

(52) CPC특허분류

C12N 15/907 (2013.01)

C12N 9/22 (2013.01)

명세서

청구범위

청구항 1

고특이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA를 포함하고,

상기 고특이성 Cas9 변이체는 Cas9 단백질의 알라닌이 아닌 하나 이상의 아미노산 잔기가 알라닌으로 치환된 것이고,

상기 매칭된 5' 뉴클레오타이드는 표적 서열의 5' 말단 뉴클레오타이드와 매칭되는 염기를 포함하는 뉴클레오타이드인,

유전자 교정용 조성물.

청구항 2

제1항에 있어서, 상기 고특이성 Cas9 변이체는 *Streptococcus pyogenes* Cas9 단백질의 K848, K1003, R1060, N497, R661, Q695, 및 Q926로 이루어진 군에서 선택된 하나 이상의 아미노산이 알라닌으로 치환된 Cas9 변이체인, 유전자 교정용 조성물.

청구항 3

제2항에 있어서, 상기 고특이성 Cas9 변이체는 *Streptococcus pyogenes* Cas9 단백질에 K848A, K1003A, 및 R1060A 변이가 도입된 eCas9-1.1, 또는 N497A, R661A, Q695A, 및 Q926A 변이가 도입된 Cas9-HF1인, 유전자 교정용 조성물.

청구항 4

제1항에 있어서, 상기 가이드 RNA는 하기의 서열 일반식으로 표현되는 sgRNA (single-guide RNA)인, 유전자 교정용 조성물:

5'-(N_{Cas9})₁-(서열번호 1)-(올리고뉴클레오타이드 링커)-(서열번호 3)-3'

상기 서열 일반식에서, (N_{Cas9})₁는 표적 서열과 혼성화 가능 표적화 서열이고, 1은 상기 표적화 서열에 포함된 뉴클레오타이드 수로서 15 내지 30의 정수이고,

상기 올리고뉴클레오타이드 링커는 A, U, C 및 G로 이루어진 군에서 각각 독립적으로 선택된 3 내지 5개의 뉴클레오타이드를 포함함.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 진핵 세포 또는 진핵 유기체에 사용하기 위한 것인, 유전체 교정용 조성물.

청구항 6

제1항 내지 제4항 중 어느 한 항의 유전자 교정용 조성물을 분리된 진핵 세포 또는 인간을 제외한 진핵 유기체에 투여하는 단계를 포함하는, 유전체 교정 방법.

청구항 7

(1) sgRNA 및

(2) 상기 sgRNA의 5' 말단, 3' 말단, 또는 양 말단에 융합된 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 포함하는, 융합 RNA 분자.

청구항 8

제7항에 있어서, 상기 자가-절단 활성을 갖는 RNA 절단효소는 해머헤드 리보자임, VS (Varkud satellite) 리보자임, 리드자임 (Leadzyme), 및 헤어핀 리보자임 (hairpin ribozyme)으로 이루어진 군에서 선택된 1종 이상인, 융합 RNA 분자.

청구항 9

sgRNA의 5' 말단, 3' 말단, 또는 양 말단에 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 융합시키는 단계를 포함하는, 표적 서열과 매칭된 5' 말단 뉴클레오타이드를 포함하는 가이드 RNA의 제조 방법.

청구항 10

제9항에 있어서, 상기 융합시키는 단계는 sgRNA의 암호화 DNA와 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 암호화하는 DNA를 하나의 벡터에서 발현시키는 단계를 포함하는 것인, 표적 서열과 매칭된 5' 말단 뉴클레오타이드를 포함하는 가이드 RNA의 제조 방법.

발명의 설명

기술 분야

[0001] 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA를 이용하여 고틱이성 Cas9 변이체의 유전자 교정 효율을 증진시키는 기술과 관련된 것으로, 고틱이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA의 복합체, 고틱이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 유전자 교정용 조성물; 및 유전자 교정용 조성물을 이용하는 유전자 교정 방법이 제공된다.

배경 기술

[0003] 박테리아와 고세균 내의 적응 면역 시스템 (adaptive immune system)에서 유래한 CRISPR-Cas9 RNA-가이드 엔도뉴클레아제는 다양한 세포 및 유기체 내에서의 표적화된 유전체 교정 용도로 용도 변경되었다. 이들 CRISPR-Cas9 RNA-가이드 엔도뉴클레아제는 염색체 DNA를 표적화된 방식으로 절단하여 부위특이적 DNA 이중 가닥 절단 (DSB; double strand break)을 생성하며, 비상동성 말단 결합 (NHEJ; non-homologous end-joining)을 통한 repair는 표적 부위에서 삽입(insertion) 또는 결실(deletion) (indels)을 유도한다. 불행히도, 표적 부위와 서열 상동성이 높은 부위에서의 표적 DNA 절단은 원하지 않는 유전체 자리에서의 돌연변이 및 염색체 재배열을 유도할 수 있다 (off-target effect). *S. pyogenes* Cas9 뉴클레아제 및 가이드 RNA (sgRNA) 모두 이러한 off-target effect를 최소화하거나 제거하기 위한 변형이 가해져 왔다. 특히, 인간 세포에서 최소화 내지는 탐지되지 않는 정도로 낮은 off-target effect를 갖는 고틱이성 Cas9 변이체가 개발되었으며, 그 예로 enhanced Cas9-1.1 (eCas9-1.1)(Slaymaker, I.M. et al. Rationally engineered Cas9 nucleases with improved specificity. *Science* **351**, 84-88 (2016)) 및 Cas9 high-fidelity variant 1 (Cas9-HF1) (Kleinstiver, B.P. et al. High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529**, 490-495 (2016))가 있다. 이들 고틱이성 Cas9 변이체들은 알려진 치환을 포함하여 Cas9 단백질과 비-표적 또는 표적 DNA 가닥 사이의 비특이적인 이온 상호작용을 약화시킨다.

[0004] 이와 같은 고틱이성 Cas9 변이체들의 유전자 교정 활성 및 특이성을 보다 증진시키기 위한 기술의 개발이 요구된다.

발명의 내용

해결하려는 과제

[0006] 본 명세서는 고틱이성 Cas9 변이체들의 오프-타겟 부위에 대한 낮은 유전자 교정 활성을 유지하면서 온-타겟 특이적 유전자 교정 활성을 보다 증진시킬 수 있는 기술을 제공한다. 보다 구체적으로, 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA (예컨대, sgRNA)를 사용하여 Cas9 변이체의 유전자 교정 효율 (예컨대, 인델 (indel) 빈도)을 증가시키는 기술을 제공한다.

[0007] 일 예는 고틱이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA(예컨대, sgRNA)의 복합체를

제공한다.

- [0008] 다른 예는 고틱이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA(예컨대, sgRNA)를 포함하는 유전자 교정용 조성물을 제공한다.
- [0009] 다른 예는 고틱이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA(예컨대, sgRNA)를 사용하는 유전자 교정 방법을 제공한다. 예컨대, 상기 방법은 상기 고틱이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA(예컨대, sgRNA)를 표적 유전자 또는 상기 표적 유전자 내에 위치하는 표적 부위 (PAM 서열을 포함하는, 10 내지 30개 뉴클레오타이드, 10 또는 25개 뉴클레오타이드, 15 내지 30개 뉴클레오타이드, 15 또는 25개 뉴클레오타이드, 17 내지 30개 뉴클레오타이드, 또는 17 또는 25개 뉴클레오타이드, 예컨대 20개 뉴클레오타이드)와 접촉시키는 단계를 포함할 수 있으며, 상기 접촉시키는 단계는 상기 복합체를 대상에 투여, 주입, 또는 도입함으로써 수행되는 것일 수 있다.
- [0010] 상기 고틱이성 Cas9 변이체는 하나 이상의 아미노산이 알라닌으로 치환되어 표적 부위에 대한 특이성이 증진된 Cas9 변이체를 의미하는 것으로, 예컨대, *Streptococcus pyogenes* Cas9 단백질의 아미노산 서열 (서열번호 4)을 기준으로, K848, K1003, R1060, N497, R661, Q695, 및 Q926로 이루어진 군에서 선택된 하나 이상의 아미노산이 알라닌으로 치환된 Cas9 변이체를 의미하는 것일 수 있다. 일 구체예에서, 상기 고틱이성 Cas9 변이체는 *Streptococcus pyogenes* Cas9 단백질(서열번호 4)에 K848A, K1003A, 및 R1060A 변이가 도입된 eCas9-1.1, 또는 N497A, R661A, Q695A, 및 Q926A 변이가 도입된 Cas9-HF1, 또는 이들의 조합일 수 있다.
- [0011] 본 명세서에서, 가이드 RNA서 매칭된 5' 뉴클레오타이드는 각 가이드 RNA가 표적하는 표적 서열의 가장 5' 말단에 위치하는 뉴클레오타이드 (5' 뉴클레오타이드라 칭함)와 매칭되는 (일치하는) 염기를 포함하는 뉴클레오타이드를 의미한다.
- [0012] 다른 예는 상기 유전자 교정 방법에 의하여 교정된 유전자를 포함하는 유전자 변형 세포를 제공한다.
- [0013] 다른 예는 상기 유전자 변형 세포로부터 얻어진 유전자 변형 동물을 제공한다.
- [0014] 상기 방법, 복합체 및 조성물은 진핵 세포 (예컨대, 분리된 인간 세포, 또는 인간을 제외한 진핵 동물 또는 진핵 식물의 세포) 또는 진핵 유기체 (예컨대, 인간, 또는 인간을 제외한 인간을 제외한 진핵 동물 또는 진핵 식물)에 적용되는 것일 수 있다.
- [0015] 다른 예는 (1) 가이드 RNA (예컨대, sgRNA) 및 (2) 상기 가이드 RNA의 5' 말단에 융합된 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 포함하는, 융합 RNA 분자를 제공한다.
- [0016] 다른 예는 가이드 RNA (예컨대, sgRNA)의 5' 말단, 3' 말단, 또는 양 말단에 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 융합시키는 단계를 포함하는, 표적 서열과 매칭된 5' 말단 뉴클레오타이드를 포함하는 가이드 RNA의 제조 방법을 제공한다. 상기 제조 방법에서 융합시키는 단계는 가이드 RNA (예컨대, sgRNA)의 암호화 DNA와 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 암호화하는 DNA를 하나의 벡터에서 발현시키는 단계를 포함할 수 있다.
- [0017] 상기 자가-절단 활성을 갖는 RNA 절단효소(리보자임)은 해머헤드 리보자임 (hammerhead ribozyme; 예컨대, Type I hammerhead ribozyme, Type II hammerhead ribozyme, TypeIII hammerhead ribozyme 등), VS (Varkud satellite) 리보자임, 리드자임 (Leadzyme), 헤어핀 리보자임 (hairpin ribozyme) 등으로 이루어진 군에서 선택된 1종 이상일 수 있으나, 이에 제한되는 것은 아니다.

과제의 해결 수단

- [0019] 진핵 세포에서 sgRNA를 발현하는데 일반적으로 사용되는 U6 프로모터는 전사를 개시하기 위하여 구아노신(G) 뉴클레오타이드를 필요로 하기 때문에 sgRNA는 전형적으로 5' 말단에 "G" 뉴클레오타이드를 포함한다. 대부분의 (평균 75%) DNA 표적 부위는 이 위치에서 미스매치 뉴클레오타이드 (즉 G가 아닌 염기 (A, T, 또는 C)를 포함하는 뉴클레오타이드)를 포함한다. 본 발명자들은 고틱이성 Cas9 변이체를 5' 말단에 G를 포함하는 가이드 RNA와 함께 5' 말단에 G를 포함하지 않는 표적 서열에 사용하는 경우, 5' 말단에 가이드 RNA와 표적 서열 간 미스매치가 발생하여, 상기 표적 서열 부위에서의 상기 고틱이성 Cas9 변이체에 의한 유전자 교정 효율이 낮아짐을 최초로 확인하여, 고틱이성 Cas9 변이체의 유전자 교정 효율에 있어서 가이드 RNA의 5' 말단과 표적 서열의 5' 말단의 매칭 여부가 중요한 역할을 가짐을 제안한다.

