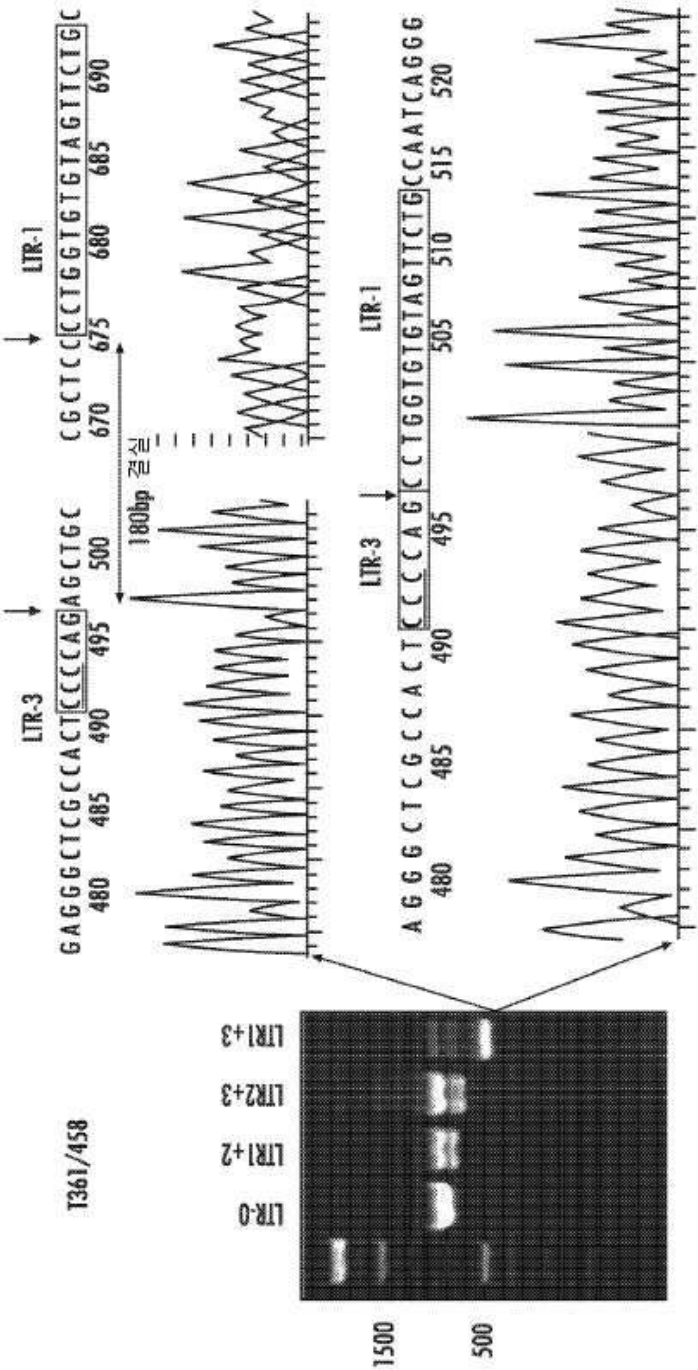
도면3i



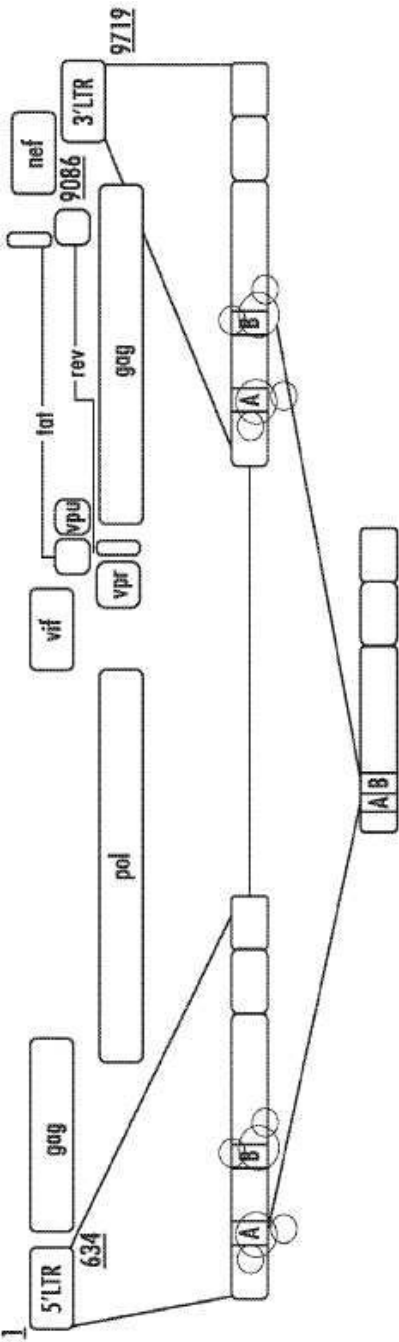
도면3j



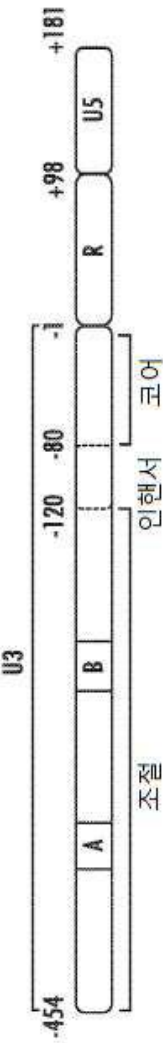
도면4



도면5a



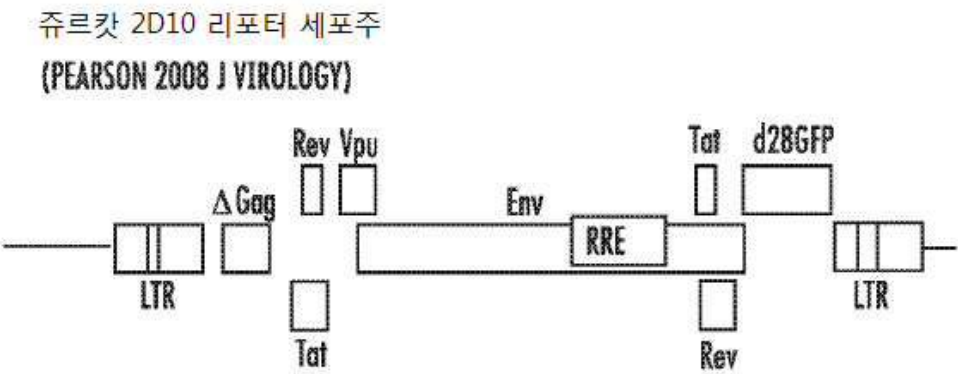
도면5b



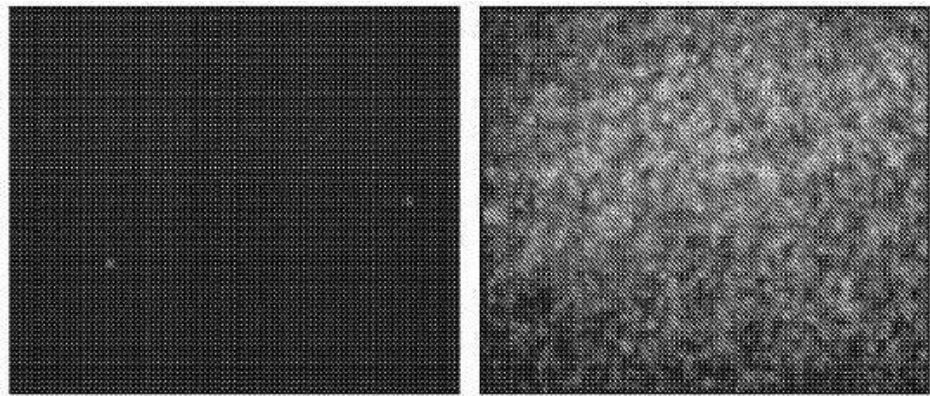
도면5c

-454 TGG AAGGGCTAATTG GTC CCAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCAVAVAVAAAGGCTACTTCCCTGATTGGCAGAACTACAC
 ACCAGGGCCAGGGGATCAGATATCCACTGACCTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAAGCAAGTAGAAGAGGCCCCAAATAAG
 GAGAGAAGAACAGCTTGTACACCCCTATGAGCCAGCATGGGATCCAGGACCCGAGGGACAAGTATTAGTGTTGGAAGTTTGACAGCCTCCTA
 GCATTTGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTCTACAAGGGACTTTCGCTGGGGGACTTTC
 CAGGGAGGTGTGGCCTGGGCGGGA CTGGGAGTGCGGAGCCCTCAGATGCTACATATAAGCAGCTGCTTTTGGCCGTACTGGGTCTCTCTGG
 TTAGACCAGATCTGAGCCTGGGAGCTCTLTGGCTAACTAGGGAAACCACTGCTTAAGCCTCAATAAAGCTTACCTTGAGTGCTCAAAAGTAGTGT
 GTGCCCGTCTGTTGTGTGTA CTCTGGTAACTAGAGATCCCTCAACCCCTTTTAGTCAGTGCTGGAAAAATCTTAGCA +181

