



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2016-0059994
 (43) 공개일자 2016년05월27일

(51) 국제특허분류(Int. Cl.)
 C12N 15/63 (2006.01) C12N 15/113 (2010.01)
 C12N 9/22 (2006.01)
 (52) CPC특허분류
 C12N 15/63 (2013.01)
 C12N 15/113 (2013.01)
 (21) 출원번호 10-2015-0162799
 (22) 출원일자 2015년11월19일
 심사청구일자 2015년11월19일
 (30) 우선권주장
 1020140161809 2014년11월19일 대한민국(KR)

(71) 출원인
 기초과학연구원
 대전광역시 유성구 유성대로1689번길 70 (전민동,KT대덕2연구센터)
 (72) 발명자
 김진수
 서울특별시 관악구 신림동 관악로 1길 교수아파트 1221동 205호
 구대영
 서울특별시 관악구 관악로 1 서울대학교 유전공학연구소 105동 411호
 (74) 대리인
 손민

전체 청구항 수 : 총 25 항

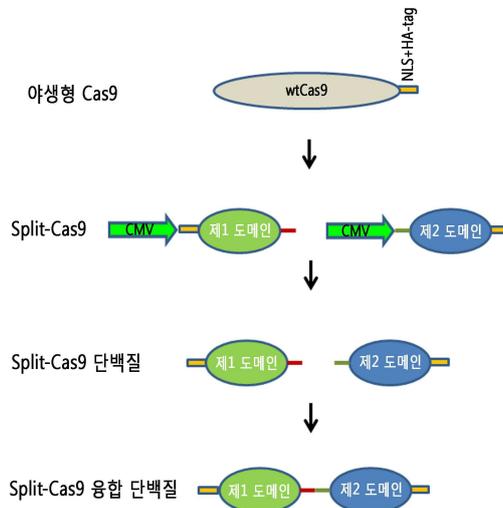
(54) 발명의 명칭 **두 개의 벡터로부터 발현된 Cas9 단백질을 이용한 유전자 발현 조절 방법**

(57) 요약

본 발명은 Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 각각 세포 내에 도입하는 단계를 포함하는, 유전자 조절 방법, 상기 재조합 벡터를 포함하는 조성물, 유전자 발현 조절용 키트, 및 Cas9 단백질의 세포 내 제조 방법에 관한 것이다.

또한, 본 발명은 상기 제1 도메인을 패키징하는 바이러스 벡터 및 제2 도메인을 패키징하는 바이러스 벡터가 도입된 형질전환 세포 및 이로부터 생산된 바이러스를 포함하는 조성물에 관한 것이다.

대표도 - 도1



(52) CPC특허분류
C12N 9/22 (2013.01)

명세서

청구범위

청구항 1

Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 각각 세포 내에 도입하는 단계를 포함하는, 유전자 발현 조절 방법으로서, 상기 Cas9 단백질은 상기 제1 도메인 및 제2 도메인으로 이루어지는 것인, 방법.

청구항 2

제1항에 있어서, 상기 유전자 발현 조절은 유전체 교정, 유전자 발현 증가 또는 유전자 발현 감소인 것인, 방법.

청구항 3

제2항에 있어서, 상기 유전체 교정은 유전자 낙인 (knock-in) 또는 낙아웃 (knock-out)인 것인, 방법.

청구항 4

제2항에 있어서, 상기 유전체 교정은 유전체 상에 표적화된 돌연변이를 도입하거나, 유전체 상의 DNA를 결실, 중복, 역위, 교체 또는 재배열시키는 것인, 방법.

청구항 5

제1항에 있어서, 상기 Cas9 단백질은 야생형 Cas9, 불활성화된 Cas9 (dCas9), 또는 Cas9 니케이즈 (nickase)인 것인, 방법.

청구항 6

제5항에 있어서, 상기 불활성화된 Cas9은 dCas9에 FokI 뉴클레아제 도메인을 연결한 RFN (RNA-guided FokI Nuclease), 또는 dCas9에 전사활성인자 (transcription activator) 또는 억제자 도메인 (repressor domain)을 연결한 것인, 방법.

청구항 7

제5항에 있어서, 상기 Cas9 니케이즈는 D10A Cas9 또는 H840A Cas9인 것인, 방법.

청구항 8

제1항에 있어서, 상기 Cas9 단백질은 스트렙토코커스 피요제네스(*Streptococcus pyogenes*), 프란시셀라 노비시다 (*Francisella novicida*), 스트렙토코커스 써모필러스 (*Streptococcus thermophilus*), 레지오넬라 뉴모필라 (*Legionella pneumophila*), 리스테리아 이노쿠아 (*Listeria innocua*), 및 스트렙토코커스 뮤탄스 (*Streptococcus mutans*)로 이루어진 군에서 선택되는 어느 하나의 유래인 것인, 방법.

청구항 9

제1항에 있어서, 상기 재조합 벡터는 플라스미드 벡터, 코즈미드 벡터, 또는 바이러스 벡터인 것인, 방법.

청구항 10

제9항에 있어서, 상기 바이러스 벡터는 레트로바이러스 (retrovirus) 벡터, 아데노바이러스 (adenovirus) 벡터, 아데노부속바이러스 (adeno-associated virus) 벡터, 및 헤르페스 심플렉스 바이러스 (herpes simplex virus) 벡터로 이루어진 군에서 선택되는 것인, 방법.

청구항 11

제1항에 있어서, 상기 도입된 각각의 벡터로부터 발현된 제1 도메인 및 제2 도메인이 융합되어, Cas9 단백질을 형성하는 단계를 추가로 포함하는 것인, 방법.

청구항 12

제1항에 있어서, 상기 제1 도메인 및 제2 도메인은 각각 400 bp (base pair) 내지 3.7 kbp (kilo base pair) 크기의 뉴클레오티드로 코딩되는 것인, 방법.

청구항 13

제1항에 있어서, 상기 제1 도메인 및 제2 도메인은 각각 NLS (nuclear localization signal), HA-태그, 스플라이싱 도너 (splicing donor) 서열, 스플라이싱 억셉터 (splicing acceptor) 서열 또는 이들의 조합을 추가로 포함하는 것인, 방법.

청구항 14

제1항에 있어서, 상기 제1 도메인은 서열번호 3의 뉴클레오티드로 코딩되는 것이고, 제2 도메인은 서열번호 5의 뉴클레오티드로 코딩되는 것인, 방법.

청구항 15

제1항에 있어서, 상기 세포 내에 도입하는 단계에서 서열 특이적인 가이드 RNA(guide RNA)를 추가로 도입하는 방법.

청구항 16

제15항에 있어서, 상기 각각의 벡터, 가이드 RNA의 도입은 동시, 순차 또는 역순으로 수행되는 것인 방법.

청구항 17

Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 포함하는 조성물로서,

상기 Cas9 단백질은 상기 제1 도메인 및 제2 도메인으로 이루어지는 것인, 조성물.

청구항 18

제17항에 있어서, 상기 조성물은 유전자 발현 조절용인 것인, 조성물.

청구항 19

제17항에 있어서, 상기 조성물은 서열 특이적인 가이드 RNA를 추가로 포함하는 것인, 조성물.

청구항 20

제17항의 조성물을 포함하는 유전자 발현 조절용 키트.

청구항 21

제20항에 있어서, 상기 키트는 서열 특이적인 가이드 RNA를 추가로 포함하는 것인, 키트.

청구항 22

Cas9 단백질의 N-말단을 포함하는 제1 도메인을 패키징하는 바이러스 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 패키징하는 바이러스 벡터가 도입된 형질전환 세포로서,

상기 Cas9 단백질은 상기 제1 도메인 및 제2 도메인으로 이루어지는 것인, 세포.

청구항 23

제22항의 세포의 배양액 또는 세포 용해물을 포함하는, 조성물.

청구항 24

제23항에 있어서, 상기 조성물은 유전자 발현 조절용인 것인, 조성물.

청구항 25

Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 각각 세포 내에 도입하는 단계를 포함하는, Cas9 단백질의 세포 내 제조 방법으로서,

상기 Cas9 단백질은 상기 제1 도메인 및 제2 도메인으로 이루어지는 것인, 방법.

발명의 설명

기술 분야

[0001] 본 발명은 Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 각각 세포 내에 도입하는 단계를 포함하는, 유전자 조절 방법, 상기 재조합 벡터를 포함하는 조성물, 유전자 발현 조절용 키트, 및 Cas9 단백질의 세포 내 제조 방법에 관한 것이다.

[0002] 또한, 본 발명은 상기 제1 도메인을 패키징하는 바이러스 벡터 및 제2 도메인을 패키징하는 바이러스 벡터가 도입된 형질전환 세포 및 이로부터 생산된 바이러스를 포함하는 조성물에 관한 것이다.

배경 기술

[0004] 현재 유전공학에서 널리 유용하게 쓰이고 있는 도구로서 제한효소는 현재 분자생물학 연구의 가장 중요한 도구의 하나이다. 그러나 게놈 크기의 DNA를 다루기에 유용한 제한효소로서 새로운 DNA 염기서열의 인지와 "rare cutter", 즉 9bp 이상의 DNA를 인지하고 자를 수 있는 제한효소의 필요성이 대두됨에 따라 여러 가지 다양한 시도가 있어왔다.

[0005] 그 일환으로, 세포 및 생물 내에서 내생 유전자의 변이, 타겟 유전자의 삽입, 및 염색체 재배열을 일으킬 수 있는 도구로서 메가뉴클레아제 (meganuclease), 징크-핑거 뉴클레아제 (zinc-finger nucleases, ZFNs) 및 TAL-이펙터 뉴클레아제 (TAL-effector nucleases, TALENs)와 같은 인공 뉴클레아제가 개발되었으며, 유전 공학에서 강력하고 다재다능한 도구로서 생명공학, 의학 등의 분야에서 유용히 사용될 수 있다. 최근 미생물의 면역체제로 알려진 CRISPR/Cas 시스템을 이용한 3세대 유전자 가위, RGEN (RNA-guided Engineered Nuclease)이 개발되면서 생명공학 분야의 모든 분야에서 새로운 발견과 혁신을 가져오고 있다 (Kim, H. etc., *Nat Rev Genet*, 2014, 15: 321-334).

[0007] 상기와 같은 인공 뉴클레아제는 세포 내에서 특이적인 표적 염기 서열을 인식하여 DNA 이중가닥 손상 (DNA double strand breaks, DSBs)을 일으킨다. 유발된 세포 내 DSB는 상동 재조합 (homologous recombination, HR)과 비상동 말단 접합 (nonhomologous end joining, NHEJ)으로 구별되는 세포의 두 가지 내재적 DNA 수선 기작에 의해 복구될 수 있는데, 이때 표적 특이적인 돌연변이 및 유전자 변형이 일어나게 된다. 진핵세포 및 생물체 내에서 상동의 DNA 제공자가 부재할 때는, 뉴클레아제에 의해 유도된 DSB는 HR에 비해 월등하게 NHEJ 기전으로 복구될 수 있다. HR에 의한 변이는 HR 제공자 DNA에 있는 서열이 정확히 복사되어 일어나지만, NHEJ에 의한 변이는 무작위로 일어나게 된다. NHEJ은 오류 발생이 쉬운 (error-prone) 수리 기전이기 때문에 DSB가 일어난 부위에서 작은 크기의 삽입/결실 (insertion/deletion: indel 돌연변이)이 일어날 수 있으며, 이는 프레임-이동 (frame-shift) 돌연변이를 유발하여 유전자 변이를 일으킨다.

[0008] 특히 CRISPR/Cas 시스템의 Cas9 단백질은 진핵세포 및 생물체 내 유전자 변형을 설계하는데 유용한 도구임에도 불구하고, 이를 코딩하는 유전자의 사이즈가 크기 때문에 바이러스 벡터 내에 삽입하여 세포 내로 전달하고자 하는 경우 바이러스 벡터의 패키징의 한계에 의해 바이러스 생산 효율 및 세포 내로의 전달 효율이 떨어지는 문제가 있다. 이에, Cas9 단백질을 바이러스 벡터를 통해 발현시키기 위한 연구가 필요한 실정이다.

[0009]

[0010] 이러한 배경 하에, 본 발명자들은 바이러스 벡터의 패키징 한계를 극복하고, 바이러스 벡터를 이용하여 Cas9 단백질 발현시킬 수 있는 시스템을 개발하기 위하여 예의 노력한 결과, Cas9 단백질을 바이러스 벡터에 패키징될 수 있는 크기의 두 도메인으로 나눈 뒤, 상기 각각의 도메인을 발현시킬 수 있는 재조합 벡터를 제작하였으며, 상기 재조합 벡터를 세포 내에 도입하여 각 도메인이 융합되어 유전체 상의 표적 DNA에 대해 Indel (insertion or deletion) 효과를 나타내는 것을 확인함으로써 본 발명을 완성하였다.

발명의 내용

해결하려는 과제

[0012] 본 발명의 목적은 Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 발현시키는 재조합 벡터를 각각 세포 내에 도입하는 단계를 포함하는, 유전자 발현 조절 방법을 제공하는 것이다.

[0013] 본 발명의 Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 포함하는 조성물을 제공하는 것이다.

