(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 September 2016 (29.09.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/154016 A2

(51) International Patent Classification:
A61K 48/00 (2006.01)

(21) International Application Number:
PCT/US2016/023170

(22) International Filing Date:
18 March 2016 (18.03.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/136,080 20 March 2015 (20.03.2015) US

(71) Applicant: **TEMPLE UNIVERSITY OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION** [US/US]; Broad Street And Montgomery Avenue, Philadelphia, PA 19122 (US).

(72) Inventor: **KHALILI, Kamel**; 190 Presidential Boulevard, #718, Bala Cynwyd, PA 19004 (US).

(74) Agents: **ZACHARIADES, Nicholas, A.** et al.; Mchale & Slavin, P.A., 2855 PGA Boulevard, Palm Beach Gardens, FL 33410 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*

WO 2016/154016 A2

(54) **Title**: TAT-INDUCED CRISPR/ENDONUCLEASE-BASED GENE EDITING

(57) **Abstract**: Compositions and methods are provided for Tat-inducible expression of a CRISPR-associated endonuclease by a truncated HIV LTR promoter containing at least a core region and a TAR region of a HIV LTR promoter. The compositions may be used as a therapeutic treatment for the treatment and/or prevention of HIV.

TAT-INDUCED CRISPR/ENDONUCLEASE-BASED GENE EDITING

FIELD OF THE INVENTION

The present invention relates to compositions that specifically cleave target sequences in retroviruses, for example human immunodeficiency virus (HIV). Such compositions, which can include nucleic acids encoding a Clustered Regularly Interspaced Short Palindromic Repeat (CRISPR) associated endonuclease and a guide RNA sequence complementary to a target sequence in a human immunodeficiency virus, can be administered to a subject having or at risk for contracting an HIV infection.

BACKGROUND OF THE INVENTION

Since the discovery of HIV-1, AIDS remains a major public health problem affecting millions of people worldwide. AIDS remains incurable due to the permanent integration of HIV-1 into the host genome. Current therapy (highly active antiretroviral therapy or HAART) for controlling HIV-1 infection and impeding AIDS development profoundly reduces viral replication in cells that support HIV-1 infection and reduces plasma viremia to a minimal level. But HAART fails to suppress low level viral genome expression and replication in tissues and fails to target latently-infected cells, for example, resting memory T cells, brain macrophages, microglia, and astrocytes, gut-associated lymphoid cells, that serve as a reservoir for HIV-1. Persistent HIV-1 infection is also linked to comorbidities including heart and renal diseases, osteopenia, and neurological disorders. There is a continuing need for curative therapeutic strategies that target persistent viral reservoirs.

The HIV-1 genome is about 9.8 kb in length, including two viral long-terminal repeats located at both ends when integrated into the host genome. The genome also includes genes that encode for the structural proteins Gag, Pol, and Env, regulatory proteins (Tat and Rev), and accessory proteins Vpu, Vpr, Vif, and Nef. The HIV-1 transactivator of transcription (Tat) is a multifunctional protein that has been proposed to contribute to several pathological consequences of HIV-1 infection. Tat not only plays an important role in viral transcription and replication, it is also capable of inducing the expression of a variety of cellular genes as well as acting as a neurotoxic protein. Tat protein is secreted by HIV-1-infected cells and acts by

diffusing through the cell membrane. It may act as a secreted, soluble neurotoxin and induces HIV-1-infected macrophages and microglia to release neurotoxic substances. Tat transcription is driven by the HIV-1 LTR promoter and is required for overall viral replication of HIV.

The clinical course of HIV infection can vary according to a number of factors, including the subject's genetic background, age, general health, nutrition, treatment received, and the HIV subtype. In general, most individuals develop flu-like symptoms within a few weeks or months of infection. The symptoms can include fever, headache, muscle aches, rash, chills, sore throat, mouth or genital ulcers, swollen lymph glands, joint pain, night sweats, and diarrhea. The intensity of the symptoms can vary from mild to severe depending upon the individual. During the acute phase, the HIV viral particles are attracted to and enter cells expressing the appropriate CD4 receptor molecules. Once the virus has entered the host cell, the HIV encoded reverse transcriptase generates a proviral DNA copy of the HIV RNA and the proviral DNA becomes integrated into the host cell genomic DNA. It is this HIV provirus that is replicated by the host cell, resulting in the release of new HIV virions which can then infect other cells.

The primary HIV infection subsides within a few weeks to a few months, and is typically followed by a long clinical "latent" period which may last for up to 10 years. The latent period is also referred to as asymptomatic HIV infection or chronic HIV infection. The subject's CD4 lymphocyte numbers rebound, but not to pre-infection levels and most subjects undergo seroconversion, that is, they have detectable levels of anti-HIV antibody in their blood, within 2 to 4 weeks of infection. During this latent period, there can be no detectable viral replication in peripheral blood mononuclear cells and little or no culturable virus in peripheral blood. During the latent period, also referred to as the clinical latency stage, people who are infected with HIV may experience no HIV-related symptoms, or only mild ones. But, the HIV virus continues to reproduce at very low levels. In subjects who have been treated with anti-retroviral therapies, this latent period may extend for several decades or more. However, subjects at this stage are still able to transmit HIV to others even if they are receiving antiretroviral therapy, although antiretroviral therapy reduces the risk of transmission.

CRISPRs (clustered regularly interspaced short palindromic repeats) are DNA loci containing short repetitions of base sequences. Each repetition is followed by short segments of "spacer DNA" from previous exposures to a virus. CRISPRs are often associated with Cas genes that code for proteins related to CRISPRs. The CRISPR/Cas system is a prokaryotic immune system that confers resistance to foreign genetic elements such as plasmids and phages and provides a form of acquired immunity. CRISPR spacers recognize and cut these exogenous genetic elements in a manner analogous to RNAi in eukaryotic organisms.

The CRISPR/Cas system has been used for gene editing (by adding, disrupting or changing the sequence of specific genes) and gene regulation in various organisms. By delivering the Cas9 protein and appropriate guide RNAs into a cell, the organism's genome can be cut at any desired location. Successful therapeutic gene editing using CRISPR/Cas9 system requires efficient and specific delivery and expression of Cas9 enzyme and guide RNAs in target cells. This is particularly challenging when the frequency of recipient cells in a tissue or cell population is low, such as in the case of certain virus-infected cells.

SUMMARY

Provided herein is a method of inactivating a human immunodeficiency virus (HIV) in a mammalian cell *in vivo* or *in vitro*. The method includes exposing the mammalian cell to a composition that includes an isolated nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated HIV LTR promoter containing at least a core region and a TAR region of a HIV LTR promoter.

In specific embodiments, the CRISPR-associated endonuclease is Cas9. The CRISPR-associated endonuclease may be optimized for expression in a human cell. The exposing of the mammalian cell to the composition may include contacting the cell. The mammalian cell may be a latently infected cell including, but not limited to, a $CD4^+$ T cell, a macrophage, a monocyte, a gut-associated lymphoid cell, a microglial cell, and an astrocyte. The mammalian cell may include a cultured cell from a subject having a HIV infection, a tissue explant, and/or a cell line. The inactivating of HIV may be performed *in vivo* or *ex vivo*.

In specific embodiments, the isolated nucleic acid may additionally encode one or more guide RNAs that are complementary to a target nucleic acid sequence in HIV. The target nucleic acid sequence in HIV may refer to a sequence within a coding and/or noncoding region and/or the long terminal repeat of HIV. The non-coding region may include a long terminal repeat of HIV. A sequence within the long terminal repeat of HIV may include a sequence within U3, R, or U5 regions that excludes any sequence of the truncated HIV LTR promoter. The composition may include a sequence that encodes a nuclear localization signal. The composition may additionally include a sequence encoding a transactivating small RNA (tracrRNA) and the tracrRNA may be fused to a sequence encoding a guide RNA. The composition may also include an enhancer region of the HIV LTR promoter.

In specific embodiments, the composition may be operably linked to an expression vector. The expression vector may be a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

Provided herein is an isolated nucleic acid sequence that includes a sequence encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing at least a core region and a TAR region of a HIV LTR promoter.

In specific embodiments, the CRISPR-associated endonuclease is Cas9. The CRISPR-associated endonuclease may be optimized for expression in a human cell.

In specific embodiments, the sequence may additionally encode one or more guide RNAs that are complementary to a target nucleic acid sequence in HIV. The target nucleic acid sequence in HIV may refer to a sequence within a coding and/or noncoding region and/or the long terminal repeat of HIV. The long terminal repeat of HIV may include a sequence within the U3, R, or U5 regions that excludes any sequence of the truncated HIV LTR promoter. The isolated nucleic acid sequence may also encode a nuclear localization signal and/or a transactivating small RNA (tracrRNA). The tracrRNA may be fused to a sequence encoding a guide RNA. The isolated nucleic acid sequence may also include an enhancer region of the HIV LTR promoter.

In specific embodiments, the isolated nucleic acid sequence may be operably linked to an expression vector. The expression vector may refer to a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

Provided herein is a pharmaceutical composition that includes a sequence encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing at least a core region and a TAR region of a HIV LTR promoter. The pharmaceutical composition may also include a pharmaceutically acceptable carrier including, but not limited to, a lipid-based or polymer-based colloid. The colloid may be a liposome, a hydrogel, a microparticle, a nanoparticle, or a block copolymer micelle. In specific embodiments, the CRISPR-associated endonuclease is Cas9. The CRISPR-associated endonuclease may be optimized for expression in a human cell.

In specific embodiments, the pharmaceutical composition may be formulated for topical application and/or contained within a condom.

In specific embodiments, the sequence may additionally encode one or more guide RNAs that are complementary to a target nucleic acid sequence in HIV. The target nucleic acid sequence in HIV may refer to a sequence within a coding and/or noncoding region and/or the long terminal repeat of HIV. The sequence within the long terminal repeat of HIV may include a sequence within the U3, R, or U5 regions that excludes any sequence of the truncated HIV LTR promoter. The sequence may encode a nuclear localization signal. The pharmaceutical composition may additionally include a sequence encoding a tracrRNA and the tracrRNA may be fused to a sequence encoding a guide RNA. The sequence may also encode an enhancer region of the HIV LTR promoter.

In specific embodiments, the sequence provided by the pharmaceutical composition may be operably linked to an expression vector. The expression vector may be a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

Provided herein is a method of treating a subject having a HIV infection. The method includes administering to the subject a therapeutically effective amount of a pharmaceutical composition that includes a sequence encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing

at least a core region and a TAR region of a HIV LTR promoter. The HIV infection being treated may be a latent infection. The method may further include identifying a subject having a HIV infection.

In specific embodiments, the CRISPR-associated endonuclease is Cas9. The CRISPR-associated endonuclease may be optimized for expression in a human cell.

In specific embodiments, the sequence may additionally encode one or more guide RNAs that are complementary to a target nucleic acid sequence in HIV. In some instances, the sequence may encode an enhancer region of the HIV LTR promoter.

In specific embodiments, an anti-retroviral agent may be administered. The anti-retroviral agent may include, but is not limited to, non-nucleoside reverse transcriptase inhibitors, protease inhibitors, and entry inhibitors. The anti-retroviral agent may include highly active antiretroviral therapy. The pharmaceutical composition may be administered topically or parenterally.

In specific embodiments, the pharmaceutical composition may be operably linked to an expression vector. The expression vector may be a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

Provided herein is a method of reducing risk of a HIV infection in a subject at risk for the HIV infection. The method may include administering to the subject a therapeutically effective amount of a pharmaceutical composition comprising a sequence encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing at least a core region and a TAR region of a HIV LTR promoter. In an embodiment, the subject is sexually active, a health care worker, and/or a first responder.

In specific embodiments, the CRISPR-associated endonuclease may be Cas9. The CRISPR-associated endonuclease may be optimized for expression in a human cell. In some instances, the pharmaceutical composition may be operably linked to an expression vector. The expression vector may be, without limitation, a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector. In an

embodiment, the sequence may also encode an enhancer region of the HIV LTR promoter.

Provided herein is a method of reducing risk of transmission of a HIV infection from a HIV-infected gestating or lactating mother to her offspring. The method includes administering to the subject a therapeutically effective amount of a pharmaceutical composition that includes a sequence encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing at least a core region and a TAR region of a HIV LTR promoter. In an embodiment, the pharmaceutical composition is administered during one or more of: prenatally, perinatally, and postnatally.

In specific embodiments, an anti-retroviral agent may be administered. The anti-retroviral agent may be, without limitation, non-nucleoside reverse transcriptase inhibitors, protease inhibitors, and entry inhibitors. The anti-retroviral agent may be highly active antiretroviral therapy. In an embodiment, a therapeutically effective amount of the composition may be administered to the offspring. In an embodiment, the sequence may also encode an enhancer region of the HIV LTR promoter.

Provided herein is a method of administering a pharmaceutical composition to prevent infection by a HIV in an uninfected subject. The method may include administering to the uninfected subject a therapeutically effective amount of the pharmaceutical composition that includes a sequence encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing at least a core region of a HIV LTR promoter and a TAR region of a HIV LTR promoter.

Provided herein is a kit that includes a measured amount of a composition that includes an isolated nucleic acid sequence that includes a sequence encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing at least a core region and a TAR region of a HIV LTR promoter, or a vector encoding the isolated nucleic acid, and one or more items of: packaging material, a package insert including instructions for use, a sterile fluid, a syringe, and a sterile container.

As envisioned in the present invention with respect to the disclosed compositions of matter and methods, in one aspect the embodiments of the invention comprise the components and/or steps disclosed herein. In another aspect, the embodiments of the invention consist essentially of the components and/or steps disclosed herein. In yet another aspect, the embodiments of the invention consist of the components and/or steps disclosed herein.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1A is a schematic representation of full length HIV-1 LTR (LTR(-454/+66)), and created LTR truncation variants LTR -120/+66, LTR -80/+66 and LTR -38/+66. The LTR elements contained in each variant are apparent from the figure.

Fig. 1B is an agarose gel electrophoresis image of PCR-amplified LTR sequences of full length HIV-1 LTR and the variants of Fig. 1A. Lane 1: full length HIV-1 LTR (pLTR(-454/+66)). Lane 2: pLTR (-120/+66). Lane 3: pLTR (-80/+66. Lane 4: pLTR (-38/+66).

Fig. 2 is a diagram of a Cas9 promoter replacement procedure according to the present invention. pX260-U6-DR-BB-DR-Cbh-NLS-hSpCas9-NLS-H1-shorttracr-PGK-puro plasmid (Addgene #42229) (marked as "CBh-Cas9"), was used as a Cas9 gene source/template. The original CBh promoter in the reference plasmid was removed by restriction enzyme digestion with the enzymes indicated in the figure and replaced with different HIV-1 LTR promoter variants (marked collectively as "LTR-Cas9").

Fig. 3A is a Western blot of Cas9, Tat and α-tubulin expression in U87 MG cells co-transfected with different amounts of plasmids expressing FLAG-labeled Cas9 under control of full length HIV-1 LTR (pLTR(-454/+66)-FLAG-Cas9) (10, 50 and 250ng), with or without Tat expressing plasmid (pCMV-Tat86, 250ng). Lane 1: pLTR(-454/+66)-Cas9 250ng, pCMV 1000ng. Lane 2: pLTR(-454/+66)-Cas9 50ng, pCMV 1200ng. Lane 3: pLTR(-454/+66)-Cas9 10ng, pCMV 1240ng. Lane 4: pLTR(-454/+66)-Cas9 250ng, pCMV 750ng, pCMV-Tat86 250ng. Lane 5: pLTR(-454/+66)-Cas9 50ng, pCMV 950ng, pCMV-Tat86 250ng. Lane 6: pLTR(-454/+66)-Cas9 10ng, pCMV 990ng, pCMV-Tat86 250ng.

Fig. 3B comprise graphs of the intensity of the bands corresponding to Cas9 and normalized to α-tubulin expression in the Western blot of Fig. 3A. The top panel show the Western blot image quantification of the Cas9 levels normalized to α-tubulin levels, with or without Tat. The bottom panel show the Western blot image quantification of the +Tat/no Tat ratio.

Fig. 4A is a Western blot of Cas9, Tat and α-tubulin expression in U87 MG cells transfected with different amounts of plasmids (5ng or 50ng) expressing FLAG-labeled Cas9 under control of the HIV-1 truncated LTR variant pLTR(-120/+66)-FLAG-Cas9 or the HIV-1 LTR variant pLTR(-80/+66)-FLAG-Cas9, with or without Tat expressing plasmid (pCMV-Tat86, 250ng). Lane 1: pLTR(-120/+66)-Cas9 5ng, pCMV 1245ng. Lane 2: pLTR(-120/+66)-Cas9 5ng, pCMV 1245ng, +rTat protein 2.5 μg/ml. Lane 3: pLTR(-120/+66)-Cas9 5ng, pCMV 995 ng, pCMV-Tat86 250ng. Lane 4: pLTR(-120/+66)-Cas9 50ng, pCMV 1200ng. Lane 5: pLTR(-120/+66)-Cas9 50ng, pCMV 1200ng, +rTat protein 2.5 μg/ml. Lane 6: pLTR(-120/+66)-Cas9 50ng, pCMV 950 ng, pCMV-Tat86 250ng. Lane 7: pLTR(-80/+66)-Cas9 5ng, pCMV 1245ng. Lane 8: pLTR(-80/+66)-Cas9 5ng, pCMV 1245ng, +rTat protein 2.5 μg/ml. Lane 9: pLTR(-80/+66)-Cas9 5ng, pCMV 995 ng, pCMV-Tat86 250ng. Lane 10: pLTR(-80/+66)-Cas9 50ng, pCMV 1200ng. Lane 11: pLTR(-80/+66)-Cas9 50ng, pCMV 1200ng, +rTat protein 2.5 μg/ml. Lane 12: pLTR(-80/+66)-Cas9 50ng, pCMV 950 ng, pCMV-Tat86 250ng.

Fig. 4B comprise graphs of the intensity of the bands corresponding to Cas9 normalized to α-tubulin expression in the Western blot of Fig, 4A. The top panel show the Western blot image quantification of the Cas9 levels normalized to α-tubulin levels, with no Tat, with rTAT or with transfected Tat. The bottom panel show the Western blot image quantification of the +Tat(transfected)/no Tat ratio.

Figs 5A-5E depict the expression of Cas9 by the HIV-1 LTR promoter is stimulated by Tat leading to cleavage of the viral promoter in the presence of gRNA. Fig. 5A: Schematic presentation of the full-length HIV-1 LTR and the various regulatory motifs within the enhancer and core regions, and the partial Gag gene. The extent of LTR deletion mutants that are created for expression of Cas9 is depicted. The positions of the gRNA target sequence and their distance from each other is

shown. Fig 5B: Co-transfection of TZMb1 cells with pX260-LTR-Cas9 containing the full-length LTR (-454/+66) or its various mutants (-120/+66 or -80/+66) along with a plasmid expressing Tat (pCMV-Tat) increased the level of Tat production as tested by Western blot (top panel). Expression of housekeeping α-tubulin (middle panel) and Tat (bottom panel) are shown. Fig. 5C: Infection of TZMb1 cells with adenovirus expressing Tat, at two different multiplicities of infection (MOI) followed by lentivirus mediated expression of Cas9 by the $LTR_{-80/+66}$ promoter and gRNAs A/B by the U6 promoter led to cleavage of the integrated HIV-1 LTR promoter DNA sequence and the appearance of a 205 bp DNA fragment in the TZMb1 cells (as tested by PCR and DNA gel analysis). Fig. 5D: SDS-PAGE illustrating the level of Cas9, β-tubulin and Tat protein expressed in TZMb1 cells as described in Fig. 5C. Fig. 5E: Luciferase assay illustrating transcriptional activity of the integrated HIV-1 LTR in TZMb1 cells after various treatments as described in Fig. 5 C.

Figs. 6A-6C show that HIV-1 infection stimulates cleavage of integrated viral DNA upon induction of Cas9. The $LTR_{-80/+66}$-Cas9 reporter TZMb1 cell line transduced with lentivirus expressing gRNA A/B (LV-gRNA A/B) or control (empty LV) was infected with three different MOI of $HIV\text{-}1_{JRFL}$ or $HIV\text{-}1_{SF162}$, and after 48 hours, cells were harvested and protein expression was determined by Western blot (Fig. 6A), the level of integrated HIV-1 LTR cleavage upon induction of Cas9 after viral infection was detected by PCR/DNA gene analysis (Fig. 6B) and transcriptional activity of the integrated HIV-1 promoter was evaluated by luciferase reporter assay (Fig. 6C).

Figs. 7A-7C show that Tat stimulation of Cas9 cleaves integrated HIV-1 DNA in T-cells encompassing the HIV-1 reporter at a latent stage. $CD4^+$ Jurkat T-cells, 2D10 cells, containing $LTR_{-80/+66}$-Cas9 gene were transduced with control (empty LV) or LV-gRNA A/B followed by transfection with pCMV or pCMV-Tat plasmids. After 48 hours, the level of various proteins, as depicted, was determined by Western blot (Fig. 7A). The genomic DNA for assessing the state of the integrated HIV-1 DNA was determined by LTR specific PCR and the excision efficiency was determined as a percentage of ratios between truncated vs. full-length amplicon (Fig 7B). The level of integrated viral promoter reactivation after cleavage was assessed by

flow cytometry and the representative scatter plots are shown (Fig. 7C). Red positive, propidium iodide stained, and dead cells were excluded from the analysis.

Figs. 8A-8C show that treatment of cells with latency reversing drugs induces Cas9 expression and cleavage of integrated viral DNA in Jurkat 2D10 cells. 2D10 cells expressing $LTR_{80/+66}$-Cas9 were treated with control (empty) or lentivirus expressing gRNAs A/B and 24 hours later they were treated with PMA (P), TSA (T) or both (P/T) for 16 hours. Protein studies for the expression of Cas9-Flag, α-tubulin and GFP (indicative of the integrated HIV-1 genome) was determined by Western blot (Fig. 8A). Genomic DNA for the detection of the level of excision within the integrated LTR DNA by Cas9 and gRNA A/B was assessed by PCR and the excision efficiency was determined as described in Fig. 7A-7C legend (Fig. 8B). GFP reporter assay, by flow cytometry, and representative scatter plot is shown (Fig. 8C).

FIG. 9 is a schematic representation of negative feedback regulation of HIV-1 by CRISPR/Cas9. At the early stage of (reactivation), basal transcription of the viral genome allows production of Tat protein ①. Upon association of TAT with the budge sequence of the viral transcript ② and recruitment of several cellular protein to associate with the loop of TAR and other transcription factors at RNA poly II in close proximity of the transcription start site, transcription of viral RNA is highly stimulated at the initiation and more importantly, elongation ③. The basal product upon viral activation also stimulates the minimum viral promoter, *ltr*, driving the Cas9 gene ③. The newly synthesized Cas9 upon association with various HIV-1 specific gRNAs, cleaves the viral genome and permanently inactivates the LTR and shuts down HIV-1 gene expression and replication. In the absence of Tat, ltr-Cas9 becomes silent. Expression of Cas9 can continue only in the presence of Tat.

Fig. 10 shows the position and nucleotide sequences of gRNA A/B targets within the LTR (highlighted in green, PAM in red) and LTR specific primers used in PCR on TZMb1 genomic DNA (highlighted in blue) in the reference HIV-1 NL4-3 genome. Sequences and sizes of LTR specific PCR products (full-length and truncated) and predicted edited fragment (SEQ ID NOS: 6-10).

Fig. 11 shows a representative agarose gel analyzing LTR specific PCR reactions used for quantification of Cas9/gRNA mediated LTR excision efficiency in experiments using the Jurkat 2D10 reporter cell line from Figures 7A-7C and 8A-8C.

Fig. 12 shows the position and nucleotide composition of LTR gRNA A/B targets (highlighted in green, PAM in red) and LTR specific primers used to analyze excision by PCR in Jurkat 2D10 cells (highlighted in blue) in the reference HIV-1 NL4-3 genome. Nucleotide sequences and sizes of amplicons (full-length and truncated LTR DNA) and predicted excised DNA fragment are shown (SEQ ID NOS: 11-21).

DEFINITIONS

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although any methods and materials similar or equivalent to those described herein can be used in the practice for testing of the present invention, the preferred materials and methods are described herein. In describing and claiming the present invention, the following terminology will be used.

It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

All genes, gene names, and gene products disclosed herein are intended to correspond to homologs from any species for which the compositions and methods disclosed herein are applicable. It is understood that when a gene or gene product from a particular species is disclosed, this disclosure is intended to be exemplary only, and is not to be interpreted as a limitation unless the context in which it appears clearly indicates. Thus, for example, for the genes or gene products disclosed herein, are intended to encompass homologous and/or orthologous genes and gene products from other species.

The articles "a" and "an" are used herein to refer to one or to more than one (i.e., to at least one) of the grammatical object of the article. By way of example, "an element" means one element or more than one element. Thus, recitation of "a cell", for example, includes a plurality of the cells of the same type. Furthermore, to

the extent that the terms "including", "includes", "having", "has", "with", or variants thereof are used in either the detailed description and/or the claims, such terms are intended to be inclusive in a manner similar to the term "comprising."

As used herein, the terms "comprising," "comprise" or "comprised," and variations thereof, in reference to defined or described elements of an item, composition, apparatus, method, process, system, etc. are meant to be inclusive or open ended, permitting additional elements, thereby indicating that the defined or described item, composition, apparatus, method, process, system, etc. includes those specified elements--or, as appropriate, equivalents thereof--and that other elements can be included and still fall within the scope/definition of the defined item, composition, apparatus, method, process, system, etc.

"About" as used herein when referring to a measurable value such as an amount, a temporal duration, and the like, is meant to encompass variations of +/- 20%, +/- 10%, +/- 5%, +/- 1%, or +/- 0.1% from the specified value, as such variations are appropriate to perform the disclosed methods. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude within 5-fold, and also within 2-fold, of a value. Where particular values are described in the application and claims, unless otherwise stated the term "about" meaning within an acceptable error range for the particular value should be assumed.

An "effective amount" as used herein, means an amount which provides a therapeutic or prophylactic benefit.

"Encoding" refers to the inherent property of specific sequences of nucleotides in a polynucleotide, such as a gene, a cDNA, or an mRNA, to serve as templates for synthesis of other polymers and macromolecules in biological processes having either a defined sequence of nucleotides (i.e., rRNA, tRNA and mRNA) or a defined sequence of amino acids and the biological properties resulting therefrom. Thus, a gene encodes a protein if transcription and translation of mRNA corresponding to that gene produces the protein in a cell or other biological system. Both the coding strand, the nucleotide sequence of which is identical to the mRNA sequence and is usually provided in sequence listings, and the non-coding strand, used

as the template for transcription of a gene or cDNA, can be referred to as encoding the protein or other product of that gene or cDNA.

The term "expression" as used herein is defined as the transcription and/or translation of a particular nucleotide sequence driven by its promoter.

"Expression vector" refers to a vector comprising a recombinant polynucleotide comprising expression control sequences operatively linked to a nucleotide sequence to be expressed. An expression vector comprises sufficient cis-acting elements for expression; other elements for expression can be supplied by the host cell or in an *in vitro* expression system. Expression vectors include all those known in the art, such as cosmids, plasmids (e.g., naked or contained in liposomes) and viruses (e.g., lentiviruses, retroviruses, adenoviruses, and adeno-associated viruses) that incorporate the recombinant polynucleotide.

"Isolated" means altered or removed from the natural state. For example, a nucleic acid or a peptide naturally present in a living animal is not "isolated," but the same nucleic acid or peptide partially or completely separated from the coexisting materials of its natural state is "isolated." An isolated nucleic acid or protein can exist in substantially purified form, or can exist in a non-native environment such as, for example, a host cell.

An "isolated nucleic acid" refers to a nucleic acid segment or fragment which has been separated from sequences which flank it in a naturally occurring state, i.e., a DNA fragment which has been removed from the sequences which are normally adjacent to the fragment, i.e., the sequences adjacent to the fragment in a genome in which it naturally occurs. The term also applies to nucleic acids which have been substantially purified from other components which naturally accompany the nucleic acid, i.e., RNA or DNA or proteins, which naturally accompany it in the cell. The term therefore includes, for example, a recombinant DNA which is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (i.e., as a cDNA or a genomic or cDNA fragment produced by PCR or restriction enzyme digestion) independent of other sequences. It also includes: a recombinant DNA which is part of a hybrid gene encoding additional polypeptide sequence, complementary DNA

(cDNA), linear or circular oligomers or polymers of natural and/or modified monomers or linkages, including deoxyribonucleosides, ribonucleosides, substituted and alpha-anomeric forms thereof, peptide nucleic acids (PNA), locked nucleic acids (LNA), phosphorothioate, methylphosphonate, and the like.

The term "variant," when used in the context of a polynucleotide sequence, may encompass a polynucleotide sequence related to a wild type gene. This definition may also include, for example, "allelic," "splice," "species," or "polymorphic" variants. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing of exons during mRNA processing. The corresponding polypeptide may possess additional functional domains or an absence of domains. Species variants are polynucleotide sequences that vary from one species to another. Of particular utility in the invention are variants of wild type gene products. Variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. Any given natural or recombinant gene may have none, one, or many allelic forms. Common mutational changes that give rise to variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

As used herein, the terms "nucleic acid sequence", "polynucleotide," are used interchangeably throughout the specification and include complementary DNA (cDNA), linear or circular oligomers or polymers of natural and/or modified monomers or linkages, including deoxyribonucleosides, ribonucleosides, substituted and alpha-anomeric forms thereof, peptide nucleic acids (PNA), locked nucleic acids (LNA), phosphorothioate, methylphosphonate, and the like. Polynucleotides include, but are not limited to, all nucleic acid sequences which are obtained by any means available in the art, including, without limitation, recombinant means, i.e., the cloning of nucleic acid sequences from a recombinant library or a cell genome, using ordinary cloning technology and PCR™, and the like, and by synthetic means.

The nucleic acid sequences may be "chimeric," that is, composed of different regions. In the context of this invention "chimeric" compounds are oligonucleotides, which contain two or more chemical regions, for example, DNA region(s), RNA region(s), PNA region(s) etc. Each chemical region is made up of at least one monomer unit, i.e., a nucleotide. These sequences typically comprise at least one region wherein the sequence is modified in order to exhibit one or more desired properties.

The term "target nucleic acid" refers to a nucleic acid (often derived from a biological sample), to which the oligonucleotide is designed to specifically hybridize. It is either the presence or absence of the target nucleic acid that is to be detected, or the amount of the target nucleic acid that is to be quantified. The target nucleic acid has a sequence that is complementary to the nucleic acid sequence of the corresponding oligonucleotide directed to the target. The term target nucleic acid may refer to the specific subsequence of a larger nucleic acid to which the oligonucleotide is directed or to the overall sequence (e.g., gene or mRNA). The difference in usage will be apparent from context.

In the context of the present invention, the following abbreviations for the commonly occurring nucleic acid bases are used, "A" refers to adenosine, "C" refers to cytosine, "G" refers to guanosine, "T" refers to thymidine, and "U" refers to uridine.

Unless otherwise specified, a "nucleotide sequence encoding an amino acid sequence" includes all nucleotide sequences that are degenerate versions of each other and that encode the same amino acid sequence. The phrase nucleotide sequence that encodes a protein or an RNA may also include introns to the extent that the nucleotide sequence encoding the protein may in some version contain an intron(s).

A "lentivirus" as used herein refers to a genus of the Retroviridae family. Lentiviruses are unique among the retroviruses in being able to infect non-dividing cells; they can deliver a significant amount of genetic information into the DNA of the host cell, so they are one of the most efficient methods of a gene delivery vector. HIV, SIV, and FIV are all examples of lentiviruses. Vectors derived from lentiviruses offer the means to achieve significant levels of gene transfer *in vivo*.

"Parenteral" administration of an immunogenic composition includes, e.g., subcutaneous (s.c.), intravenous (i.v.), intramuscular (i.m.), or intrasternal injection, or infusion techniques.

The terms "patient" or "individual" or "subject" are used interchangeably herein, and refers to a mammalian subject to be treated, with human patients being preferred. In some cases, the methods of the invention find use in experimental animals, in veterinary application, and in the development of animal models for disease, including, but not limited to, rodents including mice, rats, and hamsters, and primates.

The term "polynucleotide" is a chain of nucleotides, also known as a "nucleic acid". As used herein polynucleotides include, but are not limited to, all nucleic acid sequences which are obtained by any means available in the art, and include both naturally occurring and synthetic nucleic acids.

The terms "peptide," "polypeptide," and "protein" are used interchangeably, and refer to a compound comprised of amino acid residues covalently linked by peptide bonds. A protein or peptide must contain at least two amino acids, and no limitation is placed on the maximum number of amino acids that can comprise a protein's or peptide's sequence. Polypeptides include any peptide or protein comprising two or more amino acids joined to each other by peptide bonds. As used herein, the term refers to both short chains, which also commonly are referred to in the art as peptides, oligopeptides and oligomers, for example, and to longer chains, which generally are referred to in the art as proteins, of which there are many types. "Polypeptides" include, for example, biologically active fragments, substantially homologous polypeptides, oligopeptides, homodimers, heterodimers, variants of polypeptides, modified polypeptides, derivatives, analogs, fusion proteins, among others. The polypeptides include natural peptides, recombinant peptides, synthetic peptides, or a combination thereof.

The term "promoter" means a DNA sequence recognized by the synthetic machinery of the cell, or introduced synthetic machinery, required to initiate the specific transcription of a polynucleotide sequence. A "minimal" promoter or "truncated" promoter or "functional fragment" of a promoter includes all essential

elements of a promoter for transcriptional activation of, for example, a nucleic acid sequence operably linked or under control of the minimal promoter. In one embodiment, a truncated HIV long terminal repeat (LTR) promoter comprises at least a core region, a trans activation response element (TAR) or combinations thereof, of a HIV LTR promoter.

The term "transfected" or "transformed" or "transduced" means to a process by which exogenous nucleic acid is transferred or introduced into the host cell. A "transfected" or "transformed" or "transduced" cell is one which has been transfected, transformed or transduced with exogenous nucleic acid. The transfected/transformed/transduced cell includes the primary subject cell and its progeny.

To "treat" a disease as the term is used herein, means to reduce the frequency or severity of at least one sign or symptom of a disease or disorder experienced by a subject.

A "vector" is a composition of matter which comprises an isolated nucleic acid and which can be used to deliver the isolated nucleic acid to the interior of a cell. Examples of vectors include but are not limited to, linear polynucleotides, polynucleotides associated with ionic or amphiphilic compounds, plasmids, and viruses. Thus, the term "vector" includes an autonomously replicating plasmid or a virus. The term is also construed to include non-plasmid and non-viral compounds which facilitate transfer of nucleic acid into cells, such as, for example, polylysine compounds, liposomes, and the like. Examples of viral vectors include, but are not limited to, adenoviral vectors, adeno-associated virus vectors, retroviral vectors, and the like.

Ranges: throughout this disclosure, various aspects of the invention can be presented in a range format. It should be understood that the description in range format is merely for convenience and brevity and should not be construed as an inflexible limitation on the scope of the invention. Accordingly, the description of a range should be considered to have specifically disclosed all the possible subranges as well as individual numerical values within that range. For example, description of a range such as from 1 to 6 should be considered to have specifically disclosed

subranges such as from 1 to 3, from 1 to 4, from 1 to 5, from 2 to 4, from 2 to 6, from 3 to 6 etc., as well as individual numbers within that range, for example, 1, 2, 2.7, 3, 4, 5, 5.3, and 6. This applies regardless of the breadth of the range.

Where any amino acid sequence is specifically referred to by a Swiss Prot. or GENBANK Accession number, the sequence is incorporated herein by reference. Information associated with the accession number, such as identification of signal peptide, extracellular domain, transmembrane domain, promoter sequence and translation start, is also incorporated herein in its entirety by reference.

The term "percent sequence identity" refers to the degree of identity between any given query sequence and a subject sequence.

The term "exogenous" indicates that the nucleic acid or polypeptide is part of, or encoded by, a recombinant nucleic acid construct, or is not in its natural environment. For example, an exogenous nucleic acid can be a sequence from one species introduced into another species, i.e., a heterologous nucleic acid. Typically, such an exogenous nucleic acid is introduced into the other species *via* a recombinant nucleic acid construct. An exogenous nucleic acid can also be a sequence that is native to an organism and that has been reintroduced into cells of that organism. An exogenous nucleic acid that includes a native sequence can often be distinguished from the naturally occurring sequence by the presence of non-natural sequences linked to the exogenous nucleic acid, e.g., non-native regulatory sequences flanking a native sequence in a recombinant nucleic acid construct. In addition, stably transformed exogenous nucleic acids typically are integrated at positions other than the position where the native sequence is found.

The terms "pharmaceutically acceptable" (or "pharmacologically acceptable") refer to molecular entities and compositions that do not produce an adverse, allergic or other untoward reaction when administered to an animal or a human, as appropriate. The term "pharmaceutically acceptable carrier," as used herein, includes any and all solvents, dispersion media, coatings, antibacterial, isotonic and absorption delaying agents, buffers, excipients, binders, lubricants, gels, surfactants and the like, that may be used as media for a pharmaceutically acceptable substance.

As used herein, the term "kit" refers to any delivery system for delivering materials. Inclusive of the term "kits" are kits for both research and clinical applications. In the context of reaction assays, such delivery systems include systems that allow for the storage, transport, or delivery of reaction reagents (e.g., oligonucleotides, enzymes, etc. in the appropriate containers) and/or supporting materials (e.g., buffers, written instructions for performing the assay etc.) from one location to another. For example, kits include one or more enclosures (e.g., boxes) containing the relevant reaction reagents and/or supporting materials. As used herein, the term "fragmented kit" refers to delivery systems comprising two or more separate containers that each contains a subportion of the total kit components. The containers may be delivered to the intended recipient together or separately. For example, a first container may contain an enzyme for use in an assay, while a second container contains oligonucleotides or liposomes. The term "fragmented kit" is intended to encompass kits containing Analyte specific reagents (ASR's) regulated under section 520(e) of the Federal Food, Drug, and Cosmetic Act, but are not limited thereto. Indeed, any delivery system comprising two or more separate containers that each contains a subportion of the total kit components are included in the term "fragmented kit." In contrast, a "combined kit" refers to a delivery system containing all of the components of a reaction assay in a single container (e.g., in a single box housing each of the desired components). The term "kit" includes both fragmented and combined kits.

DETAILED DESCRIPTION

Soon after infection with HIV-1, the viral genome becomes integrated into the host chromosome and is rapidly expressed in $CD4^+$ T-cells. HIV-1 replication leads to drastic depletion of $CD4^+$ T-cells. Often, after the acute phase of infection, the virus enters a new phase called latency, where the integrated proviral DNA continues to be expressed and viral replication proceeds at very low levels. Under these circumstances, the weakened immune system caused by persistent viral replication progresses to AIDS and the development of a broad range of opportunistic infections that eventually lead to death within three years if untreated. At the molecular level, expression of the viral genome and its replication both at the acute

and chronic states is controlled by the viral promoter that spans 450 nucleotides of the 5' long terminal region (LTR). Cooperativity occurs between a series of cellular transcriptional factors that recognize DNA sequences within the U3 region of the 5'-LTR and the HIV-1 immediate early transcription activator, Tat, which interacts with the TAR RNA sequence positioned within the leader region of the viral transcript. These interactions are required for the robust initiation and efficient elongation of transcription from integrated copies of the viral DNA. While the current anti-retroviral drugs have been effective in suppressing viral infection cycles, they have yet to contain any components that inhibit viral gene expression at the transcriptional level, supporting the notion that the integrated copies of the virus may continue to express the viral genome, albeit at very low levels, in HIV-1 positive patients under active antiretroviral therapy (ART). Indeed, expression of viral genes drastically elevates upon cessation of ART and allows production of viral early regulatory proteins such as Tat to orchestrate productive replication of the viral genome.

Accordingly, embodiments of the invention are directed to compositions for conditional activation of the CRISPR/Cas at the early stage of reactivation. These compositions completely and permanently ablate virus replication prior to productive viral replication by removing a segment of the viral gene spanning the viral promoter and/or the viral coding sequence. In embodiments, a composition comprises a nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease (CRISPR/Cas) operably linked to a truncated functional viral promoter whereby the truncated viral promoter is under control of an immediate early transcriptional activator, thereby conditionally activating CRISPR/Cas at an early stage of viral replication. The isolated nucleic acid further comprises at least one guide RNA that is complementary to a target nucleic acid sequence in the virus. The CRISPR/Cas excises a segment of a viral genome, for example, the segment spanning a viral promoter and/or viral coding sequence. In these embodiments, the composition is tailored to excise any virus. In certain embodiments, the virus is a retrovirus.

A viral genome, e.g. HIV integrated into an infected host cell's genome may be eliminated from such HIV infected cells utilizing an RNA-guided clustered regularly interspaced short palindromic repeat (CRISPR)-associated endonuclease

such as a Cas9. Successful therapeutic gene editing using CRISPR/Cas9 enzyme and guide RNA requires efficient and specific delivery and expression of Cas9 enzyme and guide RNAs in target cells. This is difficult when the frequency of recipient cells in a tissue or population of cells is low, such as HIV infected cells in patients on highly active antiretroviral therapy (HAART).

According to the present invention, a CRISPR-associated endonuclease such as a Cas9 is placed under the control of a truncated Tat-responsive HIV LTR promoter. The endonuclease expression is thereby activated in cells containing the Tat protein. As demonstrated herein, both exogenously provided (e.g., by transfection) and endogenously produced (e.g., by reactivation of latent virus) Tat can activate (CRISPR)-associated endonuclease (e.g., Cas9) expression in cells lines when expression of the endonuclease is placed under the control of the truncated Tat-responsive HIV LTR promoter. In the studies presented further detail in the examples section, the compositions allow for the conditional activation of the CRISPR/Cas9 at the early stage of viral reactivation by the HIV-1 transcriptional activator, Tat. This strategy completely and permanently ablates virus replication prior to productive viral replication by removing an entire viral genome or a segment of the viral gene spanning the viral promoter and/or the viral coding sequence.

Figure 1A shows a schematic representation of the HIV LTR. It is approximately 640 bp in length. HIV-1 LTR is divided into U3, R, and U5 regions. Transcription of the HIV-1 genome is controlled by a series of cis-acting regulatory motifs spanning the long-terminal region of the viral genome at the 5' end. The U3 region of the viral promoter occupies -1 to -454 nucleotides, with respect to the transcription start site at +1 and has three sub-regions: modulatory, enhancer, and core. The enhancer contains the NF-κB binding site (-127 to -80). The core domain comprises the GC-rich and TATA box (-80 to +1). The R region (+1 to +98) of the LTR comprises TAR, a region for which the expressed RNA forms a stem-loop structure and provides a binding site for the viral transactivator (Krebs *et al*, Lentiviral LTR-directed expression, sequence variation, disease pathogenesis. Los Alamos National Laboratory HIV Sequence: Compendium, pp. 29-70.2002).

The LTRs contain all of the required signals for gene expression and are involved in the integration of a provirus into the genome of a host cell. For example, the core promoter, an enhancer, and a modulatory region are found within U3 while the TAR is found within R as shown in Figure 1A. TAR, the binding site for Tat protein and for cellular proteins, consists of approximately the first 45 nucleotides of the viral mRNAs in HIV-1 forms a hairpin stem-loop structure. In HIV-1, the U5 region includes several sub-regions, for example, including Poly A which is involved in dimerization and genome packaging, PBS or primer binding site, Psi or the packaging signal, and DIS or dimer initiation site.

According to the present invention, a composition is provided comprising an isolated nucleic acid encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing at least the core region and the TAR (transactivation response element) region of HIV LTR promoter. A truncated HIV LTR promoter refers to an operative functional promoter containing less than the full length HIV LTR promoter. Preferably, the truncated promoter contains a core region and a TAR region without all or substantially all of the modulatory and/or enhancer regions. In another embodiment, the truncated HIV LTR promoter contains the core region, the TAR region, and all or substantially all of the enhancer region, but does not contain any of the modulatory region. The truncated HIV LTR promoter is responsive to Tat protein. That is, Tat can activate the expression of the CRISPR-associated endonuclease, such as Cas9, operably linked to the truncated HIV LTR promoter. The disclosed composition may be utilized to inactivate HIV in a mammalian cell, treat a subject having a HIV infection, reduce the risk of HIV infection in a subject at risk for infection, and/or reduce the risk of transmission of HIV from a HIV-infected mother to her offspring. The therapeutic methods disclosed herein may be carried out in connection with other antiretroviral therapies such as HAART. The composition may be included as a part of a kit for diagnostic, research, and/or therapeutic applications.

Anti-retroviral therapy does not suppress low levels of viral genome expression nor does it efficiently target latently infected cells such as resting memory T cells, monocytes, macrophages, microglia, astrocytes, and gut associated lymphoid cells as described earlier. However, the methods and compositions disclosed herein

are generally useful for treatment of HIV infected subjects at any stage of infection, or to an uninfected subject who is at risk for HIV infection. In particular, the disclosed methods and compositions are useful for HIV infected subjects who are in the latent period of the infection. Moreover, when a guide RNA is associated with the CRISPR-associated endonuclease operably linked to a truncated, Tat-responsive HIV LTR promoter, as disclosed herein, the HIV genome may be excised from the host cell and eliminated.

Several advantages may be realized with the compositions containing a sequence encoding CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing the core region and the TAR region of HIV LTR promoter. The potential risk of toxic effects caused by the continuous expression may be alleviated and/or eliminated by limiting the expression of the CRISPR-associated endonuclease to cells with HIV gene expression and/or replication. For example, the potential to induce toxicity due to the immunogenicity of the CRISPR-associated endonuclease may be mitigated because of the low and/or intermittent expression of the endonuclease according to the present invention, while at the same time eliminate or cause self-destruction of the HIV genome in infected individuals. In addition, the present invention may provide a prophylactic strategy for at risk individuals because persistent expression of the CRISPR-associated endonuclease is minimized. Thus, the CRISPR-associated endonuclease driven by a truncated, Tat-responsive HIV LTR promoter may be utilized to provide a safe treatment of HIV infected subjects, and to vaccinate uninfected individuals who may be at risk of infection.

In some embodiments the promoter comprises one or more mutations, deletions, insertions, variants, derivatives or combinations thereof. The promoter may also be chimeric, comprising one or chimeric compounds.

Placing the CRISPR-associated endonuclease under control of a truncated HIV LTR promoter, as described herein, is also advantageous because a smaller-sized nucleic acid may be more readily packaged into delivery mechanisms suitable for gene therapy (e.g., retroviruses). Promoter constructs that include the modulatory region, for example, may be less suitable for gene therapy due to their size and/or variable effects on transcription of CRISPR-associated endonuclease. Further, a

composition including only the TATA box of the core region plus the full TAR region of the HIV LTR promoter is unable to adequately express Cas9 (data not shown). Compositions including the entire core region, the TAR region, and optionally the enhancer of the HIV-1 LTR promoter (see Figure 1A) are able to drive Tat-induced expression of Cas9 in a dosage dependent manner.

The truncated HIV-1 LTR promoter may comprise a nucleic acid that includes the nucleotides of positions -80 to +66 of the HIV-1 LTR promoter. In an embodiment, the truncated HIV-1 LTR promoter may comprise a nucleic acid that includes positions -120 to +66 of the HIV-1 LTR promoter. Preferably, the truncated HIV-1 LTR promoter does not contain sequences from the modulatory region.

As disclosed herein, full length and truncated HIV-1 LTR promoter sequences were obtained by PCR using pNL4-3 HIV vector (NIH AIDS Reagent program #114) as a template and the primers shown in the table below:

Primer name	**Sequence**	**SEQ ID NO.**
Kpn1-LTR(-454)-S	5'-**GGTACC**TGGAAGGGCTAATTTGG-3'	SEQ ID NO: 1
Kpn1-LTR(-120)-S	5'-**GGTACC**TCGAGCTTTCTACAAGG-3'	SEQ ID NO: 2
Xba1-LTR(-80)-S	5'-**TCTAGA**GGAGGTGTGGCCTGGGC-3'	SEQ ID NO: 3
Kpn1-LTR(-38)-S	5'-**GGTACC**AGATGCTACATATAAGC-3'	SEQ ID NO: 4
LTR(+66)-**Nco1**-AS	5'-**CCATGG**TAAGCAGTGGGTTCC-3'	SEQ ID NO: 5

The bolded nucleotides in the sequence column correspond to the cleavage site of the restriction enzyme bolded in the respective primer name column. Each

primer was utilized to generate a different-sized segment of the HIV-1 LTR promoter sequence as shown in Figure 1A. For example, LTR -454/+66 includes the entire U3 region and a portion of the R region. In contrast, LTR -80/+66 corresponds to the core region of U3 and the TAR region of R. The LTR -38/+66 nucleotide sequence was unable to adequately drive expression of Cas9 in response to Tat at a detectable level (data not shown).

The truncated HIV-1 LTR promoter of the present invention corresponds to a segment containing the core region of U3 and the TAR. The core region includes the TATA box and a GC rich region that may be a target for SP1. In some configurations, the truncated HIV-1 LTR promoter may include the enhancer at positions -120 to -80 as shown in Figure 1A.

The truncated HIV-1 LTR promoter may be utilized to drive expression of a CRISPR-associated endonuclease such as Cas9. Such endonucleases are described in PCT international application No. PCT/US2014/053441 (WO2015/031775) filed on August 29, 2014 and published on March 5, 2015, the entire disclosure of which is incorporated herein by reference. As described above, the HIV genome integrates into a host genome of an individual infected with HIV. This integrated sequence is then replicated by the host. Even in the latent period, Tat may be produced by the cell. The compositions of the present invention eliminate and/or reduce the presence of the proviral polynucleotides in the host. Because the CRISPR-associated endonuclease is driven by a Tat-responsive promoter according to the present invention, any time Tat is present (e.g., produced by an infected cell), the endonuclease is produced and degrades the nascent polynucleotides. When the virus is not active, no endonuclease is produced. Thus avoided are potential toxic effects that continual expression of the endonuclease may exert on the cell and/or host. Moreover, the amount of endonuclease produced is proportional to the amount of Tat present as described below with respect to Figures 3 and 4.

In certain embodiments, an isolated nucleic acid sequence has at least a 50% sequence similarity to any one of SEQ ID NOS: 1 to 21.

In certain embodiments, an isolated nucleic acid sequence has at least a 70% sequence similarity to any one of SEQ ID NOS: 1 to 21.

