



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년07월28일
(11) 등록번호 10-2282990
(24) 등록일자 2021년07월22일

- (51) 국제특허분류(Int. Cl.)
C12N 15/85 (2006.01) C12N 15/113 (2010.01)
C12N 15/81 (2006.01) C12N 15/82 (2006.01)
C12N 9/22 (2006.01)
- (52) CPC특허분류
C12N 15/85 (2013.01)
C12N 15/113 (2013.01)
- (21) 출원번호 10-2015-7036892
- (22) 출원일자(국제) 2014년06월04일
심사청구일자 2019년05월21일
- (85) 번역문제출일자 2015년12월29일
- (65) 공개번호 10-2016-0014036
- (43) 공개일자 2016년02월05일
- (86) 국제출원번호 PCT/US2014/040868
- (87) 국제공개번호 WO 2014/197568
국제공개일자 2014년12월11일
- (30) 우선권주장
61/830,787 2013년06월04일 미국(US)
- (56) 선행기술조사문헌
Cell., 152(5):1173-1183(2013.2.28.)*
SCIENCE., 339:823-826(2013.2.15.)*
Nat Methods., 10(3):239-242(2013.2.3.)*
SCIENCE., 337:816-821(2012.6.28.)*
*는 심사관에 의하여 인용된 문헌

- (73) 특허권자
프레지던트 앤드 펠로우즈 오브 하버드 칼리지
미합중국, 메사추세츠 02138, 캠브리지, 퀸시스트리트17
- (72) 발명자
처치, 조지 엠.
미국 02446 매사추세츠주 브루클린 켄트 스트리트 218
말리, 프라산트 지.
미국 02143 매사추세츠주 소머빌 비콘 스트리트 88
에스벨트, 케빈 엠.
미국 02138 매사추세츠주 캠브리지 아파트먼트 2
도우브리지 스트리트 20
- (74) 대리인
양영준, 김영

전체 청구항 수 : 총 48 항

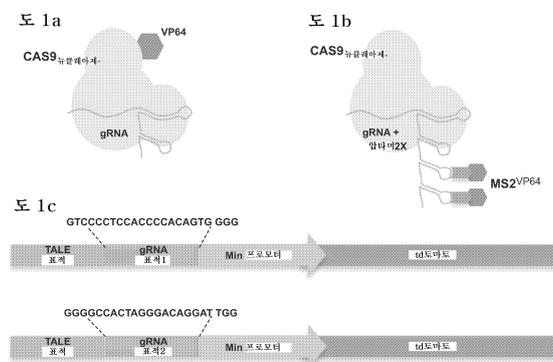
심사관 : 최성호

(54) 발명의 명칭 RNA-가이드된 전사 조절

(57) 요약

표적 핵산을 포함하는 DNA에 상보적인 1개 이상의 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, DNA에 결합하며 1개 이상의 RNA에 의해 가이드되는 뉴클레아제-기능부재 Cas9 단백질을 코딩하는 제2 외래 핵산을 세포 내로 도입하고, 전사 조절 단백질 또는 도메인을 코딩하는 제3 외래 핵산을 세포 내로 도입하는 것을 포함하며, (뒷면에 계속)

대표도



여기서 1개 이상의 RNA, 뉴클레아제-기능부재 Cas9 단백질 및 전사 조절 단백질 또는 도메인이 발현되고, 여기서 1개 이상의 RNA, 뉴클레아제-기능부재 Cas9 단백질 및 전사 조절 단백질 또는 도메인이 DNA로 공동-국제화되고, 여기서 전사 조절 단백질 또는 도메인이 표적 핵산의 발현을 조절하는 것인, 세포에서 표적 핵산의 발현을 조절하는 방법이 제공된다.

(52) CPC특허분류

C12N 15/81 (2013.01)

C12N 15/82 (2013.01)

C12N 9/22 (2013.01)

명세서

청구범위

청구항 1

표적 핵산을 포함하는 DNA에 상보적인 1개 이상의 가이드 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, 상기 DNA에 결합하며 상기 1개 이상의 가이드 RNA에 의해 가이드되는, 유형 II CRISPR 시스템의 뉴클레아제-기능부재 Cas 단백질을 코딩하는 제2 외래 핵산을 세포 내로 도입하고,

전사 조절 단백질 또는 도메인을 코딩하는 제3 외래 핵산을 세포 내로 도입하는 것

을 포함하며, 여기서 1개 이상의 가이드 RNA, 유형 II CRISPR 시스템의 뉴클레아제-기능부재 Cas 단백질 및 전사 조절 단백질 또는 도메인이 발현되고,

여기서 1개 이상의 가이드 RNA, 유형 II CRISPR 시스템의 뉴클레아제-기능부재 Cas 단백질 및 전사 조절 단백질 또는 도메인이 상기 DNA로 공동-국제화되고, 여기서 전사 조절 단백질 또는 도메인이 표적 핵산의 발현을 조절하는 것이고,

여기서 1개 이상의 RNA를 코딩하는 상기 제1 외래 핵산은 테더 서열을 추가로 코딩하고, 전사 조절 단백질 또는 도메인을 코딩하는 상기 제3 외래 핵산은 상기 전사 조절 단백질 또는 도메인에 융합된 테더 서열 결합 도메인을 추가로 코딩하는 것이고,

단, 인간 체내에서 수행되는 것이 아닌,

세포에서 표적 핵산의 발현을 조절하는 방법.

청구항 2

삭제

청구항 3

제1항에 있어서, 1개 이상의 가이드 RNA를 코딩하는 외래 핵산이 RNA-결합 도메인의 표적을 추가로 코딩하고, 전사 조절 단백질 또는 도메인을 코딩하는 외래 핵산이 전사 조절 단백질 또는 도메인에 융합된 RNA-결합 도메인을 추가로 코딩하는 것인 방법.

청구항 4

제1항에 있어서, 세포가 진핵 세포인 방법.

청구항 5

제1항에 있어서, 세포가 효모 세포, 식물 세포 또는 동물 세포인 방법.

청구항 6

제1항에 있어서, RNA가 10 내지 500개의 뉴클레오티드인 방법.

청구항 7

제1항에 있어서, RNA가 20 내지 100개의 뉴클레오티드인 방법.

청구항 8

제1항에 있어서, 전사 조절 단백질 또는 도메인이 전사 활성인자인 방법.

청구항 9

제1항에 있어서, 전사 조절 단백질 또는 도메인이 표적 핵산의 발현을 상향조절하는 것인 방법.

청구항 10

제1항에 있어서, 전사 조절 단백질 또는 도메인이 질환 또는 유해 상태를 치료하기 위해 표적 핵산의 발현을 상향조절하고, 인간이 아닌 대상(non-human subjects)에 한정되는 방법.

청구항 11

제1항에 있어서, 표적 핵산이 질환 또는 유해 상태와 연관된 것이고, 인간이 아닌 대상에 한정되는 방법.

청구항 12

제1항에 있어서, 1개 이상의 가이드 RNA가 crRNA-tracrRNA 융합체인 방법.

청구항 13

제1항에 있어서, DNA가 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA인 방법.

청구항 14

각각의 가이드 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 가이드 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고,

1개의 불활성 뉴클레아제 도메인을 가지며 상기 2개 이상의 가이드 RNA에 의해 가이드되는, 유형 II CRISPR 시스템의 1개 이상의 Cas 단백질 니카제를 코딩하는 제2 외래 핵산을 세포 내로 도입하는 것을 포함하며,

여기서 2개 이상의 가이드 RNA 및 유형 II CRISPR 시스템의 1개 이상의 Cas 단백질 니카제가 발현되고, 여기서 유형 II CRISPR 시스템의 1개 이상의 Cas 단백질 니카제가 2개 이상의 가이드 RNA와 함께 DNA 표적 핵산으로 공동-국재화되고, DNA 표적 핵산을 니킹하여 2개 이상의 인접 닉을 생성하는 것이고,

단, 인간 체내에서 수행되는 것이 아닌,

세포에서 DNA 표적 핵산을 변경시키는 방법.

청구항 15

제14항에 있어서, 2개 이상의 인접 닉이 이중 가닥 DNA의 동일 가닥 상에 있는 것인 방법.

청구항 16

제14항에 있어서, 2개 이상의 인접 닉이 이중 가닥 DNA의 동일 가닥 상에 있고, 상동 재조합을 일으키는 것인 방법.

청구항 17

제14항에 있어서, 2개 이상의 인접 닉이 이중 가닥 DNA의 상이한 가닥 상에 있는 것인 방법.

청구항 18

제14항에 있어서, 2개 이상의 인접 닉이 이중 가닥 DNA의 상이한 가닥 상에 있고, 이중 가닥 파괴를 생성하는 것인 방법.

청구항 19

제14항에 있어서, 2개 이상의 인접 닉이 이중 가닥 DNA의 상이한 가닥 상에 있고, 비상동 말단 연결을 일으키는 이중 가닥 파괴를 생성하는 것인 방법.

청구항 20

제14항에 있어서, 2개 이상의 인접 닉이 이중 가닥 DNA의 상이한 가닥 상에 있고, 서로에 대해 오프셋되는 것인 방법.

청구항 21

제14항에 있어서, 2개 이상의 인접 닉이 이중 가닥 DNA의 상이한 가닥 상에 있고, 서로에 대해 오프셋되고, 이중 가닥 파괴를 생성하는 것인 방법.

청구항 22

제14항에 있어서, 2개 이상의 인접 닉이 이중 가닥 DNA의 상이한 가닥 상에 있고, 서로에 대해 오프셋되고, 비상동 말단 연결을 일으키는 이중 가닥 파괴를 생성하는 것인 방법.

청구항 23

제14항에 있어서, 공여자 핵산 서열을 코딩하는 제3 외래 핵산을 세포 내로 도입하는 것을 추가로 포함하며, 여기서 2개 이상의 닉이 표적 핵산과 공여자 핵산 서열의 상동 재조합을 일으키는 것인 방법.

청구항 24

표적 핵산을 포함하는 DNA에 상보적인 1개 이상의 가이드 RNA를 코딩하는 제1 외래 핵산, 유형 II CRISPR 시스템의 뉴클레아제-기능부재 Cas 단백질을 코딩하는 제2 외래 핵산, 및 전사 조절 단백질 또는 도메인을 코딩하는 제3 외래 핵산을 포함하며,

여기서 1개 이상의 가이드 RNA, 유형 II CRISPR 시스템의 뉴클레아제-기능부재 Cas 단백질 및 전사 조절 단백질 또는 도메인이 표적 핵산에 대한 공동-국제화 복합체의 구성원이고, 여기서 1개 이상의 RNA를 코딩하는 상기 제1 외래 핵산은 테더 서열을 추가로 코딩하고, 전사 조절 단백질 또는 도메인을 코딩하는 상기 제3 외래 핵산은 상기 전사 조절 단백질 또는 도메인에 융합된 테더 서열 결합 도메인을 추가로 코딩하는 것인,

단리된 세포.

청구항 25

삭제

청구항 26

제24항에 있어서, 1개 이상의 가이드 RNA를 코딩하는 외래 핵산이 RNA-결합 도메인의 표적을 추가로 코딩하고, 전사 조절 단백질 또는 도메인을 코딩하는 외래 핵산이 전사 조절 단백질 또는 도메인에 융합된 RNA-결합 도메인을 추가로 코딩하는 것인 세포.

청구항 27

제24항에 있어서, 진핵 세포인 세포.

청구항 28

제24항에 있어서, 효모 세포, 식물 세포 또는 동물 세포인 세포.

청구항 29

제24항에 있어서, 가이드 RNA가 10 내지 500개의 뉴클레오티드를 포함하는 것인 세포.

청구항 30

제24항에 있어서, 가이드 RNA가 20 내지 100개의 뉴클레오티드를 포함하는 것인 세포.

청구항 31

제24항에 있어서, 전사 조절 단백질 또는 도메인이 전사 활성인자인 세포.

청구항 32

제24항에 있어서, 전사 조절 단백질 또는 도메인이 표적 핵산의 발현을 상향조절하는 것인 세포.

청구항 33

제24항에 있어서, 전사 조절 단백질 또는 도메인이 질환 또는 유해 상태를 치료하기 위해 표적 핵산의 발현을 상향조절하는 것인 세포.

청구항 34

제24항에 있어서, 표적 핵산이 질환 또는 유해 상태와 연관된 것인 세포.

청구항 35

제24항에 있어서, 1개 이상의 가이드 RNA가 crRNA-tracrRNA 융합체인 세포.

청구항 36

제24항에 있어서, DNA가 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA인 세포.

청구항 37

각각의 가이드 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 가이드 RNA를 코딩하는 제1 외래 핵산, 및

1개의 불활성 뉴클레아제 도메인을 갖는, 유형 II CRISPR 시스템의 1개 이상의 Cas 단백질 니카제를 코딩하는 제2 외래 핵산을 포함하며, 여기서 2개 이상의 가이드 RNA 및 유형 II CRISPR 시스템의 1개 이상의 Cas 단백질 니카제가 DNA 표적 핵산에 대한 공동-국제화 복합체의 구성원인

단리된 세포.

청구항 38

제37항에 있어서, 진핵 세포인 세포.

청구항 39

제37항에 있어서, 효모 세포, 식물 세포 또는 동물 세포인 세포.

청구항 40

제37항에 있어서, 가이드 RNA가 10 내지 500개의 뉴클레오티드를 포함하는 것인 세포.

청구항 41

제37항에 있어서, 가이드 RNA가 20 내지 100개의 뉴클레오티드를 포함하는 것인 세포.

청구항 42

제37항에 있어서, 표적 핵산이 질환 또는 유해 상태와 연관된 것인 세포.

청구항 43

제37항에 있어서, 2개 이상의 가이드 RNA가 crRNA-tracrRNA 융합체인 세포.

청구항 44

제37항에 있어서, DNA 표적 핵산이 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA인 세포.

청구항 45

각각의 가이드 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 가이드 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고,

1개의 불활성 뉴클레아제 도메인을 가지며 상기 2개 이상의 가이드 RNA에 의해 가이드되는, 유형 II CRISPR 시스템의 1개 이상의 Cas 단백질 니카제를 코딩하는 제2 외래 핵산을 세포 내로 도입하는 것을 포함하며,

여기서 2개 이상의 가이드 RNA 및 유형 II CRISPR 시스템의 1개 이상의 Cas 단백질 니카제가 발현되고, 여기서

유형 II CRISPR 시스템의 1개 이상의 Cas 단백질 니카제가 2개 이상의 가이드 RNA와 함께 DNA 표적 핵산으로 공동-국재화되고, DNA 표적 핵산을 니킹하여 2개 이상의 인접 니클 생성하고,

여기서 2개 이상의 인접 니클이 이중 가닥 DNA의 상이한 가닥 상에 있고, 표적 핵산의 단편화를 일으키는 이중 가닥 파괴를 생성하여 표적 핵산의 발현을 방지하는 것이고,

단, 인간 체내에서 수행되는 것이 아닌,

세포에서 DNA 표적 핵산을 변경시키는 방법.

청구항 46

제1항에 있어서, 유형 II CRISPR 시스템의 뉴클레아제-기능부재 Cas 단백질이 뉴클레아제-기능부재 Cas9 단백질인 방법.

청구항 47

제14항에 있어서, 유형 II CRISPR 시스템의 Cas 단백질 니카제가 Cas9 단백질 니카제인 방법.

청구항 48

제24항에 있어서, 유형 II CRISPR 시스템의 뉴클레아제-기능부재 Cas 단백질이 뉴클레아제-기능부재 Cas9 단백질인 세포.

청구항 49

제37항에 있어서, 유형 II CRISPR 시스템의 Cas 단백질 니카제가 Cas9 단백질 니카제인 세포.

청구항 50

제45항에 있어서, 유형 II CRISPR 시스템의 Cas 단백질 니카제가 Cas9 단백질 니카제인 방법.

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

- 청구항 75
- 삭제
- 청구항 76
- 삭제
- 청구항 77
- 삭제
- 청구항 78
- 삭제
- 청구항 79
- 삭제
- 청구항 80
- 삭제
- 청구항 81
- 삭제
- 청구항 82
- 삭제
- 청구항 83
- 삭제
- 청구항 84
- 삭제
- 청구항 85
- 삭제
- 청구항 86
- 삭제
- 청구항 87
- 삭제
- 청구항 88
- 삭제
- 청구항 89
- 삭제
- 청구항 90
- 삭제

청구항 91

삭제

청구항 92

삭제

청구항 93

삭제

청구항 94

삭제

청구항 95

삭제

청구항 96

삭제

청구항 97

삭제

청구항 98

삭제

청구항 99

삭제

청구항 100

삭제

청구항 101

삭제

청구항 102

삭제

청구항 103

삭제

청구항 104

삭제

청구항 105

삭제

청구항 106

삭제

청구항 107

삭제

청구항 108

삭제

청구항 109

삭제

청구항 110

삭제

청구항 111

삭제

청구항 112

삭제

발명의 설명

기술 분야

[0001] 관련 출원 데이터

[0002] 본원은 2013년 6월 4일에 출원된 미국 특허 가출원 번호 61/830,787을 우선권 주장하며, 이로써 상기 문헌은 모든 목적을 위해 그 전문이 본원에 참조로 포함된다.

[0003] 정부 권리의 진술

[0004] 본 발명은 미국 국립 보건원으로부터의 승인 번호 P50 HG005550 및 미국 에너지국으로부터의 DE-FG02-02ER63445 하에 정부 지원으로 이루어졌다. 정부는 본 발명에서 특정 권리를 갖는다.

배경 기술

[0005] 박테리아성 및 고세균성 CRISPR-Cas 시스템은 침입 외래 핵산 내에 존재하는 상보적 서열의 분해를 지시하는 Cas 단백질과 복합체화된 짧은 가이드 RNA에 의존한다. 문헌 [Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607 (2011); Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. 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, E2579-2586 (2012); Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821 (2012); Sapranaukas, R. et al. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. Nucleic acids research 39, 9275-9282 (2011); 및 Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annual review of genetics 45, 273-297 (2011)]을 참조한다. 최근, 에스. 피오게네스(S. pyogenes) 유형 II CRISPR 시스템을 시험관내에서 재구성한 결과, 정상적으로 트랜스-코딩된 tracrRNA ("트랜스-활성화 CRISPR RNA")와 융합된 crRNA ("CRISPR RNA")가, crRNA와 매칭되는 표적 DNA 서열을 서열-특이적으로 절단하도록 Cas9 단백질을 지시하는데 충분한 것으로 입증되었다. 표적 부위에 상동성인 gRNA의 발현은 Cas9 동원 및 표적 DNA의 분해를 유도한다. 문헌 [H. Deveau et al., Phage response to CRISPR-encoded resistance in Streptococcus thermophilus. Journal of Bacteriology 190, 1390 (Feb, 2008)]을 참조한다.

발명의 내용

[0006] 본 개시내용의 측면은 가이드 RNA, DNA 결합 단백질 및 이중 가닥 DNA 표적 서열의 복합체에 관한 것이다. 특

정 측면에 따르면, 본 개시내용의 범위 내의 DNA 결합 단백질은, 가이드 RNA와 복합체를 형성하는 단백질을 포함하는데, 여기서 가이드 RNA는 상기 복합체를 이중 가닥 DNA 서열로 가이드하고 여기서 상기 복합체가 DNA 서열에 결합한다. 본 개시내용의 이러한 측면은 이중 가닥 DNA로의 또는 그와의 RNA 및 DNA 결합 단백질의 공동-국제화로서 지칭될 수 있다. 이러한 방식으로, DNA 결합 단백질-가이드 RNA 복합체는 표적 DNA의 발현을 조절하기 위해 표적 DNA에 전사 조절 단백질 또는 도메인을 국제화시키는데 사용될 수 있다.

- [0007] 특정 측면에 따르면, 표적 핵산을 포함하는 DNA (데옥시리보핵산)에 상보적인 1개 이상의 RNA (리보핵산)를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, DNA에 결합하며 1개 이상의 RNA에 의해 가이드되는, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질을 코딩하는 제2 외래 핵산을 세포 내로 도입하고, 전사 조절 단백질 또는 도메인을 코딩하는 제3 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 1개 이상의 RNA, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질 및 전사 조절 단백질 또는 도메인이 발현되고, 여기서 1개 이상의 RNA, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질 및 전사 조절 단백질 또는 도메인이 DNA로 공동-국제화되고, 여기서 전사 조절 단백질 또는 도메인이 표적 핵산의 발현을 조절하는 것인, 세포에서 표적 핵산의 발현을 조절하는 방법이 제공된다.
- [0008] 한 측면에 따르면, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질을 코딩하는 외래 핵산은 RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질에 융합된 전사 조절 단백질 또는 도메인을 추가로 코딩한다. 한 측면에 따르면, 1개 이상의 RNA를 코딩하는 외래 핵산은 RNA-결합 도메인의 표적을 추가로 코딩하고, 전사 조절 단백질 또는 도메인을 코딩하는 외래 핵산은 전사 조절 단백질 또는 도메인에 융합된 RNA-결합 도메인을 추가로 코딩한다.
- [0009] 한 측면에 따르면, 세포는 진핵 세포이다. 한 측면에 따르면, 세포는 효모 세포, 식물 세포 또는 동물 세포이다. 한 측면에 따르면, 세포는 포유동물 세포이다.
- [0010] 한 측면에 따르면, RNA는 약 10 내지 약 500개의 뉴클레오티드이다. 한 측면에 따르면, RNA는 약 20 내지 약 100개의 뉴클레오티드이다.
- [0011] 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 전사 활성인자이다. 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 표적 핵산의 발현을 상향조절한다. 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 질환 또는 유해 상태를 치료하기 위해 표적 핵산의 발현을 상향조절한다. 한 측면에 따르면, 표적 핵산은 질환 또는 유해 상태와 연관된다.
- [0012] 한 측면에 따르면, 1개 이상의 RNA는 가이드 RNA이다. 한 측면에 따르면, 1개 이상의 RNA는 tracrRNA-crRNA 융합체이다. 한 측면에 따르면, 가이드 RNA는 스페이서 서열 및 트레이서 메이트 서열을 포함한다. 가이드 RNA는 또한, 일부가 tracr 메이트 서열에 혼성화되는 tracr 서열을 포함할 수 있다. 가이드 RNA는 또한, tracrRNA-crRNA 융합체를 생산하기 위해 트레이서 메이트 서열 및 tracr 서열을 연결하는 링커 핵산 서열을 포함할 수 있다. 스페이서 서열은, 예컨대 혼성화에 의해 표적 DNA에 결합한다.
- [0013] 한 측면에 따르면, 가이드 RNA는 말단절단된 스페이서 서열을 포함한다. 한 측면에 따르면, 가이드 RNA는 스페이서 서열의 5' 말단에 1개의 염기 말단절단을 갖는 말단절단된 스페이서 서열을 포함한다. 한 측면에 따르면, 가이드 RNA는 스페이서 서열의 5' 말단에 2개의 염기 말단절단을 갖는 말단절단된 스페이서 서열을 포함한다. 한 측면에 따르면, 가이드 RNA는 스페이서 서열의 5' 말단에 3개의 염기 말단절단을 갖는 말단절단된 스페이서 서열을 포함한다. 한 측면에 따르면, 가이드 RNA는 스페이서 서열의 5' 말단에 4개의 염기 말단절단을 갖는 말단절단된 스페이서 서열을 포함한다. 따라서, 스페이서 서열은 스페이서 서열의 5' 말단에 1 내지 4개의 염기 말단절단을 가질 수 있다.
- [0014] 특정 실시양태에 따르면, 스페이서 서열은 표적 핵산 서열에 혼성화되는 약 16 내지 약 20개의 뉴클레오티드를 포함할 수 있다. 특정 실시양태에 따르면, 스페이서 서열은 표적 핵산 서열에 혼성화되는 약 20개의 뉴클레오티드를 포함할 수 있다.
- [0015] 특정 측면에 따르면, 링커 핵산 서열은 약 4 내지 약 6개의 핵산을 포함할 수 있다.
- [0016] 특정 측면에 따르면, tracr 서열은 약 60 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 64 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 65 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 66 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 67 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 68 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 69 내지

약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 70 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 80 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 90 내지 약 500개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 100 내지 약 500개의 핵산을 포함할 수 있다.