- [0014] 본 발명의 또 다른 목적은 상기 조성물을 포함하는 유전자 발현 조절용 키트를 제공하는 것이다.
- [0015] 본 발명의 또 다른 목적은 Cas9 단백질의 N-말단을 포함하는 제1 도메인을 패키징하는 바이러스 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 패키징하는 바이러스 벡터가 도입된 형질전환 세포를 제공하는 것이다.
- [0016] 본 발명의 또 다른 목적은 상기 세포의 배양액 또는 세포 용해물을 포함하는 조성물을 제공하는 것이다.
- [0017] 본 발명의 또 다른 목적은 Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 각각 세포 내에 도입하는 단계를 포함하는, Cas9 단백질의 세포 내 제조 방법을 제공하는 것이다.

과제의 해결 수단

- [0019] 상기 목적을 달성하기 위한 본 발명의 일 구현에는 Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 각각 세포 내에 도입하는 단계를 포함하는, 유전자 발현 조절 방법을 제공한다.
- [0020] 본 발명에서 용어, "유전자 발현 조절"은 유전자의 발현을 증가시키거나 감소시키는 모든 행위를 말한다. 특히 본 발명의 목적상, 상기 유전자 발현 조절은 Cas9 단백질에 의해 수행되는 것일 수 있다. 구체적으로, Cas9 단백질을 이용하여 유전자 발현을 증가 또는 감소시키는 모든 방법이 본 발명의 범위에 제한 없이 포함될 수 있다. 예컨대, 상기 유전자 발현 조절은 유전체 교정, 유전자 발현 증가 또는 유전자 발현 감소를 의미할 수 있다.
- [0021] 본 발명에서 용어, "유전체 교정 (genome editing)"은 인간세포를 비롯한 동식물 세포의 유전체 염기서열에 표적화된 돌연변이를 도입할 수 있는 기술로서, 특정 유전자를 knockout 또는 knock-in하거나, 단백질을 생성하지 않는 비-코딩 DNA 서열에 변이를 도입하는 것을 말한다. 또한, 유전체 교정을 통해 유전체 상의 DNA를 결실, 중복, 역위, 교체 또는 재배열시킬 수 있다.
- [0022] 본 발명에서, "결실"이란, 염색체의 일부 또는 DNA 상 염기의 일부가 누락되어 일어나는 돌연변이를 말한다.
- [0023] 본 발명에서, "중복"이란, 게놈 내에 같은 유전자가 2개 또는 그 이상 존재하는 것을 말한다.
- [0024] 본 발명에서, "역위"는 게놈의 일부가 원래의 게놈과 비교할 때 거꾸로 배치된 것을 말한다.
- [0025] 본 발명에서, "교체"는 하나의 뉴클레오티드 서열이 서로 교체되는 것(즉, 정보를 지닌 서열의 교체)를 의미하며, 반드시 하나의 폴리뉴클레오티드가 다른 폴리뉴클레오티드로 화학적 또는 물리적으로 교체되는 것만을 의미하는 것은아니다.
- [0026] 본 발명에서, "재배열"이란 염색체 상의 유전자의 위치 및 순서의 변화를 일으키는 구조적 변화를 의미하며, 트랜스포존 등과 같이 전이인자가 삽입되는 것도 포함된다. 아울러, DNA 분자 내에서 염기 재배열에 의한 유전 정보의 변환을 포함할 수 있다.
- [0028] 본 발명에서, "Cas9 단백질"은 CRISPR/Cas9 시스템의 주요 단백질 구성 요소로, crRNA(CRISPR RNA) 및 tracrRNA(trans-activating crRNA)와 복합체를 형성하여 활성화된 엔도뉴클레아제 또는 nickase를 형성한다.
- [0029] Cas9 단백질 또는 유전자 정보는 NCBI(National Center for Biotechnology Information)의 GenBank와 같은 공지의 데이터 베이스에서 얻을 수 있으나, 이에 제한되지 않는다. 예를 들어, 상기 Cas9 단백질은 서열번호 1 또는 서열번호 2의 뉴클레오티드 서열로 코딩되는 것일 수 있으나, 이에 제한되는 것은 아니며 가이드 RNA와 함께 표적 특이적 뉴클레아제 활성을 가질 수 있는 것이라면 모두 본 발명의 범위에 포함될 수 있다. 또한, Cas9 단백질은 단백질 전달 도메인(protein transduction domain)과 연결될 수 있다. 상기 단백질 전달 도메인은 폴리-아르기닌 또는 HIV 유래의 TAT 단백질일 수 있으나, 이에 제한되지 않는다. 나아가, 상기 Cas9 단백질은 그 목적에 따라 당업자에 의해 추가적인 도메인이 적절하게 연결될 수 있다.
- [0030] 또한, 상기 Cas9 단백질은 야생형 Cas9 뿐만 아니라, 불활성화된 Cas9 (dCas9), 또는 Cas9 니케이즈 (nickase)와 같은 Cas9의 변이체를 모두 포함할 수 있다. 상기 불활성화된 Cas9은 dCas9에 FokI 뉴클레아제 도메인을

연결한 RFN (RNA-guided FokI Nuclease), 또는 dCas9에 전사활성인자 (transcription activator) 또는 억제자 도메인 (repressor domain)을 연결한 것일 수 있고, 상기 Cas9 니케이즈는 D10A Cas9 또는 H840A Cas9일 수 있으나, 이에 제한되는 것은 아니다.

- [0031] 본 발명의 Cas9 단백질은 그 유래에도 제한되지 않는다. 예컨대 상기 Cas9 단백질은 스트렙토코커스 피요제네스 (*Streptococcus pyogenes*), 프란시셀라 노비시다 (*Francisella novicida*), 스트렙토코커스 써모필러스 (*Streptococcus thermophilus*), 레지오넬라 뉴모필라 (*Legionella pneumophila*), 리스테리아 이노큐아 (*Listeria innocua*), 또는 스트렙토코커스 뮤탄스 (*Streptococcus mutans*) 유래일 수 있다. 본 발명의 목적상 상기 Cas9 단백질은 사이즈가 커 바이러스 벡터에서 효과적으로 발현될 수 없는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0033] 본 발명에서는 Cas9을 바이러스 벡터에서 발현시키기 위해 Cas9의 일부를 발현할 수 있는 벡터를 제작하였다. 즉, Cas9 단백질을 바이러스 벡터에서 패키징이 가능한 크기로 나누어 각각의 벡터에서 발현시키고자 하였다. 본 발명에서 Cas9 단백질의 제1 도메인 및 제2 도메인은 Cas9 단백질의 일부 부위를 지칭하는 것으로, 이들이 별도의 벡터에서 발현되어 세포 내에서 융합되도록 한 것이다. 본 발명에서는 상기와 같은 방식으로 제작된 Cas9 단백질을 "split-Cas9"이라 명명하였다 (도 1).
- [0034] 본 발명의 split-Cas9은 기존에 크기가 커서 바이러스 벡터 등을 통해 패키징되지 않던 Cas9 단백질을 패키징 가능한 크기로 나누어서 이들 각각을 벡터를 통해 발현시키더라도 세포 내에서 그 기능을 잃지 않는다는 것을 기술적 특징으로 한다.
- [0035] 본 발명에서 용어, "제1 도메인"은 상기와 같은 목적으로 절단된 Cas9의 일부분을 발현시키는 경우 원래 Cas9 단백질의 N-말단 부위를 포함하는 도메인을 말하며, "제2 도메인"은 원래 Cas9 단백질의 C-말단 부위를 포함하는 도메인을 말한다. 이들 각각의 도메인은 본 발명에서 "하프 도메인"이라는 용어와 혼용하여 사용하였다. 상기 각각의 도메인은 바이러스 벡터에서 발현시키기 위한 것이므로, 각각 바이러스 벡터에서 패키징될 수 있는 크기인 400 bp 내지 3.7 kbp의 크기일 수 있다. 구체적으로, 본 발명에서는 상기 제1 도메인 및 제2 도메인이 융합되어 원래의 Cas9 단백질 전체를 구성하므로, 하나의 도메인의 크기는 Cas9 단백질 전체의 크기에서 다른 하나의 도메인의 크기를 뺀 것이 된다.
- [0036] 본 발명의 구체적인 일 실시예에서는 제1 도메인을 2.1 kbp, 제2 도메인을 1.9 kbp로 하여 플라스미드 벡터 및 바이러스 벡터에 도입하였으며, 상기 각각의 벡터를 이용하여 세포 내에서 split-Cas9이 표적위치에서 Indel을 유도할 수 있음을 확인하였다.
- [0037] 또한, 상기 제1 도메인 및 제2 도메인은 그 목적에 따라 당업자에 의해 특정 기능을 가지는 뉴클레오티드를 추가하여 제조될 수 있으며, 예를 들어 NLS (nuclear localization signal), 태그 (tag) 서열, 또는 스플라이싱 도너 (splicing donor) / 스플라이싱 억셉터 (splicing acceptor) 서열 등을 추가로 포함할 수 있다. 나아가, 이에 제한되는 것은 아니나 상기 제1 도메인은 서열번호 3의 뉴클레오티드로 코딩되는 것일 수 있고, 제2 도메인은 서열번호 5의 뉴클레오티드로 코딩되는 것일 수 있다.
- [0039] 본 발명에서, "벡터"란, 적당한 숙주세포에서 목적 단백질을 발현할 수 있는 발현 벡터로서, 유전자 삽입물이 발현되도록 작동 가능하게 연결된 필수적인 조절 요소를 포함하는 유전자 작제물을 말한다.
- [0040] 본 발명에서 용어, "작동 가능하게 연결된(operably linked)"은 일반적 기능을 수행하도록 핵산 발현조절 서열과 목적하는 단백질을 코딩하는 핵산 서열이 기능적으로 연결(functional linkage)되어 있는 것을 말한다. 예를 들어, 본 발명의 뉴클레아제의 DNA 제1 도메인 또는 DNA 제2 도메인을 암호화하는 서열은 프로모터에 작동 가능하게 연결시킴으로써, 상기 암호화 서열의 발현은 이 프로모터의 영향 또는 조절 하에 있게 된다. 2 개의 핵산 서열(DNA 제1 도메인 또는 DNA 제2 도메인을 암호화하는 서열과 이 서열의 5' 말단의 프로모터 부위 서열)은 프로모터 작용이 유도됨으로써 상기 암호화 서열이 전사되는 경우 작동 가능하게 연결된 것이며, 상기 두 서열 사이의 연결 특성이 프레임 변경 돌연변이(frame-shift mutation)를 유도하지 않으며, 발현 조절서열이 상기 각 도메인의 발현을 지배하는 능력을 저해하지 않는 경우 작동 가능하게 연결되었다. 재조합 벡터와의 작동 가능한 연결은 당해 기술 분야에서 잘 알려진 유전자 재조합 기술을 이용하여 제조할 수 있으며, 부위-특이적 DNA 절단 및 연결은 당해 기술 분야에서 일반적으로 알려진 효소 등을 이용할 수 있다.