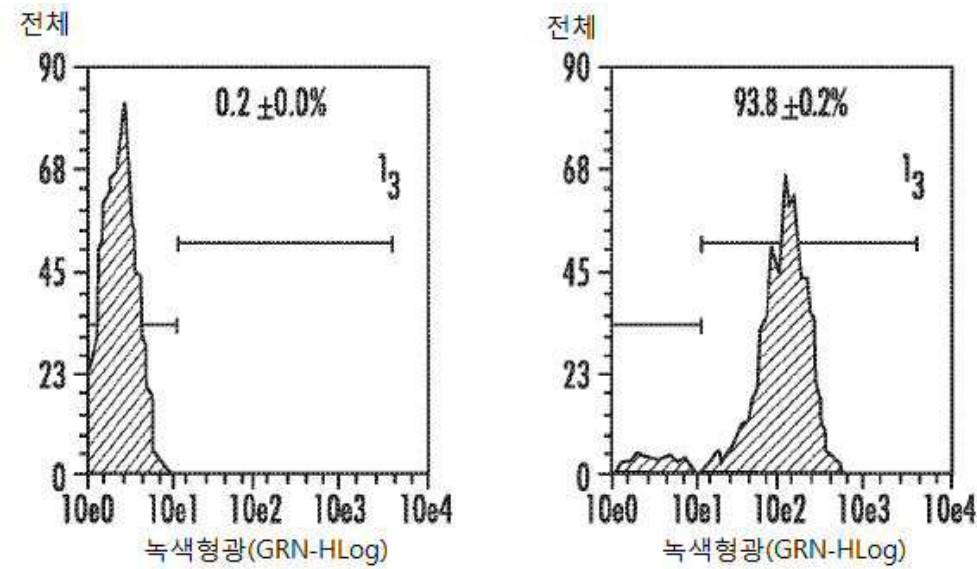
도면6a



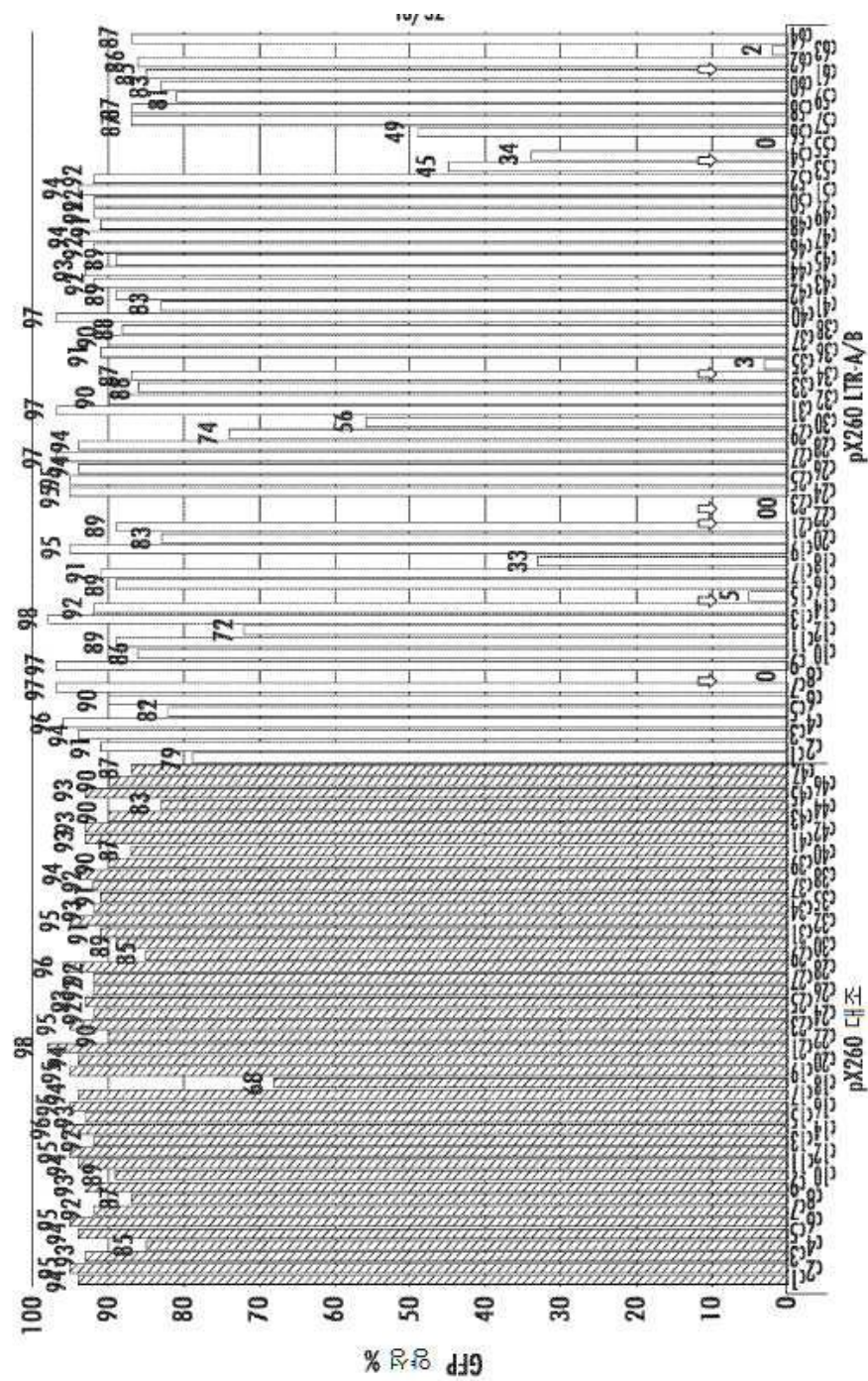
도면6b



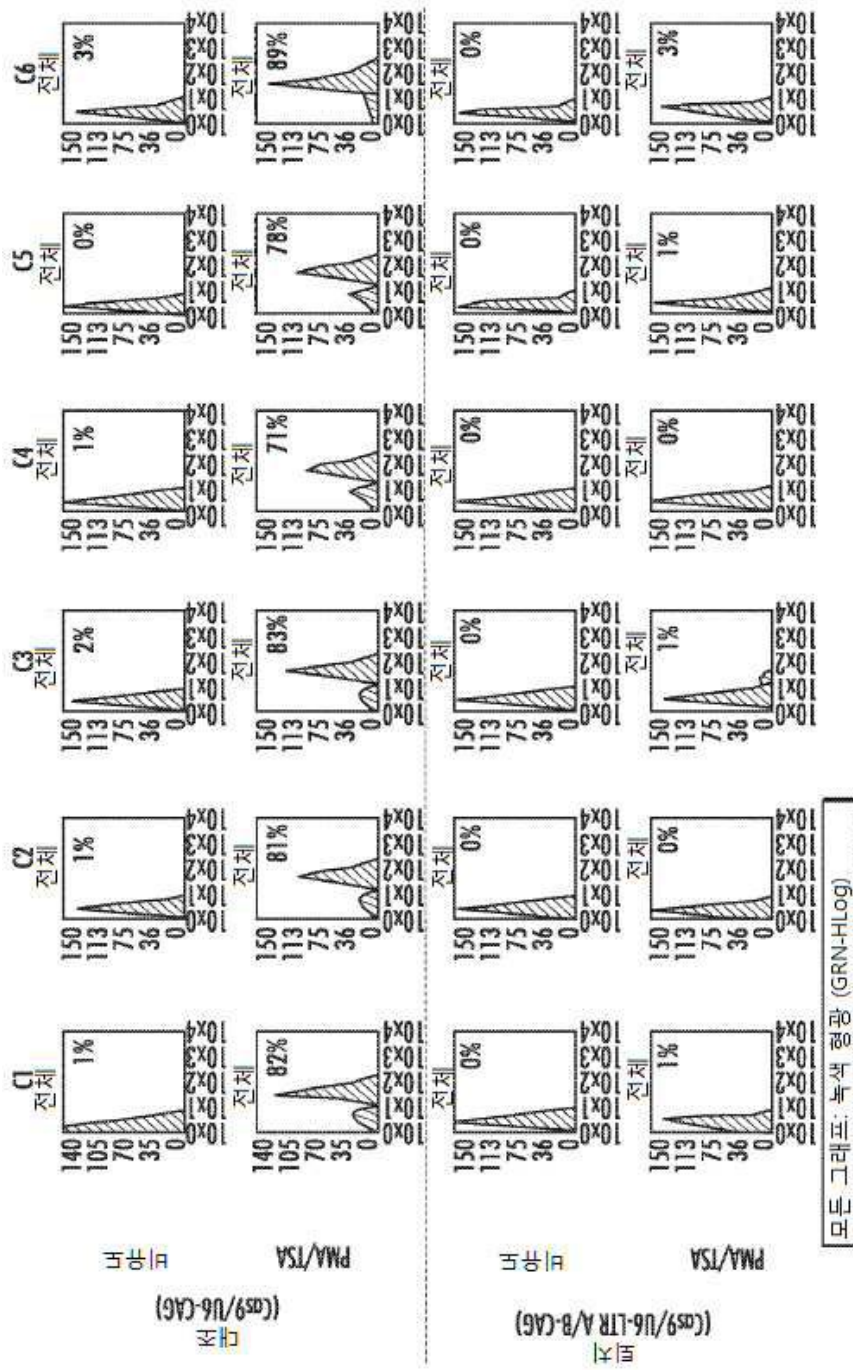
도면6c



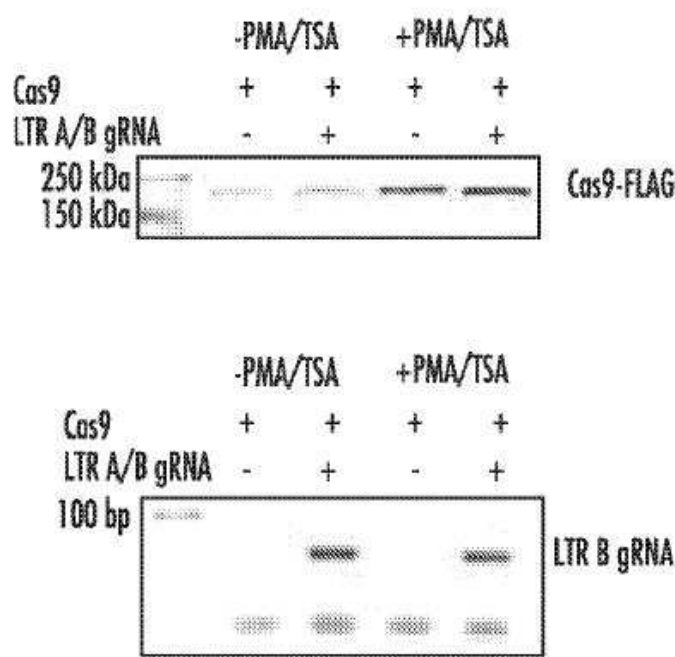
도면7a



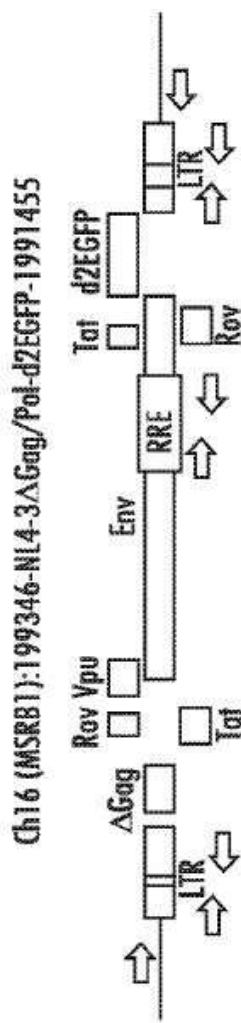
도면7b



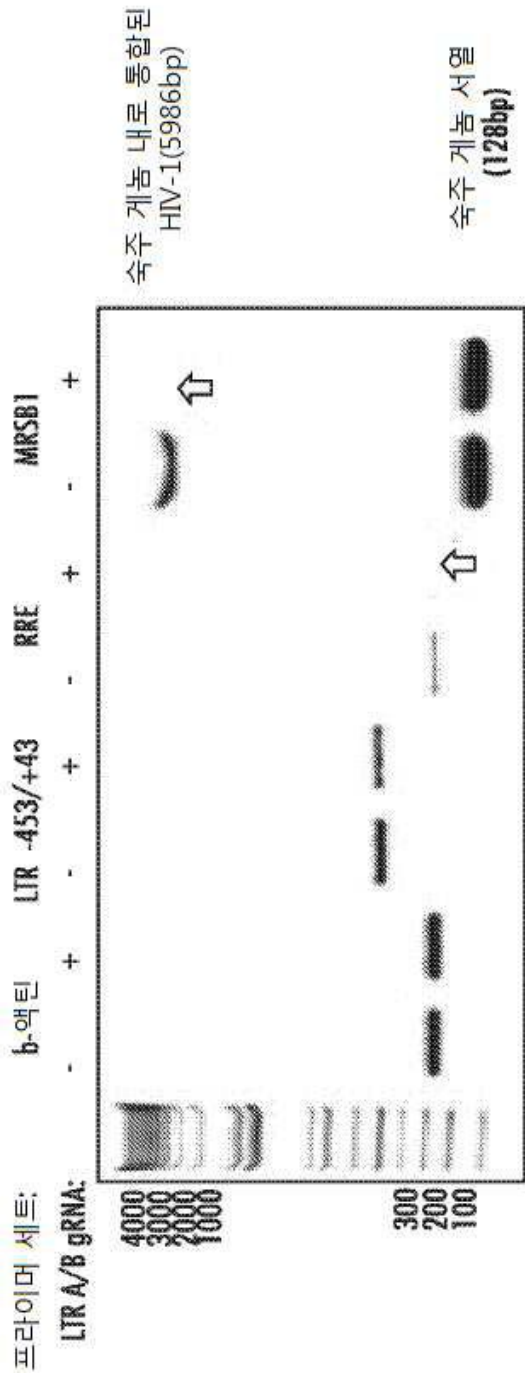
도면7c



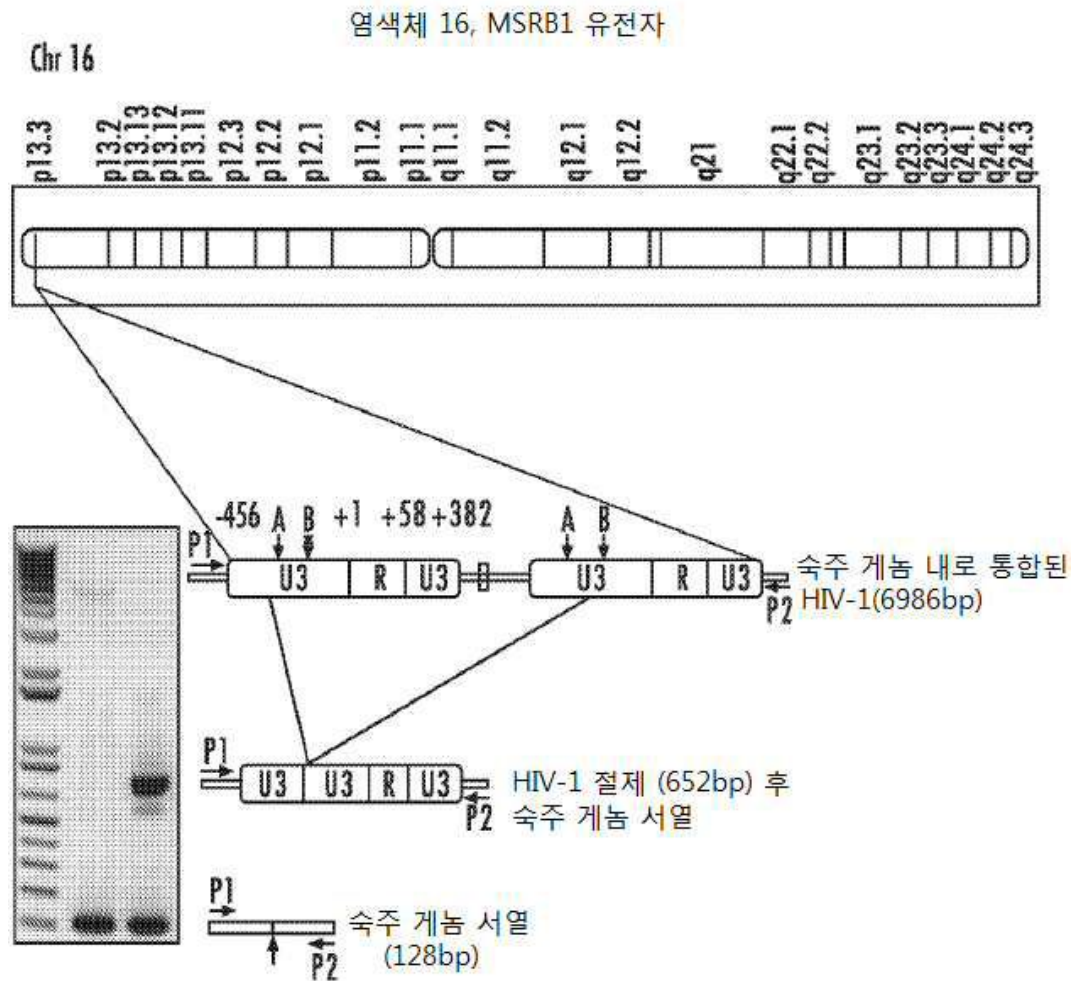
도면8a



도면8b



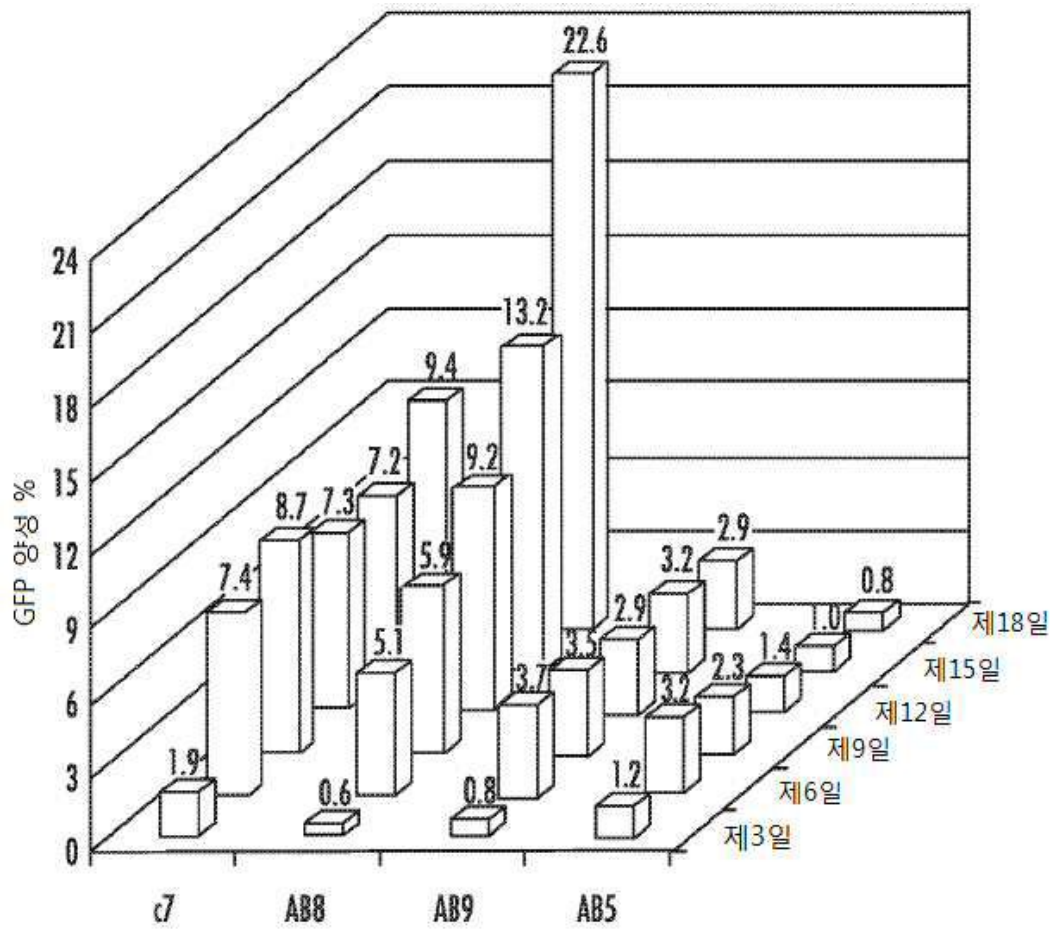
도면8c



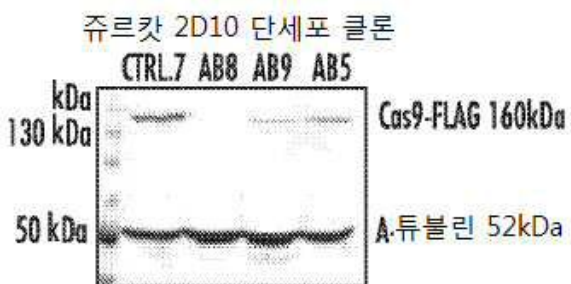
도면8d



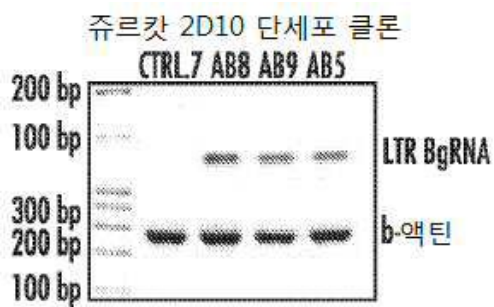
도면9a



