

(43) 공개일자 2022년10월14일

- (71) 출원인  
**프레지던트 앤드 펠로우즈 오브 하바드 칼리지**  
 미합중국, 메사추세츠 02138, 캠브리지, 퀸시스트리트17
- (72) 발명자  
**처치, 조지 엠.**  
 미국 02446 매사추세츠주 브록클린 켄트 스트리트 218  
**말리, 프라샨트**  
 미국 02143 매사추세츠주 서머빌 아파트먼트 54  
 비컨 스트리트 88  
**양, 루한**  
 미국 02143 매사추세츠주 서머빌 아파트먼트 56  
 비컨 스트리트 278
- (74) 대리인  
**양영준, 김영**

(54) 발명의 명칭 RNA-가이드된 인간 게놈 조작

대표도



(52) CPC특허분류

**C12N 15/113** (2013.01)

**C12N 15/907** (2013.01)

**C12N 9/22** (2013.01)

**C12N 2310/20** (2017.05)

---

## 명세서

### 청구범위

#### 청구항 1

진핵 세포에, crRNA 부분 및 tracrRNA 부분을 포함하는 가이드 RNA를 제공하고, 여기서 crRNA 부분은 표적 핵산 서열에 상보적인 스페이서 서열을 포함하고, crRNA 부분 및 tracrRNA 부분은 함께 혼성화되고, 가이드 RNA가 상보적인 표적 핵산 서열에 결합하는 것인 단계, 및

진핵 세포에, 가이드 RNA와 상호작용하는 Cas9 단백질을 제공하는 단계를 포함하는, 시험관내 또는 생체외에서 진핵 세포의 게놈 내 표적 핵산 서열을 변경시키는 방법.

#### 청구항 2

제1항에 있어서, Cas9 단백질이 Cas9 효소이고, 표적 핵산 서열이 절단되는 것인 방법.

#### 청구항 3

제1항에 있어서,  
세포에 가이드 RNA를 코딩하는 제1 핵산을 도입함으로써 가이드 RNA가 세포에 제공되고,  
세포에 Cas9 단백질을 코딩하는 제2 핵산을 도입함으로써 Cas9 단백질이 세포에 제공되며,  
세포가 가이드 RNA 및 Cas9 단백질을 생성하는 것인 방법.

#### 청구항 4

제3항에 있어서, 제1 핵산이 바이러스성 전달 방법을 이용하여 세포로 도입되는 것인 방법.

#### 청구항 5

제3항에 있어서, 제1 핵산 및/또는 제2 핵산이 형질감염에 의해 세포로 도입되는 것인 방법.

#### 청구항 6

제3항에 있어서, 제1 핵산이 아데노 관련 바이러스를 이용하여 세포로 도입되는 것인 방법.

#### 청구항 7

제1항에 있어서, Cas9 단백질이 Cas9 효소, Cas9 니카제, 뉴클레아제-기능 없는 Cas9, 변형된 Cas9 또는 Cas9의 상동체인 방법.

#### 청구항 8

제1항에 있어서, 진핵 세포가 효모 세포, 식물 세포 또는 포유동물 세포인 방법.

#### 청구항 9

제1항에 있어서, 세포가 인간 세포인 방법.

#### 청구항 10

제1항에 있어서,  
상이한 표적 핵산 서열에 상보적인 스페이서 서열을 각각 갖는 다수의 가이드 RNA가 진핵 세포에 제공되고, 여기서 가이드 RNA는 상이한 표적 핵산 서열에 결합하며,  
Cas9 단백질이 다수의 가이드 RNA와 상호작용하는 것인 방법.

#### 청구항 11

제1항에 있어서, 가이드 RNA가 crRNA-tracrRNA 융합체인 방법.

#### 청구항 12

제1항에 있어서, Cas9 단백질이 인간 코돈-최적화된 핵산에 의해 코딩된 것인 방법.

#### 청구항 13

제1항에 있어서, Cas9 단백질이 핵 국제화 신호를 포함하는 것인 방법.

#### 청구항 14

제1항에 있어서, 유전자와 내 표적 핵산 서열에 상보적인 가이드 RNA가 세포에 제공되며, 여기서 가이드 RNA는 표적 핵산 서열에 결합하고, Cas9 단백질이 표적 핵산 서열을 절단하는 Cas9 효소이며, 개재 핵산이 유전자좌로부터 결실되는 것인 방법.

#### 청구항 15

제1항에 있어서, Cas9 단백질이 표적 핵산 서열을 절단하는 Cas9 효소이며, 진핵 세포에 제공된 공여자 핵산이 표적 핵산 서열에 삽입되는 것인 방법.

#### 청구항 16

제1항에 있어서,

Cas9 단백질이 뉴클레아제-기능 없는 Cas9이고, 표적 핵산 서열에 결합하는 가이드 RNA와 상호작용하고,

Cas9 단백질이 이에 부착된 FokI 뉴클레아제 도메인을 갖고,

인접한 가이드 RNA 및 Cas9 단백질 복합체의 FokI 뉴클레아제 도메인 이량체화시 표적 핵산 서열이 절단되는 것인 방법.

#### 청구항 17

제1항에 있어서, 상기 가이드 RNA가 핵산 GUUUUAGAGCUAGAAUAGCAAGUUAUUAAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGUGCUUUU (서열식별번호: 45)를 포함하는 것인 방법.

#### 청구항 18

제1항에 있어서, Cas9 단백질이 뉴클레아제-기능 없는 Cas9이고, 전사 활성화 또는 억제 도메인을 포함하며, 표적 유전자 발현이 조정되는 것인 방법.

#### 청구항 19

제1항에 있어서, Cas9 단백질이 뉴클레아제-기능 없는 Cas9이고, 형광성 단백질을 포함하며, 형광성 단백질이 가시화되는 것인 방법.

#### 청구항 20

제1항에 있어서, Cas9 단백질이 뉴클레아제-기능 없는 Cas9이고, Cas9 단백질이 이에 부착된 (i) 단백질 결합된 유기 형광단, (ii) 핵산 결합된 유기 형광단, (iii) 양자 도트, (iv) 분자 비콘, (v) 에코 프로브 또는 (vi) 다 가 리간드-결합성 단백질 도메인을 포함하는 것인 방법.

#### 청구항 21

가이드 RNA 또는 상기 가이드 RNA를 코딩하는 제1 핵산이며, 여기서 가이드 RNA는 crRNA 부분 및 tracrRNA 부분을 포함하고, 여기서 crRNA 부분은 표적 핵산 서열에 상보적인 스페이서 서열을 포함하고, crRNA 부분 및 tracrRNA 부분은 함께 혼성화되고, 여기서 가이드 RNA는 상보적인 표적 핵산 서열에 결합하는 것인, 가이드 RNA

또는 상기 가이드 RNA를 코딩하는 제1 핵산; 및

가이드 RNA와 상호작용하는 Cas9 단백질을 코딩하는 제2 핵산을 포함하는, RNA-가이드된 게놈 편집 조합물.

## 청구항 22

제21항의 RNA-가이드된 게놈 편집 조합물을 함유하는 단리된 진핵 세포이며, 여기서 단리된 진핵 세포는 인간의 배아를 제외하는 것인 단리된 진핵 세포.

## 청구항 23

제1항에 있어서, 진핵 세포가 줄기 세포인 방법.

## 청구항 24

제12항에 있어서, Cas9 단백질이 서열식별번호: 22를 포함하는 서열에 의해 코딩되는 것인 방법.

## 청구항 25

제1항에 있어서, 가이드 RNA가 100개 내지 250개의 뉴클레오타이드를 포함하는 crRNA-tracrRNA 융합 전사체인 방법.

## 발명의 설명

### 기술 분야

#### [0001] 관련 출원 데이터

[0002] 본 출원은 2013년 3월 13일에 출원된 미국 가출원 번호 61/779,169 및 2012년 12월 17일에 출원된 미국 가출원 번호 61/738,355를 우선권 주장하며, 이들 가출원 각각은 모든 목적상 그 전문이 본원에 참조로 포함된다.

#### [0003] 정부 권리의 언급

[0004] 본 발명은 국립 보건 연구소가 수주한 P50 HG005550 하의 정부 지원으로 만들어졌다. 정부가 본 발명에 대한 특정 권리를 갖고 있다.

### 배경 기술

[0005] 박테리아성 및 고세균성 CRISPR 시스템은 침입 바이러스 및 플라스미드 DNA 내에 존재하는 상보적 서열의 분해를 지시하는 Cas 단백질과 복합체화된 crRNA에 좌우된다 (1-3). 최근, 에스. 피오게네스(*S. pyogenes*) 유형 II CRISPR 시스템을 시험관내에서 재구성한 결과, 정상적으로 트랜스-코딩된 tracrRNA와 융합된 crRNA가, 이러한 crRNA와 부합되는 표적 DNA 서열을 서열-특이적으로 절단시키도록 Cas9 단백질을 지시하는 데 충분한 것으로 입증되었다 (4).

### 발명의 내용

#### [0006] 개요

[0007] 본 개시내용의 끝 부분에 열거되어 있는 본 개시내용 참고문헌은 숫자로 표시되어 기록된다. 이러한 숫자에 상응하는 문헌은, 마치 그것이 완전히 인용된 바와 같이 이러한 숫자에 상응하는 지원 참고문헌으로서 본 명세서 내에 참조로 포함된다.

[0008] 본 개시내용의 한 측면에 따르면, 진핵 세포를, 게놈 DNA에 상보적인 RNA 및 이러한 RNA와 상호작용하는 효소를 포함한 2개 성분 시스템으로 형질감염시킨다. 상기 RNA 및 효소는 상기 세포에 의해 발현된다. 이어서, 상기 RNA/효소 복합체의 RNA는 상보적 게놈 DNA에 결합한다. 이어서, 상기 효소는 게놈 DNA의 절단과 같은 기능을 수행한다. RNA는 약 10개의 뉴클레오타이드 내지 약 250개의 뉴클레오타이드를 포함한다. RNA는 약 20개의 뉴클레오타이드 내지 약 100개의 뉴클레오타이드를 포함한다. 특정의 측면에 따르면, 상기 효소는 이러한 효소를 조작하기 위한 부위 특이적 방식으로 목적하는 어떠한 기능도 수행할 수 있다. 한 측면에 따르면, 진핵 세포는 효모 세포, 식물 세포 또는 포유동물 세포이다. 한 측면에 따르면, 상기 효소는 RNA 서열에 의해 표적화된 게놈 서

열을 절단함으로써 [참고문헌 (4-6) 참조], 게놈적으로 변경된 진핵 세포를 창출시킨다.

[0009] 한 측면에 따르면, 본 개시내용은 게놈 DNA에 상보적인 RNA를 코딩하는 핵산을 인간 세포의 게놈 내로 포함시키고, 게놈 DNA 상에서 목적하는 기능을 수행하는 효소를 코딩하는 핵산을 상기 세포의 게놈 내로 포함시킴으로써, 상기 인간 세포를 유전적으로 변경시키는 방법을 제공한다. 한 측면에 따르면, 상기 RNA 및 효소가 발현되고, 한 측면에 따르면, 이러한 RNA는 상보적 게놈 DNA에 혼성화된다. 한 측면에 따르면, 효소는 RNA가 상보적 게놈 DNA에 혼성하는 경우에, 부위 특이적 방식으로 목적하는 기능, 예컨대 절단을 수행하도록 활성화된다. 한 측면에 따르면, 상기 RNA 및 효소는 박테리아성 유형 II CRISPR 시스템의 성분들이다.

[0010] 한 측면에 따르면, 진핵 세포를, 진핵 세포의 게놈 DNA에 상보적인 RNA를 코딩하는 핵산으로 형질감염시키고; 진핵 세포를, 상기 RNA와 상호작용하고 게놈 DNA를 부위 특이적 방식으로 절단하는 효소를 코딩하는 핵산으로 형질감염시키는 것을 포함하며, 여기서 상기 세포는 상기 RNA 및 효소를 발현하고, RNA는 상보적 게놈 DNA에 결합하고, 효소는 게놈 DNA를 부위 특이적 방식으로 절단하는 것인, 진핵 세포를 변경시키는 방법이 제공된다. 한 측면에 따르면, 상기 효소는 Cas9 또는 변형된 Cas9 또는 Cas9의 상동체이다. 한 측면에 따르면, 진핵 세포는 효모 세포, 식물 세포 또는 포유동물 세포이다. 한 측면에 따르면, RNA는 약 10개 내지 약 250개의 뉴클레오타이드를 포함한다. 한 측면에 따르면, RNA는 약 20개 내지 약 100개의 뉴클레오타이드를 포함한다.

[0011] 한 측면에 따르면, 인간 세포를, 진핵 세포의 게놈 DNA에 상보적인 RNA를 코딩하는 핵산으로 형질감염시키고; 인간 세포를, 상기 RNA와 상호작용하고 게놈 DNA를 부위 특이적 방식으로 절단하는 효소를 코딩하는 핵산으로 형질감염시키는 것을 포함하며, 여기서 상기 인간 세포는 상기 RNA 및 효소를 발현하고, RNA는 상보적 게놈 DNA에 결합하고, 효소는 게놈 DNA를 부위 특이적 방식으로 절단하는 것인, 인간 세포를 변경시키는 방법이 제공된다. 한 측면에 따르면, 상기 효소는 Cas9 또는 변형된 Cas9 또는 Cas9의 상동체이다. 한 측면에 따르면, RNA는 약 10개 내지 약 250개의 뉴클레오타이드를 포함한다. 한 측면에 따르면, RNA는 약 20개 내지 약 100개의 뉴클레오타이드를 포함한다.

[0012] 한 측면에 따르면, 진핵 세포를, 진핵 세포의 게놈 DNA 상의 상이한 부위에 상보적인 RNA를 코딩하는 다수의 핵산으로 형질감염시키고; 진핵 세포를, 상기 RNA와 상호작용하고 게놈 DNA를 부위 특이적 방식으로 절단하는 효소를 코딩하는 핵산으로 형질감염시키는 것을 포함하며, 여기서 상기 세포는 상기 RNA 및 효소를 발현하고, RNA는 상보적 게놈 DNA에 결합하고, 효소는 게놈 DNA를 부위 특이적 방식으로 절단하는 것인, 진핵 세포를 다수의 게놈 DNA 부위에서 변경시키는 방법이 제공된다. 한 측면에 따르면, 상기 효소는 Cas9이다. 한 측면에 따르면, 진핵 세포는 효모 세포, 식물 세포 또는 포유동물 세포이다. 한 측면에 따르면, RNA는 약 10개 내지 약 250개의 뉴클레오타이드를 포함한다. 한 측면에 따르면, RNA는 약 20개 내지 약 100개의 뉴클레오타이드를 포함한다.

## 도면의 간단한 설명

### 발명을 실시하기 위한 구체적인 내용

[0014] 상세한 설명

[0015] 한 측면에 따르면, C-말단 SV40 핵 국제화 신호를 보유하고 있는 Cas9 단백질의 인간 코돈-최적화 버전을 합성하고, 이를 포유동물 발현 시스템 내로 클로닝한다 (도 1a 및 도 3a). 따라서, 도 1은 조작된 유형 II CRISPR 시스템을 이용하는 인간 세포에서의 게놈 편집에 관한 것이다. 도 1a에 도시된 바와 같이, 인간 세포에서 RNA-가이드된 유전자 표적화는 C-말단 SV40 핵 국제화 신호를 보유하고 있는 Cas9 단백질을, 인간 U6 폴리머라제 III 프로모터로부터 발현된 하나 이상의 가이드 RNA (gRNA)와 함께 공동-배양하는 것을 포함한다. Cas9는 DNA 이중체를 풀어주고, gRNA에 의한 표적 서열의 인식시 양 가닥을 절단시키지만, 이는 정확한 프로토스페이스 (protospacer)-인접한 모티프 (PAM)가 3' 말단에 존재하는 경우에만 가능하다. 형태 GN<sub>20</sub>GG의 어떠한 게놈 서열도 원칙적으로 표적화될 수 있다. 도 1b에 도시된 바와 같이, 게놈적으로 통합된 GFP 코딩 서열은 AAVS1 유전자좌로부터의 68 bp 게놈 단편 및 정지 코돈을 삽입시킴으로써 붕괴된다. 적당한 공여자 서열과의 상동 재조합 (HR)에 의해 GFP 서열을 복원시키면, FACS에 의해 정량화될 수 있는 GFP+ 세포가 생성된다. T1 및 T2 gRNA는 상기 AAVS1 단편 내의 서열을 표적으로 한다. TAL 이펙터 뉴클레아제 이중이량체 (TALEN)의 2개 절반에 대한 결합 부위에 밑줄이 그어져 있다. 도 1c에 도시된 바와 같이, 막대 그래프는 FACS에 의해 측정된 바와 같이, 표적 유전자좌에서 T1, T2 및 TALEN-매개된 뉴클레아제 활성화에 의해 유도된 HR 효율을 도시한다. 표적화

된 세포의 대표적인 FACS 플롯 및 현미경 영상이 아래에 도시된다 (눈금 막대는 100 마이크로미터이다). 데이터는 평균  $\pm$  SEM (N=3)이다.

[0016] 한 측면에 따르면, Cas9가 관심 서열을 절단하도록 지시하기 위하여, 가이드 RNA (gRNA)로서 후술되는 crRNA-tracrRNA 융합 전사체를 인간 U6 폴리머라제 III 프로모터로부터 발현시킨다. 한 측면에 따르면, gRNA는 상기 세포에 의해 직접적으로 전사된다. 이러한 측면은 유리하게, 박테리아성 CRISPR 시스템에 의해 이용된 RNA 프로세싱 기구를 재구성하는 것을 피한다 (도 1a 및 도 3b) [참고문헌 (4, 7-9) 참조]. 한 측면에 따르면, 20 bp crRNA 표적 다음의 PAM (프로토스페이서-인접한 모티프) 서열 -NGG 및 G와 함께 개시하는 U6 전사를 이용하여 게놈 DNA를 변경시키는 방법이 제공된다. 이러한 측면에 따르면, 표적 게놈 부위는 GN<sub>20</sub>GG의 형태로 존재한다 (도 3c 참조).

[0017] 한 측면에 따르면, 본원에 기재된 게놈 조작 방법의 기능성을 시험하기 위하여 기존에 보고된 바와 유사하게 [참고문헌 (10) 참조] 293T 세포에서의 GFP 리포터 검정 (도 1b)을 개발하였다. 한 측면에 따르면, 발현된 단백질 단편을 비-형광성으로 만드는 AAVS1 유전자좌로부터의 68 bp 게놈 단편 및 정지 코돈을 삽입시킴으로써 붕괴된, 게놈적으로 통합된 GFP 코딩 서열을 보유하는 안정적인 세포주를 확립하였다. 적당한 복구 공여자를 이용한 상동 재조합 (HR)이 정상 GFP 서열을 복원시킬 수 있는데, 이로써 생성된 GFP+ 세포를 유동 활성화 세포 분류 (FACS)에 의해 정량화할 줄 수 있다.

[0018] 한 측면에 따르면, 상동 재조합 (HR) 방법이 제공된다. 개재 AAVS1 단편을 표적으로 하는 2개의 gRNA T1 및 T2를 구축한다 (도 1b). 그들의 활성을, 동일한 영역을 표적으로 하는 앞서 기재된 TAL 이펙터 뉴클레아제 이중 이량체 (TALEN) [참고문헌 (11) 참조]의 활성과 비교하였다. 3가지 표적화 시약 모두를 이용하여 성공적인 HR 현상을 관찰하였는데, T1 및 T2 gRNA를 이용하는 유전자 교정율은 각각 3% 및 8%에 근접하였다 (도 1c). 이러한 RNA-매개된 편집 프로세스는 현저하게 신속하였는데, 첫 번째 검출가능한 GFP+ 세포는 형질감염 후 대략 20 시간째에 출현되었고, 이는 AAVS1 TALEN의 경우의 대략 40시간과 비교되었다. 복구 공여자, Cas9 단백질 및 gRNA를 동시에 도입한 경우에만 HR이 관찰되었는데, 이는 모든 성분이 게놈 편집에 요구된다는 것을 입증시켜 준다 (도 4). Cas9/crRNA 발현과 연관된 명백한 독성이 전혀 인지되지 않았지만, ZFN 및 TALEN을 이용한 작업 결과, 단지 1개 가닥에 흠을 내는 것이 독성을 추가로 저하시키는 것으로 나타났다. 따라서, 시험관내에서 Nickase로서 기능하는 것으로 공지되어 있는 Cas9D10A 돌연변이체를 시험하였는데, 이는 유사한 HR을 산출시켰지만, 보다 낮은 비-상동 말단 결합 (NHEJ) 비율을 나타내었다 (도 5) [참고문헌 (4, 5) 참조]. 관련 Cas9 단백질이 PAM의 상류에 있는 약 6 bp를 절단하는 것으로 밝혀진 것과 일관되게 (4), NHEJ 데이터는 대부분의 결실 또는 삽입이 표적 서열의 3' 말단에서 발생하였다는 것을 입증시켜 주었다 (도 5b). 또한, 표적 게놈 부위를 돌연변이시키면, gRNA가 그 유전자좌에서 HR을 수행하지 못하게 하는 것으로 입증되었는데, 이는 CRISPR-매개된 게놈 편집이 서열 특이적이라는 것을 명확하게 보여준다 (도 6). GFP 유전자 내의 2개의 gRNA 표적화 부위, 및 또한 DNA 메틸 트랜스퍼라제 3a (DNMT3a) 및 DNMT3b 유전자의 상동 영역으로부터의 3개의 부가 gRNA 표적화 단편이, 조작된 리포터 세포주에서 상당한 HR을 서열 특이적으로 유도시킬 수 있는 것으로 밝혀졌다 (도 7, 8). 이들 결과를 취합해 보면, 인간 세포에서의 RNA-가이드된 게놈 표적화로 인해, 다중 표적 부위 전체에 걸쳐 강건한 HR이 유도되는 것으로 입증된다.

[0019] 특정의 측면에 따르면, 천연 유전자좌를 변형시켰다. gRNA를 사용하여 염색체 19 상의 PPP1R12C 유전자 내에 위치한 AAVS1 유전자좌를 표적화하였는데, 이러한 유전자좌는 293T, K562, 및 PGP1 인간 iPS 세포 [참고문헌 (12) 참조] 내의 대부분의 조직 전반에 걸쳐 편재해서 발현되고 (도 2a), 상기 표적화된 유전자좌의 차세대 서열분석에 의해 그 결과가 분석된다. 따라서, 도 2는 다중 세포 유형 중의 천연 AAVS1 유전자좌의 RNA-가이드된 게놈 편집에 관한 것이다. 도 2a에 나타난 바와 같이, T1 (적색) 및 T2 (녹색) gRNA는 염색체 19 AAVS1 유전자좌 내의 PPP1R12C 유전자의 인트론 중의 서열을 표적으로 한다. 도 2b에 나타난 바와 같이, 차세대 서열분석에 의해 정량화된 바와 같은 Cas9와 T1 또는 T2 gRNA의 발현 후 293T, K562, 및 PGP1 iPS 세포에서 NHEJ에 의해 유발된 결실의 총 개수 및 위치가 제공된다. 적색과 녹색 점선은 T1 및 T2 gRNA 표적화 부위의 경계를 표시한다. T1 및 T2 gRNA에 대한 NHEJ 빈도는 293T에서 각각 10% 및 25%였고, K562에서는 각각 13% 및 38%였으며, PGP1 iPS 세포에서는 각각 2% 및 4%였다. 도 2c에 나타난 바와 같이, AAVS1 유전자좌에서 HR에 대한 DNA 공여자 구조, 및 성공적인 표적화 현상을 검출하기 위한 서열분석용 프라이머 (화살표)의 위치가 도시되어 있다. 도 2d에 나타난 바와 같이, 형질감염 후 3일째의 PCR 검정은 공여자, Cas9 및 T2 gRNA를 발현하는 세포만이 성공적인 HR 현상을 나타낸다는 것을 명확히 보여준다. 도 2e에 나타난 바와 같이, 성공적인 HR은, 게놈-공여자 경계와 공여자-삽입물 경계 둘 다에서 예상되는 DNA 염기가 존재한다는 것을 보여주는 PCR 앰플리콘의 생거 (Sanger) 서열분석에 의해 입증되었다. 도 2f에 나타난 바와 같이, 293T 세포의 성공적으로 표적화된 클론을 2



주 동안 푸로마이신을 이용하여 선별하였다. 2가지 대표적인 GFP+ 클론의 현미경 영상이 도시되어 있다 (눈금 막대는 100 마이크로미터이다).

[0020] GFP 리porter 검정에 대한 결과와 동일하게, 3가지 세포 유형 모두에 대한 내인성 유전자좌에서 다수의 NHEJ 현상이 관찰되었다. 2개의 gRNA T1 및 T2는 293T에서 각각 10% 및 25%의 NHEJ 비율을 달성하였고, K562에서는 각각 13% 및 38%의 NHEJ 비율을 달성하였으며, PGP1-iPS 세포에서는 각각 2% 및 4%의 NHEJ 비율을 달성하였다 (도 2b). 이들 세포 유형 중 어느 것에서도 NHEJ를 유도시키기 위해 요구되는 Cas9 및 crRNA 발현으로부터는 명백한 독성이 전혀 관찰되지 않았다 (도 9). 예상된 바와 같이, T1 및 T2에 대한 NHEJ-매개된 결실이 표적 부위 위치 주변의 중심에 있었는데, 이는 이러한 표적화 프로세스의 서열 특이성을 추가로 검증시켜 준다 (도 9, 10, 11). T1 gRNA와 T2 gRNA 둘 다를 동시에 도입하면, 개재 19 bp 단편의 고 효율 결실이 발생하였는데 (도 10), 이는 게놈 유전자좌의 다중화 편집이 이러한 접근법을 이용하여 실행가능하다는 것을 입증해준다.