In certain embodiments, an isolated nucleic acid sequence has at least a 75% sequence similarity to any one of SEQ ID NOS: 1 to 21.

In certain embodiments, an isolated nucleic acid sequence has at least an 85% sequence similarity to any one of SEQ ID NOS: 1 to 17 to about 95%, 96%, 97%, 98%, or 99% sequence similarity to any one of SEQ ID NOS: 1 to 21.

In certain embodiments, an isolated nucleic acid sequence comprises any one of SEQ ID NOS: 1 to 21 or combinations thereof.

The compositions disclosed herein may include nucleic acids encoding a CRISPR- associated endonuclease, such as Cas9. In some embodiments, one or more guide RNAs that are complementary to a target sequence of HIV may also be encoded. In bacteria, the CRISPR/Cas loci encode RNA-guided adaptive immune systems against mobile genetic elements (viruses, transposable elements and conjugative plasmids). Three types (I-III) of CRISPR systems have been identified. CRISPR clusters contain spacers, the sequences complementary to antecedent mobile elements. CRISPR clusters are transcribed and processed into mature CRISPR RNA (crRNA). The CRISPR-associated endonuclease, Cas9, belongs to the type II CRISPR/Cas system and has strong endonuclease activity to cut target DNA. Cas9 is guided by a mature crRNA that contains about 20 base pairs (bp) of unique target sequence (called spacer) and a trans-activated small RNA (tracrRNA) that serves as a guide for ribonuclease III-aided processing of pre-crRNA. The crRNA:tracrRNA duplex directs Cas9 to target DNA via complementary base pairing between the spacer on the crRNA and the complementary sequence (called protospacer) on the target DNA. Cas9 recognizes a trinucleotide (NGG) protospacer adjacent motif (PAM) to specify the cut site (the 3rd nucleotide from PAM). The crRNA and tracrRNA can be expressed separately or engineered into an artificial fusion small guide RNA (sgRNA) via a synthetic stem loop (AGAAAU) to mimic the natural crRNA/tracrRNA duplex. Such sgRNA, like shRNA, can be synthesized or *in vitro* transcribed for direct RNA transfection or expressed from U6 or H1-promoted RNA expression vector, although cleavage efficiencies of the artificial sgRNA are lower than those for systems with the crRNA and tracrRNA expressed separately.

The CRISPR-associated endonuclease can be a Cas9 nuclease. The Cas9 nuclease can have a nucleotide sequence identical to the wild type *Streptococcus pyogenes* sequence. The CRISPR-associated endonuclease may be a sequence from other species, for example other *Streptococcus species,* such as *thermophiles*. The Cas9 nuclease sequence can be derived from other species including, but not limited to: *Nocardiopsis dassonvillei*, *Streptomyces pristinaespiralis*, *Streptomyces viridochromogenes*, *Streptomyces roseum*, *Alicyclobacillus acidocaldarius*, *Bacillus pseudomycoides*, *Bacillus selenitireducens*, *Exiguobacterium sibiricum*, *Lactobacillus delbrueckii*, *Lactobacillus salivarius*, *Microscilla marina*, *Burkholderiales bacterium*, *Polaromonas naphthalenivorans*, *Polaromonas sp.*, *Crocosphaera watsonii*, *Cyanothece sp.*, *Microcystis aeruginosa*, *Synechococcus sp.*, *Acetohalobium arabaticum*, *Ammonifex degensii*, *Caldicelulosiruptor becscii*, *Candidatus desulforudis*, *Clostridium botulinum*, *Clostridium difficle*, *Finegoldia magna*, *Natranaerobius thermophilus*, *Pelotomaculum thermopropionicum*, *Acidithiobacillus caldus*, *Acidithiobacillus ferrooxidans*, *Allochromatium vinosum*, *Marinobacter sp.*, *Nitrosococcus halophilus*, *Nitrosococcus watsoni*, *Pseudoalteromonas haloplanktis*, *Ktedonobacter racemifer*, *Methanohalobium evestigatum*, *Anabaena variabilis*, *Nodularia spumigena*, *Nostoc sp.*, *Arthrospira maxima*, *Arthrospira platensis*, *Arthrospira sp.*, *Lyngbya sp.*, *Microcoleus chthonoplastes*, *Oscillatoria sp.*, *Petrotoga mobilis*, *Thermosipho africanus*, or *Acaryochloris marina*. *Psuedomona aeruginosa, Escherichia coli,* or other sequenced bacteria genomes and archaea, or other prokaryotic microogranisms may also be a source of the Cas9 sequence utilized in the embodiments disclosed herein.

The wild type *Streptococcus pyogenes* Cas9 sequence can be modified. The nucleic acid sequence can be codon optimized for efficient expression in mammalian cells, i.e., "humanized." sequence can be for example, the Cas9 nuclease sequence encoded by any of the expression vectors listed in Genbank accession numbers KM099231.1 GI:669193757; KM099232.1 GI:669193761; or KM099233.1 GI:669193765. Alternatively, the Cas9 nuclease sequence can be for example, the sequence contained within a commercially available vector such as PX330 or PX260 from Addgene (Cambridge, MA). In some embodiments, the Cas9 endonuclease can have an amino acid sequence that is a variant or a fragment of any of the Cas9

endonuclease sequences of Genbank accession numbers KM099231.1 GI:669193757; KM099232.1 GI:669193761; or KM099233.1 GI:669193765 or Cas9 amino acid sequence of PX330 or PX260 (Addgene, Cambridge, MA). The Cas9 nucleotide sequence can be modified to encode biologically active variants of Cas9, and these variants can have or can include, for example, an amino acid sequence that differs from a wild type Cas9 by virtue of containing one or more mutations (e.g., an addition, deletion, or substitution mutation or a combination of such mutations). One or more of the substitution mutations can be a substitution (e.g., a conservative amino acid substitution). For example, a biologically active variant of a Cas9 polypeptide can have an amino acid sequence with at least or about 50% sequence identity (e.g., at least or about 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, or 99% sequence identity) to a wild type Cas9 polypeptide. Conservative amino acid substitutions typically include substitutions within the following groups: glycine and alanine; valine, isoleucine, and leucine; aspartic acid and glutamic acid; asparagine, glutamine, serine and threonine; lysine, histidine and arginine; and phenylalanine and tyrosine. The amino acid residues in the Cas9 amino acid sequence can be non-naturally occurring amino acid residues. Naturally occurring amino acid residues include those naturally encoded by the genetic code as well as non-standard amino acids (e.g., amino acids having the D-configuration instead of the L-configuration). The present peptides can also include amino acid residues that are modified versions of standard residues (e.g. pyrrolysine can be used in place of lysine and selenocysteine can be used in place of cysteine). Non-naturally occurring amino acid residues are those that have not been found in nature, but that conform to the basic formula of an amino acid and can be incorporated into a peptide. These include D-alloisoleucine(2R,3S)-2-amino-3-methylpentanoic acid and Lcyclopentyl glycine (S)-2-amino-2-cyclopentyl acetic acid. For other examples, one can consult textbooks or the worldwide web (a site currently maintained by the California Institute of Technology displays structures of non-natural amino acids that have been successfully incorporated into functional proteins).

The compositions and methods of the present invention may include a sequence encoding a guide RNA that is complementary to a target sequence in HIV. The genetic variability of HIV is reflected in the multiple groups and subtypes that

have been described. A collection of HIV sequences is compiled in the Los Alamos HIV databases and compendiums (i.e., the sequence database web site is hitp://www.hiv.lani.gov). The methods and compositions of the invention can be applied to HIV from any of those various groups, subtypes, and circulating recombinant forms. These include for example, the HIV-1 major group (often referred to as Group M) and the minor groups, Groups N, 0, and P, as well as but not limited to, any of the following subtypes, A, B, C, D, F, G, H, J and K. or group (for example, but not limited to any of the following Groups, N, 0 and P) of HIV.

The guide RNA can be a sequence complimentary to a coding or a non-coding sequence (i.e., a target sequence). For example, the guide RNA can be a sequence that is complementary to a HIV long terminal repeat (LTR) region other than the portions that are utilized informing the truncated Tat-responsive promoter that is operably linked to the Cas9 gene. The guide RNA cannot target the sequence corresponding to the truncated Tat-responding HIV-1 LTR promoter as disclosed herein because it would result in degradation of the construct itself, thereby potentially removing the advantages rendered by the CRISPR-associated endonuclease driven by the truncated HIV LTR promoter. Thus, a guide RNA can include a sequence found within an HIV-1 U3, R, and/or U5 region reference sequence or consensus sequence, without selecting a sequence that is a part of the truncated Tat-responsive HIV promoter.

In some embodiments, the guide RNA can be a sequence complementary to a coding sequence such as a sequence encoding one or more viral structural proteins (e.g., gag, pol, env, and tat). Thus, the sequence can be complementary to sequence within the gag polyprotein, e.g., MA (matrix protein, p17); CA (capsid protein, p24); NC (nucleocapsid protein, p7); and P6 protein; pol, e.g., reverse transcriptase (RT) and RNase H, integrase (IN), and HIV protease (PR); env, e.g., gp160, or a cleavage product of gp160, e.g., gp120 or SU, and gp4l or TM; or tat, e.g., the 72-amino acid one-exon Tat or the 86-101 amino-acid two-exon Tat. In some embodiments, the guide RNA can be a sequence complementary to a sequence encoding an accessory protein, including for example, vif , n willef (negative factor) vpu (Virus protein U) and tev.

In some embodiments, the guide RNA sequence can be a sequence complementary to a structural or regulatory element (i.e., a target sequence) such as RRE, PE, SLIP, CRS (Cis-acting repressive sequences), and/or INS. RRE (Rev responsive element) is an RNA element encoded within the env region of HIV and includes approximately 200 nucleotides (positions 7710 to 8061 from the start of transcription in HIV-1, spanning the border of gp120 and gp41). PE (Psi element) corresponds to a set of 4 stem-loop structures preceding and overlapping the Gag start codon. SLIP is a TTTTTT "slippery site" followed by a stem-loop structure. CRS (Cis-acting repressive sequences). INS (Inhibitory/Instability RNA sequences) may be found for example, at nucleotides 414 to 631 in the gag region of HIV-1.

The guide RNA sequence can be a sense or anti-sense sequence. The guide RNA sequence generally includes a PAM. The sequence of the PAM can vary depending upon the specificity requirements of the CRISPR endonuclease used. In the CRISPR-Cas system derived from *S. pyogenes,* the target DNA typically immediately precedes a 5'-NGG proto-spacer adjacent motif (PAM). Thus, for the *S. pyogenes* Cas9, the PAM sequence can be AGG, TGG, CGG or GGG. Other Cas9 orthologs may have different PAM specificities. For example, Cas9 from *S. thermophilus* requires 5'-NNAGAA for CRISPR 1 and 5'-NGGNG for CRISPR3) and *Neiseria menigiditis* requires 5'-NNNNGATT). The specific sequence of the guide RNA may vary, but, regardless of the sequence, useful guide RNA sequences will be those that minimize off-target effects while achieving high efficiency and complete ablation of the genomically integrated HIV provirus. The length of the guide RNA sequence can vary from about 20 to about 60 or more nucleotides, for example about 20, about 21, about 22, about 23, about 24, about 25, about 26, about 27, about 28, about 29, about 30, about 31, about 32, about 33, about 34, about 35, about 36, about 37, about 38, about 39, about 40, about 45, about 50, about 55, about 60 or more nucleotides. Useful selection methods identify regions having extremely low homology between the foreign viral genome and host cellular genome including endogenous retroviral DNA, include bioinformatic screening using 12-bp+NGG target-selection criteria to exclude off-target human transcriptome or (even rarely) untranslated-genomic sites; avoiding transcription factor binding sites within the HIV-1 LTR promoter (potentially conserved in the host genome); selection of LTR-A- and -B- directed, 30-

bp guide RNAs and also pre-crRNA system reflecting the original bacterial immune mechanism to enhance specificity/efficiency versus 20-bp guide RNA-, chimeric crRNA-tracRNA-based system and WGS, Sanger sequencing and SURVEYOR assay, to identify and exclude potential off-target effects.

The guide RNA sequence can be configured as a single sequence or as a combination of one or more different sequences, e.g., a multiplex configuration. Multiplex configurations can include combinations of two, three, four, five, six, seven, eight, nine, ten, or more different guide RNAs, for example a combination of sequences in U3, R, or U5, without selecting a sequence that is a part of the truncated Tat-responsive HIV promoter. When the compositions are administered in an expression vector, the guide RNAs can be encoded by a single vector. Alternatively, multiple vectors can be engineered to each include two or more different guide RNAs. Useful configurations will result in the excision of viral sequences between cleavage sites resulting in the ablation of HIV genome or HIV protein expression. Thus, the use of two or more different guide RNAs promotes excision of the viral sequences between the cleavage sites recognized by the CRISPR endonuclease. The excised region can vary in size from a single nucleotide to several thousand nucleotides. Exemplary excised regions are described in the examples.

When the compositions are administered as a nucleic acid or are contained within an expression vector, the CRISPR endonuclease can be encoded by the same nucleic acid or vector as the guide RNA sequences. Alternatively or in addition, the CRISPR endonuclease can be encoded in a physically separate nucleic acid from the guide RNA sequences or in a separate vector. In some embodiments, the RNA molecules e.g. crRNA, tracrRNA, gRNA are engineered to comprise one or more modified nucleobases. For example, known modifications of RNA molecules can be found, for example, in Genes VI, Chapter 9 ("Interpreting the Genetic Code"), Lewis, ed. (1997, Oxford University Press, New York), and Modification and Editing of RNA, Grosjean and Benne, eds. (1998, ASM Press, Washington DC). Modified RNA components include the following: 2'-O-methylcytidine; N^4-methylcytidine; N^4-2'-O-dimethylcytidine; N^4- acetylcytidine; 5-methylcytidine; 5,2'-0-dimethylcytidine; 5-hydroxymethylcytidine; 5- formylcytidine; 2'-0-methyl-5-formaylcytidine; 3-methylcytidine; 2-thiocytidine; lysidine; 2'-0- methyluridine; 2-thiouridine; 2-thio-2'-

O-methyluridine; 3,2'-0-dimethyluridine; 3-(3-amino-3- carboxypropyl)uridine; 4-thiouridine; ribosylthymine; 5,2'-0-dimethyluridine; 5-methyl-2- thiouridine; 5-hydroxyuridine; 5-methoxyuridine; uridine 5-oxyacetic acid; uridine 5-oxyacetic acid methyl ester; 5-carboxymethyluridine; 5-methoxycarbonylmethyluridine; 5-methoxycarbonylmethyl-2'-0-methyluridine; 5-methoxycarbonylmethyl-2'-thiouridine; 5- carbamoylmethyluridine; 5-carbamoylmethyl-2'-0-methyluridine; 5-(carboxyhydroxymethyl)uridine; 5-(carboxyhydroxymethyl) uridinemethyl ester; 5-aminomethyl-2-thiouridine; 5-methylaminomethyluridine; 5-methylaminomethyl-2-thiouridine; 5-methylaminomethyl-2-selenouridine; 5-carboxymethylaminomethyluridine; 5- carboxymethylaminomethyl-2'-0-methyl-uridine; 5-carboxymethylaminomethyl-2-thiouridine; dihydrouridine; dihydroribosylthymine; 2'-methyladenosine; 2-methyladenosine; N.sup.6Nmethyladenosine; N^6, N^6-dimethyladenosine; N^6,2'-O-trimethyladenosine; 2 methylthio-N^6Nisopentenyladenosine; N^6-(cis-hydroxyisopentenyl)-adenosine; 2-methylthio-N^6-(cis-- hydroxyisopenteny1)-adenosine; N^6-glycinylcarbamoyl)adenosine; N^6 threonylcarbamoyl adenosine; N^6-methyl-N^6-threonylcarbamoyl adenosine; 2-methylthio-N^6-methyl-N^6- threonylcarbamoyl adenosine; N^6-hydroxynorvalylcarbamoyl adenosine; 2-methylthio-N^6-hydroxnorvalylcarbamoyl adenosine; 2'-O-ribosyladenosine (phosphate); inosine; 2'O-methyl inosine; 1-methyl inosine; 1;2'-0-dimethyl inosine; 2'-O-methyl guanosine; 1-methyl guanosine; N^2-methyl guanosine; N^2, N^2-dimethyl guanosine; N^2, 2'-O-dimethyl guanosine; N^2, N^2, 2'-0-trimethyl guanosine; 2'-O-ribosyl guanosine (phosphate); 7-methyl guanosine; N^2;7-dimethyl guanosine; N^2; N^2;7-trimethyl guanosine; wyosine; methylwyosine; under-modified hydroxywybutosine; wybutosine; hydroxywybutosine; peroxywybutosine; queuosine; epoxyqueuosine; galactosyl-queuosine; mannosyl-queuosine; 7-cyano-7-deazaguanosine; arachaeosine [also called 7-formamido-7-deazaguanosine]; and 7-aminomethyl-7-deazaguanosine.

Isolated nucleic acid molecules can be produced by standard techniques. For example, PCR techniques can be used to obtain an isolated nucleic acid containing a nucleotide sequence described herein, including nucleotide sequences encoding a polypeptide described herein. PCR can be used to amplify specific sequences from DNA as well as RNA, including sequences from total genomic DNA

or total cellular RNA. Various PCR methods are described in, for example, *PCR Primer: A Laboratory Manual,* Dieffenbach and Dveksler, eds., Cold Spring Harbor Laboratory Press, 1995. Generally, sequence information from the ends of the region of interest or beyond is employed to design oligonucleotide primers that are identical or similar in sequence to opposite strands of the template to be amplified. Various PCR strategies also are available by which site-specific nucleotide sequence modifications can be introduced into a template nucleic acid.

Isolated nucleic acids also can be chemically synthesized, either as a single nucleic acid molecule (e.g., using automated DNA synthesis in the 3' to 5' direction using phosphoramidite technology) or as a series of oligonucleotides. For example, one or more pairs of long oligonucleotides (e.g., >50-100 nucleotides) can be synthesized that contain the desired sequence, with each pair containing a short segment of complementarity (e.g., about 15 nucleotides) such that a duplex is formed when the oligonucleotide pair is annealed. DNA polymerase is used to extend the oligonucleotides, resulting in a single, double-stranded nucleic acid molecule per oligonucleotide pair, which then can be ligated into a vector. Isolated nucleic acids of the invention also can be obtained by mutagenesis of, e.g., a naturally occurring portion of a Cas9 -encoding DNA (in accordance with, for example, the formula above).

Two nucleic acids or the polypeptides they encode may be described as having a certain degree of identity to one another. For example, a Cas9 protein and a biologically active variant thereof may be described as exhibiting a certain degree of identity. Alignments may be assembled by locating short Cas9 sequences in the Protein Information Research (PIR) site (http://pir.georgetown.edu), followed by analysis with the "short nearly identical sequences" Basic Local Alignment Search Tool (BLAST) algorithm on the NCBI website (http://www.ncbi.nlm.nih.gov/blast).

A percent sequence identity to Cas9 can be determined and the identified variants may be utilized as a CRISPR-associated endonuclease and/or assayed for their efficacy as a pharmaceutical composition. A naturally occurring Cas9 can be the query sequence and a fragment of a Cas9 protein can be the subject sequence. Similarly, a fragment of a Cas9 protein can be the query sequence and a biologically

active variant thereof can be the subject sequence. To determine sequence identity, a query nucleic acid or amino acid sequence can be aligned to one or more subject nucleic acid or amino acid sequences, respectively, using the computer program ClustalW (version 1.83, default parameters), which allows alignments of nucleic acid or protein sequences to be carried out across their entire length (global alignment). See Chenna *et al., Nucleic Acids Res.* 31:3497-3500, 2003.

Recombinant constructs are also provided herein and can be used to transform cells in order to express Cas9 under the control of a truncated Tat-responsive HIV LTR promoter. Recombinant constructs may similarly be utilized to express a guide RNA complementary to a target sequence in HIV. A recombinant nucleic acid construct comprises a nucleic acid encoding a Cas9 and/or a guide RNA complementary to a target sequence in HIV as described herein, operably linked to a regulatory region suitable for expressing the Cas9 and/or a guide RNA complementary to a target sequence in HIV in the cell. It will be appreciated that a number of nucleic acids can encode a polypeptide having a particular amino acid sequence. The degeneracy of the genetic code is well known in the art. For many amino acids, there is more than one nucleotide triplet that serves as the codon for the amino acid. For example, codons in the coding sequence for Cas9 can be modified such that optimal expression in a particular organism is obtained, using appropriate codon bias tables for that organism.

Nucleic acids as described herein may be contained in vectors. Vectors can include, for example, origins of replication, scaffold attachment regions (SARs), and/or markers. A marker gene can confer a selectable phenotype on a host cell. For example, a marker can confer biocide resistance, such as resistance to an antibiotic (e.g., kanamycin, G418, bleomycin, or hygromycin). An expression vector can include a tag sequence designed to facilitate manipulation or detection (e.g., purification or localization) of the expressed polypeptide. Tag sequences, such as green fluorescent protein (GFP), glutathione S-transferase (GST), polyhistidine, c-myc, hemagglutinin, or FlagTM tag (Kodak, New Haven, CT) sequences typically are expressed as a fusion with the encoded polypeptide. Such tags can be inserted anywhere within the polypeptide, including at either the carboxyl or amino terminus.

Additional expression vectors also can include, for example, segments of chromosomal, non-chromosomal and synthetic DNA sequences. Suitable vectors include derivatives of SV40 and known bacterial plasmids, e.g., *E. coli* plasmids col El, pCR1, pBR322, pMal-C2, pET, pGEX, pMB9 and their derivatives, plasmids such as RP4; phage DNAs, e.g., the numerous derivatives of phage 1, e.g., NM989, and other phage DNA, e.g., M13 and filamentous single stranded phage DNA; yeast plasmids such as the 2μ plasmid or derivatives thereof, vectors useful in eukaryotic cells, such as vectors useful in insect or mammalian cells; vectors derived from combinations of plasmids and phage DNAs, such as plasmids that have been modified to employ phage DNA or other expression control sequences.

Several delivery methods may be utilized in conjunction with the truncated Tat-responsive HIV LTR promoter operably linked to the Cas9 gene for *in vitro* (cell cultures) and *in vivo* (animals and patients) systems. In one embodiment, a lentiviral gene delivery system may be utilized. Such a system offers stable, long term presence of the gene in dividing and non-dividing cells with broad tropism and the capacity for large DNA inserts. (Dull *et al*, *J Virol*, 72:8463-8471 1998). In an embodiment, adeno-associated virus (AAV) may be utilized as a delivery method. AAV is a non-pathogenic, single-stranded DNA virus that has been actively employed in recent years for delivering therapeutic gene in *in vitro* and *in vivo* systems (Choi *et al*, *Curr Gene Ther*, 5:299-310, 2005). An example non-viral delivery method may utilize nanoparticle technology. This platform has demonstrated utility as a pharmaceutical *in vivo*. Nanotechnology has improved transcytosis of drugs across tight epithelial and endothelial barriers. It offers targeted delivery of its payload to cells and tissues in a specific manner (Allen and Cullis, *Science*, 303:1818-1822, 1998).

The vector can also include a regulatory region. The term "regulatory region" refers to nucleotide sequences that influence transcription or translation initiation and rate, and stability and/or mobility of a transcription or translation product. Regulatory regions include, without limitation, promoter sequences, enhancer sequences, response elements, protein recognition sites, inducible elements, protein binding sequences, 5' and 3' untranslated regions (UTRs), transcriptional start sites, termination sequences, polyadenylation sequences, nuclear localization signals, and introns.

The term "operably linked" refers to positioning of a regulatory region and a sequence to be transcribed in a nucleic acid so as to influence transcription or translation of such a sequence. For example, to bring a coding sequence under the control of a promoter, the translation initiation site of the translational reading frame of the polypeptide is typically positioned between one and about fifty nucleotides downstream of the promoter. A promoter can, however, be positioned as much as about 5,000 nucleotides upstream of the translation initiation site or about 2,000 nucleotides upstream of the transcription start site. A promoter typically comprises at least a core (basal) promoter. A promoter also may include at least one control element, such as an enhancer sequence, an upstream element or an upstream activation region (UAR). The choice of promoters to be included depends upon several factors, including, but not limited to, efficiency, selectability, inducibility, desired expression level, and cell- or tissue-preferential expression. It is a routine matter for one of skill in the art to modulate the expression of a coding sequence by appropriately selecting and positioning promoters and other regulatory regions relative to the coding sequence.

Vectors include, for example, viral vectors (such as adenoviruses Ad, AAV, lentivirus, and vesicular stomatitis virus (VSV) and retroviruses), liposomes and other lipid-containing complexes, and other macromolecular complexes capable of mediating delivery of a polynucleotide to a host cell. Vectors can also comprise other components or functionalities that further modulate gene delivery and/or gene expression, or that otherwise provide beneficial properties to the targeted cells. As described and illustrated in more detail below, such other components include, for example, components that influence binding or targeting to cells (including components that mediate cell-type or tissue-specific binding); components that influence uptake of the vector nucleic acid by the cell; components that influence localization of the polynucleotide within the cell after uptake (such as agents mediating nuclear localization); and components that influence expression of the polynucleotide. Such components also might include markers, such as detectable and/or selectable markers that can be used to detect or select for cells that have taken up and are expressing the nucleic acid delivered by the vector. Such components can be provided as a natural feature of the vector (such as the use of certain viral vectors

which have components or functionalities mediating binding and uptake), or vectors can be modified to provide such functionalities. Other vectors include those described by Chen *et al; BioTechniques,* 34: 167-171 (2003). A large variety of such vectors is known in the art and are generally available. A "recombinant viral vector" refers to a viral vector comprising one or more heterologous gene products or sequences. Since many viral vectors exhibit size-constraints associated with packaging, the heterologous gene products or sequences are typically introduced by replacing one or more portions of the viral genome. Such viruses may become replication-defective, requiring the deleted function(s) to be provided in trans during viral replication and encapsidation (by using, e.g., a helper virus or a packaging cell line carrying gene products necessary for replication and/or encapsidation). Modified viral vectors in which a polynucleotide to be delivered is carried on the outside of the viral particle have also been described (see, e.g., Curiel, D T, *et al. PNAS* 88: 8850-8854, 1991).

Additional vectors include viral vectors, fusion proteins and chemical conjugates. Retroviral vectors include Moloney murine leukemia viruses and HIV-based viruses. One HIV based viral vector comprises at least two vectors wherein the gag and pol genes are from an HIV genome and the env gene is from another virus. DNA viral vectors include pox vectors such as orthopox or avipox vectors, herpesvirus vectors such as a herpes simplex I virus (HSV) vector [Geller, A.I. *et al., J. Neurochem,* 64: 487 (1995); Lim, F., *et al.,* in DNA Cloning: Mammalian Systems, D. Glover, Ed. (Oxford Univ. Press, Oxford England) (1995); Geller, A.I. *et al., Proc Natl. Acad. Sci.:* U.S.A.:90 7603 (1993); Geller, A.I., *et al., Proc Natl. Acad. Sci* USA: 87:1149 (1990)], Adenovirus Vectors [LeGal LaSalle *et al., Science,* 259:988 (1993); Davidson, *et al., Nat. Genet.* 3: 219 (1993); Yang, *et al., J. Virol.* 69: 2004 (1995)] and Adeno-associated Virus Vectors [Kaplitt, M.G., *et al., Nat. Genet.* 8:148 (1994)].

The polynucleotides disclosed herein may be used with a microdelivery vehicle such as cationic liposomes and adenoviral vectors. For a review of the procedures for liposome preparation, targeting and delivery of contents, see Mannino and Gould-Fogerite, *BioTechniques,* 6:682 (1988). See also, Felgner and Holm,

Bethesda Res. Lab. Focus, 11(2):21 (1989) and Maurer, R.A., Bethesda Res. Lab. Focus, 11(2):25 (1989).

Replication-defective recombinant adenoviral vectors, can be produced in accordance with known techniques. See, Quantin, *et al., Proc. Natl. Acad. Sci.* USA, 89:2581-2584 (1992); Stratford-Perricadet, *et al.,* J. Clin. Invest., 90:626-630 (1992); and Rosenfeld, *et al.,* Cell, 68:143-155 (1992).

Another delivery method is to use single stranded DNA producing vectors which can produce the expressed products intracellularly. See for example, Chen *et al, BioTechniques,* 34: 167-171 (2003), which is incorporated herein, by reference, in its entirety.

As described above, the compositions of the present invention can be prepared in a variety of ways known to one of ordinary skill in the art. Regardless of their original source or the manner in which they are obtained, the compositions disclosed herein can be formulated in accordance with their use. For example, the nucleic acids and vectors described above can be formulated within compositions for application to cells in tissue culture or for administration to a patient or subject. Any of the pharmaceutical compositions of the invention can be formulated for use in the preparation of a medicament, and particular uses are indicated below in the context of treatment, e.g., the treatment of a subject having an HIV infection or at risk for contracting and HIV infection. When employed as pharmaceuticals, any of the nucleic acids and vectors can be administered in the form of pharmaceutical compositions. These compositions can be prepared in a manner well known in the pharmaceutical art, and can be administered by a variety of routes, depending upon whether local or systemic treatment is desired and upon the area to be treated. Administration may be topical (including ophthalmic and to mucous membranes including intranasal, vaginal and rectal delivery), pulmonary (e.g., by inhalation or insufflation of powders or aerosols, including by nebulizer; intratracheal, intranasal, epidermal and transdermal), ocular, oral or parenteral. Methods for ocular delivery can include topical administration (eye drops), subconjunctival, periocular or intravitreal injection or introduction by balloon catheter or ophthalmic inserts surgically placed in the conjunctival sac. Parenteral administration includes

intravenous, intraarterial, subcutaneous, intraperitoneal or intramuscular injection or infusion; or intracranial, e.g., intrathecal or intraventricular administration. Parenteral administration can be in the form of a single bolus dose, or may be, for example, by a continuous perfusion pump. Pharmaceutical compositions and formulations for topical administration may include transdermal patches, ointments, lotions, creams, gels, drops, suppositories, sprays, liquids, powders, and the like. Conventional pharmaceutical carriers, aqueous, powder or oily bases, thickeners and the like may be necessary or desirable.

The pharmaceutical compositions may contain, as the active ingredient, nucleic acids and vectors described herein in combination with one or more pharmaceutically acceptable carriers. In making the compositions of the invention, the active ingredient is typically mixed with an excipient, diluted by an excipient or enclosed within such a carrier in the form of, for example, a capsule, tablet, sachet, paper, or other container. When the excipient serves as a diluent, it can be a solid, semisolid, or liquid material (e.g., normal saline), which acts as a vehicle, carrier or medium for the active ingredient. Thus, the compositions can be in the form of tablets, pills, powders, lozenges, sachets, cachets, elixirs, suspensions, emulsions, solutions, syrups, aerosols (as a solid or in a liquid medium), lotions, creams, ointments, gels, soft and hard gelatin capsules, suppositories, sterile injectable solutions, and sterile packaged powders. As is known in the art, the type of diluent can vary depending upon the intended route of administration. The resulting compositions can include additional agents, such as preservatives. In some embodiments, the carrier can be, or can include, a lipid-based or polymer-based colloid. In some embodiments, the carrier material can be a colloid formulated as a liposome, a hydrogel, a microparticle, a nanoparticle, or a block copolymer micelle. As noted, the carrier material can form a capsule, and that material may be a polymer-based colloid.

The nucleic acid sequences of the invention can be delivered to an appropriate cell of a subject. This can be achieved by, for example, the use of a polymeric, biodegradable microparticle or microcapsule delivery vehicle, sized to optimize phagocytosis by phagocytic cells such as macrophages. For example, PLGA (poly-lacto-co-glycolide) microparticles approximately 1-10 μm in diameter can be

used. The polynucleotide is encapsulated in these microparticles, which are taken up by macrophages and gradually biodegraded within the cell, thereby releasing the polynucleotide. Once released, the DNA is expressed within the cell. A second type of microparticle is intended not to be taken up directly by cells, but rather to serve primarily as a slow-release reservoir of nucleic acid that is taken up by cells only upon release from the micro-particle through biodegradation. These polymeric particles should therefore be large enough to preclude phagocytosis (i.e., larger than 5 μm and preferably larger than 20 μm). Another way to achieve uptake of the nucleic acid is using liposomes, prepared by standard methods. The nucleic acids can be incorporated alone into these delivery vehicles or coincorporated with tissue-specific antibodies, for example antibodies that target cell types that are commonly latently infected reservoirs of HIV infection, for example, brain macrophages, microglia, astrocytes, and gut-associated lymphoid cells. Alternatively, one can prepare a molecular complex composed of a plasmid or other vector attached to poly-L-lysine by electrostatic or covalent forces. Poly-L-lysine binds to a ligand that can bind to a receptor on target cells. Delivery of "naked DNA" (i.e., without a delivery vehicle) to an intramuscular, intradermal, or subcutaneous site, is another means to achieve *in vivo* expression. In the relevant polynucleotides (e.g., expression vectors) the nucleic acid sequence encoding an isolated nucleic acid sequence comprising a sequence encoding a CRISPR-associated endonuclease and optionally a guide RNA is operably linked to the truncated Tat-responsive HIV LTR promoter as described above.

In some embodiments, the compositions of the invention can be formulated as a nanoparticle, for example, nanoparticles comprised of a core of high molecular weight linear polyethylenimine (LPEI) complexed with DNA and surrounded by a shell of polyethyleneglycolmodiifed (PEGylated) low molecular weight LPEI.

The nucleic acids and vectors may also be applied to a surface of a device (e.g., a catheter) or contained within a pump, patch, or other drug delivery device. The nucleic acids and vectors disclosed herein can be administered alone, or in a mixture, in the presence of a pharmaceutically acceptable excipient or carrier (e.g., physiological saline). The excipient or carrier is selected on the basis of the mode and route of administration. Suitable pharmaceutical carriers, as well as pharmaceutical

necessities for use in pharmaceutical formulations, are described in Remington's Pharmaceutical Sciences (E. W. Martin), a well-known reference text in this field, and in the USP/NF (United States Pharmacopeia and the National Formulary).

In some embodiments, the compositions may be formulated as a topical gel for blocking sexual transmission of HIV. The topical gel can be applied directly to the skin or mucous membranes of the male or female genital region prior to sexual activity. Alternatively or in addition the topical gel can be applied to the surface or contained within a male or female condom or diaphragm.

In some embodiments, the compositions can be formulated as a nanoparticle encapsulating a nucleic acid encoding Cas9 or a variant Cas9 operably linked to a truncated HIV LTR promoter. The nucleic acid may additionally encode a guide RNA sequence complementary to a target HIV.

The present formulations can encompass a vector encoding Cas9 and a guide RNA sequence complementary to a target HIV. The guide RNA sequence can include a sequence complementary to a single target region or it can include any combination of sequences complementary to multiple target regions as described earlier. Alternatively the sequence encoding Cas9 driven by the truncated HIV LTR promoter and the sequence encoding the guide RNA sequence can be on separate vectors.

The compositions disclosed herein are generally and variously useful for treatment of a subject having an HIV infection. The methods are useful for targeting any HIV, for example, HIV-1 and HIV-2, and also SIV, and any circulating recombinant form thereof. A subject is effectively treated whenever a clinically beneficial result ensues. This may mean, for example, a complete resolution of the symptoms of a disease, a decrease in the severity of the symptoms of the disease, or a slowing of the disease's progression. These methods can further include the steps of a) identifying a subject (e.g., a patient and, more specifically, a human patient) who has an HIV infection; and b) providing to the subject a composition comprising a nucleic acid encoding a CRISPR-associated nuclease, e.g., Cas9, under control of the truncated Tat-responsive HIV LTR promoter. The methods may further include

providing to the subject a sequence encoding a guide RNA complementary to an HIV target sequence, e.g. an HIV LTR.

A subject can be identified using standard clinical tests, for example, immunoassays to detect the presence of HIV antibodies or the HIV polypeptide p24 in the subject's serum, or through HIV nucleic acid amplification assays. An amount of such a composition provided to the subject that results in a complete resolution of the symptoms of the infection, a decrease in the severity of the symptoms of the infection, or a slowing of the infection's progression is considered a therapeutically effective amount. The present methods may also include a monitoring step to help optimize dosing and scheduling as well as predict outcome. In some methods of the present invention, one can first determine whether a patient has a latent HIV infection, and then make a determination as to whether or not to treat the patient with one or more of the compositions described herein. Monitoring can also be used to detect the onset of drug resistance and to rapidly distinguish responsive patients from nonresponsive patients. In some embodiments, the methods can further include the step of determining the nucleic acid sequence of the particular HIV harbored by the patient and then designing the guide RNA to be complementary to those particular sequences. For example, one can determine the nucleic acid sequence of a subject's LTR U3, R, or U5 region and then design one or more guide RNAs to be precisely complementary to the patient's sequences, again without selecting a sequence that is a part of the truncated Tat-responsive HIV promoter.

The compositions are also useful for the treatment, for example, as a prophylactic treatment, of a subject at risk for having a retroviral infection, e.g., an HIV infection. These methods can further include the steps of a) identifying a subject at risk for having an HIV infection; b) providing to the subject a composition comprising a nucleic acid encoding a CRISPR-associated nuclease, e.g., Cas9, under control of a truncated Tat-responsive HIV-1 LTR promoter. The sequence may additionally encode for a guide RNA complementary to an HIV target sequence, e.g. an HIV LTR. A subject at risk for having an HIV infection can be, for example, any sexually active individual engaging in unprotected sex, i.e., engaging in sexual activity without the use of a condom; a sexually active individual having another sexually transmitted infection; an intravenous drug user; or an uncircumcised man. A

subject at risk for having an HIV infection can be, for example, an individual whose occupation may bring him or her into contact with HIV-infected populations, e.g., healthcare workers or first responders. A subject at risk for having an HIV infection can be, for example, an inmate in a correctional setting or a sex worker, that is, an individual who uses sexual activity for income employment or nonmonetary items such as food, drugs, or shelter.

The compositions can also be administered to a pregnant or lactating woman having an HIV infection in order to reduce the likelihood of transmission of HIV from the mother to her offspring. A pregnant woman infected with HIV can pass the virus to her offspring transplacentally in utero, at the time of delivery through the birth canal or following delivery, through breast milk. The compositions disclosed herein can be administered to the HIV infected mother either prenatally, perinatally or postnatally during the breast-feeding period, or any combination of prenatal, perinatal, and postnatal administration. Compositions can be administered to the mother along with standard antiretroviral therapies as described below. In some embodiments, the compositions of the invention are also administered to the infant immediately following delivery and, in some embodiments, at intervals thereafter. The infant also can receive standard antiretroviral therapy.

The compositions may be administered to an individual who is not infected with HIV to prevent infection with HIV. The composition may include delivering a therapeutically effective amount of the pharmaceutical composition. The pharmaceutical composition may include a sequence encoding a CRISPR-associated endonuclease and at least the core region of a HIV LTR promoter and a TAR region of the truncated Tat-responsive HIV LTR promoter as described above.

The methods disclosed herein can be applied to a wide range of species, e.g., humans, non-human primates (e.g., monkeys), horses or other livestock, dogs, cats, ferrets or other mammals kept as pets, rats, mice, or other laboratory animals.

The methods of the invention can be expressed in terms of the preparation of a medicament. Accordingly, the invention encompasses the use of the agents and compositions described herein in the preparation of a medicament. The compounds described herein are useful in therapeutic compositions and regimens or for the

manufacture of a medicament for use in treatment of diseases or conditions as described herein.

Any composition described herein can be administered to any part of the host's body for subsequent delivery to a target cell. A composition can be delivered to, without limitation, the brain, the cerebrospinal fluid, joints, nasal mucosa, blood, lungs, intestines, muscle tissues, skin, or the peritoneal cavity of a mammal. In terms of routes of delivery, a composition can be administered by intravenous, intracranial, intraperitoneal, intramuscular, subcutaneous, intramuscular, intrarectal, intravaginal, intrathecal, intratracheal, intradermal, or transdermal injection, by oral or nasal administration, or by gradual perfusion over time. In a further example, an aerosol preparation of a composition can be given to a host by inhalation.

The dosage required will depend on the route of administration, the nature of the formulation, the nature of the patient's illness, the patient's size, weight, surface area, age, and sex, other drugs being administered, and the judgment of the attending clinicians. Wide variations in the needed dosage are to be expected in view of the variety of cellular targets and the differing efficiencies of various routes of administration. Variations in these dosage levels can be adjusted using standard empirical routines for optimization, as is well understood in the art. Administrations can be single or multiple (e.g., 2- or 3-, 4-, 6-, 8-, 10-, 20-, 50-, 100-, 150-, or more fold). Encapsulation of the compounds in a suitable delivery vehicle (e.g., polymeric microparticles or implantable devices) may increase the efficiency of delivery.

The duration of treatment with any composition provided herein can be any length of time from as short as one day to as long as the life span of the host (e.g., many years). For example, a compound can be administered once a week (for, for example, 4 weeks to many months or years); once a month (for, for example, three to twelve months or for many years); or once a year for a period of 5 years, ten years, or longer. It is also noted that the frequency of treatment can be variable. For example, the present compounds can be administered once (or twice, three times, *etc.)* daily, weekly, monthly, or yearly.

An effective amount of any composition provided herein can be administered to an individual in need of treatment. An effective amount can be

determined by assessing a patient's response after administration of a known amount of a particular composition. In addition, the level of toxicity, if any, can be determined by assessing a patient's clinical symptoms before and after administering a known amount of a particular composition. It is noted that the effective amount of a particular composition administered to a patient can be adjusted according to a desired outcome as well as the patient's response and level of toxicity. Significant toxicity can vary for each particular patient and depends on multiple factors including, without limitation, the patient's disease state, age, and tolerance to side effects.

Any method known to those in the art can be used to determine if a particular response is induced. Clinical methods that can assess the degree of a particular disease state can be used to determine if a response is induced. The particular methods used to evaluate a response will depend upon the nature of the patient's disorder, the patient's age, and sex, other drugs being administered, and the judgment of the attending clinician.

The compositions may also be administered with another therapeutic agent, for example, an anti-retroviral agent, used in HAART. Antiretroviral agents may include reverse transcriptase inhibitors (e.g., nucleoside/nucleotide reverse transcriptase inhibitors, zidovudine, emtricitibine, lamivudine and tenoifvir; and non-nucleoside reverse transcriptase inhibitors such as efavarenz, nevirapine, rilpivirine); protease inhibitors, e.g., tipiravir, darunavir, indinavir; entry inhibitors, e.g., maraviroc; fusion inhibitors, e.g., enfuviritide; or integrase inhibitors e.g., raltegrivir, dolutegravir. Antiretroviral agents may also include multi- class combination agents for example, combinations of emtricitabine, efavarenz, and tenofivir; combinations of emtricitabine; rilpivirine, and tenofivir; or combinations of elvitegravir, cobicistat, emtricitabine and tenofivir.

Concurrent administration of two or more therapeutic agents does not require that the agents be administered at the same time or by the same route, as long as there is an overlap in the time period during which the agents are exerting their therapeutic effect. Simultaneous or sequential administration is contemplated, as is administration on different days or weeks. The therapeutic agents may be

administered under a metronomic regimen, e.g., continuous low-doses of a therapeutic agent.

Dosage, toxicity and therapeutic efficacy of such compositions can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD_{50} (the dose lethal to 50% of the population) and the ED_{50} (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD_{50}/ED_{50}.

The data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage for use in humans. The dosage of such compositions lies preferably within a range of circulating concentrations that include the ED_{50} with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any composition used in the method of the invention, the therapeutically effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound which achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography.

As described, a therapeutically effective amount of a composition (i.e., an effective dosage) means an amount sufficient to produce a therapeutically (e.g., clinically) desirable result. The compositions can be administered one from one or more times per day to one or more times per week; including once every other day. The skilled artisan will appreciate that certain factors can influence the dosage and timing required to effectively treat a subject, including but not limited to the severity of the disease or disorder, previous treatments, the general health and/or age of the subject, and other diseases present. Moreover, treatment of a subject with a therapeutically effective amount of the compositions of the invention can include a single treatment or a series of treatments.

The compositions described herein are suitable for use in a variety of drug delivery systems described above. Additionally, in order to enhance the *in vivo* serum half-life of the administered compound, the compositions may be encapsulated, introduced into the lumen of liposomes, prepared as a colloid, or other conventional techniques may be employed which provide an extended serum half-life of the compositions. A variety of methods are available for preparing liposomes, as described in, e.g., Szoka, *et al.*, U.S. Pat. Nos. 4,235,871, 4,501,728 and 4,837,028 each of which is incorporated herein by reference. Furthermore, one may administer the drug in a targeted drug delivery system, for example, in a liposome coated with a tissue specific antibody. The liposomes will be targeted to and taken up selectively by the organ.

Also provided, are methods of inactivating a retrovirus, for example a lentivirus such as a human immunodeficiency virus, a simian immunodeficiency virus, a feline immunodeficiency virus, or a bovine immunodeficiency virus in a mammalian cell. The human immunodeficiency virus can be HIV-1 or HIV-2. The human immunodeficiency virus can be a chromosomally integrated provirus. The mammalian cell can be any cell type infected by HIV, including, but not limited to $CD4^+$ lymphocytes, macrophages, fibroblasts, monocytes, T lymphocytes, B lymphocytes, natural killer cells, dendritic cells such as Langerhans cells and follicular dendritic cells, hematopoietic stem cells, endothelial cells, brain microglial cells, astrocytes and gastrointestinal epithelial cells. Such cell types include those cell types that are typically infected during a primary infection, for example, a $CD4^+$ lymphocyte, a macrophage, a monocyte or a Langerhans cell, as well as those cell types that make up latent HIV reservoirs, i.e., a latently infected cell.

The methods can include exposing and/or contacting the cell to a composition comprising an isolated nucleic acid encoding a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter containing the core region and the TAR region of the HIV LTR promoter. The isolated nucleic acid may further encode one or more guide RNAs wherein the guide RNA is complementary to a target nucleic acid sequence in the retrovirus. The contacting step can take place *in vivo*, that is, the compositions can be administered directly to a subject having HIV infection. The methods are not so limited however, and the contacting step can take

place *ex vivo*. For example, a cell or plurality of cells, or a tissue explant, can be removed from a subject having an HIV infection and placed in culture, and then contacted with a composition comprising a CRISPR-associated endonuclease operably linked to a truncated HIV LTR promoter and optionally a guide RNA wherein the guide RNA is complementary to a nucleic acid sequence in HIV. As described above, a pharmaceutical composition may include a nucleic acid encoding a CRISPR-associated endonuclease operably linked to a truncated Tat-responsive HIV LTR promoter.

The compositions are formulated in such a way as to promote uptake by the mammalian cell. Useful vector systems and formulations are described above. In some embodiments the vector can deliver the compositions to a specific cell type. The invention is not so limited however, and other methods of DNA delivery such as chemical transfection, using, for example calcium phosphate, DEAE dextran, liposomes, lipoplexes, surfactants, and perfluoro chemical liquids are also contemplated, as are physical delivery methods, such as electroporation, micro injection, ballistic particles, and "gene gun" systems.

Standard methods, for example, immunoassays to detect the CRISPR-associated endonuclease, or nucleic acid-based assays such as PCR to detect the guide RNA, can be used to confirm cell has taken up and/or expressed the protein into which it has been introduced. The engineered cells can then be reintroduced into the subject from whom they were derived as described below.

In other embodiments, the compositions comprise a cell which has been transformed or transfected with one or more Cas9/truncated Tat-responsive HIV LTR promoter vectors. In some embodiments, the methods of the invention can be applied *ex vivo*. That is, a subject's cells can be removed from the body and treated with the compositions in culture to excise HIV sequences and the treated cells returned to the subject's body. The cell can be the subject's cells or they can be haplotype matched or a cell line. The cells can be irradiated to prevent replication. In some embodiments, the cells are human leukocyte antigen (HLA)-matched, autologous, cell lines, or combinations thereof. In other embodiments the cells can be a stem cell. For example, an embryonic stem cell or an artificial pluripotent stem cell (induced

pluripotent stem cell (iPS cell)). Embryonic stem cells (ES cells) and artificial pluripotent stem cells (induced pluripotent stem cell, iPS cells) have been established from many animal species, including humans. These types of pluripotent stem cells would be the most useful source of cells for regenerative medicine because these cells are capable of differentiation into almost all of the organs by appropriate induction of their differentiation, with retaining their ability of actively dividing while maintaining their pluripotency. iPS cells, in particular, can be established from self-derived somatic cells, and therefore are not likely to cause ethical and social issues, in comparison with ES cells which are produced by destruction of embryos. Further, iPS cells, which are self-derived cell, make it possible to avoid rejection reactions, which are the biggest obstacle to regenerative medicine or transplantation therapy.

The compositions described herein can be packaged in suitable containers labeled, for example, for use as a therapy to treat a subject having a retroviral infection, for example, an HIV infection or a subject at for contracting a retroviral infection, for example, an HIV infection. The containers can include a composition comprising a nucleic acid sequence encoding a CRISPR- associated endonuclease, for example, a Cas9 endonuclease, and a truncated Tat-responsive HIV LTR promoter as described earlier. The sequence may additionally encode a guide RNA complementary to a target sequence in a HIV, or a vector encoding that nucleic acid, and one or more of a suitable stabilizer, carrier molecule, flavoring, and/or the like, as appropriate for the intended use. Accordingly, packaged products (e.g., sterile containers containing one or more of the compositions described herein and packaged for storage, shipment, or sale at concentrated or ready-to-use concentrations) and kits, including at least one of the disclosed compositions. A product can include a container (e.g., a vial, jar, bottle, bag, or the like) containing one or more compositions of the invention. In addition, an article of manufacture further may include, for example, packaging materials, instructions for use, syringes, delivery devices, buffers or other control reagents for treating or monitoring the condition for which prophylaxis or treatment is required. In some embodiments, the kits can include one or more additional antiretroviral agents, for example, a reverse transcriptase inhibitor, a protease inhibitor or an entry inhibitor. The additional agents can be packaged together in the same container as a nucleic acid sequence encoding a

CRISPR- associated endonuclease, for example, a Cas9 endonuclease, operably linked to a truncated HIV LTR promoter and optionally a guide RNA complementary to a target sequence in a HIV, or a vector encoding that nucleic acid or they can be packaged separately.

The product may also include a legend (e.g., a printed label or insert or other medium describing the product's use (e.g., an audio- or videotape)). The legend can be associated with the container (e.g., affixed to the container) and can describe the manner in which the compositions therein should be administered (e.g., the frequency and route of administration), indications therefor, and other uses. The compositions can be ready for administration (e.g., present in dose-appropriate units), and may include one or more additional pharmaceutically acceptable adjuvants, carriers or other diluents and/or an additional therapeutic agent. Alternatively, the compositions can be provided in a concentrated form with a diluent and instructions for dilution.