[0017] 특정 측면에 따르면, tracr 서열은 약 60 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 64 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 65 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 66 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 67 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 68 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 69 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 70 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 80 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 90 내지 약 200개의 핵산을 포함할 수 있다. 특정 측면에 따르면, tracr 서열은 약 100 내지 약 200개의 핵산을 포함할 수 있다.

[0018] 예시적인 가이드 RNA가 도 5b에 도시된다.

[0019] 한 측면에 따르면, DNA는 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA이다.

[0020] 특정 측면에 따르면, 표적 핵산을 포함하는 DNA (데옥시리보핵산)에 상보적인 1개 이상의 RNA (리보핵산)를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, DNA에 결합하며 1개 이상의 RNA에 의해 가이드되는 유형 II CRISPR 시스템의 RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질을 코딩하는 제2 외래 핵산을 세포 내로 도입하고, 전사 조절 단백질 또는 도메인을 코딩하는 제3 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 1개 이상의 RNA, 유형 II CRISPR 시스템의 RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질 및 전사 조절 단백질 또는 도메인이 발현되고, 여기서 1개 이상의 RNA, 유형 II CRISPR 시스템의 RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질 및 전사 조절 단백질 또는 도메인이 DNA로 공동-국제화되고, 여기서 전사 조절 단백질 또는 도메인이 표적 핵산의 발현을 조절하는 것인, 세포에서 표적 핵산의 발현을 조절하는 방법이 제공된다.

[0021] 한 측면에 따르면, 유형 II CRISPR 시스템의 RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질을 코딩하는 외래 핵산은 유형 II CRISPR 시스템의 RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질에 융합된 전사 조절 단백질 또는 도메인을 추가로 코딩한다. 한 측면에 따르면, 1개 이상의 RNA를 코딩하는 외래 핵산은 RNA-결합 도메인의 표적을 추가로 코딩하고, 전사 조절 단백질 또는 도메인을 코딩하는 외래 핵산은 전사 조절 단백질 또는 도메인에 융합된 RNA-결합 도메인을 추가로 코딩한다.

[0022] 한 측면에 따르면, 세포는 진핵 세포이다. 한 측면에 따르면, 세포는 효모 세포, 식물 세포 또는 동물 세포이다. 한 측면에 따르면, 세포는 포유동물 세포이다.

[0023] 한 측면에 따르면, RNA는 약 10 내지 약 500개의 뉴클레오티드이다. 한 측면에 따르면, RNA는 약 20 내지 약 100개의 뉴클레오티드이다.

[0024] 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 전사 활성인자이다. 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 표적 핵산의 발현을 상향조절한다. 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 질환 또는 유해 상태를 치료하기 위해 표적 핵산의 발현을 상향조절한다. 한 측면에 따르면, 표적 핵산은 질환 또는 유해 상태와 연관된다.

[0025] 한 측면에 따르면, 1개 이상의 RNA는 가이드 RNA이다. 한 측면에 따르면, 1개 이상의 RNA는 tracrRNA-crRNA 융합체이다.

[0026] 한 측면에 따르면, DNA는 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA이다.

[0027] 특정 측면에 따르면, 표적 핵산을 포함하는 DNA (데옥시리보핵산)에 상보적인 1개 이상의 RNA (리보핵산)를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, DNA에 결합하며 1개 이상의 RNA에 의해 가이드되는 뉴클레아제-기능부재 Cas9 단백질을 코딩하는 제2 외래 핵산을 세포 내로 도입하고, 전사 조절 단백질 또는 도메인을 코딩하는 제3 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 1개 이상의 RNA, 뉴클레아제-기능부재 Cas9 단백질 및 전사 조절 단백질 또는 도메인이 발현되고, 여기서 1개 이상의 RNA, 뉴클레아제-기능부재 Cas9 단백질 및 전사 조절 단백질 또는 도메인이 DNA로 공동-국제화되고, 여기서 전사 조절 단백질 또는 도메인이 표적 핵산의 발현을 조절하는 것인, 세포에서 표적 핵산의 발현을 조절하는 방법이 제공된다.

- [0028] 한 측면에 따르면, 뉴클레아제-기능부재 Cas9 단백질을 코딩하는 외래 핵산은 뉴클레아제-기능부재 Cas9 단백질에 융합된 전사 조절 단백질 또는 도메인을 추가로 코딩한다. 한 측면에 따르면, 1개 이상의 RNA를 코딩하는 외래 핵산은 RNA-결합 도메인의 표적을 추가로 코딩하고, 전사 조절 단백질 또는 도메인을 코딩하는 외래 핵산은 전사 조절 단백질 또는 도메인에 융합된 RNA-결합 도메인을 추가로 코딩한다.
- [0029] 한 측면에 따르면, 세포는 진핵 세포이다. 한 측면에 따르면, 세포는 효모 세포, 식물 세포 또는 동물 세포이다. 한 측면에 따르면, 세포는 포유동물 세포이다.
- [0030] 한 측면에 따르면, RNA는 약 10 내지 약 500개의 뉴클레오티드이다. 한 측면에 따르면, RNA는 약 20 내지 약 100개의 뉴클레오티드이다.
- [0031] 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 전사 활성인자이다. 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 표적 핵산의 발현을 상향조절한다. 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 질환 또는 유해 상태를 치료하기 위해 표적 핵산의 발현을 상향조절한다. 한 측면에 따르면, 표적 핵산은 질환 또는 유해 상태와 연관된다.
- [0032] 한 측면에 따르면, 1개 이상의 RNA는 가이드 RNA이다. 한 측면에 따르면, 1개 이상의 RNA는 tracrRNA-crRNA 융합체이다.
- [0033] 한 측면에 따르면, DNA는 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA이다.
- [0034] 한 측면에 따르면, 표적 핵산을 포함하는 DNA에 상보적인 1개 이상의 RNA를 코딩하는 제1 외래 핵산, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질을 코딩하는 제2 외래 핵산, 및 전사 조절 단백질 또는 도메인을 코딩하는 제3 외래 핵산을 포함하며, 여기서 1개 이상의 RNA, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질 및 전사 조절 단백질 또는 도메인이 표적 핵산에 대한 공동-국제화 복합체의 구성원인 세포가 제공된다.
- [0035] 한 측면에 따르면, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질을 코딩하는 외래 핵산은 RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질에 융합된 전사 조절 단백질 또는 도메인을 추가로 코딩한다. 한 측면에 따르면, 1개 이상의 RNA를 코딩하는 외래 핵산은 RNA-결합 도메인의 표적을 추가로 코딩하고, 전사 조절 단백질 또는 도메인을 코딩하는 외래 핵산은 전사 조절 단백질 또는 도메인에 융합된 RNA-결합 도메인을 추가로 코딩한다.
- [0036] 한 측면에 따르면, 세포는 진핵 세포이다. 한 측면에 따르면, 세포는 효모 세포, 식물 세포 또는 동물 세포이다. 한 측면에 따르면, 세포는 포유동물 세포이다.
- [0037] 한 측면에 따르면, RNA는 약 10 내지 약 500개의 뉴클레오티드이다. 한 측면에 따르면, RNA는 약 20 내지 약 100개의 뉴클레오티드이다.
- [0038] 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 전사 활성인자이다. 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 표적 핵산의 발현을 상향조절한다. 한 측면에 따르면, 전사 조절 단백질 또는 도메인은 질환 또는 유해 상태를 치료하기 위해 표적 핵산의 발현을 상향조절한다. 한 측면에 따르면, 표적 핵산은 질환 또는 유해 상태와 연관된다.
- [0039] 한 측면에 따르면, 1개 이상의 RNA는 가이드 RNA이다. 한 측면에 따르면, 1개 이상의 RNA는 tracrRNA-crRNA 융합체이다.
- [0040] 한 측면에 따르면, DNA는 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA이다.
- [0041] 특정 측면에 따르면, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질은 유형 II CRISPR 시스템의 RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질이다. 특정 측면에 따르면, RNA 가이드된 뉴클레아제-기능부재 DNA 결합 단백질은 뉴클레아제-기능부재 Cas9 단백질이다.
- [0042] 한 측면에 따르면, 각각의 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, 2개 이상의 RNA에 의해 가이드되는 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제를 코딩하는 제2 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 2개 이상의 RNA 및 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 발현되고, 여기서 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 2개 이상의 RNA와 함께 DNA 표적 핵산으로 공동-국제화되고, DNA 표적 핵산을 니킹하여 2개 이상의 인접 닉을 생성하는 것인, 세포에서 DNA 표적 핵산을 변경시키는 방법이 제공된다.

- [0043] 한 측면에 따르면, 각각의 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, 2개 이상의 RNA에 의해 가이드되는 유형 II CRISPR 시스템의 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제를 코딩하는 제2 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 2개 이상의 RNA 및 유형 II CRISPR 시스템의 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 발현되고, 여기서 유형 II CRISPR 시스템의 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 2개 이상의 RNA와 함께 DNA 표적 핵산으로 공동-국제화되고, DNA 표적 핵산을 니킹하여 2개 이상의 인접 틈을 생성하는 것인, 세포에서 DNA 표적 핵산을 변경시키는 방법이 제공된다.
- [0044] 한 측면에 따르면, 각각의 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, 1개의 불활성 뉴클레아제 도메인을 가지며 2개 이상의 RNA에 의해 가이드되는 1개 이상의 Cas9 단백질 니카제를 코딩하는 제2 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 2개 이상의 RNA 및 1개 이상의 Cas9 단백질 니카제가 발현되고, 여기서 1개 이상의 Cas9 단백질 니카제가 2개 이상의 RNA와 함께 DNA 표적 핵산으로 공동-국제화되고, DNA 표적 핵산을 니킹하여 2개 이상의 인접 틈을 생성하는 것인, 세포에서 DNA 표적 핵산을 변경시키는 방법이 제공된다.
- [0045] DNA 표적 핵산을 변경시키는 방법에 따르면, 2개 이상의 인접 틈은 이중 가닥 DNA의 동일한 가닥 상에 있다. 한 측면에 따르면, 2개 이상의 인접 틈은 이중 가닥 DNA의 동일한 가닥 상에 있고, 상동 재조합을 일으킨다. 한 측면에 따르면, 2개 이상의 인접 틈은 이중 가닥 DNA의 상이한 가닥 상에 있다. 한 측면에 따르면, 2개 이상의 인접 틈은 이중 가닥 DNA의 상이한 가닥 상에 있고, 이중 가닥 파괴를 생성한다. 한 측면에 따르면, 2개 이상의 인접 틈은 이중 가닥 DNA의 상이한 가닥 상에 있고, 비상동 말단 연결을 일으키는 이중 가닥 파괴를 생성한다. 한 측면에 따르면, 2개 이상의 인접 틈은 이중 가닥 DNA의 상이한 가닥 상에 있고, 서로에 대해 오프셋된다. 한 측면에 따르면, 2개 이상의 인접 틈은 이중 가닥 DNA의 상이한 가닥 상에 있고, 서로에 대해 오프셋되고, 이중 가닥 파괴를 생성한다. 한 측면에 따르면, 2개 이상의 인접 틈은 이중 가닥 DNA의 상이한 가닥 상에 있고, 서로에 대해 오프셋되고, 비상동 말단 연결을 일으키는 이중 가닥 파괴를 생성한다. 한 측면에 따르면, 방법은 공여자 핵산 서열을 코딩하는 제3 외래 핵산을 세포 내로 도입하는 것을 추가로 포함하며, 여기서 2개 이상의 틈은 표적 핵산과 공여자 핵산 서열의 상동 재조합을 일으킨다.
- [0046] 한 측면에 따르면, 각각의 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, 2개 이상의 RNA에 의해 가이드되는 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제를 코딩하는 제2 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 2개 이상의 RNA 및 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 발현되고, 여기서 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 2개 이상의 RNA와 함께 DNA 표적 핵산으로 공동-국제화되고, DNA 표적 핵산을 니킹하여 2개 이상의 인접 틈을 생성하고, 여기서 2개 이상의 인접 틈이 이중 가닥 DNA의 상이한 가닥 상에 있고, 표적 핵산의 단편화를 일으키는 이중 가닥 파괴를 생성하여 표적 핵산의 발현을 방지하는 것인, 세포에서 DNA 표적 핵산을 변경시키는 방법이 제공된다.
- [0047] 한 측면에 따르면, 각각의 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, 2개 이상의 RNA에 의해 가이드되는 유형 II CRISPR 시스템의 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제를 코딩하는 제2 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 2개 이상의 RNA 및 유형 II CRISPR 시스템의 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 발현되고, 여기서 유형 II CRISPR 시스템의 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 2개 이상의 RNA와 함께 DNA 표적 핵산으로 공동-국제화되고, DNA 표적 핵산을 니킹하여 2개 이상의 인접 틈을 생성하고, 여기서 2개 이상의 인접 틈이 이중 가닥 DNA의 상이한 가닥 상에 있고, 표적 핵산의 단편화를 일으키는 이중 가닥 파괴를 생성하여 표적 핵산의 발현을 방지하는 것인, 세포에서 DNA 표적 핵산을 변경시키는 방법이 제공된다.
- [0048] 한 측면에 따르면, 각각의 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 RNA를 코딩하는 제1 외래 핵산을 세포 내로 도입하고, 1개의 불활성 뉴클레아제 도메인을 가지며 2개 이상의 RNA에 의해 가이드되는 1개 이상의 Cas9 단백질 니카제를 코딩하는 제2 외래 핵산을 세포 내로 도입하는 것을 포함하며, 여기서 2개 이상의 RNA 및 1개 이상의 Cas9 단백질 니카제가 발현되고, 여기서 1개 이상의 Cas9 단백질 니카제가 2개 이상의 RNA와 함께 DNA 표적 핵산으로 공동-국제화되고, DNA 표적 핵산을 니킹하여 2개 이상의 인접 틈을 생성하고, 여기서 2개 이상의 인접 틈이 이중 가닥 DNA의 상이한 가닥 상에 있고, 표적 핵산의 단편화를 일으키는 이중 가닥 파괴를 생성하여 표적 핵산의 발현을 방지하는 것인, 세포에서 DNA 표적 핵산을 변경시키는 방법이 제공된다.
- [0049] 한 측면에 따르면, 각각의 RNA가 DNA 표적 핵산 내의 인접 부위에 상보적인 2개 이상의 RNA를 코딩하는 제1 외

래 핵산, 및 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제를 코딩하는 제2 외래 핵산을 포함하며, 여기서 2개 이상의 RNA 및 1개 이상의 RNA 가이드된 DNA 결합 단백질 니카제가 DNA 표적 핵산에 대한 공동-국제화 복합체의 구성원인 세포가 제공된다.

- [0050] 한 측면에 따르면, RNA 가이드된 DNA 결합 단백질 니카제는 유형 II CRISPR 시스템의 RNA 가이드된 DNA 결합 단백질 니카제이다. 한 측면에 따르면, RNA 가이드된 DNA 결합 단백질 니카제는 1개의 불활성 뉴클레아제 도메인을 갖는 Cas9 단백질 니카제이다.
- [0051] 한 측면에 따르면, 세포는 진핵 세포이다. 한 측면에 따르면, 세포는 효모 세포, 식물 세포 또는 동물 세포이다. 한 측면에 따르면, 세포는 포유동물 세포이다.
- [0052] 한 측면에 따르면, RNA는 약 10 내지 약 500개의 뉴클레오티드를 포함한다. 한 측면에 따르면, RNA는 약 20 내지 약 100개의 뉴클레오티드를 포함한다.
- [0053] 한 측면에 따르면, 표적 핵산은 질환 또는 유해 상태와 연관된다.
- [0054] 한 측면에 따르면, 2개 이상의 RNA는 가이드 RNA이다. 한 측면에 따르면, 2개 이상의 RNA는 tracrRNA-crRNA 융합체이다.
- [0055] 한 측면에 따르면, DNA 표적 핵산은 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA이다.
- [0056] 본 발명의 특정 실시양태의 추가의 특징 및 이점은 하기 실시양태의 설명 및 그의 도면, 및 청구범위로부터 보다 충분히 명백해질 것이다.

도면의 간단한 설명

- [0057] 특허 또는 출원 파일은 컬러로 제작된 도면을 포함한다. 요청 및 필요한 요금의 지불시 사무국으로부터 컬러 도면을 포함하는 본 특허 또는 특허 출원 공개의 사본이 제공될 것이다. 본 발명의 상기 및 다른 특징 및 이점은 첨부된 도면과 함께 하기 예시적 실시양태의 상세한 설명으로부터 보다 충분히 이해될 것이다.

도 1a 및 도 1b는 RNA-가이드된 전사 활성화의 개략도이다. 도 1c는 리포터 구축물의 설계이다. 도 1d는 Cas9N-VP64 융합체가 형광-활성화 세포 분류 (FACS) 및 면역형광 검정 (IF) 둘 다에 의해 검정시에 RNA-가이드된 전사 활성화를 나타낸다는 것을 입증하는 데이터를 보여준다. 도 1e는 Cas9N, MS2-VP64, 및 적절한 MS2 압타머 결합 부위를 보유하는 gRNA의 존재 하에 리포터 구축물로부터의 gRNA 서열-특이적 전사 활성화를 입증하는, FACS 및 IF에 의한 검정 데이터를 보여준다. 도 1f는 개별 gRNA 및 다중 gRNA에 의한 전사 유도를 입증하는 데이터를 도시한다.

도 2a는 Cas9-gRNA 복합체 및 TALE에 의한 표적화의 랜드스케이프를 평가하기 위한 방법론을 도시한다. 도 2b는 Cas9-gRNA 복합체가 평균적으로 그의 표적 서열에서 1-3개의 돌연변이에 내성이 있다는 것을 입증하는 데이터를 도시한다. 도 2c는 Cas9-gRNA 복합체가, PAM 서열에 국제화된 것들을 제외하고는 주로 점 돌연변이에 비감수성이라는 것을 입증하는 데이터를 도시한다. 도 2d는 2개의 염기 미스매치의 도입이 Cas9-gRNA 복합체 활성을 유의하게 손상시킨다는 것을 입증하는 열 플롯 데이터를 도시한다. 도 2e는 18-량체 TALE가 평균적으로 그의 표적 서열에서 1-2개의 돌연변이에 내성을 나타낸다는 것을 입증하는 데이터를 도시한다. 도 2f는 18-량체 TALE가, 그의 표적에서 미스매치된 단일 염기에 주로 비감수성인 Cas9-gRNA 복합체와 유사하다는 것을 입증하는 데이터를 도시한다. 도 2g는 2개의 염기 미스매치의 도입이 18-량체 TALE 활성을 유의하게 손상시킨다는 것을 입증하는 열 플롯 데이터를 도시한다.

도 3a는 가이드 RNA 설계의 개략도를 도시한다. 도 3b는 5' 오버행으로 이어지는 오프셋 닉 및 3' 오버행으로 이어지는 오프셋 닉에 대한 비-상동 말단 연결의 백분율 비율을 보여주는 데이터를 도시한다. 도 3c는 5' 오버행으로 이어지는 오프셋 닉 및 3' 오버행으로 이어지는 오프셋 닉에 대한 표적화의 백분율 비율을 보여주는 데이터를 도시한다.

도 4a는 RuvC PDB ID: 4EP4 (청색) 위치 D7에서의 금속 배위 잔기의 개략도 (좌측), 배위된 Mg²⁺-이온 (회색 구체) 및 3M7K로부터의 DNA (자주색)를 포함하는 PDB ID: 3M7K (오렌지색) 및 4H9D (시안색)로부터의 HNH 엔도뉴클레아제 도메인의 개략도 (중앙) 및 분석된 돌연변이체의 목록 (우측)이다. 도 4b는 Cas9 돌연변이체 m3 및 m4 및 또한 VP64와의 이들 각각의 융합체에 대한 검출불가능한 뉴클레아제 활성을 보여주는 데이터를 도시한다. 도 4c는 도 4b에서의 데이터의 보다 고해상도 검사이다.

도 5a는 Cas9-gRNA 활성을 결정하기 위한 상동 제조합 검정의 개략도이다. 도 5b는 무작위 서열 삽입을 갖는

가이드 RNA 및 상동 재조합의 백분율 비율을 도시한다.

도 6a는 OCT4 유전자에 대한 가이드 RNA의 개략도이다. 도 6b는 프로모터-루시페라제 리포터 구축물에 대한 전사 활성화를 도시한다. 도 6c는 내인성 유전자의 qPCR을 통한 전사 활성화를 도시한다.

도 7a는 REX1 유전자에 대한 가이드 RNA의 개략도이다. 도 7b는 프로모터-루시페라제 리포터 구축물에 대한 전사 활성화를 도시한다. 도 7c는 내인성 유전자의 qPCR을 통한 전사 활성화를 도시한다.

도 8a는 정규화된 발현 수준의 계산을 위한 고수준 특이성 분석 프로세싱 흐름 개략도를 도시한다. 도 8b는 편재된 구축물 라이브러리 내에 생성된 미스매치의 수에 대한 결합 부위의 백분율의 분포의 데이터를 도시한다. 좌측: 이론적 분포. 우측: 실제 TALE 구축물 라이브러리로부터 관찰된 분포. 도 8c는 미스매치의 수에 대한 결합 부위에 대해 집계된 태그 카운트의 백분율의 분포의 데이터를 도시한다. 좌측: 양성 대조군 샘플로부터 관찰된 분포. 우측: 비-대조군 TALE가 유도된 샘플로부터 관찰된 분포.

도 9a는 그의 표적 서열에서 1-3개의 돌연변이에 내성을 보여주는 Cas9-gRNA 복합체의 표적화 랜드스케이프의 분석에 대한 데이터를 도시한다. 도 9b는 PAM 서열에 국재된 것들을 제외하고는 점 돌연변이에 대해 비감수성을 보여주는 Cas9-gRNA 복합체의 표적화 랜드스케이프의 분석에 대한 데이터를 도시한다. 도 9c는 2개의 염기 미스매치의 도입이 활성을 유의하게 손상시킨다는 것을 보여주는 Cas9-gRNA 복합체의 표적화 랜드스케이프의 분석에 대한 열 플롯 데이터를 도시한다. 도 9d는 에스. 피오게네스 Cas9에 대한 추정 PAM이 NGG 및 또한 NAG임을 확인하는 뉴클레아제 매개 HR 검정으로부터의 데이터를 도시한다.