- [0020] 또한, 자기-절단 리보자임 (self-cleaving ribozyme)에 연결되어 생산된 매칭된 5' 뉴클레오타이드를 포함하는 sgRNA를 사용함으로써, 인간을 포함한 진핵 세포에서, 상기 고특이성 Cas9 변이체가 오프-타겟 부위에 낮은 활성을 보이는 고특이성(high-specificity)을 회생시키지 않으면서 온-타겟 부위에서의 유전자 교정 활성을 현저하게 향상시킬 수 있음을 확인하였다.
- [0021] 이에, 본 발명의 일 예는 고특이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA(예컨대, sgRNA)의 복합체를 제공한다.
- [0022] 다른 예는 고특이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA(예컨대, sgRNA)를 포함하는 유전자 교정용 조성물을 제공한다.
- [0023] 다른 예는 고특이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA(예컨대, sgRNA)를 사용하는 유전자 교정 방법을 제공한다. 예컨대, 상기 방법은 고특이성 Cas9 변이체 및 매칭된 5' 뉴클레오타이드를 포함하는 가이드 RNA(예컨대, sgRNA)를 표적 유전자 또는 상기 표적 유전자 내에 위치하는 표적 부위 (PAM 서열을 포함하는, 13 내지 30개 뉴클레오타이드, 13 또는 25개 뉴클레오타이드, 15 내지 30개 뉴클레오타이드, 15 또는 25개 뉴클레오타이드, 20 내지 30개 뉴클레오타이드, 또는 20 또는 25개 뉴클레오타이드)와 접촉시키는 단계를 포함할 수 있으며, 상기 접촉시키는 단계는 상기 복합체를 대상 (진핵 세포 또는 진핵 유기체)에 투여, 주입, 또는 도입함으로써 수행되는 것일 수 있다.
- [0024] 상기 고특이성 Cas9 변이체는 알라닌이 아닌 하나 이상의 아미노산 잔기가 알라닌으로 치환되어 표적 부위에 대한 특이성이 증진된 Cas9 변이체를 의미하는 것으로, 본 명세서에서 고충실도 (high-fidelity) Cas9 변이체라고도 명명된다. 예컨대, 상기 고특이성 Cas9 변이체는 *Streptococcus pyogenes* Cas9 단백질의 아미노산 서열 (서열번호 4)을 기준으로, K848, K1003, R1060, N497, R661, Q695, 및 Q926로 이루어진 군에서 선택된 하나 이상의 아미노산이 알라닌으로 치환된 Cas9 변이체를 의미하는 것일 수 있다. 일 구체예에서, 상기 고특이성 Cas9 변이체는 *Streptococcus pyogenes* Cas9 단백질(서열번호 4)에 K848A, K1003A, 및 R1060A 변이가 도입된 eCas9-1.1, 또는 N497A, R661A, Q695A, 및 Q926A 변이가 도입된 Cas9-HF1, 또는 이들의 조합일 수 있다.
- [0025] 본 명세서에서, 가이드 RNA의 매칭된 5' 뉴클레오타이드는 가이드 RNA가 표적하는 표적 서열의 가장 5' 말단에 위치하는 뉴클레오타이드 (5' 뉴클레오타이드라 칭함)와 부합하는 (PAM 서열이 위치하는 가닥의 표적 서열의 5' 뉴클레오타이드와 일치하는) 염기를 포함하는 뉴클레오타이드를 의미한다. 또한 미스매치 5' 뉴클레오타이드는 가이드 RNA가 표적하는 표적 서열의 5' 뉴클레오타이드와 부합하지 않는 (PAM 서열이 위치하는 가닥의 표적 서열의 5' 뉴클레오타이드와 불일치하는) 염기를 포함하는 뉴클레오타이드를 의미한다.
- [0026] 다른 예는 상기 유전자 교정 방법에 의하여 교정된 유전자를 포함하는 유전자 변형 세포를 제공한다.
- [0027] 다른 예는 상기 유전자 변형 세포로부터 얻어진 유전자 변형 동물을 제공한다.
- [0028] 상기 방법, 복합체 및 조성물은 진핵 세포 (예컨대, 분리된 인간 세포, 또는 인간을 제외한 진핵 동물 또는 진핵 식물의 세포) 또는 진핵 유기체 (예컨대, 인간, 또는 인간을 제외한 인간을 제외한 진핵 동물 또는 진핵 식물)에 적용되는 것일 수 있다.
- [0029] 다른 예는 (1) 가이드 RNA (예컨대, sgRNA) 및 (2) 상기 가이드 RNA의 5' 말단에 융합된 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 포함하는, 융합 RNA 분자를 제공한다.
- [0030] 다른 예는 가이드 RNA (예컨대, sgRNA)의 5' 말단, 3' 말단, 또는 양 말단에 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 융합시키는 단계를 포함하는, 표적 서열과 매칭된 5' 말단 뉴클레오타이드를 포함하는 가이드 RNA의 제조 방법을 제공한다. 상기 제조 방법에서 융합시키는 단계는 가이드 RNA (예컨대, sgRNA)의 암호화 DNA와 자가-절단 활성을 갖는 RNA 절단효소 또는 1 내지 6개의 tRNA를 암호화하는 DNA를 하나의 벡터에서 발현시키는 단계를 포함할 수 있다.
- [0031] 상기 자가-절단 활성을 갖는 RNA 절단효소(리보자임)은 해머헤드 리보자임 (hammerhead ribozyme; 예컨대, Type I hammerhead ribozyme, Type II hammerhead ribozyme, TypeIII hammerhead ribozyme 등), VS (Varkud satellite) 리보자임, 리드자임 (Leadzyme), 헤어핀 리보자임 (hairpin ribozyme) 등으로 이루어진 군에서 선택된 1종 이상일 수 있으나, 이에 제한되는 것은 아니다.
- [0032] 다른 예는 매칭된 5' 말단 뉴클레오타이드를 포함하는 가이드 RNA를 상기 고특이성 Cas9 변이체와 함께 진핵 세포 또는 진핵 유기체에 도입시키는 단계를 포함하는, 상기 고특이성 Cas9 변이체의 유전자 교정 효율 증진 방법을 제공한다. 상기 유전자 교정 효율 증진은 온-타겟 부위에서의 유전자 교정 효율 (예컨대, indel 빈도) 증가

및/또는 오프-타겟 부위에서의 유전자 교정 효율 (예컨대, indel 빈도) 감소를 의미하는 것일 수 있다.

- [0033] 상기 진핵 세포는 분리된 인간 세포, 또는 인간을 제외한 진핵 동물 세포 또는 진핵 식물의 세포일 수 있고, 상기 진핵 유기체는 인간, 또는 인간을 제외한 인간을 제외한 진핵 동물 또는 진핵 식물일 수 있다.
- [0034] 본 명세서에 사용된 바로서, 용어 "유전자 교정 (gene editing)"은 표적 유전자 내의 표적 부위에 이중가닥 절단 (double-stranded DNA cleavage)을 발생시켜서 하나 이상의 뉴클레오타이드의 변이 (결실, 치환, 및/또는 삽입 등)를 유발하는 작용을 의미한다. 일 예에서, 상기와 같은 유전자 교정은 표적 부위에 종료코돈을 생성시키거나, 야생형과 다른 아미노산을 코딩하는 코돈을 생성시킴으로써, 표적 유전자를 불활성화 (knock-out)시키거나, 단백질을 생성하지 않는 비코딩 DNA 서열에 변이를 도입하는 등 다양한 형태일 수 있으나, 이에 제한되는 것은 아니다.
- [0035] 본 명세서에서, 상기 유전자 교정은 생체 외 (in vitro) 또는 생체 내 (in vivo)에서 수행되는 것일 수 있다.
- [0036] 본 명세서에 사용된 바로서, 용어 "염기 서열"은 해당 염기를 포함하는 뉴클레오타이드의 서열을 의미하는 것으로, 뉴클레오타이드 서열 또는 핵산 서열과 동일한 의미로 사용될 수 있다.
- [0037] 본 명세서에 사용된 바로서,
- [0038] 용어 '표적 유전자 (target gene)'는 유전자 교정의 대상이 되는 유전자를 의미하고,
- [0039] 용어 '표적 부위 (target site or target region)'는 표적 유전자 내의 Cas9에 의한 유전자 교정이 일어나는 부위를 의미하는 것으로, 일 예에서 표적 유전자 내의 Cas9이 인식하는 서열 (PAM 서열)의 5' 말단 및/또는 3' 말단에 인접하여 위치하고, 최대 길이가 약 50bp 또는 약 40bp인 유전자 부위 (이중 가닥 또는 이중 가닥 중 어느 하나의 단일 가닥)를 의미하고,
- [0040] '표적 서열 (target sequence)'는 표적 유전자 또는 표적 유전자의 표적 부위 내의 가이드 RNA가 혼성화하는 약 15 내지 약 30개, 약 15 내지 약 35개, 약 17 내지 약 23개, 또는 약 18개 내지 약 22개, 예컨대, 약 20개의 뉴클레오타이드(nt)를 포함하는 부위의 염기서열일 수 있다.
- [0041] 또한, 가이드 RNA에 포함된 용어 '표적화 서열 (targeting sequence)'은 표적 부위 내의 연속하는 약 15 내지 약 30개, 약 15 내지 약 35개, 약 17 내지 약 23개, 또는 약 18개 내지 약 22개, 예컨대, 약 20개의 뉴클레오타이드(nt)를 포함하는 부위의 염기서열과 상보적인 염기서열을 포함하는(혼성화 가능한) 가이드 RNA의 부위일 수 있다. 상기 표적화 서열과 상보적인 염기서열을 포함하는 표적 부위의 염기서열을 '표적 서열 (target sequence)'이라고 칭할 수 있으며, 상기 표적 서열은 RNA-가이드 뉴클레아제가 인식하는 PAM 서열의 5' 말단 및/또는 3' 말단에 인접하여 위치하는 연속하는 약 15nt 내지 약 30nt, 약 15nt 내지 약 25nt, 약 17nt 내지 약 23nt, 또는 약 18nt 내지 약 22 nt, 예컨대, 약 20nt 길이의 염기서열을 의미할 수 있다.
- [0042] 상기 Cas9 단백질은 표적 유전자의 특정 서열 (PAM)을 인식하고 뉴클레오타이드 절단 활성을 가져 표적 유전자에서 인델 (insertion and/or deletion, Indel)을 야기할 수 있는 모든 Cas9들 중에서 선택된 1종 이상일 수 있다.
- [0043] 상기 Cas9 단백질은 원핵 세포, 및/또는 인간 세포를 비롯한 동식물 세포 (예컨대, 진핵 세포)의 유전체에서 특정 염기서열을 인식해 이중나선절단 (double strand break, DSB)을 일으킬 수 있다. 상기 이중나선절단은 DNA의 이중 나선을 잘라, 둔단 (blunt end) 또는 점착종단 (cohesive end)을 생성시킬 수 있다. DSB는 세포 내에서 상동재조합 (homologous recombination) 또는 비상동재조합 (non-homologous end-joining, NHEJ) 기작에 의해 효율적으로 수선될 수 있는데, 이 과정에 소망하는 변이를 표적 위치에 도입할 수 있다.
- [0044] 상기 Cas9 단백질은 유전체 DNA의 표적 부위로 안내하기 위한 표적 DNA 특이적 가이드 RNA와 함께 사용된다. 상기 가이드 RNA는 생체 외 (in vitro) 또는 세포 외에서 전사(transcribed; 예컨대 올리고뉴클레오타이드 이중가닥 또는 플라스미드 주형으로부터 전사)되거나, 생체 내 또는 세포 내에서 재조합 벡터 (발현 벡터)에 의하여 재조합적으로 생산된 것일 수 있으나, 이에 제한되지 않는다. 상기 Cas9 단백질은, 생체 (또는 세포) 외에서 또는 생체(세포) 내 전달 후, 가이드 RNA와 복합체를 형성하여 리보핵산 단백질 (RNP) 형태로 작용할 수 있다.
- [0045] Cas9 단백질은 CRISPR/Cas 시스템의 주요 단백질 구성 요소로, 활성화된 엔도뉴클레아제 또는 nickase를 형성할 수 있는 단백질이다.
- [0046] Cas9 단백질 또는 유전자 정보는 NCBI (National Center for Biotechnology Information)의 GenBank와 같은 공지의 데이터 베이스에서 얻을 수 있다. 예컨대, 상기 Cas9 단백질은,