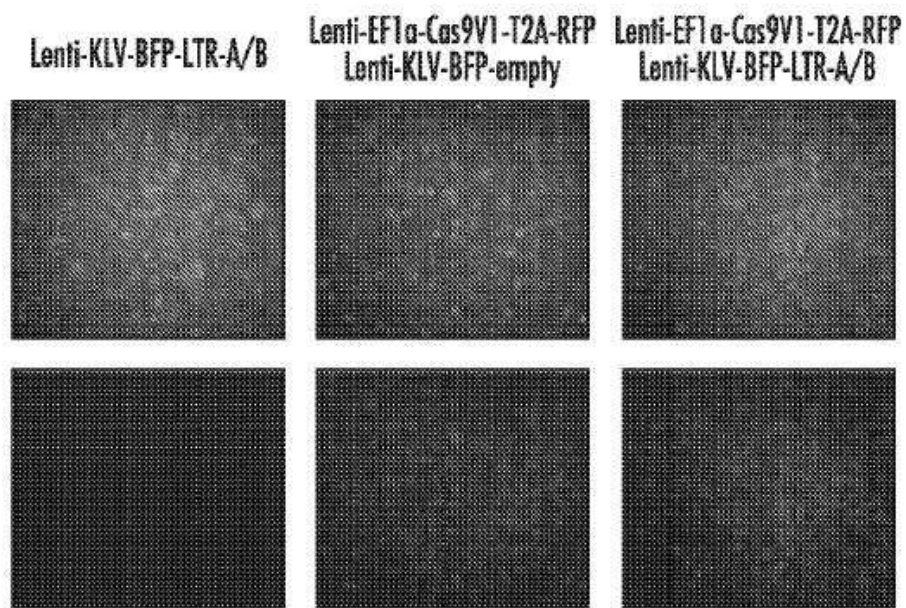
도면9b



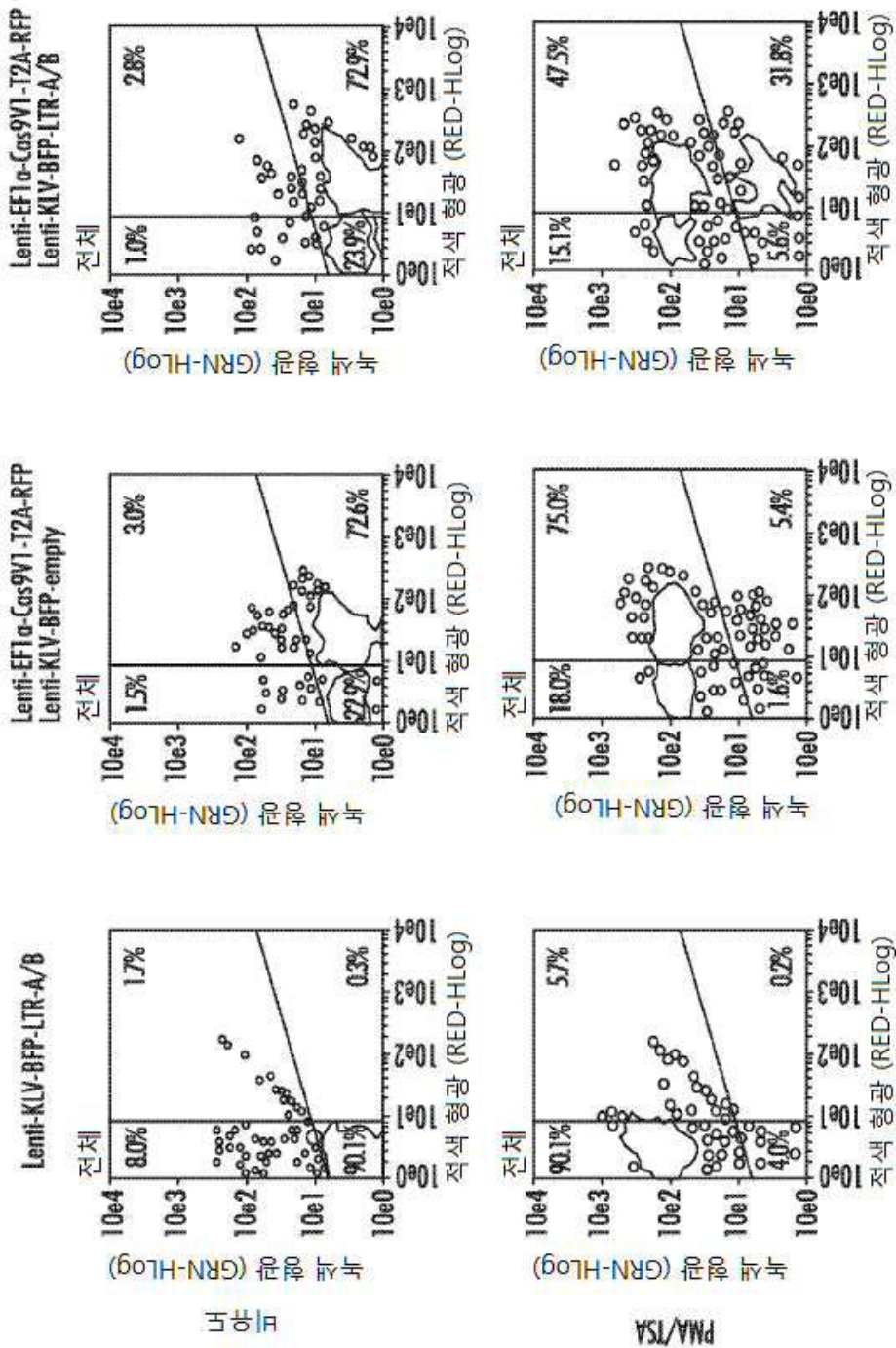
도면9c



도면10a



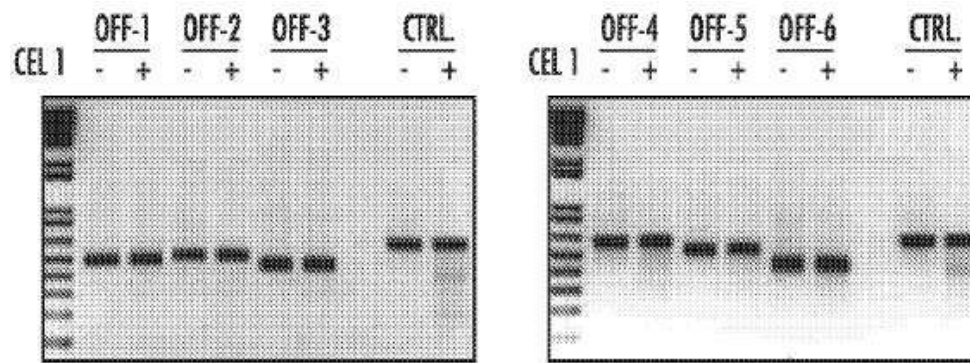
도면10b



도면11a

| | | PAM에 선행하는 시드 서열 길이(bp) | | | | | | | | |
|-------|-------------------------|------------------------|----|----|----|----|----|----|----|----|
| | gRNA 서열 | 20 | 19 | 18 | 17 | 16 | 15 | 14 | 13 | 12 |
| LTR A | ATCAGATATCCACTGACCTTTGG | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 6 | 16 |