- [0041] 본 발명의 벡터는 프로모터, 오퍼레이터, 개시코돈, 종결코돈, 폴리아데닐화 시그널, 인핸서와 같은 발현 조절 요소 외에도 막 표적화 또는 분비를 위한 신호 서열 또는 리더 서열을 포함할 수 있으며, 목적에 따라 다양하게 제조될 수 있다. 벡터의 프로모터는 구성적 또는 유도성일 수 있다. 또한, 발현벡터는 벡터를 함유하는 숙주 세포를 선택하기 위한 선택성 마커를 포함할 수 있으며, 복제 가능한 발현벡터인 경우 복제 기원을 포함할 수 있다. 벡터는 자가 복제하거나 숙주 DNA에 통합될 수 있다. 상기 벡터는 플라스미드 벡터, 코즈미드 벡터 또는 바이러스 벡터 등을 포함할 수 있으며 구체적으로는 바이러스 벡터일 수 있다. 바이러스 벡터는 레트로바이러스(Retrovirus), 예를 들어 HIV(Human immunodeficiency virus) MLV(Murine leukemia virus) ASLV(Avian sarcoma/leukosis), SNV(Spleen necrosis virus), RSV(Rous sarcoma virus), MMTV(Mouse mammary tumor virus) 등, 아데노바이러스(Adenovirus), 아데노 부속 바이러스(Adeno-associated virus), 헤르페스 심플렉스 바이러스(Herpes simplex virus) 등에서 유래한 벡터를 포함하나, 이에 제한되지 않는다.
- [0043] 본 발명에서, "세포 내에 도입하는 단계"는 당업계에 알려진 공지 어떠한 방법도 사용할 수 있으며, 형질감염(transfection) 또는 형질도입(transduction)에 의해 외래 DNA를 세포로 유입시킬 수도 있다. 형질감염은 칼슘 포스페이트-DNA 공침전법, DEAE-텍스트란-매개 형질감염법, 폴리브렌-매개 형질감염법, 전기충격법, 미세주사법, 리포좀 융합법, 리포펙타민 및 원형질체 융합법 등의 당분야에 공지된 여러 방법에 의해 수행될 수 있다.
- [0044] 본 발명의 일 실시예에서는 Cas9 단백질 중 제1 도메인과 제2 도메인을 각각 코딩하는 각각의 재조합 벡터를 제작한 뒤 (도 1), 이를 세포 내로 도입시켜 발현된 각 도메인이 세포 내에서 융합하여 온전한 형태의 Cas9 단백질로서 기능하는 것을 확인하였다. 구체적으로, 각 하프 도메인인 제1 도메인 및 제2 도메인을 각각 발현하는 재조합 벡터들로부터 각 하프 도메인이 발현 후 융합하여 형성된 Cas9 단백질이 sgRNA와 함께 작용하여 타겟 유전자 모두에 Indel(insertion or deletion)을 유도하는 것을 확인하였다 (도 2 및 도 3).
- [0045] 본 발명의 다른 일 실시예에서는 HeLa 세포 및 Hep1 세포에서 split-Cas9 단백질을 이용하여 표적 특이성을 확인한 결과 야생형 Cas9의 특이성에 비해 80 배 내지 220 배 수준으로 현저히 증가하는 것을 확인하였다 (도 4). 이는, 본 발명의 split-Cas9을 세포 내에서 발현시키는 경우 비표적 효과 (off-target)를 최소화하여 원하는 표적 위치에 작용하도록 할 수 있다는 것을 시사한다.
- [0046] 나아가, 본 발명의 다른 일 실시예에서는 split-Cas9을 아데노부속 바이러스 (adeno-associated virus) 벡터를 이용하여 발현시키고, 생산된 바이러스를 세포에 감염시킨 결과 유효하게 Indel을 유도하는 것을 확인하였다 (도 5). 이에 본 발명의 split-Cas9을 사용하여 바이러스 벡터를 통해서도 효과적으로 Cas9 단백질을 이용할 수 있음을 확인하였다.
- [0048] 구체적으로, 상기 세포 내 도입 시, 서열 특이적인 가이드 RNA (guide RNA)를 추가로 도입할 수 있다. 보다 구체적으로, 상기 각각의 벡터, 가이드 RNA의 도입은 동시, 순차 또는 역순으로 수행될 수 있다.
- [0049] 본 발명에서, "가이드 RNA(guide RNA)"는 두 개의 RNA, 즉, crRNA (CRISPR RNA) 및 tracrRNA(trans-activating crRNA)로 구성될 수 있다. 또는 crRNA 및 tracrRNA의 주요 부분의 융합으로 제조된 sgRNA (single-chain RNA)일 수 있다. 또한, 상기 가이드 RNA는 crRNA 및 tracrRNA를 포함하는 dual RNA일 수 있다.
- [0050] 3세대 유전자 가위로 알려진 RGEN은 Cas 단백질 및 dual RNA로 구성되거나, Cas 단백질 및 sgRNA로 구성될 수 있다. 가이드 RNA는 sgRNA 또는 dualRNA의 crRNA의 5' 말단에 하나 이상의 추가의 뉴클레오티드를 포함할 수 있으며, RNA 또는 상기 RNA를 코딩하는 DNA형태로 세포 내로 전달될 수 있다.
- [0052] 본 발명의 또 다른 일 구현에는 Cas9 단백질의 N-말단을 포함하는 제1 도메인을 발현하는 재조합 벡터 및 Cas9 단백질의 C-말단을 포함하는 제2 도메인을 발현하는 재조합 벡터를 포함하는 조성물을 제공한다. 상기 조성물을 세포에 도입하여 원하는 유전자의 발현을 조절할 수 있다. 상기 조성물은 서열 특이적인 가이드 RNA를 추가로 포함할 수 있다. Cas9 단백질 및 재조합 벡터는 앞서 설명한 바와 같다.
- [0054] 본 발명 일 실시예에서는 상기 제1 도메인을 발현하는 재조합 벡터 및 제2 도메인을 발현하는 재조합 벡터를 세

포 내에 도입하였다. 이는 벡터 내에 패키징이 어려운 크기를 가지는 Cas9 단백질을 보다 효율적으로 세포 내에 전달하여 발현시키기 위한 것으로, 상기 각각의 재조합 벡터를 포함하는 조성물을 이용하여 세포 내에서 Cas9 단백질을 보다 용이하게 발현시킬 수 있다. 상기 조성물은 제1 도메인을 발현하는 재조합 벡터 및 제2 도메인을 발현하는 재조합 벡터 외에도 세포를 유지할 수 있는 배지용 조성물 또는 재조합 벡터를 세포 내에 도입하는데 필요한 물질은 제한 없이 포함할 수 있다.

[0056] 본 발명의 또 다른 일 구현예는 상기 Cas9 단백질의 제1 도메인을 발현하는 재조합 벡터 및 제2 도메인을 발현하는 재조합 벡터를 포함하는, 유전자 발현 조절용 키트를 제공한다. 구체적으로 상기 키트는 서열 특이적인 가이드 RNA를 추가로 포함하는 것일 수 있다.

[0057] 또한, 상기 키트는 재조합 벡터의 발현이 이루어지거나 촉진될 수 있도록 하는 물질 또는 세포를 유지할 수 있도록 하는 배지용 조성물 뿐 아니라 재조합 벡터의 제작 또는 세포 내로의 도입 등을 용이하게 할 수 있는 조성물, 재조합 벡터의 제작 또는 세포 내로의 도입을 위한 설명서 등을 포함할 수 있다.

[0059] 본 발명의 또 다른 일 구현예는 상기 Cas9 단백질의 제1 도메인을 패키징하는 바이러스 벡터 및 제2 도메인을 패키징하는 바이러스 벡터가 도입된 형질전환 세포를 제공한다.

[0060] 본 발명에서 용어, "형질전환 세포"는 숙주세포에 목적으로 하는 폴리뉴클레오티드를 도입한 세포를 의미한다. 형질전환은 상기 "도입" 방법에 의해 이루어질 수 있고, 당 분야에서 공지된 바와 같이 숙주세포에 따라 적합한 표준 기술을 선택하여 수행할 수 있다.

[0061] 상기 "숙주 세포"는 하나 이상의 DNA 또는 벡터가 도입되는 진핵 또는 원핵 세포를 가리키며, 특정 대상 세포만이 아니라 그 자손 혹은 잠재적 자손까지도 가리키는 것으로 이해되어야 한다. 돌연변이 혹은 환경적 영향에 의해 상기 자손이 부모 세포와 완전히 동일하지 않더라도, 본 명세서에서 사용된 바와 같이 상기 용어의 범주 내에서 여전히 포함될 수 있다. 본 발명에서 상기 형질전환 세포는 각각의 하프 도메인을 코딩하는 바이러스 벡터가 도입된 것으로, 이로부터 Cas9의 각각의 도메인을 코딩하는 뉴클레오티드를 패키징하는 바이러스를 획득할 수 있다. 구체적으로, 상기 형질전환 세포의 배양액 또는 상기 세포의 용해물로부터 상기 바이러스를 획득할 수 있다.

[0062] 상기 세포는 *E. coli*와 같은 원핵세포, 이스트, 진균, 원생동물, 고등식물, 곤충 등의 진핵 세포, CHO, HeLa, HEK293, COS-1과 같은 포유류 세포 등이 포함될 수 있으며, 상기 예시에 제한되는 것은 아니다.

[0063] 또한, 체세포, 생식세포, 역분화줄기세포 (induced pluripotent stem cells), 성체줄기세포 등 인간의 모든 세포에 적용할 수 있다.

[0064] 상기 체세포는 배아 뿐 아니라 어린아이, 성인의 몸에서 얻을 수 있는 생식세포를 제외한 모든 세포를 말하며, 이들로부터 유래된 유전자변형 세포들까지 포함될 수 있다. 또한, 상기 성체줄기세포는 인간 배아 (embryo), 신생아 및 성인의 몸에서 얻을 수 있는 모든 성체줄기세포들 뿐만 아니라 제대혈유래 줄기세포 (cord blood stem cells), 태반유래 줄기세포 (placenta stem cells), 탯줄줄기세포 (Wharton's jelly stem cells), 양수줄기세포 (amniotic fluid stem cells), 양막상피줄기세포 (amniotic epithelial cells) 등 배외줄기세포 (extraembryonic stem cells)들과 이들로부터 파생된 유전자변형 세포들을 포함할 수 있다.

[0065] 또한, 상기 세포는 배양된 세포 (시험관내), 이식편 및 1차 배양물 (시험관내 및 생체외) 및 생체 내 세포일 수도 있으며, 당업계에서 통상적으로 사용되는 세포로서 그 제한을 두지 않는다.

[0067] 본 발명의 또 다른 일 구현예는 상기 형질전환 세포의 배양액 또는 세포 용해물을 포함하는 조성물을 제공한다. 형질전환 세포, 이의 배양액 및 용해물은 상기 설명한 바와 같다. 상기 조성물은 각각의 도메인을 코딩하는 뉴클레오티드를 패키징하는 바이러스를 포함하므로, 유전자 발현 조절 용도로 사용될 수 있다.

[0069] 본 발명은 Cas9 단백질에 대한 벡터 패키징의 한계를 극복하고, 상기 Cas9 단백질의 절단된 두 개의 부위 각각을 독립적으로 발현시키는 재조합 벡터를 제작하여 이를 세포 내로 전달하여 발현시킴으로써 Cas9 단백질의 세

포 내로의 전달 효율을 증가시킨 것이다. 따라서, 본 발명자들이 개발한 본 발명의 원리를 이용하여 세포의 종류, Cas9 단백질의 종류에 상관없이 적용하여 세포 내로의 전달 효율을 증가시켜 유전자 발현을 효율적으로 조절할 수 있다.

발명의 효과

[0071] 본 발명은 Cas9 단백질의 표적 특이성을 향상시킬 수 있으며, 또한 Cas9 단백질을 바이러스 벡터에도 적용할 수 있도록 하므로, Cas9 단백질을 이용한 유전자 발현 조절에 유용하게 사용될 수 있다.

도면의 간단한 설명

[0073] 도 1은 Cas9 단백질의 제1 도메인과 제2 도메인을 각각 포함하는 재조합 벡터를 제작한 후 세포 내에 도입시켜 발현시킴으로써, 이들 각각의 단백질이 세포 내에서 융합하여 온전한 Cas9 단백질(Split-Cas9으로 명명)을 형성하는 모식도를 나타낸 것이다.

도 2는 세포 내에서 Split-Cas9 단백질이 형성되어 sgRNA와 함께 작용하여 HPRT, DMD 및 CCR5 유전자 모두에 Indel을 유도하는 것을 T7E1 분석을 통해 확인한 결과를 나타낸 것이다. 1) Split-Cas9 + sgRNA, 2) 제2 도메인 + sgRNA, 3) 제1 도메인 + sgRNA, 4) 세포만.

도 3은 Split-Cas9 단백질의 표적 유전자에 대한 변이 효율을 next-generation 시퀀싱을 통해 분석한 결과를 나타낸 것이다.

도 4는 Split-Cas9 단백질의 표적 유전자에 대한 특이성을 분석한 결과를 나타낸 것이다. (a) HeLa 세포 및 (b) Hep1 세포에서 표적 위치 (on-target site)와 각각의 비표적 위치 (off-target site)에서의 유전자 변이 효율을 next-generation 시퀀싱을 통해 분석한 결과를 나타낸 것이다. 특이성은 표적 위치의 변이 효율을 4 개의 비표적 위치에서의 각각의 변이 효율로 나눈 비율 (specificity ratio) 로 분석하였다.

도 5는 split-Cas9의 벡터 제작 및 기능을 확인한 결과를 나타낸 것이다. (a) 및 (b) split-Cas9을 전달하는 아데노부속바이러스 벡터 제작을 모식화한 것이다. 아데노부속바이러스 벡터 내에 U6 프로모터, sgRNA, EFS 프로모터, 제1 도메인이 순차적으로 삽입되어 한 벡터 내에 패키징하는 바이러스 벡터를 제작하였다. 또한 제2 도메인을 포함하거나 제2 도메인과 U6 프로모터, sgRNA를 동시에 패키징할 수 있는 아데노부속바이러스를 제작하였다. (c) 아데노부속바이러스 벡터 내에 U6 프로모터, sgRNA, EFS 프로모터, 제1 도메인이 패키징되어 있는 바이러스와 제2 도메인을 패키징하는 바이러스를 각각 10, 50, 100 MOI (multiplicity of infectivity)로 HeLa 세포에 동시 감염시킨 후 5 일, 7 일, 10 일 후에 DMD 유전자에 변이를 유도하는 것을 next-generation 시퀀싱을 통해 분석한 결과를 나타낸 것이다.

발명을 실시하기 위한 구체적인 내용

[0074] 이하, 본 발명을 실시예에 의해 상세히 설명한다. 단, 하기 실시예에는 본 발명을 예시하는 것일 뿐, 본 발명이 하기 실시예에 의해 한정되는 것은 아니다.

[0076] 실시예 1. Cas9 단백질의 제1 도메인 및 제2 도메인 각각을 발현시키는 재조합 벡터의 제작

[0078] 야생형 (Wild-type, WT) Cas9 (CRISPR associated protein 9) 단백질(서열번호 2)의 중간부위에 존재하는 disordered linker(서열번호 9; agcggccagggc; SGQG 아미노산을 인코딩하는 시퀀스)의 중간 부위를 절단하여 SG 아미노산과 QG 아미노산이 제1 도메인과 제2 도메인에 각각 연결된 두 개의 하프 도메인을 제작하였다.