- [0047] 스트렙토코커스 sp. (*Streptococcus* sp.), 예컨대, 스트렙토코커스 피요제네스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질 (예컨대, SwissProt Accession number Q99ZW2(NP_269215.1));
- [0048] 캄필로박터 속, 예컨대, 캄필로박터 제주니 (*Campylobacter jejuni*) 유래의 Cas9 단백질;
- [0049] 스트렙토코커스 속, 예컨대, 스트렙토코커스 써모필러스 (*Streptococcus thermophiles*) 또는 스트렙토코커스 아우레우스 (*Streptococcus aureus*) 유래의 Cas9 단백질;
- [0050] 네이세리아 메닝기티디스 (*Neisseria meningitidis*) 유래의 Cas9 단백질;
- [0051] 파스테우렐라 (*Pasteurella*) 속, 예컨대, 파스테우렐라 물토시다 (*Pasteurella multocida*) 유래의 Cas9 단백질;
- [0052] 프란시셀라 (*Francisella*) 속, 예컨대, 프란시셀라 노비시다 (*Francisella novicida*) 유래의 Cas Cas9 단백질
- [0053] 등으로 이루어진 군에서 선택된 하나 이상일 수 있으나, 이에 제한되는 것은 아니다.
- [0054] 상기 Cas9 단백질은 미생물에서 분리된 것 또는 재조합적 방법 또는 합성적 방법 등과 같이 인위적 또는 비자연적 생산된 것(non-naturally occurring)일 수 있다. 상기 Cas9 단백질은 in vitro에서 미리 전사된 mRNA 또는 미리 생산된 단백질 형태, 또는 표적 세포 또는 생체 내에서 발현하기 위하여 재조합 벡터에 포함된 형태로 사용될 수 있다. 일 예에서, Cas9 단백질은 재조합 DNA(Recombinant DNA; rDNA)에 의하여 만들어진 재조합 단백질일 수 있다. 재조합 DAN는 다양한 유기체로부터 얻어진 이중 또는 동종 유전 물질을 포함하기 위하여 분자 클로닝과 같은 유전자 재조합 방법에 의하여 인공적으로 만들어진 DNA 분자를 의미한다. 예컨대, 재조합 DNA를 적절한 유기체에서 발현시켜 Cas9 단백질을 생산 (*in vivo* 또는 *in vitro*)하는 경우, 재조합 DNA는 제조하고자 하는 단백질을 코딩 하는 코돈들 중에서 상기 유기체에 발현하기에 최적화된 코돈을 선택하여 재구성된 뉴클레오타이드 서열을 갖는 것일 수 있다.
- [0055] 본 명세서에서 사용된 상기 Cas9 단백질은 변이된 형태의 변이 Cas9일 수 있다. 상기 변이 Cas9 단백질은 DNA 이중 가닥을 절단하는 엔도뉴클레아제 활성을 상실하도록 변이된 것을 의미할 수 있으며, 예컨대, 엔도뉴클레아제 활성을 상실하고 니카아제 활성을 갖도록 변이된 변이 표적특이적 뉴클레아제 및 엔도뉴클레아제 활성과 니카아제 활성을 모두 상실하도록 변이된 변이 표적특이적 뉴클레아제 중에서 선택된 1종 이상일 수 있다. 상기 변이 Cas9 단백질이 니카아제 활성을 갖는 것인 경우, 상기 디아미나제에 의한 염기변환(예컨대, 시티딘이 우라딘으로 변환)과 동시 또는 순서와 무관하게 순차적으로, 상기 염기 변환이 일어난 가닥 또는 그 반대 가닥 (예컨대, 염기 변환이 일어난 가닥의 반대 가닥)에서 nick이 도입될 수 있다 (예컨대, PAM 서열의 5' 말단 방향으로 3번째 뉴클레오타이드와 4번째 뉴클레오타이드 사이에 nick이 도입됨). 이와 같은 Cas9 단백질의 변이 (예컨대, 아미노산 치환 등)는 적어도 뉴클레아제의 촉매 활성 도메인 (예컨대, RuvC 촉매 도메인)에서 일어나는 것일 수 있다. 일 예에서, 상기 Cas9 단백질이 스트렙토코커스 피요제네스 유래 Cas9 단백질 (SwissProt Accession number Q99ZW2(NP_269215.1); 서열번호 4)인 경우, 상기 변이는 촉매 활성을 갖는 아스파르트산 잔기 (catalytic aspartate residue; 예컨대, 서열번호 4의 경우 10번째 위치의 아스파르트산 (D10) 등), 서열번호 4의 762번째 위치의 글루탐산 (E762), 840번째 위치의 히스티딘 (H840), 854번째 위치의 아스파라긴 (N854), 863번째 위치의 아스파라긴 (N863), 986번째 위치의 아스파르트산 (D986) 등으로 이루어진 군에서 선택된 하나 이상 임의의 다른 아미노산으로 치환된 돌연변이를 포함할 수 있다. 이 때, 치환되는 임의의 다른 아미노산은 알라닌 (alanine)일 수 있지만, 이에 제한되지 않는다.
- [0056] 다른 예에서, 상기 변이 Cas9 단백질은 야생형 Cas9 단백질과 상이한 PAM 서열을 인식하도록 변이된 것일 수 있다. 예컨대, 상기 Cas9 단백질은 스트렙토코커스 피요제네스 유래 Cas9 단백질의 1135번째 위치의 아스파르트산 (D1135), 1335번째 위치의 아르기닌 (R1335), 및 1337번째 위치의 트레오닌 (T1337) 중 하나 이상, 예컨대 3개 모두가 다른 아미노산으로 치환되어, 야생형 Cas9의 PAM 서열 (NGG)와 상이한 NGA (N은 A, T, G, 및 C 중에서 선택된 임의의 염기)를 인식하도록 변이된 것일 수 있다.
- [0057] 일 예에서, 상기 변이 Cas9 단백질은 스트렙토코커스 피요제네스 유래 Cas9 단백질의 아미노산 서열 (서열번호 4) 중,
- [0058] (1) D10, H840, 또는 D10 + H840;
- [0059] (2) D1135, R1335, T1337, 또는 D1135 + R1335 + T1337; 또는
- [0060] (3) (1)과 (2) 잔기 모두

- [0061] 에서 아미노산 치환이 일어난 것일 수 있다.
- [0062] 본 명세서에 사용된 바로서, 상기 '다른 아미노산'은, 알라닌, 이소류신, 류신, 메티오닌, 페닐알라닌, 프롤린, 트립토판, 발린, 아스파라긴산, 시스테인, 글루타민, 글리신, 세린, 트레오닌, 티로신, 아스파르트산, 글루탐산, 아르기닌, 히스티딘, 라이신, 상기 아미노산들의 공지된 모든 변형체 중에서, 야생형 단백질이 원래 변이 위치에 갖는 아미노산을 제외한 아미노산들 중에서 선택된 아미노산을 의미한다. 일 예에서, 상기 '다른 아미노산'은 알라닌, 발린, 글루타민, 또는 아르기닌일 수 있다.
- [0063] 일 예에서, 상기 변이 Cas9 단백질은 엔도뉴클레아제 활성을 상실(예컨대, 니카아제 활성을 갖거나, 엔도뉴클레아제 활성 및 니카아제 활성을 모두 상실)한 변형 Cas9 단백질, 또는 야생형 Cas9과 상이한 PAM 서열을 인식하는 것일 수 있다. 예컨대, 상기 변이 Cas9 단백질은, 스트렙토코커스 피요제네스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질(서열번호 4)에 있어서,
- [0064] (1) D10 또는 H840 위치에 돌연변이 (예컨대, 다른 아미노산으로의 치환)가 도입되어 엔도뉴클레아제 활성이 상실되고 니카아제 활성을 갖는 변형 Cas9, 또는 스트렙토코커스 피요제네스 (*Streptococcus pyogenes*) 유래의 Cas9 단백질에 D10 및 H840 위치에 모두 돌연변이 (예컨대, 다른 아미노산으로의 치환)가 도입되어 엔도뉴클레아제 활성 및 니카아제 활성을 모두 상실한 변형 Cas9 단백질;
- [0065] (2) D1135, R1335 및 T1337 중에서 하나 이상 또는 이들 모두에 돌연변이(예컨대, 다른 아미노산으로의 치환)가 도입되어 야생형과 상이한 PAM 서열을 인식하는 변형 Cas9 단백질; 또는
- [0066] (3) (1) 및 (2)의 돌연변이가 모두 도입되어 니카아제 활성을 갖고 야생형과 상이한 PAM 서열을 인식하거나, 엔도뉴클레아제 활성 및 니카아제 활성을 모두 상실하고 야생형과 상이한 PAM 서열을 인식하는 변형 Cas9 단백질
- [0067] 일 수 있다.
- [0068] 예컨대, 상기 Cas9 단백질의 D10 위치에서의 돌연변이는 D10A 돌연변이 (Cas9 단백질의 아미노산 중 10번째 아미노산인 D가 A로 치환된 돌연변이를 의미함; 이하, Cas9에 도입된 돌연변이는 동일한 방법으로 표기됨)일 수 있고, 상기 H840 위치에서의 돌연변이는 H840A 돌연변이일 수 있으며, D1135, R1335, 및 T1337 위치에서의 돌연변이는 각각 D1135V, R1335Q, 및 T1337R일 수 있다.
- [0069] 상기 Cas9 단백질은 단백질 형태, 이를 코딩하는 핵산 분자 (예컨대, DNA 또는 mRNA), 가이드 RNA와 결합된 리보핵산 단백질, 상기 리보핵산 단백질을 암호화하는 핵산 분자, 또는 상기 핵산 분자를 포함하는 재조합 벡터의 형태로 사용될 수 있다.
- [0070] 상기 Cas9 단백질을 암호화하는 핵산 분자는 핵 내로 전달, 작용, 및/또는 핵 내에서 발현될 수 있는 형태일 수 있다.
- [0071] 상기 Cas9 단백질은 세포 내로 도입되기에 용이한 형태일 수 있다. 일 예로, 상기 Cas9 단백질 또는 이를 암호화하는 유전자는 세포 침투 펩타이드 및/또는 단백질 전달 도메인 (protein transduction domain) 또는 이를 암호화하는 유전자와 연결될 수 있다. 상기 단백질 전달 도메인은 폴리-아르기닌 또는 HIV 유래의 TAT 단백질일 수 있으나, 이에 제한되지 않는다. 세포 침투 펩타이드 또는 단백질 전달 도메인은 상기 기술된 예 외에도 다양한 종류가 당업계에 공지되어 있으므로, 당업자는 상기 예에 제한되지 않고 다양한 예를 적용할 수 있다.
- [0072] 또한, 상기 Cas9 단백질, 및/또는 이들을 코딩하는 핵산 분자는 핵 위치 신호 (nuclear localization signal, NLS; 예컨대, cccaagaaga agaggaaagtc (서열번호 6))를 추가로 포함할 수 있다. 따라서, 상기 Cas9 단백질 암호화 핵산 분자를 포함하는 발현 카세트는 상기 Cas9 단백질을 발현시키기 위한 프로모터 서열 등의 조절 서열, 및, 임의로, NLS 서열을 추가로 포함할 수 있다. 상기 NLS 서열은 당업계에 잘 알려져 있다.
- [0073] 상기 Cas9 단백질, 및/또는 이를 코딩하는 핵산 분자는 분리 및/또는 정제를 위한 태그 또는 상기 태그를 코딩하는 핵산 서열과 연결될 수 있다. 일 예로, 상기 태그는 His 태그, Flag 태그, S 태그 등과 같은 작은 펩타이드 태그, GST (Glutathione S-transferase) 태그, MBP (Maltose binding protein) 태그 등으로 이루어진 군에서 적절하게 선택될 수 있으나, 이에 제한되지 않는다.
- [0074] 본 명세서에서, 용어 "가이드 RNA (guide RNA)"는 표적 유전자 내의 표적 부위 내의 특이적인 염기 서열 (표적 서열)에 혼성화 가능한 표적화 서열을 포함하는 RNA를 의미하며, 생체 외 (in vitro) 또는 생체 (또는 세포) 내에서 Cas9 단백질과 결합하여 이를 표적 유전자 (또는 표적 부위)로 인도하는 역할을 한다.
- [0075] 상기 가이드 RNA는 복합체를 형성할 Cas9 단백질의 종류 및/또는 그 유래 미생물에 따라서 적절히 선택될 수 있

다.