The practice of the invention is illustrated by the following non-limiting examples.

EXAMPLES

Example 1: Cloning of LTR-Cas9 variants

Full length and various truncated LTR promoter sequences were obtained by PCR using pNL4-3 HIV vector (NIH AIDS Reagent Program #114) as a template and the following primers (restriction sites noted in boldface):

Kpn1-LTR(-454)-S 5'-**GGTACC**TGGAAGGGCTAATTTGG-3' (SEQ ID NO: 1)

Kpn1-LTR(-120)-S 5'-**GGTACC**TCGAGCTTTCTACAAGG-3' (SEQ ID NO: 2)

Xba1-LTR(-80)-S 5'-**TCTAGA**GGAGGTGTGGCCTGGGC-3' (SEQ ID NO: 3)

Kpn1-LTR(-38)-S 5'-**GGTACC**AGATGCTACATATAAGC-3' (SEQ ID NO: 4), or

LTR(+66)-**Nco1**-AS 5'-**CCATGG**TAAGCAGTGGGTTCC-3' (SEQ ID NO: 5).

The derivation of the truncated HIV-1 LTR promoter variants is shown in diagrammatically in Fig 1A with reference to the U3, R and U5 regions of the LTR, and the LTR enhancer, core, TAR (trans-activation-responsive) and TATA box elements. Fig. 1B shows an agarose gel electrophoresis image of the PCR-amplified LTR truncation variants.

PCR products were gel purified and directly subcloned in TA vector (Invitrogen), then excised with Kpn1 or Xba1 and Nco1 restriction enzymes and ligated into Kpn1-Nco1 or Xba1-Nco1 digested pX260-U6-DR-BB-DR-Cbh-NLS-hSpCas9-NLS-H1-shorttracr-PGK-puro plasmid (Addgene #42229) (hereinafter "pX260 plasmid") as a Cas9 gene source/template. The pX260 plasmid contains a Cbh promoter (Xba1-Kpn1-Cbh-Nco1) As a result of the manipulation, the original Cbh promoter in the pX260 plasmid was removed and replaced with one of the LTR promoters (Xba1- or Kpn1-LTR-Nco1). A blueprint of the original pX260 plasmid structure in shown in Fig. 2, identified as "**Cbh-Cas9**" (from www.Addagene.org and Cong *et al.*, *Science* (2013) 339(6121):819-23). A blueprint of a modified plasmid is shown in Fig. 2 as "**LTR-Cas9**".

Example 2: Optimization of LTR/Tat Ratio for Inducing Cas9 Expression

To find an optimal ratio between Tat and LTR promoter for the best transactivation effect, cells of the human primary glioblastoma cell line U87 MG were co-transfected using Lipofectamine 2000 reagent (Invitrogen) with different amounts of plasmid expressing FLAG-labeled Cas9 under control of full length HIV-1 LTR (pLTR(-454/+66)-FLAG-Cas9) plasmid (10, 50 and 250ng), with or without Tat expressing plasmid (pCMV-Tat86, 250ng). U87 MG is an HIV-1 latency reporter cell line. The total amount of DNA was equilibrated with empty pCMV plasmid (pcDNA3.1). Forty-eight hours later, cells were lysed in TNN buffer (50mM Tris pH 7.4, 100mM NaCl, 5mM EDTA, 1% NP 40). Cas9, Tat and α-tubulin expression were then examined by Western blot. The results are shown in Fig. 3A (U87 MG WCE 50μg/well). Lane 1: pLTR(-454/+66)-Cas9 250ng, pCMV 1000ng. Lane 2: pLTR(-454/+66)-Cas9 50ng, pCMV 1200ng. Lane 3: pLTR(-454/+66)-Cas9 10ng,

pCMV 1240ng. Lane 4: pLTR(-454/+66)-Cas9 250ng, pCMV 750ng, pCMV-Tat86 250ng. Lane 5: pLTR(-454/+66)-Cas9 50ng, pCMV 950ng, pCMV-Tat86 250ng. Lane 6: pLTR(-454/+66)-Cas9 10ng, pCMV 990ng, pCMV-Tat86 250ng.

The intensity of bands corresponding to Cas9 and α-tubulin (used as a loading control) were analyzed and compared using ImageJ software. The results are shown in Fig. 3B. The top panel shows the Western blot image quantification of the Cas9 levels normalized to the levels of α-tubulin, with or without Tat. The bottom panel show the Western blot image quantification of the +Tat/no Tat ratio. The results indicate that maximal (5.3x) induction of Cas9 expression was obtained at a 1:5 ratio of pLTR-Cas9:pCMVTat86 (50ng:250ng).

Example 3: Comparison of Truncated LTR Promoters in Inducing Cas9 Expression

To test and compare truncated LTR promoters, U87 MG cells were transfected with different amounts of plasmids (5ng or 50ng) expressing FLAG-labeled Cas9 under control of the HIV-1 truncated LTR variant pLTR(-120/+66)-FLAG-Cas9 or the HIV-1 truncated LTR variant pLTR(-80/+66)-FLAG-Cas9, with or without Tat expressing plasmid (pCMV-Tat86, 250ng). Forty-eight hours later, whole cell lysates where prepared and resolved by Western blot. Intensity of bands corresponding to Cas9 and α-tubulin (used as a loading control) were analyzed and compared using ImageJ software. The results are shown in Fig. 4A. Lane 1: pLTR(-120/+66)-Cas9 5ng, pCMV 1245ng. Lane 2: pLTR(-120/+66)-Cas9 5ng, pCMV 1245ng, +rTat protein 2.5 μg/ml. Lane 3: pLTR(-120/+66)-Cas9 5ng, pCMV 995 ng, pCMV-Tat86 250ng. Lane 4: pLTR(-120/+66)-Cas9 50ng, pCMV 1200ng. Lane 5: pLTR(-120/+66)-Cas9 50ng, pCMV 1200ng, +rTat protein 2.5 μg/ml. Lane 6: pLTR(-120/+66)-Cas9 50ng, pCMV 950 ng, pCMV-Tat86 250ng. Lane 7: pLTR(-80/+66)-Cas9 5ng, pCMV 1245ng. Lane 8: pLTR(-80/+66)-Cas9 5ng, pCMV 1245ng, +rTat protein 2.5 μg/ml. Lane 9: pLTR(-80/+66)-Cas9 5ng, pCMV 995 ng, pCMV-Tat86 250ng. Lane 10: pLTR(-80/+66)-Cas9 50ng, pCMV 1200ng. Lane 11: pLTR(-80/+66)-Cas9 50ng, pCMV 1200ng, +rTat protein 2.5 μg/ml. Lane 12: pLTR(-80/+66)-Cas9 50ng, pCMV 950 ng, pCMV-Tat86 250ng.

The intensity of bands corresponding to Cas9 and α-tubulin (used as a loading control) were analyzed and compared using ImageJ software. The results are

shown in Fig. 4B. The top panel show the Western blot image quantification of the Cas9 levels normalized to the levels of α-tubulin, with no Tat, with rTAT or with transfected Tat. The bottom panel show the Western blot image quantification of the +Tat(transfected)/no Tat ratio. The results demonstrate that removing modulatory and/or enhancer regions of the LTR (those regions being schematically represented in Fig. 1A) did not significantly affect Tat-mediated transactivation of Cas9 expression. Tat-mediated expression was apparent from the pLTR(-80/+66)-FLAG-Cas9 plasmid, containing the core and TAR LTR promoters elements, but not the enhancer and modulatory regions.

Example 4: Negative Feedback Regulation of HIV-1 by Gene Editing Strategy

In the studies presented here, the gene editing composition allows conditional activation of the CRISPR/Cas9 at the early stage of viral reactivation by the HIV-1 transcriptional activator, Tat. This new strategy completely and permanently ablates virus replication prior to productive viral replication by removing a segment of the viral gene spanning the viral promoter and/or the viral coding sequence. Further, this strategy alleviates any concerns due to unforeseen complications that may arise by unnecessary and persistent expression of Cas9 at high levels in cells.

Results

The coding DNA sequence corresponding to the Cas9 gene was placed in a pX26 expression vector plasmid containing three different segments of the HIV-1 promoter spanning the U3 and R regions of the 5'-LTR to identify the minimal DNA elements of the viral promoter that remain responsive to Tat, yet lacks the sequences corresponding to gRNAs A and B that are initially used for editing HIV-1 DNA (Fig. 5A). After verification of this cloning strategy by DNA sequencing of each construct, expression of Cas9 by each vector and the level of responsiveness to Tat was examined in TZMb1 cells co-transfected with pX26 or pX26-LTR-Cas9 and CMV-Tat. Results from Western blot revealed activation of expression of Cas9 by all three constructs including the plasmid encompassing the minimal DNA promoter sequence positioned between -80 to +66 (Fig. 5B). This was particularly important for these studies as the promoter sequence resides outside of the DNA sequences corresponding

to gRNAs A and B (Fig. 5B). Next, a DNA fragment corresponding to $LTR_{(-80/+66)}$-Cas9 was cloned into a lentiviral vector (LV) and used to transduce TZMb1 cells to assess the effect of Tat protein on the editing of integrated copies of HIV-1 DNA expressing the luciferase reporter gene. Results from PCR amplification of the LTR revealed the detection of 205 bp DNA fragment in cells expressing gRNAs A and B and Tat protein (Fig. 5C, compare lanes 1-5 to lanes 6-8). The position of the primers used for PCR amplification and the expected amplicons are illustrated in Fig. 5A (also see Fig. 10). Results from sequencing verified excision of the 190 bp DNA fragment upon expression of Tat in cells transduced by LV-$LTR_{(-80/+66)}$-Cas9 plus LV-gRNAs A/B. Expression of Cas9, Tat and α-tubulin (control for equal loading) are shown in Fig. 5D.

Next, the impact of the viral DNA excision on viral promoter activity was examined by luciferase assay. Results show a gradual decrease in luciferase activity upon activation of Cas9 by Tat, corroborating the results from DNA assay, indicating that the cleavage of DNA causes inhibition of viral promoter activity in these cells (Fig. 5E). In follow-up studies, the activation of Cas9 was investigated upon infection of TZMb1 cells by HIV-1. To this end, cells were transduced by LV-$LTR_{(-80/+66)}$-Cas9 and LV-gRNAs A/B for 24 hours, after which cells were infected with HIV-1_{JRFL} or HIV-1_{SF162} at three different MOIs. After 48 hours, cells were harvested for evaluating DNA excision by PCR, expression of the integrated promoter sequence by luciferase assay, and expression of Cas9 by Western blot. Results from these experiments show the detection of a post-cleavage 205 bp DNA fragment in cells infected with HIV-1_{JRFL} and HIV-1_{SF162}, indicating that production of Tat by HIV-1_{JRFL} and HIV-1_{SF162} transactivated the $LTR_{(-80/+66)}$ promoter and production of Cas9 in these cells (Fig. 6A). Further, results from the luciferase assay revealed significant reduction of luciferase activity in the cells, again verifying the effectiveness of Cas9 activation by Tat, which is produced upon infection by HIV-1_{JRFL} or HIV-1_{SF162} in shutting down the integrated HIV-1 luciferase gene. Induction of Cas9 in the infected cells is shown in Fig. 6B. Results from Western blot showed activation of the truncated LTR promoter, $LTR_{(-80/+66)}$, upon infection of cells with HIV-1_{JRFL} and HIV-1_{SF162}, resulting in the production of Cas9 protein in the cells (Fig. 6C).

In the follow-up, the ability of Tat-mediated activation of the LTR-Cas9 was tested along with gRNAs A/B in eliminating the HIV-1 genome in the human T-lymphocytic cells line, 2D10. These cells harbor integrated copies of a single round HIV-1PNLA4-3 in a latent state, whose genome lacks a portion of the Gag and Pol genes and the Nef gene is replaced by a gene encoding the reporter green fluorescent protein (GFP). The enhanced level of Tat protein in these cells and the activation of Cas9 (shown in Fig. 7A) caused editing of the viral LTR upon activation of Cas9 in the cells transduced by LV-gRNAs A/B (Fig. 7B, also see Fig. 11, lanes 1-8). Accordingly, a significant decrease in the number of GFP positive cells was detected in the presence of Tat, indicating that activation of Tat eliminates the capacity of the cleaved promoter in expressing viral DNA, which in turn, causes suppression of GFP in these cells. The DNA sequence corresponding to the position of the gRNAs, excision of the DNA fragment and PCR primers are shown in Fig. 12.

In light of earlier observations indicating to the ability of PMA and/or TSA in stimulating integrated copies of proviral DNA in 2D10 cells, the impact of PMA and TSA on the activation of Cas9 in a latently infected T-cell model was assessed. As seen in Fig. 8A, treatment of 2D10 cells with PMA and TSA, singly or in combination, increased the level of Cas9 expression. In a parallel experiment, PCR analysis was performed for the detection of LTR DNA and showed a clear increase in the level of viral DNA excision (Fig. 8B), as evidenced by the appearance of the 205 bp DNA fragment (see Fig. 11, lanes 9-14). Examination of viral activation by measuring the level of GFP in the cells using Western blot or the quantification of green fluorescent cells, indicative of viral activation, by fluorescent microscopy (Fig. 8C) showed a drastic decrease in the level of viral gene expression. Thus, it is likely that activation of the minimal viral promoter (-88/+60) either by Tat produced upon reactivation of the silent provirus or directly by PMA and TSA that produced Cas9, have a negative impact on the expression of the latently integrated viral genome in cells containing gRNAs A and B.

Discussion

Since its discovery in 1985, the Tat protein of HIV-1 has captured significant attention due to its critical role in expression of the viral genome at the

transcriptional level and its pathogenic impact on uninfected cells. Mechanistically, Tat associates with the RNA sequence located downstream of the initiation site from transcription (nucleotides +1 to +59), so-called transcription responsive region or TAR. The association of Tat with TAR triggers a series of molecular and biochemical events leading to the formation of pre-initiation and initiation complexes of transcription in proximity to the transcription start site (nucleotide +1). This complex includes a series of cellular proteins that have the ability to phosphorylate or acetylate components of the complexes including pTEF and RNA polymerase II, thus facilitating transcriptional elongation of RNA. In addition, the interaction of Tat with various transcriptional factors including NF-κB, p300/CBP and GCN5 can affect transcription of other viral and cellular genes; all of which contribute to the disease spectrum seen in HIV-1 positive AIDS patients. Tat also plays a major role in the productive replication of latent virus in reservoirs once transcription from the reactivated viral promoter leads to an initial round of viral transcription and Tat production. The unique importance of Tat in HIV-1 replication and the pathogenesis of AIDS, provided a strong rationale for serving as a potential target for drug discovery as well as vaccine development. In fact, several potent inhibitors, some with the ability to interfere with Tat-TAR interaction and others with the capacity to prevent Tat communication with its cellular partners, have shown various degrees of efficacy in affecting HIV-1 replication.

The strategy that was utilized in this study was to recruit Tat to excise a segment of the viral genome and permanently ablate HIV-1 gene transcription and replication in cells with productive or latent HIV-1. Here a suicide path was designed for HIV-1 that is triggered by Tat and includes editing of the viral genome using CRISPR/Cas9 technology (illustrated in Fig. 9). According to this pathway, production of Tat in the cells, in addition to stimulating its own promoter with the full-length 5'-LTR sequence, potentiates expression of Cas9 through the same mechanism by a truncated minimal promoter sequence spanning the GC-rich, TATA box, and TAR (-80 to +66) regions. Production of Cas9 and its association with gRNAs designed to target the LTR DNA sequence outside of the (-80 to +66) induced InDel mutations within the full-length viral promoter and by excising a segment of the gene, can permanently eradicate HIV-1 in the cells. In addition to the expected 417 bp

DNA fragment representing the full-length LTR sequence, results from short-range amplification of LTR DNA showed a second DNA fragment of 227 bp in size only in cells expressing Tat. The 227 bp DNA fragment was generated by joining the residual 5'-LTR to the remaining 3'-LTR after cleavage by Cas9/gRNA A at either the 5'-LTR or the 3'-LTR. It is also likely that ligation of the remaining DNA fragment from the 5'-LTR with those from the 3'-LTR after cleavage by Cas9/gRNA created a new template for gene amplification and the appearance of a similar size (227 bp) amplicons. A multiplex of gRNAs were utilized that target the LTR (gRNA A) plus a region within the Gag gene with the expectation of the removal of DNA fragment between gRNA A and gRNA Gag.

The CRISPR/Cas9 gene editing strategy has received attention in biomedical research in recent years due to its extraordinary ability to edit the genome with precision and high efficiency and its simplicity and flexibility of implementation. However, there are several areas that need close attention. For example, it is important to design the most specific and effective gRNAs to avoid off-target effects. The strategy that was employed here for maximizing specificity and avoiding off-target editing was verified by ultra deep sequencing of the whole genome and various other tests. The second issue relates to the controlled expression of Cas9 to avoid unnecessary presence of the protein that may non-specifically cause injury to the host genome in the long term and/or induce an immune response. The strategy here was developed for conditional expression of Cas, only in the presence of HIV-1 Tat, which provides a novel approach for activating and silencing gene editing for eradicating HIV-1 when the virus is on the rise.

The disclosures of each and every patent, patent application, and publication cited herein are hereby incorporated herein by reference in their entirety. One skilled in the art will readily appreciate that the present invention is well adapted to carry out the objects and obtain the ends and advantages mentioned, as well as those inherent therein. While the invention has been disclosed with reference to specific embodiments, it is apparent that other embodiments and variations of this invention may be devised by others skilled in the art without departing from the true spirit and scope used in the practice of the invention. The appended claims are intended to be construed to include all such embodiments and equivalent variations

CLAIMS

What is claimed:

1. A method of inactivating a human immunodeficiency virus (HIV) *in vivo* or *in vitro*, the method comprising administering to a subject or contacting a mammalian cell with a composition comprising an isolated nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated HIV long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) of a HIV LTR promoter.

2. The method of claim 1, wherein the isolated nucleic acid sequence further encodes at least one guide RNA that is complementary to a target nucleic acid sequence in HIV.

3. The method of claim 2, wherein the target nucleic acid sequence in HIV comprises a sequence within a coding region or non-coding region of HIV.

4. The method of claim 3, wherein the non-coding region comprises a long terminal repeat sequence of HIV.

5. The method of claim 3, wherein the target nucleic acid sequence comprises a sequence within the long terminal repeat of HIV.

6. The method of claim 5, wherein the sequence within the long terminal repeat of HIV comprises a sequence within the U3, R, or U5 regions that excludes any sequence of the truncated HIV LTR promoter.

7. The method of claim 1, wherein the mammalian cell is a latently infected cell.

8. The method of claim 7, wherein the latently infected cell is selected from the group consisting of: a $CD4^+$ T cell, a macrophage, a monocyte, a gut-associated lymphoid cell, a microglial cell, and an astrocyte.

9. The method of claim 1, wherein the inactivating is *in vivo* or *ex vivo*.

10. The method of claim 9, wherein the mammalian cell comprises a cultured cell from a subject having a HIV infection, a tissue explant, or a cell line.

11. The method of claim 1, wherein the CRISPR-associated endonuclease is Cas9.

12. The method of claim 1, wherein the CRISPR-associated endonuclease is optimized for expression in a human cell.

13. The method of claim 1, wherein the composition further comprises a sequence encoding a transactivating small RNA (tracrRNA).

14. The method of claim 13, wherein the tracrRNA is fused to a sequence encoding a guide RNA.

15. The method of claim 1, wherein the composition further comprises a sequence encoding a nuclear localization signal.

16. The method of claim 1, wherein the composition is operably linked to an expression vector.

17. The method of claim 16, wherein the expression vector is selected from the group consisting of: a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

18. The method of claim 1, wherein the composition further comprises an enhancer region of the HIV-1 LTR promoter.

19. An isolated nucleic acid sequence comprising a sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated a human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) of a HIV LTR promoter.

20. The isolated nucleic acid sequence of claim 19, wherein the sequence further encodes at least one guide RNA that is complementary to a target nucleic acid sequence in HIV.

21. The isolated nucleic acid sequence of claim 20, wherein the target nucleic acid sequence in HIV comprises a sequence within a coding region or a non-coding region of HIV.

22. The isolated nucleic acid sequence of claim 21, wherein the non-coding region comprises a long terminal repeat of HIV or a sequence within the long terminal repeat of HIV.

23. The isolated nucleic acid sequence of claim 22, wherein the sequence within the long terminal repeat of HIV comprises a sequence within the U3, R, or U5 regions that excludes any sequence of the truncated HIV LTR promoter.

24. The isolated nucleic acid sequence of claim 19, wherein the CRISPR-associated endonuclease is Cas9.

25. The isolated nucleic acid sequence of claim 19, wherein the CRISPR-associated endonuclease is optimized for expression in a human cell.

26. The isolated nucleic acid sequence of claim 19, further comprising a sequence encoding a transactivating small RNA (tracrRNA).

27. The isolated nucleic acid sequence of claim 26, wherein the tracrRNA is fused to a sequence encoding a guide RNA.

28. The isolated nucleic acid sequence of claim 19, further comprising a sequence encoding a nuclear localization signal.

29. The isolated nucleic acid sequence of claim 19, wherein the isolated nucleic acid sequence is operably linked to an expression vector.

30. The isolated nucleic acid sequence of claim 29, wherein the expression vector is selected from the group consisting of: a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

31. The isolated nucleic acid sequence of claim 19, further comprising an enhancer region of the HIV LTR promoter.

32. A pharmaceutical composition comprising a nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) of a HIV LTR promoter.

33. The pharmaceutical composition of claim 32, further comprising a pharmaceutically acceptable carrier.

34. The pharmaceutical composition of claim 33, wherein the pharmaceutically acceptable carrier comprises a lipid-based or polymer-based colloid.

35. The pharmaceutical composition of claim 34, wherein the colloid is a liposome, a hydrogel, a microparticle, a nanoparticle, or a block copolymer micelle.

36. The pharmaceutical composition of claim 32, wherein the pharmaceutical composition is formulated for topical application.

37. The pharmaceutical composition of claim 32, wherein the pharmaceutical composition is contained within a condom.

38. The pharmaceutical composition of claim 32, wherein the sequence further encodes at least one guide RNA that is complementary to a target nucleic acid sequence in HIV.

39. The pharmaceutical composition of claim 38, wherein the target nucleic acid sequence in HIV comprises a sequence within a coding region or a non-coding region of HIV.

40. The pharmaceutical composition of claim 39, wherein the non-coding region comprises a long terminal repeat of HIV.

41. The pharmaceutical composition of claim 40, wherein the target nucleic acid sequence comprises a sequence within the long terminal repeat of HIV.

42. The pharmaceutical composition of claim 41, wherein the sequence within the long terminal repeat of HIV comprises a sequence within the U3, R, or U5 regions that excludes any sequence of the truncated HIV-1 LTR promoter.

43. The pharmaceutical composition of claim 32, wherein the CRISPR-associated endonuclease is Cas9.

44. The pharmaceutical composition of claim 32, wherein the CRISPR-associated endonuclease is optimized for expression in a human cell.

45. The pharmaceutical composition of claim 32, further comprising a nucleic acid sequence encoding a transactivating small RNA (tracrRNA).

46. The pharmaceutical composition of claim 45, wherein the tracrRNA is fused to a sequence encoding a guide RNA.

47. The pharmaceutical composition of claim 32, further comprising a sequence encoding a nuclear localization signal.

48. The pharmaceutical composition of claim 32, wherein the pharmaceutical composition is operably linked to an expression vector.

49. The pharmaceutical composition of claim 48, wherein the expression vector is selected from the group consisting of: a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

50. The pharmaceutical composition of claim 32, wherein the sequence further encodes an enhancer region of the HIV-1 LTR promoter.

51. A method of treating a subject having a human immunodeficiency virus (HIV) infection, the method comprising administering to the subject a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated HIV long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) of a HIV LTR promoter.

52. The method of treating the subject having the HIV infection of claim 51, wherein the sequence further encodes at least one guide RNA that is complementary to a target nucleic acid sequence in HIV.

53. The method of treating the subject having the HIV infection of claim 51, wherein HIV infection is a latent infection.

54. The method of treating the subject having the HIV infection of claim 51, wherein the pharmaceutical composition is administered topically or parenterally.

55. The method of treating the subject having the HIV infection of claim 51, further comprising administering an anti-retroviral agent.

56. The method of treating the subject having the HIV infection of claim 55, wherein the anti-retroviral agent is selected from the group consisting of non-nucleoside reverse transcriptase inhibitors, protease inhibitors, and entry inhibitors.

57. The method of treating the subject having the HIV infection of claim 56, wherein the anti-retroviral agent comprises highly active antiretroviral therapy.

58. The method of treating the subject having the HIV infection of claim 51, wherein the CRISPR-associated endonuclease is Cas9.

59. The method of treating the subject having the HIV infection of claim 57, wherein the pharmaceutical composition is operably linked to an expression vector.

60. The method of treating the subject having the HIV infection of claim 59, wherein the expression vector is selected from the group consisting of: a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

61. The method of treating the subject having the HIV infection of claim 57, wherein the sequence further encodes an enhancer region of the HIV LTR promoter.

62. A method of reducing risk of a human immunodeficiency virus (HIV) infection in a subject at risk for the HIV infection, the method comprising administering to the subject a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated HIV long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) of a HIV LTR promoter.

63. The method of reducing risk of the HIV infection in the subject at risk for the HIV infection of claim 69, wherein the subject is sexually active, a health care worker or first-responder.

64. The method of reducing risk of the HIV infection in the subject at risk for the HIV infection of claim 62, wherein the CRISPR-associated endonuclease is Cas9.

65. The method of reducing risk of the HIV infection in the subject at risk for the HIV infection of claim 62, wherein the CRISPR-associated endonuclease is optimized for expression in a human cell.

66. The method of reducing risk of the HIV infection in the subject at risk for the HIV infection of claim 62, wherein the pharmaceutical composition is operably linked to an expression vector.

67. The method of reducing risk of the HIV infection in the subject at risk for the HIV infection of claim 67, wherein the expression vector is selected from the group consisting of: a lentiviral vector, an adenoviral vector, and an adeno-associated virus vector.

68. The method of reducing risk of the HIV infection in the subject at risk for the HIV infection of claim 67, wherein the sequence further encodes an enhancer region of the HIV LTR promoter.

69. A method of reducing risk of transmission of a human immunodeficiency virus (HIV) infection from a HIV-infected gestating or lactating mother to her offspring, the method comprising administering to the subject a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated HIV long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) region of a HIV LTR promoter.

70. The method of reducing risk of transmission of the HIV infection from the HIV infected gestating or lactating mother to her offspring of claim 69, wherein the pharmaceutical composition is administered during at least one of: prenatally, perinatally, and postnatally.

71. The method of reducing risk of transmission of the HIV infection from the HIV infected gestating or lactating mother to her offspring of claim 69, further comprising administering an anti-retroviral agent.

72. The method of reducing risk of transmission of the HIV infection from the HIV infected gestating or lactating mother to her offspring of claim 71, wherein the anti-retroviral agent is selected from the group consisting of non-nucleoside reverse transcriptase inhibitors, protease inhibitors, and entry inhibitors.

73. The method of reducing risk of transmission of the HIV infection from the HIV infected gestating or lactating mother to her offspring of claim 72, wherein the anti-retroviral agent comprises highly active antiretroviral therapy.

74. The method of reducing risk of transmission of the HIV infection from the HIV infected gestating or lactating mother to her offspring of claim 69, further comprising administering a therapeutically effective amount of the composition to the offspring.

75. The method of reducing risk of transmission of the HIV infection from the HIV infected gestating or lactating mother to her offspring of claim 69, wherein the sequence further encodes an enhancer region of the HIV LTR promoter.

76. A method of administering a pharmaceutical composition to prevent infection by a human immunodeficiency virus (HIV) in an uninfected subject, the method comprising administering to the uninfected subject a therapeutically effective amount of the pharmaceutical composition comprising a nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated HIV long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) region of a HIV LTR promoter.

77. A composition comprising a nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated HIV long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) region of a HIV LTR promoter.

78. The composition of claim 77, wherein the CRISPR-associated endonuclease is optimized for expression in a human cell.

79. The composition of claim 77, wherein the CRISPR-associated endonuclease is Cas.

80. The composition of claim 79, wherein the CRISPR-associated endonuclease is Cas9 (CRISPR/Cas9).

81. The composition of claim 79, wherein the CRISPR/Cas is under control of a minimal HIV LTR promoter.

82. The composition of claim 81, wherein the minimal truncated HIV LTR promoter is activated by an immediate early transcription activator (Tat) thereby conditionally activating CRISPR/Cas at an early stage of HIV replication.

83. The composition of claim 77, further comprising at least one guide RNA that is complementary to a target nucleic acid sequence in HIV.

84. The composition of claim 77, wherein the composition further comprises a sequence encoding a transactivating small RNA (tracrRNA).

85. The composition of claim 84, wherein the tracrRNA is fused to a sequence encoding a guide RNA.

86. The composition of claim 79, wherein the CRISPR/Cas excises a segment of a viral genome, the segment spanning a viral promoter and/or viral coding sequence.

87. A composition comprising a nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease (CRISPR/Cas) operably linked to a truncated viral promoter whereby the truncated viral promoter is under control of an immediate early transcriptional activator(s), thereby conditionally activating CRISPR/Cas at an early stage of viral replication.

88. The composition of claim 87, further comprising at least one guide RNA that is complementary to a target nucleic acid sequence in the virus.

89. The composition of claim 87, wherein the CRISPR/Cas excises a segment of a viral genome, the segment spanning a viral promoter and/or viral coding sequence.

90. An isolated nucleic acid sequence having at least a 75% sequence similarity to any one of SEQ ID NOS: 1 to 21.

91. An isolated nucleic acid sequence comprising any one of SEQ ID NOS: 1 to 21 or combinations thereof.

92. A kit comprising a measured amount of a composition comprising an isolated nucleic acid sequence comprising a sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably

linked to a truncated HIV long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) region of a HIV LTR promoter, or a vector encoding the isolated nucleic acid, and at least one item selected from the group consisting of packaging material, a package insert comprising instructions for use, a sterile fluid, a syringe, and a sterile container.

Fig. 1A

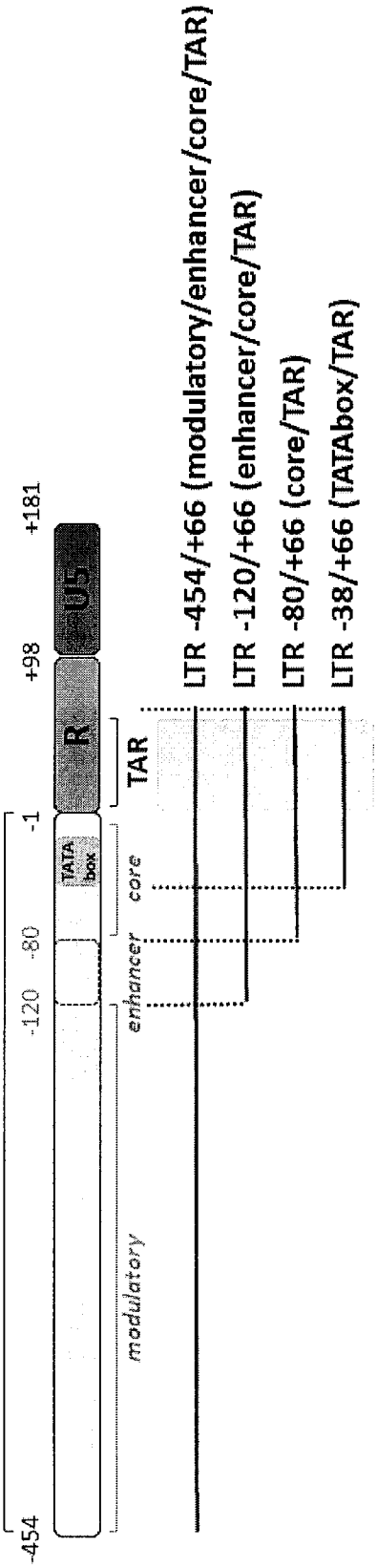
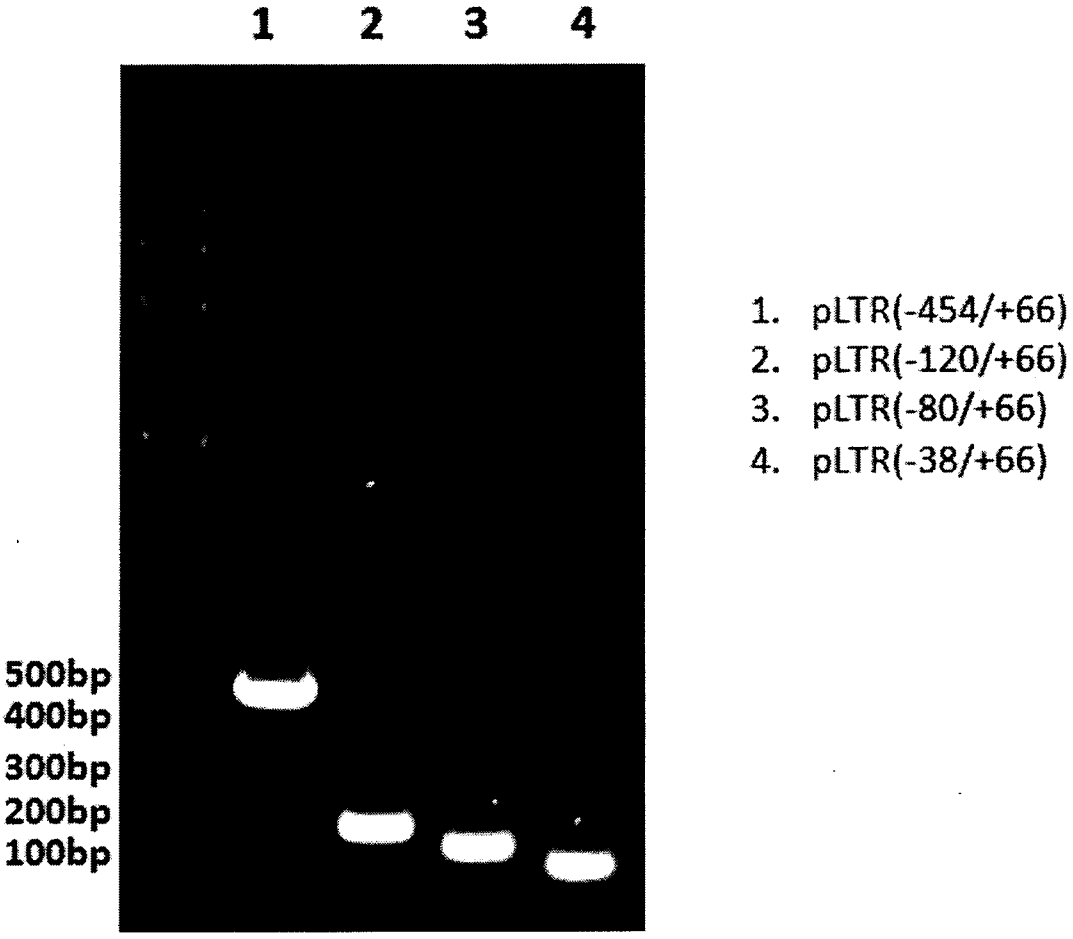

U3
R
U5
TAR
-454
-120
-80
-1
+98
+181
TATA box
modulatory
enhancer
core
LTR -454/+66 (modulatory/enhancer/core/TAR)
LTR -120/+66 (enhancer/core/TAR)
LTR -80/+66 (core/TAR)
LTR -38/+66 (TATAbox/TAR)

Fig. 1B


1
2
3
4
500bp
400bp
300bp
200bp
100bp
1. pLTR(-454/+66)
2. pLTR(-120/+66)
3. pLTR(-80/+66)
4. pLTR(-38/+66)

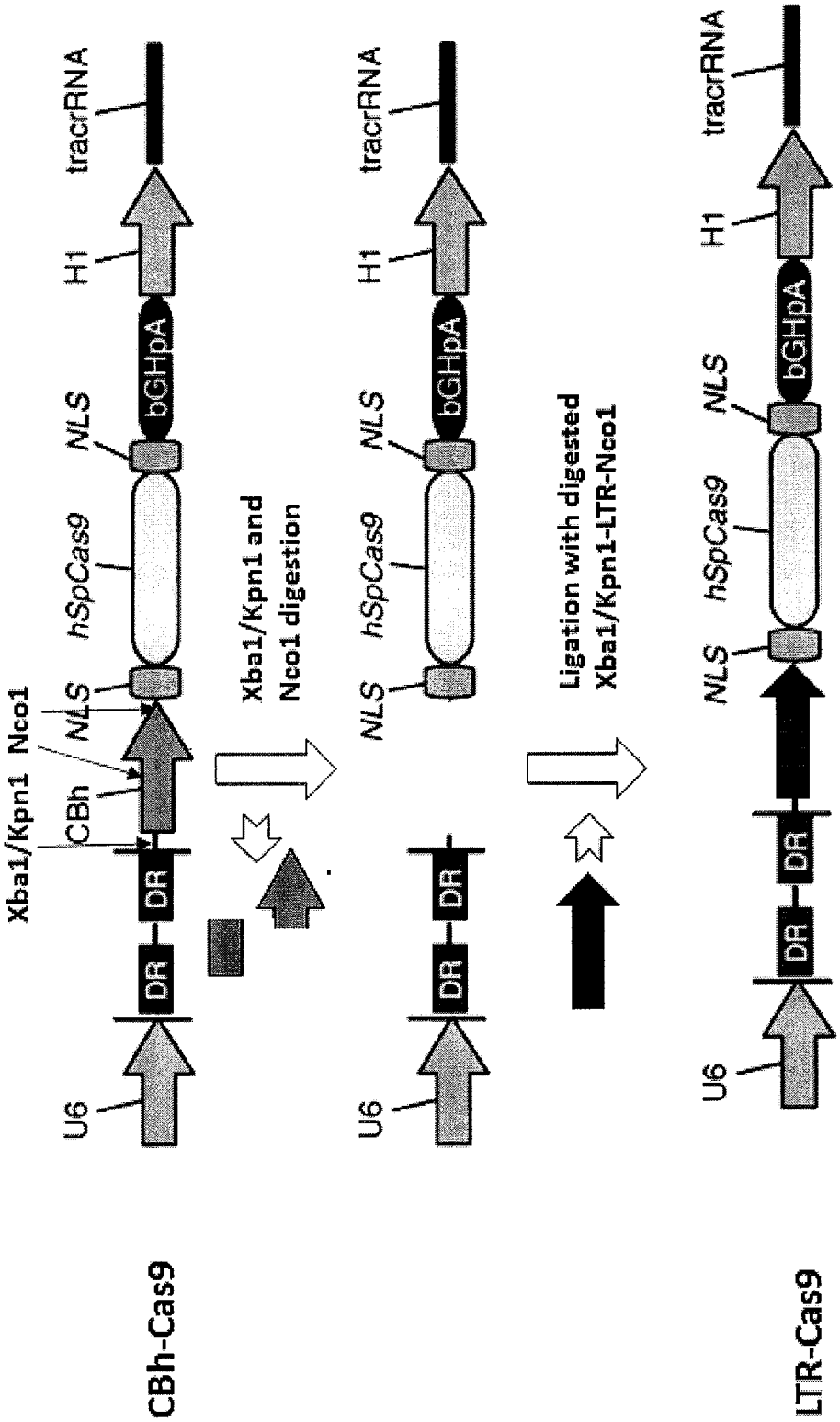
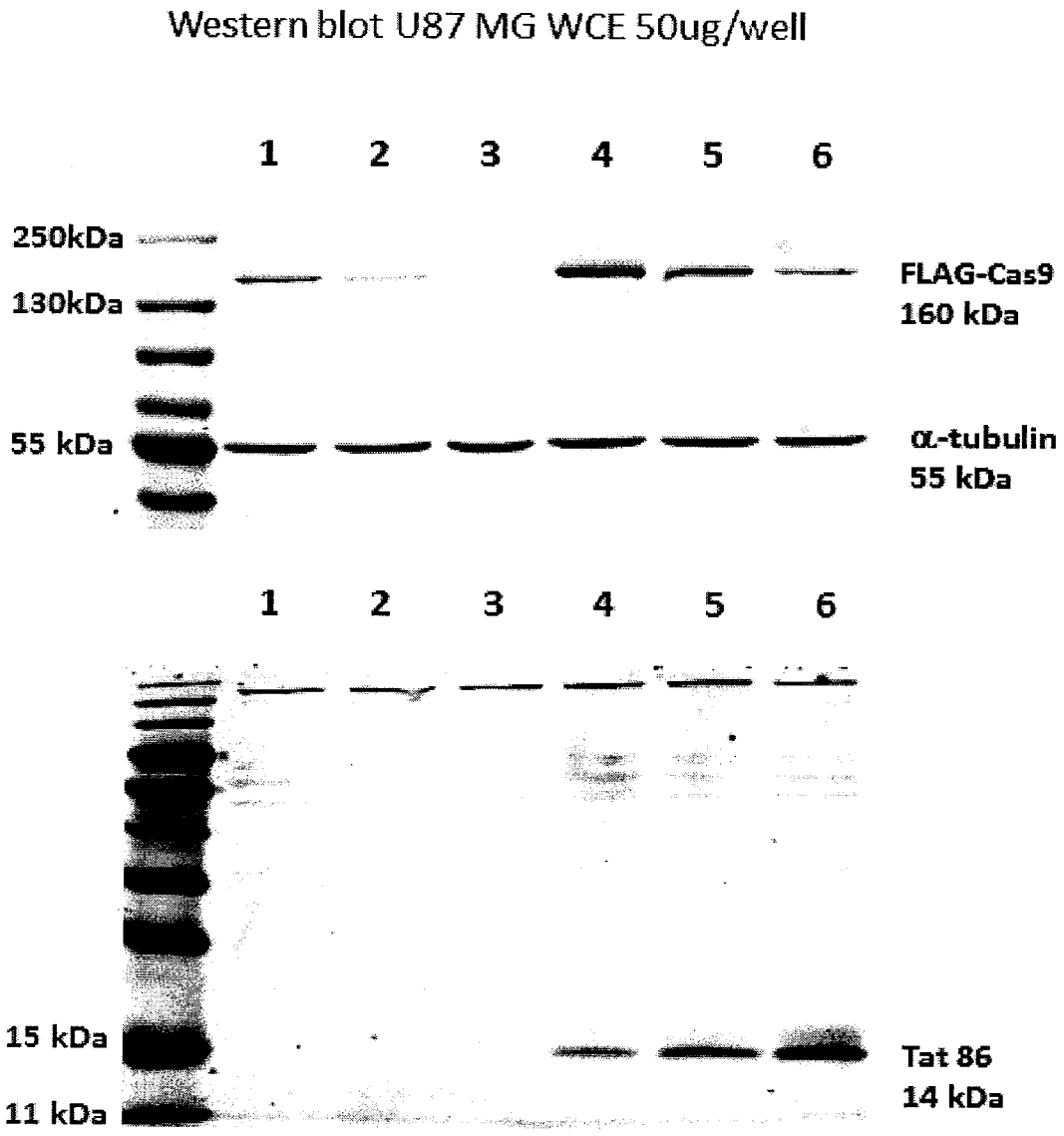

Xba1/Kpn1
Nco1
CBh
NLS
hSpCas9
NLS
bGHpA
H1
tracrRNA
U6
DR
DR
CBh-Cas9
Xba1/Kpn1 and
Nco1 digestion
Ligation with digested
Xba1/Kpn1-LTR-Nco1
LTR-Cas9

Fig. 2

Fig. 3A



1. pLTR(-454/+66)-Cas9 250ng, pCMV 1000ng
2. pLTR(-454/+66)-Cas9 50ng, pCMV 1200ng
3. pLTR(-454/+66)-Cas9 10ng, pCMV 1240ng
4. pLTR(-454/+66)-Cas9 250ng, pCMV 750ng, pCMV-Tat86 250ng
5. pLTR(-454/+66)-Cas9 50ng, pCMV 950 ng, pCMV-Tat86 250ng
6. pLTR(-454/+66)-Cas9 10ng, pCMV 990ng, pCMV-Tat86 250ng

Fig. 3B

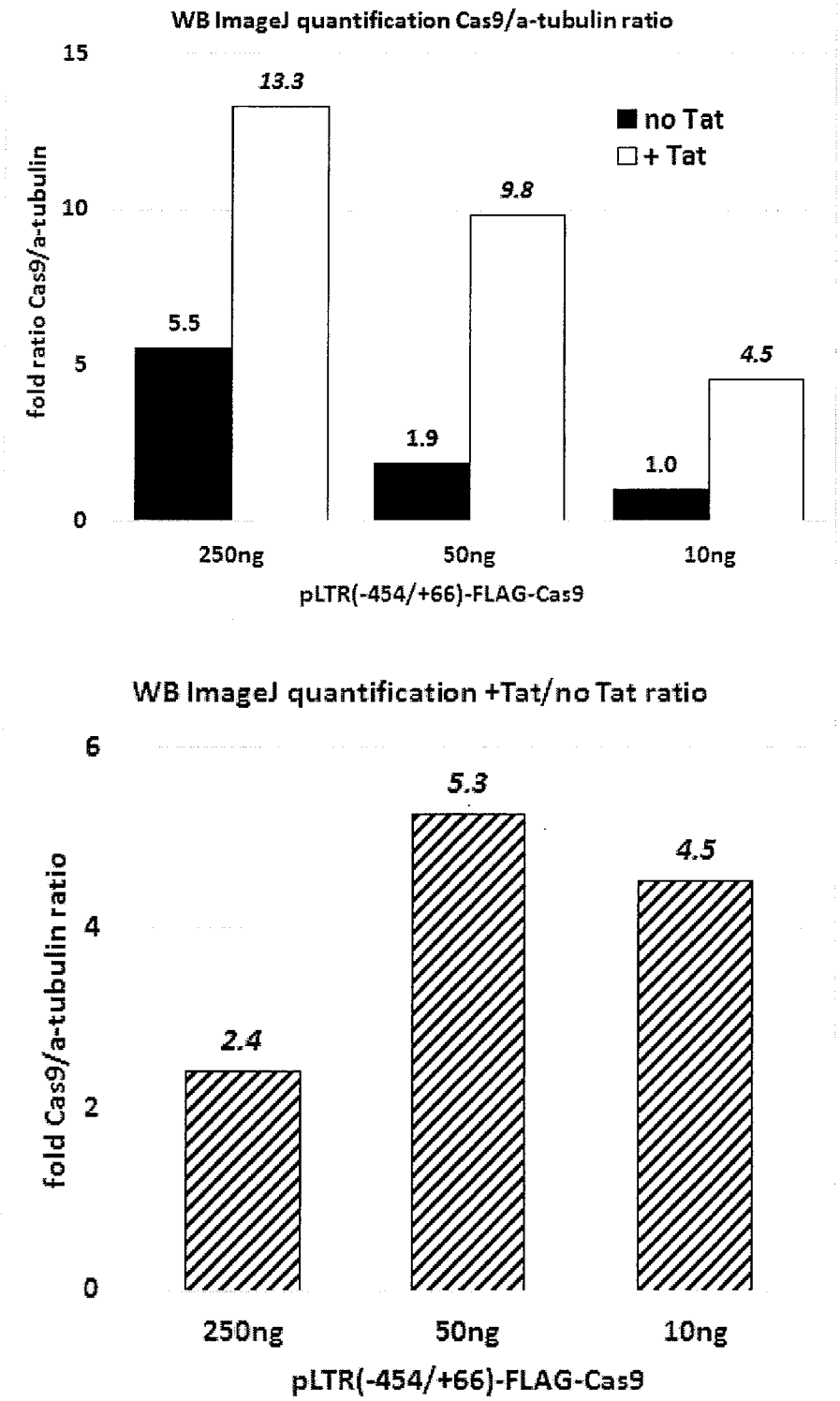
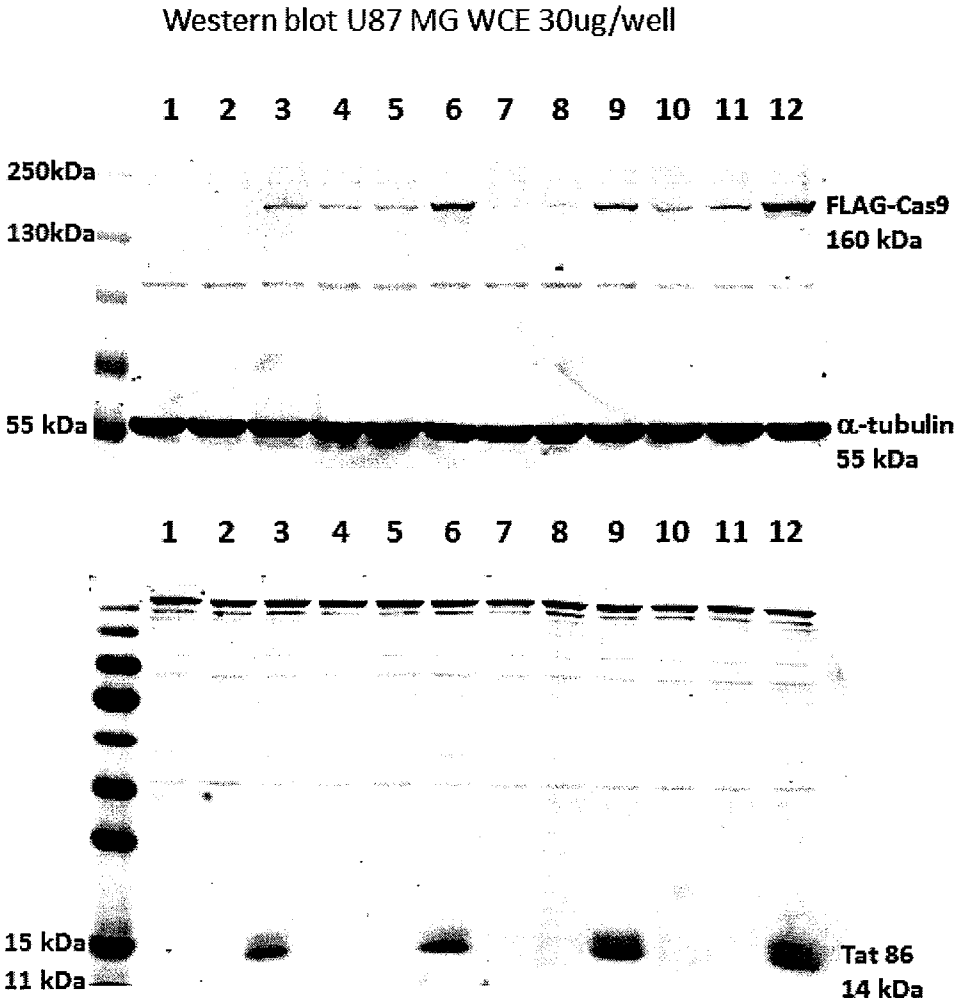
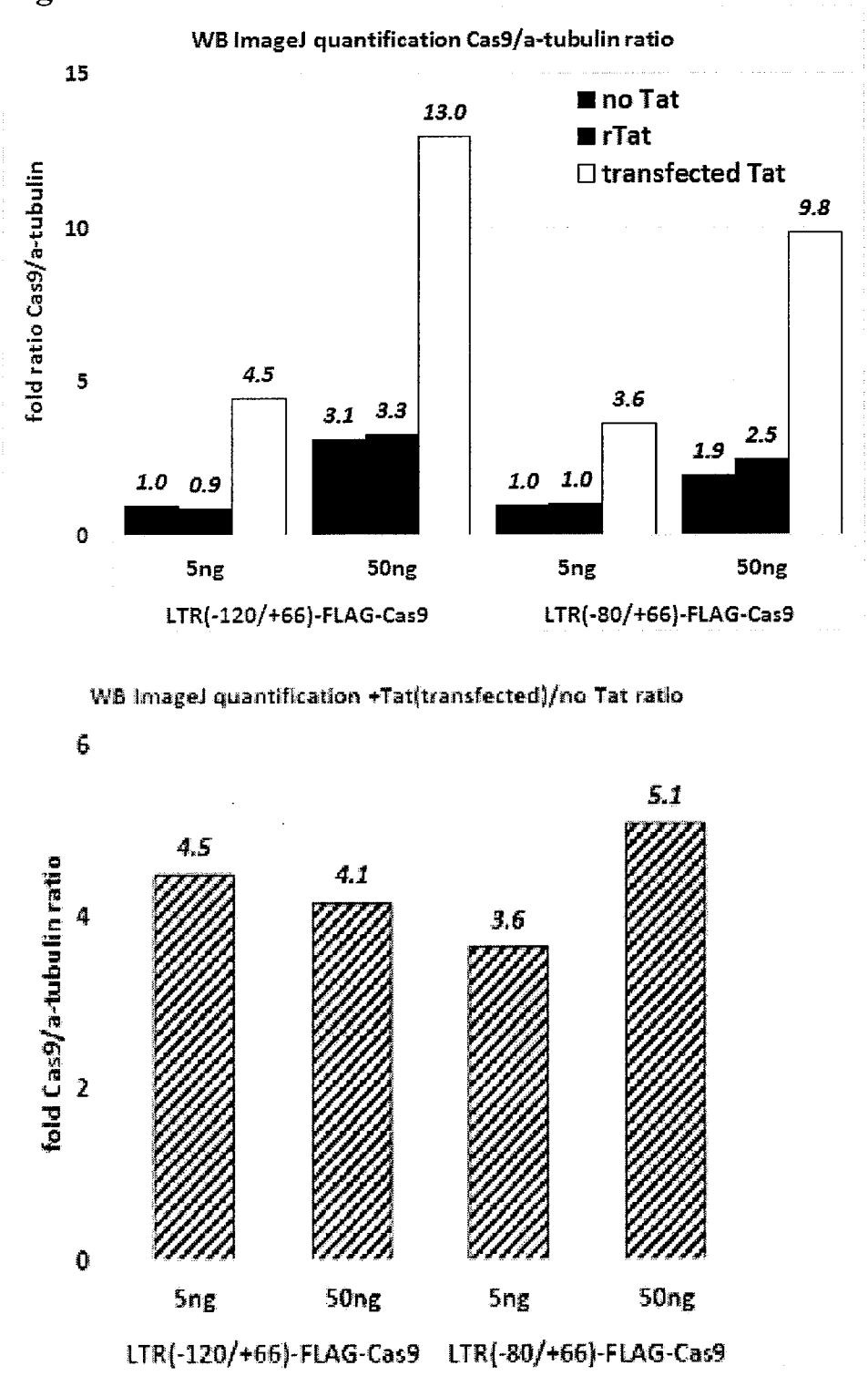