| LTR B | CAGCAGTTCGAAGTACTCCCGG | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 |

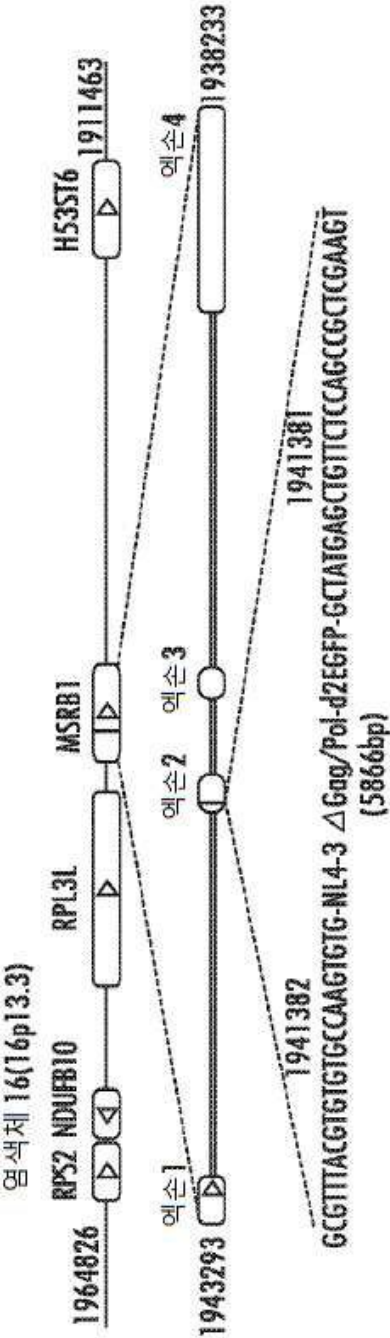
도면11b



도면11c

| | 위치 | 서열 | 시드 서열 길이(bp) | 검출된 인델 |
|-------|-------|------------------|--------------|--------|
| Off-1 | chr2 | TCCACTGACCTTTGG | 12 | 0 |
| Off-2 | chr4 | TCCACTGACCTTTGG | 12 | 0 |
| Off-3 | chr5 | TCCACTGACCTTTGG | 12 | 0 |
| Off-4 | chr6 | ATCCACTGACCTTTGG | 13 | 0 |
| Off-5 | chr11 | TCCACTGACCTTTGG | 12 | 0 |
| Off-6 | chr16 | TCCACTGACCTTTGG | 12 | 0 |