- [0076] 예컨대, 상기 가이드 RNA는,
- [0077] 표적 서열과 혼성화 가능한 부위 (표적화 서열)을 포함하는 CRISPR RNA (crRNA);
- [0078] Cas9 단백질과 상호작용하는 부위를 포함하는 *trans*-activating crRNA (tracrRNA); 및
- [0079] 상기 crRNA 및 tracrRNA의 주요 부위 (예컨대, 표적화 서열을 포함하는 crRNA 부위 및 뉴클레아제와 상호작용하는 tracrRNA의 부위)가 융합된 형태의 단일 가이드 RNA (single guide RNA; sgRNA)
- [0080] 로 이루어진 군에서 선택된 1종 이상일 수 있으며,
- [0081] 구체적으로 CRISPR RNA (crRNA) 및 *trans*-activating crRNA (tracrRNA)를 포함하는 이중 RNA (dual RNA), 또는 crRNA 및 tracrRNA의 주요 부위를 포함하는 단일 가이드 RNA (sgRNA)일 수 있다.
- [0082] 상기 sgRNA는 표적 유전자 (표적 부위) 내의 표적 서열과 상보적인 서열 (표적화 서열)을 가지는 부분 (이를 Spacer region, Target DNA recognition sequence, base pairing region 등으로도 명명함) 및 Cas 단백질 결합을 위한 hairpin 구조를 포함할 수 있다. 보다 구체적으로, 표적 유전자 내의 표적서열과 상보적인 서열(표적화 서열)을 포함하는 부분, Cas 단백질 결합을 위한 hairpin 구조, 및 Terminator 서열을 포함할 수 있다. 상기 기술된 구조는 5'에서 3' 순으로 순차적으로 존재하는 것일 수 있으나, 이에 제한되는 것은 아니다. 상기 가이드 RNA가 crRNA 및 tracrRNA의 주요 부분 및 표적 DNA의 상보적인 부분을 포함하는 경우라면 어떠한 형태의 가이드 RNA도 본 발명에서 사용될 수 있다.
- [0083] 예컨대, Cas9 단백질은 표적 유전자 교정을 위하여 두 개의 가이드 RNA, 즉, 표적 유전자의 표적 부위와 혼성화 가능한 뉴클레오타이드 서열을 갖는 CRISPR RNA (crRNA)와 Cas9 단백질과 상호작용하는 *trans*-activating crRNA (tracrRNA; Cas9 단백질과 상호작용함)를 필요로 하며, 이들 crRNA와 tracrRNA는 서로 결합된 이중 가닥 crRNA:tracrRNA 복합체 형태, 또는 링커를 통하여 연결되어 단일 가이드 RNA (single guide RNA; sgRNA) 형태로 사용될 수 있다. 일 예에서, *Streptococcus pyogenes* 유래의 Cas9 단백질을 사용하는 경우, sgRNA는 적어도 상기 crRNA의 혼성화 가능한 뉴클레오타이드 서열을 포함하는 crRNA 일부 또는 전부와 상기 Cas9의 tracrRNA의 Cas9 단백질과 상호작용하는 부위를 적어도 포함하는 tracrRNA 일부 또는 전부가 뉴클레오타이드 링커를 통하여 헤어핀 구조 (stem-loop 구조)를 형성하는 것일 수 있다 (이 때 뉴클레오타이드 링커가 루프 구조에 해당할 수 있음).
- [0084] 상기 가이드 RNA, 구체적으로 crRNA 또는 sgRNA는 표적 유전자 내 표적 서열과 상보적인 서열(표적화 서열)을 포함하며, crRNA 또는 sgRNA의 업스트림 부위, 구체적으로 sgRNA 또는 dualRNA의 crRNA의 5' 말단에 하나 이상, 예컨대, 1-10개, 1-5개, 또는 1-3개의 추가의 뉴클레오타이드를 포함할 수 있다. 상기 추가의 뉴클레오타이드는 구아닌 (guanine, G)일 수 있으나, 이에 제한되는 것은 아니다.
- [0085] 상기 가이드 RNA의 구체적 서열은 Cas9의 종류 (즉, 유래 미생물)에 따라서 적절히 선택할 수 있으며, 이는 이 발명이 속하는 기술 분야의 통상의 지식을 가진 자가 용이하게 알 수 있는 사항이다.
- [0086] 일 예에서, 표적특이적 뉴클레아제로서 *Streptococcus pyogenes* 유래의 Cas9 단백질을 사용하는 경우, crRNA는 다음의 일반식 1로 표현될 수 있다:
- [0087] 5'-(N_{cas9})_l-(GUUUUAGAGCUA)-(X_{cas9})_m-3' (일반식 1)
- [0088] 상기 일반식 1에서,
- [0089] N_{cas9}는 표적화 서열, 즉 표적 유전자(target gene)의 표적 부위(target site)의 서열에 따라서 결정되는 부위 (표적 부위의 표적 서열과 혼성화 가능)이며, l은 상기 표적화 서열에 포함된 뉴클레오타이드 수를 나타내는 것으로 15 내지 30, 17 내지 23, 또는 18 내지 22의 정수, 예컨대 20일 수 있고,
- [0090] 상기 표적화 서열의 3' 방향으로 인접하여 위치하는 연속하는 12개의 뉴클레오타이드(GUUUUAGAGCUA) (서열번호 1)를 포함하는 부위는 crRNA의 필수적 부분이고,
- [0091] X_{cas9}는 crRNA의 3' 말단쪽에 위치하는 (즉, 상기 crRNA의 필수적 부분의 3' 방향으로 인접하여 위치하는) m개의 뉴클레오타이드를 포함하는 부위로, m은 8 내지 12의 정수, 예컨대 11일 수 있으며, 상기 m개의 뉴클레오타이드들은 서로 같거나 다를 수 있으며, 각각 독립적으로 A, U, C 및 G로 이루어진 군에서 선택될 수 있다.

- [0092] 일 예에서, 상기 X_{cas9} 는 UGCGUUUUUG (서열번호 2)를 포함할 수 있으나 이에 제한되지 않는다.
- [0093] 또한, 상기 tracrRNA는 다음의 일반식 2로 표현될 수 있다:
- [0094] $5'-(Y_{cas9})_p-(UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC)-3'$ (일반식 2)
- [0095] 상기 일반식 2에서,
- [0096] 60개의 뉴클레오타이드 (UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC) (서열번호 3)로 표시된 부위는 tracrRNA의 필수적 부분이고,
- [0097] Y_{cas9} 는 상기 tracrRNA의 필수적 부분의 5' 말단에 인접하여 위치하는 p개의 뉴클레오타이드를 포함하는 부위로, p는 6 내지 20의 정수, 예컨대 8 내지 19의 정수일 수 있으며, 상기 p개의 뉴클레오타이드들은 서로 같거나 다를 수 있고, A, U, C 및 G로 이루어진 군에서 각각 독립적으로 선택될 수 있다.
- [0098] 또한, sgrRNA는 상기 crRNA의 표적화 서열과 필수적 부위를 포함하는 crRNA 부분과 상기 tracrRNA의 필수적 부분 (60개 뉴클레오타이드)를 포함하는 tracrRNA 부분이 올리고뉴클레오타이드 링커를 통하여 헤어핀 구조 (stem-loop 구조)를 형성하는 것일 수 있다 (이 때, 올리고뉴클레오타이드 링커가 루프 구조에 해당함). 보다 구체적으로, 상기 sgrRNA는 crRNA의 표적화 서열과 필수적 부위를 포함하는 crRNA 부분과 tracrRNA의 필수적 부분을 포함하는 tracrRNA 부분이 서로 결합된 이중 가닥 RNA 분자에서, crRNA 부위의 3' 말단과 tracrRNA 부위의 5' 말단이 올리고뉴클레오타이드 링커를 통하여 연결된 헤어핀 구조를 갖는 것일 수 있다.
- [0099] 일 예에서, sgrRNA는 다음의 일반식 3으로 표현될 수 있다:
- [0100] $5'-(N_{cas9})_1-(GUUUUAGAGCUA)-($ 올리고뉴클레오타이드 링커 $)-(UAGCAAGUUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGC)-3'$ (일반식 3)
- [0101] 상기 일반식 3에서, $(N_{cas9})_1$ 는 표적화 서열로서 앞서 일반식 1에서 설명한 바와 같다.
- [0102] 상기 sgrRNA에 포함되는 올리고뉴클레오타이드 링커는 3 내지 5개, 예컨대 4개의 뉴클레오타이드를 포함하는 것일 수 있으며, 상기 뉴클레오타이드들은 서로 같거나 다를 수 있고, A, U, C 및 G로 이루어진 군에서 각각 독립적으로 선택될 수 있다. 일 구체예에서, 상기 올리고뉴클레오타이드 링커는 GAAA의 핵산 서열을 포함하는 것일 수 있으나, 이에 제한되는 것은 아니다.
- [0103] 상기 crRNA 또는 sgrRNA는 5' 말단 (즉, crRNA의 타겟팅 서열 부위의 5' 말단)에 1 내지 3개의 구아닌(G)을 추가로 포함할 수 있다.
- [0104] 상기 tracrRNA 또는 sgrRNA는 tracrRNA의 필수적 부분(60nt)의 3' 말단에 3개 내지 7개, 3개 내지 5개, 또는 5개 내지 7개의 우라실 (U)을 포함하는 종결부위를 추가로 포함할 수 있다.
- [0105] 상기 가이드 RNA의 표적 서열은 표적 DNA 상의 PAM (Protospacer Adjacent Motif 서열(*S. pyogenes* Cas9의 경우, 5'-NGG-3' (N은 A, T, G, 또는 C임))의 5'에 인접하여 위치하는 약 17개 내지 약 23개 또는 약 18개 내지 약 22개, 예컨대 20개의 연속하는 핵산 서열일 수 있다.
- [0106] 상기 가이드 RNA의 표적 서열과 혼성화 가능한 가이드 RNA의 표적화 서열은 상기 표적 서열이 위치하는 DNA 가닥 (즉, PAM 서열(5'-NGG-3' (N은 A, T, G, 또는 C임))이 위치하는 DNA 가닥) 또는 이의 상보적인 가닥의 뉴클레오타이드 서열과 50% 이상, 60% 이상, 70% 이상, 80% 이상, 90% 이상, 95% 이상, 99% 이상, 또는 100%의 서열 상보성을 갖는 뉴클레오타이드 서열을 의미하는 것으로, 상기 상보적 가닥의 뉴클레오타이드 서열과 상보적 결합이 가능하다.
- [0107] 본 명세서에서, 표적 부위의 핵산 서열(표적 서열)은 표적 유전자의 해당 유전자 부위의 두 개의 DNA 가닥 중 PAM 서열이 위치하는 가닥의 핵산 서열로 표시된다. 이 때, 실제로 가이드 RNA가 결합하는 DNA 가닥은 PAM 서열이 위치하는 가닥의 상보적 가닥일 수 있으므로, 상기 가이드 RNA에 포함된 표적화 서열은, RNA 특성상 T를 U로 변경하는 것을 제외하고, 표적 서열과 동일한 핵산 서열을 가질 수 있다. 따라서, 본 명세서에서, 가이드 RNA의 표적화 서열과 표적 서열은 T와 U가 상호 변경되는 것을 제외하고 동일한 핵산 서열로 표시된다.
- [0108] 상기 가이드 RNA는 RNA 형태로 사용 (또는 상기 조성물에 포함)되거나, 이를 코딩하는 DNA를 포함하는 플라스미드 형태로 사용 (또는 상기 조성물에 포함)될 수 있다.