WB ImageJ quantification Cas9/a-tubulin ratio
fold ratio Cas9/a-tubulin
15
10
5
0
13.3
9.8
5.5
4.5
1.9
1.0
no Tat
+ Tat
250ng
50ng
10ng
pLTR(-454/+66)-FLAG-Cas9
WB ImageJ quantification +Tat/no Tat ratio
fold Cas9/a-tubulin ratio
6
4
2
0
5.3
4.5
2.4
250ng
50ng
10ng
pLTR(-454/+66)-FLAG-Cas9

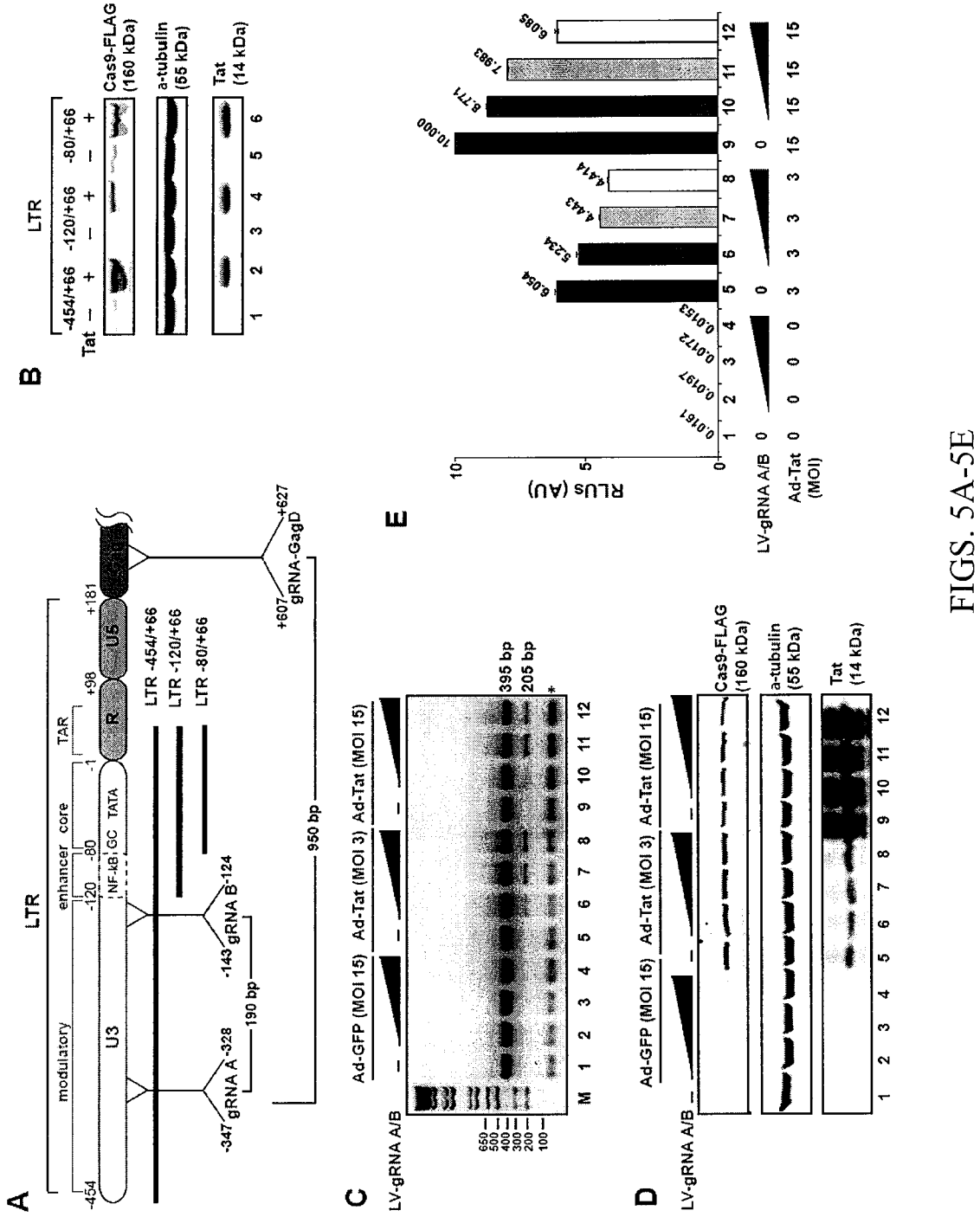
Fig. 4A



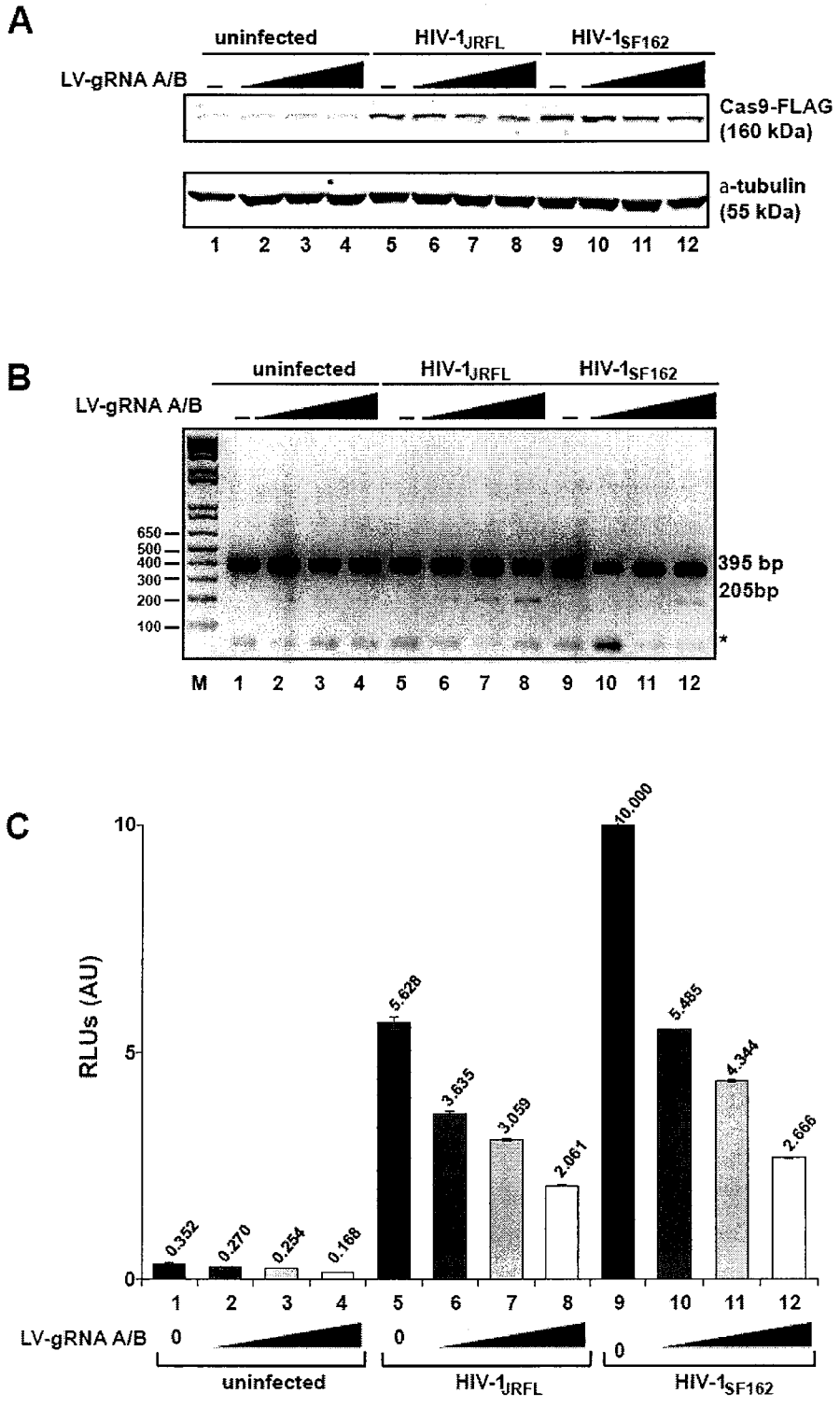
1. pLTR(-120/+66)-Cas9 5ng, pCMV 1245ng
2. pLTR(-120/+66)-Cas9 5ng, pCMV 1245ng, +rTat protein 2.5 ug/ml
3. pLTR(-120/+66)-Cas9 5ng, pCMV 995ng, pCMV-Tat86 250ng
4. pLTR(-120/+66)-Cas9 50ng, pCMV 1200ng,
5. pLTR(-120/+66)-Cas9 50ng, pCMV 1200ng, +rTat protein 2.5 ug/ml
6. pLTR(-120/+66)-Cas9 50ng, pCMV 950ng, pCMV-Tat86 250ng
7. pLTR(-80/+66)-Cas9 5ng, pCMV 1245ng
8. pLTR(-80/+66)-Cas9 5ng, pCMV 1245ng, +rTat protein 2.5 ug/ml
9. pLTR(-80/+66)-Cas9 5ng, pCMV 995ng, pCMV-Tat86 250ng
10. pLTR(-80/+66)-Cas9 50ng, pCMV 1200ng,
11. pLTR(-80/+66)-Cas9 50ng, pCMV 1200ng, +rTat protein 2.5 ug/ml
12. pLTR(-80/+66)-Cas9 50ng, pCMV 950ng, pCMV-Tat86 250ng

Fig. 4B


WB ImageJ quantification Cas9/a-tubulin ratio
fold ratio Cas9/a-tubulin
15
10
5
0
no Tat
rTat
transfected Tat
1.0
0.9
4.5
3.1
3.3
13.0
1.0
1.0
3.6
1.9
2.5
9.8
5ng
50ng
5ng
50ng
LTR(-120/+66)-FLAG-Cas9
LTR(-80/+66)-FLAG-Cas9
WB ImageJ quantification +Tat(transfected)/no Tat ratio
fold Cas9/a-tubulin ratio
6
4
2
0
4.5
4.1
3.6
5.1
5ng
50ng
5ng
50ng
LTR(-120/+66)-FLAG-Cas9
LTR(-80/+66)-FLAG-Cas9


A
LTR
modulatory
enhancer core
TAR
U3
R
U5
-454
-120
-80
-1
+98
+181
NF-kB
GC
TATA
LTR -454/+66
LTR -120/+66
LTR -80/+66
gRNA A -328
-347
190 bp
gRNA B -124
-143
+607
gRNA-GagD
+627
950 bp
B
LTR
-454/+66
-120/+66
-80/+66
Tat
Cas9-FLAG (160 kDa)
a-tubulin (55 kDa)
Tat (14 kDa)
C
Ad-GFP (MOI 15)
Ad-Tat (MOI 3)
Ad-Tat (MOI 15)
LV-gRNA A/B
395 bp
205 bp
650
500
400
300
200
100
M
D
Ad-GFP (MOI 15)
Ad-Tat (MOI 3)
Ad-Tat (MOI 15)
LV-gRNA A/B
Cas9-FLAG (160 kDa)
a-tubulin (55 kDa)
Tat (14 kDa)
E
RLUs (AU)
10
5
0
0.0161
0.0197
0.0172
0.0153
6.054
5.234
4.443
4.474
10.000
6.771
7.983
6.085
LV-gRNA A/B
Ad-Tat (MOI)

FIGS. 5A-5E


A
uninfected
HIV-1JRFL
HIV-1SF162
LV-gRNA A/B
Cas9-FLAG
(160 kDa)
a-tubulin
(55 kDa)
1 2 3 4 5 6 7 8 9 10 11 12
B
uninfected
HIV-1JRFL
HIV-1SF162
LV-gRNA A/B
650
500
400
300
200
100
395 bp
205bp
*
M 1 2 3 4 5 6 7 8 9 10 11 12
C
RLUs (AU)
10
5
0
0.352
0.270
0.254
0.168
5.628
3.635
3.059
2.061
10.000
5.485
4.344
2.666
1 2 3 4 5 6 7 8 9 10 11 12
LV-gRNA A/B
0
0
0
uninfected
HIV-1JRFL
HIV-1SF162

FIGS 6A-6C

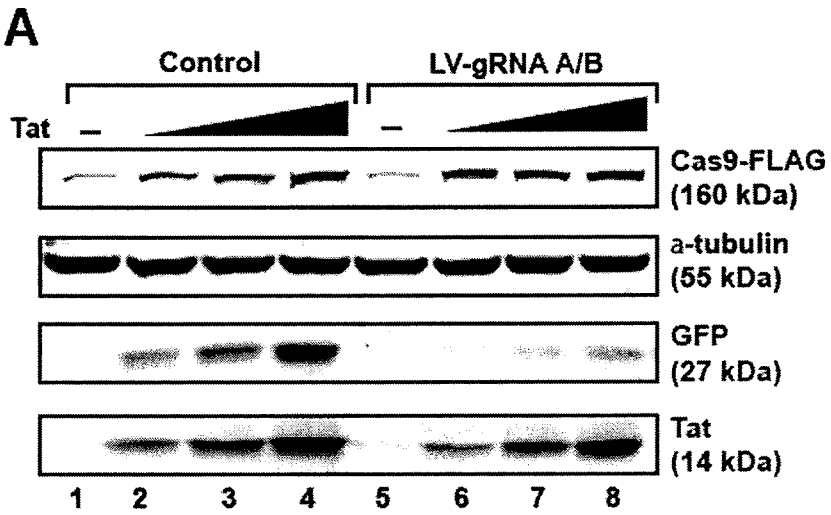

A
Control
LV-gRNA A/B
Tat
Cas9-FLAG
(160 kDa)
a-tubulin
(55 kDa)
GFP
(27 kDa)
Tat
(14 kDa)
1 2 3 4 5 6 7 8

FIG. 7A

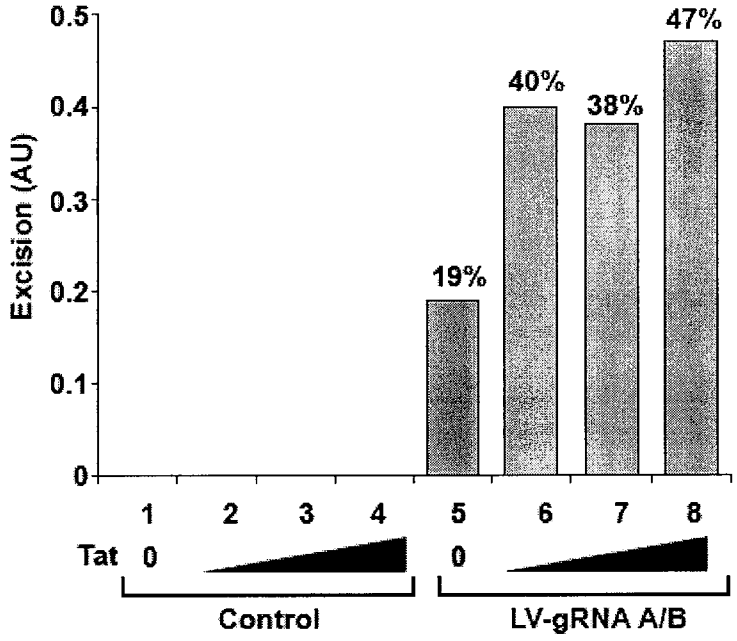

Excision (AU)
0.5
0.4
0.3
0.2
0.1
0
19%
40%
38%
47%
1 2 3 4 5 6 7 8
Tat 0 0
Control
LV-gRNA A/B

FIG.7B

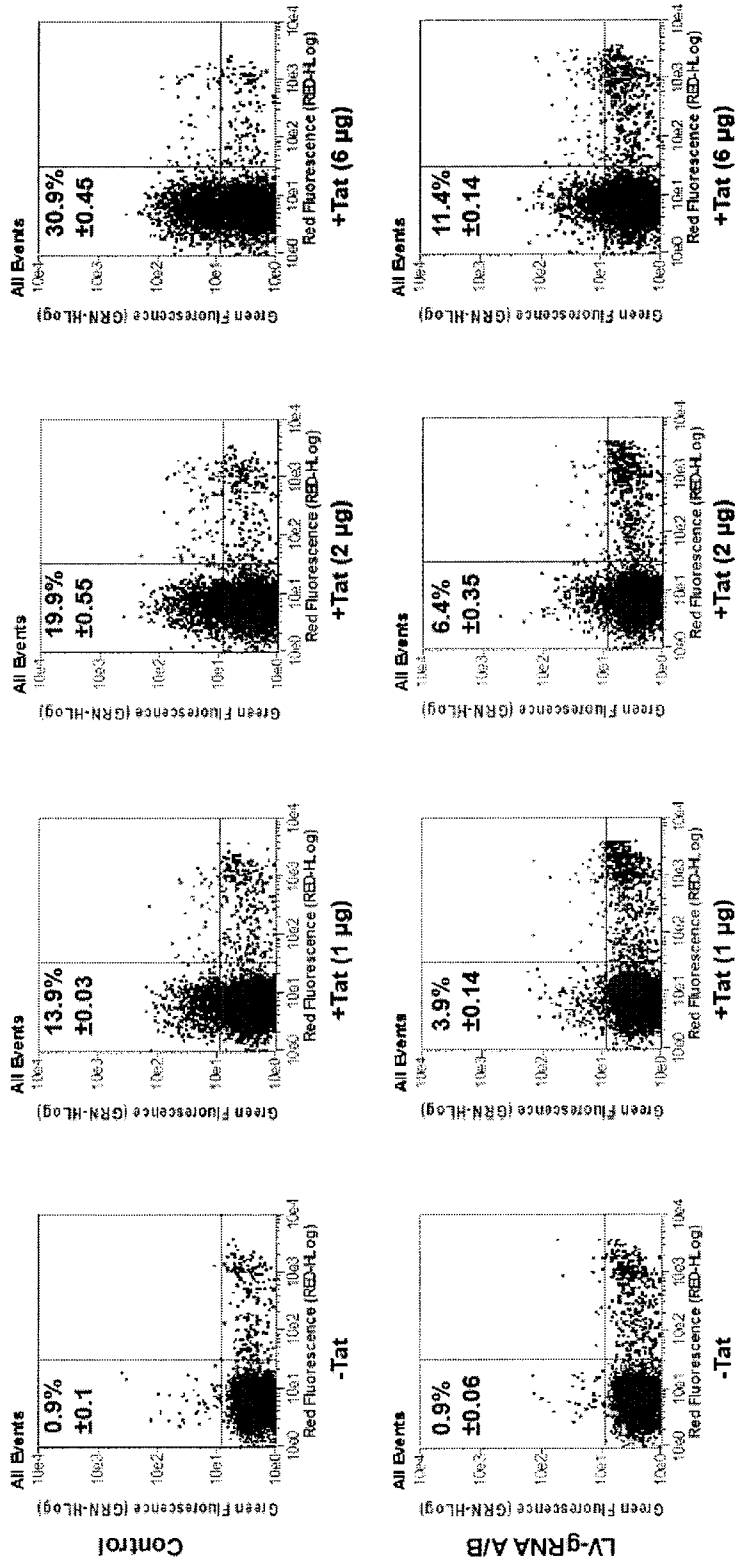

Control
LV-gRNA A/B
-Tat
+Tat (1 µg)
+Tat (2 µg)
+Tat (6 µg)
All Events
0.9% ±0.1
0.9% ±0.06
13.9% ±0.03
3.9% ±0.14
19.9% ±0.55
6.4% ±0.35
30.9% ±0.45
11.4% ±0.14
Green Fluorescence (GRN-HLog)
Red Fluorescence (RED-HLog)

FIG. 7C

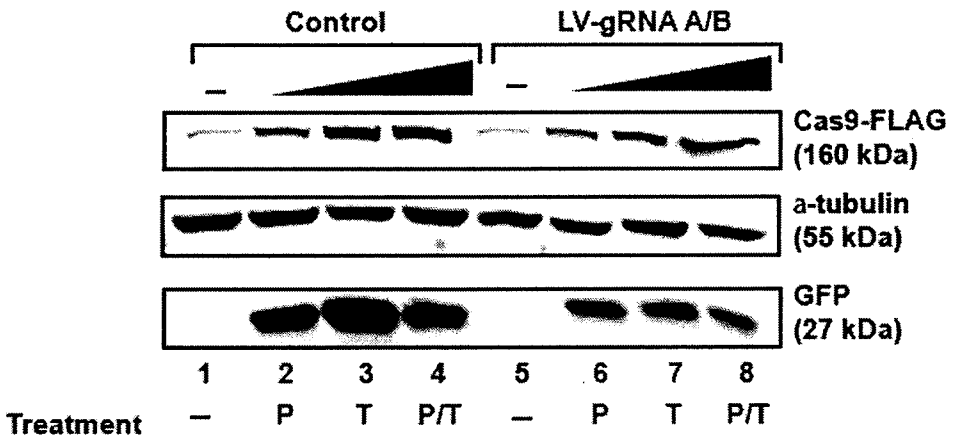

Control
LV-gRNA A/B
–
–
Cas9-FLAG
(160 kDa)
a-tubulin
(55 kDa)
GFP
(27 kDa)
1 2 3 4 5 6 7 8
Treatment – P T P/T – P T P/T

FIG. 8A

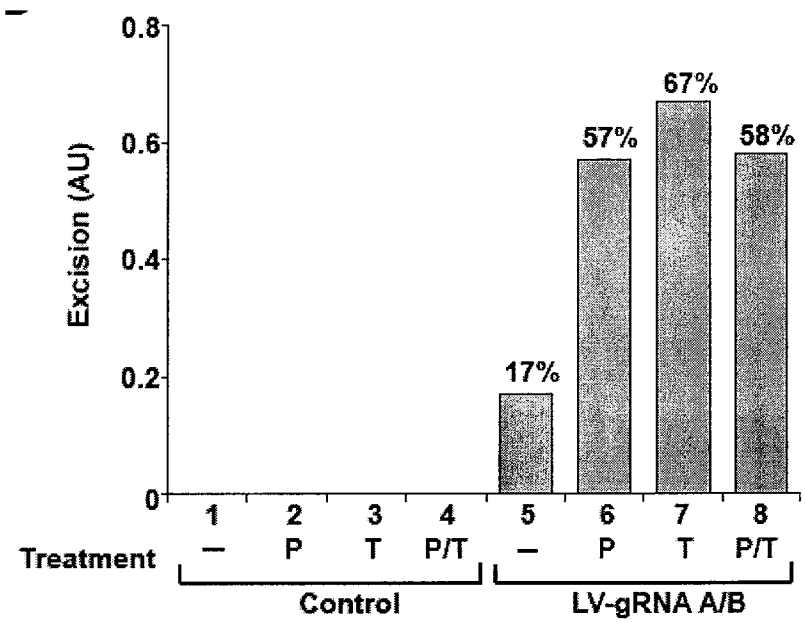

Excision (AU)
0.8
0.6
0.4
0.2
0
17%
57%
67%
58%
1 2 3 4 5 6 7 8
Treatment – P T P/T – P T P/T
Control
LV-gRNA A/B

FIG. 8B

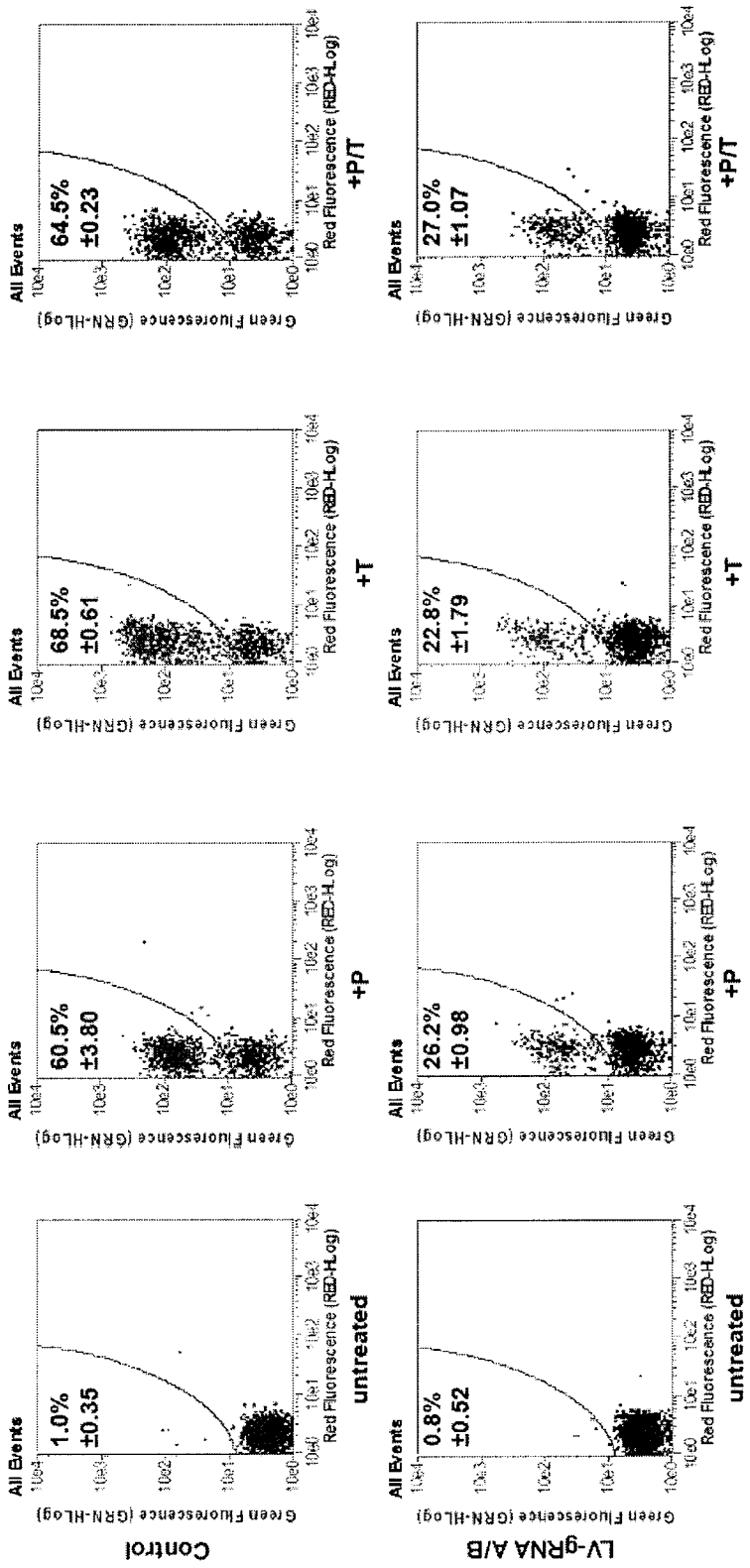

Control
LV-gRNA A/B
untreated
+P
+T
+P/T
All Events
1.0% ±0.35
60.5% ±3.80
68.5% ±0.61
64.5% ±0.23
0.8% ±0.52
26.2% ±0.98
22.8% ±1.79
27.0% ±1.07
Green Fluorescence (GRN-HLog)
Red Fluorescence (RED-HLog)

FIG. 8C

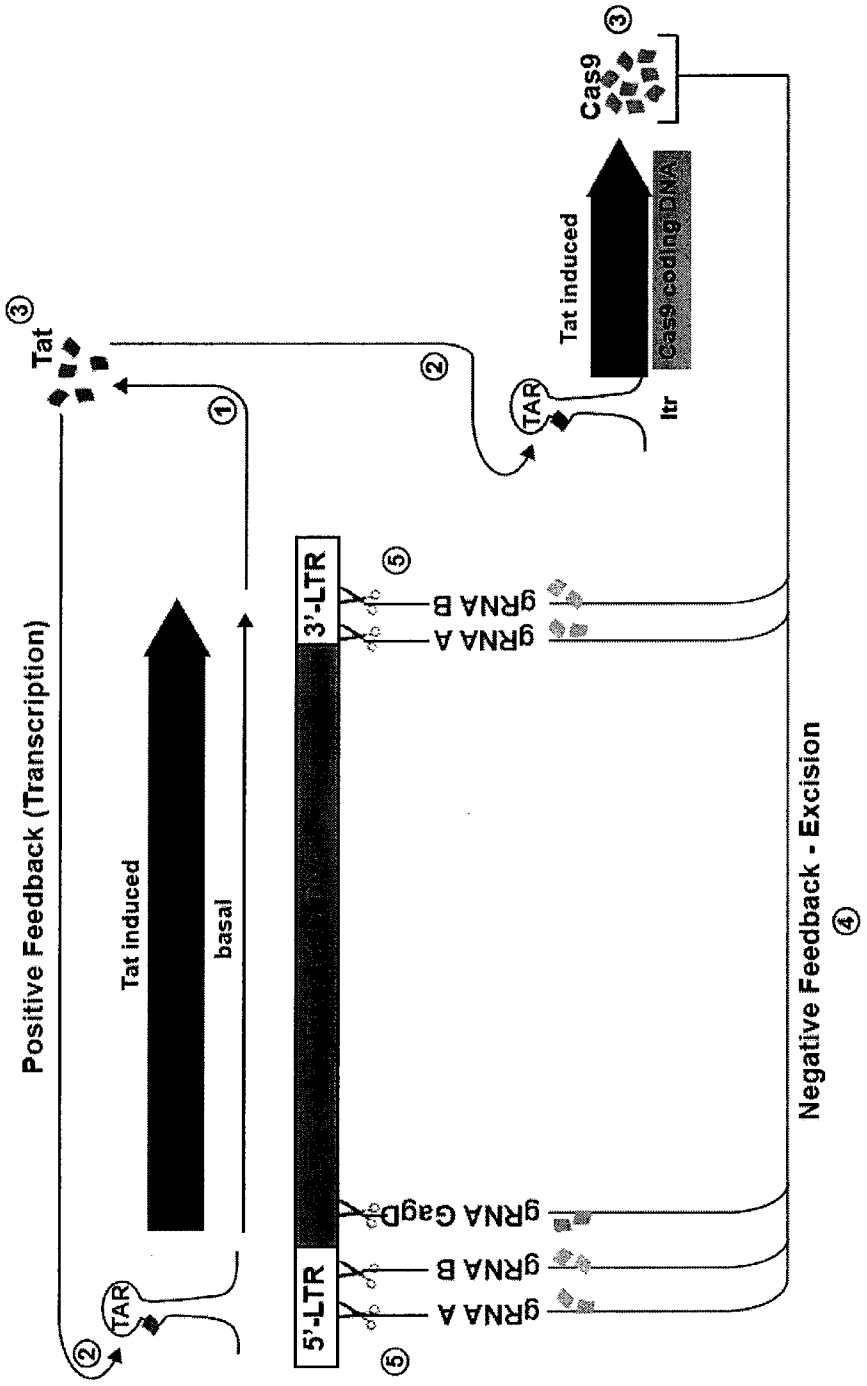

Positive Feedback (Transcription)
Tat
Tat induced
basal
TAR
5'-LTR
3'-LTR
gRNA A
gRNA B
gRNA GagD
Negative Feedback - Excision
Cas9
Tat induced
Cas9-coding DNA
ltr

FIG. 9

TZMbl (SEQ ID NO: 6)

F [-413/-391](T361)
TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTA
LTR A [-347/-328]PAM
CTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAGAGCAAGTAGAAGAGG
CCAATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGC
PAM **LTR B [-143/-124]**
ATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAG
R [-41/-19](T363)
GGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT
┌Transcription start[+1]
GCTACATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCA
ATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTG

PCR product full length LTR -413/-19: 395bp (SEQ ID NO: 7)

GATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGT
ACCAGTTGAACCAGAGCAAGTAGAAGAGGCCAATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAG
TGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCT
GGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAGC

+SpCas9/gRNA LTR A+LTR B (SEQ ID NO: 8):

GATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGAC

Edited fragment: 190bp (SEQ ID NO: 9)

↓CTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAGAGCAAGTAGAAGAGGCCAATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATG
GAGGACCCGGAGGGAGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTGCATCCGGAG↓
TACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATG
CTACATATAAGCAGC

PCR product truncated LTR -413/-19: 205bp (SEQ ID NO: 10)

GATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACTACTACAAAGACTGCTGACATCGA
GCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAGC

FIG. 10

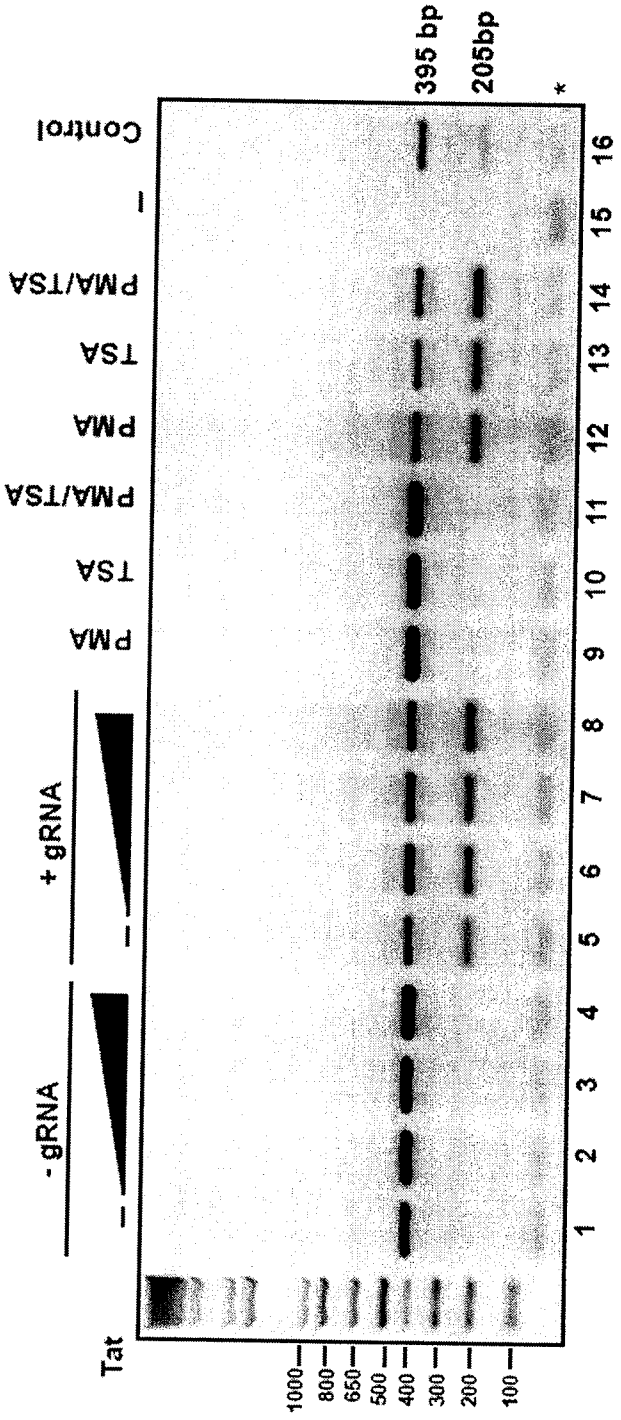

395 bp
205bp
*
Control
–
PMA/TSA
TSA
PMA
PMA/TSA
TSA
PMA
+ gRNA
– gRNA
Tat
1000
800
650
500
400
300
200
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

FIG. 11

Jurkat 2D10 (SEQ ID NO: 11):

TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTA
F [-375/-352] **LTR A [-347/-328]** PAM
CTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAAC
CAGAGCAAGTAGAAGAGGCCAATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAA
GTATTAGTGTGGAAGTTTGACAGCCTCCTAGC
PAM **LTR B [-143/-124]**
ATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAG
GGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT
Transcription start[+1] R [+19/+43]
GCTACATATAAGCAGCTGCTTTTTGCCTGTACTG**G**GTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAA
CCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTG

PCR product full length LTR -375/+43: 418bp (SEQ ID NO: 12)

TTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAGAGCAAG
TAGAAGAGGCCAATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAGTG
TGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTA
CAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAG
CTGCTTTTTGCCTGTACTG**G**GTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG

+SpCas9/gRNA LTR A+LTR B (SEQ ID NO: 13):

TTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGAC

Edited fragment: 190bp (SEQ ID NO: 14)

↓CTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAGAGCAAGTAGAAGAGGCCAATGAAGGAGAGAACAACAGCTTGTTACACCCT
ATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGA
GCTGCATCCGGAG↓

FIG. 12

PCR product truncated LTR -375/-43: 228bp (SEQ ID NO: 15):

TTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTC
CGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAGCTGCTTTTTGCC
TGTACTG**G**GTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG

SpCas9 versus SaCas9 targets in HIV-1 5'LTR and Gag:

AGGGATCAGATATCCACTGACCTTTGGATGGTGC **LTR A** [-347/-328]PAM **SpCas9 (SEQ ID NO: 16)**
AGGGATCAGATATCCACTGACCTTTGGATGGTGC **LTR 1** [-349/-329]PAM **SaCas9 (SEQ ID NO: 17)**

AAAGGATAGATGTAAAAGACACCAAGGAAGCCTT **Gag D** [+607/+627]PAM **SpCas9 (SEQ ID NO: 18)**
AAAGGATAGATGTAAAAGACACCAAGGAAGCCTT **Gag D** [+607/+627]PAM **SaCas9 (SEQ ID NO: 19)**

FIG. 12 (CONT.)

SpCas9 (PAM: NGG) target sequences (SEQ ID NO: 20):

Sequence:

F [-454/-431] F [-413/-391](T361)

TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTA

F [-375/-352] **LTR A [-347/-328]** PAM

CTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAAC

CAGAGCAAGTAGAAGAGGCCAATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAA

GTATTAGTGTGGAAGTTTGACAGCCTCCTAGC

PAM **LTR B [-143/-124]**

ATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAG

R [-41/-19](T363)

GGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT

┌**Transcription start[+1]** R [+19/+43]

GCTACATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAA

CCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCC

TCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGAGGAGATCTCTC

GACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGC

┌Gag gene start**[+336]**

GGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAAAAAATTCGGTTAAGGCCAG

GGGGAAAGAAACAATATAAACTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTTTTAGAGACA

TCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAATAGC

AGTCCTCTATTGT

Gag D [+607/+627] PAM

GTGCATCAAAGGATAGATGTAAAAGACACCAAGGAAGCCTTAGATAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCA

AGCAGCAGCTGACACAGGAAACAACAGCCAGGTCAGCCAAAATTACCCTATAGTGCAGAACCTCCAGGGGCAAATGGTACATCAGGCCA

TATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTATCA

GAAGGAGCCACCC

R [+888/+910](T458)

CACAAGATTTAAATACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGAC

FIG. 12 (CONT.)

SaCas9 (PAM: NNGRR(T/N)) target sequences (SEQ ID NO: 21):

Sequence:

PCR primer F [-413/-391](T361)

TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTA

LTR 1 [-349/-329] PAM

CTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAAC
CAGAGCAAGTAGAAGAGGCCAATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAA
GTATTAGTGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCG
AGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT

┌**Transcription start**[+1]

GCTACATATAAGCAGCTGCTTTTTGCCTGTACTG**G**GTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAA
CCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCC
TCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGAGGAGATCTCTC
GACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGC

┌Gag gene start[+336]

GGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAAAAAATTCGGTTAAGGCCAG
GGGGAAAGAAACAATATAAACTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTTTTAGAGACA
TCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAATAGC
AGTCCTCTATTGT

Gag D [+607/+627] PAM

GTGCATCAAAGGATAGATGTAAAAGACACCAAGCAAGCCTTAGATAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCA
AGCAGCAGCTGACACAGGAAACAACAGCCAGGTCAGCCAAAATTACCCTATAGTGCAGAACCTCCAGGGGCAAATGGTACATCAGGCCA
TATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTATCA
GAAGGAGCCACCC

PCR primer R [+888/+910] (T458)

CACAAGATTTAAATACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGAC

FIG. 12 (CONT.)

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
29 September 2016 (29.09.2016)

WIPO | PCT

(10) International Publication Number
WO 2016/154016 A3

(51) International Patent Classification:
A61P 31/12 (2006.01) *C12N 15/63* (2006.01)
A61P 31/18 (2006.01) *C12N 15/49* (2006.01)
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/US2016/023170

(22) International Filing Date:
18 March 2016 (18.03.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/136,080 20 March 2015 (20.03.2015) US

(71) Applicant: **TEMPLE UNIVERSITY OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION** [US/US]; Broad Street And Montgomery Avenue, Philadelphia, PA 19122 (US).

(72) Inventor: **KHALILI, Kamel**; 190 Presidential Boulevard, #718, Bala Cynwyd, PA 19004 (US).

(74) Agents: **ZACHARIADES, Nicholas, A.** et al.; Mchale & Slavin, P.A., 2855 PGA Boulevard, Palm Beach Gardens, FL 33410 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

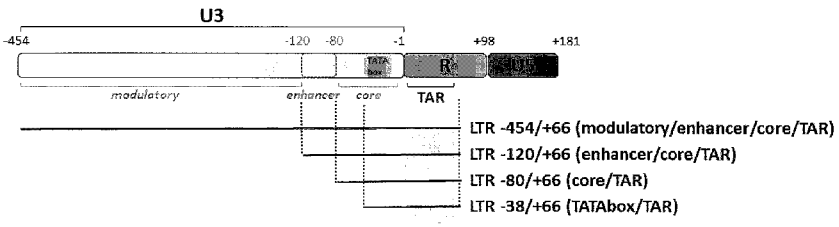
Published:

— *with international search report (Art. 21(3))*

(88) Date of publication of the international search report:
5 January 2017

(54) Title: TAT-INDUCED CRISPR/ENDONUCLEASE-BASED GENE EDITING

Fig. 1A



(57) Abstract: Compositions and methods are provided for Tat-inducible expression of a CRISPR-associated endonuclease by a truncated HIV LTR promoter containing at least a core region and a TAR region of a HIV LTR promoter. The compositions may be used as a therapeutic treatment for the treatment and/or prevention of HIV.

WO 2016/154016 A3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US16/23170

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - A61P 31/12, 31/18; C12N 9/22, 15/63, 15/49 (2016.01)
CPC - C12N 9/22, 15/635, 15/86; A61K 48/0058, 48/0066
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): A61P 31/12, 31/18; C12N 9/22, 15/63, 15/49, 15/31, 15/63, 9/22; C07K 14/155, 14/16; A61K 48/00; C07H 21/04
CPC: C07K 14/16, 14/163; C12N 9/22, 15/635, 15/86, 7/00; A61K 48/0058, 48/0066, 35/12, 45/06, 48/005, 9/0034

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); Google Scholar; Pubmed; EBSCO
Keywords: CRISPR, RNA, Cas9, HIV, virus, lentiviral, activate, mammal, human, administer, latent, infect, core, Tat, replication, inhibit, mother, infant

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	WO 2015/031775 A1 (TEMPLE UNIVERSITY OF THE COMMONWEALTH SYSTEM OF HIGHER EDUCATION) 05 March, 2015; abstract; page 2, paragraph 2; page 3, paragraph 1; page 10, paragraph 4; page 11, paragraphs 1-2; page 12, paragraph 1; page 13, paragraph 2; page 18, paragraph 2; page 21, paragraphs 1-2; page 27, paragraph 2; page 28, paragraph 2; page 29, paragraph 3; page 33, paragraph 1; page 34, paragraph 1; page 35, paragraph 1; page 36, paragraphs 1-2; page 36, paragraph 3 – page 37, paragraph 1; page 41, paragraph 2; page 42, paragraph 3page 44, paragraph 3; page 47, paragraph 3; figures 1, 13; claims 7-10, 17-18, 20-23, 26-29, 32, 35, 66-69, 77, 85	1-81, 83-86, 92 --- 82, 87-89
Y	(KARN, J et al.) Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. Cold Spring Harbor Perspectives in Medicine. February 2012; Vol. 2, No. 2; pages 1-17; abstract; page 1, column 1, paragraph 1; page 2, column 1, paragraph 1; figure 1; DOI: 10.1101/cshperspect.a006916.	82, 87-89
A	(CONG, L et al.) Multiplex Genome Engineering Using CRISPR/Cas Systems. Science. 15 February 2013; Vol. 339, No. 6121; pages 819-823; whole document; DOI: 10.1126/science.1231143	1-89, 92

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
22 August 2016 (22.08.2016)	26 SEP 2016
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US16/23170

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:
-Please See Supplemental Page-*-

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Group I: Claims 1-89, 92

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US16/23170

-***-Continued from Box No. III: Observations where unity of invention is lacking-***-

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, Claims 1-89 and 92 are directed toward an isolated nucleic acid sequence encoding a clustered regularly interspaced short palindromic repeats (CRISPR)-associated endonuclease operably linked to a truncated HIV long terminal repeat (LTR) promoter containing at least a core region and a trans activation response element (TAR) of a HIV LTR promoter; methods, nucleic acids, compositions and a kit associated therewith.

Groups II+, Claims 90, 91 and SEQ ID NO: 1 are directed toward an isolated nucleic acid sequence having at least a 75% sequence similarity to any one of SEQ ID NOS: 1 to 21; or comprising any one of SEQ ID NOS: 1 to 21 or combinations thereof.

The isolated nucleic acid will be searched to the extent that it encompasses SEQ ID NO: 1 (first exemplary nucleic acid sequence). Applicant is invited to elect additional nucleic acid sequence(s) with specified SEQ ID NO: for each CDR or specified substitution(s) at specified site(s) of a SEQ ID NO: (in order to provide a fully specified sequence), to be searched. Additional nucleic acid sequence(s) will be searched upon the payment of additional fees. It is believed that claims 90 (in-part) and 91 (in-part) encompass this first named invention and thus these claims will be searched without fee to the extent that they encompass SEQ ID NO: 1 (nucleic acid sequence). Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be a nucleic acid encompassing SEQ ID NO: 2 (first exemplary elected nucleic acid sequence).

No technical features are shared between the nucleic acid sequences of Groups II+ and, accordingly, these groups lack unity a priori.

The inventions listed as Groups I and II+ do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Group I include a CRISPR-associated nuclease, not present in any of Groups II+; the special technical features of Groups II+ include SEQ ID NO: 1, not present in Group I.

Groups I and II share the technical features including: HIV LTR sequences.

However, these shared technical features are previously disclosed by WO 2015/031775 A1 to Temple University of The Commonwealth System of Higher Education (hereinafter 'Temple').

Temple discloses HIV LTR sequences (target sequences within HIV-1 LTR U3 region (HIV LTR sequences); page 10, fourth paragraph).