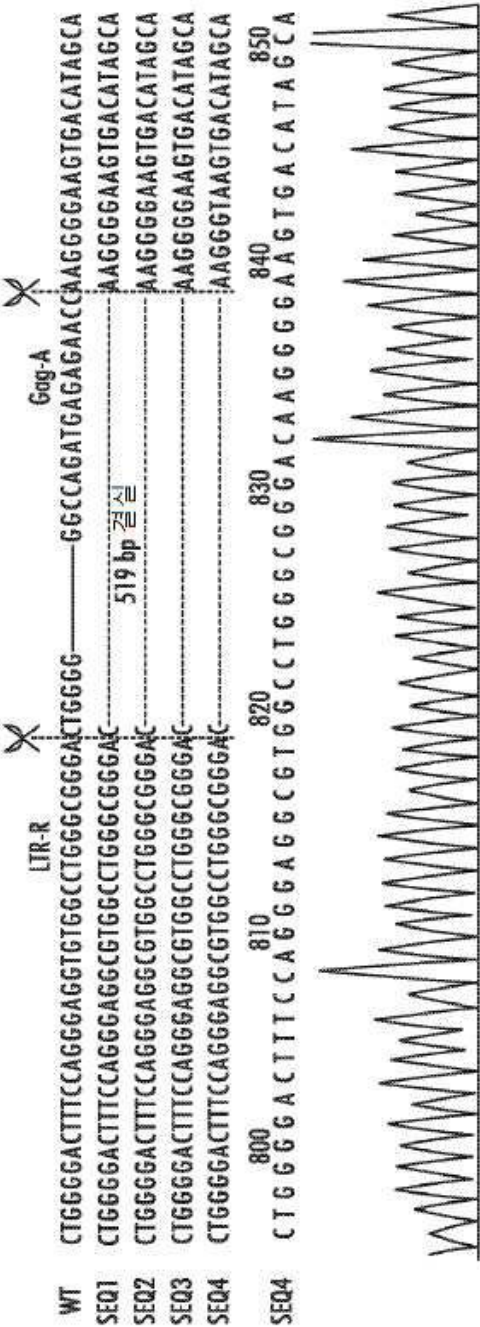
도면11d



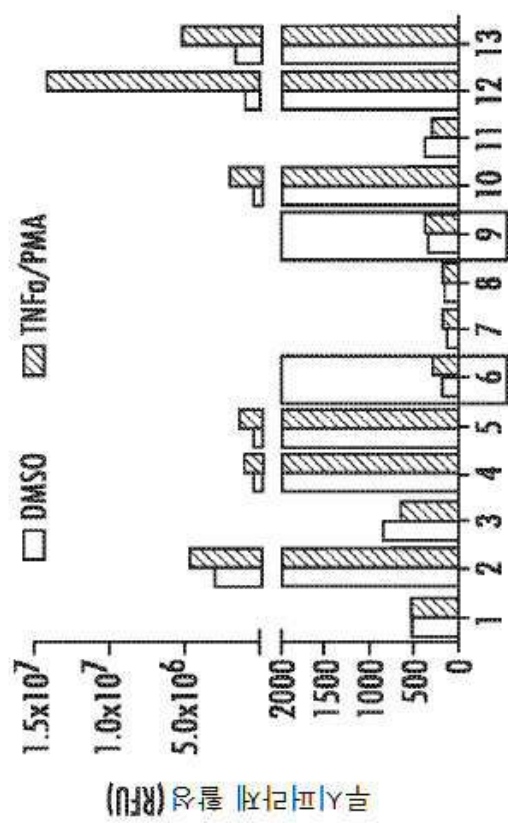
도면11e

| qRT-PCR 유전자 | 통합 부위로부터 거리 | 표적/표준 비율 | | 대조에 비교한 변화 |
|----------------|----------------|--------------|-------------------|---------------|
| | | Cas9 (대조) | Cas9+gRNA (퇴치) | |
| RPS2 | +23444 | 3.6E-01 | 4.3E-01 | +0.21 |
| NDUFB10 | +18134 | 2.9E-02 | 3.4E-02 | +0.17 |
| RPL3L | +13296 | 3.3E-06 | 4.1E-06 | +0.24 |
| MSRB1 | 0 | 1.2E-03 | 1.4E-03 | +0.21 |
| HS3ST6 | -23152 | 4.8E-07 | 1.7E-07 | -0.65 |