[0079] 각각의 하프 도메인은 CMV 프로모터에 의해 독립적인 발현이 유도될 수 있도록 하였다. 제1 도메인의 절단 3'-end 부위에 정지코돈 (Stop codon)을 PCR 클로닝을 통해 삽입함으로써 발현이 완료될 수 있도록 하였으며, 제2 도메인의 절단 5'-end 부위에 시작코돈 (Start codon)을 연결하여 발현이 시작될 수 있도록 제작하였다. 제 1 도메인의 5 -end 부위의 시작코돈 (start codon) 아래쪽으로 HA 태그 (tag)와 NLS (nuclear localization signal)이 순차적으로 삽입되어 있고 제 2 도메인의 3'end 와 정지코돈 (stop codon)사이에 NLS 부위와 HA 태그

를 순차적으로 삽입하여 핵 내로의 전달 및 HA 항체로 단백질 발현량을 측정할 수 있게 제작하였다.

[0081] 실시예 2. 각 하프 도메인을 발현시키는 재조합 벡터와 sgRNA의 세포 내로의 도입 및 타겟 유전자의 닉아웃 (knock out) 확인

[0083] 상기 실시예 1에서 제작한 각 하프 도메인을 발현시키는 재조합 벡터들과 CCR5, HPRT 및 DMD 각 유전자에 대한 sgRNA(single guide RNA)를 발현하는 각각의 플라스미드들을 리포펙타민(lipofectamin)을 이용한 형질주입으로 HeLa 세포 내에 전달하였다.

[0084] CCR5 유전자의 대립 유전자 닉아웃을 위해 사용할 수 있는 표적 서열 (target sequence)은 모든 인간에 공통적으로 존재하는 보존된 서열을 사용하였으며, CCR5 엑손 2에 존재하는 5'- TGACATCAATTATTATACATCGG -3' 서열 (서열번호 11)을 표적으로 하였다.

[0085] 또한, HPRT 유전자의 대립 유전자 닉아웃을 위해 사용할 수 있는 표적 서열로서, HPRT 엑손 3에 존재하는 5'- GCCCCCTTGAGCACACAGAGGG -3' 서열(서열번호 12)을 표적으로 하였다.

[0086] 또한, DMD 유전자의 대립 유전자 닉아웃을 위해 사용할 수 있는 표적 서열로서, DMD 엑손 51에 존재하는 5'- TCCTACTCAGACTGTTACTCTGG-3' 서열(서열번호 13)을 표적으로 하였다.

[0087] 이후 HeLa 세포로부터 유전체 DNA를 추출한 후 HPRT, DMD 및 CCR5 각 유전자 내의 표적 서열 부위를 PCR로 증폭시켰다.

[0088] 이후 Indel (insertion or deletion)이 유도되었는지를 T7E1 (T7 endonuclease I) mutation detection assay로 분석을 하였으며, 아가로스 겔 분석 결과를 도 2에 나타내었다. T7E1 분석 (T7E1 assay)은 기존에 공지된 방법으로 수행하였다. 요약하면, 제조사의 지시에 따라 유전체 DNA를 DNeasy Blood & Tissue Kit (G-DEX IIc Genomic extraction kit)를 이용하여 분리하였다.

[0090] 도 2에 나타낸 바와 같이, 세포 내에 도입된 실시예 1에서 제작한 각 하프 도메인을 발현하는 재조합 벡터들로부터 각 하프 도메인이 발현 후 융합되어 Cas9 단백질 (Split-Cas9으로 명명)을 형성함으로써 sgRNA와 함께 작용하여 HPRT, DMD 및 CCR5 유전자 모두에 Indel을 유도하는 것을 알 수 있었다.

[0092] 실시예 3. Split-Cas9과 타겟유전자 특이적 sgRNA에 의한 닉아웃 효율 분석

[0094] 타겟 유전자의 닉아웃 효율을 분석하기 위해 타겟 시퀀스 부위를 PCR로 증폭시킨 후 next generation assay로 표적 서열을 분석하였다. 분석 결과, HPRT 유전자에서 27.1 %, DMD 유전자에서 23.75 %, CCR5 유전자에서 20.27 % 의 Indel 효과가 나타났다. 대조군으로 제1 또는 제2 도메인에 대한 한 개의 하프 도메인만 세포 내에 주입하여 발현시켰으며, 이러한 경우에는 Indel이 나타나지 않았다 (도 3).

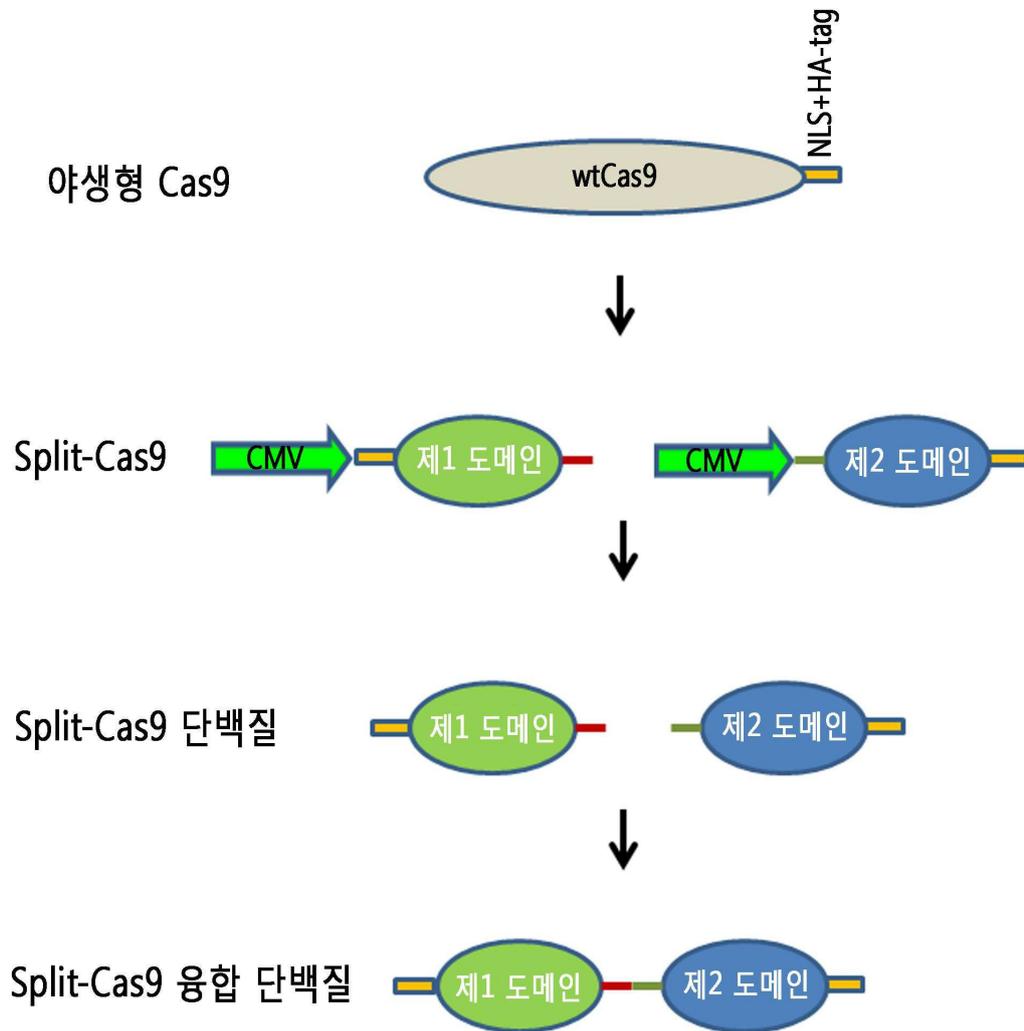
[0096] 상기와 같은 결과로부터 Cas9에 대한 각 하프 도메인을 발현시키는 재조합벡터를 세포 내에 주입한 경우, 각 하프 도메인이 정상적으로 발현 후 융합되어 온전한 Cas9 단백질을 형성함으로써 sgRNA와 함께 표적 유전자에 대해 Indel 효과를 나타낼 수 있음을 알 수 있었다. 본 발명은 바이러스 벡터 내 패키징이 가능한 크기를 넘어서는 Cas9 발현 카세트의 크기로 인해 세포 내로의 전달 효율이 떨어지는 Cas9에 대해 제1 도메인 및 제2 도메인 각각을 독립적으로 세포 내에 도입하여 각각의 발현 후 융합되어 기능할 수 있도록 함으로써 Cas9 단백질의 벡터 내로의 패키징 문제를 해결한 것이다.

[0098] 실시예 4. Split-Cas9의 표적 서열 부위 절단 특이성 분석

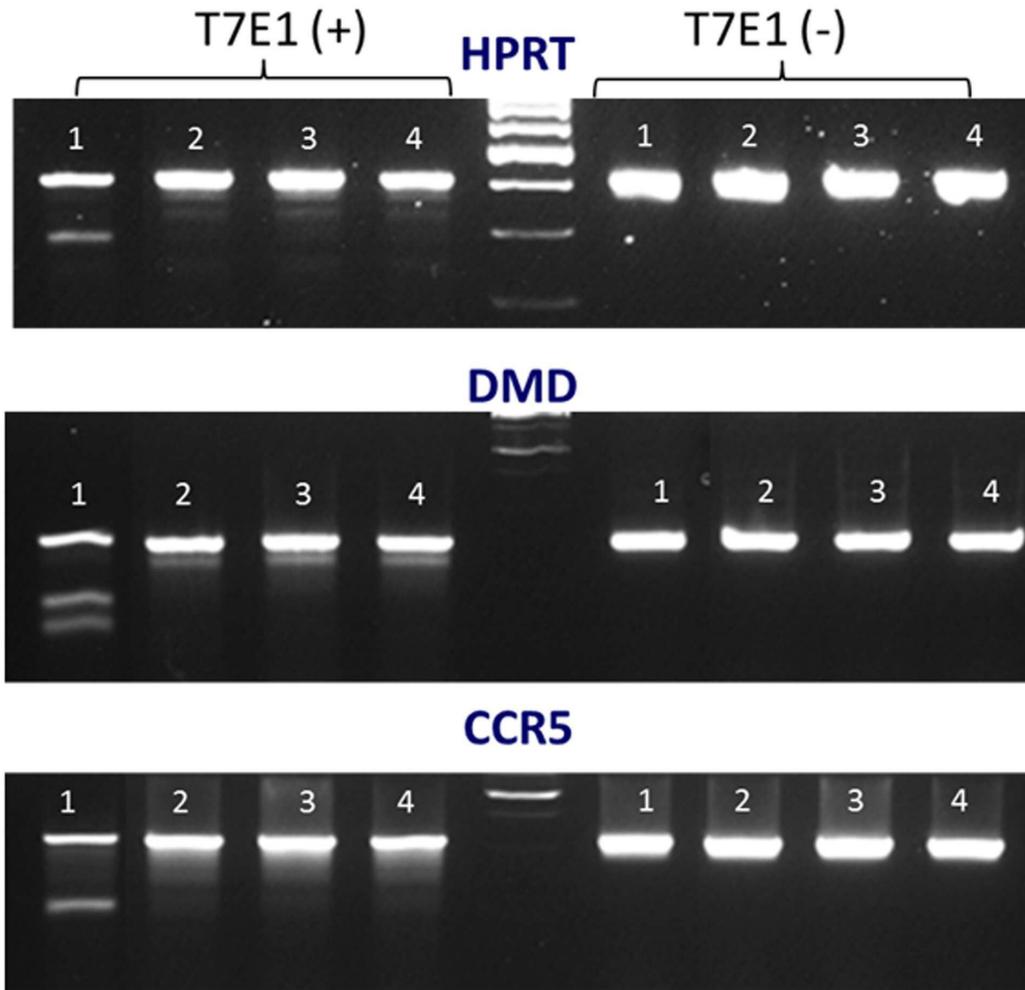
- [0100] 타겟 유전자의 비표적 (off-target) 효과를 분석하기 위해 세포에 split-Cas9과 야생형 Cas9 플라스미드를 각각 처리 후 3 일 뒤 HBB 유전자 표적 (on-target) 서열과 서열 불일치를 가지고 있는 유사 시퀀스 부위를 PCR로 증폭시킨후 next generation assay로 표적 서열을 분석하였다.
- [0101] HeLa 세포 내에서 표적 효율을 비표적 효율로 나누었을 때 split-Cas9에 의한 특이성이 야생형 Cas9의 특이성에 비해 최대 220 배 이상 증가하는 것을 확인하였다 (도 4a). 또한, Hep1 세포내에서 split-Cas9의 특이성이 야생형 Cas9에 비해 최대 80 배 증가하는 것을 확인하였다 (도 4b).
- [0103] **실시예 5. Split-Cas9을 발현하는 아데노부속바이러스에 의한 표적 서열 부위 절단 특이성 분석**
- [0105] Split-Cas9이 바이러스 벡터를 이용한 경우에도 효과적으로 작용하는가를 확인하기 위해, 제1 도메인 및 제2 도메인 각각을 아데노부속바이러스 (Adeno-associated virus) 벡터 플라스미드에 클로닝하였다 (도 5a 및 5b).
- [0106] 제1 도메인의 C-말단 부위에 스플라이싱 도너 (splicing donor)를 연결했고 제2 도메인의 N-말단 부위에 스플라이싱 억셉터 (splicing acceptor)를 연결했다. 각 하프 도메인을 패키징하는 바이러스를 생산하고, 이를 회수하여 세포 내 전달한 다음, 표적 서열 부위의 절단율을 분석했다.
- [0107] 그 결과, 각 하프 도메인이 융합된 후 full-length Cas9 단백질이 형성함으로써 유전자 절단 효과를 나타낼 수 있음을 알 수 있었다. Split-Cas9을 전달하는 아데노부속바이러스를 HeLa 세포 내 10, 50, 100 MOI (multiplicity of infectivity)로 감염시킨 뒤 5, 7, 10 일 후 표적 서열 절단율을 분석한 결과 약 5 %의 표적 서열 절단 효과를 확인할 수 있었다 (도 5c). 이에, 본 발명의 split-Cas9이 바이러스 벡터를 이용할 경우에도 효과적으로 작동하는 것을 알 수 있었다.
- [0109] 이상의 설명으로부터, 본 발명이 속하는 기술분야의 당업자는 본 발명이 그 기술적 사상이나 필수적 특징을 변경하지 않고서 다른 구체적인 형태로 실시될 수 있다는 것을 이해할 수 있을 것이다. 이와 관련하여, 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적인 것이 아닌 것으로서 이해해야만 한다. 본 발명의 범위는 상기 상세한 설명보다는 후술하는 특허 청구범위의 의미 및 범위 그리고 그 등가 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본 발명의 범위에 포함되는 것으로 해석되어야 한다.