[0021] 한 측면에 따르면, HR을 이용하여 dsDNA 공여자 구축물 [참고문헌 (13) 참조] 또는 올리고 공여자를 천연 AAVS1 유전자좌 내로 통합시킨다 (도 2c, 도 12). HR-매개된 통합은 PCR에 의한 접근법 (도 2d, 도 12)과 생거 서열 분석에 의한 접근법 (도 2e) 둘 다를 이용하여 확증하였다. 2주 간에 걸친 푸로마이신 선별을 이용하여, 변형된 세포의 풀로부터 293T 또는 iPS 클론을 용이하게 유도하였다 (도 2f, 도 12). 이들 결과는 Cas9가 인간 세포 내의 내인성 유전자좌에서 외래 DNA를 효율적으로 통합시킬 수 있다는 것을 입증해 준다. 따라서, 본 개시내용의 한 측면은 상동 재조합과 Cas9를 이용하여 외래 DNA를 세포의 게놈 내로 통합시키는 방법을 포함한다.

[0022] 한 측면에 따르면, 관심 유전자좌 내의 화합성 서열에 부합하도록 gRNA 발현 벡터의 서열을 간단히 변형시킴으로써 다른 게놈 부위를 변형시키도록 용이하게 적응시킬 수 있는 RNA-가이드된 게놈 편집 시스템이 제공된다. 이러한 측면에 따르면, 인간 게놈 내의 유전자의 약 40.5% 엑손을 표적으로 하는 190,000개의 특이적으로 gRNA-표적 가능한 서열이 생성되었다. 이들 표적 서열은 DNA 어레이 상에서의 다중화 합성과 화합성인 200 bp 포맷 내로 혼입하였다 [참고문헌 (14) 참조] (도 13). 이러한 측면에 따르면, 인간 게놈 내의 잠재적 표적 부위의 즉시 사용 가능한 게놈 규모에서의 참조와 다중화 gRNA 합성에 대한 방법론이 제공된다.

[0023] 한 측면에 따르면, 본원에 기재된 하나 이상 또는 다수의 RNA/효소 시스템을 사용하여 다수의 위치에서 세포의 게놈을 변경시킴으로써 세포 내에서의 게놈 변경을 다중화시키는 방법이 제공된다. 한 측면에 따르면, 표적 부위는 gRNA의 3' 말단에서 8 내지 12개 염기 "시드 서열" 및 PAM 서열 NGG와 완벽하게 부합된다. 특정의 측면에 따르면, 남아 있는 8 내지 12개 염기에 대해서는 완벽한 부합이 요구되지 않는다. 특정의 측면에 따르면, Cas9는 5' 말단에서 단일 부조합물과 함께 기능할 것이다. 특정의 측면에 따르면, 표적 유전자좌의 근원적인 염색질 구조와 후생적 상태는 Cas9 기능의 효율에 영향을 미칠 수 있다. 특정의 측면에 따르면, 보다 높은 특이성을 지닌 Cas9 상동체가 유용한 효소로서 포함된다. 통상의 기술자는 적합한 Cas9 상동체를 확인하거나 조작할 수 있을 것이다. 한 측면에 따르면, CRISPR-표적 가능한 서열은 상이한 PAM 요구 조건 [참고문헌 (9) 참조] 또는 지향 진화를 갖는 것을 포함한다. 한 측면에 따르면, Cas9 뉴클레아제 도메인 중 하나를 불활성화시키면, NHEJ에 대한 HR의 비가 증가되고 독성이 감소될 수 있는 반면 (도 3a, 도 5) (4, 5), 양 도메인을 불활성화시키면, Cas9가 재표적 가능한 DNA 결합성 단백질로서 기능할 수 있게 된다. 본 개시내용의 실시양태는 합성 생물학 [참고문헌 (21, 22) 참조], 유전자 네트워크의 직접적이고 다중화시킨 교란 [참고문헌 (13, 23) 참조], 및 표적화된 생체의 [참고문헌 (24-26) 참조] 및 생체내 유전자 요법 [참고문헌 (27) 참조]에 있어서 광범위하게 활용된다.

[0024] 특정의 측면에 따르면, 생물학적 발견과 생체내 스크리닝을 위한 모델 시스템으로서의 "재-조작가능한 유기체"가 제공된다. 한 측면에 따르면, 유도성 Cas9 트랜스 유전자를 보유하고 있는 "재-조작가능한 마우스"가 제공되고, gRNA 표적화 다중 유전자 또는 조절성 요소의 라이브러리를 국소성 전달하면 (예를 들어, 아테노 관련 바이러스를 이용한다), 표적 조직 유형 내에 종양을 발생시키는 돌연변이에 관하여 스크리닝할 수 있다. 이펙터 도메인 (예컨대, 활성화인자)을 보유하고 있는 Cas9 상동체 또는 뉴클레아제-기능없는 변이체를 사용하면, 생체 내에서 유전자를 다중화 활성화시키거나 억제할 수 있다. 이러한 측면에 따르면, 조직-재생, 전환 분화 등과 같은 표현형을 가능하게 해주는 요인에 대하여 스크리닝할 수 있었다. 특정의 측면에 따르면, (a) DNA-어레이를 사용하면, 규정된 gRNA 라이브러리를 다중화 합성할 수 있게 되고 (도 13 참조); (b) 크기가 작은 gRNA (도 3b 참조)를 패키징하고, 다수의 비-바이러스성 또는 바이러스성 전달 방법을 이용하여 전달한다.

[0025] 한 측면에 따르면, 게놈 조작 적용을 위하여 "닉카제"를 이용하여 관찰된 보다 낮은 독성은, Cas9 뉴클레아제 도메인 중 하나를 불활성시키고, RNA와 쌍을 형성한 DNA 가닥 염기에 홈을 내거나 또는 그의 보체에 홈을 냄으로써 달성된다. 양 도메인을 불활성화시키면, Cas9가 재-표적 가능한 DNA 결합성 단백질로서 기능할 수 있게



된다. 한 측면에 따르면, 이러한 Cas9 재-표적 가능한 DNA 결합성 단백질은

- [0026] (a) 표적 유전자 발현 (이는 염색질 재형성, 히스톤 변형, 침묵, 절연, 전사 기구와의 직접적인 상호작용을 포함하지만, 그에 제한되지 않는다)을 조정하기 위한 전사 활성화 또는 억제 도메인에 부착되거나;
- [0027] (b) 인접한 gRNA-Cas9 복합체의 이량체 형성 여부에 따라 '고도의 특이적' 게놈 편집을 가능하게 해주는 FokI과 같은 뉴클레아제 도메인에 부착되거나;
- [0028] (c) 게놈 유전자좌와 염색체 역학을 가시화하기 위하여 형광성 단백질에 부착되거나;
- [0029] (d) 기타 형광성 분자, 예컨대 단백질 또는 핵산 결합된 유기 형광단, 양자 도트, 분자 비콘 및 에코 프로브 또는 분자 비콘 대체물에 부착되거나; 또는
- [0030] (e) 게놈 규모에서의 3D 구조의 프로그램 가능한 조작을 가능하게 해주는 다가 리간드-결합성 단백질 도메인에 부착된다.
- [0031] 한 측면에 따르면, 전사 활성화 및 억제 성분은 자연적으로 또는 합성적으로 직교하는 CRISPR 시스템을 이용할 수 있으므로, gRNA는 Cas의 활성화인자 또는 억제인자 부류와만 결합된다. 이로써, gRNA의 대형 세트가 다중 표적을 조정할 수 있다.
- [0032] 특정의 측면에 따르면, gRNA를 사용하는 것이 mRNA를 사용하는 것 보다 다중화할 수 있는 능력을 제공하는데, 이는 부분적으로는 보다 작은 크기에 기인한다 (뉴클레오타이드 길이가 각각 100 대 2,000이다). 핵산 전달이 바이러스성 패키징에서와 같이 크기 제한되는 경우에 특히 유용하다. 이로써, 절단, 니킹, 활성화 또는 억제 - 또는 그의 조합의 다양한 사례가 가능해진다. 다중 조절성 표적을 용이하게 표적으로 할 수 있는 능력으로 인해, 특이적 조절성 인자의 하류에 있는 자연 조절성 회로 (예를 들어, 섬유모세포를 iPSC 내로 재프로그래밍하는 데에 사용된 4개의 mRNA)에 제한되지 않고서도 거친 또는 미세 조정 또는 조절성 네트워크가 가능하다. 다중화 적용의 예는 다음을 포함한다:
- [0033] 1. 인간 (또는 동물) 조직/기관 이식을 위한 (주요 및 소수) 조직적합성 대립 유전자, 일배체형 및 유전자형을 확립시킨다. 이러한 측면으로 인해, 예를 들어 HLA 동형접합성 세포주 또는 인간화 동물 품종 - 또는 - 상기 HLA 대립 유전자를 그 밖의 바람직한 세포주 또는 품종 상에 중첩시킬 수 있는 gRNA 세트가 생성된다.
- [0034] 2. 단일 세포 (또는 세포 컬렉션)에서 다중 시스-조절성 요소 (CRE = 전사, 스플라이싱, 번역, RNA 및 단백질 폴딩, 분해 등을 위한 신호) 돌연변이를, 정상적인 발생 또는 병리학적, 합성 또는 제약 시나리오에서 발생될 수 있는 조절성 상호작용의 복합 세트를 효율적으로 연구하는 데 사용할 수 있다. 한 측면에 따르면, 상기 CRE는 다소 직교하여 (또는 직교되도록 만들 수 있어) (즉, 혼성이 적다), 많은 것들을 한 가지 환경 하에, 예를 들어 비용이 많이 드는 동물 배아 시간 시리즈에서 시험할 수 있다. 한 가지 예시되는 적용은 RNA 형광성 계내 서열분석 (FISSeq)을 이용하는 것이다.
- [0035] 3. CRE 돌연변이 및/또는 CRE의 후생적 활성화 또는 억제의 다중 조합을 이용하여, iPSC 또는 ESC 또는 기타 줄기 세포 또는 비-줄기 세포를, 제약을 시험할 목적 (소분자, 단백질, RNA, 세포, 동물, 식물 또는 미생물 세포, 에어로솔 및 기타 전달 방법), 이식 전략, 개인화 전략 등의 목적을 위해 칩 상의 기관 또는 기타 세포 및 기관 배양물에 사용하기 위한 어떠한 세포 유형이나 세포 유형의 조합물로 변경시키거나 재프로그래밍시킬 수 있다.
- [0036] 4. 의학 유전학을 위한 진단 시험 (및/또는 DNA 서열분석)에 사용하기 위한 다중 돌연변이체 인간 세포를 제조한다. 인간 게놈 대립 유전자 (또는 후생적 표식)의 염색체 위치와 맥락이 임상 유전 진단의 정확성에 영향을 미칠 수 있는 정도로, 이소성 (트랜스제닉으로서 공지되기도 함) 위치에서 또는 합성 DNA의 별개의 조각에서 보다는 오히려 참조 게놈 내의 정확한 위치에 존재하는 대립 유전자를 갖는 것이 중요하다. 한 실시양태는 각 진단 인간 SNP 또는 구조적 변이체 당 1개의 독립적인 세포주의 시리즈이다. 또 다른 한편, 한 실시양태는 동일한 세포 내의 대립 유전자의 다중 세트를 포함한다. 일부 경우에, 하나의 유전자 (또는 다중 유전자) 내에서의 다중화 변화는 독립적인 시험의 가정 하에 바람직할 것이다. 다른 경우에는, 대립 유전자의 특별한 일배체형 조합으로 인해, 일배체형 상을 정확하게 확립시켜 주는 서열분석 (유전자형별) 방법을 시험할 수 있게 된다 (즉, 유전자의 1개 카피 또는 둘 다의 카피가 개개인 또는 체세포 유형에 영향을 미치는지에는 관계가 없다).
- [0037] 5. 반복 요소 또는 내인성 바이러스성 요소는, 미생물, 식물, 동물 또는 인간 세포에서 조작된 Cas + gRNA 시스템을 이용하여 표적화하여 해로운 전위를 저하시킬 수 있거나 또는 서열분석이나 기타 분석 게놈/전사체/단백체/진단 도구 (여기서, 거의 동일한 카피가 문제를 발생시킬 수 있다)에 도움을 줄 수 있다.

[0038] 앞서 언급된 섹션에서 숫자로써 확인된 다음 참고문헌은 그 전문이 본원에 참조로 포함된다.

1. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, *Nature* 482, 331 (Feb 16, 2012).
2. D. Bhaya, M. Davison, R. Barrangou, *Annual review of genetics* 45, 273 (2011).
3. M. P. Terns, R. M. Terns, *Current opinion in microbiology* 14, 321 (Jun, 2011).
4. M. Jinek *et al.*, *Science* 337, 816 (Aug 17, 2012).
5. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579 (Sep 25, 2012).
6. R. Sapranaukas *et al.*, *Nucleic acids research* 39, 9275 (Nov, 2011).
7. T. R. Brummelkamp, R. Bernards, R. Agami, *Science* 296, 550 (Apr 19, 2002).
8. M. Miyagishi, K. Taira, *Nature biotechnology* 20, 497 (May, 2002).
9. E. Deltcheva *et al.*, *Nature* 471, 602 (Mar 31, 2011).
10. J. Zou, P. Mali, X. Huang, S. N. Dowey, L. Cheng, *Blood* 118, 4599 (Oct 27, 2011).
11. N. E. Sanjana *et al.*, *Nature protocols* 7, 171 (Jan, 2012).
12. J. H. Lee *et al.*, *PLoS Genet* 5, e1000718 (Nov, 2009).
13. D. Hockemeyer *et al.*, *Nature biotechnology* 27, 851 (Sep, 2009).
14. S. Kosuri *et al.*, *Nature biotechnology* 28, 1295 (Dec, 2010).
15. V. Pattanayak, C. L. Ramirez, J. K. Joung, D. R. Liu, *Nature methods* 8, 765 (Sep, 2011).
16. N. M. King, O. Cohen-Haguenuer, *Molecular therapy : the journal of the American Society of Gene Therapy* 16, 432 (Mar, 2008).
17. Y. G. Kim, J. Cha, S. Chandrasegaran, *Proceedings of the National Academy of Sciences of the United States of America* 93, 1156 (Feb 6, 1996).
18. E. J. Rebar, C. O. Pabo, *Science* 263, 671 (Feb 4, 1994).
19. J. Boch *et al.*, *Science* 326, 1509 (Dec 11, 2009).
20. M. J. Moscou, A. J. Bogdanove, *Science* 326, 1501 (Dec 11, 2009).
21. A. S. Khalil, J. J. Collins, *Nature reviews. Genetics* 11, 367 (May, 2010).
22. P. E. Purnick, R. Weiss, *Nature reviews. Molecular cell biology* 10, 410 (Jun, 2009).

[0039]

23. J. Zou *et al.*, *Cell stem cell* 5, 97 (Jul 2, 2009).
24. N. Holt *et al.*, *Nature biotechnology* 28, 839 (Aug, 2010).
25. F. D. Urnov *et al.*, *Nature* 435, 646 (Jun 2, 2005).
26. A. Lombardo *et al.*, *Nature biotechnology* 25, 1298 (Nov, 2007).
27. H. Li *et al.*, *Nature* 475, 217 (Jul 14, 2011).

[0040]

[0041] 다음 실시예는 본 개시내용을 대표하는 것으로서 제시된다. 이들 실시예는 본 개시내용의 범위를 제한하는 것으로 추론되지 말아야 하는데, 이는 이들 및 기타 등가의 실시양태가 본 개시내용, 도면 및 첨부된 청구범위의 관점에서 명백할 것이기 때문이다.

[0042] 실시예 I

[0043] 유형 II CRISPR-Cas 시스템

[0044] 한 측면에 따르면, 본 개시내용의 실시양태는 진핵 세포에서 뉴클레아제에 의한 활성을 알아보기 위하여 외래 핵산을 확인하기 위해 저분자 RNA를 활용한다. 본 개시내용의 특정의 측면에 따르면, 진핵 세포는 저분자 RNA를 표적 DNA 서열에 결합시킴으로써 활성화되는 하나 이상의 뉴클레아제 및 하나 이상의 저분자 RNA를 코딩하는 그의 게놈 핵산 내에 포함되도록 변경시킨다. 특정의 측면에 따르면, 예시되는 저분자 RNA/효소 시스템, 예컨대 외래 핵산의 분해를 지시하기 위해 저분자 RNA를 이용하는 (CRISPR)/CRISPR-관련 (Cas) 시스템을 박테리아 또는 고세균 내에서 확인할 수 있다. CRISPR ["클러스터링되고 규칙적으로 사이에 공간을 남겨둔 짧은 회문식 반복 서열"] 방어는 침입 바이러스 또는 플라스미드 DNA로부터 새로운 표적화 "스페이서"를 획득하고, 이를 CRISPR 유전자좌 내로 통합시키며, 스페이서-반복 서열 단위로 이루어진 저분자 유도성 CRISPR RNA (crRNA)를 발현 및 프로세싱하고, 상기 스페이서에 상보적인 핵산 (가장 흔하게는 DNA)을 절단하는 것을 포함한다.

[0045] 3가지 부류의 CRISPR 시스템이 일반적으로 공지되어 있고, 이는 유형 I, 유형 II 또는 유형 III으로서 지칭된다. 한 측면에 따르면, 본 개시내용에 따라서 dsDNA를 절단하는 데 특히 유용한 효소는 유형 II에 공통

인 단일 이펙터 효소, Cas9이다 [참고문헌 (1) 참조]. 박테리아 내에서, 유형 II 이펙터 시스템은 스페이서-함유 CRISPR 유전자좌로부터 전사된 고분자 프리-crRNA, 다기능성 Cas9 단백질, 및 gRNA 프로세싱에 중요한 tracrRNA로 이루어진다. 이러한 tracrRNA는 상기 반복 서열 영역과 혼성화되어, 프리-crRNA의 스페이서를 분리시키고, 내인성 RNase III에 의한 dsRNA 절단을 개시한 다음, Cas9에 의해 각 스페이서 내에 제2의 절단 현상을 개시하며, tracrRNA 및 Cas9와 연합된 채로 있는 성숙한 crRNA를 생성시킨다. 한 측면에 따르면, 본 개시내용의 진핵 세포는 일반적으로 crRNA 프로세싱과 RNase III의 사용을 피하도록 조작한다 [참고문헌 (2) 참조].

[0046] 한 측면에 따르면, 본 개시내용의 효소, 예컨대 Cas9는 DNA 이중체를 풀어주고, 절단하기 위해 crRNA와 부합되는 서열에 관하여 조사한다. 표적 DNA 내의 "프로토스페이서" 서열과 crRNA 내의 나머지 스페이서 서열 간의 상보성의 검출시 표적 인식이 일어난다. 중요하게도, Cas9는 정확한 프로토스페이서-인접한 모티프 (PAM)가 3' 말단에 또한 존재하는 경우에만 DNA를 절단한다. 특정의 측면에 따르면, 상이한 프로토스페이서-인접한 모티프를 활용할 수 있다. 예를 들어, 에스. 피오게네스 시스템은 NGG 서열을 요구하는데, 여기서 N은 모든 뉴클레오티드일 수 있다. 에스. 써모필루스(*S. thermophilus*) 유형 II 시스템은 NGGNG [참고문헌 (3) 참조] 및 NNAGAAW [참고문헌 (4) 참조]를 각각 요구하는 반면, 상이한 에스. 뮤탄스(*S. mutans*) 시스템은 NGG 또는 NAAR을 용인한다 [참고문헌 (5) 참조]. 생물정보학 분석 결과, 부가의 유용한 PAM을 확인하고 CRISPR-표적 가능한 서열 세트를 확장시키기 위해 제공될 수 있는 각종 박테리아에서 CRISPR 유전자좌의 광범위한 데이터베이스가 생성되었다 [참고문헌 (6, 7) 참조]. 에스. 써모필루스에서는, Cas9가 프로토스페이서의 3' 말단 이전에 평활 말단 이중 가닥 절단물 3 bp를 생성시키는데 [참고문헌 (8) 참조], 이는 Cas9 단백질 내의 2개의 촉매적 도메인, 즉 DNA의 상보적 가닥을 절단시키는 HNH 도메인과 비-상보적 가닥을 절단시키는 RuvC-유사 도메인에 의해 매개된 프로세스이다 (도 1a 및 도 3 참조). 에스. 피오게네스 시스템이 동일한 수준의 정밀도로 성상 확인되지는 않았지만, DSB 형태가 또한, 프로토스페이서의 3' 말단을 향하여 생성된다. 상기 2개의 뉴클레아제 도메인 중 하나가 불활성화되는 경우, Cas9는 시험관내에서 [참고문헌 (2) 참조] 및 인간 세포에서 (도 5 참조) 니카제로서 기능할 것이다.

[0047] 한 측면에 따르면, gRNA를 대상으로 한 Cas9 절단의 특이성이 진핵 세포에서 게놈 조작을 위한 기전으로서 사용된다. 한 측면에 따르면, 상기 효소가 gRNA/DNA 혼성체를 인식하고 절단에 영향을 미치기 위해서 gRNA의 혼성화가 100%일 필요는 없다. 일부 빗나간 활성이 발생할 수 있었다. 예를 들어, 에스. 피오게네스 시스템은 시험관내에서 20 bp 성숙한 스페이서 서열 중에서 첫 번째 6개 염기 내에서의 부정합을 용인한다. 한 측면에 따르면, NGG와 부합되는 잠재적 빗나간 부위 (마지막 14 bp)가 gRNA에 대한 인간 참조 게놈 내에 존재하는 경우에는, 보다 큰 엄격성이 생체내에서 유리할 수 있다. 일반적으로 부정합물의 효과 및 효소 활성은 참고문헌 (9), (2), (10), 및 (4)에 기재되어 있다.

[0048] 특정의 측면에 따르면, 특이성을 개선시킬 수 있다. 간섭이 gRNA-DNA 혼성체의 융점에 대해 감수성인 경우, AT-풍부 표적 서열이 더 적은 수의 빗나간 부위를 가질 수 있다. 게놈 내의 다른 곳에 적어도 14 bp 부합 서열을 수반하는 가성-부위를 피하기 위하여 표적 부위를 조심스럽게 선택하는 것이 특이성을 개선시킬 수 있다. 보다 긴 PAM 서열을 요구하는 Cas9 변이체를 사용하면, 빗나간 부위의 빈도를 감소시킬 수 있다. 지향 진화는 빗나간 활성을 완전히 배제시키기 위해 충분한 수준으로 Cas9 특이성을 개선시킬 수 있는데, 이를 위해서는 이상적으로 최소 PAM과의 완벽한 20 bp gRNA 부합이 요구된다. 따라서, Cas9 단백질에 대한 변형이 본 개시내용의 대표적인 실시양태이다. 따라서, 짧은 기간 내에 수차례의 진화를 허용하는 신규 방법이 고려된다 [참고문헌 (11) 참조]. 본 개시내용에 유용한 CRISPR 시스템이 참고문헌 (12, 13)에 기재되어 있다.

[0049] **실시예 II**

[0050] **플라스미드 구축**

[0051] Cas9 유전자 서열을 인간 코돈 최적화시키고, IDT로부터 주문한 9개의 500 bp g블록의 계층적 융합 PCR 조립에 의해 조립하였다. 인간 세포에 대한 조작된 유형 II CRISPR 시스템에 대한 도 3a는 cas9 유전자 삽입물의 발현 포맷 및 완전 서열을 나타낸다. RuvC-유사 모티프 및 HNH 모티프, 및 C-말단 SV40 NLS는 각각, 청색, 갈색 및 오렌지색으로써 강조된다. Cas9\_D10A를 유사하게 구축하였다. 이로써 생성되는 완전한 길이의 생성물을 pcDNA3.3-TOPO 벡터 [인비트로젠(Invitrogen)] 내로 클로닝하였다. 표적 gRNA 발현 구축물은 IDT로부터 개별적 455 bp g블록으로서 직접 주문하였고, 이를 pCR-BluntII-TOPO 벡터 (인비트로젠) 내로 클로닝하거나 PCR 증폭시켰다. 도 3b는 가이드 RNA 및 예측 RNA 전사체 이차 구조에 대한 U6 프로모터 기반 발현 방식을 나타낸다. U6 프로모터를 사용하면, RNA 전사체 내의 첫 번째 위치가 'G'로 제한되므로, 형태 GN<sub>20</sub>GG의 모든 게놈 부위를 이러한 접근법으로 표적화할 수 있다. 도 3c는 사용된 7개의 gRNA를 나타낸다.

[0052] 애드젠(Addgene)으로부터의 EGIP 렌티벡터 (플라스미드 #26777) 내로 조립된 DNMT3a 및 DNMT3b 게놈 유전자좌로부터의 58 bp 단편 (도 8 참조), 또는 68 bp AAVS1 단편 (또는 그의 돌연변이체; 도 6 참조) 및 정지 코돈을 보유하고 있는 GFP 서열의 융합 PCR 조립에 의해, 절단된 GFP를 포함하는 HR 리포터 검정용 벡터를 구축하였다. 이어서, 이들 렌티벡터를 이용하여 GFP 리포터 안정한 세포주를 확립시켰다. 이러한 연구에 사용된 TALEN은 참고문헌 (14)에 기재된 프로토콜을 이용하여 구축하였다. 이 연구에서 개발된 모든 DNA 시약은 애드젠에서 입수 가능하다.