발명의 효과

[0110] 본 명세서에서 제공되는 고티성 Cas9 변이체들의 오프-타겟 부위에 대한 낮은 유전자 교정 활성을 유지하면서 온-타겟 특이적 유전자 교정 활성을 보다 증진시킬 수 있는 기술은 유전공학 분야 및 의학 분야에 광범위하게 적용될 수 있다.

도면의 간단한 설명

[0112] 도 1은 고티성도 (high-fidelity) Cas9 변이체들이 non-G 5' nucleotide를 갖는 표적 부위에서의 유전자 교정 활성을 보여주는 것으로,

1a는 야생형 Cas9(Cas9-WT), 조작된 Cas9 변이체들, 및 sgRNA를 모식적으로 보여주며, Cas9-WT와 비교하여, eCas9-1.1 또는 Cas9-HF1에 도입된 알라닌 치환을 각각 청색 또는 적색 별표로 표시하였으며, 적색 삼각형은 절단 위치를 나타내고, GX₁₉ sgRNA는 protospacer (청색줄)와 매칭된 G로 시작하는 것을 나타내고, gX₁₉ sgRNA는 5'-말단에 미스매칭된 G를 갖는 것이고, gX₂₀ sgRNA는 5'-말단에 추가의 구아닌을 포함하는 것이며, PAM (protospacer-adjacent motif)은 적색 줄로 표시된 NGG이며 (H, G이 아님 (A 또는 C 또는 T); D, C가 이님 (A 또는 G 또는 T)),

1b는 HeLa 세포에서 gX₁₉ sgRNA 또는 gX₂₀ sgRNA를 사용하여 얻어진 5' 뉴클레오타이드가 G가 아닌 온-타겟 부위에서의 Indel 빈도를 보여주는 그래프이다 (targeted deep sequencing에 의하여 측정됨, PAM 서열은 파란색으로 표시됨, Error bars, s.e.m.; n = 3).

도 2는 Hammerhead 리보자임-결합 sgRNA를 보여주는 것으로,

2a는 자가-프로세싱 리보자임(self-processing ribozyme)이 융합된 sgRNA를 모식적으로 보여주며, pre-sgRNA는 5' 말단에 Hammerhead (HH) 리보자임을 포함하고, pre-sgRNA는 자가 절단 (self-cleavage)을 거쳐 성숙 sgRNA로 방출되며, 적색 화살표는 자가 절단 위치를 나타내며,

2b는 매칭된 5' 뉴클레오타이드를 갖는 HH 리보자임-융합된 sgRNA (HH-X₂₀) 또는 미스매치 구아노신을 갖는 HH 리보자임-융합된 sgRNA (HH-gX₁₉)을 Cas9-WT 또는 high-fidelity Cas9 변이체와 각각 조합하여 HeLa 에서의 5개의 표적 부위에 대한 유전자 교정 효율을 시험하고, targeted deep sequencing를 사용하여 Indel 빈도를 측정하였으며 (PAM은 청색으로 표시되어 있고, Error bars, s.e.m.; n = 3, Statistical significances were calculated by t-test. * P < 0.05, ** P < 0.01),

2c는 5개의 표적 부위에서의 평균 indel 빈도 ± s.e.m.를 보여주는 그래프이며 (* P < 0.05, ** P < 0.01),

2d는 HH-X₂₀ sgRNA를 Cas9-WT 또는 Cas9 변이체와 함께 HeLa 세포 내로 공동 형질감염시키고, targeted deep sequencing으로 측정한 온-타겟 및 오프-타겟 부위에서의 Indel 빈도를 나타낸 그래프이다 (PAM 서열은 청색으로 나타내고, 미스매치 염기는 적색으로 표시함; specificity ratio는 Cas9 변이체 및 Cas9-WT를 사용하여 얻어진 오프-타겟 부위에서의 indel 빈도에 대한 온-타겟 부위에서의 indel 빈도의 비율 차이를 배수로 측정한 값 ([Cas9 변이체를 사용하여 얻어진 오프-타겟 부위에서의 indel 빈도에 대한 온-타겟 부위에서의 indel 빈도의 비율]/[Cas9-WT를 사용하여 얻어진 오프-타겟 부위에서의 indel 빈도에 대한 온-타겟 부위에서의 indel 빈도의 비율]); Error bars, s.e.m.; n = 3; mock transfected sample과 비교하여 유의미한 수준의 Indel 빈도는 별표로 표시함 (* P < 0.05, ** P < 0.01)).

발명을 실시하기 위한 구체적인 내용

[0113] 이하에서는 실시예를 들어 본 발명을 더욱 구체적으로 설명하고자 하나, 이는 예시적인 것에 불과할 뿐 본 발명의 범위를 제한하고자 함이 아니다. 아래 기재된 실시예들은 발명의 본질적인 요지를 벗어나지 않는 범위에서 변형될 수 있음은 당 업자들에게 있어 자명하다.

- [0115] **실시예 1: high-fidelity Cas9 변이체 암호화 플라스미드 및 HH-리보자임-융합된 sgRNA 암호화 플라스미드의 구축**
- [0116] Cas9 변이체를 암호화하는 플라스미드로서, *Streptococcus pyogenes* Cas9 단백질(서열번호 4)에 K848A, K1003A, 및 R1060A 변이가 도입된 eSpCas9-1.1를 암호화하는 플라스미드 (p3seCas9-1.1; Addgene #104172) 및 N497A, R661A, Q695A, 및 Q926A 변이가 도입된 p3s-Cas9-HF1를 암호화하는 플라스미드 (p3s-Cas9-HF1, Addgene #104173)을 각각 사용하였다.
- [0117] HH(Hammerhead)-리보자임 sgRNA 구조체 (도 2a 참조)는 HH-리보자임 핵산 서열 및 protospacer 서열을 포함하는 어닐링된 올리고뉴클레오타이드를 플라스미드 (pRG2, Addgene #104174; sgRNA가 U6 프로모터 조절 하에서 발현) 내로 ligation함으로써 클로닝하였다.
- [0118] 상기 sgRNA는 다음의 서열을 갖는다:
- [0119] 5'-(표적 서열)-(GUUUUAGAGCUA; 서열번호 1)-(뉴클레오타이드 링커)-(UAGCAAGUUAAAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGUGGC; 서열번호 3)-UUUU-3'
- [0120] (상기 표적 서열은 하기의 표 1에 표시된 표적 부위 서열(20nt)에서 "T"를 "U"로 변환한 서열이며,
- [0121] gX₁₉ sgRNA는 표 1의 모든 표적 부위 서열에 있어서 5' 말단 염기 (밑줄로 표시) 위치에 'G'를 포함하도록 제작된 sgRNA이고,
- [0122] X₂₀ sgRNA는 표 1의 각각의 표적 부위 서열의 5' 말단 염기(밑줄로 표시)에 매칭되는 염기를 포함하도록 제작된 sgRNA이며,
- [0123] 상기 뉴클레오타이드 링커는 GAAA의 뉴클레오타이드 서열을 가짐).

표 1

표적 부위 서열

Locus	Target site + PAM 서열 (굵은체 및 밑줄로 표시; 5'→3')
AAVS1	CTCCCTCCCAGGATCCTCTCT TGG (서열번호 7)
CCR5	TCATCCTGATAAACTGCAAA AGG (서열번호 8)
HBB-02	CTTGCCCCACAGGGCAGTA ACGG (서열번호 9)
HBB-03	CACGTTACCTTGCCCCAC AGG (서열번호 10)
HBB-04	CCACGTTACCTTGCCCCAC AGG (서열번호 11)
EMX1-05	TGTACTTTGTCCTCCGGTT TGG (서열번호 12)

- [0125] **실시예 2: 세포 배양 및 형질감염 (transfection)**
- [0126] HeLa cells (ATCC, CCL-2)를 100 units/mL 페니실린, 100 ug(microgram)/mL 스트렙토마이신, 0.1 mM 비필수 아미노산, 및 10%(w/v) 우태아혈청 (fetal bovine serum; FBS)이 보충된 Dulbecco's modified Eagle's medium (DMEM)에 유지시켰다. 제조자 사용 설명에 따라서 Lipofectamine 2000 (Invitrogen)를 사용하여 0.8 x 10⁵개의 HeLa 세포를 Cas9-암호화 플라스미드 (0.5ug) 및 sgRNA 발현 플라스미드 (0.5ug)로 형질감염시켰다.
- [0128] **실시예 3: Targeted deep sequencing**
- [0129] NGS 라이브러리 구축을 위하여, Phusion polymerase (Thermo Fisher Scientific)를 사용하여 온-타겟 및 오프-타겟 영역 (표 1, 도 1b, 2b, 및 2d 참조)을 PCR 증폭시켰다. MiniSeq with TruSeq HT Dual Index system (Illumina)을 사용하여 제조자 사용 설명에 따라서 Pooled PCR amplicons을 시퀀싱하였다. Indel 빈도는 Cas-Analyzer(Park, J., Lim, K., Kim, J.S. & Bae, S. Cas-analyzer: an online tool for assessing genome editing results using NGS data. *Bioinformatics* **33**, 286-288 (2017))를 사용하여 측정하였다.
- [0131] **실시예 4: 미스매칭 5' 말단 뉴클레오타이드를 갖는 가이드 RNA의 유전자 교정 효율 시험**