도면

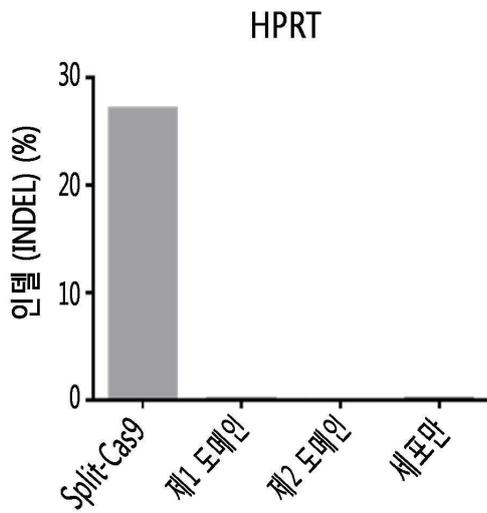
도면1



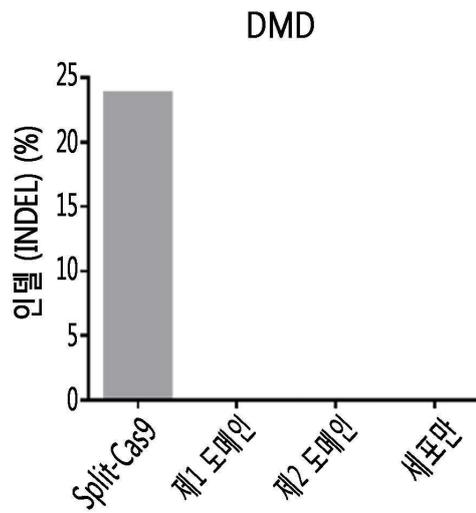
도면2



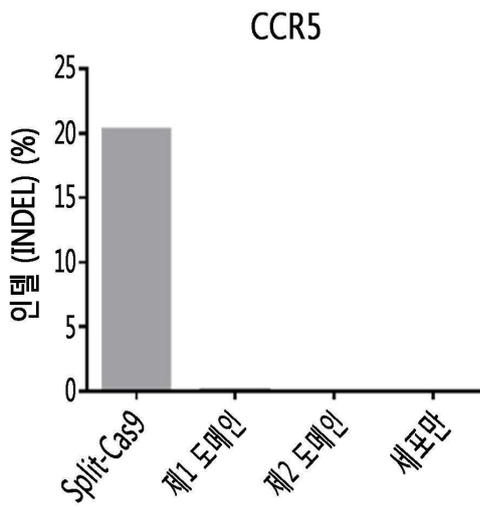
도면3



HPRT	Split-Cas9	제1도메인	제2도메인	Mock
인델 (%)	27.10	0.04	0.03	0.04

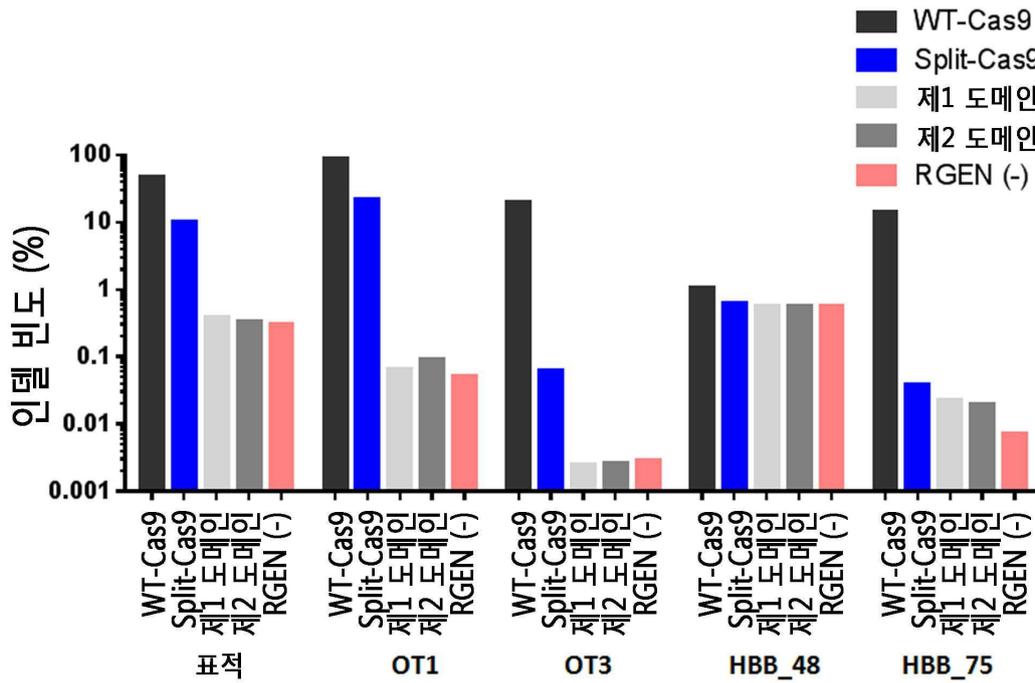


DMD	Split-Cas9	제1도메인	제2도메인	Mock
인델 (%)	23.75	0.00	0.03	0.00



CCR5	Split-Cas9	제1도메인	제2도메인	Mock
인델 (%)	20.27	0.03	0.01	0.02

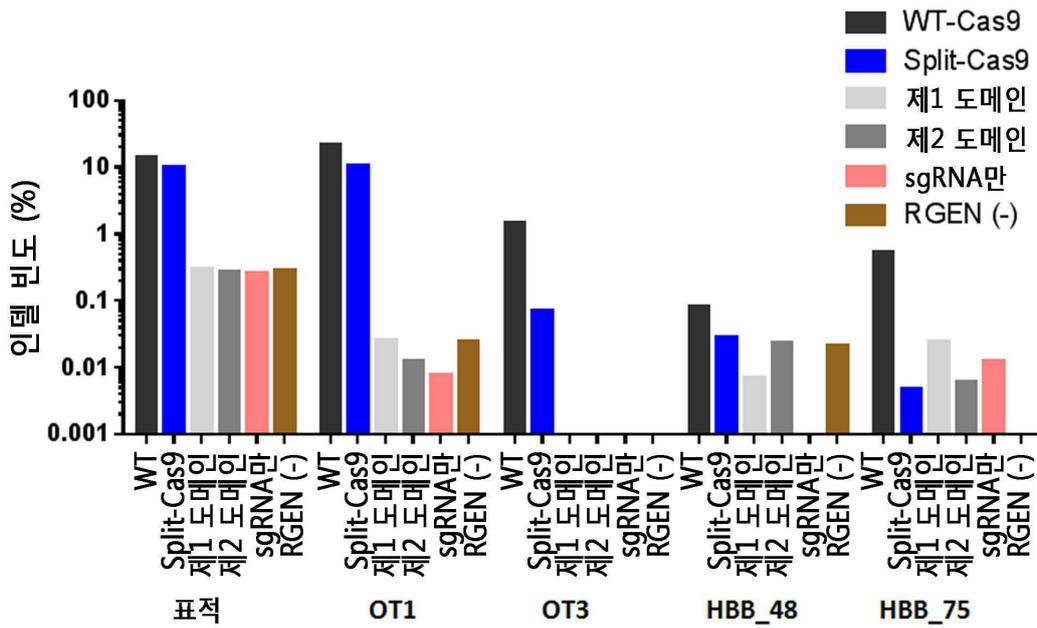
도면4a



인델 (%)	표적	OT1	OT3	HBB_48	HBB_75
Split-Cas9	10.11307	22.097	0.0619	0.6079	0.0375
WT Cas9	46.1303	89.66	19.952	1.0588	14.357
제1 도메인	0.378484	0.0642	0.0024	0.5667	0.0225
제2 도메인	0.331381	0.0922	0.0026	0.5609	0.0194
RGEN(-)	0.299536	0.0502	0.0029	0.5701	0.0072

특이성 비율 (표적/비표적)				
	OT1	OT3	HBB_48	HBB_75
Split-Cas9	0.45767598	356.924	0.10184	16.2097
WT Cas9	0.51450171	4.49377	18.8441	0.07375
Split-Cas9/WT Cas9	0.88955192	79.4264	0.0054	219.803

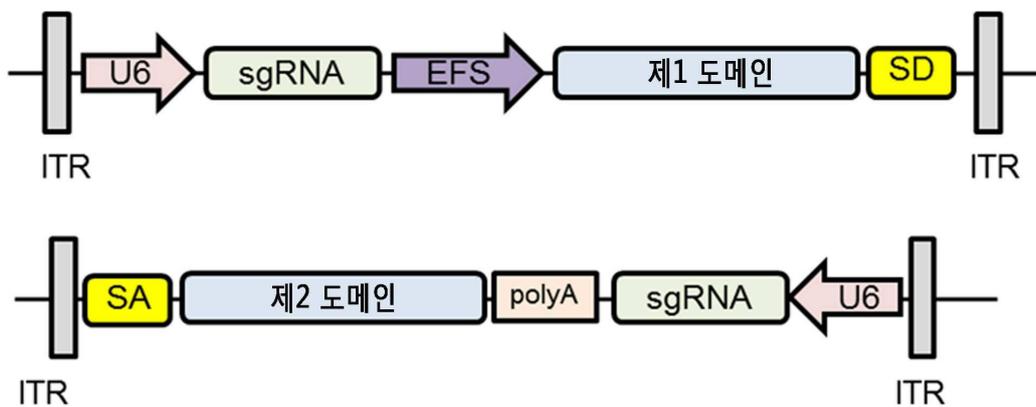
도면4b



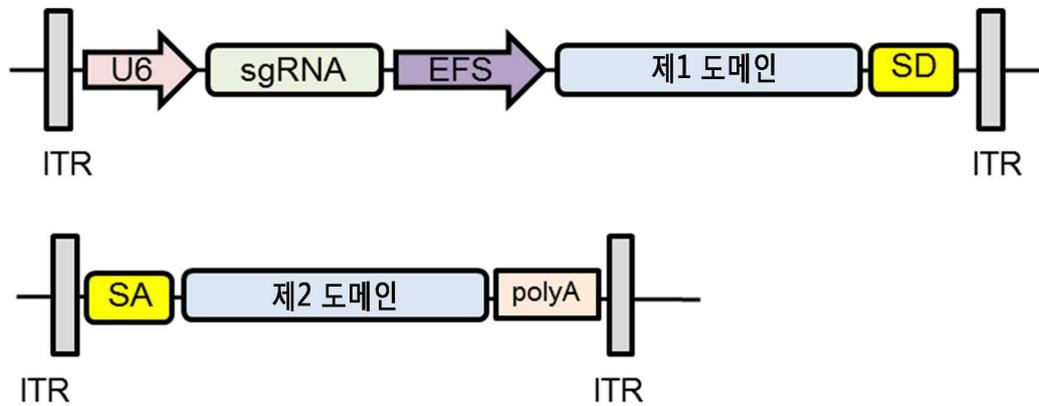
인텔 (%)	표적	OT1	OT3	HBB_48	HBB_75
Split-Cas9	9.908309	10.22706	0.071096	0.028461	0.00464
WT cas9	14.08247	21.75932	1.425269	0.082473	0.532676
제1 도메인	0.297514	0.025918	0	0.007018	0.023818
제2 도메인	0.272719	0.012307	0	0.023611	0.006136
RGEN (-)	0.278242	0.023762	0	0.021216	0

특이성 비율 (표적/비표적)				
	OT1	OT3	HBB_48	HBB_75
Split-Cas9	0.96883	139.37	348.13	2135.3
WT cas9	0.64719	9.8806	170.75	26.437
Split-Cas9/WT cas9	1.49698	14.105	2.0388	80.768