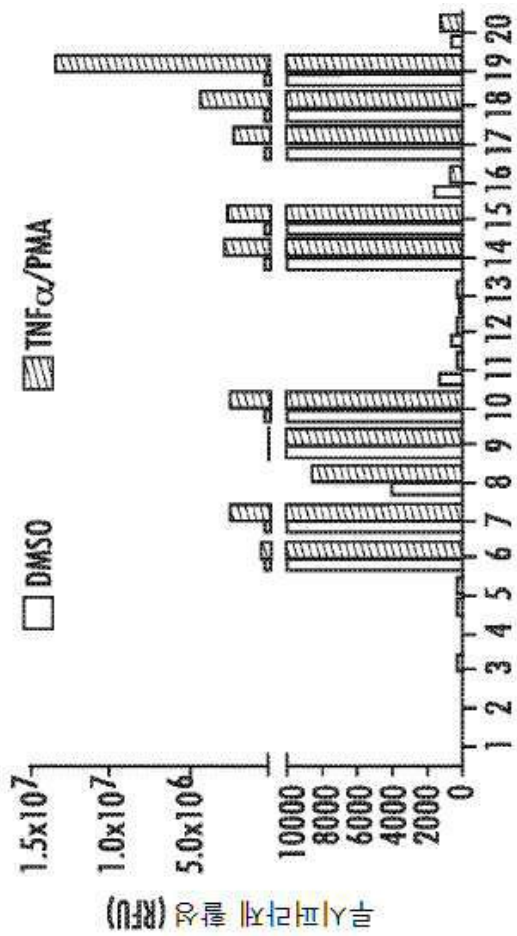
도면12a



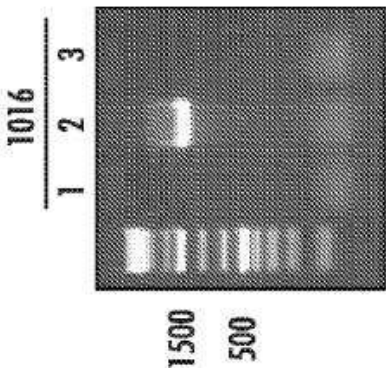
도면13a



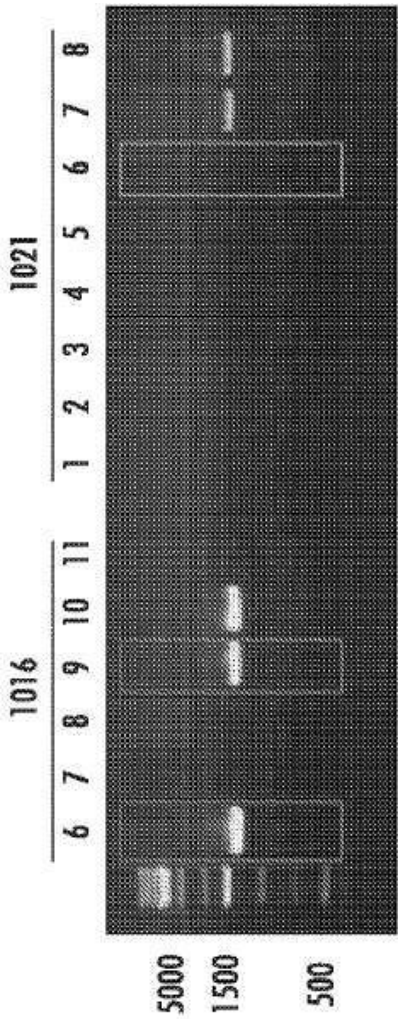
도면13b



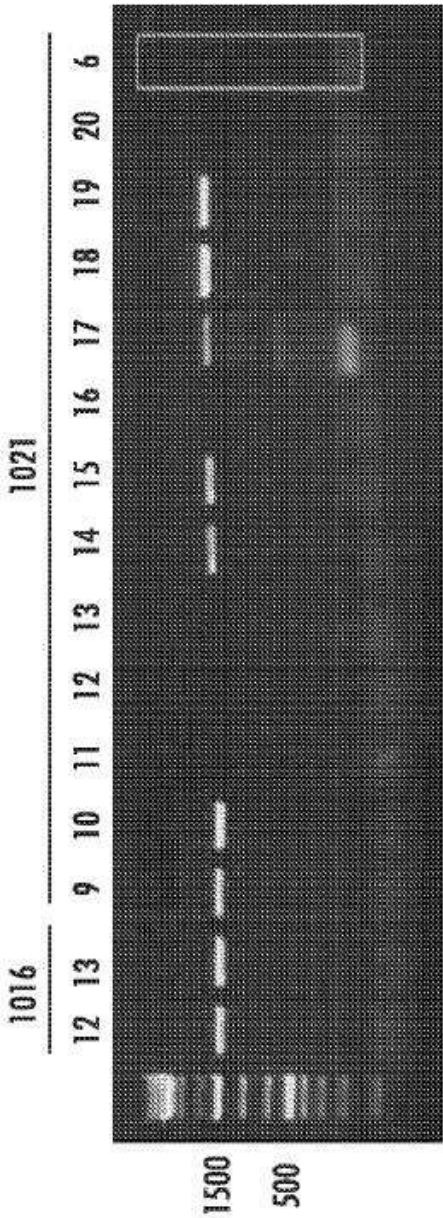
도면13c



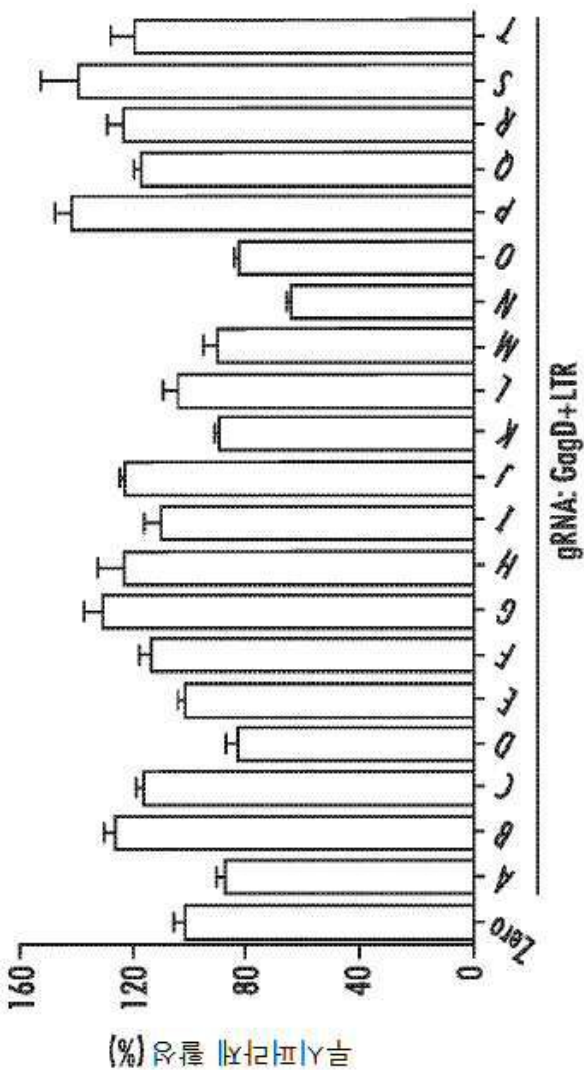
도면13d



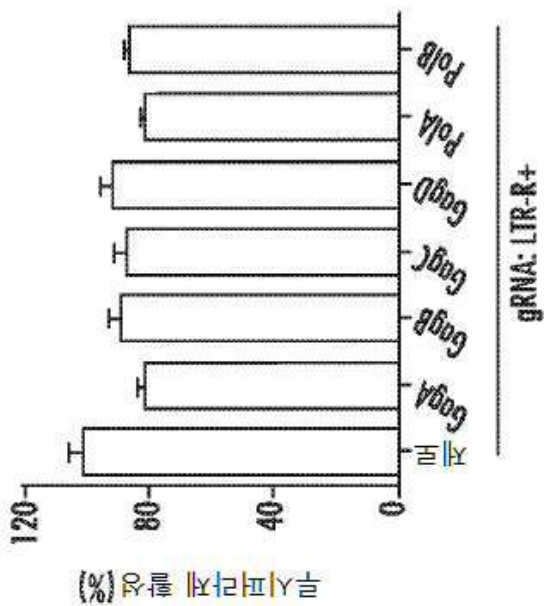
도면13e



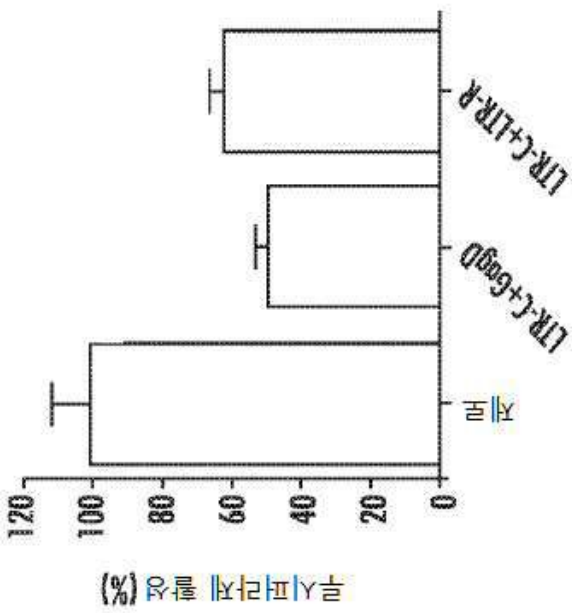
도면14a



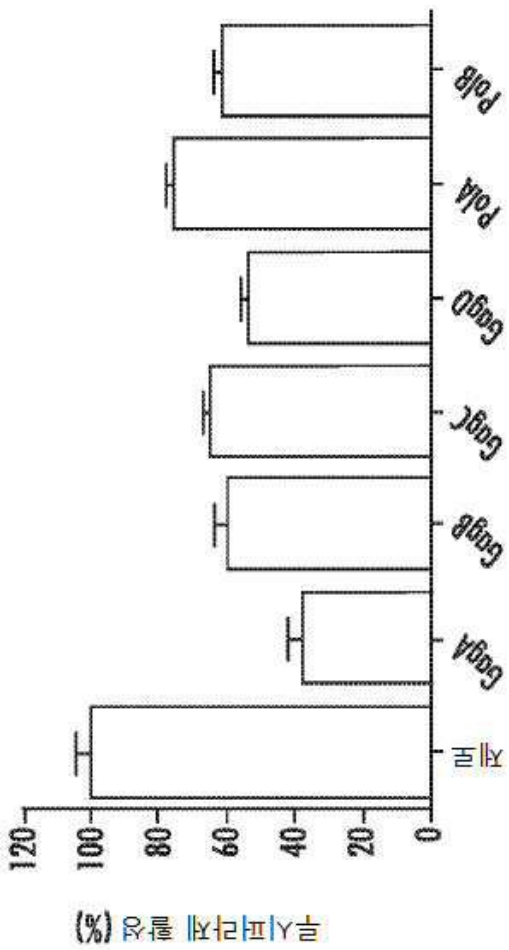
도면14b



도면14c



도면14d



서열 목록

SEQUENCE LISTING

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

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

<130> 49411.016

<140> PCT/US2016/035141

<141> 2016-06-01

<150> 62/308,320

<151> 2016-03-15

<150> 62/169,633

<151> 2015-06-02

<150> 62/169,384

<151> 2015-06-01

<160> 90

<170> PatentIn version 3.5

<210> 1

<211> 36

<212> DNA

<213> Artificial Sequence

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

primer

<400> 1

aaacagggcc agggatcaga tatccactga ccttgt

36

<210> 2

<211> 35

<212> DNA

<213> Artificial Sequence

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

primer

<400> 2

taaacaaggt cagtggatat ctgatccctg gccct

35

<210> 3

<211> 36

<212> DNA

<213> Artificial Sequence

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

<400> 3

aaacagctcg atgtcagcag ttcttgaagt actcgt 36

<210> 4

<211> 35

<212> DNA

<213> Artificial Sequence

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

<400> 4

taaacgagta cttcaagaac tgctgacatc gagct 35

<210> 5

<211> 24

<212> DNA

<213> Artificial Sequence

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

<400> 5

caccgattgg cagaactaca cacc 24

<210> 6

<211> 24

<212> DNA

<213> Artificial Sequence

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

<400> 6

aaacggtgtg tagttctgcc aatc 24

<210> 7

<211> 24

<212> DNA

<213> Artificial Sequence

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

<400> 7

caccgcgtgg cctgggcggg actg

24

<210> 8

<211> 24

<212> DNA

<213> Artificial Sequence

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

<400>

> 8

aaaccagtcc cgcccaggcc acgc

24

<210> 9

<211> 27

<212> DNA

<213> Artificial Sequence

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

<400> 9

caccgatctg tggatctacc acacaca

27

<210> 10

<211> 27

<212> DNA

<213> Artificial Sequence

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

<400> 10

aaactgtgtg tggtagatcc acagatc

27

<210> 11

<211> 27

<212> DNA

<213> Artificial Sequence

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

primer

<400> 11

caccgctgct tatatgcagc atctgag 27

<210> 12

<211> 27

<212> DNA

<213> Artificial Sequence

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

primer

<400> 12

aaacctcaga tgctgcatat aagcagc 27

<210> 13

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

><223> Description of Artificial Sequence: Synthetic

primer

<400> 13

caccgtgtgg tagatccaca gatca 25

<210> 14

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 14

aaactgatct gtggatctac cacac 25

<210> 15

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 15
caccgcaggg aagtagcctt gtgtg 25

<210> 16
<211> 25
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
primer

<400> 16
aaaccacaca aggctacttc cctgc 25

<210> 17
<211> 25
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
primer

<400> 17
caccgatcag atatccactg acctt 25

<210> 18
<211> 25
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
primer

<400> 18
aaacaaggtc agtggatata tgatc 25

<210> 19
<211> 25
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
primer

<400> 19
caccgcacac taatacttct ccctc 25

<210> 20

<211> 25

<212> DNA

<213> Artificial Sequence

<220

><223> Description of Artificial Sequence: Synthetic

primer

<400> 20

aaacgaggga gaagtattag tgtgc 25

<210> 21

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 21

caccgcctcc tagcatttcg tcaca 25

<210> 22

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 22

aaactgtgac gaaatgctag gaggc 25

<210> 23

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 23

caccgcatgg cccgagagct gcatc 25

<210> 24

<211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 24
 aaacgatgca gctctcgggc catgc 25

<210> 25
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 25
 caccgcagca gtctttgtag tactc 25

<210> 26
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 26
 aaacgagtac tacaaagact gctgc 25

<210> 27
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 ><223> Description of Artificial Sequence: Synthetic
 primer
 <400> 27
 caccgetgac atcgagcttt ctaca 25

<210> 28