도 10a는 18-량체 TALE가 그의 표적 서열에서 다중 돌연변이를 허용한다는 것을 확인하는 뉴클레아제 매개 HR 검정으로부터의 데이터를 도시한다. 도 10b는 3가지 상이한 크기 (18-량체, 14-량체 및 10-량체)의 TALE의 표적화 랜드스케이프의 분석으로부터의 데이터를 도시한다. 도 10c는 근사 단일-염기 미스매치 해상도를 나타내는 10-량체 TALE에 대한 데이터를 도시한다. 도 10d는 근사 단일-염기 미스매치 해상도를 나타내는 10-량체 TALE에 대한 열 플롯 데이터를 도시한다.

도 11a는 설계된 가이드 RNA 도시한다. 도 11b는 다양한 가이드 RNA에 대한 비상동 말단 연결의 백분율 비율을 도시한다.

도 12a는 Sox2 유전자를 도시한다. 도 12b는 Nanog 유전자를 도시한다.

도 13a-13f는 2개의 추가의 Cas9-gRNA 복합체의 표적화 랜드스케이프를 도시한다.

도 14a는 2개의 gRNA (야생형(서열식별번호:88) 및 돌연변이체(서열식별번호:89-90))의 특이성 프로파일을 도시한다. 상이한 서열은 적색으로 강조 표시하였다. 도 14b 및 14c는 이 검정이 평가 대상 gRNA에 대해 특이적임을 도시한다 (데이터는 도 13d로부터 재플롯팅됨).

도 15a-15d는 스페이서 서열 대 표적에서 단일 또는 이중-염기 미스매치 (적색으로 강조 표시됨)를 보유하는 gRNA2 (도 15a-b) 및 gRNA3 (도 15c-d)을 도시한다. 서열은 서열식별번호:91-131에서 제시된 것과 같다.

도 16a-16d는 시험된 2개의 독립된 gRNA의 뉴클레아제 검정을 도시한다: 스페이서의 5' 말단에서 말단절단된 gRNA1 (도 16a-b) 및 gRNA3 (도 16c-d). 서열은 서열식별번호:66, 185-186 및 133-140에서 제시된 것과 같다.

도 17a-17b는 에스. 피오게네스 Cas9에 대한 PAM이 NGG이고, 또한 NAG임을 보여주는 뉴클레아제 매개 HR 검정을 도시한다. 서열은 서열식별번호:67-69 및 141에서 제시된 것과 같다.

도 18a-18b는 18량체 TALE가 그의 표적 서열에 다중 돌연변이를 허용한다는 것을 확인하는 뉴클레아제 매개 HR 검정을 도시한다. 서열은 서열식별번호:70-73에서 제시된 것과 같다.

도 19a-19c는 TALE 단량체 특이성 대 TALE 단백질 특이성의 비교를 도시한다. 서열은 서열식별번호:142-150에서 제시된 것과 같다.

도 20a-20b는 오프셋 닉킹과 관련된 데이터를 도시한다. 서열은 서열식별번호:151-158에서 제시된 것과 같다.

도 21a-21c는 오프셋 닉킹 및 NHEJ 프로파일을 도시한다. 서열은 서열식별번호:159-184 및 187에서 제시된 것과 같다.

발명을 실시하기 위한 구체적인 내용

본 개시내용의 실시양태는 DNA에 전사 조절 단백질 또는 도메인을, 표적 핵산을 조절하는 방식으로 공동-국재화

[0058]

시키기 위한 DNA 결합 단백질의 사용에 기초한다. 다양한 목적을 위해 DNA에 결합하는 이러한 DNA 결합 단백질은 통상의 기술자에게 널리 공지되어 있다. 이러한 DNA 결합 단백질은 자연 발생의 것일 수 있다. 본 개시내용의 범위 내에 포함되는 DNA 결합 단백질은 본원에서 가이드 RNA로 지칭되는 RNA에 의해 가이드될 수 있는 것들을 포함한다. 이러한 측면에 따르면, 가이드 RNA와 RNA 가이드된 DNA 결합 단백질은 DNA에서 공동-국제화 복합체를 형성한다. 특정 측면에 따르면, DNA 결합 단백질은 뉴클레아제-기능부재 DNA 결합 단백질일 수 있다. 이러한 측면에 따르면, 뉴클레아제-기능부재 DNA 결합 단백질은 뉴클레아제 활성을 갖는 DNA 결합 단백질의 변경 또는 변형으로부터 생성될 수 있다. 뉴클레아제 활성을 갖는 이러한 DNA 결합 단백질은 통상의 기술자에게 공지되어 있고, 예를 들어 유형 II CRISPR 시스템으로 존재하는, Cas9 단백질과 같은, 뉴클레아제 활성을 갖는 자연 발생 DNA 결합 단백질을 포함한다. 이러한 Cas9 단백질 및 유형 II CRISPR 시스템은 관련 기술분야에 널리 문헌화되어 있다. 모든 보충 정보를 포함하여 문헌 [Makarova et al., Nature Reviews, Microbiology, Vol. 9, June 2011, pp. 467-477]을 참조하며, 이는 그 전문이 본원에 참조로 포함된다.

[0059] 뉴클레아제 활성을 갖는 예시적인 DNA 결합 단백질은 이중 가닥 DNA를 니킹하거나 컷팅하는 기능을 한다. 이러한 뉴클레아제 활성은 뉴클레아제 활성을 나타내는 하나 이상의 폴리펩티드 서열을 갖는 DNA 결합 단백질로부터 생성될 수 있다. 이러한 예시적인 DNA 결합 단백질은, 각각 이중 가닥 DNA의 특정한 가닥을 컷팅 또는 니킹하는 역할을 하는 2개의 별개의 뉴클레아제 도메인을 가질 수 있다. 뉴클레아제 활성을 갖는 예시적인 폴리펩티드 서열은 통상의 기술자에게 공지되어 있고, McrA-HNH 뉴클레아제 관련 도메인 및 RuvC-유사 뉴클레아제 도메인을 포함한다. 따라서, 예시적인 DNA 결합 단백질은 자연적으로 하나 이상의 McrA-HNH 뉴클레아제 관련 도메인 및 RuvC-유사 뉴클레아제 도메인을 함유하는 것들이다. 특정 측면에 따르면, DNA 결합 단백질은 뉴클레아제 활성이 불활성화되도록 변경되거나 달리 변형된다. 이러한 변경 또는 변형은 뉴클레아제 활성 또는 뉴클레아제 도메인을 불활성화시키기 위해 하나 이상의 아미노산을 변경시키는 것을 포함한다. 이러한 변형은 뉴클레아제 활성을 나타내는 폴리펩티드 서열 또는 폴리펩티드 서열들, 즉 뉴클레아제 도메인이 DNA 결합 단백질에 부재하도록, 뉴클레아제 활성을 나타내는 폴리펩티드 서열 또는 폴리펩티드 서열들, 즉 뉴클레아제 도메인을 제거하는 것을 포함한다. 뉴클레아제 활성을 불활성화시키기 위한 다른 변형은 본 개시내용을 기초로 하여 통상의 기술자에게 용이하게 명백할 것이다. 따라서, 뉴클레아제-기능부재 DNA 결합 단백질은 뉴클레아제 활성이 불활성화되도록 변형된 폴리펩티드 서열, 또는 뉴클레아제 활성을 불활성화시키기 위한 폴리펩티드 서열 또는 서열들의 제거를 포함한다. 뉴클레아제-기능부재 DNA 결합 단백질은 뉴클레아제 활성이 불활성화되었다고 DNA에 결합하는 능력을 유지한다. 따라서, DNA 결합 단백질은 DNA 결합에 요구되는 폴리펩티드 서열 또는 서열들을 포함하지만, 뉴클레아제 활성을 나타내는 하나 이상의 또는 모든 뉴클레아제 서열이 결여될 수 있다. 따라서, DNA 결합 단백질은 DNA 결합에 요구되는 폴리펩티드 서열 또는 서열들을 포함하지만, 불활성화된 뉴클레아제 활성을 나타내는 하나 이상의 또는 모든 뉴클레아제 서열을 가질 수 있다.

[0060] 한 측면에 따르면, 2개 이상의 뉴클레아제 도메인을 갖는 DNA 결합 단백질은 뉴클레아제 도메인 전부가 아닌 그 중 하나가 불활성화되도록 변형 또는 변경될 수 있다. 이러한 변형 또는 변경된 DNA 결합 단백질은, DNA 결합 단백질이 이중 가닥 DNA의 단 1개의 가닥만을 컷팅 또는 니킹하는 정도까지, DNA 결합 단백질 니카제로 지칭된다. RNA에 의해 DNA로 가이드되는 경우에, DNA 결합 단백질 니카제는 RNA 가이드된 DNA 결합 단백질 니카제로 지칭된다.

[0061] 예시적인 DNA 결합 단백질은 뉴클레아제 활성이 결여된 유형 II CRISPR 시스템의 RNA 가이드된 DNA 결합 단백질이다. 예시적인 DNA 결합 단백질은 뉴클레아제-기능부재 Cas9 단백질이다. 예시적인 DNA 결합 단백질은 Cas9 단백질 니카제다.

[0062] 에스. 피오게네스에서, Cas9는 단백질에서 2개의 촉매 도메인 (DNA의 상보적 가닥을 절단하는 HNH 도메인 및 비-상보적 가닥을 절단하는 RuvC-유사 도메인)에 의해 매개되는 과정을 통해 프로토스페이서-인접 모티프 (PAM)의 3bp 상류에서 평활-말단 이중-가닥 파괴를 생성한다. 문헌 [Jinke et al., Science 337, 816-821 (2012)]을 참조하며, 이는 그 전문이 본원에 참조로 포함된다. Cas9 단백질은 문헌 [Makarova et al., Nature Reviews, Microbiology, Vol. 9, June 2011, pp. 467-477]에 대한 보충 정보에서 확인되는 하기 것들을 비롯하여 다수의 유형 II CRISPR 시스템으로 존재하는 것으로 공지되어 있다: 메타노코쿠스 마리팔루디스(Methanococcus maripaludis) C7; 코리네박테리움 디프테라이에(Corynebacterium diphtheriae); 코리네박테리움 에피시엔스(Corynebacterium efficiens) YS-314; 코리네박테리움 글루타미쿰(Corynebacterium glutamicum) ATCC 13032 키타사토(Kitasato); 코리네박테리움 글루타미쿰 ATCC 13032 빌레펠트(Bielefeld); 코리네박테리움 글루타미쿰 R; 코리네박테리움 크로펜스테티이(Corynebacterium kroppenstedtii) DSM 44385; 미코박테리움 압세수스(Mycobacterium abscessus) ATCC 19977; 노카르디아 파르시니카(Nocardia farcinica) IFM10152; 로도코쿠스 에

리트로폴리스(*Rhodococcus erythropolis*) PR4; 로도코쿠스 조스티이(*Rhodococcus jostii*) RHA1; 로도코쿠스 오파쿠스(*Rhodococcus opacus*) B4 uid36573; 아시도테르무스 셀룰로리틱우스(*Acidothermus cellulolyticus*) 11B; 아르트로박터 클로로페놀리쿠스(*Arthrobacter chlorophenolicus*) A6; 크리벨라 플라비다(*Kribbella flavida*) DSM 17836 uid43465; 썬모모노스포라 쿠르바타(*Thermomonospora curvata*) DSM 43183; 비피도박테리움 덴티움(*Bifidobacterium dentium*) Bd1; 비피도박테리움 롱굼(*Bifidobacterium longum*) DJO10A; 슬라키아 헬리오트리 니레두센스(*Slackia heliotrinireducens*) DSM 20476; 페르세포넬라 마리나(*Persephonella marina*) EX H1; 박테 리오데스 프라길리스(*Bacteroides fragilis*) NCTC 9434; 카프노시토파가 오크라세아(*Capnocytophaga ochracea*) DSM 7271; 플라보박테리움 사이크로필룸(*Flavobacterium psychrophilum*) JIP02 86; 악케르만시아 무시니필라 (*Akkermansia muciniphila*) ATCC BAA 835; 로세이플렉수스 카스텐홀치이(*Roseiflexus castenholzii*) DSM 13941; 로세이플렉수스(*Roseiflexus*) RS1; 시네코시스티스(*Synechocystis*) PCC6803; 엘루시미크로비움 미누툼 (*Elusimicrobium minutum*) Pei191; 비배양된 흰개미 균 1 박테리아 계통형 Rs D17; 피브로박터 숙시노게네스 (*Fibrobacter succinogenes*) S85; 바실루스 세레우스(*Bacillus cereus*) ATCC 10987; 리스트리아 이노쿠아 (*Listeria innocua*); 락토바실루스 카세이(*Lactobacillus casei*); 락토바실루스 램노수스(*Lactobacillus rhamnosus*) GG; 락토바실루스 살리바리우스(*Lactobacillus salivarius*) UCC118; 스트렙토코쿠스 아갈락티아에 (*Streptococcus agalactiae*) A909; 스트렙토코쿠스 아갈락티아에 NEM316; 스트렙토코쿠스 아갈락티아에 2603; 스트렙토코쿠스 디스갈락티아에 에퀴시밀리스(*Streptococcus dysgalactiae equisimilis*) GGS 124; 스트렙토코 쿠스 에퀴 주에피데미쿠스(*Streptococcus equi zoepidemicus*) MGCS10565; 스트렙토코쿠스 갈롤리틱우스 (*Streptococcus gallolyticus*) UCN34 uid46061; 스트렙토코쿠스 고르도니이 칼리스(*Streptococcus gordonii* Challis) subst CH1; 스트렙토코쿠스 뮤탄스(*Streptococcus mutans*) NN2025 uid46353; 스트렙토코쿠스 뮤탄스; 스트렙토코쿠스 피오게네스(*Streptococcus pyogenes*) M1 GAS; 스트렙토코쿠스 피오게네스 MGAS5005; 스트렙토 코쿠스 피오게네스 MGAS2096; 스트렙토코쿠스 피오게네스 MGAS9429; 스트렙토코쿠스 피오게네스 MGAS10270; 스 트렙토코쿠스 피오게네스 MGAS6180; 스트렙토코쿠스 피오게네스 MGAS315; 스트렙토코쿠스 피오게네스 SSI-1; 스 트렙토코쿠스 피오게네스 MGAS10750; 스트렙토코쿠스 피오게네스 NZ131; 스트렙토코쿠스 썬모필레스 (*Streptococcus thermophiles*) CNRZ1066; 스트렙토코쿠스 썬모필레스 LMD-9; 스트렙토코쿠스 썬모필레스 LMG 18311; 클로스트리디움 보툴리눔(*Clostridium botulinum*) A3 로크 마리(Loch Maree); 클로스트리디움 보툴리눔 B 에클룬드(Eklund) 17B; 클로스트리디움 보툴리눔 Ba4 657; 클로스트리디움 보툴리눔 F 랑겔란드(Langeland); 클로스트리디움 셀룰로리틱움(*Clostridium cellulolyticum*) H10; 피네골디아 마그나(*Fingoldia magna*) ATCC 29328; 유박테리움 렉탈레(*Eubacterium rectale*) ATCC 33656; 미코플라스마 갈리셉티쿰(*Mycoplasma gallisepticum*); 미코플라스마 모빌레(*Mycoplasma mobile*) 163K; 미코플라스마 페네트란스(*Mycoplasma penetrans*); 미코플라스마 시노비아에(*Mycoplasma synoviae*) 53; 스트렙토바실루스 모닐리포르미스 (*Streptobacillus moniliformis*) DSM 12112; 브라디리조비움(*Bradyrhizobium*) BTai1; 니트로박터 함부르겐시스 (*Nitrobacter hamburgensis*) X14; 로도슈도모나스 팔루스트리스(*Rhodopseudomonas palustris*) BisB18; 로도슈 도모나스 팔루스트리스 BisB5; 파르비바쿨룸 라바멘티보란스(*Parvibaculum lavamentivorans*) DS-1; 디노로세오 박터 쉬바에(*Dinoroseobacter shibae*) DFL 12; 글루코나세토박터 디아조트로피쿠스(*Gluconacetobacter diazotrophicus*) Pal 5 FAPERJ; 글루코나세토박터 디아조트로피쿠스(*Gluconacetobacter diazotrophicus*) Pal 5 JGI; 아조스피릴룸(*Azospirillum*) B510 uid46085; 로도스피릴룸 루브룸(*Rhodospirillum rubrum*) ATCC 11170; 디아포로박터(*Diaphorobacter*) TPSY uid29975; 베르미네프로박터 에이세니아에(*Verminephrobacter eiseniae*) EF01-2; 네이세리아 메닝기티데스(*Neisseria meningitides*) 053442; 네이세리아 메닝기티데스(*Neisseria meningitides*) 알파14; 네이세리아 메닝기티데스 Z2491; 데술포비브리오 살렉시겐스(*Desulfovibrio salexigens*) DSM 2638; 캄필로박터 제주니 도일레이(*Campylobacter jejuni doylei*) 269 97; 캄필로박터 제주니 81116; 캄필로박터 제주니; 캄필로박터 라리(*Campylobacter lari*) RM2100; 헬리코박터 헤파티쿠스 (*Helicobacter hepaticus*); 윌리넬라 숙시노게네스(*Wolinella succinogenes*); 톨루모나스 아우엔시스 (*Tolomonas auensis*) DSM 9187; 슈도알테로모나스 아틀란티카(*Pseudoalteromonas atlantica*) T6c; 슈와넬라 페 알레아나(*Shewanella pealeana*) ATCC 700345; 레지오넬라 뉴모필라 파리(*Legionella pneumophila* Paris); 악티 노바실루스 숙시노게네스(*Actinobacillus succinogenes*) 130Z; 파스테우렐라 물토시다(*Pasteurella multocida*); 프란시셀라 툴라렌시스 노비시다(*Francisella tularensis novicida*) U112; 프란시셀라 툴라렌시스 홀라르크티카(*Francisella tularensis holarctica*); 프란시셀라 툴라렌시스 FSC 198; 프란시셀라 툴라렌시스 툴라렌시스; 프란시셀라 툴라렌시스 WY96-3418; 및 트레포네마 덴티콜라(*Treponema denticola*) ATCC 35405. 따라서, 본 개시내용의 측면은 본원에 기재된 바와 같이 뉴클레아제 기능부재된 또는 닉카제가 된 유형 II CRISPR 시스템으로 존재하는 Cas9 단백질에 관한 것이다.

[0063]

Cas9 단백질은 통상의 기술자에 의해 문헌에서 Csn1로서 지칭될 수 있다. 본원에 기재된 실험의 대상인 에스. 피오게네스 Cas9 단백질 서열이 하기 제시된다. 문헌 [Deltcheva et al., Nature 471, 602-607 (2011)]을 참조하며, 이는 그 전문이 본원에 참조로 포함된다.

MDKKYSIGLDIGTNSVGWAVITDEYKVPSSKFKVLGNTDRHSIKKNLIGALLFDSGETAE
 ATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFG
 NIVDEVAYHEKYPTIYHLRKKLVDSTDKADRLIYLALAHMIKFRGHFLIEGDLNPDNSD
 VDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLLENLIAQLPGEKKNGLFGN
 LIALSLGLTPNFKSNFDLAEDAKLQLSKDQYDDDLNLLAQIGDQYADLFLAAKNLSDAI
 LLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYA
 GYIDGGASQEEFYKFIKPILEKMDGTEELLVKNLREDLLRKQRTFDNGSIPHQIHLGELH
 AILRRQEDFYPLKDNREKIEKILTRIPYVVGPLARGNSRFAWMTRKSEETITPWNFEE
 VVDKGASQAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTVYNELTKVKYVTEGMRKPAFL
 SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLKLI
 IKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQLKRRRYTGWG
 RLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSL
 HEHIANLAGSPAIKKILQTVKVVDELVKVMGRHKPENIVIAMARENQTTQKGQKNSRER
 MKRIEIGIKELGSQILKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDH
 IVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNL
 TKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKS
 KLVSDFRKDFQFYKREINNYHHAHDAYLNAVVGTAIIKYPKLESEFVYGDYKVYDVR
 K
 MIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDF
 ATVRKVLSPQVNVIVKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKGFDSPVA
 YSVLVVAKVEKSKKLSVKELLGITIMERSSEFNPIDFLEAKGYKEVKKDLIIKLPK
 YSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLGSPEDNEQKQLFVE
 QHKHYLDEIIEQISEFSKRVLADANLDKVL SAYNKHDKPIREQAENIHLFTLTNLGA
 PAAFKYFDTTIDRKRYTSTKEVLDATLIHQSIITGLYETRIDLSQLGGD-

[0064]

[0065]

본원에 기재된 RNA-가이드된 게놈 조절의 방법의 특정 측면에 따르면, Cas9는 뉴클레아제 활성이 감소되거나, 실질적으로 감소되거나 또는 제거되도록 변경된다. 한 측면에 따르면, Cas9 뉴클레아제 활성은 RuvC 뉴클레아제 도메인 또는 HNH 뉴클레아제 도메인의 변경에 의해 감소되거나, 실질적으로 감소되거나 또는 제거된다. 한 측면에 따르면, RuvC 뉴클레아제 도메인은 불활성화된다. 한 측면에 따르면, HNH 뉴클레아제 도메인은 불활성화된다. 한 측면에 따르면, RuvC 뉴클레아제 도메인 및 HNH 뉴클레아제 도메인은 불활성화된다. 추가의 측면에 따르면, RuvC 뉴클레아제 도메인 및 HNH 뉴클레아제 도메인이 불활성화된 Cas9 단백질이 제공된다. 추가의 측면에 따르면, RuvC 뉴클레아제 도메인 및 HNH 뉴클레아제 도메인이 불활성화되어 있는 한, 뉴클레아제-기능부재 Cas9 단백질이 제공된다. 추가의 측면에 따르면, RuvC 뉴클레아제 도메인 또는 HNH 뉴클레아제 도메인 중 어느 하나가 불활성화되어 뉴클레아제 활성에 대한 뉴클레아제 도메인 활성이 유지되고 있는 Cas9 니카제가 제공된다. 이러한 방식으로, 이중 가닥 DNA 중 단지 1개의 가닥만이 커팅 또는 니킹된다.

[0066]

추가의 측면에 따르면, Cas9에서 1개 이상의 아미노산이 변경되거나 달리 제거되어 뉴클레아제-기능부재 Cas9 단백질을 제공하는, 뉴클레아제-기능부재 Cas9 단백질이 제공된다. 한 측면에 따르면, 아미노산은 D10 및 H840을 포함한다. 문헌 [Jinke et al., Science 337, 816-821 (2012)]을 참조한다. 추가의 측면에 따르면, 아미노산은 D839 및 N863을 포함한다. 한 측면에 따르면, D10, H840, D839 및 H863 중 1개 이상 또는 모두가 뉴클레아제 활성을 감소시키거나, 실질적으로 제거하거나 또는 제거하는 아미노산으로 치환된다. 한 측면에 따르면, D10, H840, D839 및 H863 중 1개 이상 또는 모두가 알라닌으로 치환된다. 한 측면에 따르면, D10, H840, D839 및 H863 중 1개 이상 또는 모두가 뉴클레아제 활성을 감소시키거나, 실질적으로 제거하거나 또는 제거하는 아미노산, 예컨대 알라닌으로 치환된 Cas9 단백질은 뉴클레아제-기능부재 Cas9 또는 Cas9N으로 지칭되고, 감소 또는 제거된 뉴클레아제 활성을 나타내거나, 또는 검출 수준 내에서 뉴클레아제 활성이 부재하

거나 실질적으로 부재한다. 이러한 측면에 따르면, Cas9N에 대한 뉴클레아제 활성은 공지된 검정의 이용시에 검출불가능할 수 있으며, 즉 공지된 검정의 검출 수준 미만이다.