Since none of the special technical features of the Groups I and II+ inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Temple reference, unity of invention is lacking.

(19)中华人民共和国国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 107949424 A
(43)申请公布日 2018.04.20

(21)申请号 201680029128.8

(22)申请日 2016.03.18

(30)优先权数据
62/136,080 2015.03.20 US

(85)PCT国际申请进入国家阶段日
2017.11.20

(86)PCT国际申请的申请数据
PCT/US2016/023170 2016.03.18

(87)PCT国际申请的公布数据
WO2016/154016 EN 2016.09.29

(71)申请人 天普大学-联邦高等教育系统
地址 美国宾夕法尼亚州

(72)发明人 K·哈利利

(74)专利代理机构 北京戈程知识产权代理有限公司 11314
代理人 程伟

(51)Int.Cl.
A61P 31/12(2006.01)
A61P 31/18(2006.01)
C12N 9/22(2006.01)
C12N 15/63(2006.01)
C12N 15/49(2006.01)

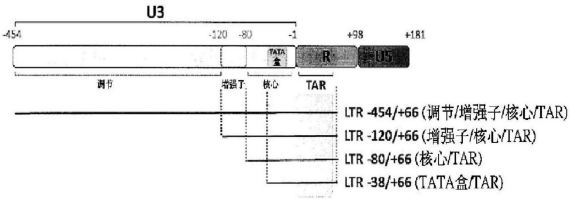
权利要求书5页 说明书30页 附图20页

(54)发明名称

TAT诱导的基于CRISPR/核酸内切酶的基因编辑

(57)摘要

提供组合物和方法，用于通过至少含有HIV LTR启动子的核心区域和TAR区域的截短的HIV LTR启动子进行的CRISPR相关核酸内切酶的Tat诱导型表达。该组合物可作为用于HIV的治疗和/预防的治疗性处理。



CN 107949424 A

1.一种在体内或体外将人免疫缺陷病毒(HIV)灭活的方法,该方法包含:向受试者给药包含分离的核酸序列的组合物或令哺乳动物细胞与该组合物接触,该分离的核酸序列编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶,该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子,而该截短的启动子至少含有HIV LTR启动子的核心区域和反式激活响应元件(TAR)。

2.如权利要求1所述的方法,其中,该分离的核酸序列进一步编码至少一种与HIV中靶标核酸序列互补的向导RNA。

3.如权利要求2所述的方法,其中,该HIV中的靶标核酸序列包含位于HIV编码区域或非编码区域内的序列。

4.如权利要求3所述的方法,其中,该非编码区域包含HIV的长末端重复序列。

5.如权利要求3所述的方法,其中,该靶标核酸序列包含位于HIV长末端重复序列内的序列。

6.如权利要求5所述的方法,其中,该位于HIV长末端重复序列内的序列包含位于U3区、R区、或U5区内的序列,而这些区域不包括该截短的HIV LTR启动子的任何序列。

7.如权利要求1所述的方法,其中,该哺乳动物细胞是潜伏性感染细胞。

8.如权利要求7所述的方法,其中,该潜伏性感染细胞选自下列所组成的群组:$CD4^+$T细胞、巨噬细胞、单核细胞、肠相关淋巴样细胞、小胶质细胞、和星形胶质细胞。

9.如权利要求1所述的方法,其中,该灭活是体内灭活或间接体内灭活。

10.如权利要求9所述的方法,其中,该哺乳动物细胞包含来自具备HIV感染的受试者的培养细胞、组织分离块、或细胞系。

11.如权利要求1所述的方法,其中,该CRISPR相关核酸内切酶是Cas9。

12.如权利要求1所述的方法,其中,该CRISPR相关核酸内切酶被优化用于人细胞内表达。

13.如权利要求1所述的方法,其中,该组合物进一步包含编码反式激活小RNA(tracrRNA)的序列。

14.如权利要求13所述的方法,其中,该tracrRNA与编码向导RNA的序列融合。

15.如权利要求1所述的方法,其中,该组合物进一步包含编码核定位信号的序列。

16.如权利要求1所述的方法,其中,该组合物可操作地链接至表达载体。

17.如权利要求16所述的方法,其中,该表达载体选自下列所组成的群组:慢病毒载体、腺病毒载体、及腺病毒群载体。

18.如权利要求1所述的方法,其中,该组合物进一步包含该HIV-1LTR启动子的增强子区域。

19.一种分离的核酸序列,其包含编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶的序列,该酶可操作地链接至截短的人免疫缺陷病毒(HIV)长末端重复序列(LTR)启动子,而该截短的启动子至少含有HIV LTR启动子的核心区域和反式激活响应元件(TAR)。

20.如权利要求19所述的分离的核酸序列,其中,该序列进一步编码至少一种与HIV中靶标核酸序列互补的向导RNA。

21.如权利要求20所述的分离的核酸序列,其中,该HIV中的靶标核酸序列包含位于HIV

编码区域或非编码区域内的序列。

22.如权利要求21所述的分离的核酸序列,其中,该非编码区域包含HIV的长末端重复序列或位于该HIV长末端重复序列内的序列。

23.如权利要求22所述的分离的核酸序列,其中,该位于HIV长末端重复序列内的序列包含位于U3区、R区、或U5区内的序列,而这些区域不包括该截短的HIV LTR启动子的任何序列。

24.如权利要求19所述的分离的核酸序列,其中,该CRISPR相关核酸内切酶是Cas9。

25.如权利要求19所述的分离的核酸序列,其中,该CRISPR相关核酸内切酶被优化用于人细胞内表达。

26.如权利要求19所述的分离的核酸序列,进一步包含编码反式激活小RNA(tracrRNA)的序列。

27.如权利要求26所述的分离的核酸序列,其中,该tracrRNA与编码向导RNA的序列融合。

28.如权利要求19所述的分离的核酸序列,进一步包含编码核定位信号的序列。

29.如权利要求19所述的分离的核酸序列,其中,该分离的核酸序列可操作地链接至表达载体。

30.如权利要求29所述的分离的核酸序列,其中,该表达载体选自下列所组成的群组:慢病毒载体、腺病毒载体、及腺病毒群载体。

31.如权利要求19所述的分离的核酸序列,进一步包含该HIV LTR启动子的增强子区域。

32.一种药物组合物,其包含编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶的核酸序列,该酶可操作地链接至截短的人免疫缺陷病毒(HIV)长末端重复序列(LTR)启动子,而该截短的启动子至少含有HIV LTR启动子的核心区域和反式激活响应元件(TAR)。

33.如权利要求32所述的药物组合物,进一步包含药学可接受的载剂。

34.如权利要求33所述的药物组合物,其中,该药学可接受的载剂包含基于脂质的胶体或基于聚合物的胶体。

35.如权利要求34所述的药物组合物,其中,该胶体是脂质体、水凝胶、微粒、纳米颗粒、或嵌段共聚物胶束。

36.如权利要求32所述的药物组合物,其中,该药物组合物配制为外用。

37.如权利要求32所述的药物组合物,其中,该药物组合物包含在避孕套内。

38.如权利要求32所述的药物组合物,其中,该序列进一步编码至少一种与HIV中靶标核酸序列互补的向导RNA。

39.如权利要求38所述的药物组合物,其中,该HIV中靶标核酸序列包含位于HIV的编码区域或非编码区域内的序列。

40.如权利要求39所述的药物组合物,其中,该非编码区域包含HIV的长末端重复序列。

41.如权利要求40所述的药物组合物,其中,该靶标核酸序列包含位于该HIV长末端重复序列内的序列。

42.如权利要求41所述的药物组合物,其中,该位于HIV长末端重复序列内的序列包含

位于U3区、R区、或U5区内的序列，而这些区域不包括该截短的HIV LTR启动子的任何序列。

43.如权利要求32所述的药物组合物，其中，该CRISPR相关核酸内切酶是Cas9。

44.如权利要求32所述的药物组合物，其中，该CRISPR相关核酸内切酶被优化用于人细胞内表达。

45.如权利要求32所述的药物组合物，进一步包含编码反式激活小RNA(tracrRNA)的核酸序列。

46.如权利要求45所述的药物组合物，其中，该tracrRNA与编码向导RNA的序列融合。

47.如权利要求32所述的药物组合物，进一步包含编码核定位信号的序列。

48.如权利要求32所述的药物组合物，其中，该药物组合物可操作地链接至表达载体。

49.如权利要求48所述的药物组合物，其中，该表达载体选自下列所组成的群组：慢病毒载体、腺病毒载体、和腺病毒群载体。

50.如权利要求32所述的药物组合物，其中，该序列进一步编码该HIV-1LTR启动子的增强子区域。

51.一种治疗具有人免疫缺陷病毒(HIV)感染的受试者的方法，该方法包含向该受试者给药治疗有效量的包含核酸序列的药物组合物，该核酸序列编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶，该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和反式激活响应元件(TAR)。

52.如权利要求51所述的治疗具有HIV感染的受试者的方法，其中，该序列进一步编码至少一种与HIV中靶标核酸序列互补的向导RNA。

53.如权利要求51所述的治疗具有HIV感染的受试者的方法，其中，该HIV感染是潜伏性感染。

54.如权利要求51所述的治疗具有HIV感染的受试者的方法，其中，该药物组合物是外用给药或肠胃外给药。

55.如权利要求51所述的治疗具有HIV感染的受试者的方法，进一步包含给药抗逆转录病毒剂。

56.如权利要求55所述的治疗具有HIV感染的受试者的方法，其中，该抗逆转录病毒剂选自非核苷逆转录酶抑制剂、蛋白酶抑制剂和进入抑制剂所组成的群组。

57.如权利要求56所述的治疗具有HIV感染的受试者的方法，其中，该抗逆转录病毒剂包含高活性的抗逆转录病毒治疗。

58.如权利要求51所述的治疗具有HIV感染的受试者的方法，其中，该CRISPR相关核酸内切酶是Cas9。

59.如权利要求57所述的治疗具有HIV感染的受试者的方法，其中，该药物组合物可操作地链接至表达载体。

60.如权利要求59所述的治疗具有HIV感染的受试者的方法，其中，该表达载体选自下列所组成的群组：慢病毒载体、腺病毒载体、和腺病毒群载体。

61.如权利要求57所述的治疗具有HIV感染的受试者的方法，其中，该序列进一步编码该HIV LTR启动子的增强子区域。

62.一种降低处于人免疫缺陷病毒(HIV)感染风险下的受试者感染HIV的风险的方法，该方法包含向该受试者给药治疗有效量的包含核酸序列的药物组合物，该核酸序列编码成

簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶,该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子,而该截短的启动子至少含有HIV LTR启动子的核心区域和反式激活响应元件(TAR)。

63.如权利要求69所述的降低处于HIV感染风险下的受试者感染HIV的风险的方法,其中,该受试者是性活跃者、医护人员或现场急救员。

64.如权利要求62所述的降低处于HIV感染风险下的受试者感染HIV的风险的方法,其中,该CRISPR相关核酸内切酶是Cas9。

65.如权利要求62所述的降低处于HIV感染风险下的受试者感染HIV的风险的方法,其中,该CRISPR相关核酸内切酶被优化用于人细胞内表达。

66.如权利要求62所述的降低处于HIV感染风险下的受试者感染HIV的风险的方法,其中,该药物组合物可操作地链接至表达载体。

67.如权利要求67所述的降低处于HIV感染风险下的受试者感染HIV的风险的方法,其中,该表达载体选自下列所组成的群组:慢病毒载体、腺病毒载体、和腺病毒群载体。

68.如权利要求67所述的降低处于HIV感染风险下的受试者感染HIV的风险的方法,其中,该序列进一步编码该HIV LTR启动子的增强子区域。

69.一种降低人免疫缺陷病毒(HIV)感染从感染HIV的妊娠期或哺乳期母亲向其子女传播的风险的方法,该方法包含向该受试者给药治疗有效量的包含核酸序列的药物组合物,该核酸序列编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶,该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子,而该截短的启动子至少含有HIV LTR启动子的核心区域和反式激活响应元件(TAR)。

70.如权利要求69所述的降低HIV感染从感染HIV的妊娠期或哺乳期母亲向其子女传播的风险的方法,其中,该药物组合物在产前、围产期和产后的至少一个阶段给药。

71.如权利要求69所述的降低HIV感染从感染HIV的妊娠期或哺乳期母亲向其子女传播的风险的方法,进一步包含给药抗逆转录病毒剂。

72.如权利要求71所述的降低HIV感染从感染HIV的妊娠期或哺乳期母亲向其子女传播的风险的方法,其中,该抗逆转录病毒剂选自非核苷逆转录酶抑制剂、蛋白酶抑制剂、和进入抑制剂所组成的群组。

73.如权利要求72所述的降低HIV感染从感染HIV的妊娠期或哺乳期母亲向其子女传播的风险的方法,其中,该抗逆转录病毒剂包含高活性的抗逆转录病毒治疗。

74.如权利要求69所述的降低HIV感染从感染HIV的妊娠期或哺乳期母亲向其子女传播的风险的方法,进一步包含向该子女给药治疗有效量的该组合物。

75.如权利要求69所述的降低HIV感染从感染HIV的妊娠期或哺乳期母亲向其子女传播的风险的方法,其中,该序列进一步编码该HIVLTR启动子的增强子区域。

76.一种给药药物组合物以防止未感染的受试者感染人免疫缺陷病毒(HIV)的方法,该方法包含向该未感染的受试者给药治疗有效量的包含核酸序列的药物组合物,该核酸序列编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶,该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子,而该截短的启动子至少含有HIV LTR启动子的核心区域和反式激活响应元件(TAR)。

77.一种包含核酸序列的组合物,该核酸序列编码成簇的规律间隔短回文重复序列

(CRISPR)相关核酸内切酶,该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子,而该截短的启动子至少含有HIV LTR启动子的核心区域和反式激活响应元件(TAR)。

78.如权利要求77所述的组合物,其中,该CRISPR相关核酸内切酶被优化用于人细胞内表达。

79.如权利要求77所述的组合物,其中,该CRISPR相关核酸内切酶是Cas。

80.如权利要求79所述的组合物,其中,该CRISPR相关核酸内切酶是Cas9(CRISPR/Cas9)。

81.如权利要求79所述的组合物,其中,该CRISPR/Cas处于最小HIV LTR启动子的控制下。

82.如权利要求81所述的组合物,其中,在HIV复制的早期,该最小截短的HIV LTR启动子被即刻早期转录激活因子(Tat)激活,从而有条件地激活CRISPR/Cas。

83.如权利要求77所述的组合物,进一步包含至少一种与HIV中靶标核酸序列互补的向导RNA。

84.如权利要求77所述的组合物,其中,该组合物进一步包含编码反式激活小RNA(tracrRNA)的序列。

85.如权利要求84所述的组合物,其中,该tracrRNA与编码向导RNA的序列融合。

86.如权利要求79所述的组合物,其中,该CRISPR/Cas切除病毒基因组的一段,该段跨越病毒启动子及/或病毒编码序列。

87.一种包含核酸序列的组合物,该核酸序列编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶(CRISPR/Cas)且该酶可操作地链接至截短的病毒启动子,借此,在病毒复制的早期,该截短的病毒启动子处于即可早期转录活化在的控制下,从而有条件地激活CRISPR/Cas。

88.如权利要求87所述的组合物,进一步包含至少一种与该病毒中靶标核酸序列互补的向导RNA。

89.如权利要求87所述的组合物,其中,该CRISPR/Cas切除病毒基因组的一段,该段跨越病毒启动子及/或病毒编码序列。

90.一种分离的核酸序列,其与SEQ ID NO:1至SEQ ID NO:21的任一者具有至少75%的序列相似性。

91.一种分离的核酸序列,其包含SEQ ID NO:1至SEQ ID NO:21的任一者或其组合。

92.一种试剂盒,其包含被测量的组合物,该组合物包含分离的核酸序列,该分离的核酸序列包含编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶的序列,该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子,而该截短的启动子含有HIV LTR启动子的至少核心区域和反式激活响应元件(TAR)区域,或该组合物包含编码该分离的核酸序列的载体;以及选自包装材料、包含使用说明书的包装插页、无菌流体、注射器、及无菌容器所组成群组的至少一项。

TAT诱导的基于CRISPR/核酸内切酶的基因编辑

技术领域

[0001] 本发明涉及特异性裂解逆转录病毒如人免疫缺陷病毒(HIV)中靶标序列的组合物。这些可包括核酸的组合物可给药至患有HIV感染或处于HIV感染风险下的受试者，其中，该核酸编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶以及与人免疫缺陷病毒中靶标序列互补的向导RNA序列。

背景技术

[0002] 自HIV-1发现以来，AIDS仍是世界范围内影响数百万人的主要公共健康问题。由于HIV-1永久性整合入宿主基因组中，AIDS尚不可治愈。目前，用于控制HIV-1感染和阻止AIDS发展的治疗(高活性的抗逆转录病毒治疗或HAART)极大地降低了细胞中支持HIV-1感染的病毒复制，并将血浆病毒血症降至最低水平。但HAART未能阻抑组织内低水平的病毒基因组表达和复制，且未能以被潜伏性感染的细胞如作为HIV-1积累池的休眠的记忆性T细胞、脑巨噬细胞、小胶质细胞、星形胶质细胞和肠相关淋巴样细胞为靶标。顽固性HIV-1感染也导致包括心脏和肾脏疾病、骨量减少、和神经障碍的并发症。对于以顽固性病毒储存库(reservoir)为靶标的有疗效的治疗策略存在持续需求。

[0003] HIV-1基因组的长度约为9.8kb，包括当整合入宿主基因组时位于两端的两个病毒长末端重复序列。该基因组也包括编码结构蛋白(Gag、Pol和Env)、调节蛋白(Tat和Rev)、以及辅助蛋白(Vpu、Vpr、Vif和Nef)的基因。HIV-1转录的反式激活因子(Tat)是一种多功能蛋白，已经发现其促成HIV-1感染的若干病理学后果。Tat不仅在病毒转录和复制中扮演重要角色，它也能诱导多种细胞基因的表达且能作为神经毒性蛋白而起作用。Tat蛋白由被HIV-1感染的细胞分泌，且通过扩散穿过细胞膜而起作用。它可作为隐藏的、可溶的神经毒素而起作用，且诱导被HIV-1感染的巨噬细胞和小胶质细胞释放神经毒性物质。Tat转录由HIV-1LTR启动子驱动，且为HIV的整体病毒复制所需。

[0004] HIV感染的临床进程可根据大量因素而变，该因素包括受试者的遗传背景、年龄、一般健康状况、营养情况、所接受的治疗、以及HIV亚型。通常，大多数个体在感染几周或几个月内发展出类似流感的症状。该症状可包括发热、头痛、肌痛、皮疹、发冷、咽喉痛、口腔或生殖器溃疡、淋巴结肿大、关节痛、盗汗、和腹泻。症状的强度可依据不同个体而从轻度到重度变化。在急性期过程中，HTV病毒颗粒被吸引至表达适宜的CD4受体分子的细胞并进入该细胞。一旦病毒进入宿主细胞，编码HIV的逆转录酶即生成HIV RNA的前病毒DNA副本，且该前病毒DNA副本变成被整合入该宿主细胞基因组DNA中。就是这一被该宿主细胞复制的HIV前病毒，导致了新的HIV病毒体的释放，该新的HIV病毒体可随后感染其它细胞。

[0005] 初次HIV感染在数周至数月内平息，但典型地，其后为可能延续长达10年的长期临床“潜伏”阶段。这一潜伏期也指代为无症状的HIV感染或慢性HIV感染。受试者的CD4淋巴细胞数回升但未至感染前的水平，且大多数受试者在感染2至4周内进行血清转化，换言之，他们体内具有可检测水平的抗HIV抗体。这一潜伏期过程中，在周边血液单核细胞中可能不存在可检测的病毒复制，且在周边血液中可能存在少量或没有可培养的病毒。该潜伏期过程

中，也指代为临床潜伏阶段，感染有HIV的人们可能不经历HIV有关症状，或仅经历轻度症状。但是，HIV病毒持续在非常低的水平上繁殖。在已经使用抗逆转录病毒治疗进行治疗的受试者体内，这一潜伏期可延续数十年甚至更久。然而，即使处于这一阶段的受试者正在接受抗逆转录病毒治疗，他们仍能将HIV传播至其他人，但抗逆转录病毒治疗降低了传播的风险。

[0006] CRISPR(成簇的规律间隔短回文重复序列)是含有短反复碱基序列的DNA基因座。每一反复后跟随从前病毒曝露至病毒的“间隔DNA”的短片段。CRISPR一般与Cas基因相关，该Cas基因编码与CRISPR相关的蛋白质。该CRISPR/Cas系统是原核生物免疫系统，其赋予对外来遗传成分如质粒和噬菌体的抗性，并提供一种获取免疫性的形式。CRISPR间隔以类似于真核生物有机体中RNAi的方式来识别并切割这些外源性遗传成分。

[0007] 该CRISPR/Cas系统已经被用于多种有机体中的基因编辑(通过加入、中断或改变具体基因的序列)和基因调节。通过将该Cas9蛋白和适宜的向导RNA输送至细胞内，可在任何所希望的位置切割该有机体的基因组。使用CRISPR/Cas9系统的成功的治疗性基因编辑需要Cas9酶和向导RNA在靶标细胞中的有效且特异性的输送和表达。当组织或细胞集群中感受细胞的频率低时，如在某些被病毒感染的细胞的情境中，这尤其是一种挑战。

发明内容

[0008] 本文提供一种在体内或体外将哺乳动物细胞内的人免疫缺陷病毒(HIV)灭活的方法。该方法包括将该哺乳动物细胞曝露于包括分离的核酸序列的组合物，分离的核酸序列编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶，该酶可操作地链接至截短的HIV LTR启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和TAR区域。

[0009] 具体实施例中，该CRISPR相关核酸内切酶是Cas9。该CRISPR相关核酸内切酶可被优化由于人细胞内表达。将该哺乳动物细胞曝露于该组合物可包括接触该细胞。该哺乳动物细胞可以是潜伏性感染的细胞，包括但不限于，$CD4^+$T细胞、巨噬细胞、单核细胞、肠相关淋巴样细胞、小胶质细胞、和星形胶质细胞。该哺乳动物细胞可包括来自具有HIV感染的受试者的培养细胞、组织分离块、及/或细胞系。该HIV的灭活可在体内实施或间接体内实施。

[0010] 具体实施例中，该分离的核酸可额外编码一种或多种与HIV中靶标核酸序列互补的向导RNA。该HIV中靶标核酸序列可指代位于HIV的编码区域及/或非编码区域及/或长末端重复序列内的序列。该非编码区域可包括HIV的长末端重复序列。该位于HIV的长末端重复序列内的序列可包括位于U3区、R区、或U5区内的序列，这些区域不包括该截短的HIV LTR启动子的任何序列。该组合物可包括编码核定位信号的序列。该组合物可额外包括编码反式激活小RNA(tracrRNA)的序列，且该tracrRNA可与一编码向导RNA的序列融合。该组合物亦可包括该HIV LTR启动子的增强子区域。

[0011] 具体实施例中，该组合物可被可操作地链接至表达载体。该表达载体可以是慢病毒载体、腺病毒载体、和腺病毒群载体。

[0012] 本文提供一种分离的核酸序列，其包括编码CRISPR相关核酸内切酶的序列，该酶可操作地链接至截短的HIV LTR启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和TAR区域。

[0013] 具体实施例中，该CRISPR相关核酸内切酶是Cas9。该CRISPR相关核酸内切酶可被

优化用于人细胞内表达。

[0014] 具体实施例中，该序列可额外编码一种或多种与HIV中靶标核酸序列互补的向导RNA。该HIV中靶标核酸序列可指代位于HIV的编码及/或非编码区域及/或长末端重复序列内的序列。该位于HIV的长末端重复序列内的序列可包括位于U3区、R区、或U5区内的序列，这些区域不包括该截短的HIV LTR启动子的任何序列。该分离的核酸序列亦可编码核定位信号及/或反式激活小RNA(tracrRNA)。该tracrRNA可与一编码向导RNA的序列融合。该分离的核酸序列亦可包括该HIV LTR启动子的增强子区域。

[0015] 具体实施例中，该分离的核酸序列可被可操作地链接至表达载体。该表达载体可指代慢病毒载体、腺病毒载体、和腺病毒群载体。

[0016] 本文提供一种包括编码CRISPR相关核酸内切酶的序列的药物组合物，该酶可操作地链接至截短的HIV LTR启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和TAR区域。该药物组合物也可包括药学可接受的载剂，包括但不限于，基于脂质的胶体或基于聚合物的胶体。该胶体可以是脂质体、水凝胶、微粒、纳米颗粒、或嵌段聚合物胶束。具体实施例中，该CRISPR相关核酸内切酶是Cas9。该CRISPR相关核酸内切酶可被优化用于人细胞内表达。

[0017] 具体实施例中，该药物组合物可配制为用于外用及/或包含在避孕套中。

[0018] 具体实施例中，该序列可额外编码一种或多种与HIV中靶标核酸序列互补的向导。该HIV中靶标核酸序列可指代位于HIV的编码及/或非编码区域内及/或长末端重复序列内的序列。该位于HIV的长末端重复序列内的序列可包括位于U3区、R区、或U5区内的序列，这些区域不包括该截短的HIV LTR启动子的任何序列。该序列可编码核定位信号。该药物组合物可额外包括编码tracrRNA的序列，且该tracrRNA可与编码向导RNA的序列融合。该序列亦可编码该HIV LTR启动子的增强子区域。

[0019] 具体实施例中，由该药物组合物提供的序列可被可操作地链接至表达载体。该表达载体可以是慢病毒载体、腺病毒载体、和腺病毒群载体。

[0020] 本文提供一种治疗具有HIV感染的受试者的方法。该方法方法包括向该受试者给药治疗有效量的药物组合物，该药物组合物包括编码CRISPR相关核酸内切酶的序列，该酶可操作地链接至截短的HIV LTR启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和TAR区域。待治疗的HIV感染可以是潜伏性感染。该方法可进一步包括鉴别具有HIV感染的受试者。

[0021] 具体实施例中，该CRISPR相关核酸内切酶是Cas9。该CRISPR相关核酸内切酶可被优化用于人细胞内表达。

[0022] 具体实施例中，该序列可额外编码一种或多种与HIV中靶标核酸序列互补的向导RNA。一些例子中，该序列可编码该HIV LTR启动子的增强子区域。

[0023] 具体实施例中，可给药抗逆转录病毒剂。该抗逆转录病毒剂可包括，但不限于，非核苷酸逆转录酶抑制剂、蛋白酶抑制剂、及进入抑制剂。该抗逆转录病毒剂可包括高活性的抗逆转录病毒治疗。该药物组合物可外用给药或肠胃外给药。

[0024] 具体实施例中，该药物组合物可被可操作地链接至表达载体。该表达载体可以是慢病毒载体、腺病毒载体、和腺病毒群载体。

[0025] 本文中提供一种降低处于HIV感染风险下的受试者感染HIV的风险的方法。该方法

可包括向该受试者给药治疗有效量的包含编码CRISPR相关核酸内切酶的序列的药物组合物，该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和TAR区域。具體实施例中，该受试者是性活跃者、医护人员及/或现场急救员。

[0026] 具体实施例中，该CRISPR相关核酸内切酶可以是Cas9。该CRISPR相关核酸内切酶可以被优化用于人细胞内表达。一些例子中，该药物组合物可以被可操作地链接至表达载体。该表达载体可以是，而非限于，慢病毒载体、腺病毒载体、和腺病毒群载体。具體实施例中，该序列也可编码该HIV LTR启动子的增强子区域。

[0027] 本文中提供一种降低HIV感染从感染HIV的妊娠期或哺乳期母亲向其子女传播的风险的方法。该方法包括向该受试者给药治疗有效量的包含编码CRISPR相关核酸内切酶的序列的药物组合物，该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和TAR区域。具體实施例中，该药物组合物在产前、围产期和产后的至少一个阶段给药。

[0028] 具体实施例中，可给药抗逆转录病毒剂。该抗逆转录病毒及可以是，而非限于，非核苷酸逆转录酶抑制剂、蛋白酶抑制剂、及进入抑制剂。该抗逆转录病毒剂可以是高活性的抗逆转录病毒治疗。具體实施例中，可将治疗有效量的该组合物给药至该子女。具體实施例中，该序列也可编码该HIV LTR启动子的增强子区域。

[0029] 本文中提供一种给药药物组合物以防止未感染的受试者感染HIV的方法。该方法可包括向该未感染的受试者给药治疗有效量的包含编码CRISPR相关核酸内切酶的序列的药物组合物，该酶可操作地链接至截短的HIV长末端重复序列(LTR)启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和HIV LTR启动子的TAR区域。

[0030] 本文中提供一种包括测得量的组合物的试剂盒，该组合物包括分离的核酸序列，该分离的核酸序列包括编码CRISPR相关核酸内切酶的序列，该酶可操作地链接至截短的HIV LTR启动子，而该截短的启动子至少含有HIV LTR启动子的核心区域和TAR区域；或编码该分离核酸的的载体；以及包装材料、包括使用说明书的包装插页、无菌流体、注射器、及无菌容器的至少一项。

[0031] 如本发明中关于所揭露的物质组合物和方法所展望，一方面，本发明的具體实施例包含本文中揭露的组分及/或步骤。另一方面，本发明的具體实施例主要由本文中揭露的组分及/或步骤组成。再一方面，本发明的具體实施例由本文中揭露的组分及/或步骤组成。

附图说明

[0032] 图1A是全长度HIV-1LTR(LTR(-454/+66))以及所创建的LTR截短变体LTR-120/+66、LTR-80/+66和LTR-38/+66的示意图。每一变体中含有的LTR元件从图中明显可见。

[0033] 图1B是图1A的全长度HIV-1LTR和变体的PCR扩增的LTR序列的琼脂糖凝胶电泳图像。道次1：全长度HIV-1LTR(pLTR(-454/+66))。道次2：pLTR(-120/+66)。道次3：pLTR(-80/+66)。道次4：pLTR(-38/+66)。

[0034] 图2是根据本发明的Cas9启动子置换过程的图解。使用pX260-U6-DR-BB-DR-Cbh-NLS-hSpCas9-NLS-Hl-短tracr-PGK-puro质粒(Addgene#42229)(标示为“CBh-Cas9”)作为Cas9基因源/模板。以图中所标注的酶进行限制酶消解而移除参考质粒中的原始CBh启动

子,并替换为不同的HIV-1LTR启动子变体(共同标示为“LTR-Cas9”)。

[0035] 图3A是在全长度HIV-1LTR(pLTR(-454/+66)-FLAG-Cas9)(10、50和250ng)的对照下,在使用不同量的表达FLAG标记的Cas9的质粒且使用或不使用Tat表达质粒(pCMV-Tat86,250ng)进行共转染的U87MG细胞中,Cas9、Tat和α-微管蛋白表达的蛋白质印迹(Western blot)图。道次1:pLTR(-454/+66)-Cas9 250ng、pCMV 1000ng。道次2:pLTR(-454/+66)-Cas9 50ng、pCMV 1200ng。道次3:pLTR(-454/+66)-Cas9 10ng、pCMV 1240ng。道次4:pLTR(-454/+66)-Cas9 250ng、pCMV 750ng、pCMV-Tat86 250ng。道次5:pLTR(-454/+66)-Cas9 50ng、pCMV 950ng、pCMV-Tat86 250ng。道次6:pLTR(-454/+66)-Cas9 10ng、pCMV 990ng、pCMV-Tat86250ng。

[0036] 图3B包含对应于Cas9且归一化至图3A蛋白质印迹图中α-微管蛋白表达的条带强度的图。上图显示Cas9水平归一化至有或无Tat的α-微管蛋白水平的蛋白质印迹图像定量。下图显示+Tat/无Tat比的蛋白质印迹图像定量。

[0037] 图4A是在HIV-1截短的LTR变体pLTR(-120/+66)-FLAG-Cas9或HIV-1LTR变体pLTR(-80/+66)-FLAG-Cas9的对照下,在使用不同量的表达FLAG标记的Cas9的质粒(5ng或50ng)且使用或不使用Tat表达质粒(pCMV-Tat86,250ng)进行转染的U87MG细胞中,Cas9、Tat和α-微管蛋白表达的蛋白质印迹图。道次1:pLTR(-120/+66)-Cas9 5ng、pCMV 1245ng。道次2:pLTR(-120/+66)-Cas9 5ng、pCMV 1245ng、+rTat蛋白2.5μg/ml。道次3:pLTR(-120/+66)-Cas9 5ng、pCMV 995ng、pCMV-Tat86 250ng。道次4:pLTR(-120/+66)-Cas9 50ng、pCMV1200ng。道次5:pLTR(-120/+66)-Cas9 50ng、pCMV 1200ng、+rTat蛋白2.5μg/ml。道次6:pLTR(-120/+66)-Cas9 50ng、pCMV 950ng、pCMV-Tat86 250ng。道次7:pLTR(-80/+66)-Cas9 5ng、pCMV 1245ng。道次8:pLTR(-80/+66)-Cas9 5ng、pCMV 1245ng、+rTat蛋白2.5μg/ml。道次9:pLTR(-80/+66)-Cas9 5ng、pCMV 995ng、pCMV-Tat86 250ng。道次10:pLTR(-80/+66)-Cas9 50ng、pCMV 1200ng。道次1 1:pLTR(-80/+66)-Cas9 50ng、pCMV 1200ng、+rTat蛋白2.5μg/ml。道次12:pLTR(-80/+66)-Cas9 50ng、pCMV 950ng、pCMV-Tat86 250ng。

[0038] 图4B包含对应于Cas9且归一化至图4A蛋白质印迹图中α-微管蛋白表达的条带强度的图。上图显示Cas9水平归一化至无Tat、有rTAT、或有经转染Tat的α-微管蛋白水平的蛋白质印迹图像定量。下图显示+Tat(经转染)/无Tat比的蛋白质印迹图像定量。

[0039] 图5A至5E说明,在gRNA的存在下,通过HIV-1LTR启动子进行的Cas9表达受到Tat的刺激,导致病毒启动子的裂解。图5A:全长度HIV-1LTR、增强子区域和核心区域内的多种调节性基序、以及部分Gag基因的图式表述。描述了被创建用于Cas9表达的LTR缺失突变体的程度。显示了gRNA靶标序列的位置及其彼此间的距离。图5B:使用含有全长度LTR(-454/+66)或其突变体(-120/+66或-80/+66)的pX260-LTR-Cas9连同表达Tat的质粒(pCMV-Tat)一起对TZMbl细胞的共转染,增加了Tat生产的水平,如通过蛋白质印迹法所测(上图)。显示了看家α-微管蛋白的表达(中间图)和Tat的表达(下图)。图5C:使用表达Tat的腺病毒以两个不同的感染复数(MOI)感染TZMbl细胞,之后通过$LTR_{-80/+66}$启动子进行慢病毒介导的Cas9表达以及通过U6启动子进行gRNA A/B的表达,导致TZMbl细胞内所整合的HIV-1LTR启动子DNA序列的裂解以及205 bp DNA片段的出现(如通过PCR和DNA凝胶分析所测)。图5D:SDS-PAGE,例示性说明TZMbl细胞内表达的Cas9、β-微管蛋白和Tat蛋白的水平,如图5C图中所示。图5E:荧光素酶试验,例示性说明在多种治疗后,TZMbl细胞内所整合的HIV-1LTR的转录活性,

如图5C中所示。

[0040] 图6A至6C显示，在Cas9的诱导下，HIV-1感染刺激了所整合的病毒DNA的裂解。以表达gRNA A/B的慢病毒(LV-gRNA A/B)或对照物(空LV)转导$LTR_{-80/+66}$-Cas9报告TZMbl细胞系，使用三种不同MOI的$HIV\text{-}1_{JRFL}$或$HIV\text{-}1_{SF162}$感染该细胞系48小时后，收获细胞，通过蛋白质印迹法测定蛋白质表达(图6A)，通过PCR/DNA基因分型检测病毒感染后在Cas9诱导下所证实的HIV-1LTR裂解的水平(图6B)，且通过荧光素酶报告基因试验评估所整合的HIV-1启动子的转录活性(图6C)。

[0041] 图7A至7C显示，在潜伏期，Tat对Cas9的刺激裂解了囊括HIV-1报告基因的T细胞内所整合的HIV-1DNA。以对照物(空LV)或LV-gRNA A/B转导含有$LTR_{-80/+66}$-Cas9基因的$CD4^{+}$ Jurkat T细胞2D10细胞，之后以pCMV或pCMV-Tat质粒转染。48小时后，通过蛋白质印迹法测定所描述的多种蛋白质的水平(图7A)。通过LTR特异性PCR测定用于评估所整合的HIV-1DNA状态的基因组DNA，并将切除效率测定为以百分比表示的截短的扩增子与全长度扩增子之比(图7B)。通过流式细胞术评估所整合的病毒启动子在裂解之后再次激活的水平，且显示了代表性散点图(图7C)。从该分析排除红色阳性的、碘化丙啶染色的、及死亡的细胞。

[0042] 图8A至8C显示，以潜伏期逆转药物治疗细胞诱发了Jurkat 2D10细胞内的Cas9表达和所整合的病毒DNA的裂解。使用对照物(空)或表达gRNAs A/B的慢病毒治疗表达$LTR_{-80/+66}$-Cas9的2D10细胞，24小时后，使用PMA(P)、TSA(T)、或两者(P/T)治疗这些细胞16小时。通过蛋白质印迹法进行蛋白质研究，测定Cas9-Flag、α-微管蛋白和GFP(表示所整合的HIV-1基因组)的表达(图8A)。通过PCR评估用于检测通过Cas9和gRNA A/B进行的所整合LTR DNA的切除水平的基因组DNA，并如图7A至7C图例中所描述测定切除效率(图8B)。显示了通过流式细胞分析进行的GFP报告基因试验以及代表性散点图(图8C)。

[0043] 图9是通过CRISPR/Cas9进行的HIV-1负反馈调节的图式表述。在再激活的早期，病毒基因组的基础转录允许Tat蛋白①的生产。当将TAT与病毒转录本②的移动序列(budge sequence)且吸收(recruitment)若干与TAR和位于RNA poly II很靠近转录起始位点处的其它转录因子的环圈联合的细胞型蛋白时，病毒RNA的转录一开始即被高度刺激，且更重要的是，该高度刺激一直延续③。在病毒激活时的基础产物也刺激最小病毒启动子Itr，驱动该Cas9基因③。当新合成的Cas9与多种HIV-1特异性gRNA联合时，前者裂解该病毒基因组，永久地灭活该LTR且终止HIV-1基因表达和复制。不存在Tat时，Itr-Cas9变为静默。Cas9的表达仅在Tat存在时可持续进行。

[0044] 图10显示，参考HIV-1NL4-3基因组中，用于TZMbl基因组DNA(蓝色高亮)的PCR中的LTR(绿色高亮，PAM为红色)和LTR特异性引物内gRNA A/B靶标的位置和核苷酸序列。LTR特异性PCR产物(全长度和截短的)及预计被编辑的片段(SEQ ID NO:6至SEQ ID NO:10)的序列和大小。

[0045] 图11显示代表性琼脂糖凝胶电泳结果，其分析在使用来自图7A至7C和8A至8C的Jurkat 2D10报告细胞系的实验中用于定量Cas9/gRNA介导的LTR切除效率的LTR特异性PCR反应。

[0046] 图12显示，参考HIV-1NL4-3基因组中，用来分析通过PCR在Jurkat 2D10细胞内进行切除(蓝色高亮)的LTR gRNA A/B靶标(绿色高亮，PAM为红色)和LTR特异性引物的位置和核苷酸组成。显示了扩增子(全长度和截短的LTR DNA)和预计切除的DNA片段的核苷酸序列

和大小(SEQ ID NO:11至SEQ ID NO:21)。

[0047] 定义

[0048] 除非定义排除,本文中使用的全部科技术语具有与本发明所属领域技术人员通常理解者相同的意义。尽管可使用与本文中揭示者相似或等效的任何方法和材料实践本发明,但优选的材料和方法揭示于本文中。将使用下述术语来描述和主张本发明。

[0049] 还应理解,本文中使用的术语仅用于揭示特定具體实施例目的,而非予以限制。

[0050] 本文中揭露的所有基因、基因名称、和基因产品旨在与来自可使用本文所揭露的组合物和方法的任何物种的同源染色体相对应。应理解,除非在上下文中明确表明,当来自特定物种的基因或基因产品被揭露时,这一揭露仅用于示例而不应视为限制。因此,举例而言,对于本文中揭露的基因或基因产品旨在涵盖来自其它物种的同源及/或直向同源基因和基因产品。

[0051] 本文中的文字“一”用来指代一个或超过一个(即,至少一个)该文字的语法宾语。举例来说,“一元件”意为一个元件或超过一个元件。因此,举例而言,“一细胞”的表述包括相同类型的多个细胞。此外,就术语“包括”(including或includes)、“具有”(having或has)、“具备”(with)或其变体在详细说明书及/或权利要求书中的使用而言,这些术语应视为包容性的,其表述方式类似于术语“包含”。

[0052] 本文中,关于项目、组合物、仪器、方法、过程、系统等的所定义或描述的元件,术语“包含”(comprising、comprise、或comprised)及其变体意为包容性的或开放性的,其准许额外元件的存在,从而表明所定义或描述的项目、组合物、仪器、方法、过程、系统等包括那些具体化的元件或,视情况而定,其等效物,且包括可被包括且仍落入所定义的项目、组合物、仪器、方法、过程、系统等的范畴/定义内的其它元件。

[0053] 本文中,当指代可测量的值如量、时距等时,“约”意为涵盖所指定值+/-20%、+/-10%、+/-5%、+/-1%、或+/-0.1%的变更,而这些变更适合于实施所揭露的方法。或者,特别是关于生物系统或过程,该术语可意为处于值的5倍数量级内以及2倍数量级内。若在申请书和权利要求书中描述了特定的值,除非明确排除,否则应假定术语“约”意为处于该特定值的可接受的误差范围内。

[0054] 本文中,“有效量”意为提供治疗性或预防性益处的量。

[0055] “编码”是指多核苷酸如基因、cDNA或mRNA中适当固有的特异性序列,其作为模板而用于生物学过程中其它聚合物和大分子的合成,具有所定义的核苷酸序列(即,rRNA、tRNA和mRNA)或所定义的氨基酸序列以及自其所得的生物学性质。因此,若对应于一基因的mRNA的转录和转译在细胞或其它生物系统中生产蛋白质,则该基因编码该蛋白质。其核苷酸序列与该mRNA序列完全相同且一般提供于序列表中的编码链,以及用作基因或cDNA转录模板的非编码链,两者均可指代为编码该蛋白质或该基因或cDNA的其它产物。

[0056] 本文中,术语“表达”定义为由其启动子驱动的特定核苷酸序列的转录及/或转译。

[0057] “表达载体”是指包含重组多核苷酸的载体,该重组多核苷酸包含可操作地链接至待表达核苷酸序列的表达控制序列。表达载体包含足够的用于表达的顺式作用元件;用于表达的其它元件可由宿主细胞或体内外表达系统供给。表达载体包括该领域中已知的全部,如合并重组多核苷酸的粘粒、质粒(如,裸质粒或脂质体中含有的质粒)和病毒(如,慢病毒、逆转录病毒、腺病毒、及腺病毒群)。

[0058] “分离的”意为自自然状态改变或移除。例如，自然存在于活体动物体内的核酸或肽不是“分离的”，但与其自然状态的共存材料部分或完全分隔开的相同的核酸或肽是“分离的”。分离的核酸或蛋白质可以实质上纯化的形式存在，或可存在于非原生环境如，举例而言，宿主细胞内。

[0059] “分离的核酸”是指核酸节段或片段，其已经与自然发生状态中位于其侧翼的序列分隔开来，即，已经从通常邻近该DNA片段的序列移除的DNA片段，该序列即该片段自然发生的基因组内邻近该片段的序列。该术语也应用于已经从自然伴随该核酸的其它组分实质上纯化的核酸，该其它组分即细胞中自然伴随该核酸的RNA或DNA或蛋白质。该术语因此包括，例如，重组DNA，其被并入载体内、被并入自主复制的质粒或病毒内、或被并入原核生物或真核生物的基因组DNA内，或其作为不依赖于其它序列的独立分子(即，作为通过PCR或限制酶消解生产的cDNA或基因组或cDNA片段)而存在。它还包括：作为编码额外核苷酸序列的杂交基因的一部分的重组DNA、互补DNA(cDNA)、天然及/或修饰单体或链接的线性或环状寡聚物或聚合物，该单体或链接包括脱氧核糖核苷酸、核糖核苷酸、其经取代的形式和α-端基异构形式、肽核酸(PNA)、锁核酸(LNA)、硫代磷酸酯、甲基磷酸酯等。

[0060] 当用于多核苷酸序列的上下文中时，术语“变体”可涵盖与野生型基因相关的多核苷酸序列。这定义亦可包括，例如，“等位基因”变体、“剪接”变体、“物种”变体、或“多态性”变体。剪接变体可具有与参考分子的显着一致性，但由于在mRNA加工过程中的外显子的交替剪接，剪接变体通常会具有更多或更少数量的多核苷酸。相应的多肽可拥有额外的功能性域或缺少某些域。物种变体是从一个物种变为另一物种的多核苷酸序列。本发明中特别有用的是野生型基因产品的变体。变体可导致该核酸序列中的至少一个突变，且可造成被改变的mRNA或造成其结构或功能可能被改变或未被改变的多肽。任何给定的天然或重组基因可不具有等位基因形式、一种等位基因形式或多种等位基因形式。引发变体的常见突变性改变通常被归因于核苷酸的天然缺失、加入或替换。在给测序列中，这些类型的改变的每一种可一次或多次地单独出现或与其它改变组合出现。

[0061] 本文中，术语“核酸序列”与“多核苷酸”通篇可互换使用，且包括互补DNA(cDNA)、天然及/或修饰单体或链接的线性或环状寡聚物或聚合物，该单体或链接包括脱氧核糖核苷酸、核糖核苷酸、其经取代的或α-端基异构形式、肽核酸(PNA)、锁核酸(LNA)、硫代磷酸酯、甲基磷酸酯等。多核苷酸包括，但不限于，通过该领域中可用的任何手段获得的所有核酸序列，该手段包括而不限于，重组手段，即，使用常规克隆技术和PCR™等从重组库(library)或细胞基因组克隆核酸序列，以及合成手段。

[0062] 核酸序列可以是“嵌合的”，换言之，由不同的区域构成。在本发明的上下文中，“嵌合的”化合物可以是寡核苷酸，其含有两个或更多个化学区域，例如，DNA区域、RNA区域、PNA区域等。每一化学区域由至少一种单体单元即核苷酸所组成。这些序列典型包含至少一个区域，其中，为了展现一种或多种所希望的性质，该序列被修饰。

[0063] 术语“靶标核酸”是指核酸(通常源自生物样品)，其是寡核苷酸被设计以特异性杂家的目标。靶标核酸的存在与否待检测，或该靶标核酸的量待定。靶标核酸的序列与被导向该靶标的相应寡核苷酸的核酸序列互补。术语“靶标核酸”可指代该寡核苷酸被导向的较大核酸的特异性子序列，或指代整体序列(如，基因或mRNA)。用法的不同从上下文中明显可见。

[0064] 本发明的上下文中，使用下述用于常见核酸碱基的缩写，“A”指腺苷，“C”指胞苷，“G”指鸟苷，“T”指胸苷，而“U”指尿苷。

[0065] 除非具体排除，“编码氨基酸序列的核苷酸序列”包括作为彼此退化版本且编码相同氨基酸序列的所有核苷酸序列。短语“编码蛋白质或RNA的核苷酸序列”也可包括内含子，其包含程度为该编码该蛋白质的核苷酸序列可以是含有内含子的一些版本。

[0066] 本文中，“慢病毒”是指逆转录病毒科的一个属。就能够感染非分化细胞而言，慢病毒在逆转录病毒中是独一无二的；它们能将大量的遗传信息输送至宿主细胞的DNA内，因此它们是基因输送载体的最有效方法之一。HIV、SIV、和FIV是慢病毒的全部实例。源自慢病毒的载体提供了实现高水平的体内基因转移的手段。免疫原性组合物的“肠胃外”给药包括，如，皮下(s.c)注射、静脉(i.v.)注射、肌肉(i.m.)注射、或胸骨内注射，或输液技术。

[0067] 本文中，输液“患者”或“个体”或“受试者”可互换使用，且指代待治疗的哺乳动物受试者，优选人类患者。一些例子中，本发明的方法在实验动物、兽医应用、及动物疾病模型的发展中找到了用途，包括但不限于，包括小鼠、大鼠和仓鼠的啮齿动物以及灵长类动物。

[0068] 术语“多核苷酸”是核苷酸的链，亦称“核酸”。本文中，多核苷酸包括但不限于，通过该领域可用的任何手段获得的所有核酸序列，且包括天然核酸和合成核酸两类。

[0069] 术语“肽”、“多肽”和“蛋白质”可互换使用，且指代由通过肽键共价键结的氨基酸残基所构成的化合物。蛋白质或肽必须含有至少两个氨基酸，且蛋白质或肽序列中可包含的最大氨基酸数并无限制。多肽包括由通过肽键彼此结合的两个或更多个氨基酸所组成的任何肽或蛋白质。本文中，该术语指代该领域中一般指代为肽、寡肽和寡聚物的短链以及该领域中通常指代为蛋白质的长链，该短链和长链均存在多种类型。“多肽”包括，例如，生物活性片段、实质上同源的多肽、寡肽、同源二聚体、多肽的变体、经修饰的多肽、衍生物、类似物、融合蛋白等。多肽包括天然多肽、重组多肽、合成多肽、或其组合。

[0070] 术语“启动子”意为一种DNA序列，其被细胞合成器械(machinery)或引入的合成器械识别且为启动多核苷酸序列的特异性转录所需。“最小”启动子或“截短的”启动子或启动子的“功能性片段”包括启动子用于转录激活的所有必需元件，例如，可操作地链接的或受控于最小启动子的核酸序列。具體实施例中，截短的HIV长末端重复序列(LTR)启动子至少包含HIV LTR启动子的核心区域、反式激活响应元件(TAR)或其组合。

[0071] 术语“经转染的”或“经转化的”或“经转导的”意为将外源性核酸转移至或引入宿主细胞内的过程。“经转染的”或“经转化的”或“经转导的”细胞是已经使用外源性核酸转染、转化或转化的细胞。经转染/转化/转化的细胞包括初代受试者细胞及其后代。