[0053] 실시예 III

[0054] 세포 배양

[0055] PGP1 iPS 세포를 mTeSR1 [스템셀 테크놀로지(Stemcell Technologies)] 중의 매트릭젤 [Matrigel (BD 바이오사이언스(Biosciences))]-코팅된 판 상에 유지시켰다. 배양물을 5 내지 7일마다 TrypLE 익스프레스(Express) (인비트로젠)를 이용하여 계대하였다. K562 세포를 성장시키고, 15% FBS를 함유하는 RPMI (인비트로젠)에서 유지시켰다. HEK 293T 세포를 10% 태아 소 혈청 (FBS, 인비트로젠), 페니실린/스트렙토마이신 (pen/strep, 인비트로젠), 및 비-필수 아미노산 (NEAA, 인비트로젠)으로 보충시킨 둘베코(Dulbecco) 변형 이글 배지 (DMEM, 인비트로젠) 고 글루코스에서 배양하였다. 모든 세포를 가습 인큐베이터에서 37°C 및 5% CO<sub>2</sub> 하에 유지시켰다.

[0056] 실시예 IV

[0057] PGP1 iPS, K562 및 293T의 유전자 표적화

[0058] PGP1 iPS 세포를 뉴클레오펙션하기 2시간 전에 Rho 키나제 (ROCK) 억제제 [칼바이오켄(Calbiochem)]에서 배양하였다. TrypLE 익스프레스 (인비트로젠)를 이용하여 세포를 채취하고 2 x 10<sup>6</sup> 개 세포를 1 µg Cas9 플라스미드, 1 µg gRNA 및/또는 1 µg DNA 공여자 플라스미드와 함께 P3 시약 [론자(Lonza)]에 재현탁시키고, 제조업자의 지시 (론자)에 따라서 뉴클레오펙션하였다. 연속해서, 세포를 처음 24시간 동안 ROCK 억제제가 보충된 mTeSR1 배지 중의 mTeSR1-코팅된 판 상에 도말하였다. K562의 경우에는, 2 x 10<sup>6</sup> 개 세포를 1 µg Cas9 플라스미드, 1 µg gRNA 및/또는 1 µg DNA 공여자 플라스미드와 함께 SF 시약 (론자)에 재현탁시키고, 제조업자의 지시 (론자)에 따라서 뉴클레오펙션하였다. 293T의 경우에는, 0.1 x 10<sup>6</sup> 개 세포를 제조업자의 프로토콜에 따라서 리포펙타민 (Lipofectamine) 2000을 이용하여 1 µg Cas9 플라스미드, 1 µg gRNA 및/또는 1 µg DNA 공여자 플라스미드로 형질감염시켰다. 내인성 AAVS1 표적화에 사용된 DNA 공여자는 dsDNA 공여자 (도 2c) 또는 90량체 올리고뉴클레오티드였다. 전자는 양옆에 위치하고 있는 짧은 상동성 암(arm)과 SA-2A-푸로마이신-CaGGS-eGFP 카세트를 갖고 있어, 성공적으로 표적화된 세포에 대해 강화시켰다.

[0059] 표적화 효율을 다음과 같이 평가하였다. 뉴클레오펙션한지 3일 후에 세포를 채취하고, 프랩GEM (ZyGEM)을 이용하여 약 1 x 10<sup>6</sup> 개 세포의 게놈 DNA를 추출하였다. PCR을 수행하여 상기 세포로부터 유래된 게놈 DNA를 수반한 표적화 영역을 증폭시켰고, 유효범위 >200,000 판독치를 수반한 MiSeq 퍼스널 서열분석기 [일루미나(Illumina)]에 의해 앰플리콘을 심층 서열분석하였다. 이러한 서열분석 데이터를 분석하여 NHEJ 효율을 추정하였다. 분석된 참조 AAVS1 서열은 다음과 같다:

```
CACTTCAGGACAGCATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGCTGGGACCA
CCTTATATTCCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCCCC
TCCACCCACAGTGGGGCCACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTT
AGGCCTCCTCCTTCTAGTCTCCTGATATTGGGTCTAACCCCCACCTCCTGTTAGGCAG
ATTCCTTATCTGGTGACACACCCCCATTTCTGGA
```

[0060]

[0061] 인간 게놈 중의 표적화 영역을 증폭시키기 위한 PCR 프라이머는 다음과 같다:

AAVS1-R CTCGGCATTCTCTGCTGAACCGCTCTTCCGATCTacaggagggtgggggtagac  
AAVS1-F.1

ACACTCTTTCCCTACACGACGCTCTTCCGATCTCGTGATtatattcccagggccgggta  
AAVS1-F.2

[0062] ACACTCTTTCCCTACACGACGCTCTTCCGATCTACATCGtatattcccagggccgggta

AAVS1-F.3  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCTAAtatattcccagggccgggta

AAVS1-F.4  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTTGGTCAtatattcccagggccgggta

AAVS1-F.5  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCACTGTtatattcccagggccgggta

AAVS1-F.6  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTATTGGCtatattcccagggccgggta

AAVS1-F.7  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCTGtatattcccagggccgggta

AAVS1-F.8  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTTCAAGTtatattcccagggccgggta

AAVS1-F.9  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTCTGATCtatattcccagggccgggta

AAVS1-F.10  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTAAGCTAtatattcccagggccgggta

AAVS1-F.11  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTGTAGCCtatattcccagggccgggta

AAVS1-F.12  
ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACAAGtatattcccagggccgggta

[0063] ACACTCTTTCCCTACACGACGCTCTTCCGATCTTACAAGtatattcccagggccgggta

[0064] 도 2c에서 DNA 공여자를 이용하여 HR 현상을 분석하기 위해 사용된 프라이머는 다음과 같다:

HR\_AAVS1-F CTGCCGTCTCTCTCCTGAGT

HR\_푸로-R GTGGGCTTGTACTCGGTCAT

[0065]

[0066] 실시예 V

[0067] 인간 엑손 CRISPR 표적을 전산처리하기 위한 생물정보학 접근법 및 그들의 다중화 합성을 위한 방법론

[0068] 인간 엑손 내의 특이적 위치를 최대 표적으로 하지만, 게놈 내의 다른 위치는 최소한도로 표적으로 하는 gRNA 유전자 서열 세트를 다음과 같이 결정하였다. 한 측면에 따르면, gRNA에 의한 최대 효율적인 표적화는 23 nt 서열에 의해 달성되는데, 이들 중 5'에 가까운 20 nt는 목적하는 위치와 정확하게 상보적인 반면, 3'에 가까운 3개의 염기는 형태 NGG여야만 한다. 부가적으로, 5'에 가까운 nt는 pol-III 전사 출발 부위를 확립시키기 위한 G여야만 한다. 그러나, 참고문헌 (2)에 따라서, 그의 게놈 표적에 대하여 20 bp gRNA의 5'에 가까운 6개의 nt의 잘못된 쌍형성은, 마지막 14 nt가 적당하게 쌍형성을 하는 한은 Cas9-매개된 절단을 폐기하지 않지만, 마지막 12 nt의 쌍형성과 함께 5'에 가까운 8개의 nt의 잘못된 쌍형성이 일어나는 반면, 5'에 가까운 7개의 nt 잘못된 쌍형성과 13개의 3' 쌍형성의 경우는 시험되지 않았다. 빗나간 효과에 관하여 보존적이 되기 위한 한 가지 조건은 5'에 가까운 7개의 잘못된 쌍형성의 경우는 6개의 경우와 마찬가지로, 절단을 허용하므로, 3'에 가까운 13 nt의 쌍형성이 절단을 위해 충분하다는 것이었다. 빗나간 절단 없이도 절단 가능해야 하는, 인간 엑손 내의 CRISPR 표적 부위를 확인하기 위하여, 형태 5'-GBBBB BBBB BBBB BBBB NGG-3' (형태 1)의 모든 23 bp 서열을 조사하였는데, 여기서 B는 엑손 위치에서의 염기를 나타내고, 그에 대하여 형태 5'-NNNNN NNBBB BBBB



BBBBB NGG-3' (형태 2)의 어떠한 서열도 인간 게놈 내의 다른 위치에 존재하지 않았다. 구체적으로 언급하면, (i) UCSC 게놈 브라우저 (15-17)로부터 GRCh37/hg19 인간 게놈의 모든 RefSeq 유전자의 코딩 영역의 위치의 BED 파일을 다운로드 받았다. 이러한 BED 파일 내의 코딩 엑손 위치는 hg19 게놈에 대한 RefSeq mRNA 승인물의 346089개 매핑 세트에 구성되었다. 그러나, 일부 RefSeq mRNA 승인물은 다중 게놈 위치에 매핑되었고 (가능한 유전자 중복), 많은 승인물은 동일한 세트의 엑손 위치의 서브세트에 매핑되었다 (동일한 유전자의 다중 이소형). 명백하게 중복된 유전자 사례를 구별하고, 다중 RefSeq 이소형 승인물에 의한 동일한 게놈 엑손 사례에 대한 다중 참조를 통합하기 위하여, (ii) 다중 게놈 위치를 갖는 705 RefSeq 승인 번호에 대한 고유한 숫자 접미사를 가하였고, (iii) BED툴즈(Tools) (18) (v2.16.2-zip-87e3926)의 mergeBed 기능을 이용하여, 중복되는 엑손 위치를 합병된 엑손 영역 내로 통합시켰다. 이들 단계는 346089개 RefSeq 엑손 위치의 초기 세트를 192783개 별개의 게놈 영역으로 감소시켰다. 합병된 모든 엑손 영역에 대한 hg19 서열은 UCSC 표 브라우저를 이용하여 다운로드시켰는데, 각 말단에 20 bp의 패딩을 부가하였다. (iv) 사용자 정의 펄(perl) 코드를 이용하여, 형태 1의 1657793개 사례들을 이러한 엑손 서열 내에서 확인하였다. (v) 이어서, 이들 서열을 대상으로 하여, 형태 2의 빛나간 발생이 존재하는 지에 관하여 필터링하였다: 합병된 각 엑손 형태 1 표적에 대해서는, 3'에 가까운 13 bp 특이적 (B) "코어" 서열을 추출하였고, 각 코어의 경우에는, 4개의 16 bp 서열 5'-BBB BBBBB BBBBB NGG-3' (N = A, C, G, 및 T)을 발생하였고, 파라미터 -l 16 -v 0 -k 2를 이용하여 보타이(Bowtie) 버전 0.12.8 (19)을 사용하여 이들 6631172개 서열에 대한 정확한 정합물을 알아보기 위하여 전체 hg19 게놈을 조사하였다. 그에 대해 2개 이상의 정합물이 존재하는 모든 엑손 표적 부위는 거절되었다. 어느 특이적 13 bp 코어 서열 다음의 서열 NGG는 단지 15 bp의 특이성을 부여하기 때문에, 무작위 약 3 Gb 서열 (양 가닥) 내의 연장된 코어 서열에 대해 평균 약 5.6개의 정합물이 존재해야 한다는 것에 주목해야 한다. 따라서, 초기에 확인된 1657793개 표적의 대부분이 거절당했지만, 189864개 서열은 이러한 필터를 통과하였다. 이들은 인간 게놈 내의 CRISPR-표적 가능한 엑손 위치 세트를 포함한다. 189864개 서열은 표적화된 엑손 영역당 약 2.4개 부위의 다중도로 78028개 합병된 엑손 영역 내의 위치를 표적으로 한다 (총 192783개의 합병된 인간 엑손 영역의 약 40.5%). 유전자 수준에서의 표적화를 평가하기 위하여, RefSeq mRNA 매핑을 클러스터링하여, 합병된 엑손 영역과 중복되는 어느 2개의 RefSeq 승인물 [(ii)에서 구별된 유전자 중복물 포함]을 단일 유전자 클러스터로서 계수하고, 189864개의 엑손 특이적 CRISPR 부위는 표적화된 유전자 클러스터당 약 11.1의 다중도로 18872개 유전자 클러스터 중 17104개를 표적으로 한다 (모든 유전자 클러스터의 약 90.6%). (이들 유전자 클러스터가, 단일 실체 내로 전사된 단일 유전자의 다중 이소형을 나타내는 RefSeq mRNA 승인물을 붕괴시키긴 하지만, 이들 클러스터는 또한, 중복되는 별개의 유전자 뿐만 아니라 안티센스 전사체를 수반한 유전자도 붕괴시킬 것이란 사실에 주목해야 한다). 본래의 RefSeq 승인물 수준에서는, 189864개 서열이, 표적화되고 매핑된 RefSeq 승인물당 약 6.2개 부위의 다중도로 RefSeq 승인물 (구별되는 유전자 중복물 포함)과 매핑된 총 43726개 중에서 30563개 내의 엑손 영역을 표적으로 하였다 (약 69.9%).

[0069] 한 측면에 따르면, 상기 데이터베이스는 gRNA와 게놈 표적 둘 다의 염기 조성 및 이차 구조 (20, 21), 및 인간 세포주에서의 이들 표적의 후생적 상태 [그에 대한 상기 정보는 입수가능하다 (22)]와 같은 요인들과 성능을 상관지음으로써 제련시킬 수 있다.

## [0071] \*실시예 VI

### [0072] 다중화 합성

[0073] DNA 어레이 상의 다중화 합성에 대해 화합성인 200 bp 포맷 내로 표적 서열을 혼입하였다 (23, 24). 한 측면에 따르면, 상기 방법으로, DNA 어레이 기반 올리고뉴클레오타이드 풀로부터 특이적 gRNA 서열 또는 gRNA 서열 풀을 표적화 검색할 수 있게 되고, 그를 통상의 발현 벡터 내로 신속하게 클로닝할 수 있게 된다 (도 13a). 구체적으로 언급하면, 커스텀어레이 인크.(CustomArray Inc.)로부터의 12 k 올리고뉴클레오타이드 풀을 합성하였다. 더욱이, 이러한 라이브러리로부터의 선택 gRNA (도 13b)를 성공적으로 검색하였다. 본 발명자들은 합성된 DNA 1000 bp당 약 4개 돌연변이의 오류 비율을 관찰하였다.

## [0074] 실시예 VII

### [0075] RNA-가이드된 게놈 편집은 성공적인 표적화를 위해 Cas9와 가이드 RNA 둘 다를 필요로 한다

[0076] 도 1b에 기재된 GFP 리포터 검정을 이용하여, 복구 DNA 공여자, Cas9 단백질 및 gRNA의 가능한 모든 조합을 대상으로 하여, 성공적인 HR을 수행할 수 있는 그들의 능력에 관하여 시험하였다 (293T에서 실시함). 도 4에 나



타넨 바와 같이, 상기 3가지 성분 모두가 존재하는 경우에만 GFP+ 세포가 관찰되었는데, 이는 이들 CRISPR 성분이 RNA-가이드된 게놈 편집에 필수적이란 것을 검증해 준다. 데이터는 평균  $\pm$  SEM (N=3)이다.

[0077] 실시예 VIII

[0078] gRNA 및 Cas9 매개된 게놈 편집의 분석

[0079] CRISPR 매개된 게놈 편집 프로세스를 (A) 앞서 기재된 바와 같은 GFP 리포터 검정 (그의 결과는 도 5a에 도시되어 있다), 및 (B) 표적화된 유전자좌의 심층 서열분석 (293T에서 실시됨) (그의 결과는 도 5b에 도시되어 있다) 중 어느 하나를 이용하여 조사하였다. 비교로서, 앞서의 보고서에 시험관내 검정에서 니카제로서 기능하는 것으로 제시된, Cas9에 대한 D10A 돌연변이체를 시험하였다. 도 5에 도시된 바와 같이, Cas9와 Cas9D10A 둘 다는 거의 유사한 비율로 성공적인 HR을 수행할 수 있었다. 그러나, 심층 서열분석 결과, Cas9는 표적화된 유전자좌에서 강건한 NHEJ를 나타내는 반면, D10A 돌연변이체는 상당히 감소된 NHEJ 비율을 나타내는 것으로 확인되었다 (단지 DNA에만 홈을 낼 수 있는 그의 추정상 능력으로부터 예상되는 바와 같다). 또한, Cas9 단백질의 공지된 생화학과 일치하여, NHEJ 데이터는 대부분의 염기-쌍 결실 또는 삽입이 표적 서열의 3' 말단 근처에서 발생하였다는 것을 확증시켜 주는데; 그 피크는 PAM 부위의 하류 약 3 내지 4개 염기이고, 중간 결실 빈도는 약 9 내지 10 bp이다. 데이터는 평균  $\pm$  SEM (N=3)이다.

[0080] 실시예 IX

[0081] RNA-가이드된 게놈 편집은 표적 서열 특이적이다

[0082] 도 1b에 기재된 GFP 리포터 검정과 유사하게, 각각 별개의 GFP 리포터 구축물을 보유하고 있는 3개의 안정한 293T 세포주를 발생시켰다. 이들은 AAVS1 단편 삽입물의 서열 (도 6에 표시된 바와 같음)로써 구별된다. 1개 세포주에는 야생형 단편이 정착된 반면, 2개의 다른 세포주는 6 bp에서 돌연변이되었다 (적색으로 강조됨). 이어서, 이들 세포주 각각은 다음 4가지 시약 중 하나에 의해 표적화되었다: 그의 표적화 서열이 양옆에 위치하고 있는 GFP 단편 내에 있으므로, 세포주를 따라 존재하기 때문에 모든 세포 유형을 표적으로 할 수 있는 GFP-ZFN 쌍; 다른 2개의 세포주에서의 돌연변이로 인해, 좌측 TALEN이 그들의 부위와 결합될 수 없게 해야 하기 때문에 wt-AAVS1 단편만을 잠재적으로 표적으로 할 수 있었던 AAVS1 TALEN; 그의 표적 부위가 또한, 2개의 돌연변이체 주에서 붕괴되기 때문에, wt-AAVS1 단편만을 잠재적으로 표적으로 할 수 있는 T1 gRNA; 및 최종적으로, T1 gRNA와는 달리, 그의 표적 부위가 상기 3가지 세포주 중에서 변경되지 않기 때문에, 모든 3가지 세포주를 표적으로 할 수 있어야 하는 T2 gRNA. ZFN은 3가지 세포 유형 모두를 변형시켰고, AAVS1 TALEN 및 T1 gRNA는 단지 wt-AAVS1 세포 유형을 표적으로 하였으며, T2 gRNA는 3가지 세포 유형 모두를 성공적으로 표적으로 한다. 이들 결과를 취합해 보면, 가이드 RNA 매개된 편집이 표적 서열 특이적인 것으로 확인된다. 데이터는 평균  $\pm$  SEM (N=3)이다.

[0083] 실시예 X

[0084] GFP 서열을 표적으로 한 가이드 RNA는 강건한 게놈 편집을 가능하게 한다

[0085] AAVS1 삽입물을 표적으로 하는 2개의 gRNA 이외에도, (293T에서) 도 1b에 기재된 리포터의 양옆에 위치하고 있는 GFP 서열을 표적으로 하는 2개의 부가 gRNA를 시험하였다. 도 7에 도시된 바와 같이, 이들 gRNA는 또한, 상기 조작된 유전자좌에서 강건한 HR을 수행할 수 있었다. 데이터는 평균  $\pm$  SEM (N=3)이다.

[0086] 실시예 XI

[0087] RNA-가이드된 게놈 편집은 표적 서열 특이적이고, ZFN 또는 TALEN과 유사한 표적화 효율을 입증해 준다

[0088] 도 1b에 기재된 GFP 리포터 검정과 유사하게, 각각 별개의 GFP 리포터 구축물을 보유하고 있는 2개의 안정한 293T 세포주를 발생시켰다. 이들은 단편 삽입물의 서열 (도 8에 표시된 바와 같음)로써 구별된다. 1개 세포주에는 DNMT3a 유전자로부터의 58 bp 단편이 정착된 반면, 다른 세포주는 DNMT3b 유전자로부터의 상동 58 bp 단편을 보유하고 있다. 그 서열 차이가 적색으로 강조된다. 이어서, 이들 세포주 각각은 다음 6가지 시약 중 하나에 의해 표적화되었다: 그의 표적화 서열이 양옆에 위치하고 있는 GFP 단편 내에 있으므로, 세포주를 따라 존재하기 때문에 모든 세포 유형을 표적으로 할 수 있는 GFP-ZFN 쌍; DNMT3a 또는 DNMT3b 단편을 잠재적으로 표적으로 하는 한 쌍의 TALEN; DNMT3a 단편만을 잠재적으로 표적으로 할 수 있는 한 쌍의 gRNA; 및 최종적으로, DNMT3b 단편만을 잠재적으로 표적으로 해야 하는 gRNA. 도 8에 표시된 바와 같이, ZFN은 3가지 세포 유형 모두를 변형시켰고, TALEN 및 gRNA는 이들 각각의 표적만을 표적화하였다. 더욱이, 표적화의 효율은 상기 6가지 표적화 시약 전반에 걸쳐 거의 동등한 수준이었다. 이들 결과를 취합해 보면, RNA-가이드된 편집이 표적 서열 특이적인

것으로 확증되고, ZFN 또는 TALEN과 유사한 표적화 효율을 입증해 준다. 데이터는 평균  $\pm$  SEM (N=3)이다.

[0089] 실시예 XII

[0090] 인간 iPS 세포에서 RNA-가이드된 NHEJ

[0091] 인간 iPS 세포 (PGP1)를 도 9의 좌측 패널에 표시된 구축물로 뉴클레오펙션하였다. 뉴클레오펙션한지 4일 후에, 심층 서열분석함으로써 이중 가닥 절단물 (DSB)에서 게놈 결실 및 삽입 비율을 평가함으로써 NHEJ 비율을 측정하였다. **패널 1:** 표적화 영역에서 검출된 결실 비율. 적색 점선: T1 RNA 표적화 부위의 경계; 녹색 점선: T2 RNA 표적화 부위의 경계. 각 뉴클레오타이드 위치에서의 결실 발생률은 흑색 라인으로 플롯팅하였고, 결실을 수반하는 판독치의 %로서의 결실 비율을 계산하였다. **패널 2:** 표적화 영역에서 검출된 삽입 비율. 적색 점선: T1 RNA 표적화 부위의 경계; 녹색 점선: T2 RNA 표적화 부위의 경계. 첫 번째 삽입 연결부가 검출된 게놈 위치에서의 삽입 발생률은 흑색 라인으로 플롯팅하였고, 삽입을 수반하는 판독치의 %로서의 삽입 비율을 계산하였다. **패널 3:** 결실 크기 분포도. NHEJ 전체 집단 중에서 상이한 크기 결실의 빈도를 플롯팅하였다. **패널 4:** 삽입 크기 분포도. NHEJ 전체 집단 중에서 상이한 크기 삽입의 빈도를 플롯팅하였다. 양 gRNA에 의해 표적화되는 iPS는 효율적이고 (2 내지 4%), 서열 특이적이며 (NHEJ 결실 분포도의 위치 상의 이동으로써 제시된 바와 같음), 도 4의 결과를 재차 확인하였으며, NGS-기반 분석은 또한, Cas9 단백질과 상기 gRNA 둘 다 표적 유전자좌에서 NHEJ 현상에 필수적이라는 것을 나타낸다.

[0092] 실시예 XIII

[0093] K562 세포에서 RNA-가이드된 NHEJ

[0094] K562 세포를 도 10의 좌측 패널에 표시된 구축물로 뉴클레오펙션하였다. 뉴클레오펙션한지 4일 후에, 심층 서열분석함으로써 DSB에서 게놈 결실 및 삽입 비율을 평가함으로써 NHEJ 비율을 측정하였다. **패널 1:** 표적화 영역에서 검출된 결실 비율. 적색 점선: T1 RNA 표적화 부위의 경계; 녹색 점선: T2 RNA 표적화 부위의 경계. 각 뉴클레오타이드 위치에서의 결실 발생률은 흑색 라인으로 플롯팅하였고, 결실을 수반하는 판독치의 %로서의 결실 비율을 계산하였다. **패널 2:** 표적화 영역에서 검출된 삽입 비율. 적색 점선: T1 RNA 표적화 부위의 경계; 녹색 점선: T2 RNA 표적화 부위의 경계. 첫 번째 삽입 연결부가 검출된 게놈 위치에서의 삽입 발생률은 흑색 라인으로 플롯팅하였고, 삽입을 수반하는 판독치의 %로서의 삽입 비율을 계산하였다. **패널 3:** 결실 크기 분포도. NHEJ 전체 집단 중에서 상이한 크기 결실의 빈도를 플롯팅하였다. **패널 4:** 삽입 크기 분포도. NHEJ 전체 집단 중에서 상이한 크기 삽입의 빈도를 플롯팅하였다. 양 gRNA에 의해 표적화되는 K562는 효율적이고 (13 내지 38%), 서열 특이적이다 (NHEJ 결실 분포도의 위치 상의 이동으로써 제시된 바와 같음). 중요하게도, 결실 크기의 관찰된 빈도의 히스토그램 내의 피크로써 입증된 바와 같이, T1 가이드 RNA와 T2 가이드 RNA를 동시에 도입하면, 개재 19 bp 단편의 고 효율 결실이 발생되었는데, 이는 게놈 유전자좌의 다중화 편집이 또한, 이러한 접근법을 이용하여 실행가능하다는 것을 입증해 준다.