[0067] 한 측면에 따르면, 뉴클레아제 기능부재 Cas9 단백질은, DNA에 결합하고 RNA에 의해 가이드되는 단백질의 능력을 유지하는 그의 상동체 및 오르토로그를 포함한다. 한 측면에 따르면, 뉴클레아제 기능부재 Cas9 단백질은 D10, H840, D839 및 H863 중 1개 이상 또는 모두가 알라닌으로 치환된, 에스. 피오게네스로부터의 자연 발생 Cas9에 대해 제시된 서열, 및 상기 서열에 대해 적어도 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% 또는 99% 상동성을 갖는 단백질 서열을 포함하며, 이는 DNA 결합 단백질, 예컨대 RNA 가이드된 DNA 결합 단백질이다.

[0068] 한 측면에 따르면, 뉴클레아제 기능부재 Cas9 단백질은 RuvC 뉴클레아제 도메인 및 HNH 뉴클레아제 도메인의 단백질 서열이 제외된 에스. 피오게네스로부터의 자연 발생 Cas9에 대해 제시된 서열, 및 또한 상기 서열에 대해 적어도 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 98% 또는 99% 상동성을 갖는 단백질 서열을 포함하며, 이는 DNA 결합 단백질, 예컨대 RNA 가이드된 DNA 결합 단백질이다. 이러한 방식으로, 본 개시내용의 측면은, DNA 결합, 예를 들어 가이드 RNA와의 공동-국재화 및 DNA에 대한 결합의 역할을 하는 단백질 서열 및 그에 상동성인 단백질 서열을 포함하며, 뉴클레아제 기능부재 Cas9 단백질을 생산하기 위해 RuvC 뉴클레아제 도메인 및 HNH 뉴클레아제 도메인은 자연 발생 Cas9 단백질의 단백질 서열로부터 불활성화 또는 제거될 수 있기 때문에, 이들 도메인에 대한 단백질 서열을 포함할 필요는 없다 (DNA 결합에 필요하지 않은 정도까지).

[0069] 본 개시내용의 목적을 위해, 도 4a는 Cas9에 대한 상동성을 갖는 공지된 단백질 구조에서의 금속 배위 잔기를 도시한다. 잔기는 Cas9 서열의 위치를 기초로 표지된다. 좌측: Cas9 서열에서 D10에 상응하는 RuvC 구조, PDB ID: 4EP4 (청색) 위치 D7은 Mg-이온 배위 위치가 강조되어 있다. 중앙: 배위된 Mg-이온 (회색 구체) 및 3M7K로부터의 DNA (자주색)를 포함하는 PDB ID: 3M7K (오렌지색) 및 4H9D (시안색)로부터의 HNH 엔도뉴클레아제 도메인의 구조. Cas9 아미노산 D839 및 N863에 대한 서열 상동성을 갖는 3M7K 및 4H9D 위치 D53 및 N77에서의 잔기 D92 및 N113을 막대로 나타내었다. 우측: 제조하여 뉴클레아제 활성에 대해 분석된 돌연변이체의 목록: Cas9 야생형; D10이 알라닌으로 치환된 Cas9_{m1}; D10이 알라닌으로, H840이 알라닌으로 치환된 Cas9_{m2}; D10이 알라닌으로, H840이 알라닌으로, D839가 알라닌으로 치환된 Cas9_{m3}; 및 D10이 알라닌으로, H840이 알라닌으로, D839가 알라닌으로, N863이 알라닌으로 치환된 Cas9_{m4}.

[0070] 도 4b에 나타낸 바와 같이, Cas9 돌연변이체: m3 및 m4, 및 또한 VP64와의 그의 각각의 융합체는 표적화된 유전자좌에서 심층 서열분석시에 검출불가능한 뉴클레아제 활성을 나타내었다. 플롯은 게놈 위치에 대한 돌연변이 빈도를 보여주며, 여기서 적색 선은 gRNA 표적과의 경계이다. 도 4c는 도 4b에서의 데이터의 보다 고해상도 검사이고, 돌연변이 랜드스케이프가 비변형된 유전자좌와 대등한 프로파일을 보여준다는 것이 확인된다.

[0071] 한 측면에 따르면, 뉴클레아제-기능부재 Cas9에 또는 가이드 RNA에 대한 전사 활성화 도메인의 테더링에 의해 인간 세포에서 RNA-가이드된 게놈 조절을 가능하게 하는 조작된 Cas9-gRNA 시스템이 제공된다. 본 개시내용의 한 측면에 따르면, 하나 이상의 전사 조절 단백질 또는 도메인 (상기 용어는 상호교환가능하게 사용됨)은 뉴클레아제-결합 Cas9 또는 하나 이상의 가이드 RNA (gRNA)에 연결되거나 또는 달리 결합된다. 전사 조절 도메인은 표적화된 유전자좌에 상응한다. 따라서, 본 개시내용의 측면은 전사 조절 도메인을 Cas9N에 또는 gRNA에 융합, 결합 또는 연결시킴으로써 상기 도메인을 표적화된 유전자좌에 국재화시키기 위한 방법 및 물질을 포함한다.

[0072] 한 측면에 따르면, 전사 활성화가 가능한 Cas9N-융합 단백질이 제공된다. 한 측면에 따르면, VP64 활성화 도메인 (전문이 본원에 참조로 포함되는 문헌 [Zhang et al., Nature Biotechnology 29, 149-153 (2011)] 참조)은 Cas9N의 C 말단에 연결, 융합, 결합 또는 달리 테더링된다. 한 방법에 따르면, 전사 조절 도메인은 Cas9N 단백질에 의해 표적 게놈 DNA의 부위에 제공된다. 한 방법에 따르면, 전사 조절 도메인에 융합된 Cas9N은 하나 이상의 가이드 RNA와 함께 세포 내에 제공된다. 전사 조절 도메인이 융합되어 있는 Cas9N은 표적 게놈 DNA에 또는 그 부근에 결합된다. 하나 이상의 가이드 RNA는 표적 게놈 DNA에 또는 그 부근에 결합된다. 전사 조절 도메인은 표적 유전자의 발현을 조절한다. 구체적 측면에 따르면, Cas9N-VP64 융합체는 프로모터 부근의 서열을 표적화하는 gRNA와의 조합시에 리포터 구축물의 전사를 활성화하여, RNA-가이드된 전사 활성화를 나타낸다.

[0073] 한 측면에 따르면, 전사 활성화가 가능한 gRNA-융합 단백질이 제공된다. 한 측면에 따르면, VP64 활성화 도메인은 gRNA에 연결, 융합, 결합 또는 달리 테더링된다. 한 방법에 따르면, 전사 조절 도메인은 gRNA에 의해 표적 게놈 DNA의 부위에 제공된다. 한 방법에 따르면, 전사 조절 도메인에 융합된 gRNA는 Cas9N 단백질과 함께 세포 내에 제공된다. Cas9N은 표적 게놈 DNA에 또는 그 부근에 결합된다. 전사 조절 단백질 또는 도메인이 융합되어 있는 하나 이상의 가이드 RNA는 표적 게놈 DNA에 또는 그 부근에 결합된다. 전사 조절 도메인은 표적

유전자의 발현을 조절한다. 구체적 측면에 따르면, 전사 조절 도메인과 융합된 Cas9N 단백질 및 gRNA는 리포터 구축물의 전사를 활성화시켜, RNA-가이드된 전사 활성화를 나타낸다.

[0074] 무작위 서열을 gRNA 내로 삽입하고 Cas9 기능에 대해 검정하여 gRNA의 어느 영역이 변형에 내성을 갖는지를 확인함으로써 전사 조절이 가능한 gRNA 테더를 구축하였다. 키메라 gRNA의 crRNA 부분의 5' 말단 또는 tracrRNA 부분의 3' 말단에 무작위 서열 삽입을 보유하는 gRNA는 기능을 유지하는 반면, 키메라 gRNA의 tracrRNA 스키펴드 부분 내로의 삽입은 기능 상실을 일으킨다. 무작위 염기 삽입에 대한 gRNA 유연성을 요약하는 도 5a-b를 참조한다. 도 5a는 Cas9-gRNA 활성을 결정하기 위한 상동 재조합 (HR) 검정의 개략도이다. 도 5b에 나타난 바와 같이, 키메라 gRNA의 crRNA 부분의 5' 말단 또는 tracrRNA 부분의 3' 말단에 무작위 서열 삽입을 보유하는 gRNA는 기능을 유지하는 반면, 키메라 gRNA의 tracrRNA 스키펴드 부분 내로의 삽입은 기능 상실을 일으킨다. gRNA 서열 내의 삽입 지점은 적색 뉴클레오티드에 의해 표시된다. 과학 이론에 얽매이기를 원하지 않으므로, 5' 말단에서의 무작위 염기 삽입시에 증가된 활성은 보다 긴 gRNA의 증가된 반감기로 인한 것일 수 있다.

[0075] gRNA에 VP64를 부착시키기 위해, RNA 줄기-루프에 결합하는 MS2 박테리오파지 코트-단백질의 2개 카피를 gRNA의 3' 말단에 첨부하였다. 전문이 본원에 참조로 포함되는 문헌 [Fusco et al., Current Biology: CB13, 161-167 (2003)]을 참조한다. 이들 키메라 gRNA를 Cas9N 및 MS2-VP64 융합 단백질과 함께 발현시켰다. 리포터 구축물로부터의 서열-특이적 전사 활성화는 3가지 성분 모두의 존재 하에 관찰되었다.

[0076] 도 1a는 RNA-가이드된 전사 활성화의 개략도이다. 도 1a에 나타난 바와 같이, 전사 활성화가 가능한 Cas9N-융합 단백질을 생성하기 위해, VP64 활성화 도메인을 직접적으로 Cas9N의 C 말단에 테더링하였다. 도 1b에 나타난 바와 같이, 전사 활성화가 가능한 gRNA 테더를 생성하기 위해, RNA 줄기-루프에 결합하는 MS2 박테리오파지 코트-단백질의 2개 카피를 gRNA의 3' 말단에 첨부하였다. 이들 키메라 gRNA를 Cas9N 및 MS2-VP64 융합 단백질과 함께 발현시켰다. 도 1c는 전사 활성화를 검정하는데 사용된 리포터 구축물의 설계를 보여준다. 2개의 리포터는 독특한 gRNA 표적 부위를 보유하고, 대조 TALE-TF 표적 부위를 공유한다. 도 1d에 나타난 바와 같이, Cas9N-VP64 융합체는 형광-활성화 세포 분류 (FACS) 및 면역형광 검정 (IF) 둘 다에 의해 검정시에 RNA-가이드된 전사 활성화를 나타낸다. 구체적으로, 대조 TALE-TF는 리포터 둘 다를 활성화시킨 반면, Cas9N-VP64 융합체는 gRNA 서열 특이적 방식으로 리포터를 활성화시킨다. 도 1e에 나타난 바와 같이, Cas9N, MS2-VP64 및 적절한 MS2 압타머 결합 부위를 보유하는 gRNA의 3개 성분 모두의 존재 하에서만 리포터 구축물로부터의 gRNA 서열-특이적 전사 활성화가 FACS 및 IF 둘 다에 의해 관찰되었다.

[0077] 특정 측면에 따르면, Cas9N, 하나 이상의 gRNA 및 전사 조절 단백질 또는 도메인을 사용하여 내인성 유전자를 조절하는 방법이 제공된다. 한 측면에 따르면, 내인성 유전자는 표적 유전자로서 본원에 지칭되는 임의의 바람직한 유전자일 수 있다. 하나의 예시적인 측면에 따르면, 조절을 위한 유전자 표적은 ZFP42 (REX1) 및 POU5F1 (OCT4)을 포함하며, 이들은 둘 다 다능성의 유지에 관여하는 엄격하게 조절된 유전자이다. 도 1f에 나타난 바와 같이, 전사 출발 부위의 DNA 상류의 ~5kb 스트레치를 표적화하는 10개의 gRNA (DNase 과민 부위를 녹색으로 강조함)를 REX1 유전자에 대해 설계하였다. 프로모터-루시페라제 리포터 구축물을 사용하여 (전문이 본원에 참조로 포함되는 문헌 [Takahashi et al., Cell 131 861-872 (2007)] 참조) 또는 내인성 유전자의 qPCR을 통해 직접적으로 전사 활성화를 검정하였다.

[0078] 도 6a-c는 Cas9N-VP64를 사용한 RNA-가이드된 OCT4 조절에 관한 것이다. 도 6a에 나타난 바와 같이, 전사 출발 부위의 DNA 상류의 ~5kb 스트레치를 표적화하는 21개의 gRNA를 OCT4 유전자에 대해 설계하였다. DNase 과민 부위는 녹색으로 강조된다. 도 6b는 프로모터-루시페라제 리포터 구축물을 사용한 전사 활성화를 보여준다. 도 6c는 내인성 유전자의 qPCR을 통한 직접적 전사 활성화를 보여준다. 개별 gRNA의 도입은 전사를 중간 정도로 자극하지만, 다중 gRNA는 상승작용적으로 작용하여 강건한 다중-배수 전사 활성화를 자극한다.

[0079] 도 7a-c는 Cas9N, MS2-VP64 및 gRNA+2X-MS2 압타머를 사용한 RNA-가이드된 REX1 조절에 관한 것이다. 도 7a에 나타난 바와 같이, 전사 출발 부위의 DNA 상류의 ~5kb 스트레치를 표적화하는 10개의 gRNA를 REX1 유전자에 대해 설계하였다. DNase 과민 부위는 녹색으로 강조된다. 도 7b는 프로모터-루시페라제 리포터 구축물을 사용한 전사 활성화를 보여준다. 도 7c는 내인성 유전자의 qPCR을 통한 직접적 전사 활성화를 보여준다. 개별 gRNA의 도입은 전사를 중간 정도로 자극하지만, 다중 gRNA는 상승작용적으로 작용하여 강건한 다중-배수 전사 활성화를 자극한다. 한 측면에서, gRNA에서의 2X-MS2 압타머의 부재는 전사 활성화를 일으키지 않는다. 각각의 전문이 본원에 참조로 포함되는 문헌 [Maeder et al., Nature Methods 10, 243-245 (2013) 및 Perez-Pinera et al., Nature Methods 10, 239-242 (2013)]을 참조한다.

[0080] 따라서, 방법은 다중 가이드 RNA와 Cas9N 단백질 및 표적 유전자의 발현을 조절하기 위한 전사 조절 단백질 또

는 도메인의 사용에 관련된다.

[0081] Cas9 및 gRNA 테더링 접근법은 둘 다 효과적이었고, 전자의 경우 ~1.5-2배 더 높은 효력을 나타내었다. 이러한 차이는 3-성분 복합체 어셈블리와 대조적으로 2-성분에 대한 요건으로 인한 것일 수 있다. 그러나, gRNA 테더링 접근법은 원칙적으로 각각의 gRNA가 상이한 RNA-단백질 상호작용 쌍을 사용하는 한 상이한 이펙터 도메인이 별개의 gRNA에 의해 동원되도록 할 수 있다. 전문이 본원에 참조로 포함되는 문헌 [Karyer-Bibens et al. *Biology of the Cell / Under the Auspices of the European Cell Biology Organization* 100, 125- 138 (2008)]을 참조한다. 본 개시내용의 한 측면에 따르면, 다양한 표적 유전자는 특이적 가이드 RNA 및 일반적 Cas9N 단백질, 즉 다양한 표적 유전자와 동일하거나 유사한 Cas9N 단백질을 사용하여 조절될 수 있다. 한 측면에 따르면, 멀티플렉스 유전자 조절의 방법은 동일하거나 유사한 Cas9N을 사용하여 제공된다.

[0082] 본 개시내용의 방법은 또한 Cas9N 단백질 및 본원에 기재된 가이드 RNA를 사용하여 표적 유전자를 편집함으로써 인간 세포의 멀티플렉스 유전적 및 후성적 조작을 제공하는 것에 관한 것이다. Cas9-gRNA 표적화를 목표로 (전문이 본원에 참조로 포함되는 문헌 [Jiang et al., *Nature Biotechnology* 31, 233-239 (2013)] 참조), 표적 서열 변이의 매우 큰 공간에 대한 Cas9 친화도의 심층 조사를 위한 방법이 제공된다. 따라서, 본 개시내용의 측면은 dsDNA 절단 독성에 의해 도입된 복잡성 및 천연 뉴클레아제-활성 Cas9를 사용한 특이성 시험에 의해 초래된 돌연변이 유발 복구를 회피하면서, 인간 세포에서의 Cas9 표적화의 직접적 고처리량 관독을 제공한다.

[0083] 본 개시내용의 추가 측면은 일반적으로 표적 유전자의 전사 조절을 위한 DNA 결합 단백질 또는 시스템의 용도에 관한 것이다. 통상의 기술자는 본 개시내용을 기초로 예시적인 DNA 결합 시스템을 용이하게 확인할 것이다. 이러한 DNA 결합 시스템은 자연 발생 Cas9 단백질과 같이, 어떠한 뉴클레아제 활성도 가질 필요가 없다. 따라서, 이러한 DNA 결합 시스템은 뉴클레아제 활성을 불활성화시킬 필요가 없다. 하나의 예시적인 DNA 결합 시스템은 TALE이다. 게놈 편집 도구로서, 통상적으로 TALE-FokI 이량체가 사용되고, 게놈 조절을 위해 TAEL-VP64 융합체가 고도로 효과적인 것으로 밝혀졌다. 한 측면에 따르면, TALE 특이성을 도 2a에 나타낸 방법론을 이용하여 평가하였다. 라이브러리의 각각의 성분이 d토마토 형광 단백질을 유도하는 최소 프로모터를 포함하는 구축물 라이브러리를 설계한다. 전사 출발 부위 m의 하류에 24bp (A/C/G) 무작위 전사체 태그를 삽입하고, 2개의 TF 결합 부위를 프로모터의 상류에 위치시켰다: 하나는 모든 라이브러리 요소에 의해 공유되는 불변 DNA 서열이고, 두번째 것은 돌연변이의 많은 조합이 프로그램가능 DNA 표적화 복합체가 결합하도록 설계된 표적 서열로부터 떨어져 존재하는 서열의 다수의 집합을 포괄하도록 조작된 결합 부위의 '편제된' 라이브러리를 보유하는 가변 특징부이다. 이것은 표적 서열 뉴클레오티드가 79% 빈도로 나타나고 각각의 다른 뉴클레오티드가 7% 빈도로 발생하도록 각각의 위치에서의 뉴클레오티드 빈도를 보유하도록 조작된 축중성 올리고뉴클레오티드를 사용하여 달성된다. 전문이 본원에 참조로 포함되는 문헌 [Patwardhan et al., *Nature Biotechnology* 30, 265-270 (2012)]을 참조한다. 이어서, 리포터 라이브러리를 서열분석하여, 라이브러리 요소에서 24bp d토마토 전사체 태그와 그의 상응하는 '편제된' 표적 부위 사이의 회합을 밝혀냈다. 전사체 태그의 큰 다양성은 다양한 표적 사이의 태그의 공유가 극히 드물 것이라는 것을 보장하는 반면, 표적 서열의 편제된 구축은 소수의 돌연변이를 갖는 부위가 더 많은 돌연변이를 갖는 부위보다 더 많은 태그와 회합될 것이라는 것을 의미한다. 다음으로, d 토마토 리포터 유전자의 전사가 공유된 DNA 부위에 결합하도록 조작된 대조-TF, 또는 표적 부위에 결합하도록 조작된 표적-TF로 자극된다. 각각의 발현된 전사체 태그의 존재비가 자극된 세포 상에서 RNAseq를 수행함으로써 각각의 샘플에서 측정되고, 이어서 보다 먼저 확립된 회합 표를 사용하여 그의 상응하는 결합 부위로 재맵핑한다. 대조-TF는 그의 결합 부위가 모든 라이브러리 요소에 걸쳐 공유되기 때문에 동등하게 모든 라이브러리 구성원을 여기시킬 것으로 예상되고, 반면에 표적-TF는 발현된 성분의 분포가 그에 의해 우선적으로 표적화되는 것들에 대해 치우칠 것으로 예상된다. 이러한 가정은 표적-TF에 대해 수득된 태그 카운트를 대조-TF에 대해 수득된 것들로 나눔으로써 각각의 결합 부위에 대한 정규화된 발현 수준을 계산하기 위해 단계 5에서 사용된다.

[0084] 도 2b에 나타낸 바와 같이, Cas9-gRNA 복합체의 표적화 랜드스케이프는 그가 평균적으로 그의 표적 서열에서 1-3개의 돌연변이에 내성이 있다는 것을 나타낸다. 도 2c에 나타낸 바와 같이, Cas9-gRNA 복합체는 또한 PAM 서열에 국제화된 것들을 제외하고는 주로 점 돌연변이에 비감수성이다. 두드러지게, 이러한 데이터는 에스. 피오 케네스 Cas9에 대한 추정 PAM이 NGG 뿐만 아니라 NAG임을 나타낸다. 도 2d에 나타낸 바와 같이, 2개의 염기 미스매치의 도입은 Cas9-gRNA 복합체 활성을 유의하게 손상시키지만, 단지 이들이 gRNA 표적 서열의 3' 말단에 보다 가까운 8-10개의 염기에 국제화된 경우에만 그러하다 (열 플롯에서 표적 서열 위치는 5' 말단으로부터 출발하여 1-23 표지됨).

[0085] 또 다른 널리 사용되는 게놈 편집 도구인, TALE 도메인의 돌연변이 내성을 본원에 기재된 전사 특이성 검정을 이용하여 결정하였다. 도 2e에 나타낸 바와 같이, 18-량체 TALE에 대한 TALE 오프-표적화 데이터는 그가 평균

적으로 그의 표적 서열에서 1-2개의 돌연변이에 내성일 수 있으며, 그의 표적에서 대다수의 3개의 염기 미스매치 변이체를 활성화하는데 실패한 것을 나타낸다. 도 2f에 나타낸 바와 같이, 18-량체 TALE는 그의 표적에서 미스매치된 단일 염기에 주로 비감수성인 Cas9-gRNA 복합체와 유사하다. 도 2g에 나타낸 바와 같이, 2개의 염기 미스매치의 도입은 18-량체 TALE 활성을 유의하게 손상시킨다. TALE 활성은 그의 표적 서열의 5' 말단에 보다 가까운 미스매치에 보다 감수성이다 (열 플롯에서 표적 서열 위치는 5' 말단으로부터 출발하여 1-18 표지됨).