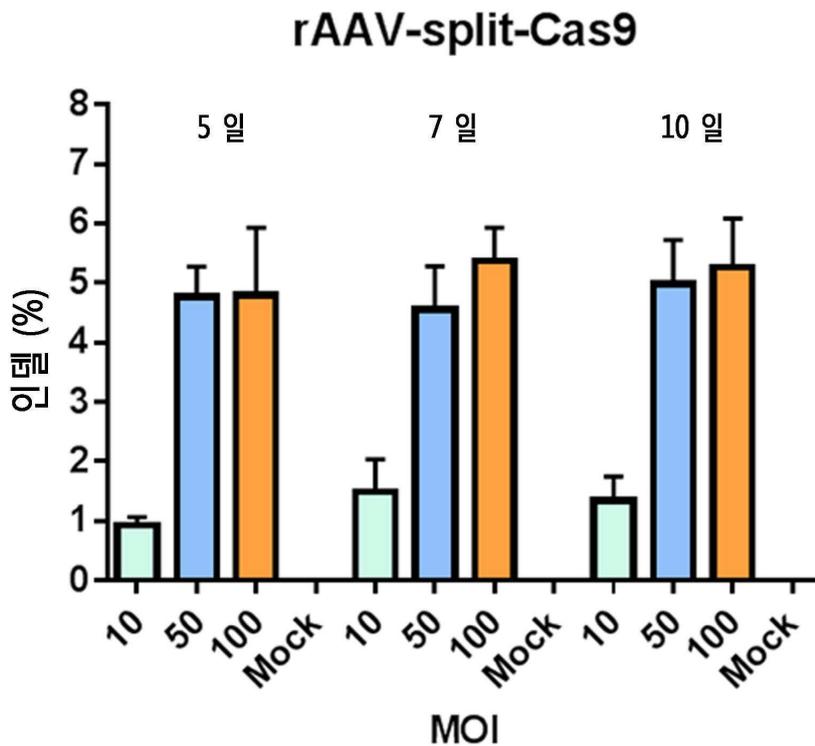
도면5a



도면5b



도면5c



인텔(%)	MOI 10	MOI 50	MOI 100	Mock
5 일	0.94	4.78	4.81	0.00
7 일	1.49	4.56	5.37	0.00
10 일	1.35	4.99	5.26	0.00

서열 목록

<110> INSTITUTE FOR BASIC SCIENCE

<120> A method for regulation of gene expression by expressing Cas9 protein from the two independent vector

<130> KPA141046-KR-P1

<150> KR 2014-0161809
 <151> 2014-11-19
 <160> 54
 <170> Kopatent In 2.0
 <210> 1
 <211> 4104
 <212> DNA
 <213> Artificial Sequence
 <220><223> Cas9
 <400> 1

atggacaaga agtacagcat cggcctggac atcggtagca acagcgtggg ctgggccgtg 60
 atcacccgacg agtacaaggt gcccagcaag aagttcaagg tgctgggcaa caccgaccgc 120

 cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag 180
 gccacccgcc tgaagcgcac cgcccgccgc cgctacacc gccgcaagaa cgcactctgc 240
 tactgcagg agatcttcag caacgagatg gccaaaggtg acgacagctt cttccaccgc 300
 ctggaggaga gcttctctgt ggaggaggac aagaagcacg agcggccacc catcttcggc 360
 aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgcgcaag 420
 aagctggtgg acagcaccga caaggccgac ctgcgcctga tctacctggc cctggcccac 480
 atgatcaagt tccgcggeca cttctctgatc gagggcgacc tgaaccccga caacagcgac 540

 gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc 600
 atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgctgag caagagccgc 660
 cgcttgaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggaac 720
 ctgatcgccc tgagcctggg cctgaccccc aacttcaaga gcaacttca cctggccgag 780
 gacgccaage tgcagctgag caaggacacc tacgacgacg acctggaaa cctgctggcc 840
 cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgcatc 900
 ctgctgagcg acatcctgcg cgtgaacacc gagatcacca aggccccct gagcgccagc 960

 atgatcaagc gctacgacga gcaccaccag gacctgacce tgctgaagge cctggtgcgc 1020
 cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc 1080
 ggctacatcg acggcggcgc cagccaggag gatttctaca agttcatca gccatcctg 1140
 gagaagatgg acggcaccga ggagctgctg gtgaagetga accgcgagga cctgctgcgc 1200
 aagcagcgca ccttcgaaa cggcagcatc ccccaccaga tccacctggg cgagctgcac 1260
 gccatcctgc gccgccagga ggaattctac cccttctga aggacaaccg cgagaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ccctggcccg cggcaacagc 1380

 cgcttcgcct ggatgaccgg caagagcgag gagaccatca cccctggaa cttcgaggag 1440
 gtggtggaca agggcgccag cgcccagagc ttcacgagc gcatgaccaa cttcgacaag 1500
 aacctgccc aacgagaagg gctgcccagg cacagcctgc tgtacgagta cttcacctg 1560
 tacaacgagc tgaccaaggt gaagtacgtg accgagggca tgcgcaagcc cgccttctg 1620
 agcggcgagc agaagaaggc catcgtggac ctgctgttca agaccaaccg caaggtgacc 1680
 gtgaagcagc tgaaggagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc 1740
 agcggcgtgg aggaccgctt caacgccagc ctgggcacct accacgacct gctgaagatc 1800

 atcaaggaca aggacttctt ggacaacgag gagaacgagg acatcctgga ggacatcgtg 1860
 ctgaccctga cctgtttcga ggaccgagc atgatcgagg agcgcctgaa gacctacgcc 1920
 cacctgttcg acgacaaggt gatgaagcag ctgaagcgcc gccgctacac cggctggggc 1980
 cgcctgagcc gcaagcttat caacggcatc cgcgacaagc agagcggcaa gaccatcctg 2040
 gacttctcga agagcgacgg cttcgccaac cgcaacttca tgcagctgat ccacgacgac 2100
 agcctgacct tcaaggagga catccagaag gccaggtga gcccagggg cgacagcctg 2160
 cacgagcaca tcgccaacct ggccggcagc cccgccatca agaagggcat cctgcagacc 2220

 gtgaaggtgg tggacgagct ggtgaaggtg atgggccgcc acaagcccga gaacatcgtg 2280
 atcgagatgg cccgcgagaa ccagaccacc cagaagggcc agaagaacag ccgcgagcgc 2340
 atgaagcgca tcgaggaggg catcaaggag ctgggcagcc agatcctgaa ggagcacccc 2400
 gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccgc 2460
 gacatgtacg tggaccagga gctggacatc aaccgcctga gcgactacga cgtggaccac 2520
 atcgtgcccc agagcttctt gaaggacgac agcatcgaca acaaggtgct gacccgcagc 2580
 gacaagaacc gggcaagag cgacaacgtg cccagcgagg aggtggtgaa gaagatgaag 2640

 aactactggc gccagctgct gaacgccaag ctgatcacc agcgaagtt cgacaacctg 2700
 accaaggccg agcgcggcgg cctgagcgag ctggacaagg ccggcttcat caagcggcag 2760
 ctggtggaga cccgccagat caccaagcac gtggcccaga tcctggacag ccgatgaac 2820
 accaagtacg acgagaacga caagctgatc cgcgaggtga aggtgatcac cctgaagagc 2880
 aagctggtga gcgacttccg caaggacttc cagttctaca aggtgcgca gatcaacaac 2940
 taccaccagc cccacgagc ctacctgaac gccgtggtgg gcaccgcct gatcaagaag 3000
 taccccaagc tggagagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcgcaag 3060

atgatcgcca agagcgagca ggagatcggc aaggccaccg ccaagtactt cttctacagc 3120
aacatcatga acttcttcaa gaccgagatc accctggcca acggcgagat ccgcaagcgc 3180
ccctgatcg agaccaacgg cgagaccggc gagatcgtgt gggacaaggg ccgcgacttc 3240
gccaccgtgc gcaaggtgct gagcatgccc caggtgaaca tcgtgaagaa gaccgaggtg 3300
cagaccggcg gcttcagcaa ggagagcatc ctgccaagc gcaacagcga caagctgatc 3360
gcccgcaaga aggactggga cccaagaag tacggcggct tcgacagccc caccgtggcc 3420
tacagcgtgc tgggtggggc caaggtggag aagggaaga gcaagaagct gaagagcgtg 3480

aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac 3540
ttcctggagg ccaagggcta caaggaggtg aagaaggacc tgatcatcaa gctgccaag 3600
tacagcctgt tcgagctgga gaacggccgc aagcgcagtc tggccagcgc cggcgagctg 3660
cagaaggga acgagctggc cctgcccagc aagtacgtga acttctgta cctggccagc 3720
cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag 3780
cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg 3840
atcctggcgc acgccaacct ggacaaggtg ctgagcgcct acaacaagca ccgcgacaag 3900

cccatccgcg agcaggccga gaacatcatc cacctgttca ccctgaccaa cctgggcgcc 3960
cccgccgct tcaagtactt cgacaccacc atcgaccgca agcgtacac cagcaccaag 4020
gaggtgctgg acgccacct gatecaccag agcataccg gtctgtacga gaccgcgcatc 4080
gacctgagcc agctgggcgg cgac 4104

- <210> 2
- <211> 4173
- <212> DNA
- <213> Artificial Sequence
- <220><223> Cas9-modified
- <400> 2

atggacaaga agtacagcat cggcctggac atcggtacca acagcgtggg ctgggccgtg 60
atcaccgacg agtacaaggt gcccagcaag aagttaagg tgctgggcaa caccgaccgc 120
cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag 180
gccaccgcc tgaagcgcac cgcccggcgc cgtacaccc gccgcaagaa ccgcatctgc 240
tacctgcagg agatcttcag caacgagatg gccaaaggtg acgacagctt cttccaccgc 300
ctggaggaga gcttctctgt ggaggaggac aagaagcagc agcggcacc cactcttcggc 360
aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgcgcaag 420

aagctggtgg acagcacga caaggccgac ctgcgcctga tctacctggc cctggcccac 480

atgatcaagt tccgcgcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540
gtggacaage tgttcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc 660
cgcttgaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggaac 720
ctgatcgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780
gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggaaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc 900

ctgctgagcg acatcctgcg cgtgaacacc gagatcacca aggccccct gagcgccagc 960
atgatcaagc gctacgacga gcaccaccag gacctgacct tgctgaaggc cctggtgcgc 1020
cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacatcg acggcggcgc cagccaggag gatttctaca agttcatcaa gccatcctg 1140
gagaagatgg acggcaccga ggagctgctg gtgaagctga acccgagga cctgctgcgc 1200
aagcagcgca cttcgaaa cggcagcatc ccccaccaga tccacctggg cgagctgac 1260
gccatcctgc gccccagga ggacttctac cccttctga aggacaaccg cgagaagatc 1320

gagaagatcc tgaccttcg catcccctac tacgtgggccc cctggcccg cggcaacagc 1380
cgcttcgccc ggatgacccc caagagcgag gagaccatca cccctggaa cttcgaggag 1440
gtggtggaca agggcgccc cgcccagagc ttcacgagc gcatgacaa cttcgacaag 1500
aacctgccc aagagaaggt gctgcccag cacagcctgc tgtacgagta cttaccctg 1560
tacaacgagc tgaccaaggt gaagtacgtg accgagggca tgcgcaagcc cgccttctg 1620
agcggcgagc agaagaagc catcgtggac ctgctgttca agaccaaccg caaggtgacc 1680
gtgaagcagc tgaaggagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc 1740

agcggcgtgg aggaccgctt caaccccagc ctgggcacct accacgacct gctgaagatc 1800
atcaaggaca aggacttctt ggacaacgag gagaacgagg acatcctgga ggacatcgtg 1860
ctgacctga cctgtttcga ggaccgag atgatcgagg agcgcctgaa gacctacgcc 1920
cacctgttc acgacaaggt gatgaagcag ctgaagcgc gccgctacac cggctggggc 1980
cgctgagcc gcaagcttat caacggcatc cgcgacaagc agagcggcaa gacctcctg 2040
gacttctga agagcgacgg cttcgccaac cgcaacttca tgcagctgat ccacgacgac 2100
agcctgacct tcaaggagga catccagaag gccaggtga gcggccaggc cgacagcctg 2160

cacgagcaca tcgccaacct ggccggcagc cccgccatca agaagggcat cctgcagacc 2220
 gtgaaggtgg tggacgagct ggtgaaggtg atgggccgcc acaagcccga gaacatcgtg 2280
 atcgagatgg cccgcgagaa ccagaccacc cagaagggcc agaagaacag ccgcgagcgc 2340
 atgaagcgca tcgaggaggg catcaaggag ctgggcagcc agatcctgaa ggagcacccc 2400
 gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccgc 2460
 gacatgtacg tggaccagga gctggacatc aaccgcctga gcgactacga cgtggaccac 2520
 atcgtgcccc agagcttctt gaaggacgac agcatcgaca acaaggtgct gacccgcagc 2580