[0095] 실시예 XIV

[0096] 293T 세포에서 RNA-가이드된 NHEJ

[0097] 293T 세포를 도 11의 좌측 패널에 표시된 구축물로 뉴클레오펙션하였다. 뉴클레오펙션한지 4일 후에, 심층 서열분석함으로써 DSB에서 게놈 결실 및 삽입 비율을 평가함으로써 NHEJ 비율을 측정하였다. **패널 1:** 표적화 영역에서 검출된 결실 비율. 적색 점선: T1 RNA 표적화 부위의 경계; 녹색 점선: T2 RNA 표적화 부위의 경계. 각 뉴클레오타이드 위치에서의 결실 발생률은 흑색 라인으로 플롯팅하였고, 결실을 수반하는 판독치의 %로서의 결실 비율을 계산하였다. **패널 2:** 표적화 영역에서 검출된 삽입 비율. 적색 점선: T1 RNA 표적화 부위의 경계; 녹색 점선: T2 RNA 표적화 부위의 경계. 첫 번째 삽입 연결부가 검출된 게놈 위치에서의 삽입 발생률은 흑색 라인으로 플롯팅하였고, 삽입을 수반하는 판독치의 %로서의 삽입 비율을 계산하였다. **패널 3:** 결실 크기 분포도. NHEJ 전체 집단 중에서 상이한 크기 결실의 빈도를 플롯팅하였다. **패널 4:** 삽입 크기 분포도. NHEJ 전체 집단 중에서 상이한 크기 삽입의 빈도를 플롯팅하였다. 양 gRNA에 의해 표적화되는 293T는 효율적이고 (10 내지 24%), 서열 특이적이다 (NHEJ 결실 분포도의 위치 상의 이동으로써 제시된 바와 같음).

[0098] 실시예 XV

[0099] dsDNA 공여자 또는 저분자 올리고뉴클레오타이드 공여자를 이용하는 내인성 AAVS1 유전자좌에서의 HR

[0100] 도 12a에 도시된 바와 같이, PCR 스크린 (도 2c를 참조로 함) 결과, 무작위로 골라낸 24개의 293T 클론 중에서 21개가 성공적으로 표적화된 것으로 확증되었다. 도 12b에 도시된 바와 같이, 유사한 PCR 스크린 결과, 무작위

로 골라낸 7개의 PGP1-iPS 클론 중에서 3개가 또한, 성공적으로 표적화되었다는 것이 확인되었다. 도 12c에 도시된 바와 같이, 저분자 90량체 올리고가 또한, 내인성 AAVS1 유전자좌 (K562 세포에 대해 본원에 제시됨)에서 강력한 표적화를 수행할 수 있었다.

[0101] 실시예 XVI

[0102] 인간 게놈 중의 가이드 RNA 표적화 유전자의 다중화 합성, 검색 및 U6 발현 벡터 클로닝을 위한 방법론

[0103] 인간 게놈 중의 유전자의 모든 엑손의 약 40.5%를 표적으로 하는, 약 190 k의 생물정보학적으로 전산처리된 고유한 gRNA 부위의 자원을 생성시켰다. 도 13a에 도시된 바와 같이, DNA 어레이 상의 다중화 합성에 대해 화합성인 200 bp 포맷 내로 gRNA 표적 부위를 혼입하였다. 구체적으로 언급하면, 상기 설계로 인해, (i) DNA 어레이 올리고뉴클레오타이드 풀로부터 특이적 gRNA 표적 또는 gRNA 표적 풀을 표적화 검색할 수 있게 되고 (도식으로 나타낸 도면에 표시된 바와 같이 중첩된 PCR을 연속해서 3차례 수행함), (ii) AflIII 부위를 이용한 선형화시 gRNA 삽입물 단편의 김슨(Gibson) 조립 매개된 혼입에 대한 수용자로서 제공되는 통상의 발현 벡터 내로 신속하게 클로닝할 수 있게 된다. 도 13b에 도시된 바와 같이, 상기 방법을 이용하여 커스텀어레이 인크.에 의해 합성된 12 k 올리고뉴클레오타이드 풀로부터 10개의 고유한 gRNA의 표적화 검색을 달성하였다.

[0104] 실시예 XVII

[0105] CRISPR 매개된 RNA-가이드된 전사 활성화

[0106] CRISPR-Cas 시스템은 박테리아에서 적응 면역 방어 시스템을 갖고 있고 침입 핵산을 '절단'시키는 기능을 한다. 한 측면에 따르면, CRISPR-CAS 시스템은 인간 세포에서 기능하고 게놈 DNA를 '절단'시키도록 조작된다. 이는 Cas9 단백질 (뉴클레아제 기능을 갖고 있다)이 가이드 RNA 내의 스페이서에 상보적인 표적 서열을 대상으로 하는 저분자 가이드 RNA에 의해 달성된다. DNA를 '절단'시킬 수 있는 능력으로, 게놈 편집과 관계된 다수의 적용이 가능해지고, 또한 표적화 게놈 조절이 가능해진다. 이 무렵, Mg<sup>2+</sup> (RuvC-유사 및 HNH-유사 도메인의 뉴클레아제 기능에 대해 중요한 것으로 공지됨)와의 커플링을 폐기시키는 것으로 예상되는 돌연변이를 도입함으로써, Cas9 단백질이 뉴클레아제-기능없게 만들도록 이를 돌연변이시켰다: 구체적으로 언급하면, D10A, D839A, H840A 및 N863A 돌연변이의 조합을 도입하였다. 이어서, 이로써 생성된 Cas9 뉴클레아제-기능없는 단백질 (서열분석에 의해 DNA를 절단할 수 없는 그의 능력으로써 확인된 바와 같음) (Cas9R-H-로서 후술됨)을 전사 활성화 도메인 (본 실시예에서 VP64)과 커플링시켜, CRISPR-cas 시스템이 RNA 유도된 전사 인자로서 기능하도록 할 수 있다 (도 14 참조). Cas9R-H-+VP64 융합으로, 제시된 2개의 리포터에서 RNA-가이드된 전사 활성화가 가능해진다. 구체적으로 언급하면, FACS 분석과 면역형광 영상화 둘 다는 상기 단백질이 상응하는 리포터의 gRNA 서열 특이적 표적화를 가능하게 하고, 더욱이 d토마토 형광성 단백질의 발현에 의해 검정된 바와 같은, 상기와 같이 하여 생성된 전사 활성화는 통상적인 TALE-VP64 융합 단백질에 의해 유도된 것과 유사한 수준이었다는 것을 입증해준다.

[0107] 실시예 XVIII

[0108] gRNA 서열 가요성 및 그의 적용

[0109] 서열 삽입물을 설계하기 위한 gRNA 스캐폴드 서열의 가요성은, gRNA의 5' 부분, 중간 부분 및 3' 부분 상에서 특정 범위의 무작위 서열 삽입물을 체계적으로 검정함으로써 결정하였다: 구체적으로 언급하면, 1 bp, 5 bp, 10 bp, 20 bp, 및 40 bp 삽입물을 gRNA의 5' 말단, 중간 및 3' 말단에서 gRNA 서열 내에 만들었다 (이러한 삽입물의 정확한 위치는 도 15에서 '적색'으로 강조된다). 이어서, 이러한 gRNA를 대상으로 하여, GFP 리포터 검정에서 HR을 유도시킬 수 있는 그의 능력에 의해 기능성에 관하여 시험하였다 (본원에 기재된 바와 같음). gRNA가 5' 말단과 3' 말단 상의 서열 삽입물에 대해 가요성이라는 것은 분명하다 (보유된 HR 유도성 활성화에 의해 측정된 바와 같음). 따라서, 본 개시내용의 측면은 gRNA 활성 또는 gRNA 가시화의 발생을 촉발시킬 수 있는 소분자 반응성 RNA 앵타머의 태그화에 관한 것이다. 부가적으로, 본 개시내용의 측면은 혼성화를 통하여 ssDNA 공여자를 gRNA에 테더링함으로써, 오류 유발 비-상동 말단 결합에 비해 상동 재조합 비율을 증진시킬 수 있는 복구 주형의 즉시 신체적 국제화 및 게놈 표적 절단의 커플링을 가능하게 하는 것에 관한 것이다. 한 측면에 따르면, 스캐폴드 서열은 핵산 서열

GUUUUAGAGCUAGAAUAGCAAGUAAAAUAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGUGCUUUU(서열식별번호: 45)를 포함할 수 있다.

[0110] 숫자로서 실시예 섹션에서 확인된 다음 참고문헌은 모든 목적상 그들의 전문이 본원에 참조로 포함된다.

[0111]

참고문헌

1. K. S. Makarova *et al.*, Evolution and classification of the CRISPR-Cas systems. *Nature reviews. Microbiology* 9, 467 (Jun, 2011).
2. M. Jinek *et al.*, A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337, 816 (Aug 17, 2012).
3. P. Horvath, R. Barrangou, CRISPR/Cas, the immune system of bacteria and archaea. *Science* 327, 167 (Jan 8, 2010).
4. H. Deveau *et al.*, Phage response to CRISPR-encoded resistance in *Streptococcus thermophilus*. *Journal of bacteriology* 190, 1390 (Feb, 2008).
5. J. R. van der Ploeg, Analysis of CRISPR in *Streptococcus mutans* suggests frequent occurrence of acquired immunity against infection by M102-like bacteriophages. *Microbiology* 155, 1966 (Jun, 2009).
6. M. Rho, Y. W. Wu, H. Tang, T. G. Doak, Y. Ye, Diverse CRISPRs evolving in human microbiomes. *PLoS genetics* 8, e1002441 (2012).
7. D. T. Pride *et al.*, Analysis of streptococcal CRISPRs from human saliva reveals substantial sequence diversity within and between subjects over time. *Genome research* 21, 126 (Jan, 2011).
8. G. Gasiunas, R. Barrangou, P. Horvath, V. Siksnys, Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proceedings of the National Academy of Sciences of the United States of America* 109, E2579 (Sep 25, 2012).
9. R. Sapranaukas *et al.*, The *Streptococcus thermophilus* CRISPR/Cas system provides immunity in *Escherichia coli*. *Nucleic acids research* 39, 9275 (Nov, 2011).
10. J. E. Garneau *et al.*, The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* 468, 67 (Nov 4, 2010).
11. K. M. Esvelt, J. C. Carlson, D. R. Liu, A system for the continuous directed evolution of biomolecules. *Nature* 472, 499 (Apr 28, 2011).
12. R. Barrangou, P. Horvath, CRISPR: new horizons in phage resistance and strain identification. *Annual review of food science and technology* 3, 143 (2012).
13. B. Wiedenheft, S. H. Sternberg, J. A. Doudna, RNA-guided genetic silencing systems in bacteria and archaea. *Nature* 482, 331 (Feb 16, 2012).
14. N. E. Sanjana *et al.*, A transcription activator-like effector toolbox for genome engineering. *Nature protocols* 7, 171 (Jan, 2012).
15. W. J. Kent *et al.*, The human genome browser at UCSC. *Genome Res* 12, 996 (Jun, 2002).

[0112]

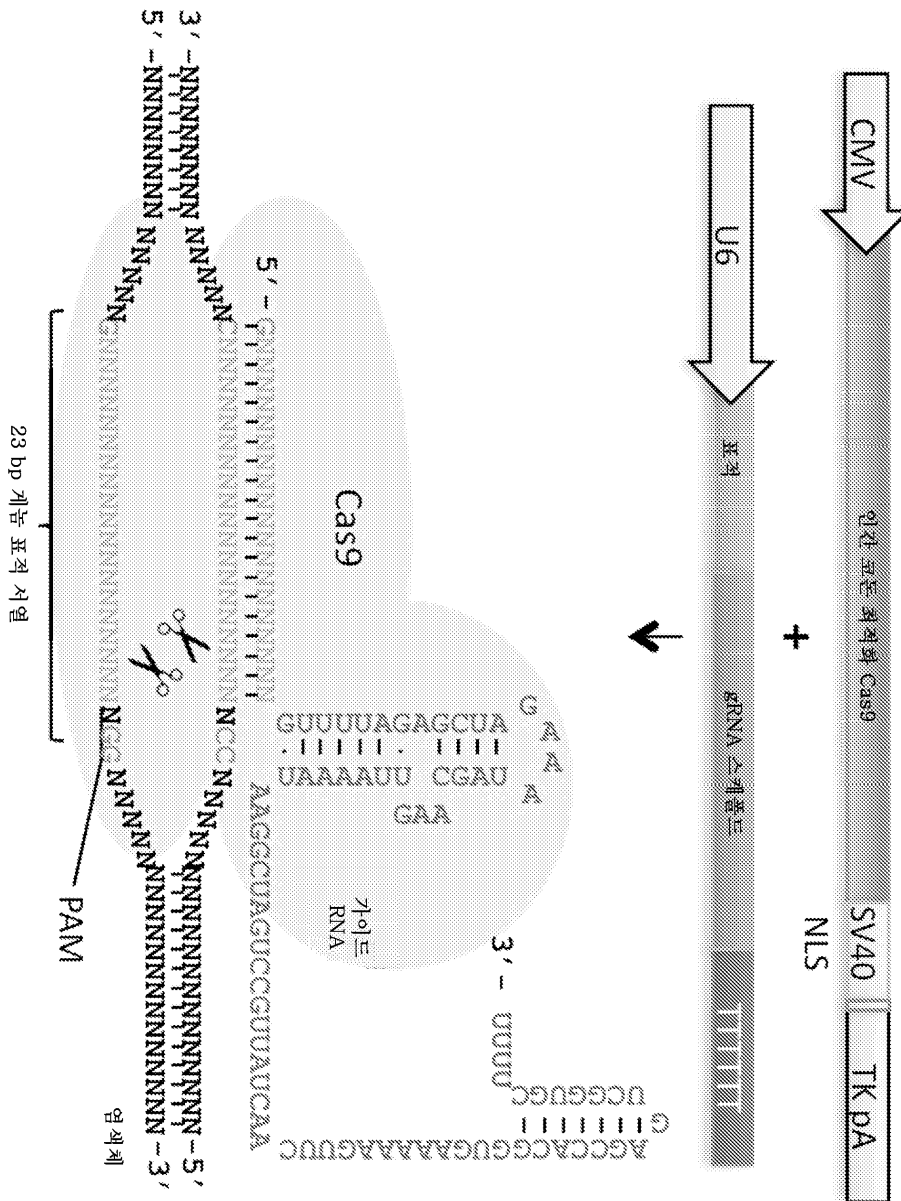
16. T. R. Dreszer *et al.*, The UCSC Genome Browser database: extensions and updates 2011. *Nucleic Acids Res* 40, D918 (Jan, 2012).
17. D. Karolchik *et al.*, The UCSC Table Browser data retrieval tool. *Nucleic Acids Res* 32, D493 (Jan 1, 2004).
18. A. R. Quinlan, I. M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841 (Mar 15, 2010).
19. B. Langmead, C. Trapnell, M. Pop, S. L. Salzberg, Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10, R25 (2009).
20. R. Lorenz *et al.*, ViennaRNA Package 2.0. *Algorithms for molecular biology : AMB* 6, 26 (2011).
21. D. H. Mathews, J. Sabina, M. Zuker, D. H. Turner, Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *Journal of molecular biology* 288, 911 (May 21, 1999).
22. R. E. Thurman *et al.*, The accessible chromatin landscape of the human genome. *Nature* 489, 75 (Sep 6, 2012).
23. S. Kosuri *et al.*, Scalable gene synthesis by selective amplification of DNA pools from high-fidelity microchips. *Nature biotechnology* 28, 1295 (Dec, 2010).
24. Q. Xu, M. R. Schlabach, G. J. Hannon, S. J. Elledge, Design of 240,000 orthogonal 25mer DNA barcode probes. *Proceedings of the National Academy of Sciences of the United States of America* 106, 2289 (Feb 17, 2009).

[0113]



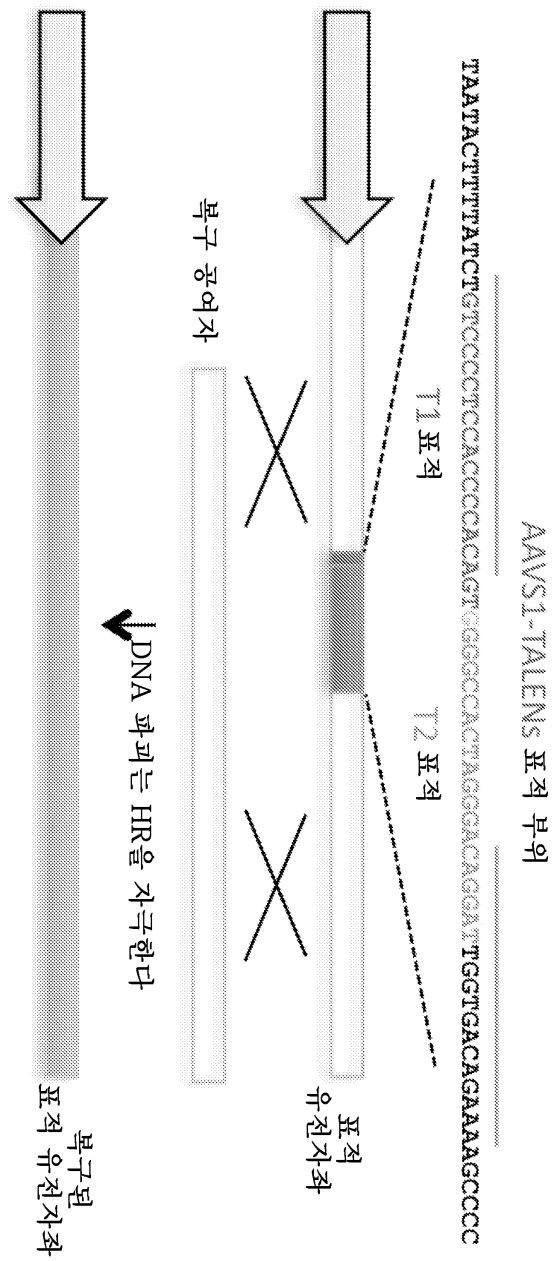
도면

도면1a





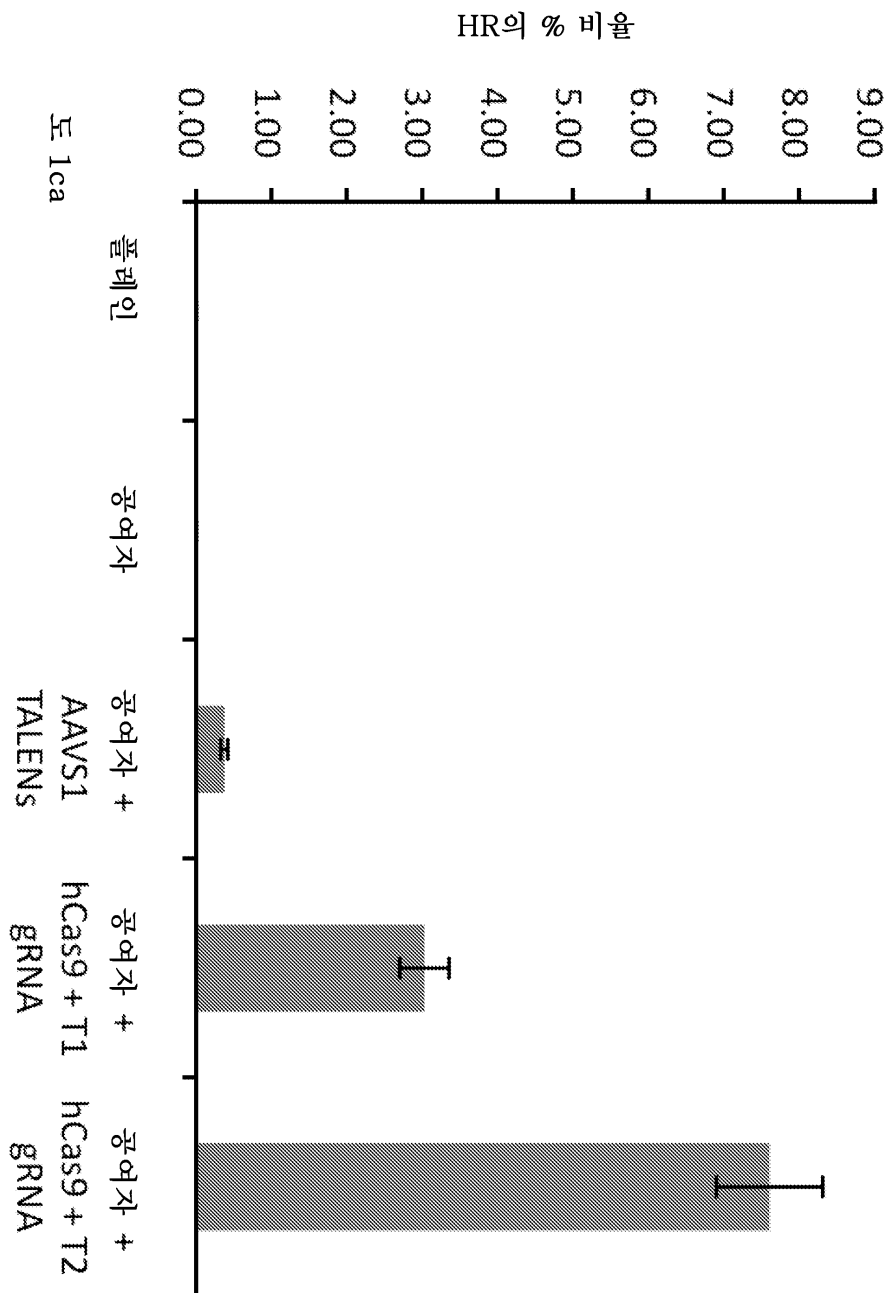
도면1b



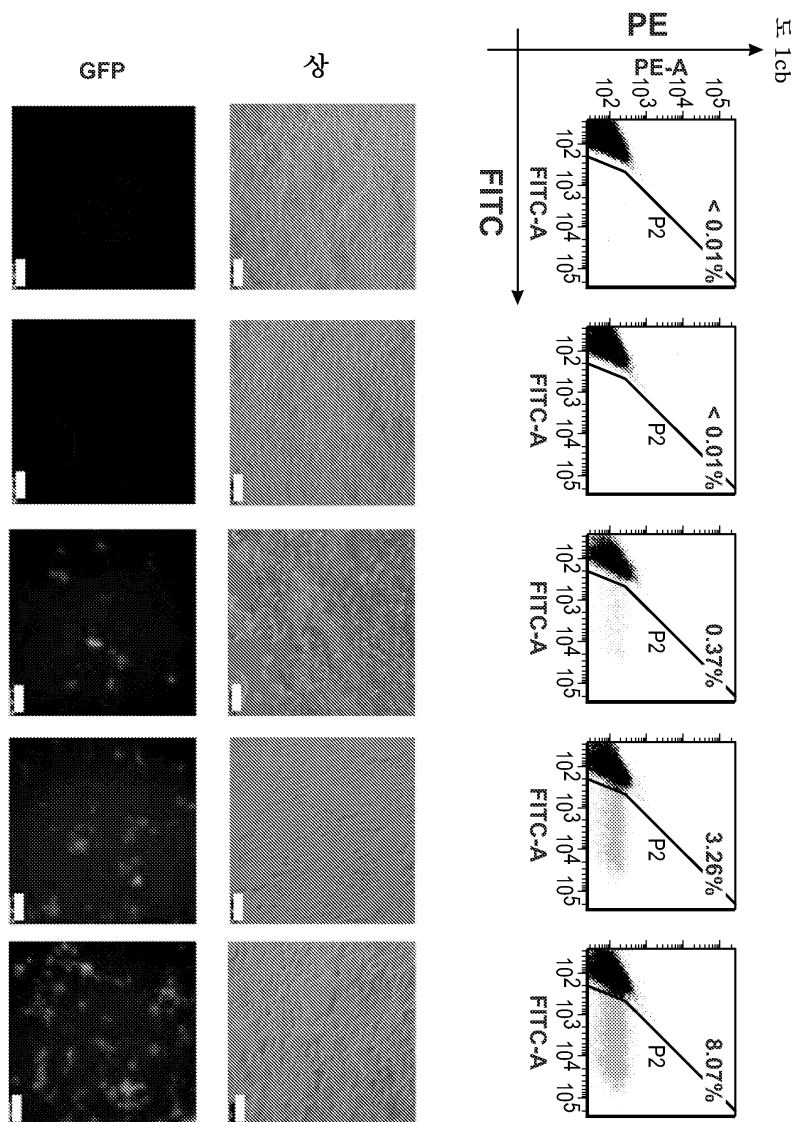
도 1ca
도 1cb

도 1c

도면1ca



도면1cb



도면2a

내인성 '천연' hAAVS1 유전자좌 서열

..TTATCTGTCCCTCCACCCCACTGGGCCACTAGGACAGGATTGCTGA..