[0086] 결과를, 다양한 크기의 TALE에 의한 표적화의 랜스케이프를 평가하는 것에 관련된 도 10a-c의 대상인 뉴클레아제 검정으로 표적화된 실험을 이용하여 확인하였다. 도 10a에 나타낸 바와 같이, 뉴클레아제 매개 HR 검정을 이용하여, 18-량체 TALE가 그의 표적 서열에서 다중 돌연변이를 허용한다는 것을 확인하였다. 도 10b에 나타낸 바와 같이, 도 2에서 기재된 접근법을 이용하여, 3가지 상이한 크기 (18-량체, 14-량체 및 10-량체)의 TALE의 표적화 랜스케이프를 분석하였다. 보다 짧은 TALE (14-량체 및 10-량체)는 그의 표적화에서 점점 더 특이적이지만, 또한 거의 한 자릿수만큼 활성이 감소된다. 도 10c 및 10d에 나타낸 바와 같이, 10-량체 TALE는 근사 단일-염기 미스매치 해상도를 나타내며, 2개의 미스매치를 보유하는 표적에 대해 거의 모든 활성이 손실된다 (열 플롯에서 표적 서열 위치는 5' 말단에서 출발하여 1-10 표지됨). 종합하여, 이들 데이터는 보다 짧은 TALE를 조작하는 것이 게놈 조작 적용에서 보다 높은 특이성을 생성할 수 있는 반면에 TALE 뉴클레아제 적용에서의 FokI 이량체화에 대한 요건이 오프-표적 효과를 회피하는데 필수적이라는 것을 나타낸다. 각각의 전문이 본원에 참조로 포함되는 문헌 [Kim et al., Proceedings of the National Academy of Sciences of the United States of America 93, 1 156-1 160 (1996) 및 Pattanayak et al., Nature Methods 8, 765-770 (2011)]을 참조한다.

[0087] 도 8a-c는 실험 데이터로부터의 예로 예시된 정규화된 발현 수준의 계산을 위한 고수준 특이성 분석 프로세싱 흐름에 관한 것이다. 도 8a에 나타낸 바와 같이, 구축물 라이브러리는 리포터 유전자 전사체 내로 혼입될 결합 부위 서열의 편재 분포 및 무작위 서열 24bp 태그를 갖도록 생성된다 (상단). 전사된 태그는 그들이 Cas9 또는 TALE 결합 서열에 다 대 일로 맵핑되어야 하도록 고도로 축증된다. 구축물 라이브러리를 서열분석하여 (제3 수준, 좌측) 태그가 결합 부위와 공동-발생한다는 것을 확립함으로써, 결합 부위 대 전사된 태그의 회합 표를 생성하였다 (제4 수준, 좌측). 라이브러리 바코드 (여기서 담청색 및 담황색에 의해 나타냄; 수준 1-4, 좌측)를 사용하여 다양한 결합 부위에 대해 구축된 다중 구축물 라이브러리를 한 번에 서열분석할 수 있다. 이어서, 구축물 라이브러리를 세포 집단 내로 형질감염시키고, 일련의 다양한 Cas9/gRNA 또는 TALE 전사 인자를 집단의 샘플에서 유도한다 (제2 수준, 우측). 하나의 샘플은 항상 구축물 내의 고정된 결합 부위 서열 (상단 수준, 녹색 박스)에 표적화된 고정된 TALE 활성인자에 의해 유도되고; 이러한 샘플은 양성 대조군 (녹색 샘플, 또한 + 부호로 나타냄)의 역할을 한다. 이어서, 유도된 샘플에서 리포터 mRNA 분자로부터 생성된 cDNA를 서열분석하고, 샘플에서 각각의 태그에 대한 태그 카운트를 얻기 위해 분석한다 (제3 및 제4 수준, 우측). 구축물 라이브러리 서열분석과 같이, 양성 대조군을 포함하는 다중 샘플을 서열분석하고, 샘플 바코드를 첨부함으로써 함께 분석하였다. 여기서 담적색은 서열분석되고 양성 대조군 (녹색)과 함께 분석된 하나의 비-대조 샘플을 나타낸다. 각각의 리드에 구축물 결합 부위가 아닌 전사된 태그만이 나타나기 때문에, 이어서, 구축물 라이브러리 서열분석으로부터 수득한 결합 부위 대 태그 회합 표를 사용하여 각각의 샘플에서 각각의 결합 부위로부터 발현된 태그의 총 카운트를 총계한다 (제5 수준). 이어서, 각각의 비-양성 대조군 샘플에 대한 총계를 양성 대조군 샘플에서 수득한 총계로 나눔으로써 이를 각각의 결합 부위에 대한 정규화된 발현 수준으로 변환한다. 미스매치의 수에 대한 정규화된 발현 수준의 플롯의 예가 도 2b 및 2e, 및 도 9a 및 도 10b에 제공된다. 잘못된 태그에 대한, 구축물 라이브러리와 회합될 수 없는 태그에 대한, 및 다중 결합 부위와 명백히 공유된 태그에 대한 필터링의 몇몇 수준이 상기 전반적인 공정 흐름에 포함되지 않는다. 도 8b는 편재된 구축물 라이브러리 내에 생성된 미스매치의 수에 대한 결합 부위의 백분율의 예시적 분포를 도시한다. 좌측: 이론적 분포. 우측: 실제 TALE 구축물 라이브러리로부터 관찰된 분포. 도 8c는 미스매치의 수에 대한 결합 부위에 대해 집계된 태그 카운트의 백분율의 예시적 분포를 도시한다. 좌측: 양성 대조군 샘플로부터 관찰된 분포. 우측: 비-대조 TALE가 유도된 샘플로부터 관찰된 분포. 양성 대조군 TALE가 구축물 내의 고정된 부위에 결합하기 때문에, 집계된 태그 카운트의 분포는 도 8b에서의 결합 부위의 분포를 밀접하게 반영하고, 반면에 더 적은 미스매치를 갖는 부위가 더 높은 발현 수준을 유도하기 때문에 분포는 비-대조 TALE 샘플의 경우에 좌측으로 치우친다. 하단: 표적-TF에 대해 수득된 태그 카운트를 대조-TF에 대해 수득된 것으로 나눔으로써 이들 사이의 상대적 풍부화를 계산하는 것은 표적 부위에서의 돌연변이의 수에 대한 평균 발현 수준을 나타낸다.

[0088] 이들 결과는 다양한 Cas9-gRNA 복합체를 사용하여 생성된 특이성 데이터에 의해 추가로 재확인된다. 도 9a에 나타낸 바와 같이, 다양한 Cas9-gRNA 복합체는 그의 표적 서열에서 1-3개의 돌연변이에 내성이 있다. 도 9b에

나타낸 바와 같이, Cas9-gRNA 복합체는 또한 PAM 서열에 국재화된 것들을 제외하고는 주로 점 돌연변이에 대해 비감수성이다. 그러나, 도 9c에 나타낸 바와 같이, 2개의 염기 미스매치의 도입은 활성을 유의하게 손상시킨다 (열 플롯에서 표적 서열 위치는 5' 말단으로부터 출발하여 1-23 표지됨). 도 9d에 나타낸 바와 같이, 뉴클레아제 매개 HR 검정을 이용하여 에스. 피오케네스 Cas9에 대한 추정 PAM이 NGG 및 또한 NAG임을 확인하였다.

[0089] 특정 측면에 따르면, 결합 특이성은 본원에 기재된 방법에 따라 증가된다. 다중 복합체 사이의 상승작용은 Cas9N-VP64에 의한 표적 유전자 활성화의 요인이기 때문에, 개별 오프-표적 결합 이벤트는 최소 효과를 가져야 하고, 따라서 Cas9N의 전사 조절 적용은 자연적으로 상당히 특이적이다. 한 측면에 따르면, 오프셋 Nick은 게놈-편집의 방법에 사용된다. 대다수의 Nick은 드물게 NHEJ 이벤트를 일으키고 (전문이 본원에 참조로 포함되는 문헌 [Certo et al., Nature Methods 8, 671-676 (2011)] 참조), 따라서 오프-표적 Nicking의 효과를 최소화한다. 대조적으로, 오프셋 Nick을 유도하여 이중 가닥 파괴 (DSB)를 생성하는 것은 유전자 파괴를 유도하는데 있어서 고도로 효과적이다. 특정 측면에 따르면, 5' 오버행은 3' 오버행과는 대조적으로 보다 유의한 NHEJ 이벤트를 생성한다. 유사하게, 3' 오버행은 NHEJ 이벤트에 비해 HR을 선호하지만, HR 이벤트의 총 수는 5' 오버행이 생성하는 경우보다 유의하게 더 낮다. 따라서, 상동 재조합을 위해 Nick을 사용하고, 오프-표적 Cas9-gRNA 활성의 효과가 최소화되도록 이중 가닥 파괴를 생성하기 위해 오프셋 Nick을 사용하는 방법이 제공된다.

[0090] 도 3a-c는 멀티플렉스 오프셋 Nicking 및 가이드 RNA와의 오프-표적 결합을 감소시키는 방법에 관한 것이다. 도 3a에 나타낸 바와 같이, 신호등 리포터를 사용하여 표적화된 Nick 또는 파괴의 도입시의 HR 및 NHEJ 이벤트에 대해 동시에 검정하였다. HDR 경로를 통해 해상된 DNA 절단 이벤트는 GFP 서열을 복원하는 반면, 돌연변이유발 NHEJ는 GFP를 프레임 밖으로, 하류 m체리 서열을 프레임 내로 프레임시프트되게 한다. 검정을 위해, 센스 가닥을 표적화하는 7개 (U1-7) 및 안티센스 가닥을 표적화하는 7개 (D1-7)의, 200bp 스트레치의 DNA를 포함하는 14개의 gRNA를 설계하였다. 상보적 가닥을 Nicking하는 Cas9D 10A 돌연변이체를 사용하여, gRNA의 다양한 2-원 조합을 사용하여 다수의 프로그래밍된 5' 또는 3' 오버행을 유도하였다 (14개의 gRNA에 대한 Nicking 부위를 표시함). 도 3b에 나타낸 바와 같이, 오프셋 Nick을 유도하여 이중 가닥 파괴 (DSB)를 생성하는 것은 유전자 파괴를 유도하는데 있어서 고도로 효과적이다. 두드러지게, 5' 오버행으로 이어지는 오프셋 Nick은 3' 오버행과는 대조적으로 보다 많은 NHEJ 이벤트를 일으킨다. 도 3c에 나타낸 바와 같이, 3' 오버행의 생성은 또한 NHEJ 이벤트에 비해 HR의 비가 우세하지만, HR 이벤트의 총 수는 5' 오버행이 생성하는 경우보다 유의하게 더 낮다.

[0091] 도 11a-b는 Cas9D10A Nickase 매개 NHEJ에 관한 것이다. 도 11a에 나타낸 바와 같이, 신호등 리포터를 사용하여 표적화된 Nick 또는 이중-가닥 파괴의 도입시의 NHEJ 이벤트를 검정하였다. 간략하게, DNA 절단 이벤트의 도입시에, 파괴가 돌연변이유발 NHEJ를 통해 진행된다면, GFP는 프레임 밖으로 번역되고, 하류 m체리 서열은 프레임 내에서 적색 형광을 생성하게 된다. 센스 가닥을 표적화하는 7개 (U1-7) 및 안티센스 가닥을 표적화하는 7개 (D1-7)의, 200bp 스트레치의 DNA를 포함하는 14개의 gRNA를 설계하였다. 도 11b에 나타낸 바와 같이, 모든 표적에 걸쳐 DSB 및 강건한 NHEJ를 일으키는 야생형 Cas9와 달리, (Cas9D 10A 돌연변이체 사용시) 대부분의 Nick이 드물게 NHEJ 이벤트를 일으키는 것으로 관찰되었다. 모든 14개의 부위는 DNA의 인접 200bp 스트레치 내에 위치하고, 표적화 효율에서 10배가 넘는 차이가 관찰되었다.

[0092] 특정 측면에 따르면, 1개 이상, 2개 이상 또는 다수의 외래 핵산을 세포 내로 도입하는 것을 포함하는, 세포에서 표적 핵산의 발현을 조절하는 방법이 본원에 기재된다. 세포 내로 도입된 외래 핵산은 가이드 RNA 또는 가이드 RNA들, 뉴클레아제-기능부재 Cas9 단백질 또는 단백질들, 및 전사 조절 단백질 또는 도메인을 코딩한다. 함께, 가이드 RNA, 뉴클레아제-기능부재 Cas9 단백질 및 전사 조절 단백질 또는 도메인은 공동-국재화 복합체로서 지칭되며, 상기 용어는 가이드 RNA, 뉴클레아제-기능부재 Cas9 단백질 및 전사 조절 단백질 또는 도메인이 DNA에 결합하고 표적 핵산의 발현을 조절한다는 정도로 통상의 기술자에 의해 이해된다. 특정의 추가의 측면에 따르면, 세포 내로 도입된 외래 핵산은 가이드 RNA 또는 가이드 RNA들 및 Cas9 단백질 Nickase를 코딩한다. 함께, 가이드 RNA 및 Cas9 단백질 Nickase는 공동-국재화 복합체로서 지칭되며, 상기 용어는 가이드 RNA 및 Cas9 단백질 Nickase가 DNA에 결합하고 표적 핵산을 Nicking한다는 정도로 통상의 기술자에 의해 이해된다.

[0093] 본 개시내용에 따른 세포는 외래 핵산이 본원에 기재된 바와 같이 도입되고 발현될 수 있는 임의의 세포를 포함한다. 본원에 기재된 본 개시내용의 기본 개념은 세포 유형에 제한되지 않는 것으로 이해되어야 한다. 본 개시내용에 따른 세포는 진핵 세포, 원핵 세포, 동물 세포, 식물 세포, 진균 세포, 고세균 세포, 유박테리아 세포 등을 포함한다. 세포는 진핵 세포, 예컨대 효모 세포, 식물 세포 및 동물 세포를 포함한다. 특정한 세포는 포유동물 세포를 포함한다. 또한, 세포는 표적 핵산을 조절하는데 유익하거나 바람직한 임의의 것을 포함한다. 이러한 세포는 질환 또는 유해 상태를 유발하는 특정한 단백질의 발현이 결핍된 것들을 포함할 수 있다. 이러한 질환 또는 유해 상태는 통상의 기술자에게 용이하게 공지되어 있다. 본 개시내용에 따르면, 특정한 단백질

의 발현에 원인이 되는 핵산은, 표적 핵산 및 특정한 단백질의 상응하는 발현의 상향조절이 일어나도록 본원에 기재된 방법 및 전사 활성인자에 의해 표적화될 수 있다. 이러한 방식으로, 본원에 기재된 방법은 치유적 치료를 제공한다.

[0094] 표적 핵산은, 본원에 기재된 바와 같은 공동-국제화 복합체가 조절 또는 너킹에 유용할 수 있는 임의의 핵산 서열을 포함한다. 표적 핵산은 유전자를 포함한다. 본 개시내용의 목적을 위해, DNA, 예컨대 이중 가닥 DNA는 표적 핵산을 포함할 수 있고, 공동-국제화 복합체는 표적 핵산에서 또는 그에 인접하여 또는 그 부근에서 공동-국제화 복합체가 표적 핵산에 대해 바람직한 효과를 가질 수 있는 방식으로 DNA에 결합할 수 있거나, 또는 그와 달리 공동-국제화될 수 있다. 이러한 표적 핵산은 내인성 (또는 자연 발생) 핵산 및 외인성 (또는 외래) 핵산을 포함할 수 있다. 본 개시내용을 기초로 하여 통상의 기술자는 표적 핵산을 포함하는 DNA에 공동-국제화되는 가이드 RNA 및 Cas9 단백질을 용이하게 확인 또는 설계할 수 있을 것이다. 통상의 기술자는 또한, 마찬가지로 표적 핵산을 포함하는 DNA에 공동-국제화되는 전사 조절 단백질 또는 도메인을 확인할 수 있을 것이다. DNA는 게놈 DNA, 미토콘드리아 DNA, 바이러스 DNA 또는 외인성 DNA를 포함한다.

[0095] 외래 핵산 (즉, 세포의 천연 핵산 조성물의 부분이 아닌 것들)이, 이러한 도입에 대해 통상의 기술자에게 공지된 임의의 방법을 이용하여 세포 내로 도입될 수 있다. 이러한 방법은 형질감염, 형질도입, 바이러스 형질도입, 미세주사, 리포펙션, 뉴클레오펙션, 나노입자 포격, 형질전환, 접합 등을 포함한다. 통상의 기술자는 용이하게 확인가능한 문헌 공급원을 이용하여 이러한 방법을 용이하게 이해하고 적합화할 것이다.

[0096] 전사 활성인자인 전사 조절 단백질 또는 도메인은 VP16 및 VP64, 및 본 개시내용을 기초로 하여 통상의 기술자에 의해 용이하게 확인가능한 다른 것들을 포함한다.

[0097] 질환 및 유해 상태는 특정한 단백질의 발현의 비정상적 손실을 특징으로 하는 것들이다. 이러한 질환 또는 유해 상태는 특정한 단백질의 상향조절로 치료될 수 있다. 따라서, 본원에 기재된 바와 같은 공동-국제화 복합체가 표적 핵산을 포함하는 DNA에 회합 또는 달리 결합되고, 공동-국제화 복합체의 전사 활성인자가 표적 핵산의 발현을 상향조절하는 것인, 질환 또는 유해 상태를 치료하는 방법이 제공된다. 예를 들어, PRDM16 및 갈색 지방 분화를 촉진하고 대사 섭취를 증가시키는 다른 유전자의 상향조절이 대사 증후군 또는 비만을 치료하는데 사용될 수 있다. 항염증 유전자의 활성화는 자가면역 및 심혈관 질환에 유용하다. 종양 억제 유전자의 활성화는 암의 치료에 유용하다. 통상의 기술자는 본 개시내용을 기초로 하여 이러한 질환 및 유해 상태를 용이하게 확인할 것이다.

[0098] 하기 실시예는 본 개시내용을 대표하는 것으로서 제시된다. 이들 실시양태 및 다른 등가의 실시양태가 본 개시내용, 도면 및 첨부된 청구범위의 관점에서 명백할 것이기 때문에, 이들 실시예는 본 개시내용의 범위를 제한하는 것으로 해석되어서는 안된다.

[0099] 실시예 I

[0100] Cas9 돌연변이체

[0101] 기지의 구조를 갖는 Cas9에 상동성인 서열을 검색하여, 그의 RuvC 및 HNH 도메인의 본래의 활성을 제거할 수 있는 Cas9에서의 후보 돌연변이를 확인하였다. HHpred (www.toolkit.tuebingen.mpg.de/hhpred)를 이용하여, Cas9의 전체 서열을 전체 단백질 데이터뱅크(Protein Data Bank) (January 2013)와 대조하여 조회하였다. 다시, Cas9의 HNH 도메인에 대해 유의한 서열 상동성을 갖는 2종의 상이한 HNH 엔도뉴클레아제인 PacI 및 추정 엔도뉴클레아제 (각각 PDB ID: 3M7K 및 4H9D)를 검색하였다. 이들 단백질을, 마그네슘 이온 배위에 관여하는 잔기를 찾기 위해 조사하였다. 이어서, 상응하는 잔기를 Cas9에 대한 서열 정렬에서 확인하였다. Cas9의 동일한 아미노산 유형에 대해 정렬된 각각의 구조 내에서 2개의 Mg-배위 측쇄가 확인되었다. 이들은 3M7K D92 및 N113, 및 4H9D D53 및 N77이다. 이들 잔기는 Cas9 D839 및 N863에 상응하였다. 또한, PacI 잔기 D92 및 N113의 알라닌으로의 돌연변이가 뉴클레아제를 촉매작용이 결여되도록 만든다고 보고되었다. 이러한 분석에 기초하여 Cas9 돌연변이 D839A 및 N863A를 만들었다. 또한, HHpred는 Cas9와 썬무스 썬포필루스(Thermus thermophilus) RuvC (PDB ID: 4EP4)의 N-말단 사이의 상동성을 예측하기도 한다. 이러한 서열 정렬은 Cas9에서 RuvC 도메인의 기능을 제거한 이전의 보고된 돌연변이 D10A를 포함한다. 이것이 적절한 돌연변이인지 확인하기 위해, 금속 결합 잔기를 이전과 같이 결정하였다. 4EP4에서, D7은 마그네슘 이온의 배위를 돕는다. 이 위치가 Cas9 D10에 상응하는 서열 상동성을 가지므로, 이러한 돌연변이가 금속 결합의 제거를 도와, Cas9 RuvC 도메인으로부터의 촉매 활성을 제거한다는 것을 확인하였다.