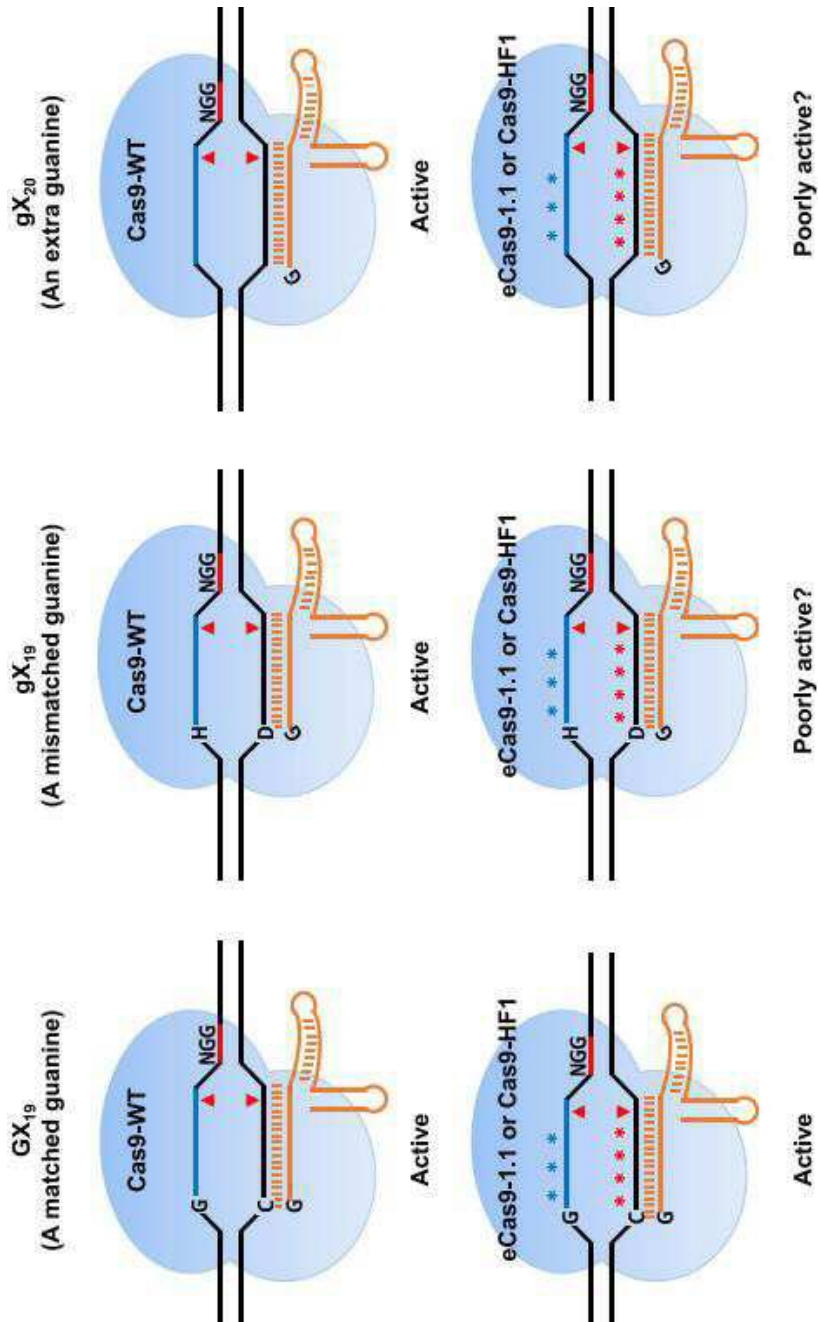
- [0132] 진핵 세포에서 sgRNA를 발현하는데 일반적으로 사용되는 U6 프로모터는 전사를 개시하기 위하여 구아노신(G) 뉴클레오타이드를 필요로 하기 때문에 sgRNA는 전형적으로 5' 말단에 G 뉴클레오타이드를 포함한다. 대부분의 (평균 75%) DNA 표적 부위는 이 위치에서 미스매치를 포함하므로, 이들 부위에서의 gX₁₉ sgRNA (도 1a; 도 1a에서 "g"는 미스매칭된 구아노신을, "G"는 매칭된 구아노신을 각각 나타내고, 각각의 N 또는 X는 A, T, G, 및 C를 포함하는 핵산 염기들 중에서 독립적으로 선택됨)와의 복합체화된 감쇠(attenuated) Cas9 변이체에 의한 세포 내 교정 수준은 낮아질 수 있다.
- [0133] 이와 같은 가설을 입증하기 위하여, gX₁₉ sgRNA를 사용하여, HeLa 세포에서 5' 말단 뉴클레오타이드가 구아노신이 아닌 5개 유전자 부위에서의 eCas9-1.1과 Cas9-HF1의 유전자 교정 활성을 야생형 Cas9 단백질 (Cas9-WT)과 비교하여 그 결과를 도 1b에 나타내었다. 유전자 교정 활성은 실시예 3에 기재된 방법을 참조하여 indel 빈도를 측정하여 확인하였다.
- [0134] 도 1b에 나타난 바와 같이, Cas9-WT는 5' 말단의 뉴클레오타이드 미스매치에 민감하지 않아서 48% 내지 78% (평균 70±5%)의 높은 빈도로 indels를 유도했다. eCas9-1.1은 5개의 사이트 중 4개의 사이트 (33±14%)에서 1.4~45% (22±10%)의 매우 낮은 indel 빈도를 보였다. Cas9-HF1은 3개의 Cas9 뉴클레아제 중에서 가장 활성이 낮았으며, 빈도는 0.2 내지 20% (8.3±4.1%)의 indel 빈도를 보였다. CCR5 표적 부위에서, 2개의 감쇠된 Cas9 단백질은 모두 비활성이었다 (1% 미만의 indel 빈도를 보임). 또한 5' 말단에 추가의 구아노신을 갖는 gX₂₀ sgRNA를 사용하여 유전자 교정 활성 (indel 빈도)을 시험하였다. 이들 sgRNA는 5' 말단에 매칭된 뉴클레오타이드를 갖는다. gX₂₀ sgRNA는 AAVS1 및 HBB-02 부위에서 Cas9 변이체의 유전자 교정 활성을 증가시켰지만, 다른 세 부위에서는 gX₁₉ sgRNA와 비교하여 Cas9 변이체의 유전자 교정 활성을 감소시켰다.
- [0136] **실시예 5: 매칭된 5' 뉴클레오타이드를 포함하는 sgRNA의 제작 및 이를 이용한 Cas9 변이체의 유전자 교정 효율 시험**
- [0137] 높은 충실도의 Cas9 변이체로 타겟팅할 수 있는 유전자 부위의 수를 확대하기 위하여, 자가 절단 리보자임 (self-cleaving ribozyme)을 이용하여 5' 뉴클레오타이드가 표적 DNA 서열과 매칭되는 sgRNA를 생산하였다. 각각의 sgRNA는 5' 말단에서 Hammerhead (HH) 리보자임에 융합되어 (Gao, Y. & Zhao, Y. Self-processing of ribozyme-flanked RNAs into guide RNAs in vitro and in vivo for CRISPR-mediated genome editing. *Journal of integrative plant biology* **56**, 343-349 (2014)), 자가 절단 후 성숙한 20-뉴클레오타이드 (X₂₀) sgRNAs를 생성하였다. 상기 과정을 도 2a에 모식적으로 나타내었다.
- [0138] 매칭된 5' 뉴클레오타이드를 갖는 HH 리보자임-융합 sgRNA (HH-X₂₀으로 명명) 또는 미스매치 5' 구아노신 뉴클레오타이드를 갖는 HH 리보자임-융합 sgRNA (HH-gX₁₉로 명명)를 Cas9-WT 또는 높은 충실도의 Cas9 변이체와 조합하여 HeLa 세포에서의 유전자 교정 활성을 시험하여, 그 결과를 도 2b 및 도 2c에 나타내었다. 유전자 교정 활성은 실시예 3에 기재된 방법을 참조하여 indel 빈도를 측정하여 확인하였다.
- [0139] 도 2b에 나타난 바와 같이, HH-X₂₀ sgRNA를 사용함으로써 시험된 2개의 Cas9 변이체 (eCas9-1.1 및 Cas9-HF1) 모두 5개의 모든 표적 부위에서 교정 활성을 나타낼 수 있게 되었다. HH-X₂₀ sgRNA와 함께 eCas9-1.1 (69±5%) 또는 Cas9-HF1 (59±7%)를 사용하여 얻은 indel 빈도는 HH-X₂₀ 또는 HH-gX₁₉ sgRNA와 함께 Cas9-WT (각각 71±6% 또는 70±3%)를 사용하여 얻은 결과와 거의 유사하게 나타났다 (도 2c). Cas9 변이체 eCas9-1.1 및 Cas9-HF1는 HH-gX₁₉ sgRNA와 함께 사용되는 경우 낮은 교정 활성을 보였으며, 이러한 결과는 높은 충실도의 변이체는 리보자임과의 융합 그 자체 보다는 5' 말단에 매칭된 뉴클레오타이드의 존재 때문에 높은 유전자 교정 활성을 유지할 수 있음을 보여준다.
- [0140] 다음으로, HeLa 세포에서 알려진 오프-타겟 위치에서의 돌연변이 빈도를 측정하여, HH-X₂₀ sgRNAs와 함께 사용되는 경우의 2가지 Cas9 변이체의 표적 특이도를 비교하였다 (CCR5 특이적 sgRNA에 대해서는 오프-타겟 위치가 알려지지 않아서 분석에서 제외함). 상기 얻어진 결과를 도 2d에 나타내었다. 도 2d에 나타난 바와 같이, 각각의 on-target site와 1 내지 3개의 뉴클레오타이드가 상이한 off-target site의 대부분에서, 2종의 Cas9 변이체 모두 Cas9-WT보다 낮은 indel 빈도를 나타냈다. Cas9-HF1은 하나의 뉴클레오타이드 미스매치를 갖는 3개의 오프-타겟 사이트 (하나의 HBB-03 오프-타겟 사이트 및 2개의 HBB-04 오프-타겟 사이트)를 구별할 수 있

었다.

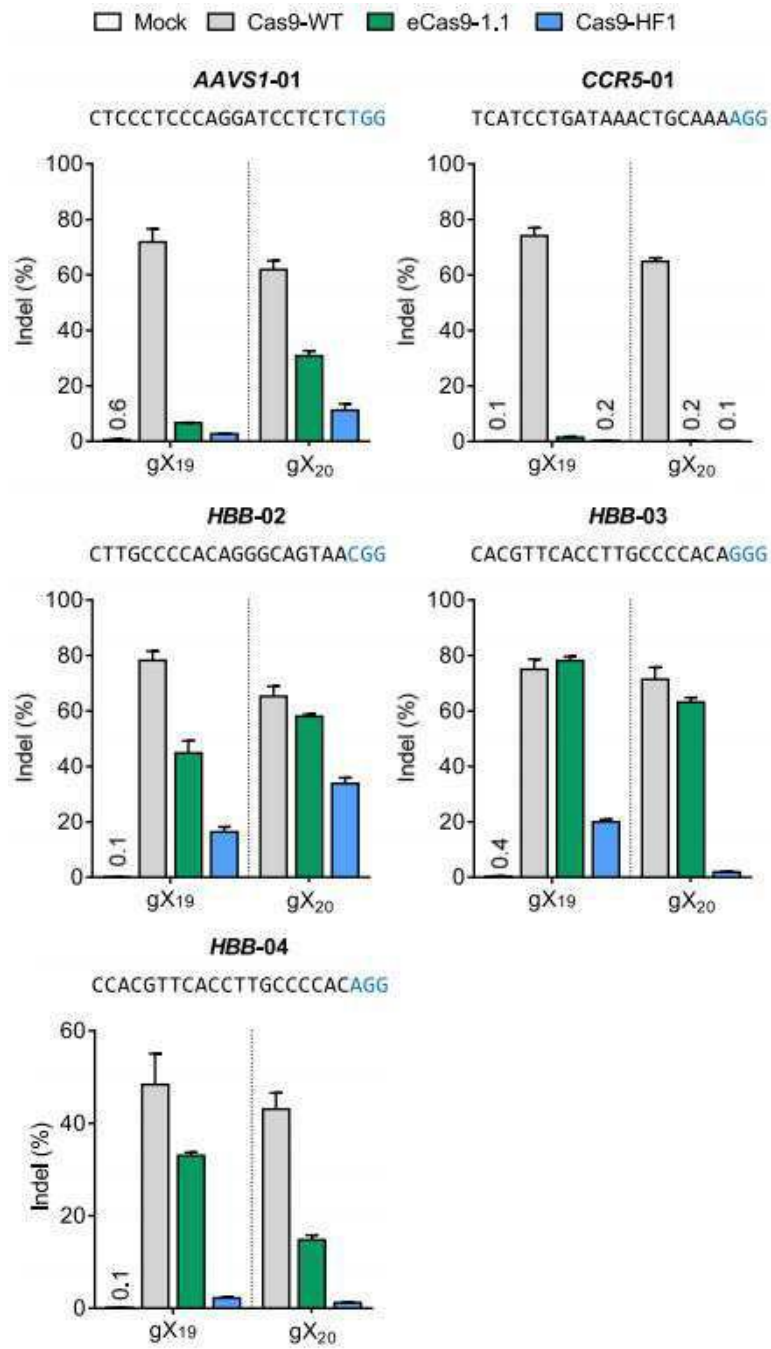
- [0142] 상기 결과를 요약하면, 야생형 단백질과 달리, 새롭게 개발된 고티성 Cas9 변이체가 5' 말단에서 미스매칭을 갖는 표적 부위에서 교정 효율이 낮은 경우가 많음을 보여주며, 이러한 결과는 5' 뉴클레오타이드가 원핵 세포 또는 인간 세포와 같은 진핵 세포에서 CRISPR-Cas9의 높은 특이성에 기여함을 최초로 보여주는 것이다. HHH-리보자임 융합의 자기 분해 활성화에 의하여 sgRNA의 첫 번째 뉴클레오타이드를 표적 DNA 서열과 매칭시킴으로써, 온-타겟 교정 효율에 불리한 영향 없이 유전체 교정의 높은 특이성을 달성할 수 있다. HHH-리보자임 융합을 사용하는 것에 대한 대안으로서, tRNA 융합(Port, F. & Bullock, S.L. Augmenting CRISPR applications in *Drosophila* with tRNA-flanked sgRNAs. *Nature methods* **13**, 852-854 (2016)) 또는 화학적 합성 (Hendel, A. et al. Chemically modified guide RNAs enhance CRISPR-Cas genome editing in human primary cells. *Nature biotechnology* **33**, 985-989 (2015))을 통해 5' 말단에 G 이외의 매칭된 뉴클레오타이드 (5' non-G nucleotide)를 갖는 sgRNA를 제작하고 상기한 2 종의 high-fidelity Cas9 변이체와 조합하여 사용될 수 있다. Cas9- 및 sgRNA-암호화 플라스미드를 사용하는 경우와 비교하여, 사전 조립된 (생체 외 또는 세포 외에서 조립(제작)된) Cas9 변이체 리보핵산단백질(Kim, S., Kim, D., Cho, S.W., Kim, J. & Kim, J.S. Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome research* (2014))의 전달에 의하여 genome-wide target specificity를 보다 향상시킬 수 있다.