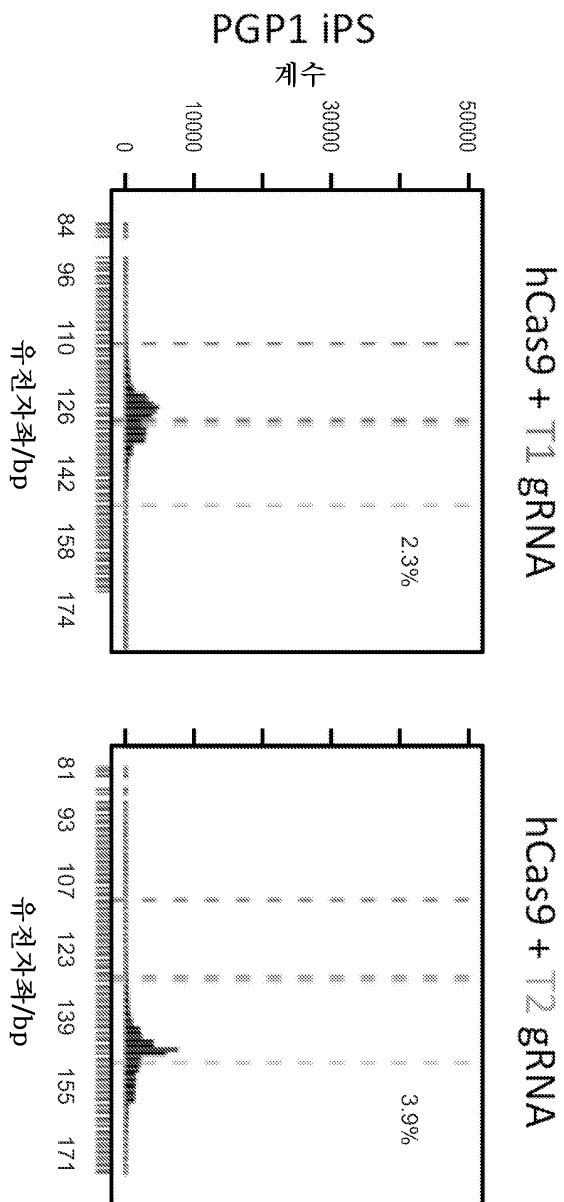
T1 표적                      T2 표적

도 2ba
도 2bb
도 2bc

도 2b

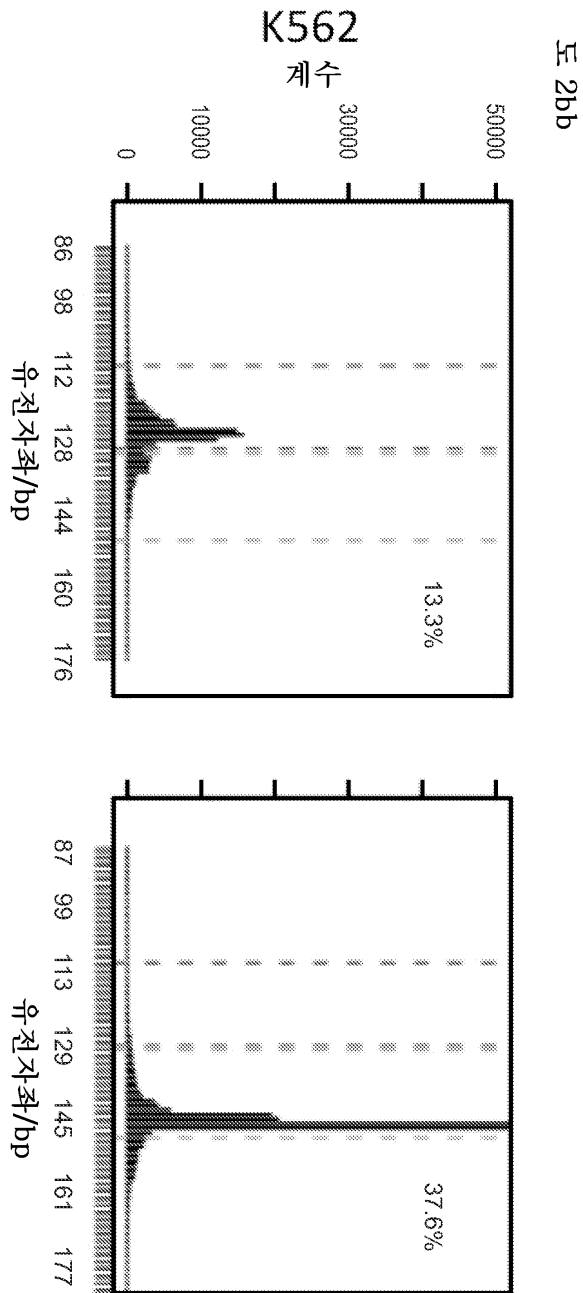
도 2ba

표적화된 AAVS1 유전자좌의 NGS를 이용하여 평가된 NHEJ 비율



도면 2ba

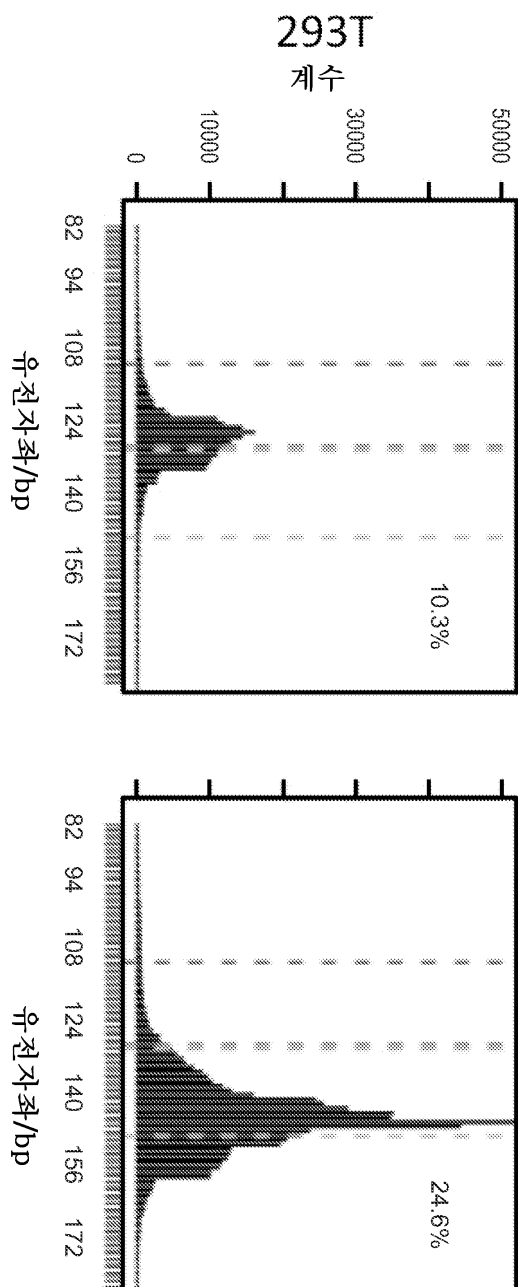
도면2bb



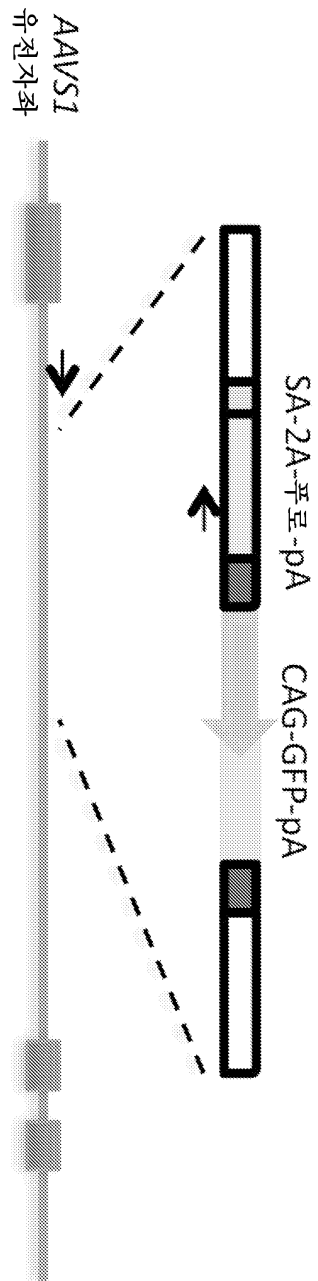


도면2bc

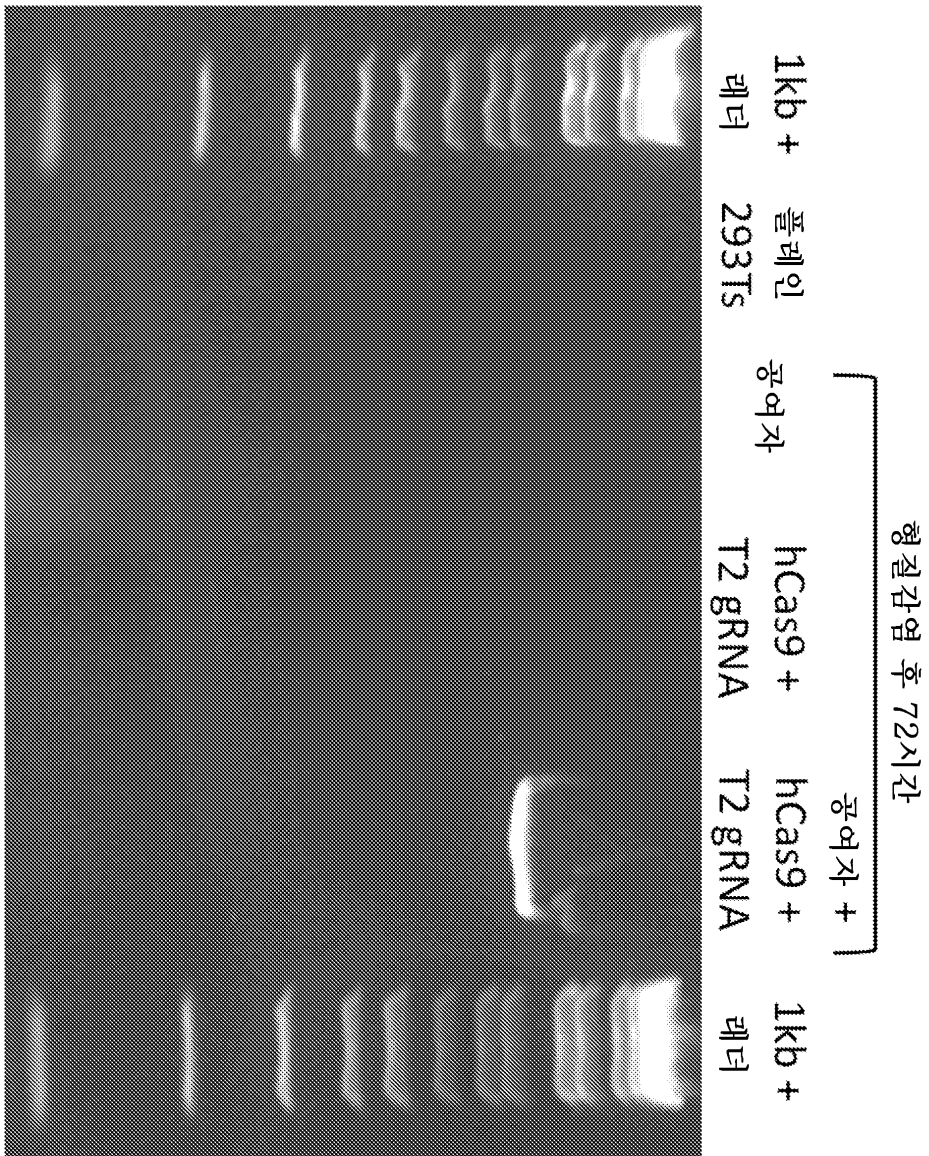
도 2bc



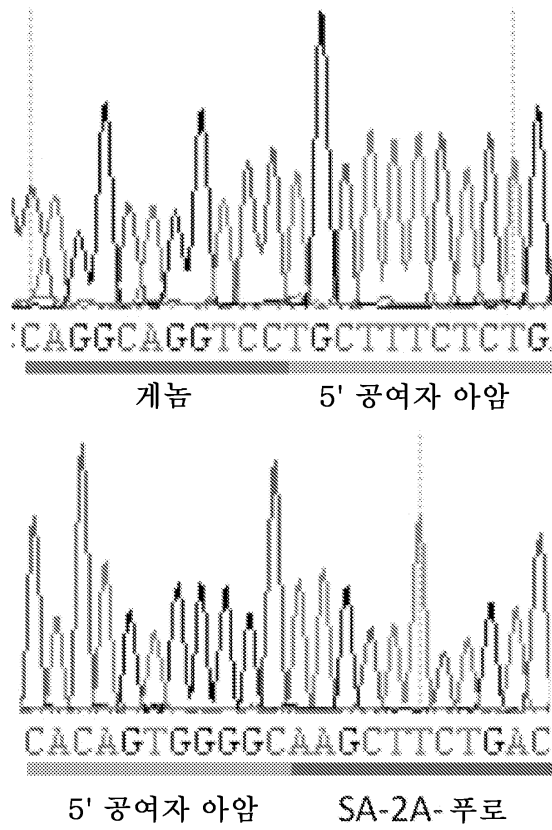
도면2c



도면2d

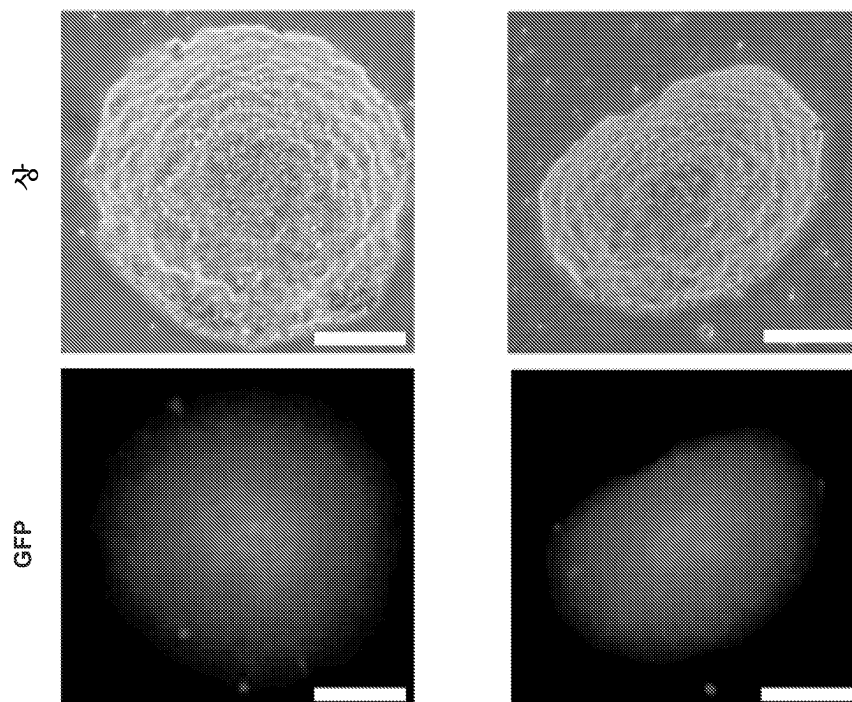


도면2e

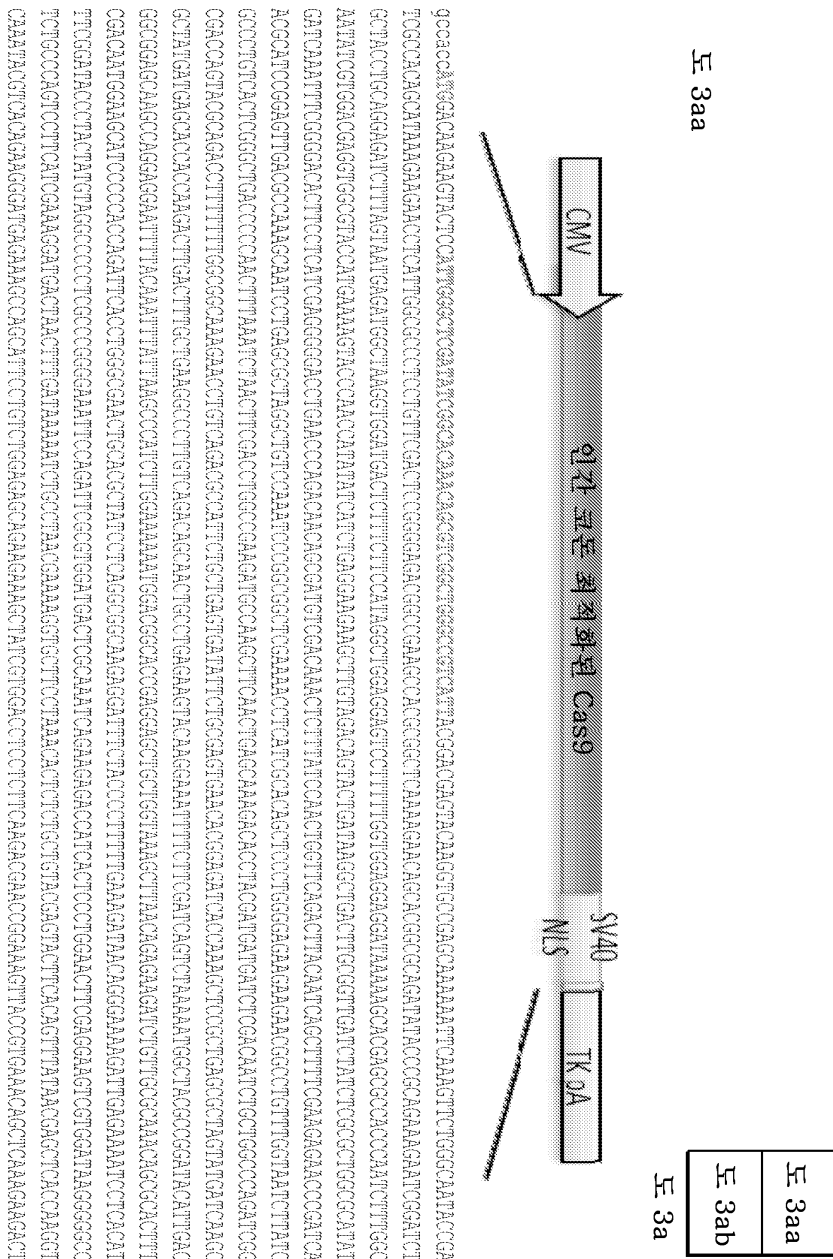


도면2f

표적화된 293T 클론



도면 3aa





도면3ab

도 3ab

ATTTCAAAAAGATTGATGCTTTTCGACTCTGTTGAAATCAGCGGAGTGGAGATGCTTCAACGCATCCCTGGCAAGCTATCAGATCTCTCTGAAATCATTTAAAGACAAGGACTTCCTGGAC  
 AATCAGAGAGACAGAGACATTTCTTGAAGACATTTCTCTACCCCTTACGTTGTTTGAAATAGGAGATGATTTGAAGAAACGCTTGAAACTTACGCTCATCTTCGACGCAAAAGTCATGAA  
 ACAGCTCAGAGAGCGCGGATATACAGGATGGGGGGGGCTGTCAAGAAACTGATCAANTGGGATCCAGACAGAGTGGAAAGACATTCCTGGATTTCCTTAAGTCGATGGATTTTCCCA  
 ACUGGAACTTCATGAGTGTGATCCATGATGATCTCTCACCTTAAAGAGGACATCCAGAAACACAACTTCTGGCCAGGGGACACTTTCAGAGCAATCGCTAATCTTGCAGGTAGC  
 CCAGCTATCAAAAAGGGATTACTGCAGACCCCTTAAGCTGTGGATGAGATGATCTGTAAGTATGGGAGGCAATAAGCCGAGATTCCTTATCGAGATGGCCGAGAGACCAACTACCCA  
 GAAGGACAGAAACACAGTAGGGAAAGGTGAGAGGATTAAGAGGGTTAAAGAGCTGGGCTCCAAATCCTTAAAGAAACACCCAGTTGAAGAACACCCAGCTTCAGATGAGAGCTCT  
 ACCTGATCTACCTGCAGAACGCGAGGACATGTAGTGGATCAGGAACTGGACATCAATCGGCTCCAGCTACAGCTGATCAATCTGCTCCAGCTTTCTCAAAATGATTTCTTAT  
 GATATTAATGCTTTGCAAGATCCGATTAATTAAGAGGGAAAGGTGATTAACGTCCCTCAGAGAACTTGTCAAGAAATGAATAATTATTGGCGCAGCTGCTGAAGCCAAACTGATCAC  
 ACAAAGAACTTCGATTAATCTGACTAAGCTCAACGACCTCCCTGTGTGATTTGAAGCGGCTTCATCAAAAGGCACTTTGACACAGCGCAATCAACAAGCACTGGCCCAAA  
 TTCTCGATTCACGCATGACACCACTAGATGAGATGAATAAGCAAACTGATTCGAGCTGAAGTTATTACTCGAAATGAGCTGATCGAGTTTCAGAAAGCATTCAGATTTTATAG  
 GTAGAGAGATCAACAATTTACACCATGGCATGTGCTTACCTGATTCAGTGGTAGGACCTGCTATCAAAAAATATCCAGCTTAATCTGATTTGTTTACGGAGCTATAAGT  
 GTAGATGTTTGAATAATATGCGAACTGTGAGCAGGAAATAGGCAAGCGCACCGCTAAAGTCTCTTTTACAGCAATATTATGAATTTTTCAGACCGAGTTACACTGGCCATGAG  
 AGATTGCGAAGCGACACTTATCGAAACAAAGGAGAAACAGGAATGTGTGGGACAGGGTAGGATTTCCGACAGTCGGAAAGGTCTGTCCATGTCGCAAGTGAACATGCTTAA  
 AAACCGAAGTACAGACCGAGGCTTCTCCAGGAAGTATCTCCGAAAGAGAACGCAAGCTGATCGACAGCAAAAGATTTGGACCCCAAGAAATACGGCGATTGATTCCTCC  
 TACGTGCTTACAGTGTACTGCTGTGTGGCAAGTCSAGAAAGGAAATCTTAAAAATCAAAACCGTCAAGAAATCTGTGGCATCAATATGAGAGATCAAGCTTCAAAAAAAC  
 CCTTCGACTTTCTCGAGCGCAAGATTAAGAGCTAAAAAAGACCTCATCTTAACTTCCAGTACTCTCTCTTGAAGCTTGAANAAGCGCGAAACGATGCTGCTAGTGGCGC  
 GAGGTGAGAAAGGTACAGAGCTGGACCTGGCCCTTAATATCTTAATTTCTGATCTGGCCAGCAATAGAAAGCTCAAGGGCTCTCCGAGATATAGCAAGAGAGAGCTGTTGCT  
 GGAACACACAAACACTACCTTGAATGATCATCGCAAAATTAAGCAATTTCCAAAGAGTGAATCTCCGCGCAGCTTAACCTGATTAAGTGTCTTTGCTTACATTAAGCACAGGATA  
 AGGCCATCAGGAGAGCAGAAAAAATTAACACTTTTACTGTGACCAACTTGGGCGCGCTCCAGCCTTCAAGTACTTGGACACCAACATAGACAGAAAGCGGTACACTCTACAAG  
 GAGTCTCTGAGCGCCACACTGATTCATGATCAATTAAGGGCTCTTGAACAGAGATGACCTCTCACTGTGTGGAGCAAGCGCCCTTAACCCCAAGAAAGAGAGAGCTGCA





U6 프로모터 + 표적 RNA + 가이드 RNA 스캐폴드:

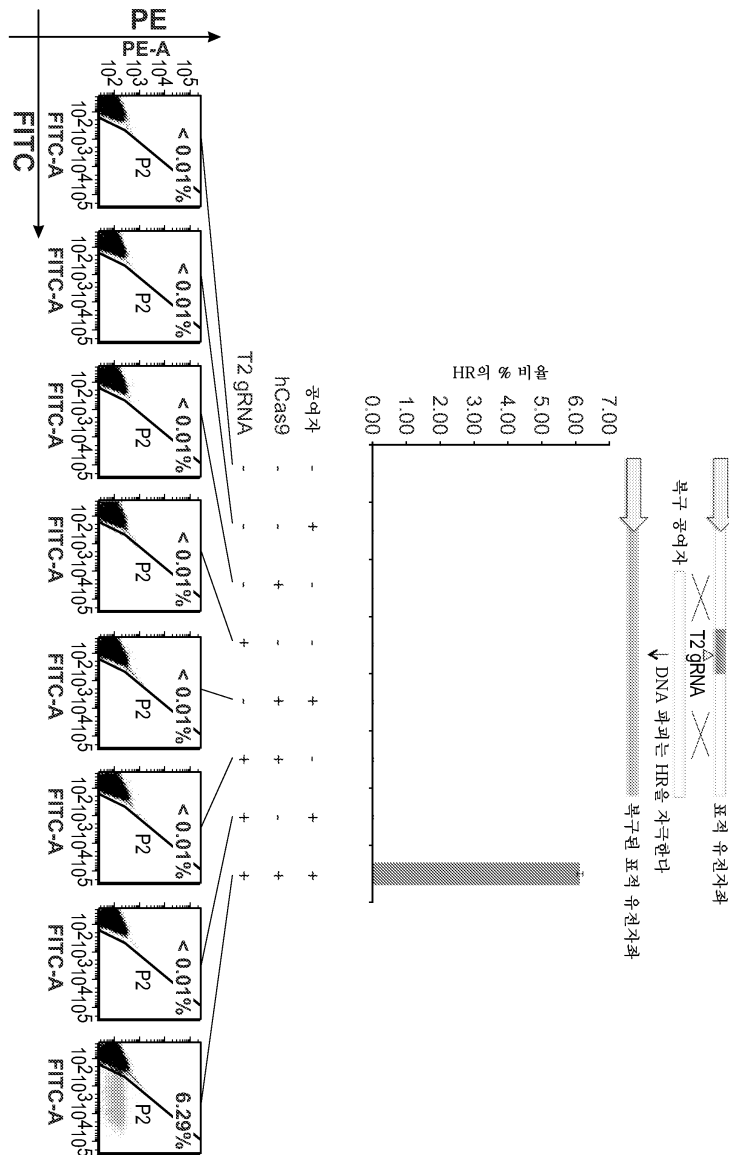
[illegible]