[0102] 실시예 II

- [0103] 플라스미드 구축
- [0104] Cas9 돌연변이체를 퀵체인지(Quikchange) 키트 (애질런트 테크놀로지스(Agilent technologies))를 사용하여 제조하였다. 표적 gRNA 발현 구축물은 (1) IDT로부터 개별 지블록(gBlock)으로 직접 주문하여 pCR-BluntII-TOPO 벡터 (인비트로젠(Invitrogen))에 클로닝한 것이거나; 또는 (2) 진위즈(Genewiz)에 의해 주문 합성된 것이거나; 또는 (3) 올리고뉴클레오티드의 깃슨(Gibson) 어셈블리를 사용하여 gRNA 클로닝 벡터 (플라스미드 #41824)에 어셈블리된 것이다. 봉괴된 GFP가 수반되는 HR 리포터 검정에 사용하기 위한 벡터를, 정지 코돈을 보유하는 GFP 서열과 애드진(Addgene)으로부터의 EGIP 렌티벡터 (플라스미드 #26777)에 어셈블리된 적절한 단편의 융합 PCR 어셈블리에 의해 구축하였다. 이어서, 이러한 렌티벡터를 사용하여 GFP 리포터 안정한 라인을 확립하였다. 이 연구에서 사용되는 TALEN은 표준 프로토콜을 이용하여 구축하였다. 본원에 그 전문이 참조로 포함된 문헌 [Sanjana et al., Nature Protocols 7, 171-192 (2012)]을 참조한다. 표준 PCR 융합 프로토콜 절차를 이용하여 Cas9N과 MS2 VP64를 융합시켰다. OCT4 및 REX1에 대한 프로모터 루시페라제 구축물 (플라스미드 #17221 및 플라스미드 #17222)을 애드진으로부터 입수하였다.
- [0105] 실시예 III
- [0106] 세포 배양 및 형질감염
- [0107] HEK 293T 세포를 10% 태아 소 혈청 (FBS, 인비트로젠), 페니실린/스트렙토마이신 (pen/strep, 인비트로젠), 및 비-필수 아미노산 (NEAA, 인비트로젠)이 보충된 돌베코 변형 이글 배지(Dulbecco's modified Eagle's medium) (DMEM, 인비트로젠) 고 글루코스에서 배양하였다. 세포를 가슴 인큐베이터에서 37°C 및 5% CO₂에서 유지하였다.
- [0108] 뉴클레아제 검정에 수반되는 형질감염은 다음과 같다: 리포펙타민(Lipofectamine) 2000을 제조업체의 프로토콜에 따라 사용하여 0.4 x 10⁶ 개 세포를 2 µg Cas9 플라스미드, 2 µg gRNA 및/또는 2 µg DNA 공여자 플라스미드로 형질감염시켰다. 형질감염시키고 3일 후에 세포를 수거하여, FACS에 의해 분석하거나, 또는 게놈 절단물의 직접 검정을 위해 DNAeasy 키트 (퀴아젠(Qiagen))를 사용하여 ~1 X 10⁶ 개 세포의 게놈 DNA를 추출하였다. 이러한 경우, 세포로부터 유래된 게놈 DNA로 PCR을 수행하여 표적화 영역을 증폭시키고, 앰플리콘을 MiSeq 퍼스널 시퀀서(MiSeq Personal Sequencer) (일루미나(Illumina))에 의해 유효범위 >200,000개 판독물로 심층 서열분석하였다. 이러한 서열분석 데이터를 분석하여 NHEJ 효율을 추정하였다.
- [0109] 전사 활성화 검정에 수반되는 형질감염의 경우: 0.4 x 10⁶ 개 세포를 (1) 2 µg Cas9N-VP64 플라스미드, 2 µg gRNA 및/또는 0.25 µg의 리포터 구축물; 또는 (2) 2 µg Cas9N 플라스미드, 2 µg MS2-VP64, 2 µg gRNA-2XMS2압타머 및/또는 0.25 µg의 리포터 구축물로 형질감염시켰다. 형질감염시키고 24-48시간 후에 세포를 수거하여, FACS 또는 면역형광 방법을 이용하여 검정하거나, 또는 그의 전체 RNA를 추출하고, 이후에 RT-PCR에 의해 분석하였다. 여기서 OCT4 및 REX1에 대한 표준 택맨(taqman) 프로브 (인비트로젠)가 사용되었으며, 각 샘플을 GAPDH에 대해 정규화하였다.
- [0110] Cas9-gRNA 복합체 및 TALE의 특이성 프로파일에 대한 전사 활성화 검정에 수반되는 형질감염의 경우: 0.4 x 10⁶ 개 세포를 (1) 2 µg Cas9N-VP64 플라스미드, 2 µg gRNA 및 0.25 µg의 리포터 라이브러리; 또는 (2) 2 µg TALE-TF 플라스미드 및 0.25 µg의 리포터 라이브러리; 또는 (3) 2 µg 대조-TF 플라스미드 및 0.25 µg의 리포터 라이브러리로 형질감염시켰다. 형질감염시키고 24시간 후에 (리포터의 자극이 포화 모드가 되는 것을 피하기 위해) 세포를 수거하였다. RNAeasy-플러스 키트 (퀴아젠)를 사용하여 전체 RNA를 추출하고, 슈퍼스크립트(Superscript)-III (인비트로젠)을 사용하여 표준 RT-pcr을 수행하였다. 전사체-태그의 표적화된 pcr 증폭에 의해 차세대 서열분석을 위한 라이브러리가 생성되었다.
- [0111] 실시예 IV
- [0112] Cas9-TF 및 TALE-TF 리포터 발현 수준의 계산을 위한 전산 및 서열 분석
- [0113] 이러한 과정을 위한 고수준 논리 흐름도가 도 8a에 도시되고, 추가의 세부사항이 여기에 제시된다. 구축물 라이브러리 구성에 대한 세부사항은, 도 8a (레벨 1) 및 8b를 참조한다.
- [0114] 서열분석: Cas9 실험의 경우, 구축물 라이브러리 (도 8a, 레벨 3, 좌측) 및 리포터 유전자 cDNA 서열 (도 8a, 레벨 3, 우측)을 일루미나 MiSeq 상에서 150bp 중첩 쌍형성된 말단 판독물로서 수득하고, TALE 실험의 경우, 상

응하는 서열을 일루미나 HiSeq 상에서 51bp 비-중첩 쌍형성된 말단 판독물로서 획득하였다.

- [0115] 구축물 라이브러리 서열 처리: 정렬: Cas9 실험의 경우, 노보얼라인(novoalign) V2.07.17 (www.novocraft.com/main/index/php)을 이용하여, 8bp 라이브러리 바코드의 쌍에 의해 플랭킹된 234bp의 구축물에 상응하는 250bp 참조 서열의 세트에 대해 쌍형성된 판독물을 정렬시켰다 (도 8a, 제3 레벨, 좌측 참조). 노보얼라인에 제공된 참조 서열에는, 구축물 라이브러리 바코드가 명확하게 제시되면서 23bp 축중성 Cas9 결합 부위 영역 및 24bp 축중성 전사체 태그 영역 (도 8a, 제1 레벨 참조)이 N으로 명시되어 있다. TALE 실험의 경우, 참조 서열의 길이가 203bp이고, 축중성 결합 부위 영역의 길이가 18bp 대 23bp인 것을 제외하고 동일한 절차를 이용하였다. 유효성 검사: 각각의 판독물 쌍에 대한 좌측 및 우측 판독물이 참조 서열에 대해 개별적으로 정렬되도록 구성된 파일에 대한 노보얼라인 결과. 둘 모두가 참조 서열에 대해 특유하게 정렬된 판독물 쌍만이 추가의 유효성 조건에 적용되었으며, 이러한 모든 조건을 통과한 판독물 쌍만을 남겨두었다. 유효성 조건은 다음을 포함하였다:
- [0116] (i) 2개의 구축물 라이브러리 바코드는 각각 참조 서열 바코드에 대해 적어도 4군데 위치에서 정렬되어야 하고, 2개의 바코드는 동일한 구축물 라이브러리에 대한 바코드 쌍에 대해 정렬되어야 한다.
- [0117] (ii) 참조 서열의 N 영역에 대해 정렬된 모든 염기는 노보얼라인에 의해 A, C, G 또는 T로 지명되어야 한다. Cas9 실험 뿐만 아니라 TALE 실험에서도 좌측 및 우측 판독물이 참조 N 영역에서 중첩되지 않았으며, 이에 따라 이러한 N 염기가 노보얼라인에 의해 불분명하게 지명될 가능성이 발생하지 않았다는 점을 주목한다. (iii) 마찬가지로, 노보얼라인-지명된 삽입부 또는 결실부가 이들 영역에서 나타나지 않아야 한다. (iv) (이러한 무작위 서열은 오직 A, C, 및 G로부터 생성되므로) T가 전사체 태그 영역에서 나타나지 않아야 한다. 이러한 조건 중 어느 하나를 위반하는 판독물 쌍을 불합격 판독물 쌍 파일에 수집하였다. 이러한 유효성 검사를 주문형 perl 스크립트를 사용하여 시행하였다.
- [0118] 유도된 샘플 리포터 유전자 cDNA 서열 처리: 정렬: 먼저 SeqPrep (www.github.com/jstjohn/SeqPrep로부터 다운로드됨)을 사용하여 중첩 판독물 쌍을 79bp 공통 절편에 병합시킨 후에, 노보얼라인 (상기 버전)을 사용하여 이들 79bp 공통 절편들을, (구축물 라이브러리 서열분석의 경우) 샘플 바코드가 명확하게 제시되면서 24bp 축중성 전사체 태그가 N으로 명시된 참조 서열의 세트에 대해 쌍형성되지 않은 단일 판독물로서 정렬시켰다 (도 8a, 제3 레벨, 우측 참조). TALE 및 Cas9 cDNA 서열 영역은 둘 다 8bp 샘플 바코드 서열의 쌍에 의해 플랭킹된 cDNA의 동일한 63bp 영역에 상응하였다. 유효성 검사: 하기를 제외한 동일한 조건이 구축물 라이브러리 서열분석에 적용되었다 (상기 참조): (a) 여기서, 판독물 쌍의 이전의 SeqPrep 병합으로 인해, 유효성 처리는 판독물 쌍 내의 두 판독물의 특유한 정렬을 필터링하는 것이 아니라 오직 병합된 판독물의 특유한 정렬을 필터링해야 한다. (b) 유효성 처리가 오직 참조 서열의 전사체 태그 영역에만 적용되고, 별도의 결합 부위 영역에는 적용되지 않도록, 오직 전사체 태그만이 cDNA 서열 판독물에서 나타나야 한다.
- [0119] 결합 부위 대 전사체 태그 회합의 표의 어셈블리: 주문형 perl을 사용하여, 유효 구축물 라이브러리 서열로부터 상기 표를 작성하였다 (도 8a, 제4 레벨, 좌측). A, C, 및 G 염기로 구성된 24bp 태그 서열은 구축물 라이브러리의 전반에 걸쳐 본질적으로 특유해야 하지만 (공유 확률 = $\sim 2.8e^{-11}$), 결합 부위 대 태그 회합의 초기 분석에서 태그 서열의 무시할 수 없는 분획이 실제로 다중 결합 서열에 의해 공유되는 것으로 밝혀졌으며, 이는 아마도 주로 결합 서열에서의 서열 오류, 또는 구축물 라이브러리를 생성하는데 사용되는 올리고에서의 올리고 합성 오류의 조합에 의해 유발된 것일 것이다. 태그 공유에 더하여, 유효 판독물 쌍에서 결합 부위와 회합되어 발견되는 태그는 바코드 미스매치로 인해 이들이 어느 구축물 라이브러리로부터 유래된 것인지 분명하지 않은 경우에 구축물 라이브러리 판독물 쌍 불합격 파일에서 발견될 수도 있을 것이다. 마지막으로, 태그 서열은 그 자체가 서열 오류를 함유할 수 있다. 이러한 예외의 근원을 처리하기 위해, 태그를 3가지 속성으로 분류하였다: (i) 안전 대 불안, 여기서 불안은 태그가 구축물 라이브러리 불합격 판독물 쌍 파일에서 발견될 수 있음을 의미한다; 공유 대 비공유, 여기서 공유는 태그가 다중 결합 부위 서열과 회합되어 발견되었다는 것을 의미한다; 및 2+ 대 1-유일, 여기서 2+는 태그가 유효 구축물 라이브러리 서열 가운데서 적어도 2회 나타나고, 이에 따라 서열 오류를 함유할 가능성이 더 작은 것으로 간주됨을 의미한다. 이러한 3가지 기준을 조합하여, 각 결합 부위와 회합된 태그의 8가지 클래스를 만들었다: 가장 많이 안전한 (그러나 가장 덜 풍부한) 클래스는 오직 안전, 비공유, 2+ 태그만을 포함하고, 가장 덜 안전한 (그러나 가장 많이 풍부한) 클래스는 안전성, 공유여부, 또는 발생 횟수에 관계없이 모든 태그를 포함한다.
- [0120] 정규화된 발현 수준의 계산: 주문형 perl 코드를 사용하여 도 8a, 레벨 5-6에 지시된 단계를 시행하였다. 먼저, 각각의 유도된 샘플에서 획득한 태그 수를, 구축물 라이브러리에 대하여 이전에 계산된 결합 부위 대 전

사체 태그 표를 사용하여 각 결합 부위에 대해 집계하였다 (도 8c 참조). 이어서, 각 샘플에 대하여, 각 결합 부위에 대해 집계된 태그 수를 양성 대조군 샘플에 대해 집계된 태그 수로 나누어, 정규화된 발현 수준을 생성하였다. 이러한 계산과 관련된 추가의 고려사항은 다음을 포함한다:

[0121] 1. 각 샘플에서, 유효성-검사된 cDNA 유전자 서열 가운데서 결합 부위 대 전사체 태그 회합 표에서 찾아볼 수 없었던 "신규" 태그의 하위세트가 발견되었다. 이러한 태그는 후속 계산에서 무시하였다.

[0122] 2. 결합 부위 대 전사체 태그 회합 표에 상기 기재된 태그의 8개 클래스 각각에 대하여 상기 기재된 태그 수의 집계를 수행하였다. 구축물 라이브러리 내의 결합 부위가 중앙 서열에 유사한 서열은 빈번하게 생성하지만, 증가하는 개수의 미스매치를 갖는 서열은 점점 더 드물게 생성하는 편재성을 나타내기 때문에, 소수의 미스매치를 갖는 결합 부위는 일반적으로 태그의 큰 수로 집계되는 한편, 보다 많은 미스매치를 갖는 결합 부위는 보다 적은 수로 집계된다. 따라서, 일반적으로는 가장 안전한 태그 클래스의 사용이 바람직할 수 있으나, 2개 이상의 미스매치를 갖는 결합 부위의 평가는 결합 부위당 적은 수의 태그에 기초할 수 있을 것이며, 이로 인해 안전한 수 및 비는, 태그 그 자체는 보다 더 신뢰할만하더라도, 통계적으로 보다 덜 신뢰할만하게 되었다. 이러한 경우에, 모든 태그가 사용되었다. 이러한 고려사항은, n개 미스매칭 위치에 대해 별도로 집계된 태그 수의 개수

가 미스매칭 위치의 조합의 개수 ($\binom{L}{n} 3^n$)에 따라 상승하여, n에 따라 현저하게 증가하고; 따라서 상이한 개수(n)의 미스매치에 대해 집계된 태그 수의 평균 (도 2b, 2e, 및 도 9a 및 10b에 제시됨)은 $n \geq 2$ 에 대해 집계된 태그 수의 통계적으로 매우 큰 세트에 기초한다는 사실에 의거하여 약간 보정된다.

[0123] 3. 마지막으로, TALE 구축물 라이브러리에 구축된 결합 부위는 18bp였고, 태그 회합은 이들 18bp 서열에 기초하여 할당되었으나, 일부 실험은 18bp 구축물 결합 부위 영역 내의 중앙 14bp 또는 10bp 영역에 결합하도록 프로그램화된 TALE를 사용하여 수행하였다. 이러한 TALE에 대한 발현 수준의 계산에서, 태그를 회합 표 내의 18bp 결합 부위의 상응하는 영역에 기초하여 결합 부위에 집합시켜, 이러한 영역 외부의 결합 부위 미스매치는 무시하였다.

[0124] 실시예 V

[0125] Cas9_N-VP64를 사용하는 RNA-가이드된 SOX2 및 NANOG 조절

[0126] 본원에 기재된 sgRNA (암타머-변형된 단일 가이드 RNA) 테더링 접근법은 상이한 이펙터 도메인이 별개의 sgRNA에 의해 동원되는 것을 허용하며, 다만, 각각의 sgRNA는 상이한 RNA-단백질 상호작용 쌍을 사용하고, 동일한 Cas9_N-단백질을 사용하는 멀티플렉스 유전자 조절을 가능하게 하는 것이다. 도 12a SOX2 및 도 12b NANOG 유전자에 대하여, 전사 개시 부위의 상류 ~1kb DNA 스트레치를 표적화하는 10개의 gRNA를 설계하였다. DNase 과민성 부위는 녹색으로 강조 표시된다. 내인성 유전자의 qPCR을 통해 전사 활성화를 검증하였다. 두 경우에, 개별 gRNA의 도입은 전사를 중간 정도로 자극하였으나, 다중 gRNA는 수-배의 강한 전사 활성화를 상승작용적으로 자극하도록 작용하였다. 데이터는 평균 +/- SEM (N=3)이다. 도 12a-b에 도시된 바와 같이, 2개의 추가의 유전자, SOX2 및 NANOG는 프로모터 DNA의 상류 ~1kb 스트레치 이내를 표적화하는 sgRNA에 의해 조절되었다. 전사 개시 부위에 근접한 sgRNA가 강한 유전자 활성화를 발생시켰다.

[0127] 실시예 VI

[0128] Cas9-gRNA 복합체에 의한 표적화의 랜스케이프 평가

[0129] 도 2에 기재된 접근법을 이용하여, 2개의 추가의 Cas9-gRNA 복합체 (도 13a-c) 및 (도 13d-f)의 표적화 랜스케이프를 분석하였다. 2개의 gRNA는, gRNA2는 최대 2-3개의 미스매치를 허용하고 gRNA3은 최대 1개만을 허용하는 매우 상이한 특이성 프로파일을 갖는다. 이러한 측면은 1개의 염기 미스매치 (도 13b, 13e) 및 2개의 염기 미스매치 플롯 (도 13c, 13f) 둘 다에서 반영된다. 도 13c 및 13f에서, 데이터 디스플레이를 개선하기 위해 정규화된 발현 수준이 컬러 스케일(color scale)의 상한을 넘어서는 이상치인 미스매치 쌍을 별표 '*'가 함유된 황색 박스로 표시하면서, 정규화된 발현 수준을 계산하는데 이용가능한 데이터가 충분하지 않은 염기 미스매치 쌍을 'x'가 함유된 회색 박스로 표시한다. 통계적 유의성 기호는 다음과 같다: P<.0005/n에 대하여 ***, P<.005/n에 대하여 **, P<.05/n에 대하여 *, 및 P>=.05/n에 대하여 N.S. (비-유의함), 여기서 n은 비교물의 개수이다 (표 2 참조).

[0130] 실시예 VII

- [0131] 확인, 리포터 검정의 특이성
- [0132] 도 14a-c에 도시된 바와 같이, 특이성 데이터가 2가지 상이한 sgRNA:Cas9 복합체를 사용하여 생성되었다. 상응하는 돌연변이체 sgRNA가 리포터 라이브러리를 자극하지 못하였으므로, 이 검정이 평가 대상 sgRNA에 대해 특이적임을 확인하였다. 도 14a: 2개의 gRNA (야생형 및 돌연변이체; 상이한 서열은 적색으로 강조 표시됨)의 특이성 프로파일을, 야생형 gRNA 표적 서열에 대해 설계된 리포터 라이브러리를 사용하여 평가하였다. 도 14b: 상응하는 돌연변이체 gRNA가 리포터 라이브러리를 자극하지 못하였으므로, 이 검정이 평가 대상 gRNA에 대해 특이적임을 확인하였다 (데이터는 도 13d로부터 재플로팅됨). 통계적 유의성 기호는 다음과 같다: $P < .0005/n$ 에 대하여 ***, $P < .005/n$ 에 대하여 **, $P < .05/n$ 에 대하여 *, 및 $P \geq .05/n$ 에 대하여 N.S. (비-유의함), 여기서 n은 비교물의 개수이다 (표 2 참조). 상이한 sgRNA는 상이한 특이성 프로파일을 가질 수 있으며 (도 13a, 13d), 구체적으로, sgRNA2는 최대 3개의 미스매치를 허용하고, sgRNA3은 최대 1개만을 허용한다. 다른 위치에서의 미스매치가 또한 활성에 영향을 미치는 것으로 관찰되었을지라도, 미스매치에 대한 최대의 감도는 스페이서의 3' 말단에 국한되었다.
- [0133] 실시예 VIII
- [0134] 확인, 단일 및 이중-염기 gRNA 미스매치
- [0135] 도 15a-d에 도시된 바와 같이, 표적화 실험에 의해, 검정된 sgRNA에서의 스페이서의 3' 말단의 12 bp 내의 단일-염기 미스매치는 검출가능한 표적화를 발생시킨다는 것을 확인하였다. 그러나, 이 영역에서의 2 bp 미스매치는 활성의 유의한 손실을 초래하였다. 뉴클레아제 검정을 이용하여, 2개의 독립된 gRNA를 시험하였다: 스페이서 서열 대 표적에서 단일 또는 이중-염기 미스매치 (적색으로 강조 표시됨)를 보유하는 gRNA2 (도 15a-b) 및 gRNA3 (도 15c-d). 검정된 gRNA에서의 스페이서의 3' 말단의 12bp 내의 단일-염기 미스매치는 검출가능한 표적화를 발생시키지만, 이 영역에서의 2bp 미스매치는 빠른 활성 손실을 초래한다는 것을 확인하였다. 이들 결과는 도 13의 결과와 일치하는, 상이한 gRNA 사이의 특이성 프로파일의 차이를 더 강조한다. 데이터는 평균 +/- SEM (N=3)이다.
- [0136] 실시예 IX
- [0137] 확인, 5' gRNA 말단절단
- [0138] 도 16a-d에 도시된 바와 같이, 스페이서의 5' 부분에서의 말단절단은 sgRNA 활성을 유지하였다. 뉴클레아제 검정을 이용하여, 2개의 독립된 gRNA를 시험하였다: 스페이서의 5' 말단에서 말단절단된 gRNA1 (도 16a-b) 및 gRNA3 (도 16c-d). 1-3 bp의 5' 말단절단은 잘 허용되지만, 보다 큰 결실은 활성의 손실을 초래하는 것으로 관찰되었다. 데이터는 평균 +/- SEM (N=3)이다.
- [0139] 실시예 X
- [0140] 확인, 에스. 피오케네스 PAM
- [0141] 도 17a-b에 도시된 바와 같이, 뉴클레아제 매개 HR 검정을 이용하여, 에스. 피오케네스 Cas9에 대한 PAM가 NGG이고, 또한 NAG임을 확인하였다. 데이터는 평균 +/- SEM (N=3)이다. 추가의 연구에 따르면, 표적화 서열의 마지막 13 nt를 공유하는 대안적 NGG 표적을 갖지 않는 인간 엑손에서 생성된 약 190K Cas9 표적의 세트를 대안적 NAG 부위의 존재 여부 또는 이전의 13 nt에 미스매치를 갖는 NGG 부위에 대해 스캐닝하였다. 단지 0.4%만이 이러한 대안적 표적을 갖지 않는 것으로 밝혀졌다.
- [0142] 실시예 XI
- [0143] 확인, TALE 돌연변이
- [0144] 뉴클레아제 매개 HR 검정을 이용하여 (도 18a-b), 18량체 TALE가 그의 표적 서열에 다중 돌연변이를 허용한다는 것을 확인하였다. 도 18a-b에 도시된 바와 같이, 표적의 중앙부에서의 특정 돌연변이는, 뉴클레아제 검정에서 표적화된 실험을 통해 결정된 바와 같이 보다 높은 TALE 활성을 발생시켰다.
- [0145] 실시예 XII
- [0146] TALE 단량체 특이성 대 TALE 단백질 특이성
- [0147] 개개의 반복부-가변성 이잔기(diresidue) (RVD)의 역할을 분리시키기 위해, RVD의 선택이 염기 특이성에 기여하였으며, 또한 TALE 특이성이 전체적으로 단백질의 결합 에너지의 함수임을 확인하였다. 도 19a-c는 TALE 단량

체 특이성 대 TALE 단백질 특이성의 비교를 보여준다. 도 19a: 도 2에 기재된 접근법의 변형을 이용하여, 6개의 NI 또는 6개의 NH 반복부의 인접한 세트를 보유하는 2가지 14량체 TALE-TF의 표적화 랜드스케이프를 분석하였다. 이러한 접근법에서, 중앙부에 축중성 6량체 서열을 보유하는 리포터의 축소된 라이브러리를 만들어, TALE-TF 특이성 검정에 사용하였다. 도 19b-c: 두 경우에, 예상된 표적 서열이 풍부화된 것을 주목하였다 (즉, NI 반복부의 경우 6개의 A를 보유하고, NH 반복부의 경우 6개의 G를 보유함). 이러한 TALE는 각각 여전히 중앙의 6량체 표적 서열에 1-2개의 미스매치를 허용한다. 단량체의 선택이 염기 특이성에 기여하며, 또한 TALE 특이성은 전체적으로 단백질의 결합 에너지의 함수이다. 한 측면에 따르면, 보다 짧은 조작된 TALE 또는 고친화도 및 저친화도 단량체의 조성물을 보유하는 TALE는 게놈 조작 응용에서 보다 높은 특이성을 일으키고, 뉴클레아제 응용에서의 FokI 이량체화는 보다 짧은 TALE를 사용하는 경우에 오프-표적 영향을 더 감소시킨다.