도면

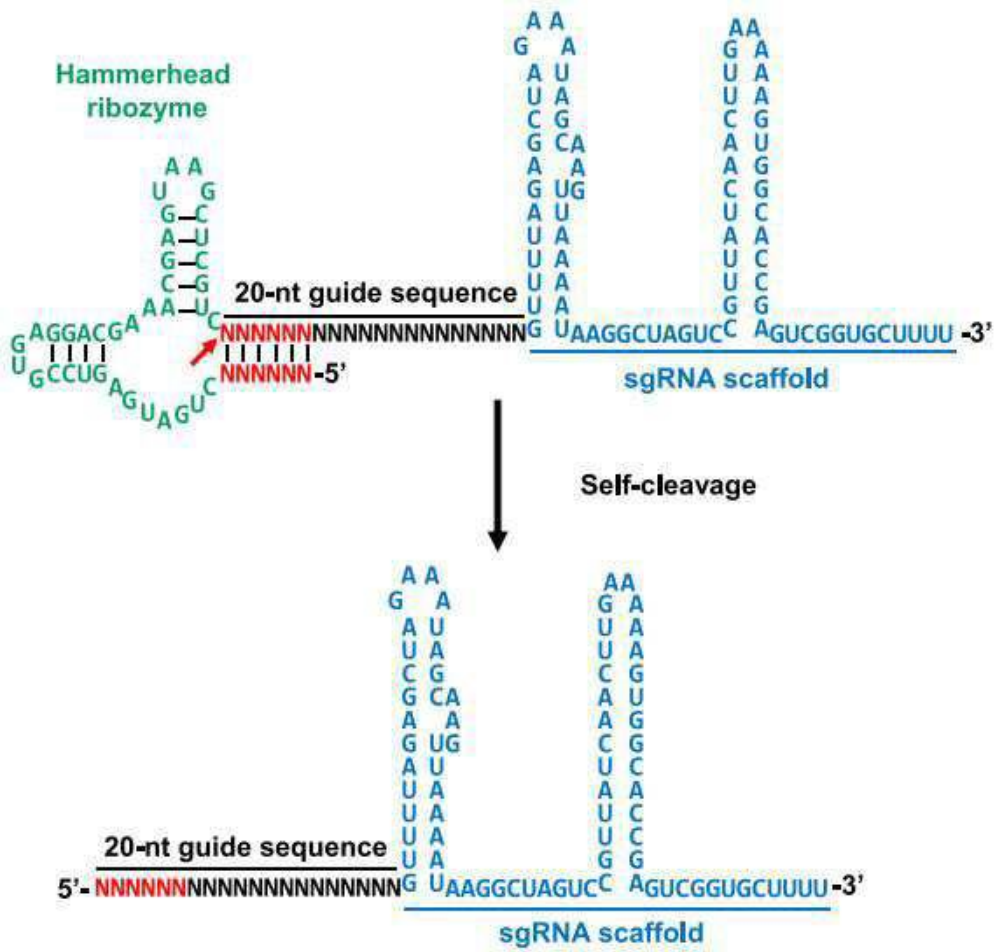
도면1a



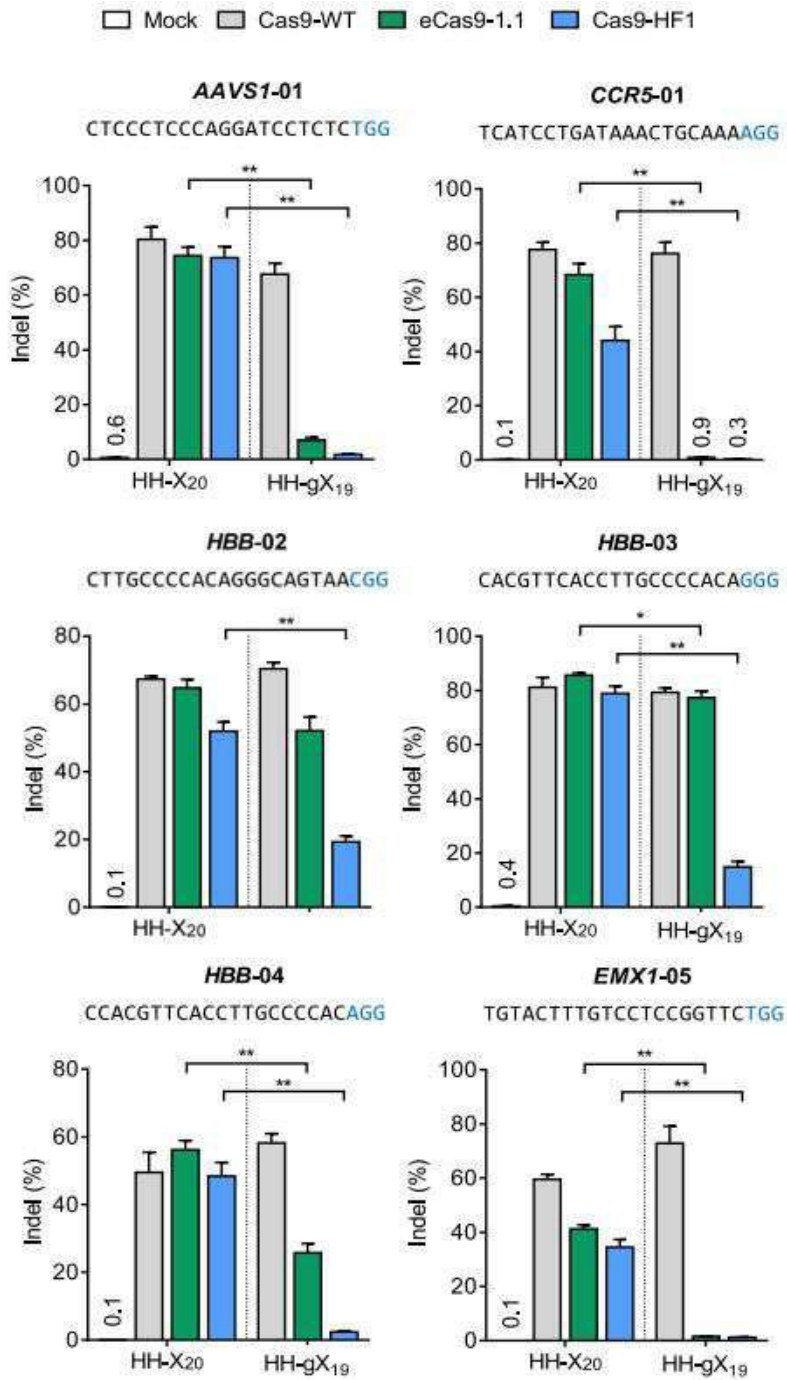
도면1b



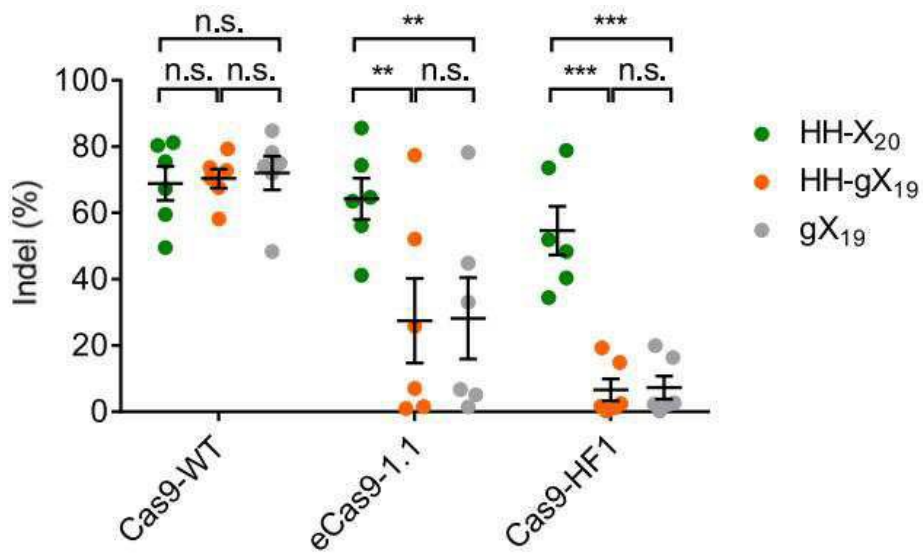
도면2a



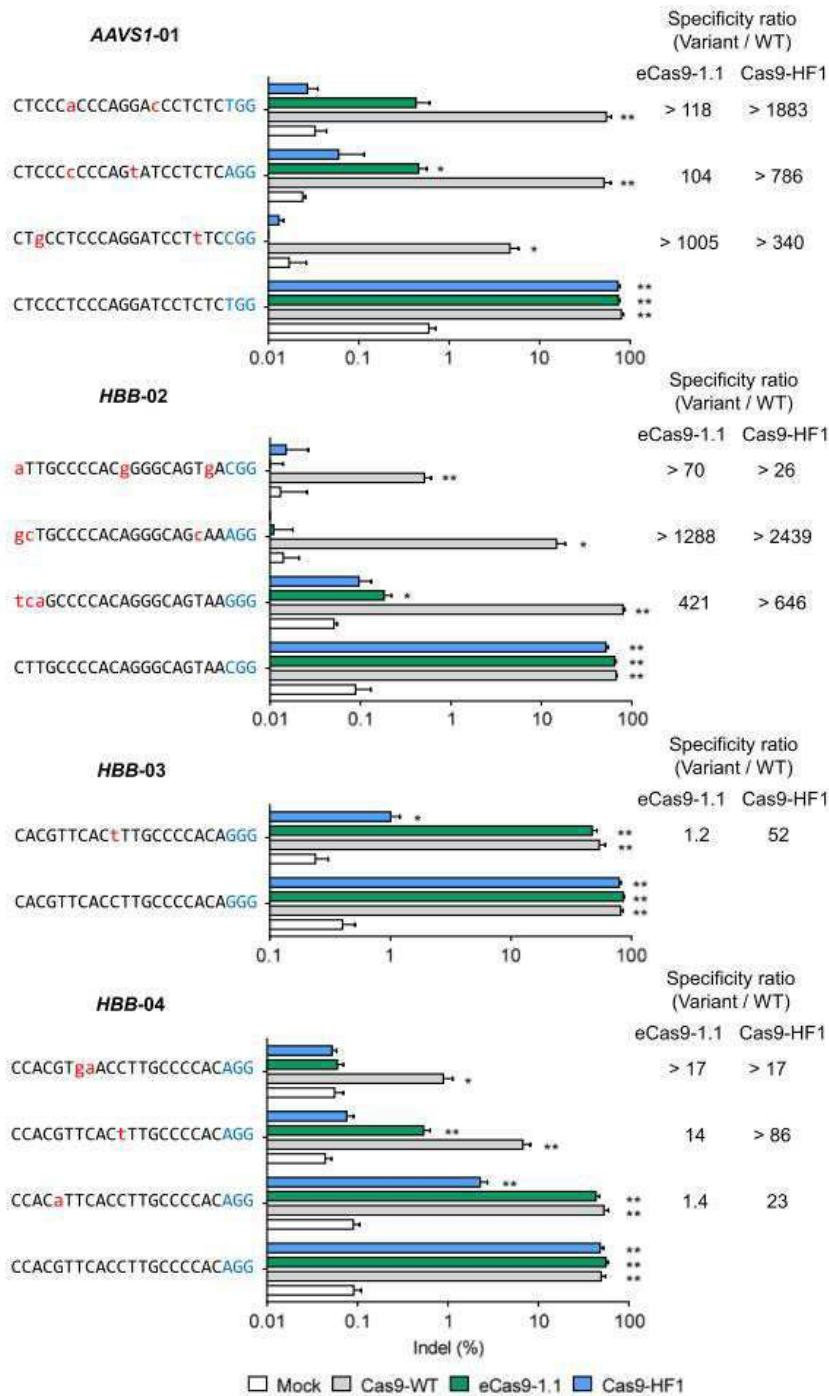
도면2b



도면2c



도면2d



서열목록

- <110> INSTITUTE FOR BASIC SCIENCE
- <120> Gene editing composition comprising sgRNAs with matched 5' nucleotide and gene editing method using the same
- <130> DPP20181455KR
- <150> KR 10-2017-0064332
- <151> 2017-05-24

<160> 12
 <170> KopatentIn 3.0
 <210> 1
 <211> 12
 <212> RNA
 <213> Artificial Sequence
 <220><223> Essential part of crRNA
 <400> 1
 guuuuagagc ua 12
 <210> 2
 <211> 10

 <212> RNA
 <213> Artificial Sequence
 <220><223> 3'-terminal part of crRNA
 <400> 2
 ugcuguuuug 10
 <210> 3
 <211> 60
 <212> RNA
 <213> Artificial Sequence
 <220><223> Essential part of tracrRNA
 <400> 3
 uagcaaguua aaauaaggcu aguccguuau caacuugaaa aaguggcacc gagucggugc 60
 60

 <210> 4
 <211> 1368
 <212> PRT
 <213>
 > Artificial Sequence
 <220><223> Cas9 from Streptococcus pyogenes
 <400> 4
 Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val
 1 5 10 15
 Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
 275 280 285
 Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
 290 295 300
 Ile Leu Arg Val Asn Thr Glu Ile Thr Lys Ala Pro Leu Ser Ala Ser
 305 310 315 320
 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335
 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380
 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400
 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560

 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575
 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590
 Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605
 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala

 625 630 635 640
 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 645 650 655
 Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670
 Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 675 680 685
 Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 690 695 700

 Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 705 710 715 720
 His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735
 Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
 740 745 750
 Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
 755 760 765
 Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile

1025 1030 1035 1040
 Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn Gly Glu
 1045 1050 1055
 Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr Gly Glu Ile

 1060 1065 1070
 Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser
 1075 1080 1085
 Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
 1090 1095 1100
 Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
 1105 1110 1115 1120
 Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
 1125 1130 1135

 Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
 1140 1145 1150
 Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
 1155 1160 1165
 Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
 1170 1175 1180
 Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
 1185 1190 1195 1200
 Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser

 1205 1210 1215
 Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr
 1220 1225 1230
 Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
 1235 1240 1245
 Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
 1250 1255 1260
 Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
 1265 1270 1275 1280

Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
 1285 1290 1295
 His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
 1300 1305 1310
 Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
 1315 1320 1325
 Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
 1330 1335 1340
 Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile

1345 1350 1355 1360
 Asp Leu Ser Gln Leu Gly Gly Asp
 1365

<210> 5
 <211> 4107
 <212> DNA
 <213> Artificial Sequence
 <220><223> Cas9-coding sequence
 <400> 5

atggacaaga agtacagcat cggcctggac atcggtagca acagcgtggg ctgggccgtg 60
 atcaccgacg agtacaaggt gccagcaag aagttcaagg tgctgggcaa caccgaccgc 120
 cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag 180
 gccaccgcc tgaagcgac cgcccggcg cgctacacc gccgcaagaa ccgcatctgc 240

 tacctgcagg agatcttcag caacgagatg gccaaagtgg acgacagctt cttccaccgc 300
 ctggaggaga gttcctggt ggaggaggac aagaagcag agcgccacc catcttcggc 360
 aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgcgcaag 420
 aagctggtgg acagcaccga caaggccgac ctgcgcctga tctacctggc cctggcccac 480
 atgatcaagt tccgcgcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540