도면 3b

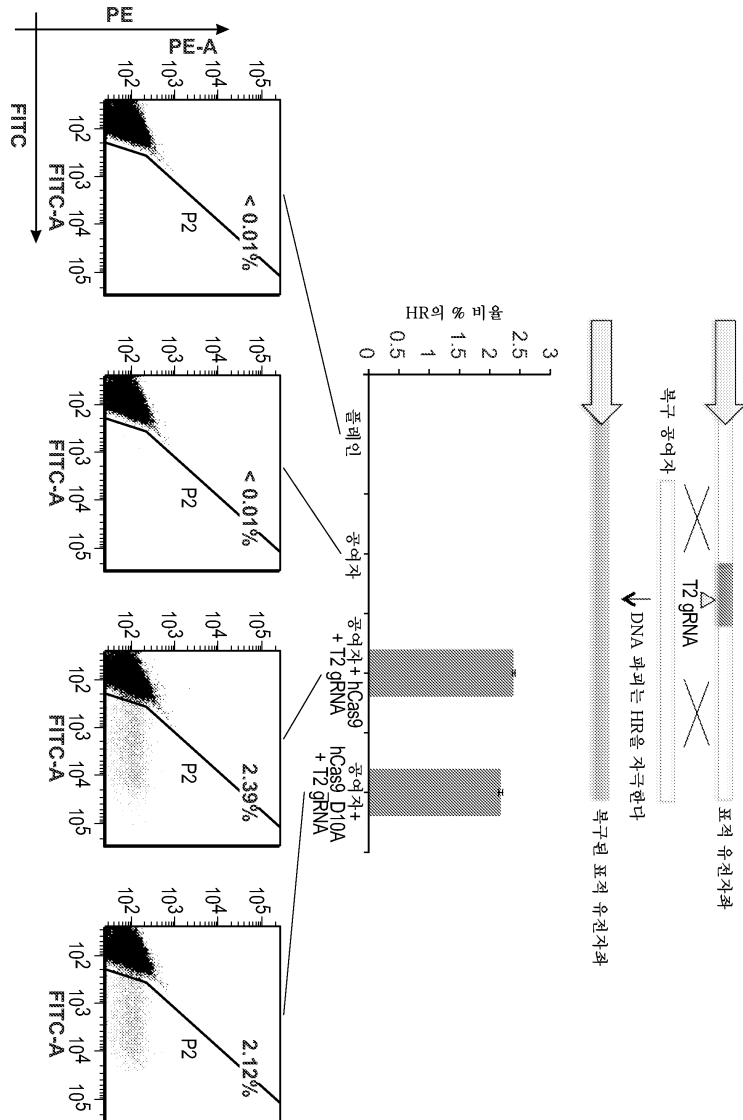
## PAM CGG 표적 서열의 일반적 형태

명칭	표적 서열
GFP gRNA 표적 1	GTGAACCGCATCGAGCTGAAGGG
GFP gRNA 표적 2	GGAGCGCAACATCTTCTTCAAGG
AAVS1 gRNA 표적 1	GTCCCCCTCCACCCCCCACTGGGG
AAVS1 gRNA 표적 2	GGGGCCACTAGGACACAGGATTGG
DNMT3a gRNA 표적 1	GCATGATGCGCGGCCCAAGGAGG
DNMT3a gRNA 표적 2	GAGATGATCGCCCCCTTCTTCTGG
DNMT3b gRNA 표적	GAATTACTCAAGGCCCAAGGAGG

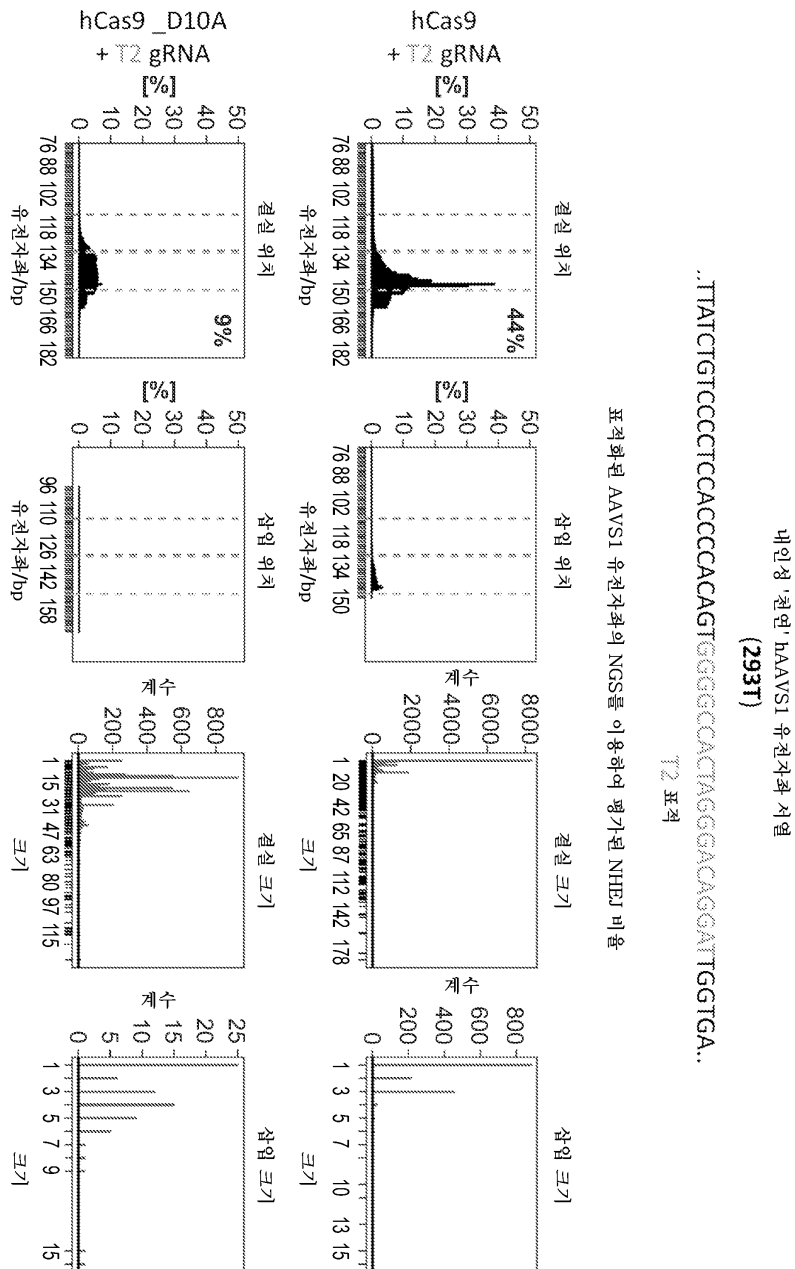
도면4



도면5a



도면5b

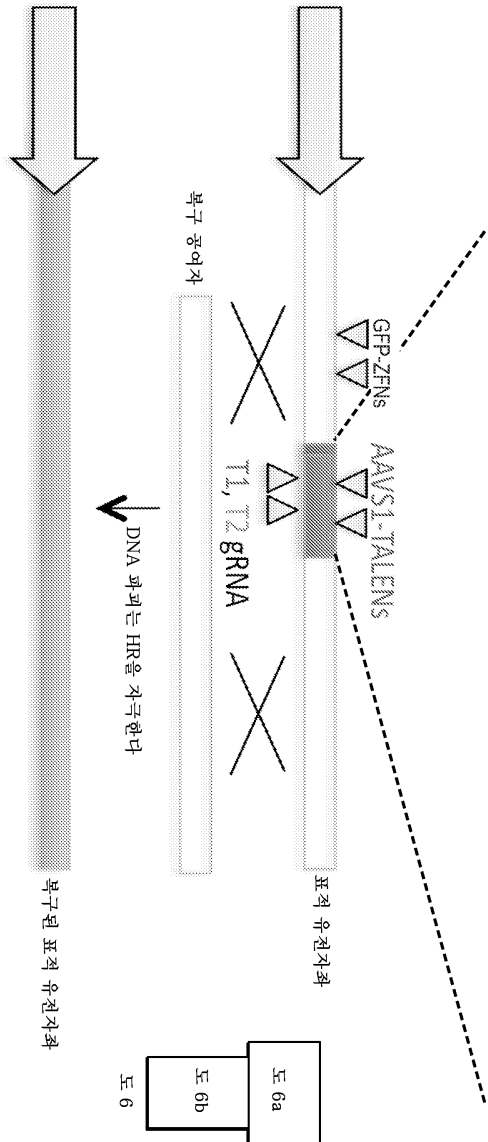


# AAVS1 유전자좌로부터의 68 bp DNA 단편

WT 세포주 유전자좌 서열 : TTAATACTTTTATCTGTCTCCCTCCACCCACAGTGGGGCCACTTAGGACAGGATTGGTGAACAGAAAGCCCC

mut(A) 세포주 유전자좌 서열 : TTAATACTTTTATCTGTCTCAAAAACCCACAGTGGGGCCACTTAGGACAGGATTGGTGAACAGAAAGCCCC

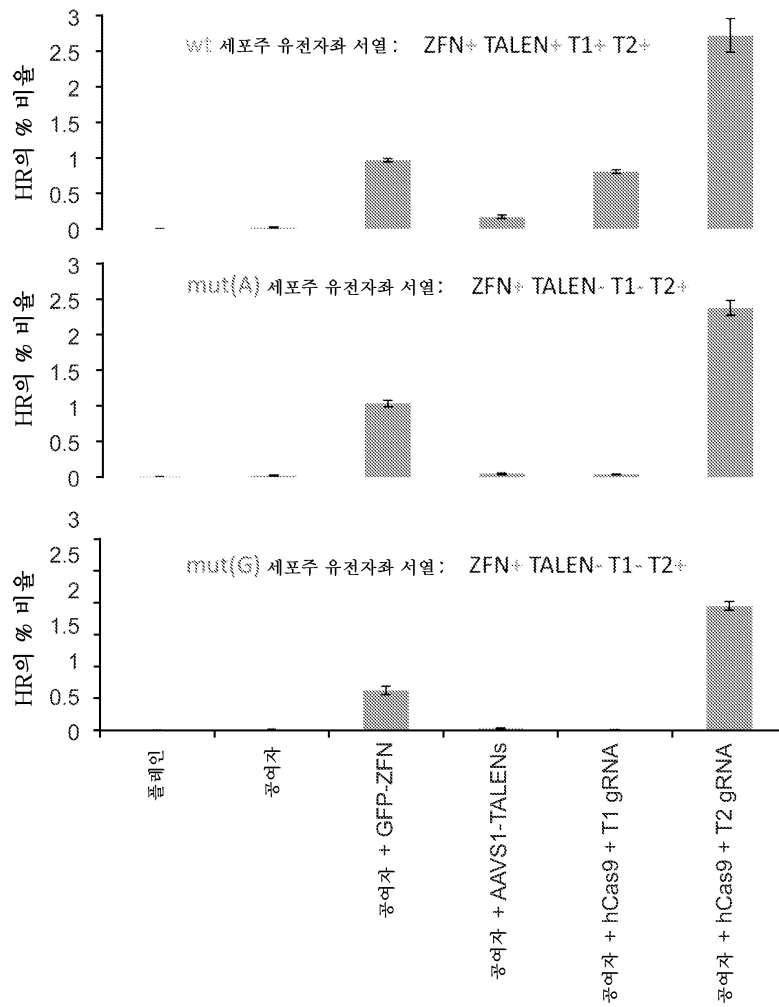
mut(G) 세포주 유전자좌 서열 : TTAATACTTTTATCTGTCTGGGGGAACCCACAGTGGGGCCACTTAGGACAGGATTGGTGAACAGAAAGCCCC



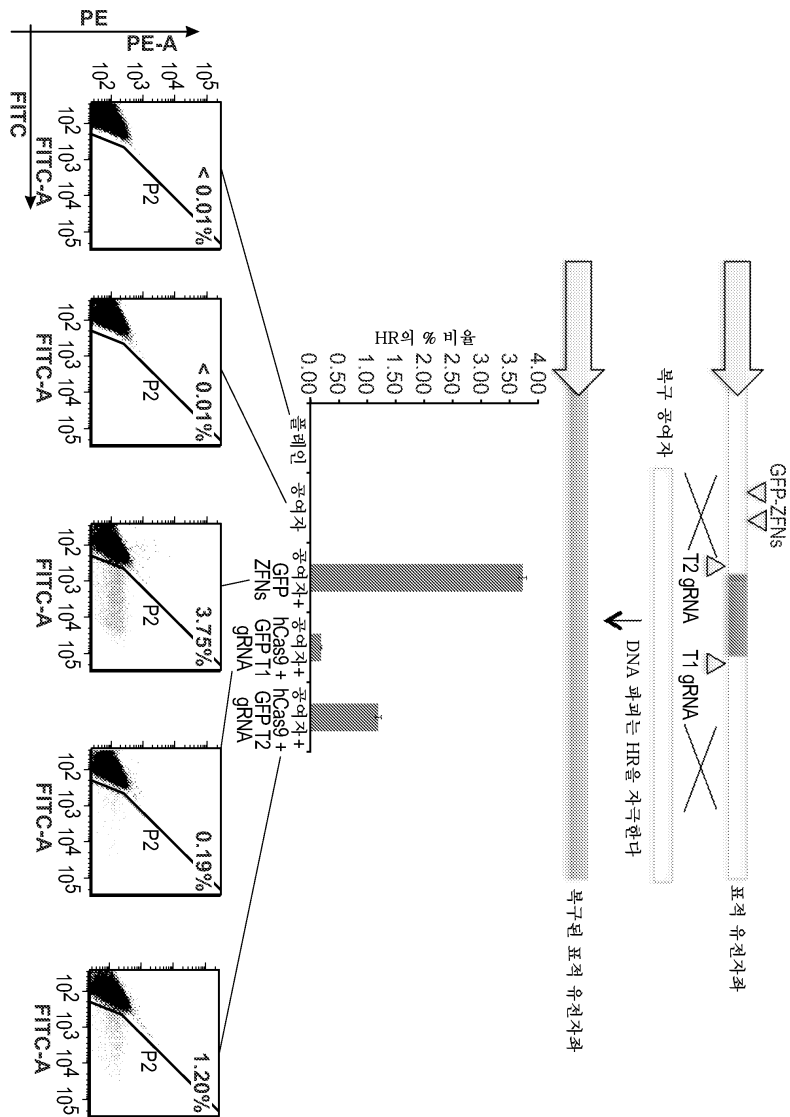
도면6a



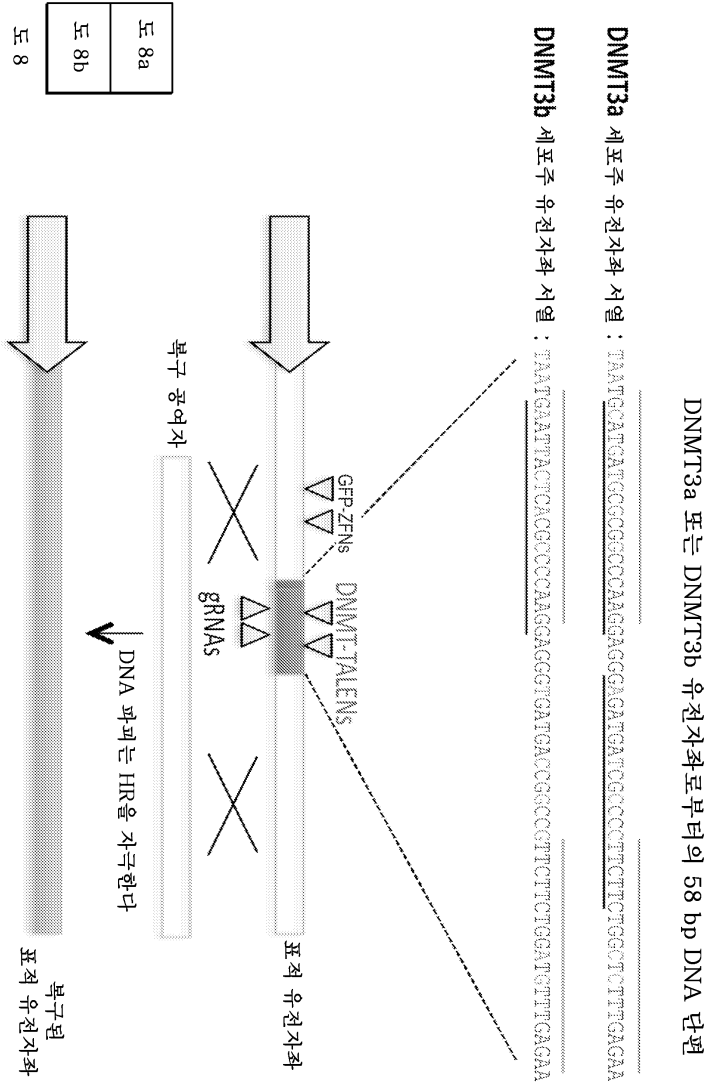
도면6b



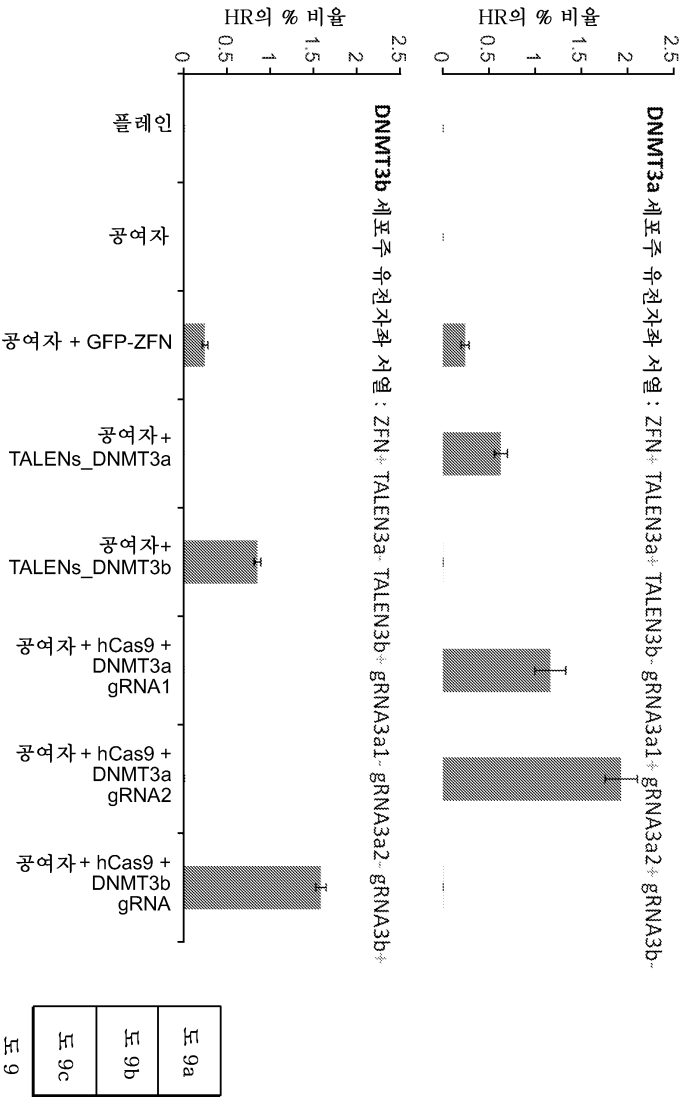
도면7



도면8a



도면8b



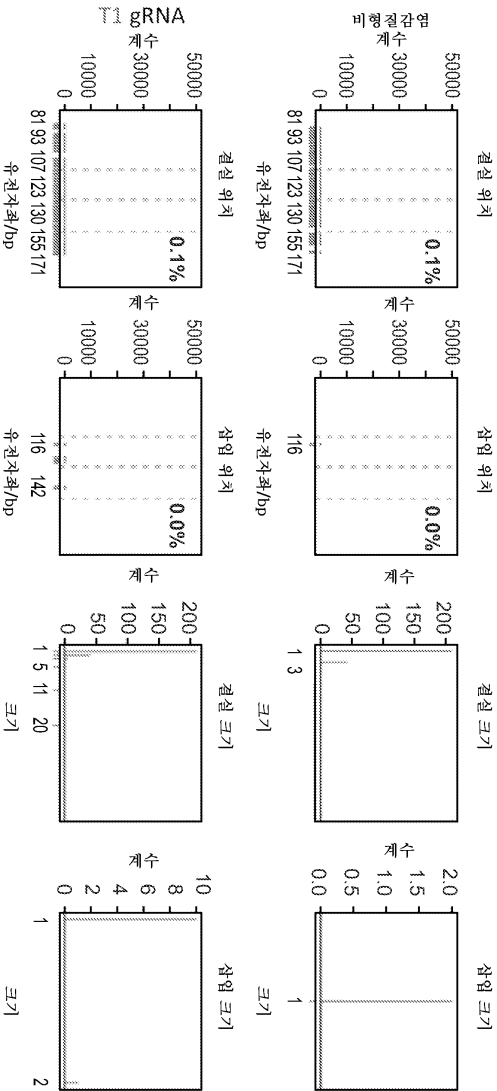
내인성 '원연' hAAVS1 유전자좌 서열  
(PGP1 IPS)

..TTAICTGTCCCTCCACCCACAGTGGGGCCACTAGGGACAGGATTGGTGA..