 gacaagaacc gcgcaagag cgacaactg cccagcgagg aggtggtgaa gaagatgaag 2640
 aactactggc gccagctgct gaacccaag ctgatcacc agcgcaagtt cgacaacctg 2700
 accaaggccg agcgcgccgg cctgagcgag ctggacaagg ccggcttcat caagcgcag 2760
 ctggtggaga cccgccagat caccaagcac gtggcccaga tcctggacag ccgcatgaac 2820
 accaagtacg acgagaacga caagctgac cgcgaggtga aggtgatcac cctgaagagc 2880
 aagctggtga gcgacttccg caaggacttc cagtctaca aggtgcgca gatcaacaac 2940
 taccaccag cccacgacgc ctacctgaac gccgtggtgg gcaccgcctt gatcaagaag 3000

 taccccaagc tggagagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcgcaag 3060
 atgatcgcca agagcgagca ggagatcggc aaggccaccg ccaagtactt ctctacagc 3120
 aacatcatga acttcttcaa gaccgagatc accctggcca acggcgagat ccgcaagcgc 3180
 ccctgatcg agaccaacgg cgagaccggc gagatcgtgt gggacaaggg ccgcgacttc 3240
 gccaccgtgc gcaaggtgct gagcatgccc caggtgaaca tcgtgaagaa gaccgaggtg 3300
 cagaccggcg gcctcagcaa ggagagcatc ctgcccaagc gcaacagcga caagctgatc 3360
 gcccgaaga aggactggga cccaagaag tacggcggct tcgacagccc caccgtggcc 3420

 tacagcgtgc tgggtggtggc caaggtggag aagggaaga gcaagaagct gaagagcgtg 3480
 aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa cccatcgac 3540
 ttctggagg ccaagggcta caaggaggtg aagaaggacc tgatcatcaa gctgcccaag 3600
 tacagcctgt tcgagctgga gaacggccgc aagcgcagc tggccagcgc cggcgagctg 3660
 cagaaggga acgagctggc cctgcccagc aagtacgtga acttctgta cctggccagc 3720
 cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag 3780
 cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg 3840

 atcctggccg acgccaacct ggacaaggtg ctgagcgctt acaacaagca ccgcgacaag 3900
 cccatccgag agcaggccga gaacatcatc cacctgttca cctgaccaa cctgggcgcc 3960
 cccgccctt tcaagtactt cgacaccacc atcgaccgca agcgctacac cagcaccaag 4020

gaggtgctgg acgccacct gatccaccag agcatcaccg gtctgtacga gacccgcatc 4080
gacctgagcc agctgggctg cgacggcggc tccggacctc caaagaaaa gagaaaagta 4140
taccctacg acgtgcccga ctacgcctaa taa 4173

<210> 3
<211> 2145
<212> DNA

<213> Artificial Sequence
<220><223> domain 1
<400> 3

atggacaaga agtacagcat cggcctggac atcggtagca acagcgtggg ctgggccgtg 60
atcaccgacg agtacaaggt gccagcaag aagtccaagg tgctgggcaa caccgaccgc 120
cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag 180
gccacccgcc tgaagcgcac cgcccggcgc cgctacacc gccgcaagaa ccgcatctgc 240
tacctgcagg agatcttcag caacagatg gccaaagtgg acgacagctt ctccaccgc 300
ctggaggaga gcttcctggt ggaggaggac aagaagcacg agcgccaccc catcttcggc 360

aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgcgcaag 420
aagctggtgg acagcacgca caaggccgac ctgcgcctga tctacctggc cctggcccc 480
atgatcaagt tccgcgcca cttctgatc gagggcgacc tgaaccccga caacagcgac 540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc 600
atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc 660
cgctggaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggaac 720
ctgatcgccc tgagcctggg cctgaccccc aacttcaaga gcaacttca cctggccgag 780

gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggaaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc 900
ctgctgagcg acatcctgcg cgtgaacacc gagatcacca aggccccct gagcgccagc 960
atgatcaagc gtiacgacga gcaccaccag gacctgacc tgctgaaggc cctggtgcgc 1020
cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggtacgcc 1080
ggctacatcg acggcggcgc cagccaggag gatttctaca agttcatca gcccacctg 1140
gagaagatgg acggcaccga ggagctgctg gtgaagctga accgcgagga cctgctgcgc 1200

aagcagcga ccttcgaaa cggcagcatc ccccaccaga tccacctggg cgagctgcac 1260
gccatcctgc gccccagga ggacttctac cccttctga aggacaaccg cgagaagatc 1320

gagaagatcc tgaccttccg catcccctac tacgtgggcc ccctggcccg cggcaacagc 1380
 cgcttcgcct ggatgacccg caagagcgag gagaccatca cccctggaa cttcgaggag 1440
 gtggtggaca agggcgccag cgcccagagc ttcacgagc gcatgaccaa cttcgacaag 1500
 aacctgcccc acgagaaggt gctgccaag cacagcctgc tgtacgagta cttaccctg 1560
 tacaacgagc tgaccaaggt gaagtacgtg accgagggca tgcgcaagcc cgccttctg 1620

 agcggcgagc agaagaagc catcgtggac ctgctgttca agaccaaccg caagtgacc 1680
 gtgaagcagc tgaaggagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc 1740
 agcggcgtgg aggaccgctt caacgccagc ctgggcacct accacgacct gctgaagatc 1800
 atcaaggaca aggacttctt ggacaacgag gagaacgagg acatcctgga ggacatcgtg 1860
 ctgacctga ccctgttcga ggaccgagc atgatcgagg agcgcctgaa gacctacgcc 1920
 cacctgttcg acgacaaggt gatgaagcag ctgaagcgcc gccgctacac cggtggggc 1980
 cgctgagcc gcaagcttat caacggcatc cgcgacaagc agagcggcaa gaccatcctg 2040

 gacttctga agagcgacgg cttcgccaac cgcaacttca tgcagctgat ccacgacgac 2100
 agcctgacct tcaaggagga catccagaag gccaggtga gcggc 2145

 <210> 4
 <211> 2238
 <212> DNA
 <213> Artificial Sequence
 <220><223> domain 1-modified
 <400> 4

 atggtgtacc cctacgactg gcccgactac gccgaattgc ctcaaaaaa gaagagaaag 60
 gtagggatcc gaattcccgg ggaaaaaccg gacaagaagt acagcatcgg cctggacatc 120
 ggtaccaaca gcgtgggctg ggccgtgatc accgacgagt acaaggtgcc cagcaagaag 180

 ttcaaggtgc tgggcaacac cgaccgccac agcatcaaga agaacctgat cggcgcctg 240
 ctgttcgaca gggcgagac cgccgaggcc acccgctga agcgcaccgc cgcgcgccg 300
 tacaccgcc gcaagaaccg catctgctac ctgcaggaga tcttcagcaa cgagatggcc 360
 aaggtggacg acagttctt ccaccgctg gaggagagct tcttgggtga ggaggacaag 420
 aagcagagc gccaccceat cttcggaac atcgtggacg aggtggccta ccacgagaag 480
 taccaccaca tctaccact gcgcaagaag ctggtgaca gcaccgaca gcccgacctg 540
 cgctgatct acctggccct ggcccacatg atcaagtcc gcggccactt cctgatcgag 600

 ggcgacctga accccgaca cagcgactg gacaagctgt tcatccagct ggtgcagacc 660

tacaaccagc tgttcgagga gaaccccatc aacccagcg gcgtggacgc caaggccatc 720
ctgagcgcgc gcctgagcaa gagccgccgc ctggagaacc tgatcgcca gctgcccggc 780
gagaagaaga acggcctggt cggcaacctg atcgccctga gcctgggcct gacccccaac 840
ttcaagagca acttcgacct ggccgaggac gccaagctgc agctgagcaa ggacacctac 900
gacgacgacc tggacaacct gctggcccag atcgccgacc agtacgccga cctgttcctg 960
gccgccaaga acctgagcga cgccatcctg ctgagcgaca tctgcgctg gaacaccgag 1020

atcaccaagg cccccctgag cgccagcatg atcaagcgt acgacgagca ccaccaggac 1080
ctgacctgc tgaaggccct ggtgcccag cagctgccg agaagtaca ggagatcttc 1140
ttcgaccaga gcaagaacgg ctacccggc tacatcgacg gcggcgccag ccaggaggag 1200
ttctacaagt tcatcaagcc catcctggag aagatggacg gcaccgagga gctgctggtg 1260
aagctgaacc gcgaggacct gctgccaag cagcgacct tcgacaacgg cagcatcccc 1320
caccagatcc acctggcgga gctgcacgcc atcctgcgcc gccaggagga cttctacccc 1380
ttcctgaagg acaaccgca gaagatcgag aagatcctga ccttccgat ccctactac 1440

gtgggcccc tgccccggc caacagcgc ttgcctgga tgaccgcaa gagcgaggag 1500
accatcacc cctggaactt cgaggaggtg gtggacaagg gcgccagcgc ccagagcttc 1560
atcgagcga tgaccaactt cgacaagaac ctgccaacg agaaggtgct gcccaagcac 1620
agcctgctgt acgagtactt caccgtgtac aacgagctga ccaaggtgaa gtacgtgacc 1680
gagggcatgc gcaagccgc cttcctgagc ggcgagcaga agaaggccat cgtggacctg 1740
ctgttcaaga ccaaccgcaa ggtgacctg aagcagctga aggaggacta cttcaagaag 1800
atcgagtgt tcgacagcgt ggagatcagc ggctggagg accgcttcaa cgccagcctg 1860

ggcacctacc acgacctgt gaagatcatc aaggacaagg acttcctgga caacaggag 1920
aacgaggaca tctggagga catcgtgtg acctgacct tgttcgagga ccgagatg 1980
atcgaggagc gcctgaagac ctaccccac ctgttcgacg acaaggtgat gaagcagctg 2040
aagcgcgcgc gctacaccgg ctggggccgc ctgagccgca agcttatcaa cggcatccgc 2100
gacaagcaga gggcaagac catcctggac ttctgaaga gcgacggctt cgccaaccgc 2160
aacttcatgc agctgatcca cgacgacagc ctgacctca aggaggacat ccagaaggcc 2220
caggtgagcg gctaataa 2238

- <210> 5
- <211> 1959
- <212> DNA
- <213> Artificial Sequence

<220><223> domain 2

<400> 5

cagggcgaca gcctgcacga gcacatcgcc aacctggccg gcagccccgc catcaagaag	60
ggcatcctgc agaccgtgaa ggtggtggac gagctggtga aggtgatggg cgccacaag	120
cccgagaaca tcgtgatcga gatggccccg gagaaccaga ccaccagaa gggccagaag	180
aacagccgag agcgcatgaa gcgcatcgag gagggcatca aggagctggg cagccagatc	240
ctgaaggagc accccgtgga gaacaccag ctgcagaacg agaagctgta cctgtactac	300
ctgcagaacg gccgcatat gtacgtggac caggagctgg acatcaaccg cctgagcgac	360
tacgacgtgg accacatcgt gccccagagc ttctgaagg acgacagcat cgacaacaag	420
gtgctgacc gcagcgacaa gaaccgcccg aagagcgaca acgtgccag cgaggaggtg	480
gtgaagaaga tgaagaacta ctggcgccag ctgctgaacg ccaagctgat caccagcgc	540
aagttcgaca acctgaccaa ggccgagcgc ggccgctga gcgagctgga caaggccggc	600
ttcatcaagc gccagctggt ggagaccgc cagatcacca agcacgtggc ccagatcctg	660
gacagccgca tgaacaccaa gtacgacgag aacgacaagc tgatccgca ggtgaaggtg	720
atcacctga agagcaagct ggtgagcgac ttccgcaagg acttccagtt ctacaaggtg	780
cgcgagatca acaactacca ccaccccc gacgcctacc tgaacgccgt ggtgggcacc	840
gccctgatca agaagtacc caagctggag agcgagttcg tgtacggcga ctacaaggtg	900
tacgacgtgc gcaagatgat cgccaagagc gacgaggaga tcggcaagc caccgccaag	960
tactttttt acagcaacat catgaacttc ttcaagaccg agatcacct ggccaacggc	1020
gagatccgca agcgcacct gatcgagacc aacggcgaga ccggcgagat cgtgtgggac	1080
aagggccgag acttcgccac cgtgcaag gtgctgagca tgccccaggt gaacatcgtg	1140
aagaagaccg aggtgcagac cggcggcttc agcaaggaga gcatcctgcc caagcgcaac	1200
agcgacaagc tgatgcccc caagaaggac tgggaccca agaagtacgg cggcttcgac	1260
agccccaccg tggcctacag cgtgctggtg gtggccaagg tggagaagg caagagcaag	1320
aagctgaaga gcgtgaagga gctgctgggc atcacatca tggagcgag cagcttcgag	1380
aagaaccca tcgacttct ggaggccaag ggctacaagg aggtgaagaa ggacctgatc	1440
atcaagctgc ccaagtacag cctgttcgag ctggagaacg gccgcaagc catgctggcc	1500
agcgcggcag agctgcagaa gggcaacgag ctggccctgc ccagcaagta cgtgaacttc	1560
ctgtacctgg ccagccacta cgagaagctg aagggcagcc ccgaggacaa cgagcagaag	1620
cagctgttcg tggagcagca caagcactac ctggacgaga tcatcgagca gatcagcgag	1680
ttcagcaagc gcgtgatcct ggccgacgcc aacctggaca aggtgctgag gcctacaac	1740

aagcaccgcg acaagcccat ccgcgagcag gccgagaaca tcatccacct gttcacctg 1800
 accaactgg gcgccccgc gccttcaag tacttcgaca ccaccatga ccgaagcgc 1860
 tacaccagca ccaaggaggt gctggacgcc acctgatcc accagagcat caccggtctg 1920
 tacgagacc gcatcgacct gagccagctg ggcggcgac 1959