 <211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 28

aaactgtaga aagctcgatg tcagc 25

<210> 29

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 29

caccgtctac aagggaacttt ccgct 25

<210> 30

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 30

aaacagcggg aagtccttg tagac 25

<210> 31

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 31

caccgctttc cgctggggac ttcc 25

<210> 32

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 32

aaacggaaag tccccagcgg aaagc

25

<210> 33

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 33

caccgcctcc ctggaaagtc cccag

25

<210> 34

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

><223> Description of Artificial Sequence: Synthetic

primer

<400> 34

aaacctgggg actttccagg gaggc

25

<210> 35

<211> 24

<212> DNA

<213> Artificial Sequence

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

primer

<400> 35

caccgcctgg gcgggactgg ggag

24

<210> 36

<211> 24

<212> DNA

<213> Artificial Sequence

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

primer

<400> 36

aaacctcccc agtcccgccc aggc

24

<210> 37

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 37

caccgtccat cccatgcagg ctcac

25

<210> 38

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 38

aaacgtgagc ctgcatggga tggac

25

<210> 39

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 39

caccgcggag agagaagtat tagag

25

<210> 40

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 40
aaacctctaa tacttctctc tccgc 25
<210> 41
<211> 24
<212> DNA
<213> Artificial Sequence
<220>
><223> Description of Artificial Sequence: Synthetic
primer
<400> 41
caccggccag atgagagaac caag 24
<210> 42
<211> 24
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
primer
<400> 42
aaaccttggt tctctcatct ggcc 24
<210> 43
<211> 26
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
primer
<400> 43
caccgccttc ccacaaggga aggcca 26
<210> 44
<211> 26
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
primer
<400> 44

aaactggcct tcccttgtgg gaaggc 26

<210> 45

<211> 26

<212> DNA

<213> Artificial Sequence

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

primer

<400> 45

caccgcgaga gcgtcggtat taagcg 26

<210> 46

<211> 26

<212> DNA

<213> Artificial Sequence

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

primer

<400> 46

aaaccgctta ataccgacgc tctcgc 26

<210> 47

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 47

caccggatag atgtaaaaga cacca 25

<210> 48

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

><223> Description of Artificial Sequence: Synthetic

primer

<400> 48

aaactggtgt cttttacatc tatcc 25

<210> 49

<211> 24

<212> DNA

<213> Artificial Sequence

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

<400> 49

caccgcagga tatgtaactg acag 24

<210> 50

<211> 24

<212> DNA

<213> Artificial Sequence

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

<400> 50

aaacctgtca gttacatatc ctgc 24

<210> 51

<211> 24

<212> DNA

<213> Artificial Sequence

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

<400> 51

caccgcatgg gtaccagcac acaa 24

<210> 52

<211> 24

<212> DNA

<213> Artificial Sequence

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

<400> 52

aaacttgtgt gctggtaccc atgc 24

<210> 53

<211> 25

<212> DNA

<213> Artificial Sequence

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

primer

<400> 53

caccgcttta ttgaggctta agcag 25

<210> 54

<211> 24

<212> DNA

<213> Artificial Sequence

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

primer

<400> 54

aaacgagtca cacaacagac gggc 24

<210> 55

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

><223> Description of Artificial Sequence: Synthetic

primer

<400> 55

tggaatgcag tggcgcgac ttggc 25

<210> 56

<211> 23

<212> DNA

<213> Artificial Sequence

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

primer

<400> 56

cacagcatca agaagaacct gat 23

<210> 57

<211> 23

<212> DNA

<213> Artificial Sequence

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

<400> 57

tgaagatctc ttgcagatag cag 23

<210> 58

<211> 23

<212> DNA

<213> Artificial Sequence

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

<400> 58

gatctgtgga tctaccacac aca 23

<210> 59

<211> 23

<212> DNA

<213> Artificial Sequence

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

<400> 59

cccactgtgt ttagcatggg att 23

<210> 60

<211> 101

<212> RNA

<213> Artificial Sequence

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

<220><221> modified_base

<222> (1)..(20)