T1 표적

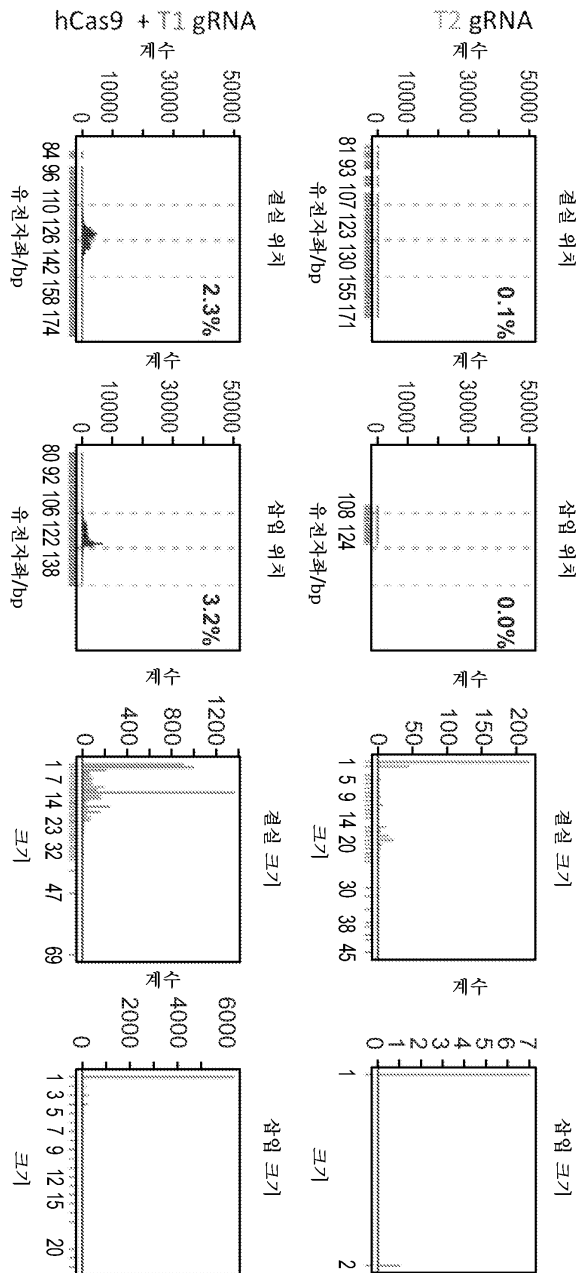
T2 표적

표적화된 AAVS1 유전자좌의 NGS를 이용하여 평가된 NHEJ 비율



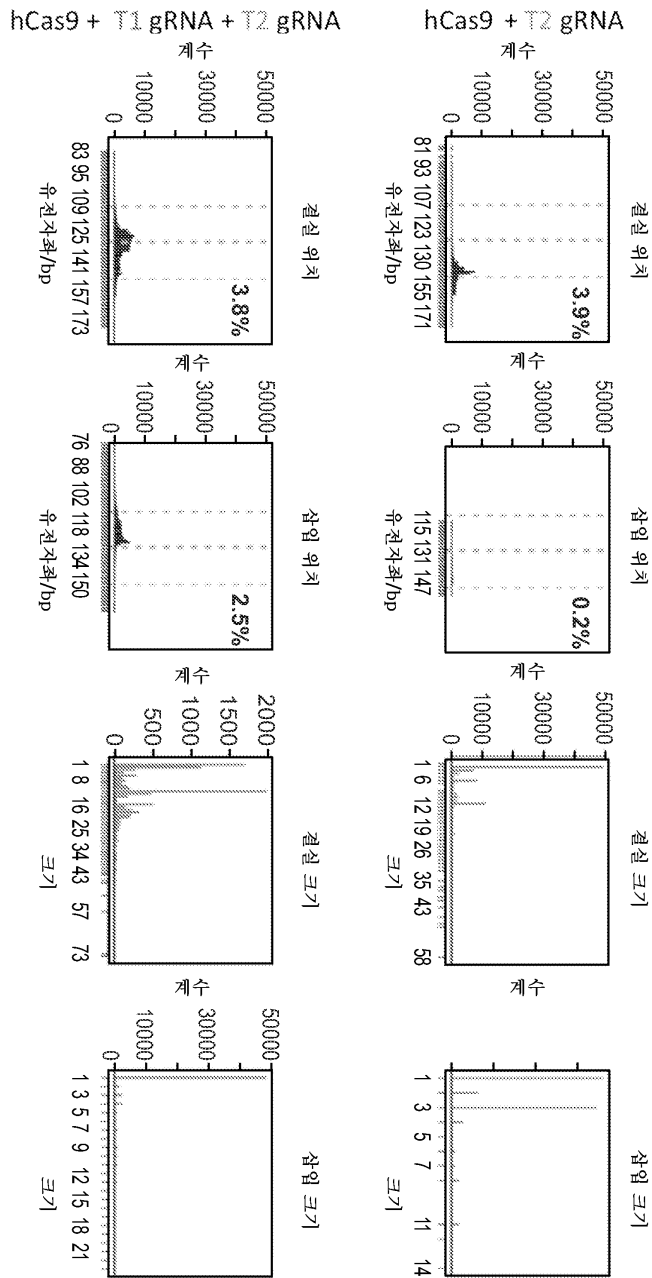
도면9a

도면9b

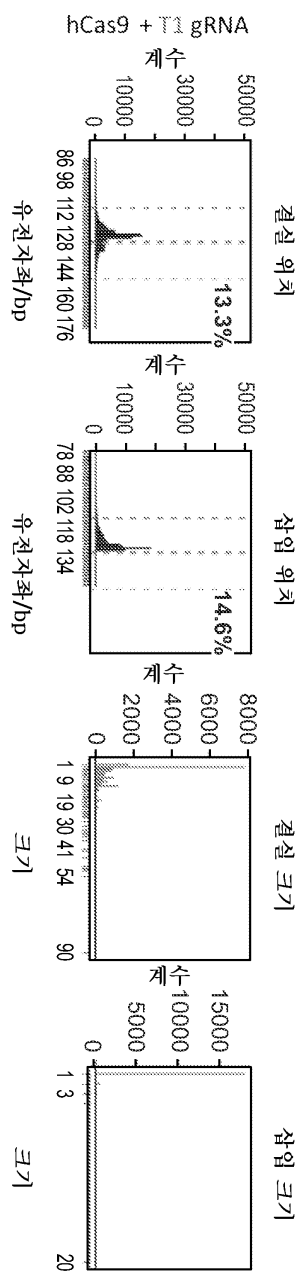




도면9c



도면10a



표적화된 AAVS1 유전자좌의 NGS를 이용하여 평가된 NHEJ 비율

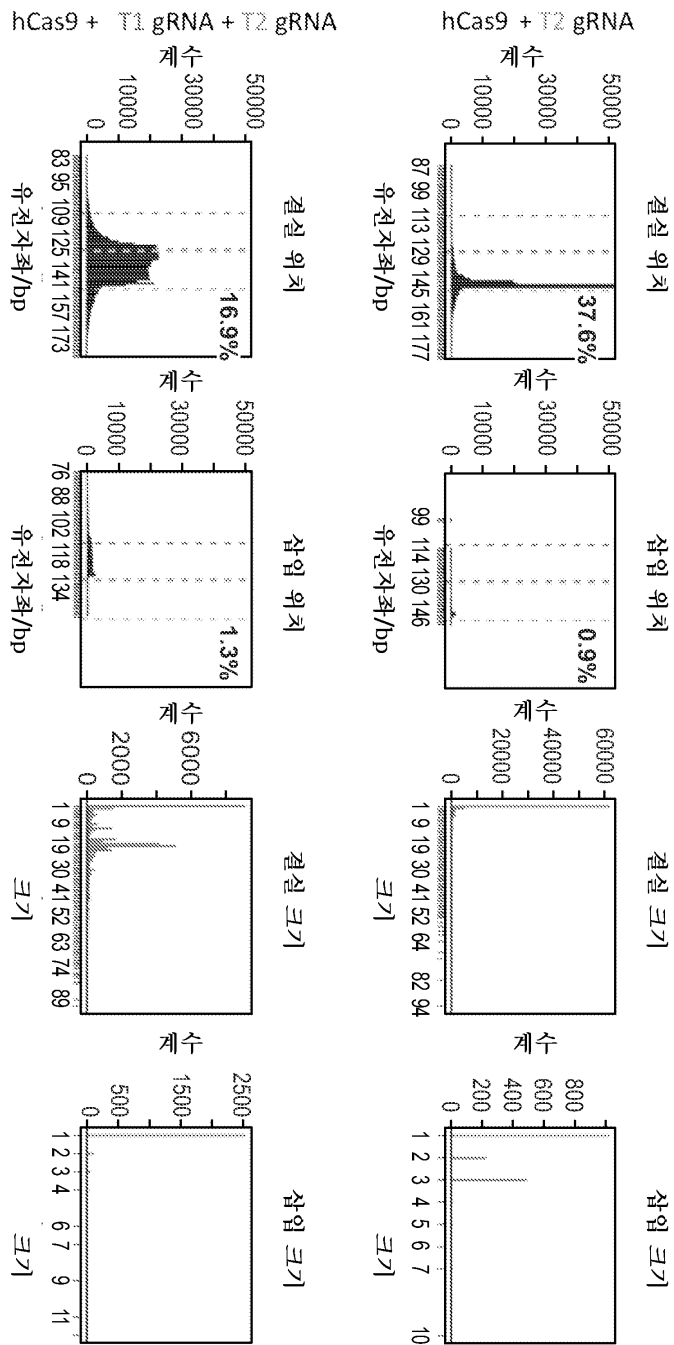
..TTAICTGTCCCCCTCCACCCACACAGTGGGGCCACTAGGGACACAGGATTGGTGA..  
T1 표적 T2 표적

내인성 '천연' hAAVS1 유전자좌 서열  
(K562)

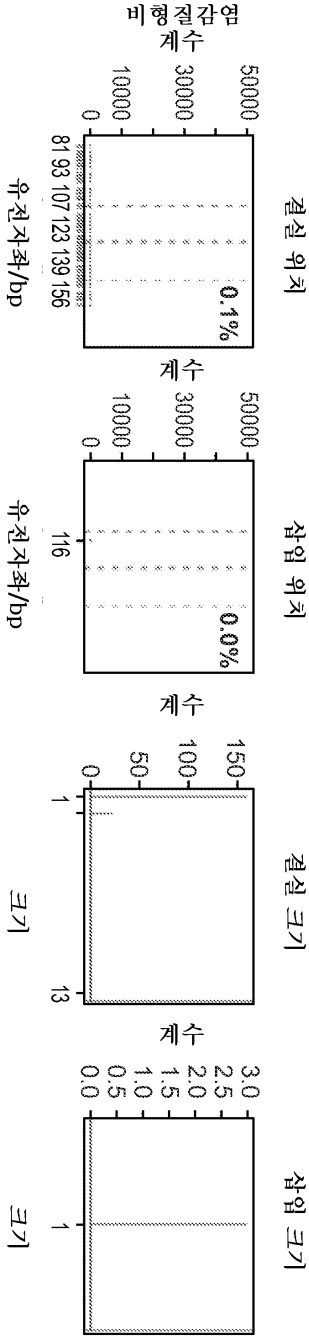
도 10a
도 10b

도 10

도면10b



도면11a



표적화된 AAVS1 유전자좌의 NGS를 이용하여 평가된 NHEJ 비환

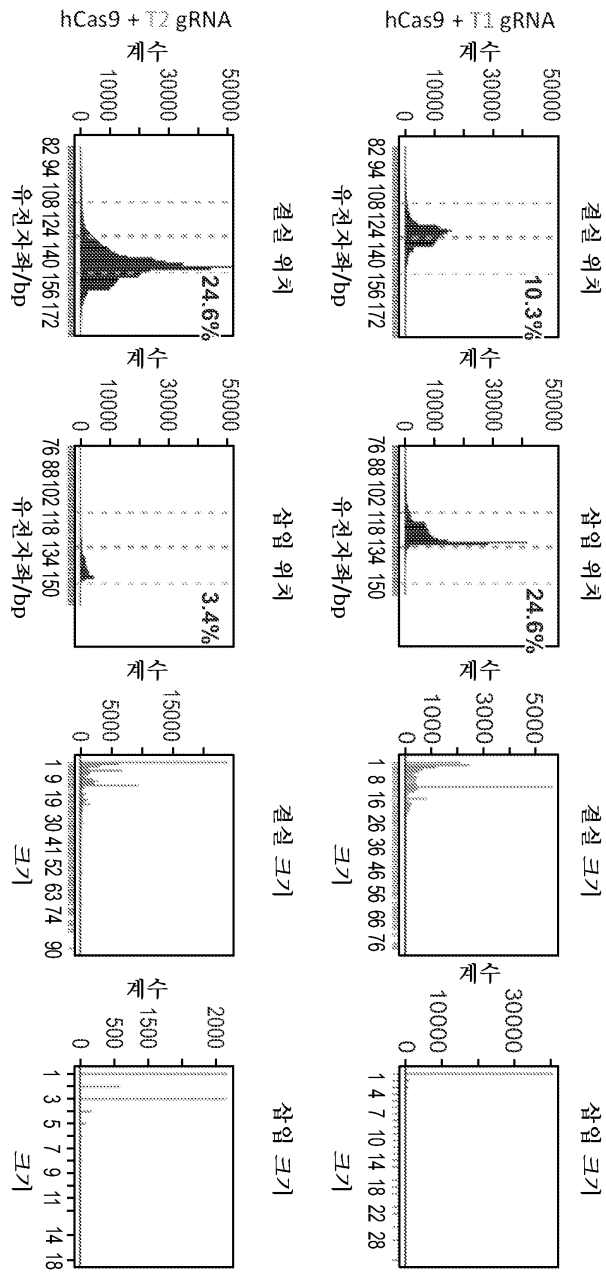
..TTATCTGTCCTCCACCCACACAGTGGGGCCACTAGGGACACAGGATTGGTGA..  
T1 표적 T2 표적

내인성 '결실' hAAVS1 유전자좌 서열  
(293T)

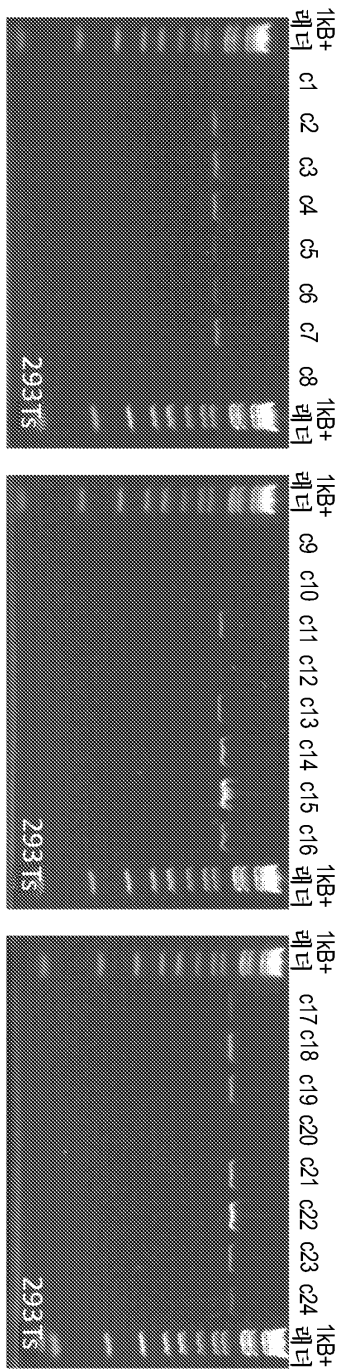
도 11a
도 11b

도 11

도면11b

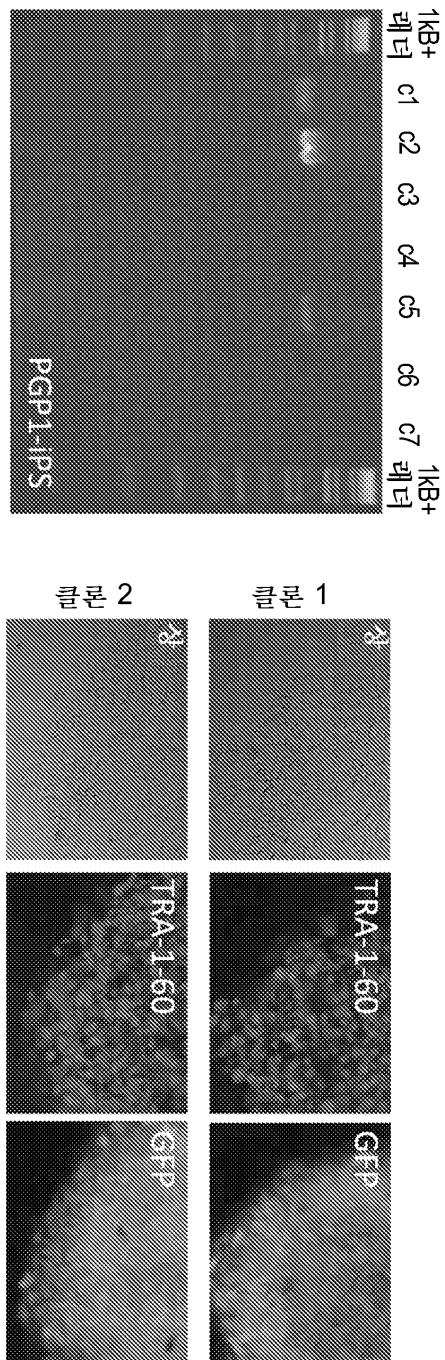


도면12a





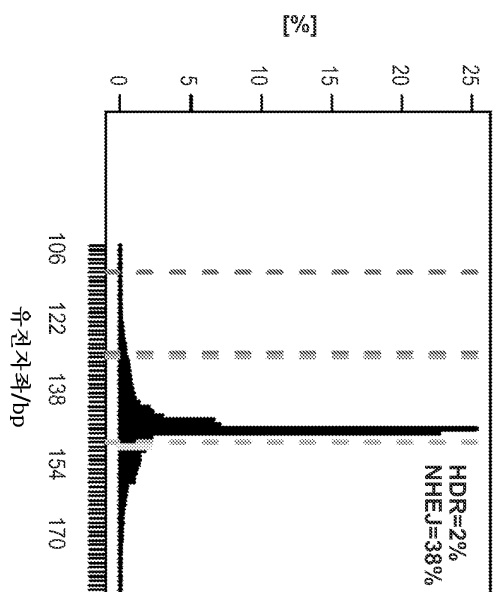
도면12b



내인성 천연 hAAVS1 유전자좌 서열 - AAVS1 T2 gRNA 표적 부위가 강조되어 있다

...GGTACTTTTATCTGTCCCTCCACCCCAAGTGGGCGCATTTAGGACAGGATTTGTGACAGAAAGCCCATCTCTTTAGGCGTCTCTTCTAGTCTCTCTGATATTGGTC...

K562에서 AAVS1 유전자좌의 표적화: hCas9 + AAVS1 T2 gRNA + 90량체 올리고 공여자



도면 12c

13a

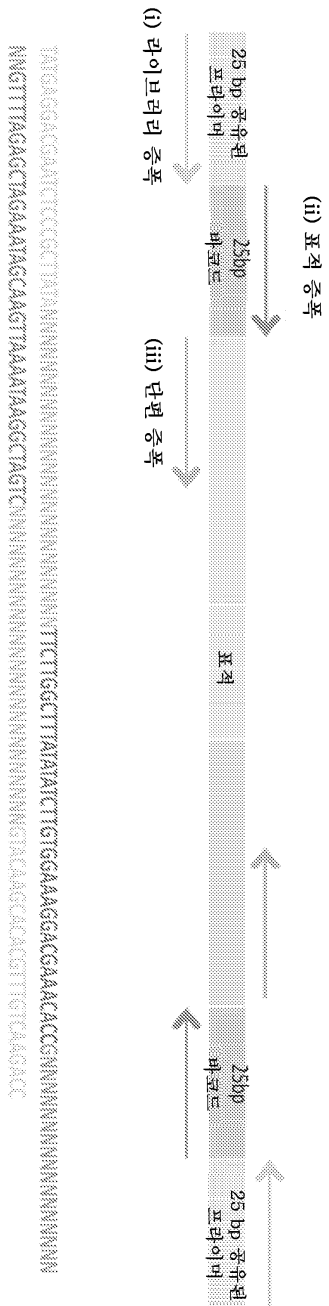
# 인간 게놈 중의 유전자에 대한 gRNA 표적 부위

.....

200 bp DNA 아레이 기반  
함성 포맷 내로 혼입됨

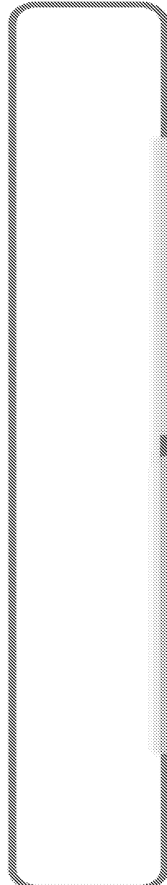
$\mathbb{F}_{13aa}$
$\mathbb{F}_{13ab}$
$\mathbb{F}_{13a}$

도면 13aa

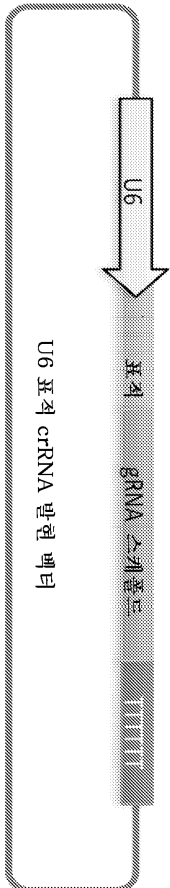


도 13ab

+ 동상의 수용자 벡터의 AflII 부위 내로  
표적 단편 앵플리콘의 동온 조립

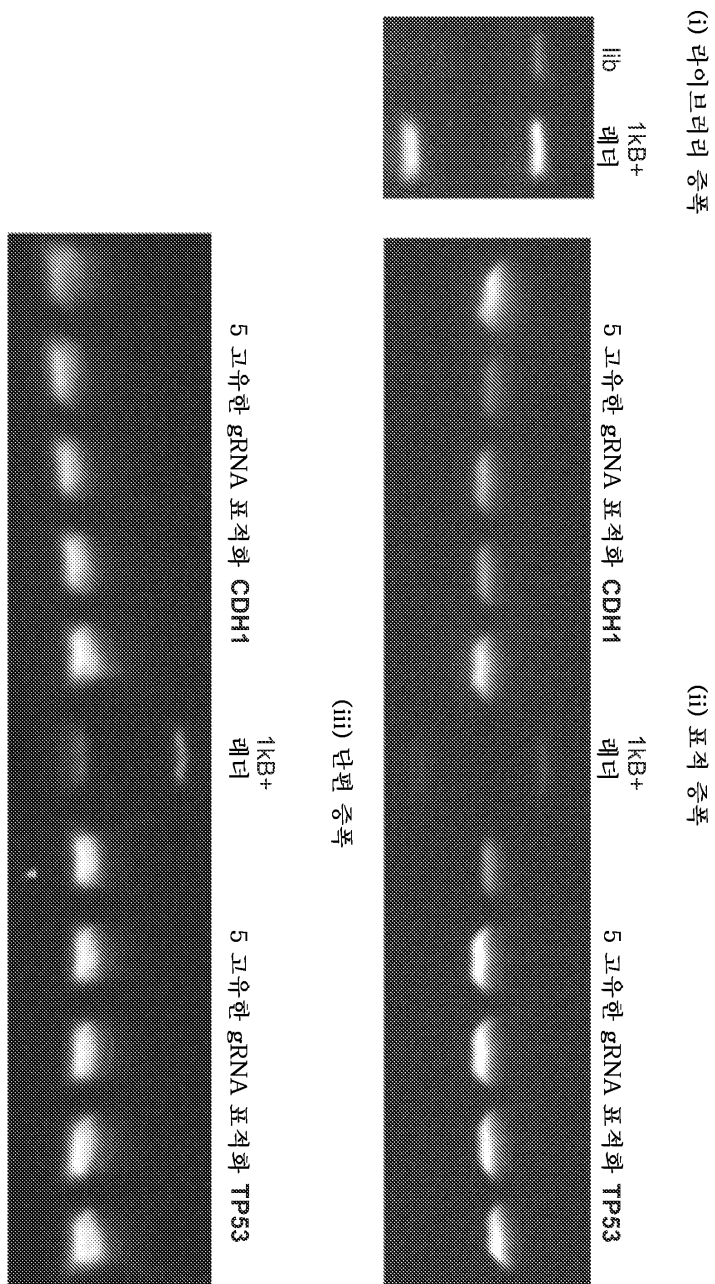


...ATTCGCCCTTGTACAAAAAGCAGCTTAAAGGAACCAATTCAGTCGATCGGTACCAAGTCGGCAGGAGAGAGGCGCTATTCATGATCTTCAATTTGC  
ATATCGATACAGGCTGTAGAGAGATATAGATTTATTTGACTGAACACAAAGATATAGTACAAATATGACGTAGAAAGTATATTTCTTGGTATGTTGCAGTT  
TTAAATATGTTTAAATGACTATCATAGCTTACCGTACTGAAAGTATTCGATTTCTGCTTATATATCTTAAAGTTAAATAAGGCTAGTCTTATCACTTGAAAAA  
CTGCGACCGAGTGGCTTTTCTAGACCGACTTCTGTACAAAGTTGCGATTAAAGGCGAAT..