<210> 6
 <211> 2031
 <212> DNA

<213> Artificial Sequence
 <220><223> domain 2-modified
 <400> 6

atgcaggcgc acagcctgca cgagcacatc gccaacctgg ccggcagccc cgccatcaag 60
 aaggcatcc tgcagacct gaagtggtg gacgagctgg tgaagtgat gggccgccac 120
 aagcccgaga acatcgtgat cgagatggcc cgcgagaacc agaccacca gaaggccag 180
 aagaacagcc gcgagcgcga gaagcgcac gaggaggcca tcaaggagct gggcagccag 240
 atcctgaagg agcaccctgt ggagaacacc cagctgcaga acgagaagct gtacctgtac 300
 tacttcgaga acggccgcga catgtactgt gaccaggagc tggacatcaa ccgctgagc 360

gactacgacg tggaccacat cgtgccccag agcttctga aggacgacag catcgacaac 420
 aagtgctga cccgagcga caagaaccgc ggcaagagcg acaactgcc cagcgaggag 480
 gtggtgaaga agatgaagaa ctactggcgc cagctgctga acgccaagct gatcaccag 540
 cgcaagtctg acaacctgac caaggccgag cgcggcggcc tgagcgagct ggacaaggcc 600
 ggcttcatca agcgcacct ggtggagacc cgcagatca ccaagcacgt ggcccagatc 660
 ctggacagcc gcatgaacac caagtacgac gagaacgaca agctgatccg cgaggtgaag 720
 gtgatcacc tgaagagcaa gctggtgagc gacttccga aggacttcca gttctacaag 780

gtgcgcgaga tcaacaacta ccaccagcc cacgacgct acctgaacgc cgtggtgggc 840
 accgcctga tcaagaagta cccaagctg gagagcgagt tcgtgtacgg cgactacaag 900
 gtgtacgacg tgcgaagat gatcgccaag agcagcagg agatcggcaa ggccaccgcc 960
 aagtacttct tctacagcaa catcatgaac ttcttcaaga ccgagatcac cctggccaac 1020
 ggcgagatcc gcaagcgcct cctgatcgag accaacggcg agaccggcga gatcgtgtgg 1080
 gacaagggcc gcgacttcgc caccgtgctg aagtgctga gcatcccca ggtgaacatc 1140
 gtgaagaaga ccgaggtgca gaccggcggc ttacgcaagg agagcatcct gcccaagcgc 1200

aacagcgaca agctgatcgc ccgcaagaag gactgggacc ccaagaagta cggcggcttc 1260

gacagcccca ccgtggccta cagcgtgctg gtggtggcca aggtggagaa gggcaagagc 1320
aagaagctga agagcgtgaa ggagctgctg ggcatcacca tcatggagcg cagcagcttc 1380
gagaagaacc ccatcgactt cctggaggcc aagggctaca aggaggtgaa gaaggacctg 1440
atcatcaagc tgcccaagta cagcctgttc gagctggaga acggccgcaa gcgcatgctg 1500
gccagcgccg gcgagctgca gaagggcaac gagctggccc tgcccagcaa gtacgtgaac 1560
ttcctgtacc tggccagcca ctacgagaag ctgaaggga gccccgagga caacgagcag 1620

aagcagctgt tcgtggagca gcacaagcac tacctggacg agatcatcga gcagatcagc 1680
gagttcagca agcgcgtgat cctggccgac gccaacctgg acaaggtgct gagcgcctac 1740
aacaagcacc gcgacaagcc catccgagcag caggccgaga acatcatcca cctgttcacc 1800
ctgaccaacc tgggcgcccc cgccgccttc aagtacttcg acaccacat cgaccgcaag 1860
cgctacacca gcaccaagga ggtgctggac gccaccctga tccaccagag catcaccggt 1920
ctgtacgaga cccgcatcga cctgagccag ctgggcggcg acggcggctc cggacctcca 1980
aagaaaaaga gaaaagtata ccctacgac gtgcccgact acgcctaata a 2031

<210> 7

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> nuclear localization signal

<400> 7

ccaaagaaaa agagaaaagt a 21

<210> 8

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> HA tag

<400> 8

taccctacg acgtgccga ctacgcc 27

<210> 9

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> linker

<400> 9

agcggccagg gc

12

<210> 10

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> linker

<400> 10

Ser Gly Gln Gly

1

<210> 11

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> CCR5 exon 2 target sequence

<400> 11

tgacatcaat tattatacat cgg

23

<210> 12

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> HPRT exon 3 target sequence

<400> 12

gcccccttg agcacacaga ggg

23

<210> 13

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> DMD exon 51 target sequence

<400> 13

tcctactcag actgttactic tgg

23

<210> 14

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> CCR5 exon 2 target forward primer for T7E1 assay PCR

<400> 14

ctccatggtg ctatagagca 20

<210> 15

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> CCR5 exon 2 target reverse primer for T7E1 assay PCR

<400> 15

gcctgtcaa gattgacac 20

<210> 16

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> CCR5 exon 2 target forward primer for T7E1 assay nested PCR

<400> 16

gagccaagct ctccatctag t 21

<210> 17

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> CCR5 exon 2 target reverse primer for T7E1 assay nested PCR

<400> 17

gcctgtcaa gattgacac 20

<210> 18

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> CCR5 exon 2 target forward primer for deep sequencing PCR

<400> 18

ctccatggtg ctatagagca 20

<210> 19

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> CCR5 exon 2 target reverse primer for deep sequencing PCR
 <400> 19
 gccctgtcaa gagttgacac 20

<210> 20
 <211> 59
 <212> DNA
 <213> Artificial Sequence
 <220><223> CCR5 exon 2 target forward primer for deep sequencing nested PCR
 <400> 20
 acactctttc cctacacgac gctcttccga tctaagatca ctttttattt atgcacagg 59

<210> 21
 <211> 54
 <212> DNA
 <213> Artificial Sequence
 <220><223> CCR5 exon 2 target reverse primer for deep sequencing nested PCR
 <400> 21
 gtgactggag ttcagacgtg tgctcttccg atctatgttg cccacaaaac caaa 54

<210> 22
 <211>
 > 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> HPRT exon 3 target forward primer for T7E1 assay PCR
 <400> 22
 ccaggttggt gtggaagttt 20

<210> 23
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> HPRT exon 3 target reverse primer for T7E1 assay PCR
 <400> 23

ggacttttga ctccccacaa 20

<210> 24

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> HPRT exon 3 target forward primer for T7E1 assay nested PCR

<400> 24

ccaggttggt gtggaagttt 20

<210> 25

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> HPRT exon 3 target reverse primer for T7E1 assay nested PCR

<400> 25

tggtttgcag agattcaaag aa 22

<210> 26

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> HPRT exon 3 target forward primer for deep sequencing PCR

<400> 26

ccaggttggt gtggaagttt 20

<210> 27

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> HPRT exon 3 target reverse primer for deep sequencing PCR

<400> 27

ggacttttga ctccccacaa 20

<210> 28

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> HPRT exon 3 target forward primer for deep sequencing nested PCR
 <400> 28
 acactctttc cctacacgac gctcttccga tcttgctcga gatgtgatga agg 53

<210> 29
 <211> 56
 <212> DNA
 <213> Artificial Sequence

<220><223> HPRT exon 3 target reverse primer for deep sequencing nested PCR
 <400> 29
 gtgactggag ttcagacgtg tgctcttccg atctaagaaa acctactgtt gccact 56

<210> 30
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220><223> DMD exon 51 target forward primer for T7E1 assay PCR
 <400> 30
 acttgtccag gcatgagaat gagca 25

<210> 31
 <211> 25
 <
 212> DNA
 <213> Artificial Sequence

<220><223> DMD exon 51 target reverse primer for T7E1 assay PCR
 <400> 31
 gctgcgtagt gccaaaacaa acagt 25

<210> 32
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220><223> DMD exon 51 target forward primer for T7E1 assay nested PCR
 <400> 32
 ccagcatga gaatgagcaa aatcg 25

<210> 33
 <211> 24

<212> DNA
 <213> Artificial Sequence

<220><223> DMD exon 51 target reverse primer for T7E1 assay nested PCR
 <400> 33
 ggtaagttct gtccaagccc gggt 24
 <210> 34
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220><223> DMD exon 51 target forward primer for deep sequencing PCR
 <400> 34
 acttgccag gcatgagaat gagca 25
 <210> 35
 <211> 25
 <212> DNA
 <213> Artificial Sequence

<220><223> DMD exon 51 target reverse primer for deep sequencing PCR
 <400> 35
 gctgcgtagt gccaaaacaa acagt 25
 <210> 36
 <211> 56
 <212> DNA
 <213> Artificial Sequence

<220><223> DMD exon 51 target forward primer for deep sequencing nested PCR
 <400> 36
 acactctttc cctacacgac gctcttccga tctggctctt tagcttgtgt ttctaa 56
 <210> 37
 <211> 58
 <212> DNA
 <213> Artificial Sequence

<220><223> DMD exon 51 target reverse primer for deep sequencing nested PCR
 <400> 37

gtgactggag ttcagacgtg tgctcttccg atctggtaag ttctgtccaa gcccggtt 58

<210> 38
 <211> 57
 <212> DNA
 <213> Artificial Sequence
 <220><223> On-target (HBB gene) forward primer for deep sequencing nested
 PCR
 <400> 38

acactctttc cctacacgac gctcttccga tctgtcttg taacttgat accaacc 57

<210> 39
 <211> 54
 <212> DNA
 <213> Artificial Sequence
 <220><223> On-target (HBB gene) reverse primer for deep sequencing nested
 PCR
 <400> 39

gtgactggag ttcagacgtg tgctcttccg atctgcaacc tcaaacagac acca 54

<210> 40
 <211> 53
 <212> DNA
 <213> Artificial Sequence
 <220><223> OT1 target forward primer for deep sequencing nested PCR
 <400> 40

acactctttc cctacacgac gctcttccga tctgaaaggg gaagatcca gag 53

<210> 41
 <211> 54
 <212> DNA
 <213> Artificial Sequence
 <220><223> OT1 target reverse primer for deep sequencing nested PCR
 <400> 41

gtgactggag ttcagacgtg tgctcttccg atctatttcc aggetatget teca 54

<210> 42
 <211> 53

<212>

DNA

<213> Artificial Sequence

<220><223> OT3 target forward primer for deep sequencing nested PCR

<400> 42

acactctttc cctacacgac gctcttccga tcttttgtgt gggatgctga gag 53

<210> 43

<211> 56

<212> DNA

<213> Artificial Sequence

<220><223> OT3 target reverse primer for deep sequencing nested PCR

<400> 43

gtgactggag ttcagacgtg tgctcttccg atctagctac cacggtgaca gtaaca 56

<210> 44

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> HBB_48 target forward primer for deep sequencing nested PCR

<400> 44

acactctttc cctacacgac gctcttccga tctaggaggt gagagtccag tcg 53

<210> 45

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> HBB_48 target reverse primer for deep sequencing nested PCR

<400> 45

gtgactggag ttcagacgtg tgctcttccg atctgaacca gctgctcag actt 54

<210> 46

<211> 53

<212> DNA

<213> Artificial Sequence

<220><223> HBB_75 target forward primer for deep sequencing nested PCR

<400> 46

acactctttc cctiacacgac gctcttccga tcttgggtgg tagatgagga tga 53

<210> 47

<211> 54

<212> DNA

<213> Artificial Sequence

<220><223> HBB_75 target reverse primer for deep sequencing nested PCR

<400> 47

gtgactggag ttcagacgtg tgctcttccg atctcctggc aaaagtgttt ggat 54

<210> 48

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> On-target sequence (HBB gene)

<400> 48

cttgccccac agggcagtaa cgg 23

<210> 49

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> off-target sequence (OT1)

<400> 49

tcagccccac agggcagtaa ggg 23

<210> 50

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> off-target sequence (OT3)

<400> 50

gctgccccac agggcagcaa agg 23

<210> 51

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> off-target sequence (HBB_48)

<400> 51

attgccccac ggggcagtga cgg 23

<210> 52

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> off-target sequence (HBB_75)

<400> 52

gtggccccac agggcaggaa tgg 23

<210> 53

<211> 50

<212> DNA

<213> Artificial Sequence

<220><223> Splicing donor in adeno-associated virus vector carrying domain 1
of spCas9

<400> 53

gtaagtatca aggttacaag acaggtttaa ggagaccaat agaaactggg 50

<210> 54

<211> 83

<212> DNA

<213> Artificial Sequence

<220><223> Splicing acceptor in adeno-associated virus vector carrying
domain 2 of spCas9

<400> 54

cttgtcgaga cagagaagac tcttgcgttt ctgataggca cctattggtc ttactgacat 60

ccactttgcc tttctctcca cag 83