[0072] 作为本文中使用的术语，“治疗”疾病意为降低该受试者所经历的疾病或病变的至少一种症状的频率或严重性。

[0073] “载体”是物质的组合物，其包含分离的核酸且可用来将该分离的核酸输送至细胞内部。载体的实例包括但不限于，线性多核苷酸、与离子化合物或两亲化合物联合的多核苷酸、质粒、和病毒。因此，术语“载体”包括自主复制的质粒或病毒。该术语也视为包括促进将核酸转移至细胞内的非质粒化合物和非病毒化合物如，举例而言，聚赖氨酸化合物、脂质体等。病毒载体的实例包括，但不限于，腺病毒载体、腺病毒群载体、逆转录病毒载体等。

[0074] 范围：贯穿本公开，本发明的多个方面可以范围格式呈现。应理解，范围格式的说明仅仅为了表述的方便和简洁，而不应视为本发明该范畴的刚性限制。据此，对范围的说明

应视为具有具体揭露的所有可能的子范围以及该范围内的个体数值。例如，对范围如从1至6的说明应视为具有具体揭露的子范围，如从1至3、从1至4、从1至5、从2至4、从2至6、从3至6等，以及该范围内的个体数值，例如，1、2、2.7、3、4、5、5.3、和6。这一规则可无视该范围的广度而使用。

[0075] 若通过Swiss Prot.或GENBANK登录号指代氨基酸序列，则该序列通过引用而并入本文。与该登录号相关联的信息，如信号肽、胞外域、跨膜域、启动子序列和转译开始的鉴定，也通过引用而以其整体并入本文。

[0076] 术语“序列一致性百分比”指代任何给定的查询序列与目标序列之间的一致性的程度。

[0077] 术语“外源性”表明，该核酸或多肽是重组核酸构造的一部分或由该重组核酸编码，或并非出于其自然环境中。例如，外源性核酸可以是来自被引入一种物种的另一物种的序列，即异源性核酸。典型地，这一外源性核酸经由重组核酸构造而被引入其它物种内。外源性核酸可以是一有机体原生的且已经被再次引入该有机体细胞内的序列。包括原生序列的外源性核酸通常可通过链接至该外源性核酸的非天然序列的存在而与天然出现的序列区分开来，该非天然序列例如重组核酸构造中位于原生序列侧翼的非原生调节性序列。此外，稳定转化的外源性核酸典型被整合在除发现原生序列处之外的位置。

[0078] 术语“药学可接受的”(或“药理学可接受的”)是指，当酌情给药至动物或人时，不产生副作用、过敏反应或其它不良反应的分子整体和组合物。本文中使用的术语“药学可接受的载剂”包括任何和所有溶剂、分散截止、涂层、抗菌剂、等渗剂和吸收延迟剂、缓冲剂、赋形剂、粘合剂、润滑剂、凝胶、表面活性剂等，它们可用作药学可接受的物质的介质。

[0079] 本文中，术语“试剂盒”是指用于输送材料的任何输送系统。术语“试剂盒”所包括的是用于研究和临床应用的两类试剂盒。在反应试验情况下，这些输送系统包括允许反应试剂(如，置于适宜容器内的寡核苷酸、酶等)及/或支持材料(如，缓冲剂、用于实施该试验的书面说明书等)存储以及从一个地点向另一个地点运输或输送的系统。例如，试剂盒包括一个或多个含有相关反应试剂及/或支持材料的密封件(如，盒子)。本文中，术语“分散式试剂盒”是指包含两个或多个独立容器的输送系统，每一容器含有总试剂盒组分的小部分。该容器可被一起或单独输送至预定接收者。例如，第一容器可含有试验中使用的酶，而第二容器含有寡核苷酸或脂质体。术语“分散式试剂盒”倾向于涵盖含有分析质特异性试剂(ASR)但并不限于此的试剂盒，该试剂是联邦食品、药品及化妆品法案第520(e)条下规定的试剂。事实上，任何包含两个或更多个独立容器且每一容器含有总试剂盒组分的一部分的输送系统均包括在术语“分散式试剂盒”中。相反，“组合式试剂盒”是指在一个容器内含有反应试验所有组分(如，一个盒子内容纳每一所希望的组分)的输送系统。术语“试剂盒”包括分散式试剂盒和组合式试剂盒两种。

具体实施方式

[0080] 以HIV-1感染后不久，该病毒基因组变为被整合入宿主染色体并迅速在$CD4^+$T细胞内表达。HIV-1复制导致$CD4^+$T细胞的剧烈损耗。通常，在感染的急性期后，该病毒进入被称为潜伏期的新阶段，在该阶段，被整合的前病毒DNA基序被表达，且病毒复制在非常低的水平进行。在这些情况下，由病毒持续复制所造成的被唤醒的免疫系统向AIDS发展，且发展出

宽范围的机会性感染，而该感染若不经治疗，则最终导致在三年内死亡。在分子水平上，急慢性状态的病毒基因组的表达及其复制处于病毒启动子的控制下，而该病毒启动子跨越5’长末端区域(LTR)的450个核苷酸。一系列识别5’-LTR的U3区域内的DNA序列的细胞转录因子与HIV-1即刻早期转录活化因子Tat之间出现协同性，该Tat与位于该病毒转录本的前导区内的TAR RNA序列相互作用。这些相互作用为来自所整合的病毒DNA副本的强健启动和有效延伸所需。尽管目前抗逆转录病毒药物已经有效地阻抑了病毒感染循环，它们仍必须含有在转录水平上抑制病毒基因表达的任何组分，支持所整合的病毒副本可继续在进行活性抗逆转录病毒治疗(ART)的HIV-1阳性患者体内表达病毒基因组的概念，虽然表达的水平非常低。事实上，病毒基因的表达在ART中断时即剧烈提升，且允许病毒早期调节性蛋白如Tat的生产以统筹该病毒基因组的生产性复制。

[0081] 据此，本发明的具體实施例指向用于在再激活的早期有条件地激活CRISPR/Cas的组合物。这些组合物在生产性病毒复制之前，通过移除该病毒基因的跨越该病毒启动子及/或病毒编码序列的片段而完全且永久地废除病毒复制。具体实施例中，组合物包含编码成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶(CRISPR Cas)的核酸序列，该酶可操作地链接至截短的功能性病毒启动子，借此，该截短的病毒启动子处于即刻早期转录活化因子的控制下，从而在病毒复制的早期有条件地激活CRISPR/Cas。该分离的核酸进一步包含至少一种与病毒内的靶标核酸序列互补的向导RNA。CRISPR/Cas切除病毒基因组的一段，例如，跨越病毒启动子及/或病毒编码序列的一段。这些具體实施例中，调整该组合物以适于切除任何病毒。某些具體实施例中，该病毒是逆转录病毒。

[0082] 可以使用RNA引导的成簇的规律间隔短回文重复序列(CRISPR)相关核酸内切酶如Cas9，将整合入被感染的宿主细胞基因组的病毒基因组如HIV从这些HIV感染的细胞中消除。使用CRISPR/Cas9酶和向导RNA的成功的治疗性基因编辑需要在靶标细胞内有效且特异性的输送和表达Cas9酶和向导RNA。当组织或细胞集群中感受细胞的频率较低时，如在进行高活性的抗逆转录病毒(HAART)治疗的患者体内的HIV感染细胞内，这难以进行。

[0083] 根据本发明，将CRISPR相关核酸内切酶如Cas9置于截短的、响应Tat的HIV LTR启动子的控制下。因此在含有该Tat蛋白的细胞内激活该核酸内切酶的表达。如本文中所表明，当将该核酸内切酶的表达置于该截短的、响应Tat的HIV LTR启动子的控制下时，外源性提供的(如，通过转染)和内源性生产的(如，通过潜伏病毒的再激活)Tat可激活CRISPR相关核酸内切酶(如，Cas9)在细胞系内的表达。在具體实施例部分中进一步详细呈现的研究中，该组合物考虑到了在病毒再激活的早期通过HIV-1转录活化因子Tat有条件地激活CRISPR/Cas9。通过移除整体病毒基因组或该病毒基因的跨越病毒启动子及/或病毒编码序列的一段，这一策略在生产性病毒复制之前完全且永久地废除了病毒复制。

[0084] 图1A显示了HIV LTR的示意图。其长度大约为640bp。HIV-1LTR分为U3区、R区、和U5区。通过一系列跨越该病毒基因组5’端的长末端区域的顺式作用调节性基序来控制HIV-1基因组的转录。将转录起始位点设为+1，则该病毒启动子的U3区占据了-1至-454核苷酸且具有三个子区域：调节性区域、增强子区域和核心区域。该增强子含有NF-κB结合位点(-127至-80)。该核心域包含该富GC区和TATA盒(-80至+1)。该LTR的R区(+1至+98)包含TAR，TAR是所表达的RNA形成茎环结构所用的且提供用于病毒反式激活因子的结合位点的区域(Krebs et al.,Lentiviral LTR-directed expression,sequence variation,disease

pathogenesis.Los Alamos National Laboratory HIVsequence:Compendium,pp.29-70.2002)。

[0085] LTR含有基因表达所需的所有信号,且牵涉入前病毒至宿主细胞基因组内的整合中。例如,发现核心启动子、增强子、和调节性区域处于U3内,而发现TAR处于R内,如图1A中所示。TAR,作为Tat蛋白或细胞蛋白的结合位点,大概由HIV-1的病毒mRNA的前45个核苷酸组成且形成发夹茎环结构。在HIV-1中,U5区包括若干子区域,例如,包括牵涉如二聚体化和基因组封装的Poly A、PBS或引物结合位点、Psi或封装信号、以及DIS或二聚体起始位点。

[0086] 根据本发明,提供包含分离的核酸的组合物,该分离的核酸编码CRISPR相关核酸内切酶,该酶可操作地链接至截短的HIV LTR启动子,而该截短的启动子至少含有HIV LTR启动子的核心区域和TAR(反式激活响应元件)区域。截短的HIV LTR启动子是指操作性功能性启动子,其含有的内容少于全长度HIV LTR启动子。该截短的启动子优选含有核心区域和TAR区域而实质上不含或完全不含调节性区域及/或增强子区域。另一具體实施例中,该截短的HIV LTR启动子含有核心区域、TAR区域、以及全部或实质上全部的增强子区域,但不含有调节性区域的任何部分。该截短的HIV LTR启动子可响应Tat蛋白。换言之,Tat可激活可操作地链接至截短的HIV LTR启动子的CRISPR相关核酸内切酶如Cas9的表达。所揭露的组合物可用来灭活哺乳动物细胞内的HIV、治疗具有HIV感染的受试者、降低处于HIV感染风险下的受试者感染HIV的风险、及/或降低HIV从感染HIV的母亲向其子女传播的风险。本文中揭露的治疗性方法可与其它抗逆转录病毒治疗如HAART合用。该组合物可被包括而作为用于诊断、研究、及/或治疗性应用的试剂盒的一部分。

[0087] 抗逆转录病毒治疗既不阻抑低水平的病毒基因组表达,如早前所述,也未有效地以潜伏性感染的细胞如休眠记忆T细胞、单核细胞、巨噬细胞、小胶质细胞、星形胶质细胞、和肠相关淋巴样细胞为靶标。然而,本文揭露的方法和组合物一般可用于治疗任何感染阶段的感染HIV的受试者,或用于处于HIV感染风险下的未感染的受试者。特别地,所揭露的方法和组合物可用于处于感染潜伏期的感染HIV的受试者。此外,当将向导RNA与可操作地链接至截短的、响应Tat的HIV LTR启动子的CRISPR相关核酸内切酶相关联时,如本文所揭露,可从宿主细胞切除和消除该HIV基因组。

[0088] 使用该组合物可实现若干优势,其中,该组合物含有编码CRISPR相关核酸内切酶的序列,该酶可操作地链接至截短的HIV LTR启动子,而该截短的启动子含有HIV LTR启动子的核心区域和TAR区域。该持续表达所造成的潜在的毒性效应风险,可通过将CRISPR相关核酸内切酶的表达限制在具有HIV基因表达及/或复制的细胞内而得以缓解及/或消除。例如,根据本发明,因为该核酸内切酶的低及/或间歇性表达,可以缓和缘于CRISPR相关核酸内切酶的免疫原性的诱导毒性的潜力,同时消除或造成被感染个体体内HIV基因组的自我毁灭。此外,因为CRISPR相关核酸内切酶的持续表达被最小化,本发明可提供用于处于风险下的个体的预防性策略。因此,由截短的、响应Tat的HIV LTR启动子所驱动的CRISPR相关核酸内切酶可用来提供对HIV感染者的安全治疗,以及用来对可能处于风险下的未感染个体注射疫苗。

[0089] 一些具體实施例中,该启动子包含一个或多个突变、缺失、插入、变体、衍生物或其组合。该启动子亦可为嵌合的,包含一个化合物或嵌合化合物。

[0090] 如本文所述,因为较小尺寸的核酸可以更容易地封装在适用于基因治疗的输送机

制(如,逆转录病毒)内,将CRISPR相关核酸内切酶置于截短的HIV LTR启动子的控制下也具备优势。例如,由于包括该调节性区域的启动子构造的尺寸及/或对CRISPR相关核酸内切酶转录的可变效果,该启动子构造可能较不适用于基因治疗。再者,仅包括HIV LTR启动子的核心区域的TATA盒加上全TAR区域的组合物不能充分地表达Cas9(数据未显示)。包括该HIV-1LTR启动子的整个核心区域、TAR区域、以及任选的增强子(见图1A)的组合物能以剂量依赖方式驱动Tat诱导的Cas9表达。

[0091] 该截短的HIV-1LTR启动子可包含核酸,而该核酸包括HIV-1LTR启动子的位置-80至+66的核苷酸。具體实施例中,该截短的HIV-1LTR启动子可包含核酸,而该核酸包括该HIV-1LTR启动子的位置-120至+66。优选地,该截短的HIV-1LTR启动子并不含有来自该调节性区域的序列。

[0092] 如本文中揭露,通过使用pNL4-3HIV载体(NIH AIDS反应物程序#114)作为模板和下表中所示引物的PCR获得全长度的和截短的HIV-1LTR启动子序列:

[0093]

引物名称	序列	SEQ ID NO.
Kpn1-LTR(-454)-S	5'-**GGTACC**TGGAAGGGCTAATTTGG-3'	SEQ ID NO: 1
Kpn1-LTR(-120)-S	5'-**GGTACC**TCGAGCTTTCTACAAGG-3'	SEQ ID NO: 2
Xba1-LTR(-80)-S	5'-**TCTAGA**GGAGGTGTGGCCTGGGC-3'	SEQ ID NO: 3
Kpn1-LTR(-38)-S	5'-**GGTACC**AGATGCTACATATAAGC-3'	SEQ ID NO: 4
LTR(+66)-**Nco1**-AS	5'-**CCATGG**TAAGCAGTGGGTTCC-3'	SEQ ID NO: 5

[0094] 序列栏中加粗的核苷酸对应于各自引物名栏中加粗的限制酶。每一引物用来生成如图1A中所示的不同尺寸的HIV-1LTR启动子序列段。例如,LTR-454/+66包括整个U3区和该R区的一部分。相比之下,LTR-80/+66对应于U3的核心区域和R的TAR区域。最为对Tat的响应,LTR-38/+66核苷酸序列能充分驱动可检测水平的Cas9表达(数据未显示)。

[0095] 本发明的截短的HIV-1LTR启动子对应于含有U3的核心区域以及TAR的节段。该核心区域包括TATA盒和富GC区域,它可以是SP1的靶标。一些构型中,该截短的HIV-1LTR启动子可包括如图1A所示的位于位置-120至-80的增强子。

[0096] 截短的HIV-1LTR启动子可用来驱动CRISPR相关核酸内切酶如Cas9的表达。这些核酸内切酶揭示在2014年8月29日提交且在2015年3月5日公开的PCT国际申请PCT/US 2014/053441(WO 2015/031775)中,该申请的整体公开内容通过引用而并入本文。如上所述,HIV

基因组整合入被HIV感染的个体的宿主基因组中。随后，这一整合的序列被宿主复制。设置在潜伏期内，Tat即可由细胞生产。本发明的组合物消除及/或降低了宿主体内前病毒多核苷酸的存在。因为CRISPR相关核酸内切酶由根据本发明的响应Tat(Tat-responsive)的启动子驱动，在Tat存在(如，由被感染的细胞生产)的任何时间内，该核酸内切酶被生产出来并降解初生的多核苷酸。当病毒不活动时，没有核酸内切酶产生。因此，核酸内切酶的连续表达可能施加于该细胞及/或宿主的潜在毒性效应得以避免。

[0097] 此外，所生产的核酸内切酶的量与存在的Tat的量成正比，如下文中关于图3和图4所述。

[0098] 某些具體实施例中，分离的核酸序列与SEQ ID NO:1至SEQ ID NO:21的任一者的序列相似性为至少50％。

[0099] 某些具體实施例中，分离的核酸序列与SEQ ID NO:1至SEQ ID NO:21的任一者的序列相似性为至少70％。某些具體实施例中，分离的核酸序列与SEQ ID NO:1至SEQ ID NO:21的任一者的序列相似性为至少75％。

[0100] 某些具體实施例中，分离的核酸序列与SEQ ID NO:1至SEQ ID NO:17的任一者的序列相似性为至少85％，至与SEQ ID NO:1至SEQ ID NO:21的任一者的序列相似性为约95％、96％、97％、98％、或99％。

[0101] 某些具體实施例中，分离的核酸序列包含SEQ ID NO:1至SEQ ID NO:21的任一者或其组合。

[0102] 本文中揭露的组合物可包括编码CRISPR相关核酸内切酶如Cas9的核酸。一些具體实施例中，一种或多种与HIV靶标序列互补的向导RNA也可以被编码。在细菌中，CRISPR/Cas基因座编码RNA引导的对抗可动遗传因子(病毒、转位因子和接合质粒)的适应性免疫系统。已经鉴别了三种类型(I至III)的CRISPR系统。CRISPR簇含有间隔序列，与前驱可动因子互补的序列。CRISPR簇被转录和发展为成熟的CRISPR RNA(crRNA)。CRISPR相关核酸内切酶Cas9属于II型CRISPR/Cas系统，且具有强烈的切割靶标DNA的核酸内切酶活性。Cas9由成熟的crRNA引导，该成熟的crRNA含有约20个碱基对(bp)的独特靶标序列(称为间隔序列)和用作前crRNA的核糖核酸酶III辅助加工导引的反式激活的小RNA(tracrRNA)。该crRNA:tracrRNA双链引导Cas9经由crRNA上该间隔序列与靶标DNA上该互补序列(称为前间隔序列)之间的互补性碱基配对而以该DNA为靶标。Cas9识别三核苷酸(NGG)前间隔序列邻近基序(PAM)以指定切割位点(自PAM起的第3个核苷酸)。该crRNA和tracrRNA可独立被表达，或经由合成性茎环(AGAAAU)而被加工为人工融合小向导RNA(sgRNA)以模拟天然crRNA/tracrRNA双链。这种sgRNA如shRNA可经合成、或体外转录以引导RNA转染或从U6或H1促进的RNA表达载体表达，然而仍sgRNA的裂解效率低于独立表达crRNA和tracrRNA的系统的裂解效率。

[0103] CRISPR相关核酸内切酶可以是Cas9核酸酶。Cas9核酸酶可具有与野生型酿脓链球菌(Streptococcus pyogenes)序列完全一致的核苷酸序列。CRISPR相关核酸内切酶可以是来自其它物种例如其它链球菌种如嗜热菌的序列。Cas9核酸酶序列可以源自其它物种，该其它物种包括但不限于：达松维尔拟诺卡氏菌(Nocardiopsis dassonvillei)、始旋链霉菌(Streptomyces pristinaespiralis)、绿产色链霉菌(Streptomyces viridochromogenes)、淡红链霉菌(Streptomyces roseum)、酸热脂环酸杆菌

(Alicyclobacillus acidocaldarius)、假蕈状芽孢杆菌(Bacillus pseudomycoides)、硒还原芽孢杆菌(Bacillus selenitireducens)、西伯利亚微小杆菌(Exiguobacterium sibiricum)、德氏乳酸杆菌(Lactobacillus delbrueckii)、唾液乳酸杆菌(Lactobacillus salivarius)、微颤菌(Microscilla marina)、伯克氏菌目细菌(Burkholderiales bacterium)、萘降解单胞菌(Polaromonas naphthalenivorans)、单胞菌属(Polaromonas sp.)、海洋固氮蓝藻(Crocosphaera watsonii)、蓝丝菌属(Cyanothece sp.)、铜绿微囊藻(Microcystis aeruginosa)、聚球藻属(Synechococcus sp.)、阿拉伯糖醋盐杆菌(Acetohalobiumarabaticum)、丹氏制氨菌(Ammonifex degensii)、热角军纤维素菌(Caldicelulosiruptor becscii)、金矿菌(Candidatus desulforudis)、肉毒梭状芽孢杆菌(Clostridiumbotulinum)、艰难梭菌(Clostridiumdifficle)、大芬戈尔德菌(Finegoldia magna)、嗜热盐碱厌氧菌(Natranaerobius thermophilus)、热丙酸盐暗色厌氧香肠状菌(Pelotomaculum thermopropionicum)、嗜酸性喜温硫杆菌(Acidithiobacillus caldus)、嗜酸氧化亚铁硫杆菌(Acidithiobacillus ferrooxidans)、异着色菌(Allochromatiumvinosum)、海杆菌属(Marinobacter sp.)、嗜盐亚硝化球菌(Nitrosococcus halophilus)、海洋亚硝化球菌(Nitrosococcus watsoni)、游海假交替单胞菌(Pseudoalteromonas haloplanktis)、消旋纤线杆菌(Ktedonobacter racemifer)、甲烷盐菌(Methanohalobiumevestigatum)、多变鱼腥藻(Anabaena variabilis)、泡沫节球藻(Nodularia spumigena)、念珠藻属(Nostoc sp.)、极大螺旋藻(Arthrospira maxima)、钝顶螺旋藻(Arthrospira platensis)、节螺藻属(Arthrospira sp.)、林氏藻属(Lyngbya sp.)、原型微鞘藻(Microcoleus chthonoplastes)、颤藻属(Oscillatoria sp.)、石袍菌(Petrotoga mobilis)、非洲栖热腔菌(Thermosipho africanus)、或深海单细胞蓝细菌(Acaryochloris marina)。绿脓假单胞菌(Psuedomona aeruginosa)、大肠杆菌(Escherichia coli)、或其它测序的细菌基因组和古生菌、或其它原核微生物也可以是本文中揭露的具體实施例中使用的Cas9序列的来源。

[0104] 野生型酿脓链球菌Cas9序列可经修饰。该核酸序列可以是经优化用于在哺乳动物细胞内有效表达即“人源化”的密码子。该序列可以是，例如，由Genbank登录号KM099231.1GI:669193757、KM099232.1GL669193761、或KM099233.1GI:669193765中所列的任何表达载体编码的Cas9核酸酶序列。或者，该Cas9核酸酶序列可以是，例如，可商购的载体如来自Addgene(Cambridge,MA)的PX330或PX260内含有的序列。一些具體实施例中，Cas9核酸内切酶可具有氨基酸序列，该氨基酸序列是Genbank登录号KM099231.1GI:669193757、KM099232.1GI:669193761、或KM099233.1GI:669193765的任何Cas9核酸内切酶序列或PX330或PX260的Cas9氨基酸序列(Addgene,Cambridge,MA)的变体或片段。可修饰Cas9核苷酸序列以编码Cas9的生物学活性变体，且这些变体可具有或可包括，例如，凭借含有一个或多个突变(如，加入、缺失、或取代突变，或这些突变的组合)而不同于野生型Cas9的序列。一个或多个取代突变可以是取代(如，保守氨基酸取代)。例如，Cas9多肽的生物学活性变体可具有一氨基酸序列，该序列与野生型Cas9多肽的序列一致性为至少或约50％(如，序列一致性为至少或约50％、55％、60％、65％、70％、75％、80％、85％、90％、95％、97％、98％、或99％)。保守氨基酸取代典型包括下述组内的取代：甘氨酸和丙氨酸；缬氨酸、异亮氨酸、和亮氨酸；天冬氨酸和谷氨酸；天冬酰氨、谷氨酰氨、丝氨酸和苏氨酸；赖氨酸、组氨酸和精氨

酸;以及苯丙氨酸和酪氨酸。Cas9氨基酸序列中的氨基酸残基可以是非自然出现的氨基酸残基。自然出现的氨基酸残基包括那些有遗传密码自然编码的氨基酸残基以及非标准氨基酸(如,具有替代L-构型的D-构型的氨基酸)。本发明的肽也可包括作为标准残基的修饰版本的氨基酸残基(如,吡咯赖氨酸可用来替代赖氨酸,而硒半胱氨酸可用来替代半胱氨酸)。非自然出现的氨基酸残基是那些自然界未发现但遵循氨基酸基本分子式且可并入肽中的氨基酸残基。这些非自然出现的氨基酸包括D-别异亮氨酸(2R,3S)-2-氨基-3-甲基戊酸和L-环戊基甘氨酸(S)-2-氨基-2-环戊基乙酸。至于其它实例,可查阅手册或网站(目前由加州理工学院维护的网站,其显示已经被成功并入功能性蛋白质的非自然氨基酸的结构)。

[0105] 本发明的组合物和方法可包括编码与HIV中靶标序列互补的向导RNA的序列。HIV的遗传变异性体现在已经揭示的多个组和亚型中。许多HIV序列收集在Los Alamos HIV数据库和纲要中(即,序列数据库网址为hitp://www.hiv.lani.gov)。本发明的方法和组合物可用于来自哪些各种组别、亚型和循环重组形式的任何HIV。这些HIV包括,例如,HIV-1主要组别(一般指代为M组)的HIV和次要组别的N组、O组和P组HIV,以及,但不限于,下述任何亚型:A、B、C、D、F、G、H、J和K组(例如,但不限于,N组、O组和P组中任何组)的HIV。

[0106] 该向导RNA可以是与编码序列或非编码序列(即,靶标序列)互补的序列。例如,该向导RNA可以是与HIV长末端重复序列(LTR)区域互补的序列,其中该HIV长末端重复序列(LTR)区域不同于使用在通知(inform)该可操作地链接至Cas9基因的截短的、响应Tat启动子的部分。因为将会造成其自身构造的降解,该向导RNA不能以对应于本文中揭露的截短的、响应Tat的HIV-1LTR启动子的序列为靶标,从而潜在地移除了通过由截短的HIV LTR启动子驱动CRISPR相关核酸内切酶所获得的优势。因此,向导RNA可包括在HIV-1U3、R、及/或U5区域参考序列或共有序列中发现的序列,而不选择作为该截短的、响应Tat的HIV启动子的一部分的序列。

[0107] 一些具體实施例中,该向导RNA可以是与编码序列如编码一种或多种病毒结构蛋白(如,gag、pol、env、和tat)的序列互补的序列。因此,该序列可与gag多蛋白如MA(基质蛋白,p17)、CA(衣壳蛋白,p24)、NC(核壳蛋白,p7)、和P6蛋白;pol如逆转录酶(RT)和RNase H、整合酶(IN)、及HIV蛋白酶(PR);env如g 160、或gp160的裂解产物如gp120或SU、以及gp41或TM;或tat如72个氨基酸一个外显子的Tat或86至101个氨基酸两个外显子的Tat内的序列互补。一些具體实施例中,该向导RNA可以是与编码辅助蛋白的序列互补的序列,该辅助蛋白包括,例如,vif、n willef(负因子)、vpu(病毒蛋白U)和tev。

[0108] 一些具體实施例中,该向导RNA序列可以是与结构性或调节性元件(即,靶标序列)如RRE、PE、SLIP、CRS(顺式作用阻抑性序列)、及/或INS互补的序列。RRE(Rev响应元件)是在HIV的env区域内被编码的RNA元件,且包括大约200个核苷酸(HIV-1中从转录起始计的位置7710至8061,跨越gp120与gp41的边界)。PE(Psi元件)对应于处在该Gag起始密码子之前且与该密码子重叠的一组4个茎环结构。SLIP是其后跟随有茎环结构的TTTTTT“湿滑位点”(slippery site)。CRS(顺式作用阻抑性序列)。例如,INS(抑制性/不稳定性RNA序列)可见于HIV-1的gag区域中的核苷酸414至631。

[0109] 该向导RNA序列可以是正义序列或反义序列。该向导RNA序列通常包括PAM。该PAM的序列可依据所使用的CRISPR核酸内切酶的特异性需求而变。在源自酿脓链球菌的CRISPR-Cas系统中,靶标DNA典型位于紧邻5'-NGG前间隔相邻基序(PAM)之前的位置。因此,

对于酿脓链球菌Cas9，该PAM序列可以是AGG、TGG、CGG或GGG。其它Cas9直系同源物可具有不同的PAM特异性。例如，来自嗜热链球菌的Cas9需要5'-NNAGAA用于CRISPR 1且需要5'-NGGNG用于CRISPR3；以及，来自脑膜炎奈瑟氏菌(Neiseria menigiditis)的Cas9需要5'-NNNNGATT。该向导RNA的特异性序列可变，但无论该序列如何改变，可用的向导RNA序列将是那些最小化脱靶效应同时高效且完全废除基因组地整合的HIV前病毒的序列。该向导RNA序列的长度可从约20至约60或更多个核苷酸变化，例如，约20、约21、约22、约23、约24、约25、约26、约27、约28、约29、约30、约31、约32、约33、约34、约35、约36、约37、约38、约39、约40、约45、约50、约55、约60或更多个核苷酸。可用的选择方法鉴定了具有极低的外来病毒基因组与包括内源性逆转录病毒DNA的宿主细胞基因组之间的同源性的区域，该方法包括使用12-bp+NGG靶标选择准则进行生物信息学筛选，以排除脱靶人转录组或(甚至罕见地)未转译的基因组位点；避免HIV-1LTR启动子(潜在地保全于宿主基因组内)内的转录因子结合位点；选择LTR-A-和-B-指向的、30-bp向导RNA以及选择反映原始细菌免疫机制的前-crRNA系统，以提升对比基于20-bp向导RNA、嵌合crRNA-tracRNA的系统的特异性/有效性；以及WGS、Sanger测序和SURVEYOR试验，以鉴定且排除潜在的脱靶效应。

[0110] 该向导RNA序列的构型可以是单一序列，或一种或多种不同序列的组合，如多态构型。多态构型可包括2、3、4、5、6、7、8、9、10或更多中向导RNA的组合，例如，U3、R或U5内序列的组合，而不选择作为该截短的、响应Tat的HIV启动子的一部分的序列。当该组合物于表达载体内给药时，该向导RNA可由单一载体编码。或者，可将多个载体加工为各自包括两种或更多种不同的向导RNA。可用的构型将导致裂解位点之间的病毒序列的切除，导致HIV基因组或HIV蛋白表达的废除。因此，使用两种或更多种不同的向导RNA，促进了被CRISPR核酸内切酶识别的裂解位点之间的病毒序列的切除。被切除区域的尺寸可从一个核苷酸至几千个核苷酸内改变。例示性的被切除区域揭示于具體实施例中。

[0111] 当该组合物作为核酸而给药或包含在表达载体内时，CRISPR核酸内切酶和该向导RNA序列者可由相同的核酸或载体编码。或者或此外，CRISPR核酸内切酶可在与该向导RNA序列物理上分离的核酸中或独立的载体中被编码。一些具體实施例中，将RNA分子如crRNA、tracrRNA、gRNA加工为包含一个或多个修饰核酸碱基。例如，已知的对RNA分子的修饰可见于Lewis编撰的《基因VI》第9章“遗传密码的诠释”(Genes VI,Chapter 9,Interpreting the Genetic Code)(1997,Oxford University Press,New York)以及Grosjean和Benne编撰的《RNA的修饰与编辑》(Modification and Editing of RNA)(1998,ASM Press,Washington DC)。经修饰的RNA组分包括下列：2'-O-甲基胞苷；N^4-甲基胞苷；N^4-2'-O-二甲基胞苷；N^4-乙酰胞苷；5-甲基胞苷；5,2'-O-二甲基胞苷；5-羟甲基胞苷；5-甲酰胞苷；2'-O-甲基-5-甲酰胞苷；3-甲基胞苷；2-硫胞苷；赖胞苷(lysidine)；2'-O-甲基尿苷；2-硫尿苷；2-硫-2'-O-甲基尿苷；3,2'-O-二甲基尿苷；3-(3-氨基-3-羧丙基)尿苷；4-硫尿苷；核糖基胸苷；5,2'-O-二甲基尿苷；5-甲基-2-硫尿苷；5-羟基尿苷；5-甲氧基尿苷；尿苷-5-氧乙酸；尿苷-5-氧乙酸甲酯；5-羧甲基尿苷；5-甲氧羰基甲基尿苷；5-甲氧羰基甲基-2'-O-甲基尿苷；5-甲氧羰基甲基-2'-硫尿苷；5-氨甲酰基甲基尿苷；5-氨甲酰基甲基-2'-O-甲基尿苷；5-(羧基羟甲基)尿苷；5-(羧基羟甲基)尿苷甲酯；5-氨基甲基-2-硫尿苷；5-甲基氨基甲基尿苷；5-甲基氨基甲基-2-硫尿苷；5-甲基氨基甲基-2-硒尿苷；5-羧甲基氨基甲基尿苷；5-羧甲基氨基甲基-2'-O-甲基-尿苷；5-羧甲基氨基甲基-2-硫尿苷；二氢尿苷；二氢核糖基胸

腺嘧啶;2'-甲基腺苷;2-甲基腺苷;N^6,N-甲基腺苷;N^6,N^6-二甲基腺苷;N^6,2'-O-三甲基腺苷;2-甲基硫-N^6,N-异戊烯基腺苷;N^6-(顺-羟基异戊烯基)-腺苷;2-甲基硫-N^6-(顺-羟基异戊烯基)-腺苷;N^6-(环氧丙基氨基甲酰基)-腺苷;N^6-苏氨酰氨基甲酰基腺苷;N^6-甲基-N^6-苏氨酰氨基甲酰基腺苷;2-甲基硫-N^6-甲基-N^6-苏氨酰氨基甲酰基腺苷;N^6-羟基正缬氨酰氨基甲酰基腺苷;2-甲基硫-N^6-羟基正缬氨酰氨基甲酰基腺苷;2'-O-核糖基腺苷(磷酸酯);肌苷;2'-O-甲基肌苷;1-甲基肌苷;1,2'-O-二甲基肌苷;2'-O-甲基鸟苷;1-甲基鸟苷;N^2-甲基鸟苷;N^2,N^2-二甲基鸟苷;N^2,2'-O-二甲基鸟苷;N^2,N^2,2'-O-三甲基鸟苷;2'-O-核糖基鸟苷(磷酸酯);7-甲基鸟苷;N^2,7-二甲基鸟苷;N^2,N^2,7-三甲基鸟苷;怀俄苷(wyosine);甲基怀俄苷;修饰不足的羟基怀丁苷;怀丁苷(wybutosine);羟基怀丁苷;过氧怀丁苷;辫苷(queuosine);环氧辫苷;半乳糖基辫苷;甘露糖基辫苷;7-氰基-7-去氮鸟苷;古生菌苷(arachaeosine)[亦称7-甲酰氨基-7-去氮鸟苷];和7-氨基甲基-7-去氮鸟苷。

[0112] 可通过标准技术生产分离的核酸分子。例如,PCR技术可用来获取含有本文中揭示的核苷酸序列的分离的核酸,包括编码本文中揭示的多肽的核苷酸序列。PCR可用来扩增来自DNA以及RNA的特异性序列,包括来自全基因组DNA或总细胞RNA的序列。多种PCR方法揭示于例如Dieffenbach和Dveksler编撰的《PCR引物:实验室手册》(PCR Primer:A Laboratory Manual,Cold Spring Harbor Laboratory Press,1995)。通常,采用来自感兴趣区域或超出该区域的末端的序列信息来设计寡核苷酸引物,该寡核苷酸引物在序列上与待扩增目标的相反链一致或相似。可使用多种PCR策略,通过该策略,可将位点特异性核苷酸序列修饰引入模板核酸中。

[0113] 分离的核酸也可化学合成为单一核酸分子(如,使用亚磷酰氨技术在3'至5'方向进行自动DNA合成)或合成为一系列寡核苷酸。例如,可将一对或多对长寡核苷酸(如,>50至100个核苷酸)合成为含有所希望的序列,且每一对含有互补性短节段(如,约15个核苷酸),因此,当将该寡核苷酸对退火时形成双螺旋链。使用DNA聚合酶来延伸该寡核苷酸,每对寡核苷酸得到单个的双链核酸分子,随后,该双链核酸分子被连接(ligate)入载体内。本发明的分离的核酸亦可通过诸如Cas9编码DNA的自然出现部分的突变(例如,根据上式)而获得。

[0114] 两个核酸或它们编码的多肽可揭示为彼此具有某种程度的一致性。例如,Cas9蛋白及其生物学活性变体可揭示为展现某种程度的一致性。通过将短Cas9序列在蛋白质信息研究(PIR)网站(http://pir.georgetown.edu)内检索而获得检索结果,之后通过NCBI网站(http://www.ncbi.nlm.mh.gov/blast)上的"短近乎一致序列"局部序列排比检索基本工具(BLAST)算法进行分析。

[0115] 可测定相对于Cas9的序列一致性百分比,且所鉴别的变体可用作CRISPR相关核酸内切酶及/或可试验其作为药物组合物的功效。自然出现的Cas9可以是查询序列,而Cas9蛋白的片段可以是受试序列。

[0116] 同样,Cas9蛋白的片段可以是查询序列,而其生物学活性变体可以是受试序列。为了测测序列一致性,可使用计算机程序ClustalW(版本1.83,默认参数),将存疑核酸或氨基酸序列分别对准一个或多个受试核酸或氨基酸序列,该程序令核酸或蛋白质序列的对准在其整体长度上进行(全局对准)。见Chenna et al.,Nucleic acids Res.31:3497-3500,2003。

[0117] 为了在截短的、响应Tat的HIV LTR启动子控制下表达Cas9,本文中也提供重组构

造，该重组构造可用来转化细。重组构造可类似地用来表达与HIV中靶标序列互补的向导RNA。重组核酸构造包含如本文中揭示的编码Cas9的核酸及/或与HIV中靶标序列互补的向导RNA，可操作地链接至适用于在细胞内表达Cas9及/或与HIV中靶标序列互补的向导RNA的调节性区域。可获知，大量核酸可编码具有特定氨基酸序列的多肽。遗传密码的退化是该领域中广为人知的。对于多数氨基酸，存在不止一个用作氨基酸密码子的核苷酸三元组。例如，可使用适宜的用于特定有机体的密码子偏移表来修饰Cas9编码序列中的密码子，因此获得在该有机体内的优化表达。

[0118] 本文中揭示的核酸可包含在载体中。载体可包括，例如，复制的起源、核骨架附着区域(SAR)、及/或标记物。标记物基因可赋予宿主细胞以可选择的表型。例如，标记物可赋予杀生物剂抗性，如对抗生素(如，卡那霉素、G418、博来霉素、或潮霉素)的耐药性。表达载体可包括被设计来促进操控或检测(如，纯化或定位)所表达的多肽的标签序列。标签序列，如绿色荧光蛋白(GFP)、谷胱甘肽-S-转移酶(GST)、多组氨酸、c-myc、血细胞凝集素、或FlagTM标签(Kodak,New Haven,CT)序列，典型被表达为与所编码的多肽融合。这些标签被插入该多肽内的任意处，包括羧基端或氨基端。

[0119] 也可包括额外的表达载体，例如，染色体系列、非染色体序列或合成DNA序列的节段。适当的载体包括SV40的衍生物和已知的细菌质粒，如大肠杆菌质粒col E1、pCR1、pBR322、pMal-C2、pET、pGEX、pMB9及其衍生物，质粒如RP4；噬菌体DNA，如噬菌体1许多衍生物如NM989，以及其它噬菌体DNA如M13和丝状单链DNA；酵母质粒，如2μ质粒或其衍生物；可用于真核细胞内的载体，如可用于昆虫或哺乳动物细胞内的载体；源自质粒与噬菌体DNA的组合的载体，如已经被修饰以采用噬菌体DNA或其它表达控制序列的质粒。

[0120] 若干输送方法可与该可操作地链接至Cas9基因的截短的、响应Tat的HIV LTR启动子联合用于体外(细胞培养物)和体内(动物和患者)系统。具體实施例中，可使用慢病毒基因输送系统。这一系统在具有广泛趋向性和大DNA插入能力的分裂细胞和非分裂细胞中提供稳定、长期的基因存在(Dull et al.,J Virol,72:8463-8471 1998)。具體实施例中，腺病毒群(AAV)可用作输送方法。AAV是非病原性的单链DNA病毒，近年来，已经被积极地用于在体外和体内系统中输送治疗性基因(Choi et al.,Cnrr Gene Ther,5:299-310,2005)。一种例示性的非病毒输送方法可使用纳米颗粒技术。这一平台已经表明其作为体内药物的实用性。纳米技术已经改善了药物跨越紧密的上皮与内皮屏障的转胞吞作用。该技术提供以特异性方式进行其负载至细胞的靶向输送(Allen and Cullis,Science,303:1818-1822,1998)。

[0121] 该载体亦可包括调节性区域。术语“调节性区域”是指影响转录或转译启动和速率、以及转录或转译产物的稳定性及/或移动性的核苷酸序列。调节性区域包括而不限于，启动子序列、增强子序列、响应元件、蛋白质识别位点、可诱导的元件、蛋白质结合序列、5’和3’未转译区域(UTR)、转录开始位点、终止序列、聚腺苷酸化序列、核定位信号、以及内含子。

[0122] 术语“可操作地链接”是指将核酸中的调节性区域和待转录的序列定位，从而影响该序列的转录或转译。例如，为了将编码序列置于启动子的控制下，典型将该多肽的转译阅读框的转译起始位点定位在该启动子下游的一个至约50个核苷酸之间。然而，可将启动子定位在该转译起始位点上游多达约5,000个核苷酸处或定位在转录起始位点上游约2,000

个核苷酸处。启动子典型包含至少一个核心(基础)启动子。启动子亦可包括至少一个控制元件,如增强子序列,其是上游元件或上游激活区域(UAR)。对待包括的启动子的选择取决于若干因素,包括但不限于,效率、可选择性、可诱导性、所希望的表达水平、以及细胞或组织优先性表达。通过适宜地选择并定位启动子和与编码序列相关的其它调节性区域而调节该编码序列的表达,是该领域技术人员的基本常识。

[0123] 载体包括,例如,病毒载体(如,腺病毒Ad、AAV、慢病毒、以及水泡性口炎病毒(VSV)和逆转录病毒)、脂质体和其它含有脂质的络合物,以及能介导多核苷酸至宿主细胞的输送的其它大分子络合物。载体也可包含其它进一步调节基因输送和/或基因表达的组分或功能性,或反而对靶标细胞提供有益性质的组分或功能性。如下文中更详细揭示和说明,这些其它组分包括,例如,影响结合至细胞或以细胞为靶向的组分(包括介导细胞类型或组织特异性结合的组分);影响细胞摄取该载体核酸的组分;影响该多核苷酸在被细胞摄取后在该细胞内丁烷的组分(如,介导核定位的剂);以及影响该多核苷酸表达的组分。这些组分亦可包括标记物,如可用来检测或选择已经摄取并表达该载体所输送的核酸的细胞的可检测及/或可选择的标记物。这些组分可提供作为该载体的天然特征(如,使用某些具有介导结合和摄取的组分或功能性的病毒载体),或可修饰载体以提供这些功能性。其它载体包括Chen et al.,BioTechniques,34:167-171(2003)中揭示的那些。各种各样的此类载体是该领域中已知且通常可用的。“重组病毒载体”是指包含一种或多种异源性基因产物或序列的病毒载体。由于很多病毒载体展现了与封装相关联的尺寸约束性,典型通过替换该病毒基因组的一个或多个部分而引入该异源性基因产物或序列。这些病毒可变为具有复制缺陷,需要在病毒复制和衣壳化过程中被反向提供所缺失的功能(如,通过使用携带复制及/或衣壳化所必需的基因产物的辅助病毒或封装细胞系)。也已经揭示了待输送的多核苷酸被携带在病毒载体外侧的经修饰的病毒颗粒(见,如,Curiel,D T,et al.PNAS 88:8850-8854,1991)。

[0124] 另外的载体包括病毒载体、融合蛋白和化学接合物。逆转录病毒载体包括莫洛尼鼠白血病病毒和基于HIV的病毒。一个基于HIV的病毒载体包含至少两个载体,其中,gag基因和pol基因来自HIV基因组,而env基因来自另一病毒。DNA病毒载体包括痘病毒载体,如天花病毒载体或禽痘病毒载体、疱疹病毒载体如I型单纯性疱疹病毒(HSV)载体[Geller,A.I.et al.,J.Neurochem,64:487(1995);Lim,F.,et al.,in DNA Cloning:Mammalian Systems,D.Glover,Ed.(Oxford Univ.Press,Oxford England)(1995);Geller,A.I.et al.,Proc Natl.Acad.Sci.:U.S.A.:90 7603(1993);Geller,A.I.,et al.,Proc Natl.Acad.Sci USA:87:1 149(1990)];腺病毒载体[LeGal LaSalle et al.,Science,259:988(1993);Davidson,et al.,Nat.Genet.3:219(1993);Yang,et al.,J.Virol.69:2004(1995)];以及腺病毒群载体[Kaplitt,M.G.,et al.,Nat.Genet.8:148(1994)]。

[0125] 本文中揭露的多核苷酸可与微输送载具如阳离子脂质体和腺病毒载体合用。关于对脂质体制备、靶向和内容为输送的综述,见Mannino and Gould-Fogerite,BioTechniques,6:682(1988)。也见,Feigner and Holm,Bethesda Res.Lab.Focus,1 1(2):21(1989)和Maurer,R.A.,Bethesda Res.Lab.Focus,1 1(2):25(1989)。

[0126] 可根据已知技术生产复制缺陷型重组腺病毒载体。见,Quantin,et al.,Proc.Natl.Acad.Sci.USA,89:2581-2584(1992);Stratford-Perricadet,et al.,

J.Clin.Invest.,90:626-630(1992)；和Rosenfeld,et al.,Cell,68:143-155(1992)。

[0127] 另一输送方法是使用生产单链DNA的载体，该载体可生产在细胞内表达的产物。见，例如，Chen et al.,BioTechniqiies,34:167-171(2003)，该文献通过引用而以其整体并入本文。

[0128] 如上所述，可以该领域技术人员所知的多种途径制备本发明的组合物。无论其原始来源或获得途径如何，可根据其用途配制本文中揭露的组合物。例如，上述核酸和载体可配制在应用于组织培养的细胞中或用于给药至患者或受试者的组合物中。本发明的任何药物组合物可配制为用于药剂的制备中，且特别用途为下述治疗情况中所表明，如，对具有HIV感染的受试者或处于接触HIV和HIV感染风险下的受试者的治疗。当作为药物而使用时，任何该核酸和载体均可以药物组合物的形式给药。这些组合物可以制药领域广为人知的方式制备，且可通过多种途径给药，给药途径取决于所需者是否为局部或全身治疗以及待治疗的面积。给药可以是外用给药(包括眼部给药和给药至粘膜，包括经鼻腔输送、经阴道输送、和经直肠输送)、肺部给药(如，通过粉末或气溶胶的吸入或吹入，包括通过喷雾器给药；气管内、鼻腔、上皮、和透皮给药)、眼部给药、口服给药或肠胃外给药。用于眼部输送的方法可包括外用给药(滴眼液)，结膜下、眼周或玻璃体内注射，或通过以外科手术置于结膜囊内的气囊导管或眼科插入物引入。肠胃外给药包括静脉、动脉、皮下、腹膜、或肌肉注射或输液；或颅内如鞘内或心室内给药。肠胃外给药可以是单次推注剂量(single bolus dose)的形式，或可以是，例如，通过连续灌流泵给药。用于外用给药的药物组合物和制剂可包括透皮贴剂、软膏剂、洗剂、乳液、凝胶剂、滴剂、栓剂、喷雾剂、液剂、粉剂等。传统的药学载剂、水性的、粉末或油性基质、增稠剂等可能是必需的或所希望的。

[0129] 该药物组合物可含有，作为活性成分，本文中揭示的核酸和载体与一种或多种药学可接受的载剂合用。在作成本发明的组合物时，该或形成的典型与赋形剂混合，通过赋形剂稀释或密封在例如胶囊、片剂、小袋、纸张或其它容器形式的载剂内。当该赋形剂用作稀释剂时，它可以是固体、半固体或液体材料(如，生理盐水)，其用作该活性成分的载具、载剂或介质。因此，该组合物可以是片剂、丸剂、粉剂、锭剂、袋剂、小锭剂、酏剂、悬浮剂、乳剂、溶液剂、糖浆剂、气溶胶(作为固体或处于液体介质中)、洗剂、乳液、软膏剂、凝胶剂、软胶囊和硬胶囊、栓剂、无菌注射液、和无菌封装粉末的形式。如该领域已知的，稀释剂的类型可依据预设的给药途径而变。所得组合物可包括另外的剂，如防腐剂。一些具體实施例中，该载剂可以是或可包括基于脂质的胶体或基于聚合物的胶体。一些具體实施例中，该载剂材料可以是配制为脂质体、水凝胶、微颗粒、纳米颗粒或嵌段共聚物胶束的胶体。如标注，载剂材料可形成胶囊，且该材料可以是基于聚合物的胶体。

[0130] 本发明的核酸序列可输送至受试者的适宜细胞。这可通过例如使用聚合物的、生物可降解微颗粒或微胶囊输送载具而实现，其中，该载具的尺寸为适于吞噬细胞如巨噬细胞吞噬的最佳尺寸。例如，可使用直径约为1至10μm的PLGA(聚-丙交酯-共-乙交酯)微颗粒。该多核苷酸被封装在这些微颗粒内，该微颗粒被巨噬细胞摄取并在该细胞里逐渐被生物降解，从而释放该多核苷酸。一旦被释放，该DNA即在该细胞内表达。第二种类型的微颗粒倾向于不被细胞直接摄取，但主要用作核酸的缓释蓄积容器，该核酸仅在通过生物降解从该微颗粒中释放时才被细胞摄取。这些聚合性颗粒应因此足够大，以阻碍吞噬作用(即，大于5μm，且优选大于20μm)。另一实现该核酸的摄取的途径是使用通过标准方法制备的脂质体。该