[0148] 실시예 XIII

[0149] 오프셋 니킹, 천연 유전자좌

[0150] 도 20a-b는 오프셋 니킹과 관련된 데이터를 보여준다. 게놈-편집과 관련하여, 오프셋 니킹을 만들어 DSB를 생성하였다. 대다수의 니킹은 비-상동 말단 접합 (NHEJ) 매개 indel을 생성하지 않으며, 이에 따라 오프셋 니킹을 유도하는 경우에, 오프-표적 단일 니킹 사건은 아마도 매우 낮은 indel 비율을 생성할 것이다. 오프셋 니킹을 유도하여 DSB를 생성하는 것은 통합된 리포터 유전자좌 및 천연 AAVS1 게놈 유전자좌 둘 다에서 유전자 파괴를 유도하는데 효과적이다. 도 20a: 200bp DNA 스트레치를 커버하는 8개의 gRNA (4개는 센스 가닥을 표적화하고 (s1-4), 4개는 안티센스 가닥을 표적화함 (as1-4))를 갖는 천연 AAVS1 유전자좌를 표적화하였다. 상보적 가닥에 니킹이 생성된 Cas9D10A 돌연변이체를 사용하여, gRNA의 다양한 2-원 조합을 이용하여 소정 범위의 프로그램화된 5' 또는 3' 오버행을 유도하였다. 도 20b: 생어(Sanger) 서열분석 기반 검정을 이용하여, 단일 gRNA는 검출가능한 NHEJ 사건을 유도하지 않는 반면에, 오프셋 니킹을 유도하여 DSB를 생성하는 것은 유전자 파괴를 유도하는데 매우 효과적이라는 것을 관찰하였다. 주목할만한게, 5' 오버행을 생성하는 오프셋 니킹은 3' 오버행과는 대조적으로 보다 많은 NHEJ 사건을 발생시킨다. 생어 서열분석 클론의 개수가 막대 위에 강조 표시되고, 예측되는 오버행 길이가 상응하는 x-축 범례 아래 표시된다.

[0151] 실시예 XIV

[0152] 오프셋 니킹, NHEJ 프로파일

[0153] 도 21a-c는 오프셋 니킹 및 NHEJ 프로파일에 관한 것이다. 3가지 상이한 오프셋 니킹 조합의 대표적인 생어 서열분석 결과가 박스로 강조 표시된 표적화 gRNA의 위치와 함께 제시된다. 또한, 상동 재조합 (HR) 매개 복구를 위한 표준 모델과 일치하게, 오프셋 니킹을 통한 5' 오버행의 조합은 3' 오버행보다 더 강한 NHEJ 사건을 발생시켰다 (도 3b). NHEJ의 자극에 더하여, 강한 HR 유도가 5' 오버행이 생성되었을 때 관찰되었다. 3' 오버행의 생성은 HR 비율을 개선시키지 않았다 (도 3c).

[0154] 실시예 XV

[0155] 표 1

[0156] 내인성 유전자 조절을 위한 gRNA 표적

[0157] Cas9-gRNA 매개 활성화 실험에 사용되는 REX1, OCT4, SOX2 및 NANOG 프로모터 내의 표적이 열거된다.

gRNA 명칭	gRNA 표적
REX1 1	ctggcggatcactcgcggtt agg
REX1 2	cctcggcctccaaaagtgtc agg
REX1 3	acgctgattcctgcagatca ggg
REX1 4	ccaggaatacgtatccacca ggg
REX1 5	gccacacccaagcgatcaaa tgg
REX1 6	aaataatacattctaaggtg agg
REX1 7	gctactggggaggctgagggc agg
REX1 8	tagcaatacagtcacattaa tgg
REX1 9	ctcatgtgatcccccgctc cgg
REX1 10	ccgggcagagagtgaacgcg cgg
OCT4 1	ttccttcctcctcctccgtgct tgg
OCT4 2	tctctgcaaagcccctggag agg
OCT4 3	aatgcagttgccgagtgag tgg
OCT4 4	cctcagcctcctaaagtgtc ggg
OCT4 5	gagtccaaatcctctttact agg
OCT4 6	gagtgctctggatttgggata agg
OCT4 7	cagcacctcatctcccagtg agg
OCT4 8	tctaaaaaccagggatcat ggg
OCT4 9	cacaaggcagccagggatcc agg
OCT4 10	gatggcaagctgagaaacac tgg
OCT4 11	tgaaatgcacgcatacaatt agg
OCT4 12	ccagtcagacctggccttc tgg
OCT4 13	cccagaaaaacagaccctga agg
OCT4 14	aagggttgagcacttggtta ggg
OCT4 15	atgtctgagttttggttgag agg
OCT4 16	ggtcccttgaaggggaagta ggg
OCT4 17	tggcagtctactcttgaaga tgg
OCT4 18	ggcacagtgcagaggtctg tgg
OCT4 19	taaaaataaaaaaactaaca ggg
OCT4 20	tctgtgggggacctgcactg agg
OCT4 21	ggccagaggtcaaggctagt ggg
SOX2 1	cacgaccgaaacccttctta cgg
SOX2 2	gttgaatgaagcagcttag tgg
SOX2 3	taagaacagagcaagttacg tgg
SOX2 4	tgtaaggtaagagaggagag cgg
SOX2 5	tgacacaccaactcctgcac tgg
SOX2 6	tttaccactctccttcgaaa agg
SOX2 7	gtggctggcaggtggctct ggg
SOX2 8	ctccccggcctcccccgcg cgg
SOX2 9	caaaaccggcagcgaggct ggg
SOX2 10	aggagccgcccgcgctgat tgg
NANOG 1	cacacacacccacagat ggg
NANOG 2	gaagaagctaaagagccaga ggg
NANOG 3	atgagaatttcaataacctc agg
NANOG 4	tcccgtctctgttgcccaggc tgg
NANOG 5	cagacacccaccatgctg tgg
NANOG 6	tcccaatttactgggattac agg
NANOG 7	tgatttaaaagttgaaacg tgg
NANOG 8	tctagttccccacctagtct ggg
NANOG 9	gattaactgagaattcaca ggg
NANOG 10	cgccaggaggggtgggtcta agg

[0158]

[0159] 실시예 XVI

[0160] 표 2

[0161] Cas9-gRNA 및 TALE 특이성 데이터의 통계적 분석의 개요

[0162] 표 2(a) 특정한 개수의 표적 부위 돌연변이를 갖는 표적 서열에 결합하는 TALE 또는 Cas9-VP64 활성화계의 정규화된 발현 수준의 비교에 대한 P-값. 정규화된 발현 수준을 도면 칼럼 내에 표시된 도면 내의 박스플롯으로 표시하였고, 여기서 박스는 표적 부위로부터의 미스매치의 개수에 대한 이들 수준의 분포를 나타낸다. P-값을 각각의 박스플롯 내의 미스매치의 각 연속적 쌍의 개수에 대한 t-테스트 (t-테스트는 1 샘플 또는 2 샘플 t-테스트임)를 이용하여 계산하였다 (방법 참조). 통계적 유의성을 본페로니(Bonferroni)-보정된 P-값 역치 (보정은 각각의 박스플롯 내의 비교물의 개수에 기초함)를 사용하여 평가하였다. 통계적 유의성 기호는 다음과 같다: $P < .0005/n$ 에 대하여 ***, $P < .005/n$ 에 대하여 **, $P < .05/n$ 에 대하여 *, 및 $P \geq .05/n$ 에 대하여 N.S. (비-유의함), 여기서 n은 비교물의 개수이다. 표 2(b) 도 2d에서의 시드 영역의 통계적 특성화: 20bp 표적 부위의 3' 말단의 후보 시드 영역 내에서 돌연변이된 위치 쌍 대 모든 다른 위치 쌍에 대한, 2개의 돌연변이를 갖는 표적 서열에 결합하는 Cas9N VP64+gRNA에 대한 발현 값들 사이의 분리 정도를 나타내는 $\log_{10}(P\text{-값})$. 최대 $-\log_{10}(P\text{-값})$ (위에 강조 표시됨)에 의해 지시되는 최대 분리가 표적 부위의 마지막 8-9bp에서 발견된다. 이들 위치

는 이 표적 부위의 "시드" 영역이 시작되는 부분을 나타내는 것으로 해석될 수 있다. P-값을 계산하는 방법에 대한 정보에 대해 방법 내의 섹션 "시드 영역의 통계적 특성화"를 참조한다.

a					b				
도면	발현 수준 비교: 돌연변이 대 돌연변이		t-검정	P-값	기호	시드 출발 위치	위치 쌍 개수		-log10 P-값
	돌연변이	대 돌연변이					시드에 둘 다 존재	시드에 둘 다 부재	
2b	0	1	1-samp	7.8E-05	***	2	171	19	3.11
	1	2	2-samp	1.4E-06	***	3	153	37	1.46
	2	3	2-samp	4.0E-61	***	4	136	54	2.01
	3	4	2-samp	0	***	5	120	70	3.34
	4	5	2-samp	0	***	6	105	85	5.65
	5	6	2-samp	1.0E-217	***	7	91	99	7.34
	6	7	2-samp	1.7E-43	***	8	78	112	6.61
	7	8	2-samp	3.7E-02	N.S.	9	66	124	7.10
2e	0	1	1-samp	8.9E-01	N.S.	10	55	135	9.72
	1	2	2-samp	1.9E-06	***	11	45	145	9.83
	2	3	2-samp	5.0E-147	***	12	36	154	10.44
	3	4	2-samp	0	***	13	28	162	10.72
	4	5	2-samp	0	***	14	21	169	8.97
	5	6	2-samp	4.2E-62	***	15	15	175	5.61
	6	7	2-samp	1.6E-03	*	16	10	180	3.34
	7	8	2-samp	4.7E-01	N.S.	17	6	184	2.26
S7a	0	1	1-samp	5.2E-02	N.S.	18	3	187	1.16
	1	2	2-samp	2.8E-05	***				
	2	3	2-samp	3.5E-21	***				
	3	4	2-samp	1.4E-58	***				
	4	5	2-samp	8.3E-101	***				
	5	6	2-samp	6.8E-94	***				
	6	7	2-samp	1.8E-61	***				
	7	8	2-samp	8.1E-24	***				
S7d 및 S8d	0	1	1-samp	2.3E-18	***				
	1	2	2-samp	2.4E-08	***				
	2	3	2-samp	6.2E-54	***				
	3	4	2-samp	4.0E-141	***				
	4	5	2-samp	1.9E-20	***				
	5	6	2-samp	1.2E-03	*				
	6	7	2-samp	3.8E-05	***				
	7	8	2-samp	9.4E-01	N.S.				
S8c	0	1	1-samp	7.2E-03	N.S.				
	1	2	2-samp	5.0E-01	N.S.				
	2	3	2-samp	3.9E-84	***				
	3	4	2-samp	8.5E-153	***				
	4	5	2-samp	8.6E-76	***				
	5	6	2-samp	1.6E-03	*				
	6	7	2-samp	7.1E-01	N.S.				
	7	8	2-samp	7.8E-02	N.S.				
S13a (좌측)	0	1	1-samp	7.3E-01	N.S.				
	1	2	2-samp	2.4E-06	***				
	2	3	2-samp	7.2E-140	***				
	3	4	2-samp	0	***				
	4	5	2-samp	0	***				
	5	6	2-samp	1.0E-72	***				
S13a (중앙)	6	7	2-samp	4.0E-03	*				
	0	1	1-samp	9.4E-02	N.S.				
	1	2	2-samp	5.2E-09	***				
	2	3	2-samp	7.9E-86	***				
	3	4	2-samp	2.9E-53	***				
S13a (우측)	4	5	2-samp	3.5E-10	***				
	0	1	1-samp	1.3E-13	***				
	1	2	2-samp	1.1E-04	***				
	2	3	2-samp	3.7E-08	***				

[0163]

[0164] 실시예 XVII

[0165] 실시예에서의 단백질 및 RNA의 서열

[0166] A. m4 돌연변이체를 기초로 하는 Cas9_N-VP64 활성인자 구축물의 서열을 하기 나타내었다. 3개 버전을 구축하였고 Cas9_{m4}^{VP64} 및 Cas9_{m4}^{VP64}_N 융합 단백질 포맷이 가장 높은 활성을 보여주었다. m3 및 m2 돌연변이체에 상응하는 벡터 (도 4a)를 또한 구축하였다 (NLS 및 VP64 도메인을 강조함).

[0167]

>Cas9_{m4}^{VP64}

gccaccATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAAACAGCGTCGGCTGG
 GCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAAT
 ACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCCTGTTCTGACTCCGGGG
 AGACGGCCGAAGCCACGCGGCTCAAAGAACAGCACGGCGCAGATATACCCGCAGAA
 AGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGA
 CTCTTTCTTCATAGGCTGGAGGAGTCCTTTTGGTGGAGGAGGATAAAAAGCACGAG
 CGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAA
 CCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAAGGCTGACTTGCAGTT
 GATCTATCTCGCGCTGGCGCATATGATCAAATTTGGGGGACACTTCCTCATCGAGGGG
 GACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTATCCAAGTGGTTCAGACTT
 ACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAA
 TCCTGAGCGCTAGGCTGTCCAAATCCCAGCGGCTCGAAAACTCATCGCACAGCTCCC
 TGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACC
 CCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAG
 ACACCTACGATGATGATCTCGACAATCTGCTGGCCAGATCGGCGACCAGTACGCAGA
 CCTTTTTTTGGCGGCAAAGAACCTGTGACAGCCATTCTGCTGAGTGATATTCTGCGAG
 TGAACACGGAGATCACAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATG
 AGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTGACAGCAACTGCCTGAGAA
 GTACAAGGAAATTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGC
 GGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGAAAAAATGGAC
 GGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGC
 ACTTTCGACAATGGAAGCATCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCC
 TCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAAGATTGAGA
 AAATCCTCACATTTTCGATACCCACTATGTAGGCCCCCTCGCCGGGGAAATTCAG
 ATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAAGTTCGAGGAA
 GTCGTGGATAAGGGGGCCTCTGCCAGTCCTTCATCGAAAGGATGACTAACTTTGATA
 AAAATCTGCCTAACGAAAAGGTGCTTCCTAAACACTCTCTGCTGTACGAGTACTTCAC
 AGTTTATAACGAGCTACCAAGGTCAAATACGTACAGAAAGGATGAGAAAGCCAGC
 ATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGG
 AAAGTTACCGTGAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACT

[0168]

CTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGA
TCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT
TCTTGAGGACATTGTCTCACCTTACGTTGTTGAAGATAGGGAGATGATTGAAGAA
CGCTTGAAAACCTTACGCTCATCTCTTCGACGACAAAGTCATGAAACAGCTCAAGAGGC
GCCGATATACAGGATGGGGGCGGCTGTCAAGAAAACGTCAATGGGATCCGAGACA
AGCAGAGTGGAAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAA
CTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAAGCA
CAAGTTTCTGGCCAGGGGGACAGTCTTACGAGCACATCGCTAATCTTGCAGGTAGCC
CAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTCGTGGATGAACTCGTCAAAG
TAATGGGAAGGCATAAGCCCAGAAATATCGTTATCGAGATGGCCCCGAGAGAACCCAAA
CTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAGGATTGAAGAGGGT
ATAAAAGAAGTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTT
CAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTGGAT
CAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGCTGCTATCGTGCCCCAGT
CTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTGACAAGATCCGATAAAgcTAGA
GGGAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGG
CGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAG
GCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAAGGCAGCTTG
TTGAGACACGCCAGATCACCAGCACGTGGCCCAAATTCGATTACGCATGAACAC
CAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCT
AAGCTGGTCTCAGATTTAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACA
ATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAA
AAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTT
AGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTC
TTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGA
TTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACA
AGGGTAGGGATTTCCGACAGTCCGGAAGGTCCTGTCCATGCCGACGGTGAACATCGT
TAAAAAGACCGAAGTACAGACCCGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAG
GAACAGCGACAAGCTGATCGCACGCAAAAAGATTGGGACCCCAAGAAATACGGCGG
ATTGATTTCTCTACAGTCGTTACAGTGTACTGGTTGTGGCCAAAAGTGGAGAAAAGG
AAGTCTAAAAAACTCAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGA
TCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGGCGAAAGGATATAAAGAGGTC
AAAAAAGACCTCATCATTAAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAAAACGGCC
GGAAACGAATGCTCGTAGTGGGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGC
CCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT
CCCGAAGATAATGAGCAGAAGCAGCTGTTCTGTGGAACAACACAAACACTACCTTGAT

[0169]

GAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAAC
CTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGG
CAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCTGCAGCCTTCAA
GTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGAGGTCCTGGA
CGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCT
CAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCAG
CGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGAC
GCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCT
CGACATGCTCGGCAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTAACCTCT
AGATGA

[0170]

[0171] >Cas9_{m4}^{VP64}N 서열

gcccaccATGCCCAAGAAGAAGAGGAAGGTGGGAAGGGGGATGGACAAGAAGTACTCCA
 TTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGACGAGTACAA
 GGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAA
 GAACCTCATTGGCGCCCTCCTGTTTCGACTCCGGGGAGACGGCCGAAGCCACGCGGCTC
 AAAAGAACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAG
 GAGATCTTTAGTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGG
 AGTCCTTTTTGGTGGAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATAT
 CGTGGACGAGGTGGCGTACCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAA
 GCTTGTAGACAGTACTGATAAGGCTGACTTTCGGTTGATCTATCTCGCGCTGGCGCAT
 ATGATCAAATTTTCGGGGACACTTCCTCATCGAGGGGGACCTGAACCCAGACAACAGC
 GATGTCGACAAACTCTTTATCCAACCTGTTTCAGACTTACAATCAGCTTTTCGAAGAGA
 ACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAATCCTGAGCGCTAGGCTGTCCAA
 ATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGGGAGAAGAAGAACGGCCT
 GTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCAACTTTAAATCTAACTTCG
 ACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATGATGATCTCG
 ACAATCTGCTGGCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGGCAAAGAA
 CCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAGTGAACACGGAGATCACCAAA
 GCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTT
 TGCTGAAGGCCCTTGTGAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTTCTTCGA
 TCAGTCTAAAAATGGCTACGCCGATACATTGACGGCGGAGCAAGCCAGGAGGAATT
 TTACAAATTTATTAAGCCCATCTTGGAAAAAATGGACGGCACCGAGGAGCTGCTGGTA
 AAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATC

[0172]

CCCCACCAGATTCACCTGGGCGAACTGCACGCTATCCTCAGGCGGCAAGAGGATTTCT
 ACCCCTTTTTGAAAGATAACAGGGAAAAAGATTGAGAAAATCCTCACATTTCCGATACC
 CTAATATGTAGGCCCCCTCGCCCGGGGAAATTCCAGATTCGCGTGGATGACTCGCAA
 TCAGAAGAGACCATCACTCCCTGGAACCTCGAGGAAGTCGTGGATAAGGGGGCCTCT
 GCCCAGTCCTTCATCGAAAGGATGACTAACTTTGATAAAAATCTGCCTAACGAAAAGG
 TGCTTCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTTATAACGAGCTCACCAAG
 GTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCTGTCTGGAGAGCAGAAG
 AAAGCTATCGTGGACCTCTCTTCAAGACGAACCGGAAAAGTTACCGTGAAACAGCTCA
 AAGAAGACTATTTCAAAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAGTGGA
 GGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGAC
 AAGGACTTCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCACCC
 TTACGTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACCTTACGCTCATCT
 CTTGACGACAAAAGTCATGAAACAGCTCAAGAGGCGCCGATATACAGGATGGGGGGCG
 GCTGTCAAGAAAAGTCAATGGGATCCGAGACAAGCAGAGTGGAAGACAATCCT
 GGATTTTCTTAAAGTCCGATGGATTTGCCAACCAGGAACTTCATGCAGTTGATCCATGATG
 ACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAG
 TCTTACGAGCACATCGCTAATCTTGCAGGTAGCCAGCTATCAAAAAGGGAATACTG
 CAGACCGTTAAGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAG
 AATATCGTTATCGAGATGGCCCGAGAGAACCAAACTACCCAGAAGGGACAGAAGAAC
 AGTAGGGAAAGGATGAAGAGGATTGAAGAGGGTATAAAAAGAACTGGGGTCCCAAAT
 CCTTAAAGGAACACCCAGTTGAAAACACCCAGCTTCAGAATGAGAAGCTCTACCTGTAC
 TACCTGCAGAACGGCAGGGACATGTACGTGGATCAGGAACTGGACATCAATCGGCTC
 TCCGACTACGACGTGGCTGCTATCGTGCCCCAGTCTTTTCTCAAAGATGATTCTATTGA
 TAATAAAGTGTGACAAGATCCGATAAAgcTAGAGGGAAGAGTGATAACGTCCCCTCA
 GAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTGCTGAACGCCAAACTG
 ATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGGCCTGTCTGAGT
 TGGATAAAGCCGGCTTCATCAAAAAGGCAGCTTGTTGAGACACGCCAGATCACCAAGC
 ACGTGGCCCAAATTCCTCGATTACGCATGAACACCAAGTACGATGAAAATGACAAACT
 GATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTAGAAAAG
 GACTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCT
 ACCTGAATGCAGTGGTAGGCACTGCACTTATCAAAAAATATCCCAAGCTTGAATCTGA
 ATTTGTTTACGGGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAG
 CAGGAAATAGGCAAGGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTT
 TCAAGACCGAGATTACACTGGCCAATGGAGAGATTCGGAAGCGACCACTTATCGAAA
 CAAACGGAGAAACAGGAGAAATCGTGTGGGACAAGGGTAGGGATTTTCGCGACAGTCC
 GGAAGGTCCTGTCCATGCCGAGGTGAACATCGTTAAAAAGACCGAAGTACAGACCG

[0173]

GAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAAGGAACAGCGACAAGCTGATCGCAC
 GCAAAAAAGATTGGGACCCCAAGAAATACGGCGGATTTCGATTCTCTACAGTCGCTTA
 CAGTGTACTGGTTGTGGCCAAAGTGAGAAAAGGAAGTCTAAAAAACTCAAAAAGCGT
 CAAGGAACTGCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAAAAAACCCCAT
 CGACTTTCTCGAGGCGAAAAGGATATAAAGAGGTCAAAAAAGACCTCATCATTAAAGCTT
 CCCAAGTACTCTCTTTGAGCTTGAAAACGGCCGAAACGAATGCTCGCTAGTGCGG
 GCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTA
 TCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCTCCCGAAGATAATGAGCAGAAGCA
 GCTGTTCGTGGAACAACACAAACTACCTTGATGAGATCATCGAGCAAATAAGCGA
 ATTCTCCAAAAGAGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTCTGCTTAC
 AATAAGCACAGGGATAAGCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTT
 ACTCTGACCAACTTGGGCGCGCTGCAGCCTTCAAGTACTTCGACACCACCATAGACA
 GAAAGCGGTACACCTCTACAAAGGAGGTCTTGACGCCCACTGATTCATCAGTCAAT
 TACGGGGCTCTATGAAACAAGAATCGACCTCTCTCAGCTCGGTGGAGACAGCAGGGCT
 GACCCCAAGAAGAAGAGGAAGGTGGAGGCCAGCGTTCCGGACGGGCTGACGCATTG
 GACGATTTTGATCTGGATATGCTGGGAAGTGACGCCCTCGATGATTTTGACCTTGACA
 TGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGACATGCTCGGCAGTGACGCCCTT
 GATGATTTGACCTGGACATGCTGATTAACTCTAGATGA