 gtggacaagc tgitcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc 600
 atcaacgcca gcggcgtgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc 660

 cgctggaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggcaac 720
 ctgatcgcgc tgagcctggg cctgaccccc aacttcaaga gcaacttca cctggccgag 780
 gacgccaagc tgcagctgag caaggacacc tacgacgac acctggacaa cctgctggcc 840

cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc 900
 ctgctgagcg acatcctgcg cgtgaacacc gagatcacca agggccccct gagcgccagc 960
 atgatcaagc gctacgacga gcaccaccag gacctgacct tgctgaaggc cctggtgcgc 1020
 cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggttacgcc 1080

 ggctacatcg acggcggcgc cagccaggag gagttctaca agttcatcaa gcccatcctg 1140
 gagaagatgg acggcaccca ggagctgctg gtgaagctga accgcgagga cctgctgcgc 1200
 aagcagcgca ccttcgacaa cggcagcatc ccccaccaga tccacctggg cgagctgcac 1260
 gccatcctgc gccgccagga ggacttctac cccttcctga aggacaaccg cgagaagatc 1320
 gagaagatcc tgaccttccg catcccctac tacgtgggcc ccctggcccg cggcaacagc 1380
 cgcttcgctt ggatgacctg caagagcgag gagaccatca cccctggaa cttcgaggag 1440
 gtggtggaca agggcgccag cgcccagagc ttcatcgagc gcatgaccaa cttcgacaag 1500

 aacctgccc acgagaaggt gctgccaag cacagcctgc tgtacgagta cttcacctg 1560
 tacaacgagc tgaccaaggt gaagtacgtg accgagggca tgcgcaagcc cgccttctg 1620
 agcggcgagc agaagaagc 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
 ctgacctga ccctgttcga ggaccgagc atgatcgagg agcgctgaa gacctacgcc 1920

 cacctgttcg acgacaaggt gatgaagcag ctgaagcgcc gccgctacac cggctggggc 1980
 cgctgagcc gcaagcttat caacggcatc cgcgacaagc agagcggcaa gacctcctg 2040
 gacttctga agagcgacgc cttcgcacac cgcaacttca tgcagctgat ccacgacgac 2100
 agcctgacct tcaaggagga catccagaag gccaggtga gcggccaggg cgacagcctg 2160
 cacgagcaca tcgccaacct ggccggcagc cccgccatca agaaggcat cctgcagacc 2220
 gtgaaggtgg tggacgagct ggtgaaggtg atgggccgcc acaagcccga gaacatcgtg 2280
 atcgagatgg cccgagaaa ccagaccacc cagaagggcc agaagaacag ccgagagcgc 2340

 atgaagcgca tcgaggagg catcaaggag ctgggcagcc agatcctgaa ggagcacc 2400
 gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccgc 2460
 gacatgtacg tggaccagga gctggacatc aaccgctga gcgactacga cgtggaccac 2520
 atcgtgcccc agagcttctt gaaggacgac agcatcgaca acaaggtgct gacccgcagc 2580
 gacaagaacc gcggcaagag cgacaacctg cccagcgagg aggtggtgaa gaagatgaag 2640
 aactactggc gccagctgct gaacccaag ctgatcacc agcgcaagtt cgacaacctg 2700

accaaggccg agcgcggcgg cctgagcgag ctggacaagg ccggcttcat caagcgccag 2760

 ctggtggaga cccgccagat caccaagcac gtggcccaga tcctggacag ccgcatgaac 2820
 accaagtacg acgagaacga caagetgata cgcgaggtga aggtgatcac cctgaagagc 2880
 aagctggtga gcgacttccg caaggacttc cagttctaca aggtgcgcga gatcaacaac 2940
 taccaccacg cccacgacgc ctacctgaac gccgtggtgg gcaccgcctt gatcaagaag 3000
 taccccaage 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 gagcatgcc caggtgaaca tcgtgaagaa gaccgaggtg 3300
 cagaccggcg gcttcagcaa ggagagcatc ctcccaagc gcaacagcga caagctgatc 3360
 gcccgaaga aggactggga cccaagaag tacggcggct tcgacagccc caccgtggcc 3420
 tacagcgtgc tgggtggggc caaggtggag aagggaaga gcaagaagct gaagagcgtg 3480
 aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac 3540
 ttctggagg ccaagggcta caaggaggtg aagaaggacc tgatcatcaa gctgccaag 3600

 tacagcctgt tcgagctgga gaacggccgc aagcgcacgc tggccagcgc cggcgagctg 3660
 cagaagggca acgagctggc cctgcccagc aagtacgtga acttctgta cctggccagc 3720
 cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag 3780
 cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg 3840
 atcctggccg acgccaacct ggacaagggt ctgagcgcct acaacaagca ccgcgacaag 3900
 cccatccgag agcaggccga gaacatcatc cacctgttca ccctgacaa cctgggcgcc 3960
 cccgcccct tcaagtactt cgacaccacc atcgaccgca agcgtacac cagcaccaag 4020

 gaggtgctgg acgccacct gatccaccag agcataccg gtctgtacga gaccgcacac 4080
 gacctgagcc agctgggagg cgactaa 4107

 <210> 6
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> NLS
 <400> 6
 cccaagaaga agaggaaagt c 21

<210> 7
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Target sequence on AAVS1 and PAM
 <400> 7
 ctccctccca ggatcctctc tgg 23

<210> 8
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Target sequence on CCR5 and PAM
 <400> 8
 tcctcctgat aaactgcaaa agg 23

<210> 9
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Target sequence on HBB-02 and PAM
 <400> 9
 cttgccccac agggcagtaa cgg 23

<210> 10
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Target sequence on HBB-03 and PAM
 <400> 10
 cacgttcacc ttgccccaca ggg 23

<210> 11
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Target sequence on HBB-04 and PAM

<400> 11
ccacgttcac cttgccccac agg 23
<210> 12
<211> 23
<212> DNA
<213> Artificial Sequence
<220><223> Target sequence on EMX1-05 and PAM
<400> 12
tgtactttgt cctccggtg tgg 23