核酸可单独并入其输送载具中，或与组织特异性抗体一起并入载具中，该抗体为例如以一般作为HIV感染的潜伏性感染蓄积容器的细胞类型为靶标的抗体，该细胞为例如脑巨噬细胞、小胶质细胞、星形胶质细胞、和肠相关淋巴样细胞。或者，可制备由通过静电力或共价力附着至聚-L-赖氨酸的质粒或其它载体构成的分子络合物。聚-L-赖氨酸结合配体，而该配体可结合至靶标细胞上的受体。“裸DNA”(即没有输送载具)至肌肉内、真皮内或皮下位点的输送，是另一实现体内表达的手段。如上文所述，在相关的多核苷酸(如，表达载体)中，编码分离的核酸序列的核酸序列被可操作地链接至该截短的、响应Tat-的HIV LTR启动子，其中，该分离的核酸序列包含编码CRISPR相关核酸内切酶和人选向导RNA的序列。

[0131] 一些具體实施例中，本发明的组合物可配制为纳米颗粒，例如，由与DNA络合的高分子量线性聚乙烯亚氨(LPEI)核心和环绕该核心的聚乙二醇修饰(PEG化)的低分子量LPEI外壳构成的纳米颗粒。

[0132] 该核酸和载体也可应用于装置(如，导管)的表面，或包含在泵、贴、或其它药物输送装置内。在药学可接受的赋形剂或载剂(如，生理盐水)的存在下，本文中揭露的核酸和载体可单独给药或以混合物形式给药。该赋形剂或载剂是基于给药模式和途径选择的。用于药学制剂的适当的药学载剂以及药学必需品揭示于本领域广为人知的参考书《雷明顿药物科学》(Remington’s Pharmaceutical Sciences(E.W.Martin))中以及USP/NF(《美国药典和国家处方集》(United States Pharmacopeia and the National Formulary))中。

[0133] 一些具體实施例中，该组合物可配制为用于阻断HIV的性传播的外用凝胶。该外用凝胶可在性活动之前，直接施加至男性或女性生殖器区域的皮肤或黏膜上。或者或此外，该外用凝胶可施加至男用或女用避孕套或隔膜的表面上或其内部。

[0134] 一些具體实施例中，该组合物可配制为纳米颗粒，该纳米颗粒内封装有编码可操作地链接至截短的HIV LTR启动子的Cas9或Cas9变体的核酸。该核酸还可编码与靶标HIV互补的向导RNA序列。

[0135] 本发明的制剂可囊括编码Cas9和与靶标HIV互补的向导RNA序列的载体。该向导RNA序列可包括与单一靶标区互补的序列，或其可包括任何与早前揭示的多个靶标区互补的序列的任意组合。或者，该编码由截短的HIV LTR启动子驱动的Cas9的序列与编码该向导RNA序列的序列可以位于分开的载体上。

[0136] 本文中揭露的组合物通常且在多方面可用于治疗具有HIV感染的受试者。该方法可用于以任何HIV如HIV-1和HIV-2、SIV、及其任何循环重组形式为靶向。每当临床有益的结果发生时，则受试者被有效治疗。这可能意味着，例如，疾病症状的完全消退、疾病症状的严重性下降、或疾病进展的减慢。这些方法可进一步包括步骤:a)鉴别具有HIV感染的受试者(如，患者，且更具体而言，人类患者)；和b)向该受试者提供包含核酸的组合物，其中，该核酸编码处于截短的、响应Tat的HIV LTR启动子控制下的CRISPR相关核酸酶如Cas9。该方法可进一步包括向该受试者提供编码与HIV靶标序列如HIV LTR互补的向导RNA的序列。

[0137] 可使用标准临床测试鉴别受试者，例如，使用免疫试验来检测该受试者血清内是否存在HIV抗体或HIV多肽p24，或通过HIV核酸扩增试验进行测试。将提供至该受试者且造成感染症状完全消退、感染的严重性下降、或感染的进程减缓的该组合物的量视为治疗有效量。本发明的方法亦可包括监控步骤，以帮助优化剂量、给药时间表、以及预期成果。本发明的一些方法中，可首先测定患者是否具有潜伏性的HIV感染，随后测定是否以一种或多种

本文中揭示的组合物治疗该患者。监控也可用来检测耐药性的发作，以及快速区分敏感患者与不明患者。一些具體实施例中，该方法可进一步包括检测该患者携带的特定HIV的核酸序列，以及随后设计待与那些特测序列互补的向导RNA的步骤。例如，可测定受试者的LTR U3、R、或U5区域的核酸序列，并随后设计与该患者的序列精确互补的一种或多种向导RNA，再一次地，无需选择作为该截短的、响应Tat的HIV启动子的一部分的序列。

[0138] 该组合物亦可用于对处于逆转录病毒感染如HIV感染的风险下的受试者的治疗，例如，作为对该受试者的预防性治疗。这些方法可进一步包括步骤：a)鉴别具有HIV感染的受试者；和b)向该受试者提供包含核酸的组合物，其中，该核酸编码处于截短的、响应Tat的HIV LTR启动子控制下的CRISPR相关核酸酶如Cas9。该序列可额外地编码与HIV靶标序列如HIV LTR互补的向导RNA。处于HIV感染的风险下的受试者可以是，例如，任何参与不安全性行为的性活跃个体，即参与不使用避孕套的性活动者；具有另一性传播感染的性活跃个体；静脉吸毒者；或未受割礼的男性。处于HIV感染的风险下的受试者可以是，例如，其职业可能令其与感染HIV的群体接触的个体，如医护工作者或现场急救员。处于HIV感染的风险下的受试者可以是，例如，监狱环境的同被收容者或性工作者，换言之，使用性活动获得收益或非货币项目如食品、药品或庇护所的个体。

[0139] 该组合物亦可给药至具有HIV感染的怀孕期或哺乳期妇女，以减轻HIV从母亲向其子女传播的可能性。在通过产道分娩时，感染有HIV的孕妇可经胎盘将病毒传递至子宫内的子女，或在分娩后，通过乳汁将病毒传递至子女。本文中揭露的组合物可在产前、围产期、或产后的哺乳期给药至感染HIV的母亲，或以产前、围产期、或产后给药的任意组合给药。该组合物可与下文揭示的标准抗逆转录病毒疗法一起给药。一些具體实施例中，本发明组合物也在分娩后立即给药至婴儿，以及，一些具體实施例中，在其后不时地给药。该婴儿也可接受标准抗逆转录病毒治疗。

[0140] 该组合物可给药至未感染HIV的个体，以防止其感染HIV。该组合物可包括输送治疗有效量的药物组合物。该药物组合物可包括编码CRISPR相关核酸内切酶的序列，以及至少上述HIV LTR启动子的核心区域和该截短的、响应Tat的HIV LTR启动子的TAR区域。

[0141] 本文中揭露的方法可施加至多个物种，如人类、非人灵长类(如，猴子)、马或其它家畜、狗、猫、雪貂或其它作为宠物豢养的哺乳动物、大鼠、小鼠、或其它实验室动物。

[0142] 本发明的方法可表示为医药制剂。据此，本发明涵盖本文中揭示的剂和组合物在医药制剂中的用途。本文中揭示的化合物可用于治疗性组合物和方案中，或可用于制造治疗本文中揭示的疾病或病症的医药制品。

[0143] 本文中揭示的任何组合物可给药至宿主身体的任何部位，以进行后续至靶标细胞的输送。组合物可输送至而非限于，哺乳动物的脑、脑脊液、关节、鼻黏膜、血液、肺、肠、肌肉组织、皮肤、或腹膜腔。就输送途径而言，组合物可通过静脉、颅内、腹膜、肌肉、皮下、肌肉、直肠、阴道、鞘内、气管内、皮内、或透皮注射给药，通过口服或鼻腔给药，或通过随时间渐变的灌注给药。其它实例中，可通过吸入将组合物的气溶胶制剂给至宿主。

[0144] 所需的剂量将取决于给药途径；制剂的特性；患者疾病的特性；患者的身高、体重、表面积、年龄和性别；正在给药的其它药物；以及临床主治医生的判断。就细胞靶标的多样性和多种给药途径的不同效率而言，预期所需的剂量变化范围广。可使用标准经验性优化途径如该领域中广为人知的途径，调节这些剂量水平的变化。给药可以是单剂或多剂(如，2

倍或3倍、4倍、6倍、8倍、10倍、20倍、50倍、100倍、150倍、或更多倍)。将化合物封装在适当地输送载具(如,聚合物微颗粒或可植入的装置)中可增加输送效率。

[0145] 使用本文中提供的任何组合物进行治疗的持续时间可以是从短至一天到长达宿主整个寿命(如,很多年的任意长度的时间。例如,化合物可给药每周一次(共给药例如4周至很多月或很多年);每月一次(共给药例如3至12个月或很多年);或每年一次,给药5年、10年或更长的时间。也应注意,治疗的频率可变。例如,本发明的化合物可每天、每周、每月、或每年给药一次(或两次、三次等)。

[0146] 可将有效量的本文中提供的任何组合物给药至需要治疗的个体。可通过评估患者在给药已知量的特定组合物后的响应来测定有效量。此外,若有毒性,则可通过评估患者在给药已知量的特定组合物之前和之后的临床症状而测定该毒性的水平。注意,给药至患者的特定组合物的有效量可根据所希望的结果以及患者的响应和毒性水平进行调节。对于每一特定患者的重大毒性可能有变化,且该变化取决于多种因素,包括但不限于,该患者的疾病状态、年龄以及对副作用的耐受性。

[0147] 该领域中已知的任何方法均可用来测定是否诱发了特定的响应。可评估特定疾病状态的程度的临床方法可用来测定是否诱发了响应。用来估量响应的特定方法将取决于该患者病变的特性、患者的年龄和性别、正在服用的其它药物、以及主治医生的判断。

[0148] 该组合物也可与其它治疗剂如用于HAART中的抗逆转录病毒剂一起给药。抗逆转录病毒剂可包括逆转录酶抑制剂(如,核苷/核苷酸逆转录酶抑制剂、齐多夫定、恩曲他滨(emtricitibine)、拉米夫定和替诺福韦(tenoifvir);以及非核苷逆转录酶抑制剂如依法韦仑(efavarenz)、奈韦拉平、利匹韦林(rilpivirine));蛋白酶抑制剂,如替匹拉韦(tipiravir)、地瑞那韦(darunavir)、茚地那韦;进入抑制剂,如马拉维若(maraviroc);融合抑制剂,如恩夫韦地(enfuviritide);或整合酶抑制剂,如雷特格韦(raltegrivir)、度鲁特韦(dolutegravir)。抗逆转录病毒剂亦可包括该剂的多类别组合,例如,恩曲他滨、依法韦仑与替诺福韦的组合;恩曲他滨、利匹韦林与替诺福韦的组合;或埃替拉韦(elvitegravir)、可比司他(cobicistat)、恩曲他滨与替诺福韦的组合。

[0149] 两种或多种治疗剂的并行给药并不需要将该等剂在相同时间或通过相同途径给药,只要在该等剂发挥其治疗功效的时间段存在重叠即可。同步或依序给药预计为在不同日或不同周内给药。该治疗剂可以节奏性方案给药,如持续给药低剂量的治疗剂。

[0150] 可通过标准药学过程在细胞培养物中或实验动物中测定这些组合物的剂量、毒性和疗效,如测定LD_{50}(将群体的50%致死的剂量)和ED_{50}(对群体的50%治疗上有效的剂量)。毒性与疗效之间的剂量比是治疗指数,且可表示为LD_{50}/ED_{50}比。

[0151] 从细胞培养试验和动物研究获得的数据可用于制订在人体中使用的剂量范围。这一组合物的剂量优选处于包括具有低毒或无毒的ED_{50}的循环浓度范围内。该剂量可依据所采取的剂型和所使用的给药途径在这一范围内变化。对于在本发明方法中使用的任何组合物,最初可从细胞培养试验中估算治疗有效剂量。可在动物模型中制订剂量,以实现包括在细胞培养物中测得的IC_{50}(即,测试化合物实现对症状的半最大抑制的浓度)的循环血浆浓度范围。这一信息可用来更准确地测定可在人体内使用的剂量。可例如通过高效液相色谱测量血浆中的水平。

[0152] 如上所述,治疗有效量的组合物(即,有效剂量)意为足以产生治疗上(如,临床上)

所希望结果的量。该组合物的给药频率为每天一次或多次至每周一次或多次;包括每两天一次。该领域技术人员应了解,某些因素可影响有效治疗受试者所需的剂量和时机,该因素包括但不限于,疾病或病症的严重性、先前治疗、该受试者的一般健康状况及/或年龄、以及存在的其它疾病。此外,使用治疗有效量的本发明的组合物治疗受试者可包括单一治疗或一系列治疗。

[0153] 本文中揭示的组合物适用于上揭的多种药物输送系统。另外,为了提升所给药化合物的体内血清半衰期,该组合物可将胶囊化、引入脂质体内腔、制备为胶体,或可采用向该组合物提供延长的半衰期的其它传统技术。多种方法可用于制备脂质体,如在Szoka等人的美国专利US 4,235,871、US 4,501,728和US 4,837,028中揭示的,这些专利各自通过引用而并入本文。再者,人们可在靶向药物输送系统中给药该药物,例如,在涂覆有组织特异性抗体的脂质体中给药。该脂质体将会以某一器官为靶向并被该器官选择性地摄取。

[0154] 还提供令哺乳动物细胞中的逆转录病毒如慢病毒如人免疫缺陷病毒、猿免疫缺陷病毒、猫免疫缺陷病毒、牛免疫缺陷病毒失活的方法。该人免疫缺陷病毒可以是HIV-1或HIV-2。该人免疫缺陷病毒可以是染色体整合的前病毒。该哺乳动物细胞可以是被HIV感染的任何细胞类型,包括但不限于,$CD4^+$淋巴细胞、巨噬细胞、纤维母细胞、单核细胞、T淋巴细胞、B淋巴细胞、天然杀手细胞、树突细胞如朗格汉斯细胞和滤泡树突细胞、造血干细胞、内皮细胞、脑小胶质细胞、星形胶质细胞和胃肠上皮细胞。这些细胞类型包括那些典型在初次感染过程中被感染的细胞类型,例如,$CD4^+$淋巴细胞、巨噬细胞、单核细胞或朗格汉斯细胞,以及那些组成潜伏性HIV逆转录病毒的细胞类型,即潜伏性感染细胞。

[0155] 该方法可包括将细胞暴露于及/或接触包含编码CRISPR相关核酸内切酶的分离的核酸的组合物,其中该酶可操作地链接至截短的HIV LTR启动子,且该截短的启动子含有HIV LTR启动子的核心区域和TAR区域。该分离的核酸可进一步编码一种或多种向导RNA,其中,该向导RNA与该逆转录病毒中的靶标核酸序列互补。该接触步骤可发生在体内,换言之,该组合物可直接给药至具有HIV感染的受试者。但该方法并非受限于此,且该接触步骤可以间接体内方式发生。例如,可从具有HIV感染的受试者移除一个细胞或多个细胞、或组织外植体并培养,随后令其与一组合物接触,其中,该组合物包含可操作地链接至截短的HIV LTR启动子的CRISPR相关核酸内切酶以及任选的向导RNA,其中该向导RNA与HIV中的核酸序列互补。如上所述,药物组合物可包括编码CRISPR相关核酸内切酶的核酸,其中该酶可操作地链接至截短的、响应Tat的HIV LTR启动子。

[0156] 该组合物以促进其被哺乳动物细胞摄取的方式配制。可用的载体系统和制剂如上所述。一些具體实施例中,载体可将该组合物输送至特异的细胞类型。但本发明并不限于此,也可采用其它DNA输送方法如使用诸如磷酸钙、DEAE右旋糖酐、脂质体、脂质复合体、表面活性剂和全氟化学液体的化学转染,以及物理输送方法如电穿孔系统、微量注射系统、弹道颗粒(ballistic particle)系统、及“基因枪”系统。

[0157] 可使用标准方法如,用以检测CRISPR相关核酸内切酶的免疫试验、或用以检测向导RNA的基于核酸的试验如PCR,来证实细胞已经摄取及/或表达已经被引入该细胞内的蛋白。随后,如下所示,所加工的细胞可再次引入作为其来源的受试者体内。

[0158] 其它具體实施例中,该组合物包含已经使用一种或多种Cas9/截短的、响应Tat的HIV LTR启动子载体转化或转染的细胞。一些具體实施例中,本发明的方法可以间接体内方

式应用。换言之，受试者的细胞可自其身体移除，并使用该组合物在培养中除了以切除HIV序列，再将治疗后的细胞送回该受试者身体内。该细胞可以是受试者的细胞，或它们可以是单体型匹配的细胞或细胞系。可辐射该细胞以防止复制。一些具體实施例中，该细胞是人白细胞抗原(HLA)匹配的细胞、自体细胞、细胞系、或其组合。其它具體实施例中，该细胞可以是干细胞。例如，胚胎干细胞或人工多能干细胞(诱导多能干细胞(iPS cell))。已经评价了来自很多动物物种包括人类的胚胎干细胞(ES细胞)和人工多能干细胞(诱导多能干细胞，iPS细胞)。这些类型的多能干细胞将会是用于再生医学的最有用的细胞来源，因为这些细胞能通过对其分化的适宜诱导而分化为几乎全部器官，且在维持其多能性的同时保留其旺盛的分化能力。特别地，与通过摧毁配体而生产的ES细胞相比，iPS细胞可从自源体细胞建立且因此不太可能造成伦理和社会问题。再者，作为自源性细胞，iPS细胞可能避免排异反应，而配以反映是再生医学或移植治疗的最大障碍。

[0159] 本文中揭示的组合物可保证在适当的标记容器内，例如，用于作为治疗来处理逆转录病毒感染如HIV感染的受试者或处于逆转录病毒感染如HIV感染风险下的受试者。该容器可包括组合物，该组合物包含早前揭示的编码CRISPR相关核酸内切酶如Cas9核酸内切酶的核酸序列以及截短的、响应Tat的HIV LTR启动子。该序列可额外地编码与HIV中靶标序列互补的向导RNA，或编码该核酸的载体、以及适用于预期用途的一种或多种适当的稳定剂、载剂分子、芳香剂(flavoring)等。据此，包装的产品(如，含有本文中揭示的一种或多种组合物的无菌容器，并被包装用于储存、运输或以浓缩或易于使用的浓度销售)以及试剂盒包括至少一种所揭露的组合物。产物可包括含有一种或多种本发明组合物的容器(如，小瓶、罐、瓶、袋等)。此外，可进一步包括加工件，例如，封装材料、使用说明书、注射器、输送装置、缓冲剂或用于处理或监控预防或治疗所需条件的其它控制试剂。一些具體实施例中，该试剂盒可包括一种或多种额外的抗逆转录病毒剂，例如，逆转录酶抑制剂、蛋白酶抑制剂、或进入抑制剂。该额外的剂可与核酸序列或编码该核酸的载体一起封装在相同容器内，或它们可单独封装，其中，该核酸编码可操作地链接至截短的HIV LTR启动子的CRISPR相关核酸内切酶如Cas9以及任选的与HIV中靶标序列互补的向导RNA。

[0160] 该产品亦可包括图例(如，描述该产品用途的印刷标记或插页或其它介质(如，录音带或录影带))。该图例可与该容器关联(如，粘附于该容器)，且可揭示该容器内的组合物应采取的给药方式(如，给药频率和途径)、其说明、及其它用途。该组合物可易于给药(如，以剂量适当的单元形式存在)，且可包括一种或多种额外的药学可接受的佐剂、载剂或其它稀释剂、及/或额外的治疗剂。或者，该组合物可提供为具有稀释剂和稀释说明的浓缩形式。

[0161] 通过下述非限制性实施例对本发明实践进行例示性说明。

[0162] [实施例]

[0163] 实施例1:LTR-Cas9变体的克隆

[0164] 使用pNL4-3HIV载体(NIH AIDS试剂程序#114)作为模板和下述引物(限制位点加粗标注)的PCR获得全长度的和多种截短的LTR启动子序列：

[0165] Kpnl-LTR(-454)-S 5'-GGTACCTGGAAGGGCTAATTTGG-3'(SEQ ID NO:1)

[0166] Kpnl-LTR(-120)-S 5'-GGTACCTCGAGCTTTCTACAAGG-3'(SEQ ID NO:2)

[0167] Xbal-LTR(-80)-S 5'-TCTAGAGGAGGTGTGGCCTGGGC-3'(SEQ ID NO:3)

[0168] Kpnl-LTR(-38)-S 5'-GGTACCAGATGCTACATATAAGC-3'(SEQ ID NO:4),or

[0169] LTR(+66)-Ncol-AS 5'-CC ATGGTAAGC AGTGGGTTCC-3'(SEQ ID NO:5).

[0170] 参照LTR的U3区、R区和U5区以及LTR增强子元件、核心元件、TAR(反射激活响应)元件和TATA盒元件,截短的HIV-1LTR启动子变体的由来概略地显示于图1A中。图1B显示经PCR扩增的LTR截断变体的琼脂糖凝胶电泳图像。

[0171] PCR产物经凝胶纯化并直接在TA载体(Invitrogen)中亚克隆,随后使用Kpnl或Xbal及Ncol限制性内切酶切除并连接入作为Cas9基因来源及/或模板的经Kpnl-Ncol或Xbal-Ncol消解的pX260-U6-DR-BB-DR-Cbh-NLS-hSpCas9-NLS-Hl-短shorttracr-PGK-puro质粒(Addgene#42229)(后文称"pX260"质粒)中。pX260质粒含有Cbh启动子(Xbal-Kpnl-Cbh-Ncol)。作为该操作的结果,pX260质粒中的原始Cbh启动子被移除并替换为一种LTR启动子(Xbal-或Kpnl-LTR-Ncol)。图2中显示的原始pX260质粒结构的蓝图被认定为"Cbh-Cas9"(来自www.Addagene.org和Cong et al.,Science(2013)339(6121):819-23)。经修饰质粒的蓝图作为"LTR-Cas9"显示在图2中。

[0172] 实施例2:用于诱导Cas9表达的LTR/Tat比的优化

[0173] 为了找到并优化具有最佳反式激活效果的Tat与LTR启动子的比值,在具有或不具有Tat表达质粒(pCMV-Tat86,250ng)的全长度HIV-1LTR(pLTR(-454/+66)-FLAG-Cas9)质粒(10、50和250ng)的对照下,使用具有不同量的表达FLAG标记Cas9的Lipofectamine 2000试剂(Invitrogen)共转染人初代成胶质细胞瘤细胞系U87MG的细胞。U87MG是HIV-1潜伏细胞系。以空pCMV质粒(pcDNA3.1)平衡DNA的总量。48小时后,在TNN缓冲液(50mM Tris pH 7.4、100mM NaCl、5mM EDTA、1%NP 40)中裂解细胞。随后通过蛋白质印迹法检查Cas9、Tat和α微管蛋白的表达。结果显示于图3A中(U87MG WCE 50μg/孔)。道次1:pLTR(-454/+66)-Cas9 250ng、pCMV 1000ng。道次2:pLTR(-454/+66)-Cas9 50ng、pCMV 1200ng。道次3:pLTR(-454/+66)-Cas910ng、pCMV 1240ng。道次4:pLTR(-454/+66)-Cas9 250ng、pCMV 750ng、pCMV-Tat86 250ng。道次5:pLTR(-454/+66)-Cas9 50ng、pCMV 950ng、pCMV-Tat86 250ng。道次6:pLTR(-454/+66)-Cas910ng、pCMV 990ng、pCMV-Tat86 250ng。

[0174] 分析并使用ImageJ软件比较对应于Cas9和α-微管蛋白(用作载荷对照)的谱带的强度。结果显示于图3B中。上图显示归一化至α-微管蛋白水平的具有或不具有Tat的Cas9水平的蛋白质印迹图像定量。下图显示+Tat/无Tat比的蛋白质印迹图像定量。结果表明,以1:5的pLTR-Cas9:pCMVTat86(50ng:250ng)比获得了对Cas9表达的最大(5.3x)诱导。

[0175] 实施例3:多种截短的LTR启动子在诱导Cas9表达中的比较

[0176] 为了测试和比较截短的LTR启动子,在使用或不使用表达Tat的质粒(pCMV-Tat86,250ng)下,使用不同量的在HIV-1截短的LTR变体pLTR(-120/+66)-FLAG-Cas9或HIV-1截短的LTR变体pLTR(-80/+66)-FLAG-Cas9的控制下表达FLAG标记的Cas9的质粒(5ng或50ng)转染U87MG细胞。48小时后,制备全细胞裂解液并通过蛋白质印迹法分解。分析并使用ImageJ软件比较对应于Cas9和α-微管蛋白(用作载荷对照)的谱带的强度。结果显示于图4A中。道次1:pLTR(-120/+66)-Cas9 5ng、pCMV 1245ng。道次2:pLTR(-120/+66)-Cas9 5ng、pCMV 1245ng、+rTat蛋白2.5μg/ml。道次3:pLTR(-120/+66)-Cas9 5ng、pCMV 995ng、pCMV-Tat86 250ng。道次4:pLTR(-120/+66)-Cas9 50ng、pCMV 1200ng。道次5:pLTR(-120/+66)-Cas9 50ng、pCMV 1200ng、+rTat蛋白2.5μg/ml。道次6:PLTR(-120/+66)-Cas9 50ng、pCMV 950ng、pCMV-Tat86 250ng。道次7:pLTR(-80/+66)-Cas9 5ng、pCMV 1245ng。道次8:pLTR(-80/+

66)-Cas9 5ng、pCMV 1245ng、+rTat蛋白2.5μg/ml。道次9:pLTR(-80/+66)-Cas9 5ng、pCMV 995ng、PCMV-Tat86 250ng。道次10:pLTR(-80/+66)-Cas9 50ng、pCMV 1200ng。道次11:pLTR(-80/+66)-Cas9 50ng、pCMV 1200ng、+rTat蛋白2.5μg/ml。道次12:pLTR(-80/+66)-Cas9 50ng、pCMV 950ng、PCMV-Tat86 250ng。

[0177] 分析并使用ImageJ软件比较对应于Cas9和α-微管蛋白(用作载荷对照)的谱带的强度。结果显示于图4B中。上图显示归一化至α-微管蛋白水平的不具有Tat、具有rTAT、或具有经转染Tat的Cas9水平的蛋白质印迹图像定量。下图显示+Tat(经转染)/无Tat比的蛋白质印迹图像定量。结果表明,移除LTR的调节区域及/或增强子区域(那些区域按照图式表示在图1A中)并未显着影响Tat介导的反式激活的Cas9表达。Tat介导的表达从pLTR(-80/+66)-FLAG-Cas9质粒明显可见,该质粒含有该核心元件和TAR LTR启动子元件而不含有增强子区域和调节区域。

[0178] 实施例4:通过基因编辑策略对HIV-1的负反馈调节

[0179] 在本文呈递的研究中,该基因编辑组合物允许在病毒再次激活的早期,通过HIV-1转录活化因子Tat对CRISPR/Cas9进行有条件的激活。这一新策略在生产性病毒复制之前,通过移除该病毒基因的跨越该病毒启动子及/或该病毒编码序列的节段,完全且永久地废除了病毒复制。再者,这一策略减轻了由于可能由Cas9在细胞内的非必要且持续的高水平表达引起的意料之外的并发症而获得关注的点。

[0180] 结果

[0181] 将对应于Cas9基因的编码DNA序列置于含有该HIV-1启动子的跨越5'-LTR的U3区和R区的三个不同节段的pX26表达载体质粒内,以鉴别病毒启动子的最小DNA元件,该最小DNA元件保留对Tat的响应性但缺少对应于最初用于编辑HIV-1DNA的gRNA A和gRNA B的序列(图5A)。在通过对每一构造进行DNA测序而核实这一克隆策略后,检测通过每一载体进行的Cas9表达以及在pX26或pX26-LTR-Cas9和CMV-Tat共转染的TZMb1细胞中对Tat的响应性水平。蛋白质印迹结果显示,全部三种包括该涵盖位于-80至+66之间位置的最小DNA启动子序列的质粒的构造均激活了Cas9的表达(图5B)。由于该启动子序列处于与gRNA A和gRNA B相对应的DNA序列之外,这一激活对于这些研究尤为重要(图5B)。此后,将对应于$LTR_{(-80/+66)}$-Cas9的DNA片段克隆入慢病毒载体(LV)内并用来转导TZMb1细胞,以评估Tat蛋白在编辑表达荧光素酶报告基因的整合HIV-1DNA复本上的效果。LTR的PCR扩增结果显示,检测表达gRNA A和gRNA B以及Tat蛋白的细胞内的205bp DNA片段的结果(图5C,将道次1至5与道次6至8相比较)。图5A(也见图10)中例示性说明了用于PCR扩增的引物的位置和预期的扩增子。测序结果证实,一旦Tat在通过LV-$LTR_{(-80/+66)}$-Cas9加上LV-gRNA A/B转导的细胞内被表达,190bp DNA片段即被切除。Cas9、Tat和α-微管蛋白(作为相同负载的对照)的表达显示在图5D中。

[0182] 此后,通过荧光素酶试验检查病毒DNA切除对病毒启动子活性的影响。结果显示,当通过Tat激活Cas9时,荧光素酶活性逐步下降,证实了DNA试验的结果,表明DNA的裂解造成了对这些细胞内病毒启动子活性的抑制(图5E)。随访研究中,考察了TZMb1细胞被HIV-1感染时的Cas9的激活。为了这一目的,细胞以LV-$LTR_{(-80/+66)}$-Cas9和LV-gRNA A/B转导24小时,之后,使用$HIV\text{-}1_{JRFL}$或$HIV\text{-}1_{SF162}$以三种不同的MOI感染细胞。48小时后,收获细胞,通过PCR评估DNA切除,通过荧光素酶试验评估整合启动子序列的表达,并通过蛋白质印迹法评

估Cas9的表达。这些实验的结果显示，在以HIV-1_{JRFL}和HIV-1_{SF162}感染的细胞内检测出裂解后的205bp DNA片段，表明在这些细胞中，通过HIV-1_{JRFL}和HIV-1_{SF162}反式激活LTR$_{(-80/+66)}$启动子产生了Tat，以及Cas9在这些细胞中的产生(图6A)。再者，荧光素酶试验的结果显示，这些细胞内荧光素酶活性显着降低，再次证实通过Tat激活的Cas9对于关闭所整合的HIV-1荧光素酶基因的有效性，而Tat是通过HIV-1_{JRFL}或HIV-1_{SF162}感染时产生的。将Cas9引入被感染细胞的动作显示于图6B中。蛋白质印迹法的结果显示，当使用HIV-1_{RFL}和HIV-1_{SF162}感染细胞时，截短的LTR启动子LTR$_{(-80/+66)}$被激活，导致该细胞内产生Cas9蛋白(图6C)。

[0183] 随访研究中，测试了Tat介导的激活LTR-Cas9连同gRNA A/B一起消除人T-淋巴细胞系2D10中的HIV-1基因组的能力。这些细胞荷有潜伏状态的单边HIV-1PNLA4-3的整合复本，该单边HIV-1PNLA4-3的基因组缺失了Gag基因和Pol基因的部分，且Nef基因被替换为编码绿色荧光报告蛋白的基因(GFP)。这些细胞内Tat蛋白质水平的提升和Cas9的活化(显示于图7A中)造成，在通过LV-gRNA A/B转导的细胞内的Cas9被激活时，编辑该病毒LTR(图7B，亦见图11，道次1至8)。据此，检测到了Tat存在时GFP阳性细胞数目的显着下降，说明Tat的激活消除了被裂解的启动子表达病毒DNA的能力，而反过来造成对这些细胞中GFP的阻抑。对应于gRNA位置、DNA片段的切除、以及PCR引物的DNA序列显示于图12中。

[0184] 鉴于早前的观察表明了PMA及/或TSA在刺激2D10细胞内前病毒DNA整合复本中的能力，对PMA和TSA对潜伏性感染的T细胞模型中Cas9的激活的影响进行评估。如图8A中所见，以单独或组合形式的PMA和TSA处理2D10细胞，增加了Cas9表达的水平。平行实验中，进行用于检测LTR DNA的PCR分析，该PCR分析显示了病毒DNA切除水平的明显增加(图8B)，这被205bp DNA片段的出现所证明(见图11，道次9至14)。通过使用蛋白质印迹法测量细胞内GFP的水平而检查病毒的激活，或通过荧光显微镜对标识着病毒激活的绿色荧光细胞进行定量(图8C)，显示了病毒基因表达水平的显着下降。因此，通过由静默前病毒被再次激活时产生的Tat或直接通过产生Cas9的PMA和TSA对最小病毒启动子(-88/+60)的激活，可能对于含有gRNA A和gRNA B的细胞内的潜伏性整合病毒基因组的表达具有负面影响。

[0185] 探讨

[0186] 自1985年被发现以来，由于HIV-1的Tat蛋白在病毒基因组以转录水平表达中及其对未感染细胞的致病性冲击中的关键角色，该蛋白已经收到重大关注。机理上，Tat与位于转录起始位点下游的RNA序列(核苷酸+1至+59)即所谓的转录响应区域或TAR相关联。Tat与TAR的关联触发一系列分子事件和生化事件，导致接近转录起始位点(核苷酸1)的转录起始复合物的启动前和启动。这一复合物包括一系列具有令该复合物的组分磷酸化或乙酰化的能力的细胞蛋白，包括pTEF和RNA聚合酶II，因此促进RNA的转录延伸。此外，Tat与多种转录因子的相互作用可影响其它病毒和细胞基因的转录；其中，该转录因子包括NF-κB、p300/CBP和GCN5，且所有因子均促成见于HIV-1阳性AIDS患者体内的疾病谱。一旦来自再次激活的病毒启动子的转录导致了病毒转录和Tat产生的初始循环，则储存库内潜伏性病毒即进行生产性复制，而Tat也在此过程中扮演重要角色。Tat在HIV-1复制和AIDS发病机理中独一无二的重要性，提供了用作病毒输送以及疫苗研发的潜在靶标的强有力的基本原理。事实上，若干潜在的抑制剂，一些具有干扰Tat-TAR相互作用的能力，而其它具有防止Tat与其细胞性伙伴通讯的能力，已经显示了多种程度的影响HIV-1复制的效能。

[0187] 本研究中使用的策略是采用Tat来切除病毒基因组的节段并永久性地废除具有生

产性或潜伏性HIV-1的细胞内的HIV-1基因转录和复制。这里，设计了一种用于HIV-1的由Tat触发的自杀路径，该路径包括使用CRISPR/Cas9技术编辑病毒基因组(在图9中说明)。根据这一途径，Tat在细胞内的产生，除了刺激其具有全长度5'-LTR序列的自身启动子外，还通过相同机理借由跨越该富GC、TATA盒和TAR(-80至+66)区域的截短的最小启动子序列而加强Cas9的表达。借由切除基因节段，Cas9的产生极其与gRNA的联合可永久性地根除细胞内的HIV-1，其中，该gRNA设计为以(-80至+66)外的诱发全长度病毒启动子内的InDel突变的LTR DNA序列为靶标。除了所预期的代表全长度LTR序列的417bp DNA片段之外，LTR DNA的短范围扩增结果还显示了仅存在于表达Tat的细胞内的体积为227bp的DNA片段。通过将残留的5'-LTR与通过Cas9/gRNA A裂解5'-LTR或3'-LTR后而保留的3'-LTR结合，生成该227bp DNA片段。也可能是，来自通过Cas9/gRNA裂解后的5'-LTR的保留DNA片段与来自3'-LTR的DNA片段连接，创建了新的用于基因扩增和小体积(227bp)扩增子呈现的模板。使用以LTR加上Gag基因内一区域为靶标的多工gRNA(gRNA A)，且具有gRNA A与gRNA Gag之间的DNA片段被移除的预期。

[0188] 近年来，由于该CRISPR/Cas9基因编辑策略在以高精度和高效率编辑基因组方面的非凡能力及其实施的简单性和灵活性，该策略已经在生物医学研究中受到关注。但是，若干方面需要密切关注。例如，设计最具特异性且有效的gRNA以避免脱靶效应是重要的。通过全基因组的超深度测序和多种其它测试，证实本文所采用的策略可用于最大化特异性且避免脱靶编辑。第二课题涉及Cas9的受控表达，以避免可能非特异性地造成宿主基因组长期受损及/或诱导免疫响应的蛋白质的非必要存在。本策略被发展用于Cas仅在HIV-1Tat存在下的有条件的表达，其提供一种用于在HIV-1呈上升趋势时激活和静默用于根除该病毒的基因编辑的新颖途径。

[0189] 本文中引用的每一专利、专利申请案和公开案的内容通过引用而以其整体并入本文。该领域技术人员将轻易了解，本发明很好地适应了实践并获得所提及的结果和优点及该实践所固有的结果和优点。尽管本发明是参考具体实施例而揭露的，但很明显，该领域技术人员可设计本发明的其它具體实施例和变更而不悖离本发明实践中所使用的真实精神和范畴。倾向于将所附权利要求书解释为包括全部具體实施例及等效变更。

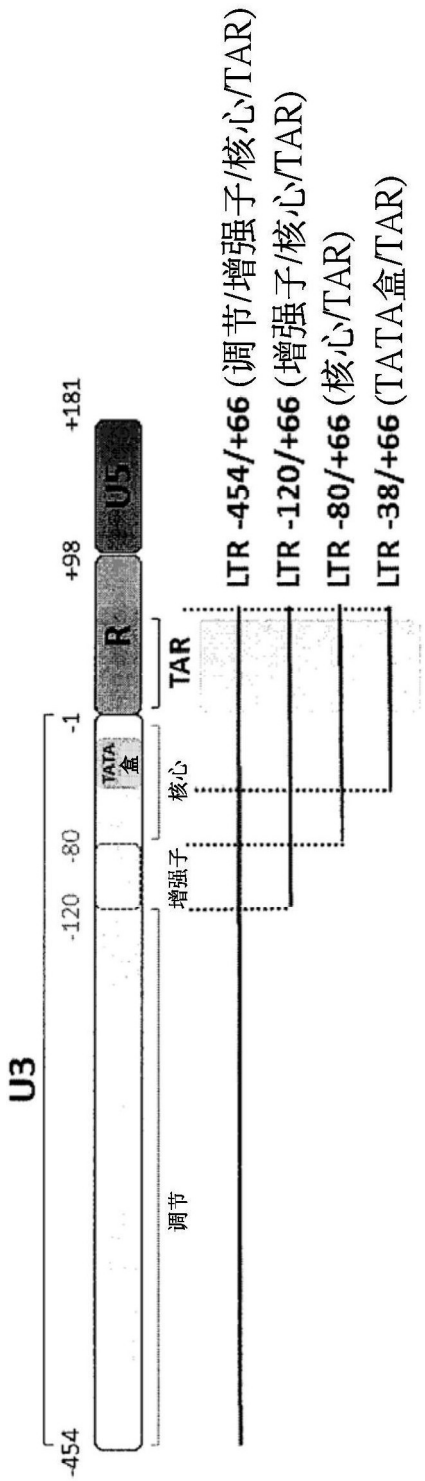

LTR -454/+66 (调节/增强子/核心/TAR)
LTR -120/+66 (增强子/核心/TAR)
LTR -80/+66 (核心/TAR)
LTR -38/+66 (TATA盒/TAR)
U3
R
U5
TAR
TATA盒
核心
增强子
调节
-454
-120
-80
-1
+98
+181

图1A

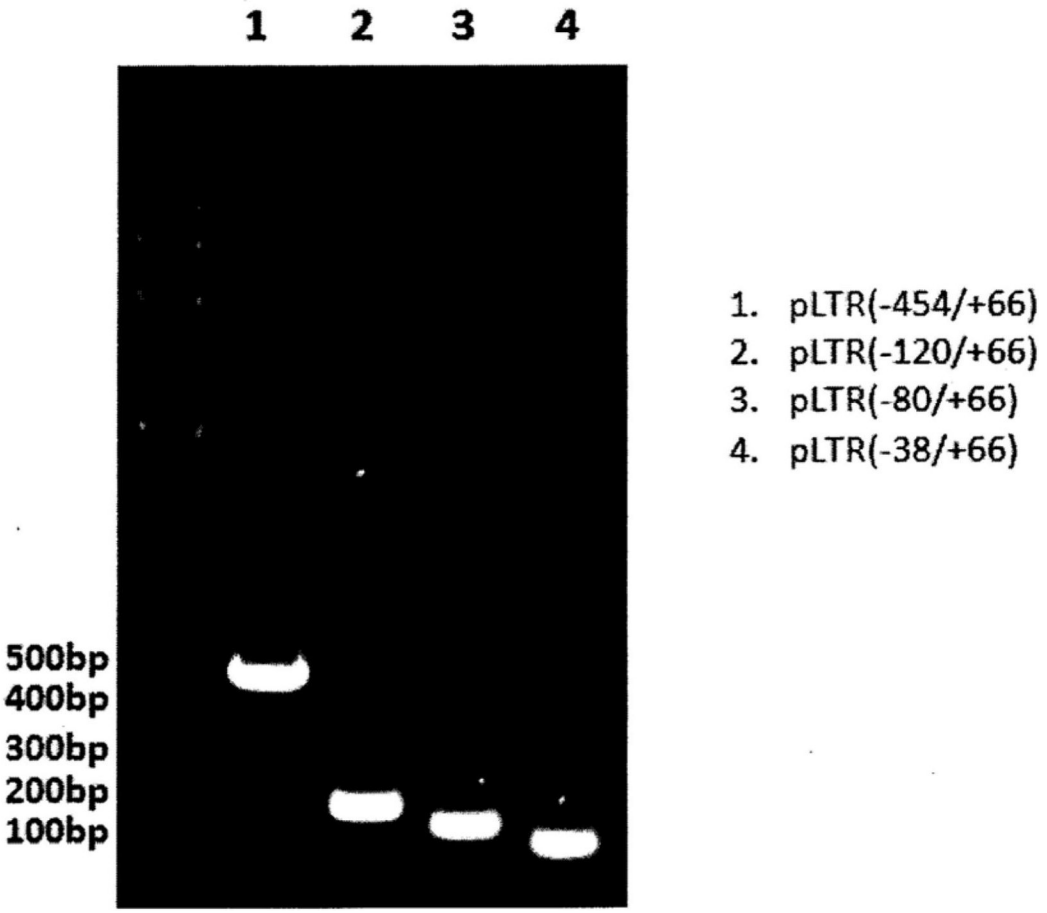

1
2
3
4
500bp
400bp
300bp
200bp
100bp
1. pLTR(-454/+66)
2. pLTR(-120/+66)
3. pLTR(-80/+66)
4. pLTR(-38/+66)

图1B

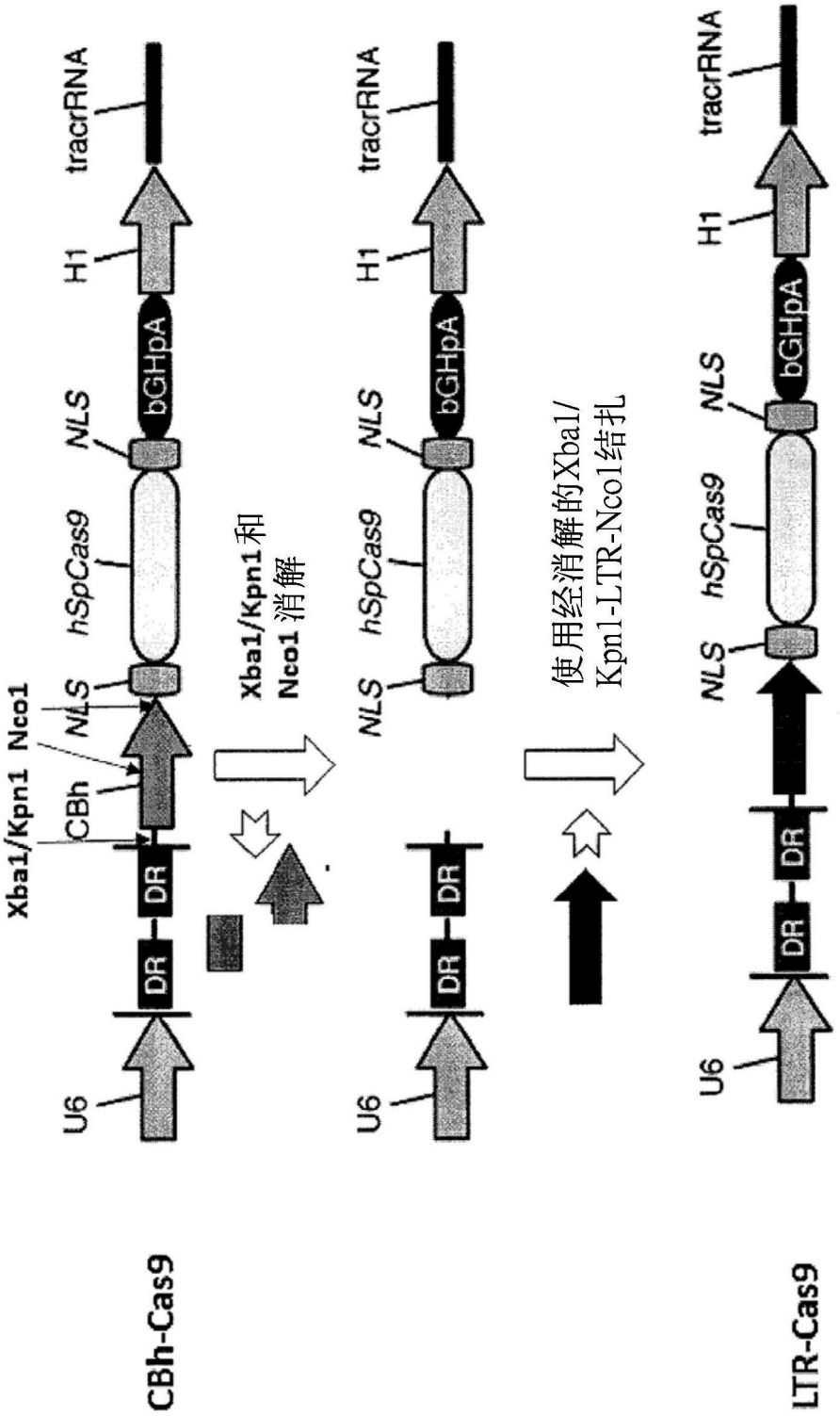
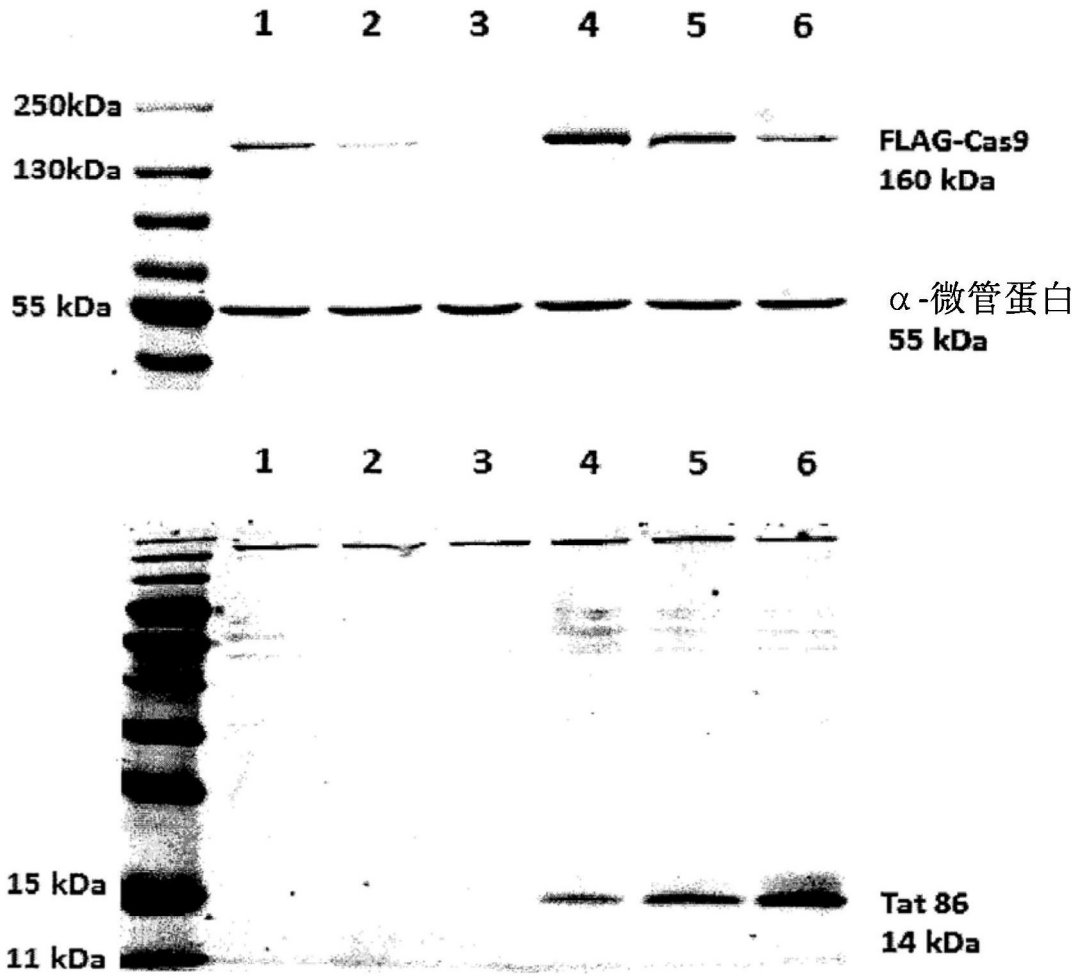

U6
DR
DR
CBh
Xba1/Kpn1
Nco1
NLS
hSpCas9
NLS
bGHpA
H1
tracrRNA
CBh-Cas9
Xba1/Kpn1和
Nco1消解
使用经消解的Xba1/
Kpn1-LTR-Nco1结扎
LTR-Cas9

图2

蛋白质印迹法U87 MG WCE 50 μg/孔



1. pLTR(-454/+66)-Cas9 250ng, pCMV 1000ng
2. pLTR(-454/+66)-Cas9 50ng, pCMV 1200ng
3. pLTR(-454/+66)-Cas9 10ng, pCMV 1240ng
4. pLTR(-454/+66)-Cas9 250ng, pCMV 750ng, pCMV-Tat86 250ng
5. pLTR(-454/+66)-Cas9 50ng, pCMV 950 ng, pCMV-Tat86 250ng
6. pLTR(-454/+66)-Cas9 10ng, pCMV 990ng, pCMV-Tat86 250ng