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

<400> 60

nnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaaagc aaguuaaaau aaggcuaguc 60

cguuaucac uugaaaaaag uggcaccga gucggugcuuu u 101

<210> 61
 <211> 400
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 61
 tccttgatct gtggatctac cacacacaag gctacttccc tgattggcag aactacacac 60

 cagggccagg gatcagatat ccactgacct ttggatgggtg cttcaagtta gtaccagttg 120
 aaccagagca agtagaagag gccaatgaag gagagaacaa cagcttgta caccctatga 180
 gccagcatgg gatggaggac ccggagggag aagtattagt gtggaagttt gacagcctcc 240
 tagcatttcg tcacatggcc cgagagctgc atccggagta ctacaaagac tgctgacatc 300
 gagctttcta caagggactt tccgctgggg actttccagg gaggtgtggc ctgggcggga 360
 ctggggagtg gcgagccctc agatgctaca tataagcagc 400
 <210> 62
 <211> 26
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 62
 gagggctcgc cactccccag agctgc 26
 <210> 63
 <211> 25
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 63
 cgctcccctg gtgtgtagtt ctgcc 25
 <210> 64
 <211> 46
 <212> DNA
 <213> Human immunodeficiency virus 1
 <400> 64
 agggctcgcc actccccagc ctggtgtgta gttctgcca tcaggg 46
 <210> 65
 <211> 633
 <212> DNA

<213> Human immunodeficiency virus 1

<400> 65

tggaagggt aatttggtcc caaaaaagac aagagatcct tgatctgtgg atctaccaca 60

cacaaggcta cticcctgat tggcagaact acacaccagg gccagggatc agatatccac 120

tgacctttgg atggtgcttc aagtttagtac cagttgaacc agagcaagta gaagaggcca 180

aataaggaga gaagaacagc ttgttacacc ctatgagcca gcatgggatg gaggacccgg 240

agggagaagt attagtgtgg aagtttgaca gcctcctagc atttcgtcac atggcccgag 300

agctgcatcc ggagtactac aaagactgct gacatcgagc tttctacaag ggactttccg 360

ctggggactt tccagggagg tgtggcctgg gcgggactgg ggagtggcga gccctcagat 420

gctacatata agcagctgct ttttgctgt actgggtctc tctggttaga ccagatctga 480

gcctgggagc tctctggeta actaggaac ccactgetta agcctcaata aagcttgctt 540

tgagtgtca aagtagtgtg tgcccgtctg ttgtgtgact ctggtaacta gagatccctc 600

aaccctttta gtcagtgtgg aaaatctcta gca 633

<210> 66

<211> 49

<212> DNA

<213> Human immunodeficiency virus 1

<400> 66

gccgagagct gcatccggag ctaagtttta ctcaagaac tgctgacat 49

<210> 67

<211> 23

<212> DNA

<213> Human immunodeficiency virus 1

<400> 67

atcagatata cactgacett tgg 23

<210> 68

<211> 23

<212> DNA

<213> Human immunodeficiency virus 1

<400> 68

cagcagttcg aagtactccc cgg 23

<210> 69

| | |
|---|----|
| <211> 15 | |
| <212> DNA | |
| <213> Human immunodeficiency virus 1 | |
| <400> 69 | |
| tccactgacc tttgg | 15 |
| <210> 70 | |
| <211> 16 | |
| <212> DNA | |
| <213> Human immunodeficiency virus 1 | |
| <400> 70 | |
| atccactgac ctttgg | 16 |
| <210> 71 | |
| <211> 24 | |
| <212> DNA | |
| <213> Human immunodeficiency virus 1 | |
| <400> 71 | |
| gcgtttacgt gtgtgccaag tgtg | 24 |
| <210> 72 | |
| <211> 30 | |
| <212> DNA | |
| <213> Human immunodeficiency virus 1 | |
| <400> 72 | |
| gctatgagct gttctccagc cgctcgaagt | 30 |
| <210> 73 | |
| <211> 78 | |
| <212> DNA | |
| <213> Human immunodeficiency virus 1 | |
| <400> 73 | |
| ctggggactt tccagggagg tgtggcctgg gcgggactgg ggggccagat gagagaacca | 60 |
| aggggaagtg acatagca | 78 |
| <210> 74 | |
| <211> 56 | |
| <212> DNA | |

<213> Artificial Sequence

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

<400> 74

ctggggactt tccaggagg cgtggcctgg gcgggacaag gggaagtgac atagca 56

<210> 75

<211> 56

<212> DNA

<213> Artificial Sequence

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

<400> 75

ctggggactt tccaggagg cgtggcctgg gcgggacaag ggtaagtgac atagca 56

<210> 76

<211> 65

<212> DNA

<213> Human immunodeficiency virus 1

<400> 76

acgtggcccg agagctgcat ccggggatag atgtaaaaga caccaaggaa gccttagata 60
agata 65

<210> 77

<211> 42

<212> DNA

<213> Artificial Sequence

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

<400> 77

acgtggcccg agagctgcac caaggaagcc ttagataaga ta 42

<210> 78

<211> 41

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 78

acatggcccg agagctgccc aaggaagcct tagataagat a 41

<210> 79

<211> 40

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 79

acgtggccag agagctgcca aggaagcctt agataagata 40

<210> 80

<211> 42

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 80

acatggcccg agagctgcac caaggaagcc ttagataaga ta 42

<210> 81

<211> 78

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 81

acgtggcccg agagctgcag gtgtgccacc atgccagct agtttttgta tttaccaag 60

gaagccttag aaaagata 78

<210> 82

<211> 233

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 82

acatggcccc agagctgccca ggccggtgat gctctggtgg atcagggtgg cgtccagcac 60

ctctttggtg ctggtgtacc tcttcgggtc gatggtggtg tcaaagtact tgaaggcggc 120

aggggctccc agattggtca gggtaaacag gtggatgata ttctcggcct gctctctgat 180

gggcttatcc cgggtccttg ttaggcgga ccaaggaagc ctagataag ata 233

<210> 83

<211> 328

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 83

acgtggcccc agagctgcat ccggactgta ctgagtctct ctggttagac cagatctgag 60

cctgggagct ctctggctaa ctaggaacc cactgcttaa gcctcaataa agcttgctt 120

gagtgttca agtagtgtgt gccctctgt tgtgtgattc tggtactag agatccctca 180

gaccctttta gtcagtgtgg aaaatctcta gcagggcccg tttaaaccg ctgacagcc 240

tcgactgtgc cttctagtta ccagccatct gttgtttgcc cctccccgt gccttccttg 300

accctgcaag gaagccttag ataagata 328

<210> 84

<211> 86

<212> DNA

<213> Human immunodeficiency virus 1

<400> 84

agctgcatcc ggagtactac aaagactgct gacatcgagc tttctggata gatgtaaaag 60

acaccaagga agccttagat aagata 86

<210> 85

<211> 69

<212> DNA

<213> Artificial Sequence

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

oligonucleotide

<400> 85

agctgcatcc ggagtactac aaagactgct gacatcgagc tttctaccaa ggaagcctta 60
gataagata 69
<210> 86
<211> 67
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 86
agctgcatcc ggagtactac aaagactgct gacatcgagc tttaccaagg aagccttaga 60

taagata 67
<210> 87
<211> 70
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide
<400> 87
agctgcatcc ggagtactac aaagactgct gacatcgagc tttctcacca aggaagcctt 60
agataagata 70
<210> 88
<211> 68
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 88
agctgcatcc ggagtactac aaagactgct gacatcgagc tttctccaag gaagccttag 60
ataagata 68
<210> 89
<211> 398
<212> DNA
<213> Artificial Sequence

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

polynucleotide

<400> 89

| | |
|--|-----|
| agctgcatcc ggagctgtac tgggtctctc tggtttagacc agatctgagc ctgggagctc | 60 |
| tctggctaac tagggaaccc actgcttaag cctcaataaa gcttgccctt agtgcttcaa | 120 |
| gtagtggtg cccgtctgtt gtgtgactct ggtaactaga gatccctcag acccttttag | 180 |
| | |
| tcagtgtgga aaatctctag cagagcccg ttaaaccgc tgatcagcct cgactgtgcc | 240 |
| ttctagtgc cagccatctg ttgtttgccc ctccccgtg ccttccttga cctggaagg | 300 |
| tgccactccc actgtccttt cctaataaaa tgaggaaatt gcatcgatt gtctgagtag | 360 |
| gtgtcattct attaccaag gaagccttag ataagata | 398 |

<210> 90

<211> 225

<212> DNA

<213> Artificial Sequence

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

polynucleotide

<400> 90

| | |
|---|-----|
| agctgcatcc ggagcactac aaagactgct gacatcgagc ttactacaa agactgctga | 60 |
| | |
| catcgagctt taattcactc ccaaagaagt caagatctgc tttttgctg tactgggtct | 120 |
| ctctggtttag accagagtct ctctggtttag accagatctg agcctgggag ctctctggct | 180 |
| aactagggaa cccactgctt aaccaaggaa gccttagata agata | 225 |