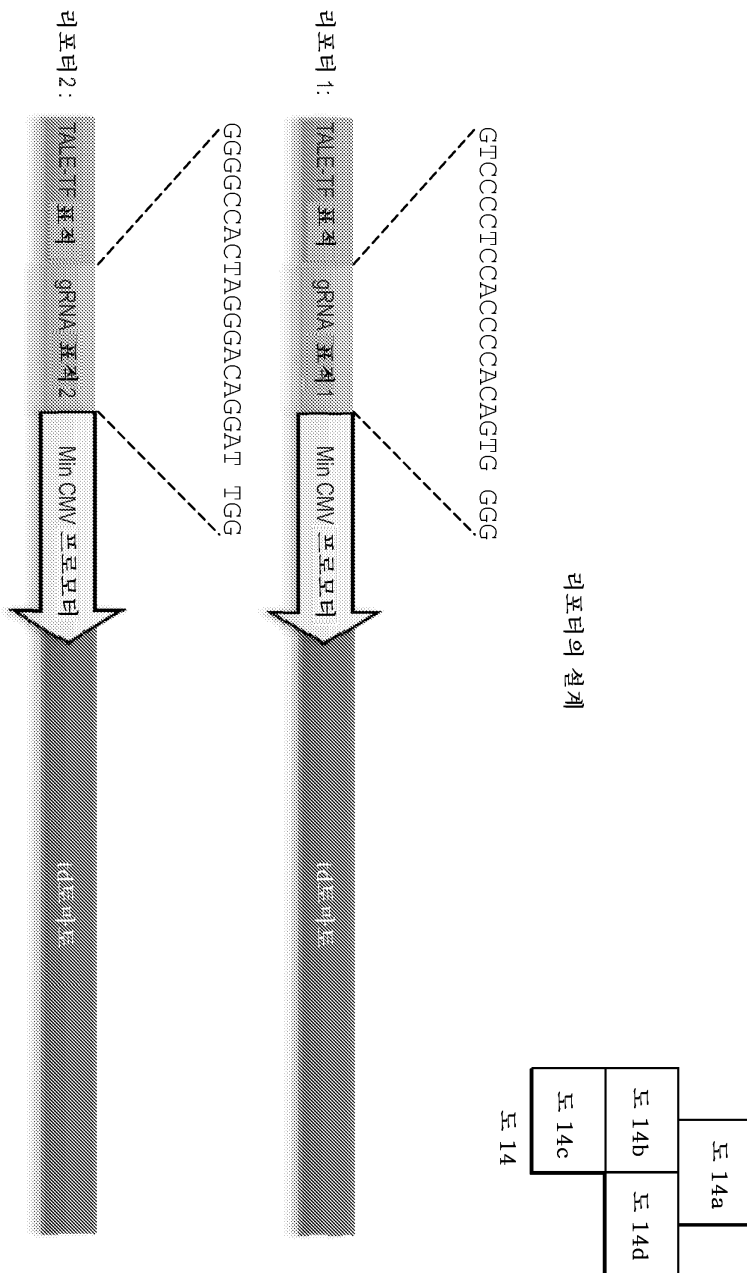


도면 13ab

도면13b

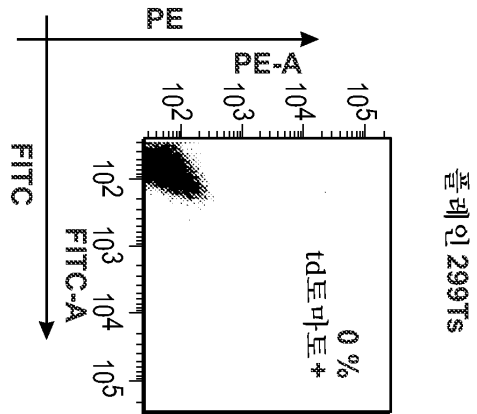


도면14a



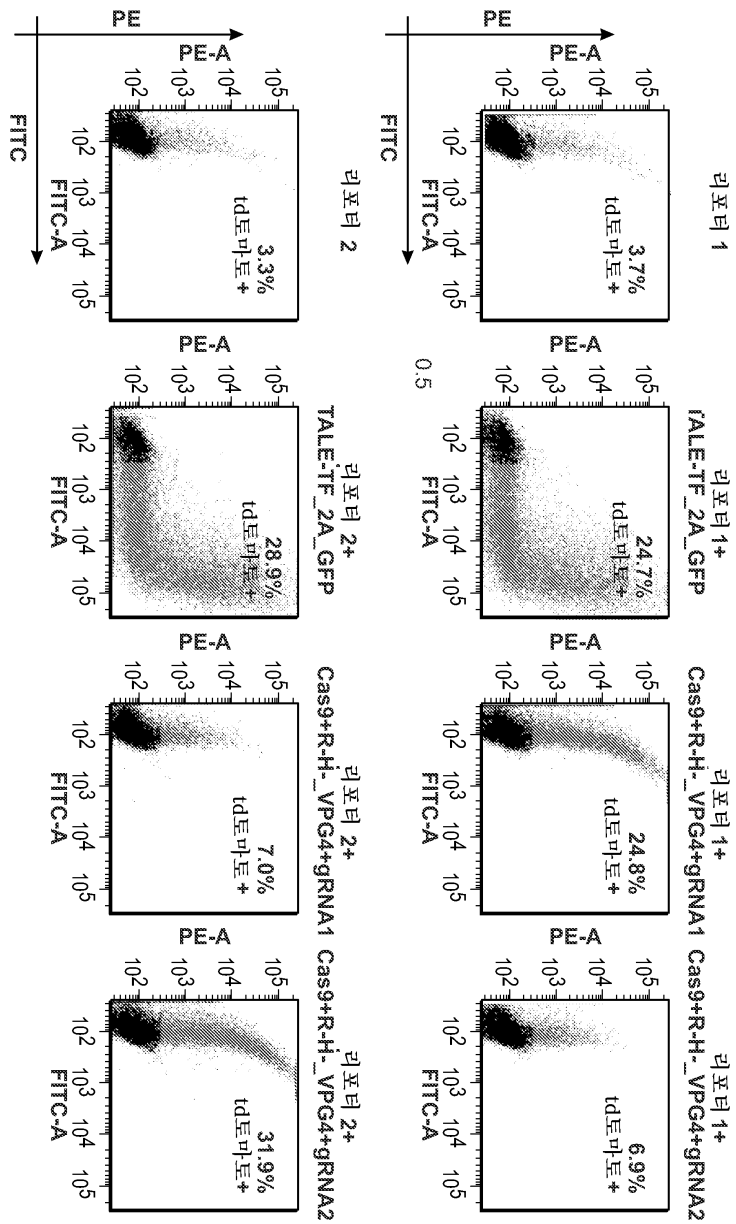


도면14b

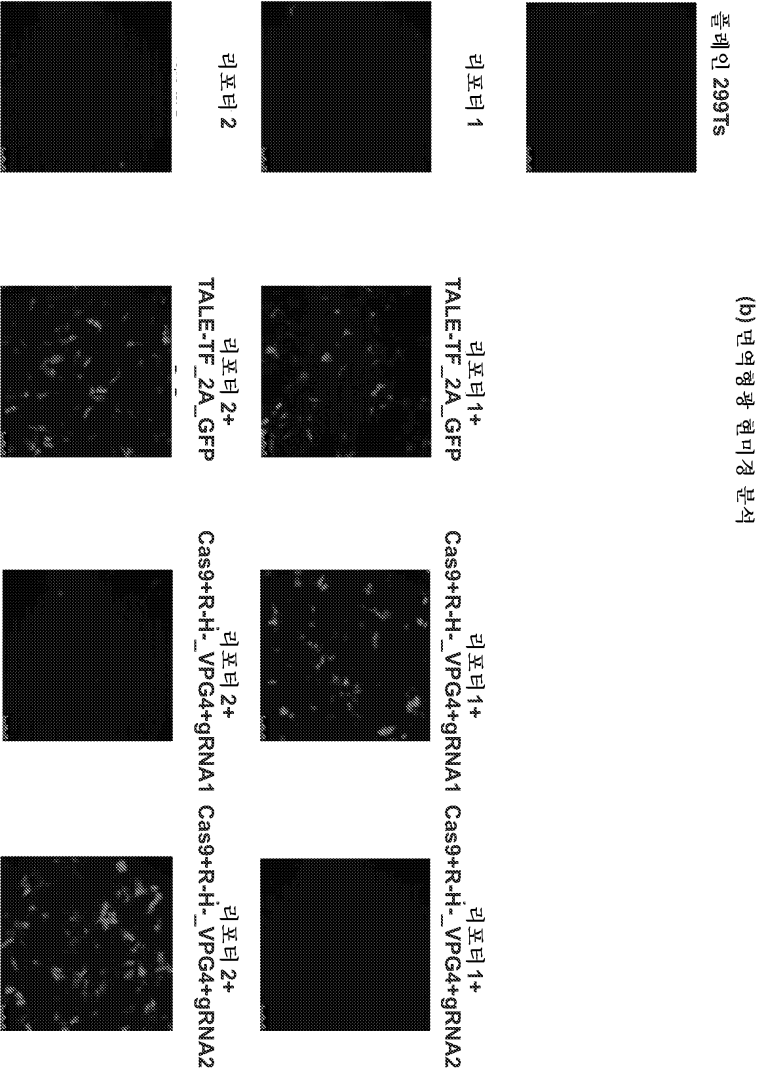


(a) 형광 활성화 세포 분류 (FACS) 분석

도면14c



도면14d



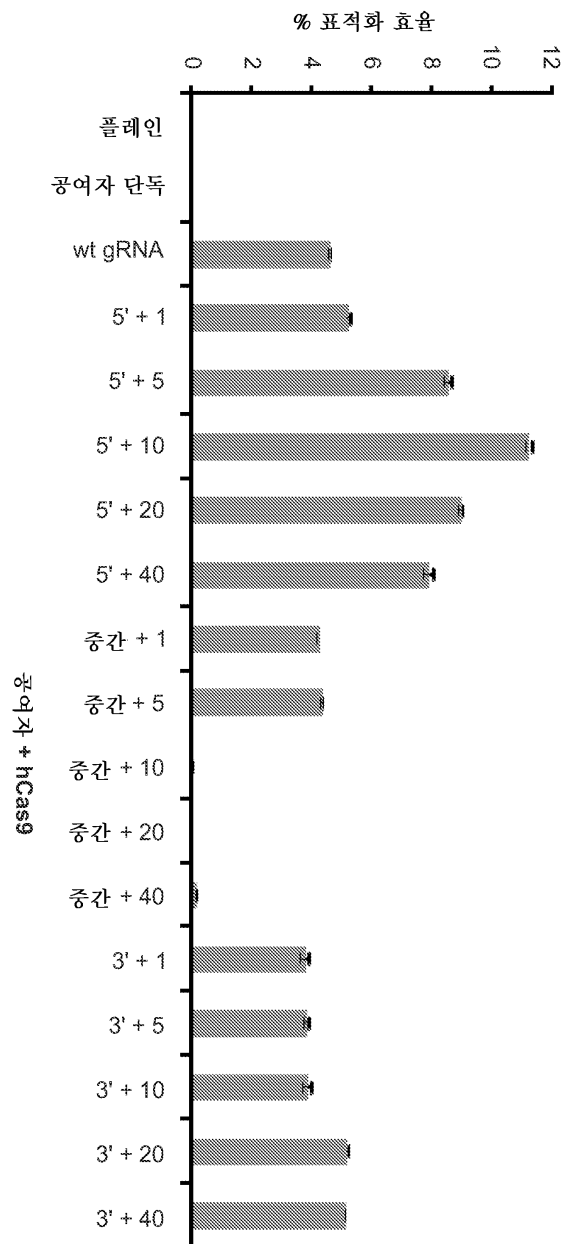
도면15a



도 15a
도 15b

도 15

도면15b



## 서열 목록

### SEQUENCE LISTING

<110> Church, George

Mali, Prashant

Yang, Luhan

<120> RNA-Guided Human Genome Engineering

<130> 010498-00463

<140> PCT/US13/075317

<141> 2013-12-16

<150> US 61/738,355

<151> 2012-12-17

<150> US 61/779,169

<151> 2013-03-13

<160> 45

<170> PatentIn version 3.5

<210> 1

<211> 268

<212> DNA

<213> Homo sapiens

<400> 1

catttcagga cagcatgttt gctgcctcca gggatcctgt gtccccgagc tgggaccacc 60

ttatatccccc agggccgggtt aatgtggctc tggttctggg tactttttatc tgtcccctcc 120

acccacagct ggggccacta gggacaggat tggtagacaga aaagcccat ccttaggcct 180

cctccttctc agtctcctga tattgggtct aacccccacc tctgttagg cagattcctt 240

atctggtgac acaccccat ttcctgga 268

<210> 2

<211> 53

<212> DNA

<213> Artificial

<220><223> PCR primer

<400> 2

ctcggcattc ctgctgaacc gctcttcga tctacaggag gtgggggtta gac 53

<210> 3

<211> 59

<212> DNA

<213> Artificial

<220><223> PCR primer

<400> 3

acactcttc cctacacgac gctcttcga tctcgtgatt atattccag ggccggtta 59

<210> 4

<211> 59  
 <212> DNA  
 <213> Artificial  
 <220><223> PCR primer  
 <400> 4  
 acactctttc cctacacgac gctcttccga tctacatcgt atattcccag ggccgggta 59  
 <210> 5  
 <211> 59  
 <212> DNA  
 <213> Artificial  
 <220><223> PCR primer  
 <400> 5  
 acactctttc cctacacgac gctcttccga tctgcctaata atattcccag ggccgggta 59  
 <210> 6  
 <211> 59  
 <212> DNA  
 <213> Artificial  
 <220><223> PCR primer  
 <400> 6  
 acactctttc cctacacgac gctcttccga tcttggtcat atattcccag ggccgggta 59  
 <210> 7  
 <211> 59  
 <212> DNA  
 <213> Artificial  
 <220><223> PCR primer  
 <400> 7  
 acactctttc cctacacgac gctcttccga tctcactggt atattcccag ggccgggta 59  
 <210> 8  
 <211> 59  
 <212> DNA  
 <213> Artificial  
 <220><223> PCR primer  
 <400> 8  
 acactctttc cctacacgac gctcttccga tctattggct atattcccag ggccgggta 59



<210> 9

<211> 59

<212> DNA

<213> Artificial

<220><223> PCR primer

<400> 9

acactctttc cctacacgac gctcttccga tctgatctgt atattcccag ggccggtta 59

<210> 10

<211>

59

<212> DNA

<213> Artificial

<220><223> PCR primer

<400> 10

acactctttc cctacacgac gctcttccga tcttcaagtt atattcccag ggccggtta 59

<210> 11

<211> 59

<212> DNA

<213> Artificial

<220><223> PCR primer

<400> 11

acactctttc cctacacgac gctcttccga tctctgatct atattcccag ggccggtta 59

<210> 12

<211> 59

<212> DNA

<213> Artificial

<220><223> PCR primer

<400> 12

acactctttc cctacacgac gctcttccga tctaagctat atattcccag ggccggtta 59

<210> 13

<211>

59

<212> DNA

<213> Artificial

<220><223> PCR primer

<400> 13	
acactctttc cctacacgac gctcttccga tctgtagcct atattcccag ggccggtta	59
<210> 14	
<211> 59	
<212> DNA	
<213> Artificial	
<220><223> PCR primer	
<400> 14	
acactctttc cctacacgac gctcttccga tcttacaagt atattcccag ggccggtta	59
<210> 15	
<211> 20	
<212> DNA	
<213> Artificial	
<220><223> PCR primer	
<400> 15	
ctgccgtctc tctcctgagt	20
<210> 16	
<211>	
20	
<212> DNA	
<213> Artificial	
<220><223> PCR primer	
<400> 16	
gtgggcttgt actcggtcat	20
<210> 17	
<211> 80	
<212> RNA	
<213> Artificial	
<220><223> guide RNA	
<400> 17	
guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu	60
ggcaccgagu cggugcuuuu	80
<210> 18	
<211> 71	
<212> DNA	

<213> Homo sapiens  
 <400> 18  
 taatactttt atcttgtccc ctccacccca cagtggggcc actagggaca ggattggtga 60

cagaaagccc c 71  
 <210> 19  
 <211> 52  
 <212> DNA  
 <213> Homo sapiens  
 <400> 19  
 ttatctgtcc cctccacccc acagtggggc cactagggaa caggattggt ga 52

<210> 20  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <400> 20  
 caggcaggtc ctgctttctc tg 22

<210> 21  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <400> 21  
 cacagtgggg caagcttctg ac 22

<210> 22  
 <211> 4146  
 <212> DNA  
 <213> Homo sapiens  
 <400> 22  
 gccacatgg acaagaagta ctccattggg ctcgatatcg gcacaaacag cgtcggctgg 60  
 gccgtcatta cggacgagta caagtgccg agcaaaaaat tcaaagttct gggcaatacc 120  
 gatcgccaca gcataaagaa gaacctcatt ggcgccctcc tgttcgactc cggggagacg 180  
 gccgaagcca cgcggctcaa aagaacagca cggcgcagat atacccgag aaagaatcgg 240  
 atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc 300

cataggctgg aggagtcctt ttgtgtggag gaggataaaa agcacgagcg ccaccaatc	360
tttggcaata tcgtggacga ggtggcgtag catgaaaagt acccaacat atatcatctg	420
aggaagaagc tttagacag tactgataag gctgacttgc ggttgatcta tctcgcgctg	480
gcgcatatga tcaaatttcg gggacacttc ctcatcgagg gggacctgaa cccagacaac	540
agcgatgtcg acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag	600
aaccgatca acgcatccgg agttgacgcc aaagcaatcc tgagcgctag gctgtccaaa	660
tcccgccggc tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt	720
ggtaatctta tcgccctgtc actcgggctg accccaact ttaaattctaa cttcgacctg	780
gccgaagatg ccaagcttca actgagcaaa gacacctacg atgatgatct cgacaatctg	840
ctggcccaga tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac	900
gccattctgc tgagtgatat tctgcgagtg aacacggaga tcaccaaagc tccgctgagc	960
gctagtatga tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt	1020
gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc	1080
tacgccggat acattgacgg cggagcaagc caggaggaat tttaaaatt tattaagccc	1140
atcttggaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg	1200
ttgcgcaaac agcgcacttt cgacaatgga agcatcccc accagattca cctgggcgaa	1260
ctgcacgcta tctcaggcg gcaagaggat ttctaccct tttgaaaga taacagggaa	1320
aagattgaga aaatctcac atttcggata ccctactatg taggccccct cggccgggga	1380
aattccagat tcgctggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc	1440
gaggaagtgc tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt	1500
gataaaaatc tgcctaacga aaaggtgctt cctaaacact ctctgctgta cgagtacttc	1560
acagtttata acgagctcac caaggtcaaa tacgtcacag aagggatgag aaagccagca	1620
ttctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa	1680
gttacctga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt	1740
gaaatcagcg gagtggagga tcgcttcaac gcatccctgg gaacgtatca cgatctcctg	1800
aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac	1860
attgtctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact	1920
tacgtcatc tcttcagca caaagtcatg aaacagctca agaggcgccg atatacagga	1980
tggggcgccg tgtcaagaaa actgatcaat gggatccgag acaagcagag tggaaagaca	2040

atcctggatt ttcttaagtc cgatggattt gccaacccga acttcatgca gttgatccat	2100
gatgactctc tcacctttta ggaggacatc cagaaagcac aagtttctgg ccagggggac	2160
agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatactg	2220
cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccagagaat	2280
atcgttatcg agatggcccc agagaaccaa actaccaga agggacagaa gaacagtagg	2340
gaaaggatga agaggattga agagggtata aaagaactgg ggtcccaaat ccttaaggaa	2400
caccagttg aaaacacca gcttcagaat gagaagctct accgtacta cctgcagaac	2460
ggcagggaca tgtactgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg	2520
gatcatatcg tgccccagtc ttttctcaaa gatgattcta ttgataataa agtgttgaca	2580
agatccgata aaaatagagg gaagagtga aacgtccctt cagaagaagt tgtcaagaaa	2640
atgaaaaatt attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat	2700
aatctgacta aggtcgaacg aggtggcctg tctgagttgg ataaagccgg ctatcatcaa	2760
aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaattctt cgattcacgc	2820
atgaacacca agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg	2880
aagtctaagc tggctctcaga tttcagaaag gactttcagt ttataaagg gagagagatc	2940
aacaattacc accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttacc	3000
aaaaaatatc ccaagcttga atctgaattt gtttacggag actataaagt gtacgatgtt	3060
aggaaaaatga tcgcaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt	3120
tacagcaata ttatgaattt tttcaagacc gagattacac tggccaatgg agagattcgg	3180
aagcgaccac ttatcgaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg	3240
gatttcgcga cagtcgggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc	3300
gaagtacaga ccggaggttt ctccaaggaa agtatcctcc cgaaaaggaa cagcgacaag	3360
ctgatcgac gcaaaaaaga ttgggacccc aagaaatagc gcggattcga ttctcttaca	3420
gtcgtttaca gtgtactggt tgtggccaaa gtggagaaa ggaagtctaa aaaactcaa	3480
agcgtcaagg aactgctggg catcacaatc atggagcgat caagcttcga aaaaaacccc	3540
atcgactttc tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt	3600
cccaagtact ctctctttga gcttgaaaac ggccggaaac gaatgctcgc tagtgcgggc	3660
gagctgcaga aaggtaacga gctggcactg ccctctaaat acgttaattt cttgtatctg	3720
gccagccact atgaaaagct caaagggtct cccgaagata atgagcagaa gcagctgttc	3780
gtggaacaac acaaacacta ctttgatgag atcatcgagc aaataagcga attctccaaa	3840
agagtgatcc tcgccgacgc taacctgat aagggtgcttt ctgcttaca taagcacagg	3900

gataagccca tcagggagca ggcagaaaac attatccact tgtttactct gaccaacttg 3960

ggcgcgcctg cagccttcaa gtacttcgac accaccatag acagaaagcg gtacacctct 4020

acaaaggagg tcctggacgc cacactgatt catcagtcaa ttacggggct ctatgaaaca 4080

agaatcgacc tccttcagct cggtggagac agcagggtg accccaagaa gaagaggaag 4140

gtgtga 4146

<210> 23

<211> 455

<212> DNA

<213> Homo sapiens

<220><221> misc\_feature

<222> (320)..(338)

<223> n is a, c, g, or t

<400> 23

tgtacaaaa agcaggcttt aaaggaacca attcagtcga ctggatccgg taccaaggtc 60

gggcaggaag agggcctatt tcccatgatt ccttcattat tgcatatacg atacaaggct 120

gttagagaga taattagaat taatttgact gtaaacacaa agatattagt acaaaatagc 180

tgacgtagaa agtaataatt tcttgggtag ttgcagttt taaaattatg ttttaaatg 240

gactatcata tgcttaccgt aacttgaaag tatttcgatt tcttggcttt atatatcttg 300

tggaaaggac gaaacaccgn nnnnnnnnnn nnnnnnnngt tttagagcta gaaatagcaa 360

gttaaaataa ggctagtcgg ttatcaactt gaaaaagtgg caccgagtcg gtgctttttt 420

tctagacca gctttcttgt acaagttgg catta 455

<210> 24

<211> 99

<212> RNA

<213> Artificial

<220><223> guide RNA

<220><221> misc\_feature

<222> (2)..(20)

<223> wherein N is G, A, C or U

<400> 24

gnnnnnnnnn nnnnnnnnnn guuuuagagc uagaaaagc aaguuaaag aaggguagug 60

gguuaucaac uugaaaaagc ggcaccgagu caauacuuu 99

<210> 25  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <400> 25  
 gtgaaccgca tcgagctgaa ggg 23  
 <210> 26  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <400> 26  
 ggagcgcacc atcttcttca agg 23  
 <210> 27  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <400> 27  
 gtccctcca cccacagtg ggg 23  
 <210> 28  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <400> 28  
 ggggccacta gggacaggat tgg 23  
 <210> 29  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <400> 29  
 gcatgatgcg cggcccaagg agg 23  
 <210> 30  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens



<400> 30  
gagatgatcg ccccttcttc tgg 23

<210> 31  
<211> 23  
<212> DNA  
<213> Homo sapiens  
<400> 31

gaattactca cgccccaagg agg 23

<210> 32  
<211> 72  
<212> DNA  
<213> Homo sapiens  
<400> 32

taatactttt atcttgtccc ctccacccca cagtggggcc actagggaca ggattggtga 60  
cagaaaagcc cc 72

<210> 33  
<211> 70  
<212> DNA  
<213> Homo sapiens  
<400> 33

taatactttt atctgtcaaa aaaacccac agtggggcca ctaggacagg attggtgaca 60  
gaaaagcccc 70

<210> 34  
<211> 71  
<212> DNA  
<213> Homo sapiens  
<400> 34

taatactttt atctgtcggg gggacccac agtggggcca ctagggacag gattggtgac 60  
agaaaagccc c 71

<210> 35  
<211> 61  
<212> DNA  
<213> Homo sapiens

<400> 35  
 taatgcatga tgcgcggccc aaggaggag atgacgccc cttcttctgg ctctttgaga 60  
 a 61  
 <210> 36  
 <211> 61  
 <212> DNA  
 <213> Homo sapiens  
 <400>  
 > 36  
 taatgaatta ctcacgcccc aaggagggtg atgaccggcc gttcttctgg atgtttgaga 60  
 a 61  
 <210> 37  
 <211> 110  
 <212> DNA  
 <213> Homo sapiens  
 <400> 37  
 ggtactttta tctgtccct ccacccaca gtggccact cgggacagga ttggtgacag 60  
 aaaagcccca tccttaggcc tcctccttcc tagtctctg atattgggtc 110  
 <210> 38  
 <211> 90  
 <212> DNA  
 <213> Homo sapiens  
 <400> 38  
 ttatctgtcc cctccacccc acagtggggc cactaggac aggaaagggtg acagaaaagc 60  
 cccatcctta ggcctcctcc ttctagtct 90  
 <210> 39  
 <211> 23  
 <212> DNA  
 <213> Homo sapiens  
 <220><221> misc\_feature  
 <222> (2)..(21)  
 <223> wherein N is G, A, T or C  
 <400> 39  
 gnnnnnnnnn nnnnnnnnnn ngg 23

<210> 40

<211> 201

<212> DNA

<213> Homo sapiens

<220><221> misc\_feature

<222> (27)..(51)

<223> wherein N is G, A, T or C

<220><221> misc\_feature

<222> (93)..(111)

<223> wherein N is G, A, T or C

<220><221> misc\_feature

<222> (152)..(176)

<223> wherein N is G, A, T or C

<400> 40

tatgaggacg aatctcccg cttatannnn nnnnnnnnnn nnnnnnnnn ntctttggct	60
ttatatatct tgtggaaagg acgaaaacac cgnnnnnnnn nnnnnnnnn ngtttttagag	120
ctagaaatag caagttaaaa taaggctagt cnnnnnnnnn nnnnnnnnn nnnnnngtac	180
aagcacacgt ttgtcaagac c	201

<210> 41

<211> 415

<212> DNA

<213> Homo sapiens

<400> 41

attcgccctt tglacaaaa agcaggcttt aaaggaacca attcagtcga ctggatccgg	60
taccaaggtc gggcaggaag agggcctatt tcccatgatt ccttcattt tcatatacga	120
tacaaggctg ttagagagat aataagaatt aatttgactg taaacacaaa gatattagta	180
caaaatagct gacgtagaaa gtaataattt ctgggtagt ttgcagtttt aaaattatgt	240
tttaaaatgg actatcatat gcttaccgta acttgaaagt atttcgattt ctggcttta	300
tatatcttaa gttaaaataa ggctagtccg ttatcaactt gaaaaagtgg caccgagtcg	360
gtgctttttt tctagacca gctttcttgt acaaagtgg cattaaaggg cgaat	415

<210> 42

<211> 23

<212> DNA

<213> Artificial

<220><223> reporter oligonucleotide

<400> 42

gtccccctcca cccacagtg ggg 23

<210> 43

<211> 23

<212> DNA

<213> Artificial

<220><223> reporter oligonucleotide

<400> 43

ggggccacta gggacaggat tgg 23

<210> 44

<211> 100

<212> DNA

<213> Artificial

<220><223> guide RNA

<400> 44

ggggccacta gggacaccat guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc 60

cguuaucaac uucaaaaaca cccacccaga cggugcuuuu 100

<210> 45

<211> 80

<212> DNA

<213> Artificial

<220><223> guide RNA scaffold

<400> 45

guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60

ggcaccgagu cggugcuuuu 80