图3A

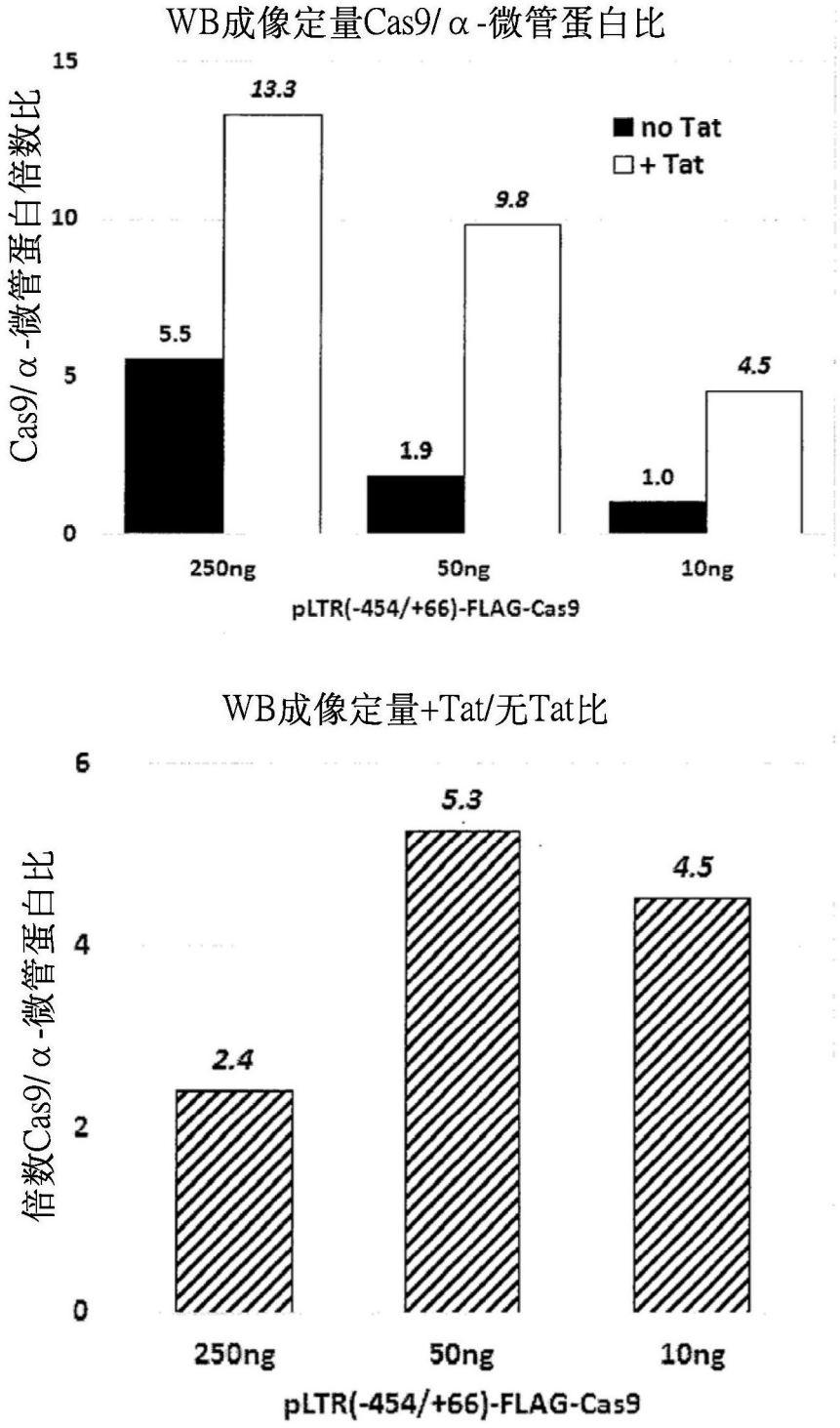
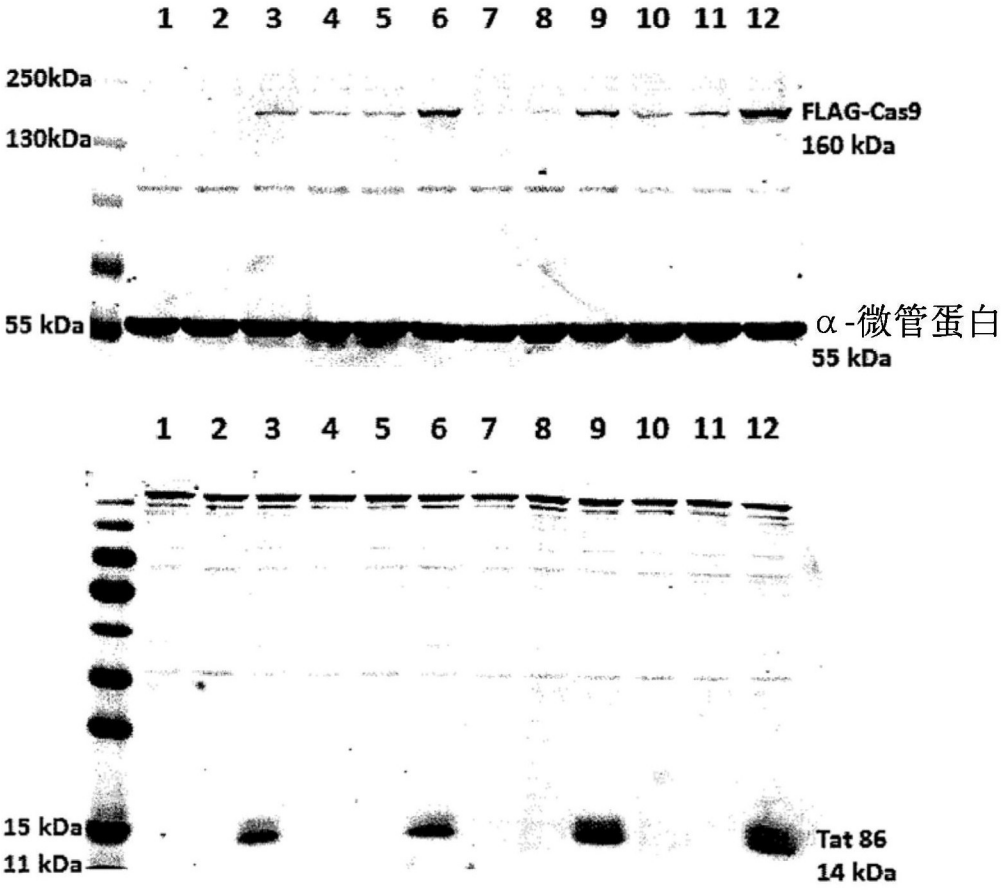

WB成像定量Cas9/α-微管蛋白比
Cas9/α-微管蛋白倍数比
15
10
5
0
13.3
9.8
5.5
4.5
1.9
1.0
no Tat
+ Tat
250ng
50ng
10ng
pLTR(-454/+66)-FLAG-Cas9
WB成像定量+Tat/无Tat比
倍数Cas9/α-微管蛋白比
6
4
2
0
5.3
4.5
2.4
250ng
50ng
10ng
pLTR(-454/+66)-FLAG-Cas9

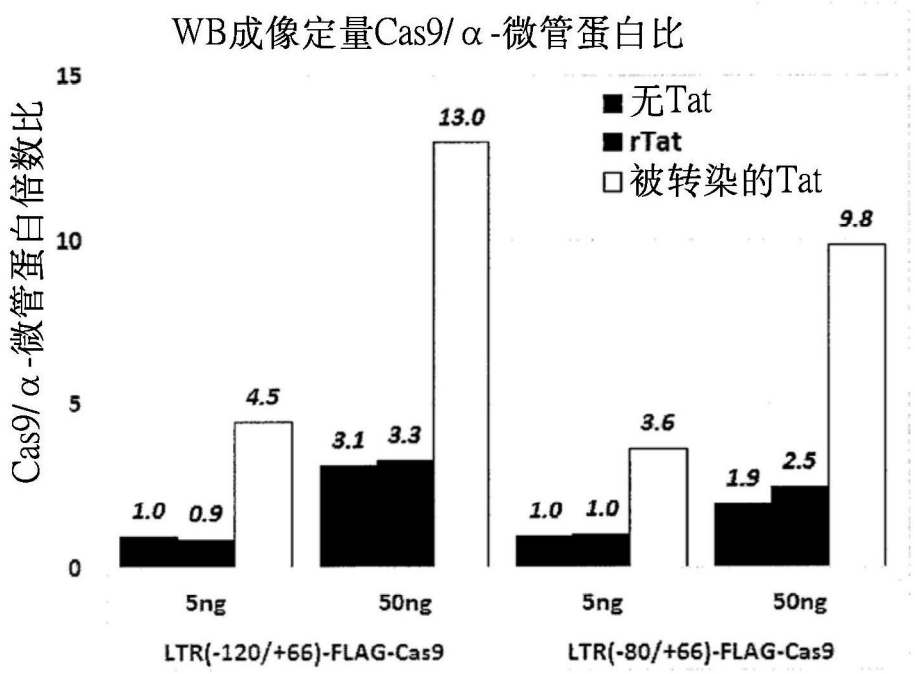
图3B

蛋白质印迹法U87 MG WCE 30 μg/孔



1. pLTR(-120/+66)-Cas9 5ng, pCMV 1245ng
2. pLTR(-120/+66)-Cas9 5ng, pCMV 1245ng, +rTat protein 2.5 ug/ml
3. pLTR(-120/+66)-Cas9 5ng, pCMV 995ng, pCMV-Tat86 250ng
4. pLTR(-120/+66)-Cas9 50ng, pCMV 1200ng,
5. pLTR(-120/+66)-Cas9 50ng, pCMV 1200ng, +rTat protein 2.5 ug/ml
6. pLTR(-120/+66)-Cas9 50ng, pCMV 950ng, pCMV-Tat86 250ng
7. pLTR(-80/+66)-Cas9 5ng, pCMV 1245ng
8. pLTR(-80/+66)-Cas9 5ng, pCMV 1245ng, +rTat protein 2.5 ug/ml
9. pLTR(-80/+66)-Cas9 5ng, pCMV 995ng, pCMV-Tat86 250ng
10. pLTR(-80/+66)-Cas9 50ng, pCMV 1200ng,
11. pLTR(-80/+66)-Cas9 50ng, pCMV 1200ng, +rTat protein 2.5 ug/ml
12. pLTR(-80/+66)-Cas9 50ng, pCMV 950ng, pCMV-Tat86 250ng

图4A


WB成像定量Cas9/α-微管蛋白比
Cas9/α-微管蛋白倍数比
15
10
5
0
无Tat
rTat
被转染的Tat
1.0
0.9
4.5
3.1
3.3
13.0
1.0
1.0
3.6
1.9
2.5
9.8
5ng
50ng
5ng
50ng
LTR(-120/+66)-FLAG-Cas9
LTR(-80/+66)-FLAG-Cas9

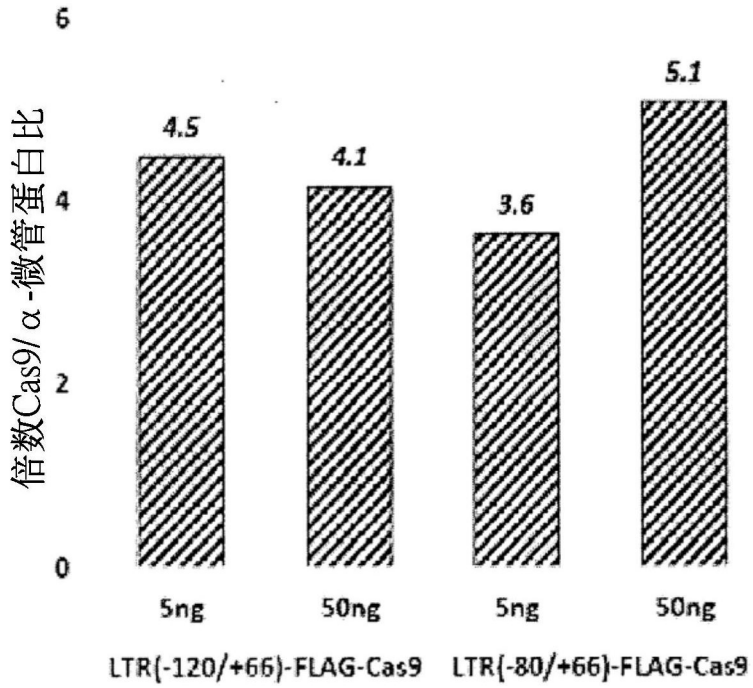

WB成像定量+Tat (被转染)/无Tat比
倍数Cas9/α-微管蛋白比
6
4
2
0
4.5
4.1
3.6
5.1
5ng
50ng
5ng
50ng
LTR(-120/+66)-FLAG-Cas9
LTR(-80/+66)-FLAG-Cas9

图4B

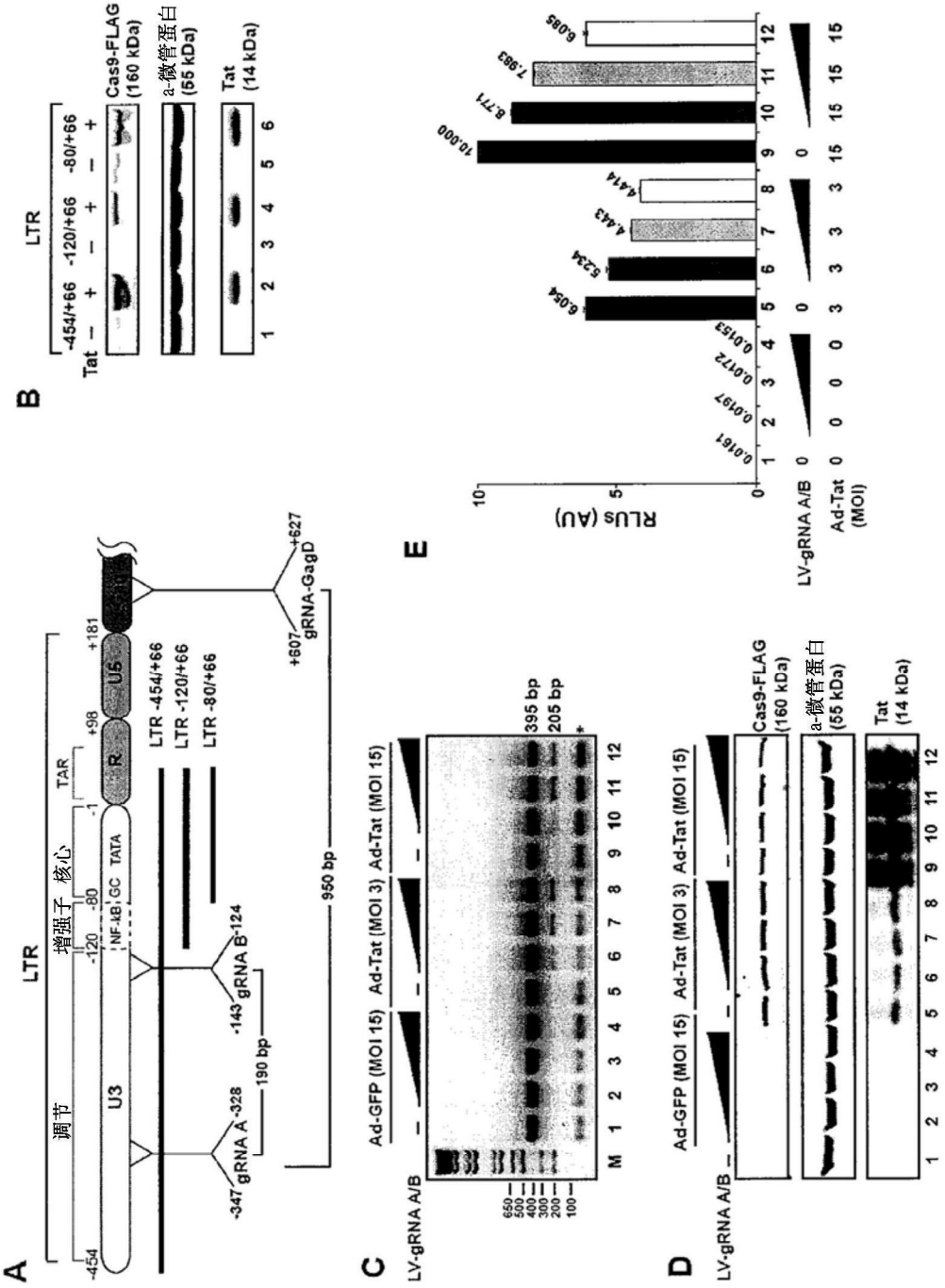

A
LTR
调节
增强子
核心
TAR
U3
R
U5
TATA
NF-κB
GC
-454
-120
-80
-1
+98
+181
LTR -454/+66
LTR -120/+66
LTR -80/+66
gRNA A -328
gRNA B -124
-347
-143
190 bp
950 bp
+607
+627
gRNA-GagD
B
LTR
-454/+66
-120/+66
-80/+66
Tat
Cas9-FLAG (160 kDa)
a-微管蛋白 (55 kDa)
Tat (14 kDa)
C
LV-gRNA A/B
Ad-GFP (MOI 15)
Ad-Tat (MOI 3)
Ad-Tat (MOI 15)
395 bp
205 bp
M
D
LV-gRNA A/B
Ad-GFP (MOI 15)
Ad-Tat (MOI 3)
Ad-Tat (MOI 15)
Cas9-FLAG (160 kDa)
a-微管蛋白 (55 kDa)
Tat (14 kDa)
E
RLUs (AU)
6.054
5.234
4.443
4.414
10.000
8.771
7.983
6.085
0.0161
0.0197
0.0172
0.0153
LV-gRNA A/B
Ad-Tat (MOI)

图5

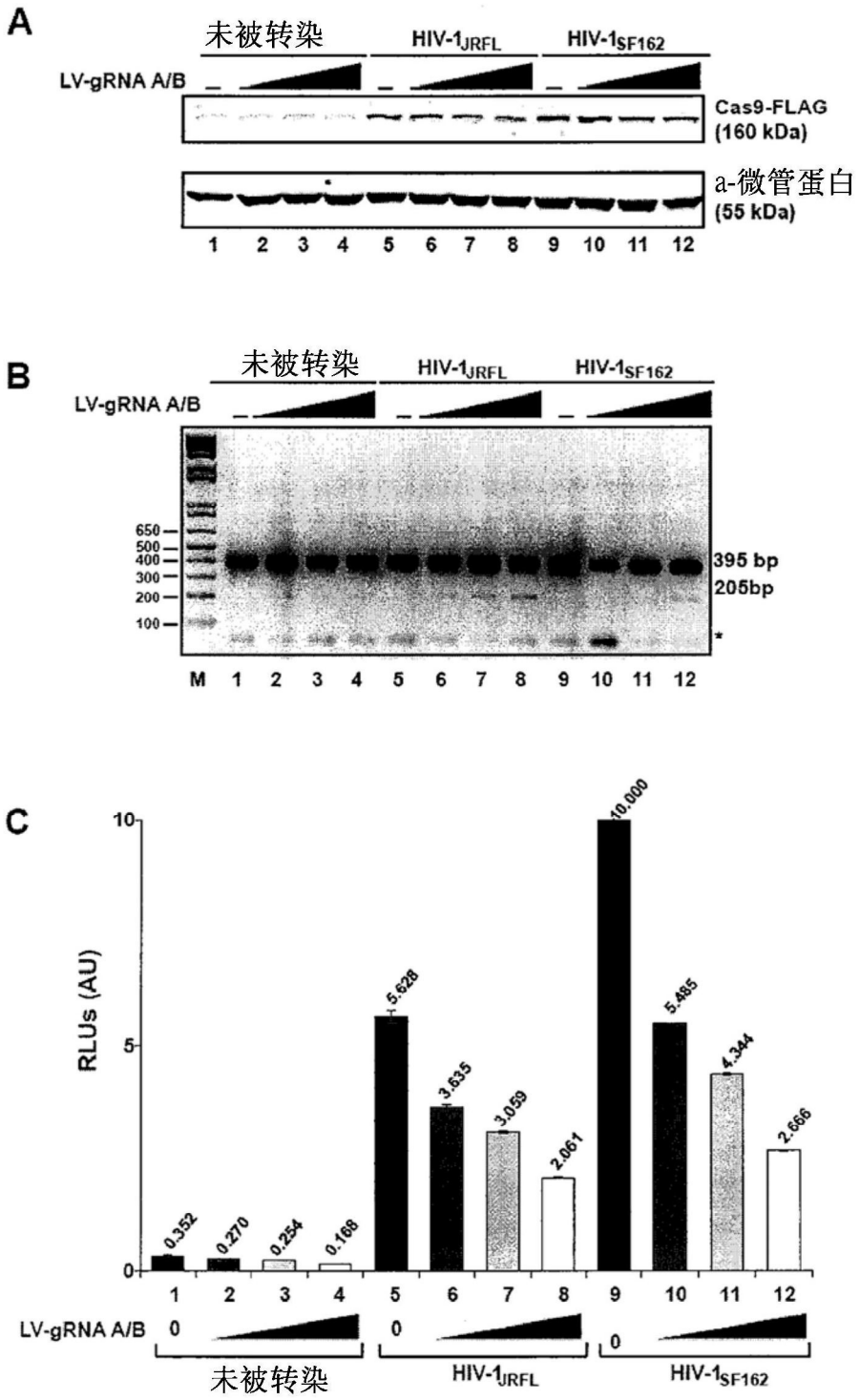

A
未被转染
HIV-1JRFL
HIV-1SF162
LV-gRNA A/B
Cas9-FLAG
(160 kDa)
a-微管蛋白
(55 kDa)
1 2 3 4 5 6 7 8 9 10 11 12
B
未被转染
HIV-1JRFL
HIV-1SF162
LV-gRNA A/B
650
500
400
300
200
100
395 bp
205bp
*
M 1 2 3 4 5 6 7 8 9 10 11 12
C
RLUs (AU)
10
5
0
0.352
0.270
0.254
0.168
5.628
3.635
3.059
2.061
10.000
5.485
4.344
2.666
1 2 3 4 5 6 7 8 9 10 11 12
LV-gRNA A/B
0
0
0
未被转染
HIV-1JRFL
HIV-1SF162

图6

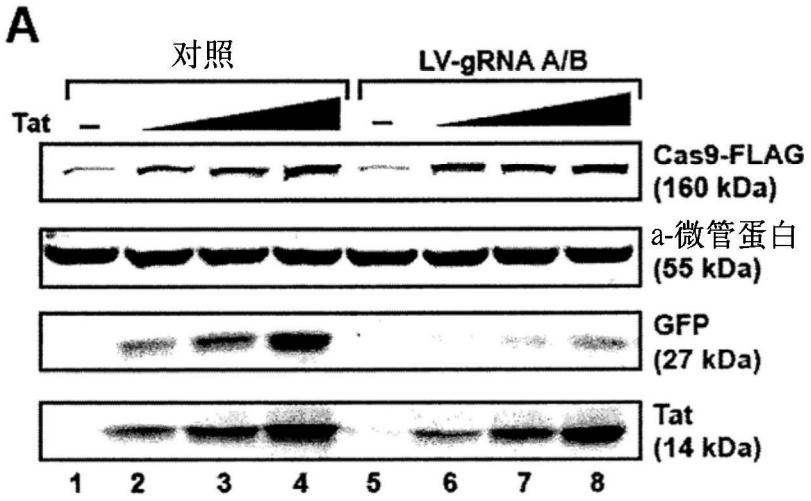

A
对照
LV-gRNA A/B
Tat
–
–
Cas9-FLAG
(160 kDa)
a-微管蛋白
(55 kDa)
GFP
(27 kDa)
Tat
(14 kDa)
1 2 3 4 5 6 7 8

图7A

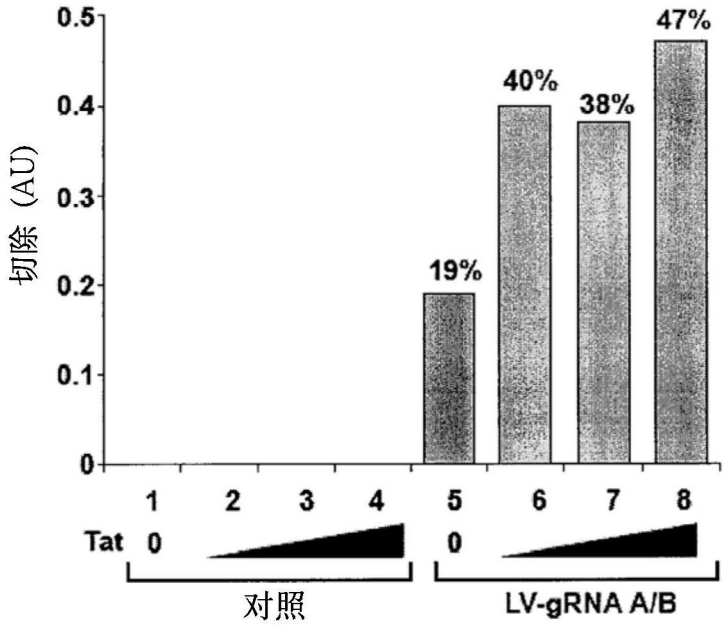

0.5
0.4
0.3
0.2
0.1
0
切除 (AU)
19%
40%
38%
47%
1 2 3 4 5 6 7 8
Tat 0 0
对照
LV-gRNA A/B

图7B

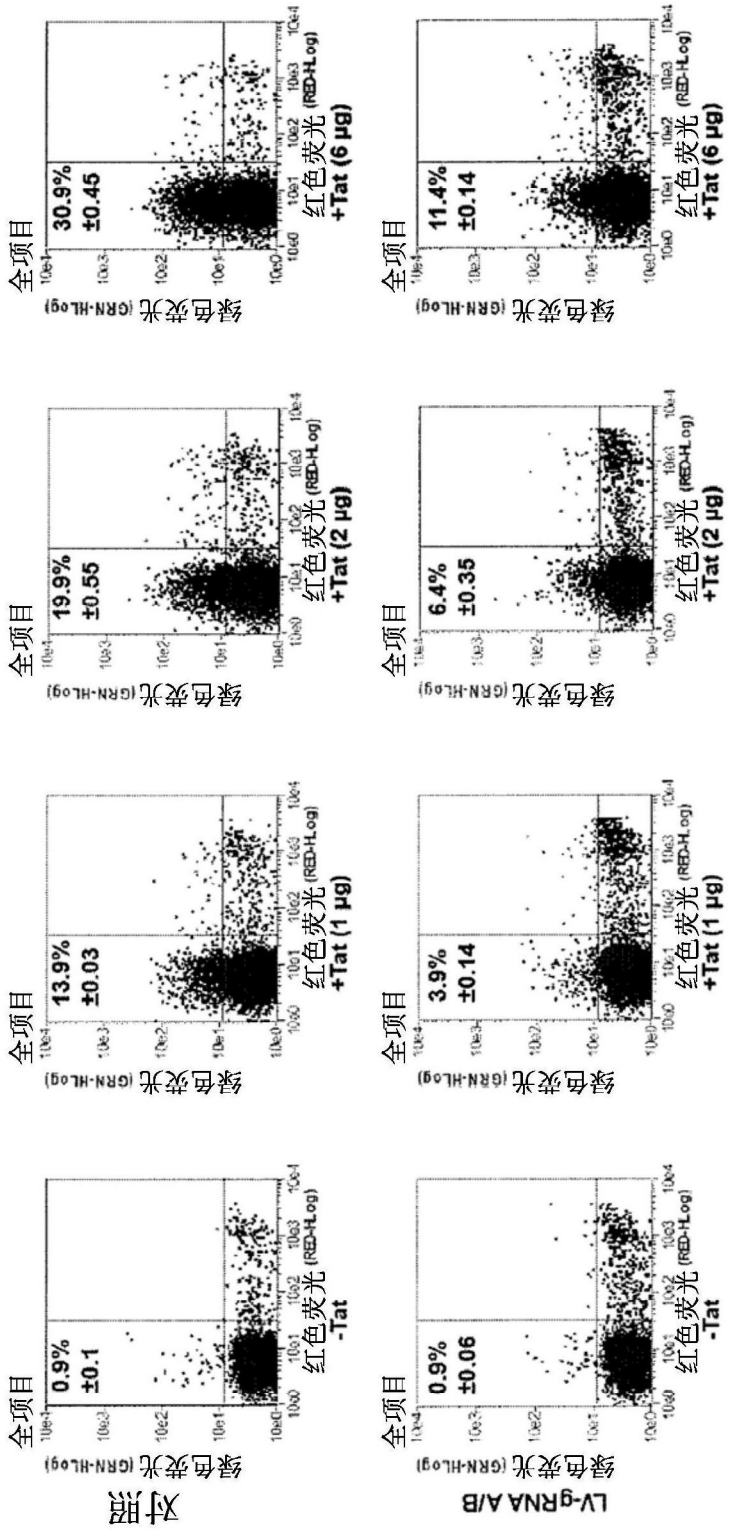

对照
LV-gRNA A/B
全项目
绿色荧光 (GRN-HLog)
红色荧光 (RED-HLog)
-Tat
+Tat (1 μg)
+Tat (2 μg)
+Tat (6 μg)
0.9%
±0.1
0.9%
±0.06
13.9%
±0.03
3.9%
±0.14
19.9%
±0.55
6.4%
±0.35
30.9%
±0.45
11.4%
±0.14

图7C

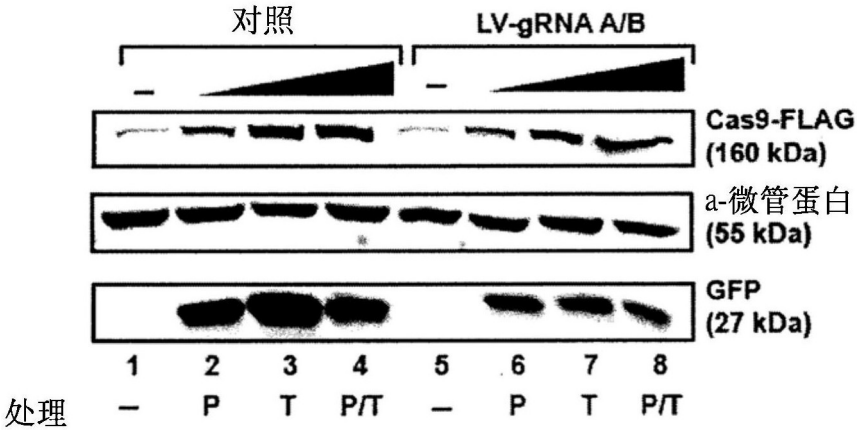

对照
LV-gRNA A/B
Cas9-FLAG
(160 kDa)
a-微管蛋白
(55 kDa)
GFP
(27 kDa)
1 2 3 4 5 6 7 8
处理 — P T P/T — P T P/T

图8A

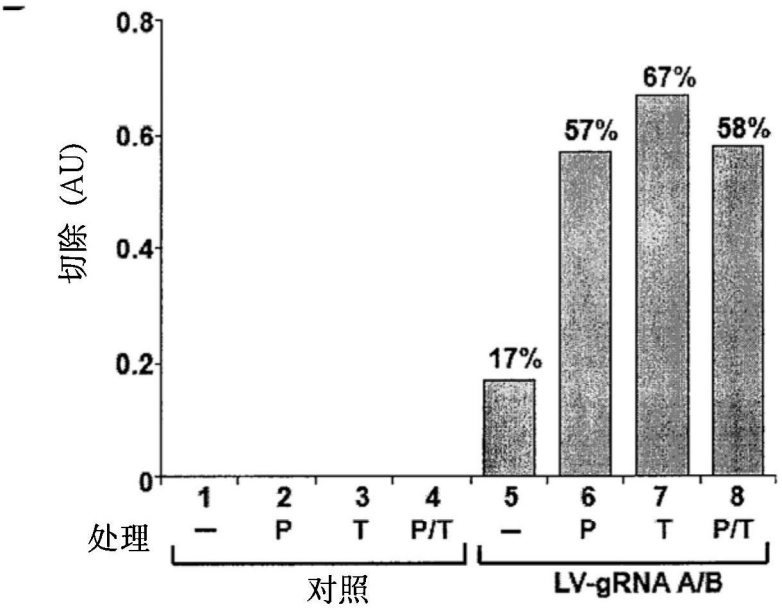

0.8
0.6
0.4
0.2
0
切除 (AU)
17%
57%
67%
58%
1 2 3 4 5 6 7 8
处理 — P T P/T — P T P/T
对照
LV-gRNA A/B

图8B

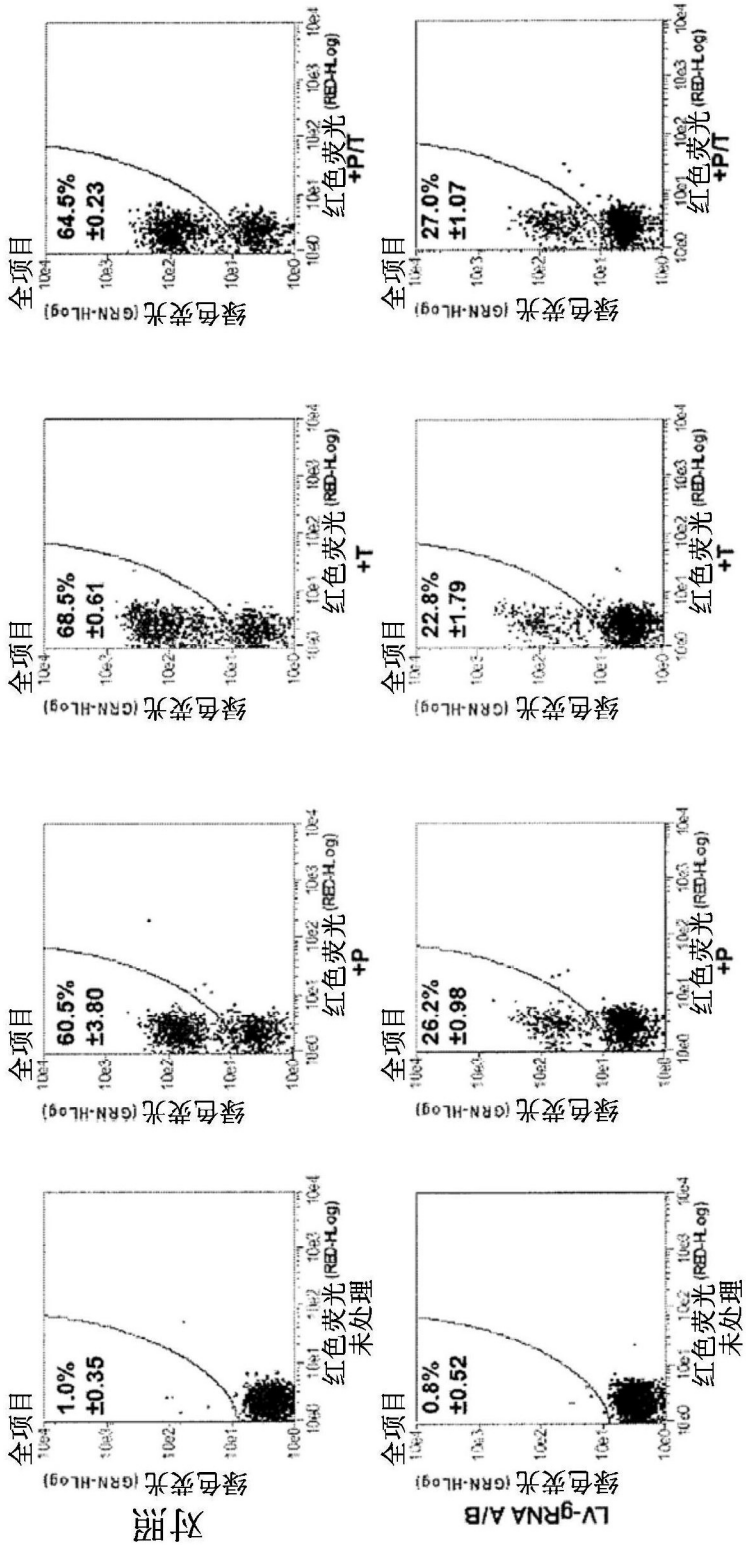

对照
LV-gRNA A/B
未处理
+P
+T
+P/T
全项目
1.0%
±0.35
60.5%
±3.80
68.5%
±0.61
64.5%
±0.23
0.8%
±0.52
26.2%
±0.98
22.8%
±1.79
27.0%
±1.07
绿色荧光 (GRN-HLog)
红色荧光 (RED-HLog)

图8C

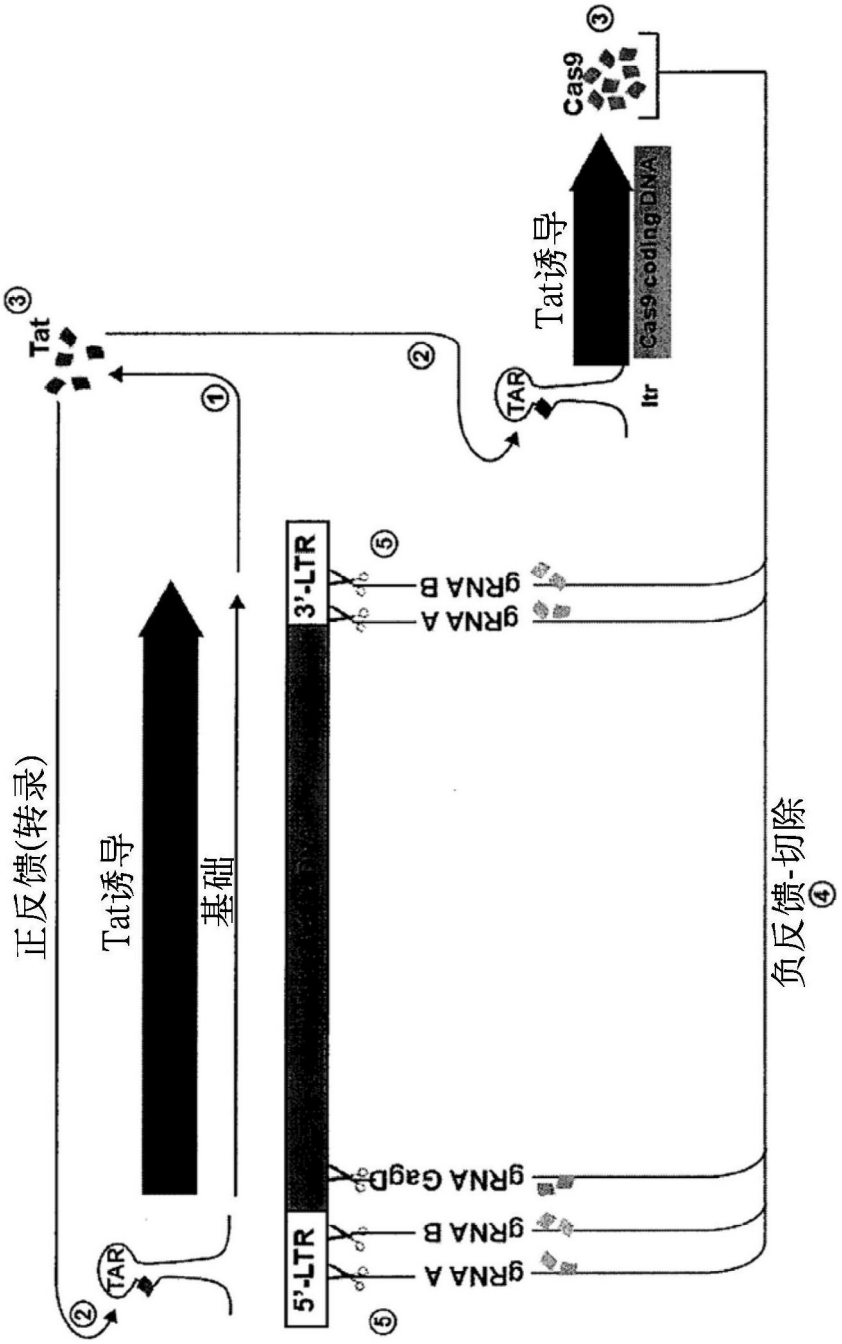

Cas9
Tat诱导
Cas9 coding DNA
Itr
TAR
Tat
3'-LTR
gRNA B
gRNA A
5'-LTR
gRNA GagD
gRNA B
gRNA A
正反馈(转录)
Tat诱导
基础
负反馈-切除

图9

TZMbl (SEQ ID NO: 6)

F [-413/-391](T361)
TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTA
LTR A [-347/-328]PAM
CTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAGAGCAAGTAGAAGAGG
CCAATGAAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGC
PAM **LTR B [-143/-124]**
ATTTCGTCACATGGCCCGAGAGCTGCATCGGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAG
R [-41/-19](T363)
GGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT
Transcription start[+1]
GCTACATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAACCCACTGCTTAAGCCTCA
ATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTG

PCR产物全长度 **LTR -413/-19: 395bp (SEQ ID NO: 7)**

GATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGT
ACCAGTTGAACCAGAGCAAGTAGAAGAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAG
TGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTGCATCGGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCT
GGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAGC

+SpCas9/gRNA LTR A+LTR B (SEQ ID NO: 8):

GATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGAC

经编辑的片段: **190bp (SEQ ID NO: 9)**

↓CTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAGAGCAAGTAGAAGAGGAGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATG
GAGGACCCGGAGGGAGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTGCATCGGGAG↓
TACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATG
CTACATATAAGCAGC

PCR产物截短的 **LTR -413/-19: 205bp (SEQ ID NO: 10)**

GATCTGTGGATCTACCACACACAAGGCTACTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACTACTACAAAGACTGCTGACATCGA
GCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAGC

图10

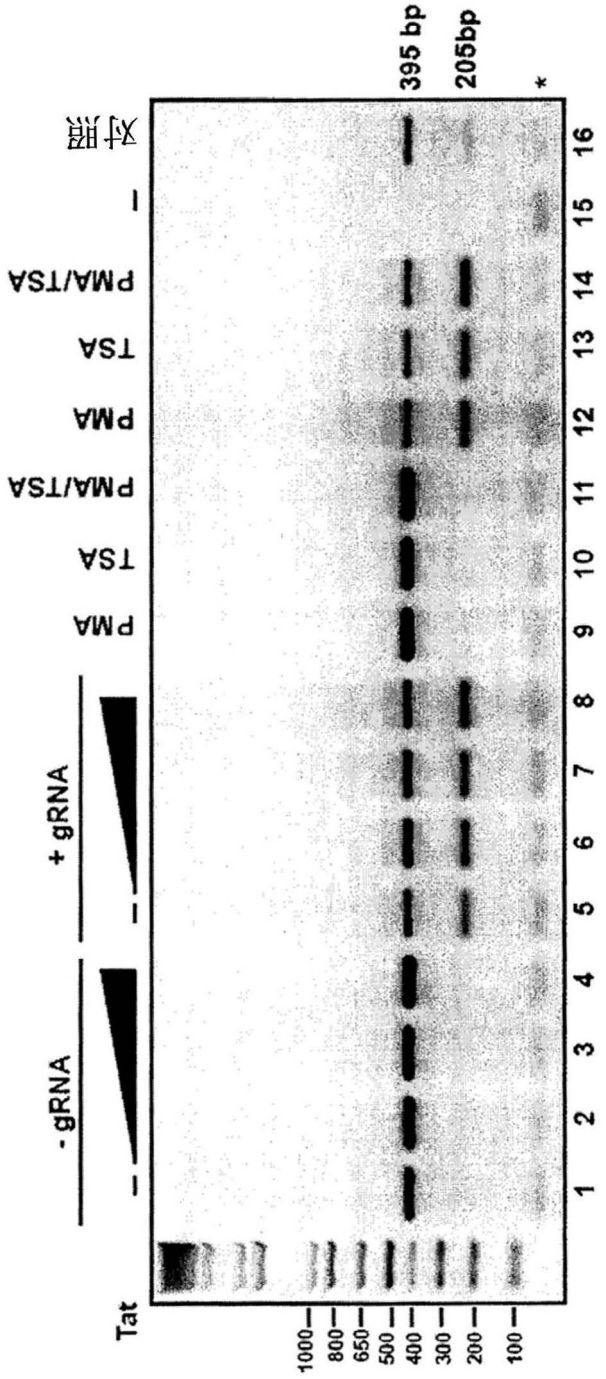

395 bp
205bp
*
对照
-
PMA/TSA
TSA
PMA
PMA/TSA
TSA
PMA
+ gRNA
- gRNA
Tat
1000
800
650
500
400
300
200
100
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16

图11

Jurkat 2D10 (SEQ ID NO: 11):

TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTA

F [-375/-352] **LTR A [-347/-328]** PAM

CTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAAC

CAGAGCAAGTAGAAGAGGCCAATGAAGGAGAAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAA

GTATTAGTGTGGAAGTTTGACAGCCTCCTAGC

PAM **LTR B [-143/-124]**

ATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAG

GGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT

Transcription start[+1] R [+19/+43]

GCTACATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAA

CCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTG

PCR产物全长度 **LTR -375/+43: 418bp (SEQ ID NO: 12)**

TTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAGAGCAAG

TAGAAGAGGCCAATGAAGGAGAAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAGTG

TGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTA

CAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAG

CTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG

+SpCas9/gRNA LTR A+LTR B (SEQ ID NO: 13):

TTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGAC

经编辑的片段: **190bp (SEQ ID NO: 14)**

↓CTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAACCAGAGCAAGTAGAAGAGGAGAGAACAACAGCTTGTTACACCCT

ATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAAGTATTAGTGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGA

GCTGCATCCGGAG↓

图12

PCR产物截短的 LTR -375/-43: 228bp (SEQ ID NO: 15):

TTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACTACTACAAAGACTGCTGACATCGAGCTTTCTACAAGGGACTTTC
CGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATGCTACATATAAGCAGCTGCTTTTTGCC
TGTACTGGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGG

SpCas9 versus SaCas9 targets in HIV-1 5'LTR and Gag:

AGGGATCAGATATCCACTGACCTTTGGATGGTGC **LTR A** [-347/-328] PAM **SpCas9 (SEQ ID NO: 16)**
AGGGATCAGATATCCACTGACCTTTGGATGGTGC **LTR 1** [-349/-329] PAM **SaCas9 (SEQ ID NO: 17)**

AAAGGATAGATGTAAAAGACACCAAGGAAGCCTT **Gag D** [+607/+627] PAM **SpCas9 (SEQ ID NO: 18)**
AAAGGATAGATGTAAAAGACACCAAGGAAGCCTT **Gag D** [+607/+627] PAM **SaCas9 (SEQ ID NO: 19)**

图12(接续)

SpCas9 (PAM: NGG) 靶标序列 (SEQ ID NO: 20):

Sequence:

F [-454/-431] F [-413/-391](T361)

TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTA

F [-375/-352] **LTR A [-347/-328]** PAM

CTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAAC

CAGAGCAAGTAGAAGAGGCCAATGAAGGAGAAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAA

GTATTAGTGTGGAAGTTTGACAGCCTCCTAGC

PAM **LTR B [-143/-124]**

ATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCGAGCTTTCTACAAG

R [-41/-19](T363)

GGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT

Transcription start[+1] R [+19/+43]

GCTACATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAA

CCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCC

TCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGAGGAGATCTCTC

GACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGC

Gag gene start[+336]

GGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAAAAAATTCGGTTAAGGCCAG

GGGGAAAGAAACAATATAAACTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTTTTAGAGACA

TCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAATAGC

AGTCCTCTATTGT

Gag D [+607/+627] PAM

GTGCATCAAAGGATAGATGTAAAAGACACCAAGGAAGCCTTAGATAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCA

AGCAGCAGCTGACACAGGAAACAACAGCCAGGTCAGCCAAAATTACCCTATAGTGCAGAACCTCCAGGGGCAAATGGTACATCAGGCCA

TATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTATCA

GAAGGAGCCACCC

R [+888/+910](T458)

CACAAGATTTAAATACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGAC

图12(接续)

SpCas9 (PAM: NNGRR(T/N)) 靶标序列 (SEQ ID NO: 21):

Sequence:

PCR primer F [-413/-391](T361)

TGGAAGGGCTAATTTGGTCCCAAAAAAGACAAGAGATCCTTGATCTGTGGATCTACCACACACAAGGCTA

LTR 1 [-349/-329] PAM

CTTCCCTGATTGGCAGAACTACACACCAGGGCCAGGGATCAGATATCCACTGACCTTTGGATGGTGCTTCAAGTTAGTACCAGTTGAAC

CAGAGCAAGTAGAAGAGGCCAATGAAGGAGAACAACAGCTTGTTACACCCTATGAGCCAGCATGGGATGGAGGACCCGGAGGGAGAA

GTATTAGTGTGGAAGTTTGACAGCCTCCTAGCATTTCGTCACATGGCCCGAGAGCTGCATCCGGAGTACTACAAAGACTGCTGACATCG

AGCTTTCTACAAGGGACTTTCCGCTGGGGACTTTCCAGGGAGGTGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGAT

Transcription start[+1]

GCTACATATAAGCAGCTGCTTTTTGCCTGTACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAGGGAA

CCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCTCAAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCC

TCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCAGTGGCGCCCGAACAGGGACTTGAAAGCGAAAGTAAAGCCAGAGGAGATCTCTC

GACGCAGGACTCGGCTTGCTGAAGCGCGCACGGCAAGAGGCGAGGGGCGGCGACTGGTGAGTACGCCAAAAATTTTGACTAGC

Gag gene start[+336]

GGAGGCTAGAAGGAGAGAGATGGGTGCGAGAGCGTCGGTATTAAGCGGGGGAGAATTAGATAAATGGGAAAAAATTCGGTTAAGGCCAG

GGGGAAAGAAACAATATAAAACTAAAACATATAGTATGGGCAAGCAGGGAGCTAGAACGATTCGCAGTTAATCCTGGCCTTTTAGAGACA

TCAGAAGGCTGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAAGAACTTAGATCATTATATAATACAATAGC

AGTCCTCTATTGT

Gag D [+607/+627] PAM

GTGCATCAAAGGATAGATGTAAAGACACCAAGGAAGCCTTAGATAAGATAGAGGAAGAGCAAAACAAAAGTAAGAAAAAGGCACAGCA

AGCAGCAGCTGACACAGGAAACAACAGCCAGGTCAGCCAAAATTACCCTATAGTGCAGAACCTCCAGGGGCAAATGGTACATCAGGCCA

TATCACCTAGAACTTTAAATGCATGGGTAAAAGTAGTAGAAGAGAAGGCTTTCAGCCCAGAAGTAATACCCATGTTTTCAGCATTATCA

GAAGGAGCCACCC

PCR primer R [+888/+910] (T458)

CACAAGATTTAAATACCATGCTAAACACAGTGGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGAC

图12(接续)