[0174]

[0175] >Cas9_{m4}^{VP64}C

gccaccATGGACAAGAAGTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGG
 GCCGTCATTACGGACGAGTACAAGGTGCCGAGCAAAAAATTCAAAGTTCTGGGCAAT
 ACCGATCGCCACAGCATAAAGAAGAACCTCATTGGCGCCCTCTGTTGACTCCGGGG
 AGACGGCCGAAGCCACGCGGCTCAAAGAACAGCACGGCGCAGATATACCCGCAGAA
 AGAATCGGATCTGCTACCTGCAGGAGATCTTTAGTAATGAGATGGCTAAGGTGGATGA
 CTCTTTCTTCATAGGCTGGAGGAGTCCTTTTTGGTGGAGGAGGATAAAAAAGCACGAG
 CGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTACCATGAAAAGTACCCAA
 CCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGGCTGACTTGCGGTT
 GATCTATCTCGCGCTGGCGCATATGATCAAATTCGGGGACACTTCCTCATCGAGGGG
 GACCTGAACCCAGACAACAGCGATGTCGACAACTCTTTATCCAAGTTGACTTTCAGACTT
 ACAATCAGCTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAA
 TCCTGAGCGCTAGGCTGTCCAAATCCCGGCGGCTCGAAAACCTCATCGCACAGCTCCC
 TGGGGAGAAGAAGAACGGCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACC
 CCCAACTTTAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAG
 ACACCTACGATGATGATCTCGACAATCTGCTGGCCAGATCGGCGACCAGTACGCAGA
 CCTTTTTTTGGCGGCAAAGAACCTGTCAGACGCCATTCTGCTGAGTGATATTCTGCGAG

[0176]

TGAACACGGAGATCACCAAAGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATG
 AGCACCACCAAGACTTGACTTTGCTGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAA
 GTACAAGGAAATTTTCTTCGATCAGTCTAAAAATGGCTACGCCGGATACATTGACGGC
 GGAGCAAGCCAGGAGGAATTTTACAAATTTATTAAGCCCATCTTGAAAAAATGGAC
 GGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAGATCTGTTGCGCAAACAGCGC
 ACTTTCGACAATGGAAGCATCCCCACCAGATTCACCTGGGCGAACTGCACGCTATCC
 TCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGAAAAGATTGAGA
 AAATCCTCACATTTCCGATAACCTACTATGTAGGCCCCCTCGCCCGGGGAAATCCAG
 ATTCGCGTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACCTTCGAGGAA
 GTCGTGGATAAGGGGGCCTCTGCCAGTCCTTCATCGAAAGGATGACTAACTTTGATA
 AAAATCTGCCTAACGAAAAGGTGCTTCCCTAAACTCTCTGCTGTACGAGTACTTCAC
 AGTTTATAACGAGCTCACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGC
 ATTCCTGTCTGGAGAGCAGAAGAAAGCTATCGTGGACCTCCTTTCAAGACGAACCGG
 AAAGTTACCGTGAAACAGCTCAAAGAAGACTATTTCAAAAAGATTGAATGTTTCGACT
 CTGTTGAAATCAGCGGAGTGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGA
 TCTCCTGAAAATCATTAAAGACAAGGACTTCCTGGACAATGAGGAGAACGAGGACAT
 TCTTGAGGACATTGTCTCACCCCTTACGTTGTTGAAGATAGGGAGATGATTGAAGAA
 CGCTTGAAAACCTACGCTCATCTCTTCGACGACAAAAGTCATGAAACAGCTCAAGAGGC
 GCCGATATACAGGATGGGGGGCGGCTGTCAAGAAAAGTCAATGGGATCCGAGACA
 AGCAGAGTGGAAAGACAATCCTGGATTTTCTTAAGTCCGATGGATTTGCCAACCGGAA
 CTTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGGACATCCAGAAAGCA
 CAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTTGCAGGTAGCC
 CAGCTATCAAAAAGGGAATACTGCAGACCGTTAAGGTTCGTGGATGAACTCGTCAAAG
 TAATGGGAAGGCATAAGCCCAGAAATATCGTTATCGAGATGGCCCGAGAGAACCAAA
 CTACCCAGAAGGGACAGAAGAACAGTAGGGAAAAGGATGAAGAGGATTGAAGAGGGT
 ATAAAAGAAGTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACACCCAGCTT
 CAGAATGAGAAGCTCTACCTGTAACCTGCAGAACGGCAGGGACATGTACGTGGAT
 CAGGAACTGGACATCAATCGGCTCTCCGACTACGACGTGGCTGCTATCGTGCCCCAGT
 CTTTTCTCAAAGATGATTCTATTGATAATAAAGTGTGACAAGATCCGATAAAgcTAGA
 GGGAAAGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGG
 CGGCAGCTGCTGAACGCCAAACTGATCACACAACGGAAGTTCGATAATCTGACTAAG
 GCTGAACGAGGTGGCCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAAGGCAGCTTG
 TTGAGACACGCCAGATCACCAAGCACGTGGCCCAAATTCGATTACGCATGAACAC
 CAAGTACGATGAAAATGACAAACTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCT
 AAGCTGGTCTCAGATTTTCAGAAAGGACTTTCAGTTTTATAAGGTGAGAGAGATCAACA
 ATTACCACCATGCGCATGATGCCTACCTGAATGCAGTGGTAGGCACTGCACTTATCAA

[0177]

AAAAATATCCCAAGCTTGAATCTGAATTTGTTTACGGAGACTATAAAGTGTACGATGTT
 AGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAAGGCCACCGCTAAGTACTTC
 TTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACACTGGCCAATGGAGAGA
 TTCGGAAGCGACCACTTATCGAAACAAACGGAGAAACAGGAGAAATCGTGTGGGACA
 AGGGTAGGGATTTTCGCGACAGTCCGGAAGGTCCTGTCCATGCCGCAGGTGAACATCGT
 TAAAAAGACCGAAGTACAGACCCGGAGGCTTCTCCAAGGAAAGTATCCTCCCGAAAAG
 GAACAGCGACAAGCTGATCGCACGCAAAAAAGATTGGGACCCCAAGAAATACGGCGG
 ATTCGATTCTCCTACAGTCGTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGG
 AAGTCTAAAAAACTCAAAAGCGTCAAGGAACTGCTGGGCATCACAATCATGGAGCGA
 TCAAGCTTCGAAAAAAACCCCATCGACTTTCTCGAGGGCGAAAAGGATATAAAAGAGGTC
 AAAAAAGACCTCATCATTAAGCTTCCCAAGTACTCTCTCTTTGAGCTTGAACCGCC
 GGAAACGAATGCTCGTAGTGC GGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGC
 CCTCTAAATACGTTAATTTCTTGTATCTGGCCAGCCACTATGAAAAGCTCAAAGGGTCT
 CCCGAAGATAATGAGCAGAAGCAGCTGTTTCGTGGAACAACACAAACACTACCTTGAT
 GAGATCATCGAGCAAATAAGCGAATTCTCCAAAAGAGTGATCCTCGCCGACGCTAAC
 CTCGATAAGGTGCTTTCTGCTTACAATAAGCACAGGGATAAGCCCATCAGGGAGCAGG
 CAGAAAAATTATCCACTTGTTTACTCTGACCAACTTGGGCGCGCCTGCAGCCTTCAA
 GTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAAGGAGGTCCTGGA
 CGCCACACTGATTATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACCTCTCT
 CAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGGAGGCCAG
 CGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGAC
 GCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCT
 CGACATGCTCGGCAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTAAGTCT
 AGAGCGGCCGCAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAA
 GGTAGATCCAAAAAAGAAGAGAAAGGTAGATACGGCCGCATAG

[0178]

[0179] B. MS2-활성인자 구축물의 서열, 및 2X MS2 압타머 도메인을 갖는 상응하는 gRNA 백본 벡터를 하기 제공하였다

[0180] (NLS, VP64, gRNA 스페이서, 및 MS2-결합 RNA 스템 루프 도메인을 강조함). 전자의 2개 버전을 구축하였고 MS2_{VP64N} 융합 단백질 포맷이 가장 높은 활성을 보여주었다.

[0181] >MS2_{VP64N}

gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTTCTAGAAATGGCTTCTAA
 CTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTGCCCCA
 AGCAACTTCGCTAACGGGATCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTT
 ACAAAGTAACCTGTAGCGTTTCGTAGAGCTCTGCGCAGAATCGCAAATACACCATCAA
 AGTCGAGGTGCCTAAAGGCGCCTGGCGTTCGACTTAAATATGGAACCTAACCATTCCA
 ATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCTCCTAA
 AAGATGGAAACCCGATTCCTCAGCAATCGCAGCAAACTCCGGCATCTACGAGGCCA
 GCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGA
 CGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACC
 TCGACATGCTCGGCAGTGACGCCCTTGATGATTTTCGACCTGGACATGCTGATTAAGTCT
 AGATGA

[0182]

[0183] >MS2_{VP64C}

gccaccATGGGACCTAAGAAAAAGAGGAAGGTGGCGGCCGCTTCTAGAATGGCTTCTAA
 CTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCA
 AGCAACTTCGCTAACGGGATCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTT
 ACAAAGTAACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAATCGCAAATACACCATCAA
 AGTCGAGGTGCCTAAAGGCGCCTGGCGTTCGTAATAATATGGAACCTAACCATTCCA
 ATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGTCTCCTAA
 AAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACTCCGGCATCTACGAGGCCA
 GCGGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGA
 CGCCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACC
 TCGACATGCTCGGCAGTGACGCCCTTGATGATTTGACCTGGACATGCTGATTAACCTCT
 AGAGCGGCCCGCAGATCCAAAAAGAAGAGAAAGGTAGATCCAAAAAGAAGAGAAA
 GGTAGATCCAAAAAGAAGAGAAAGGTAGATACGGCCGCATAG

[0184]

[0185] >gRNA_{2MS2}

TGTACAAAAAAGCAGGCTTTAAAGGAACCAATTCAGTCGACTGGATCCGGTACCAAG
 GTCGGGCAGGAAGAGGGCCTATTTCCCATGATTCCTTCATATTTGCATATACGATACA
 AGGCTGTTAGAGAGATAATTAGAATTAATTTGACTGTAAACACAAAGATATTAGTACA
 AAATACGTGACGTAGAAAGTAATAATTTCTTGGGTAGTTTGCAGTTTTAAAAATTATGTT
 TAAAAATGGACTATCATATGCTTACCGTAACTTGAAAGTATTTGATTTCTTGGCTTTA
 TATATCTTGTGGAAGGACGAAACACCGNNNNNNNNNNNNNNNNNNNNNGTTTTAGAG
 CTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC
 GAGTCGGTGTCTGTCAGGTCGACTCTAGA**AAACATGAGGATCACCCATGTCT**GTCAGTA
 TTCCCGGGTTCATTAGATCCTAAGGTACCTAATTGCCTAGAA**AAACATGAGGATCACCC**
ATGTCTGTCAGGTCGACTCTAGAAATTTTTTCTAGAC

[0186]

[0187] C. d토마토 형광 기반 전사 활성화 리포터 서열을 하기 열거하였다

[0188] (**ISceI** 대조-TF 표적, **gRNA** 표적, **minCMV** 프로모터 및 **FLAG** 태그 + **d 토마토** 서열 을 강조함).

[0189] >TF 리포터 1

TAGGGATAACAGGGTAATAGTGTCCCCTCCACCCACAGTGGGGCGAGGTAGGCGTG
TACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAG
AATTCgcccaccatgGACTACAAGGATGACGACGATAAAACTTCCGGTGGCGGACTGGGTTC
CACCGTGAGCAAGGGCGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCAT
GGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCC
CTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTT
CGCCTGGGACATCCTGTCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGAAGCAC
CCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGC
GCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCA
GGACGGCACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCGACGG
CCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCC
CCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACGGCGG
CCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCC
GGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCA
TCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCTGTACGGCATGGA
CGAGCTGTACAAGTAA

[0190]

[0191] >TF 리포터 2

TAGGGATAACAGGGTAATAGTGGGGCCACTAGGGACAGGATTGGCGAGGTAGGCGTG
TACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAG
AATTCgcccaccatgGACTACAAGGATGACGACGATAAAACTTCCGGTGGCGGACTGGGTTC
CACCGTGAGCAAGGGCGAGGAGGTCATCAAAGAGTTCATGCGCTTCAAGGTGCGCAT
GGAGGGCTCCATGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCC
CTACGAGGGCACCCAGACCGCCAAGCTGAAGGTGACCAAGGGCGGCCCCCTGCCCTT
CGCCTGGGACATCCTGTCCCCCAGTTCATGTACGGCTCCAAGGCGTACGTGAAGCAC
CCCGCCGACATCCCCGATTACAAGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGC
GCGTGATGAACTTCGAGGACGGCGGTCTGGTGACCGTGACCCAGGACTCCTCCCTGCA
GGACGGCACGCTGATCTACAAGGTGAAGATGCGCGGCACCAACTTCCCCCCGACGG
CCCCGTAATGCAGAAGAAGACCATGGGCTGGGAGGCCTCCACCGAGCGCCTGTACCC
CCGCGACGGCGTGCTGAAGGGCGAGATCCACCAGGCCCTGAAGCTGAAGGACGGCGG
CCACTACCTGGTGGAGTTCAAGACCATCTACATGGCCAAGAAGCCCGTGCAACTGCC
GGCTACTACTACGTGGACACCAAGCTGGACATCACCTCCCACAACGAGGACTACACCA
TCGTGGAACAGTACGAGCGCTCCGAGGGCCGCCACCACCTGTTCTGTACGGCATGGA
CGAGCTGTACAAGTAA

[0192]

[0193] D. TALE 및 Cas9-gRNA 특이성 검정에 사용된 리포터 라이브러리의 일반적 포맷을 하기 제공하였다

[0194] (ISceI 대조-TF 표적, gRNA/TALE 표적 부위 (gRNA의 경우 23bp 및 TALE의 경우 18bp),

[0195] minCMV 프로모터, RNA 바코드, 및 d 토마토 서열을 강조함).

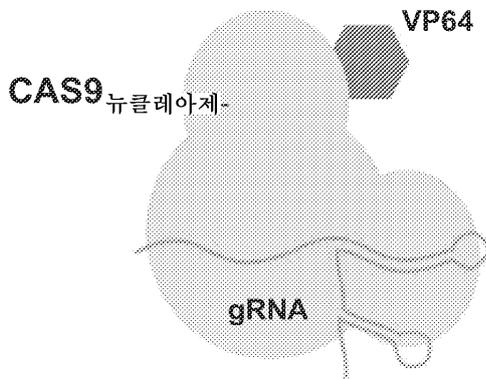
[0196] > 특이성 리포터 라이브러리

TAGGGATAACAGGGTAATAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNCGAGGTAGGCGT
 GTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGA
 GAATTCgcccaccatgGACTACAAGGATGACGACGATAAANNNNNNNNNNNNNNNNNNNNN
 NNNNACTTCCGGTGGCGGACTGGGTTCCACCGTGAGCAAGGGCGAGGAGGTCATCAA
 AGAGTTCATGCGCTTCAAGGTGCGCATGGAGGGCTCCATGAACGGCCACGAGTTCGA
 GATCGAGGGCGAGGGCGAGGGCCGCCCTACGAGGGCACCCAGACCGCCAAGCTGAA
 GGTGACCAAGGGCGGCCCTGCCCTTCGCCTGGGACATCCTGTCCCCCAGTTCATG
 TACGGCTCCAAGGCGTACGTGAAGCACCCCGCCGACATCCCCGATTACAAGAAGCTGT
 CCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGACGGCGGTCTGGT
 GACCGTGACCCAGGACTCCTCCCTGCAGGACGGCACGCTGATCTACAAGGTGAAGAT
 GCGCGGCACCAACTTCCCCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTG
 GGAGGCCTCCACCGAGCGCCTGTACCCCGCGACGGCGTGCTGAAGGGCGAGATCCA
 CCAGGCCCTGAAGCTGAAGGACGGCGGCCACTACCTGGTGGAGTTCAAGACCATCTA
 CATGGCCAAGAAGCCCGTGCAACTGCCCGGCTACTACTACGTGGACACCAAGCTGGA
 CATCACCTCCCACAACGAGGACTACCCATCGTGGAACAGTACGAGCGCTCCGAGGG
 CCGCCACCACCTGTTCCTGTACGGCATGGACGAGCTGTACAAGTAAGAATTC

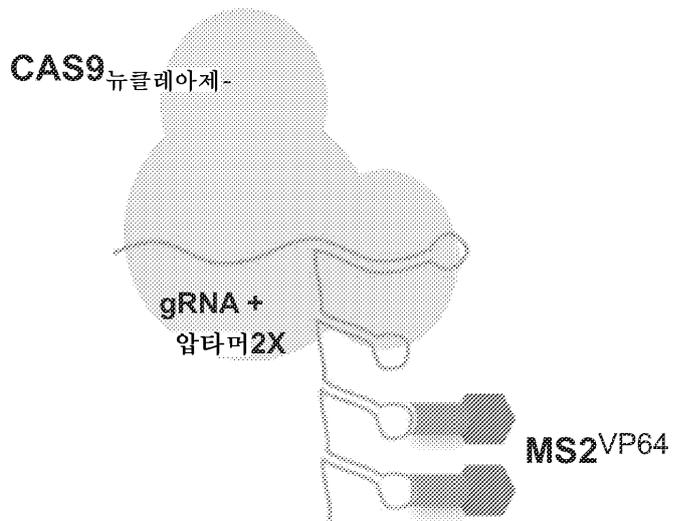
[0197]

도면

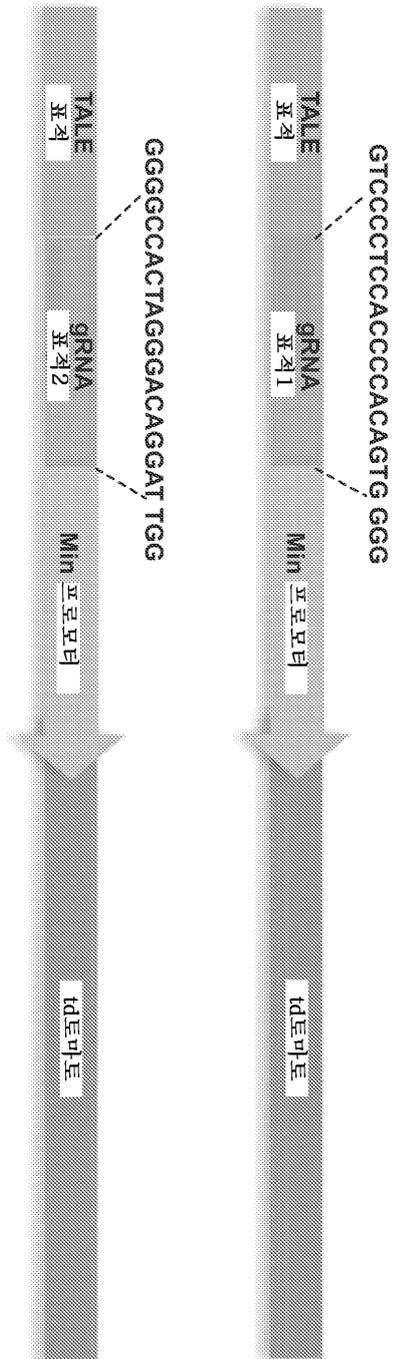
도면 1a



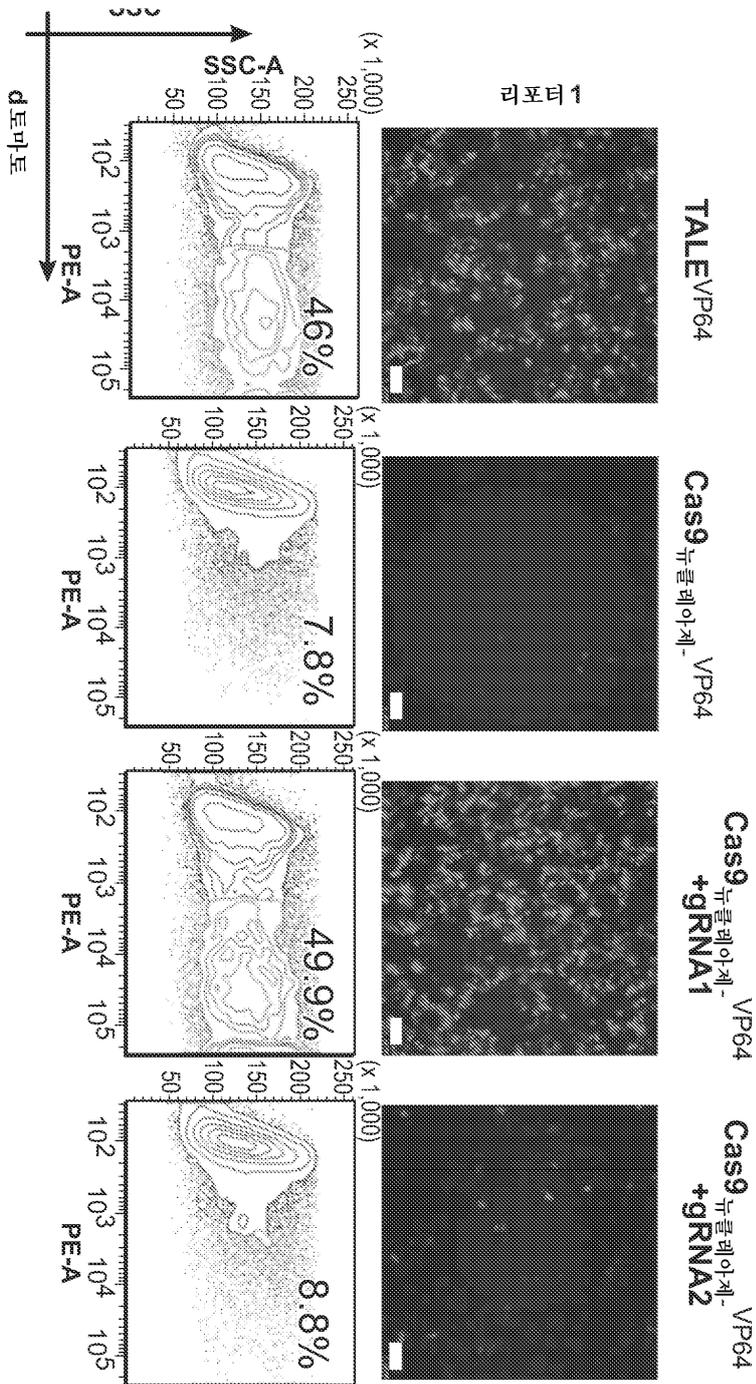
도면1b



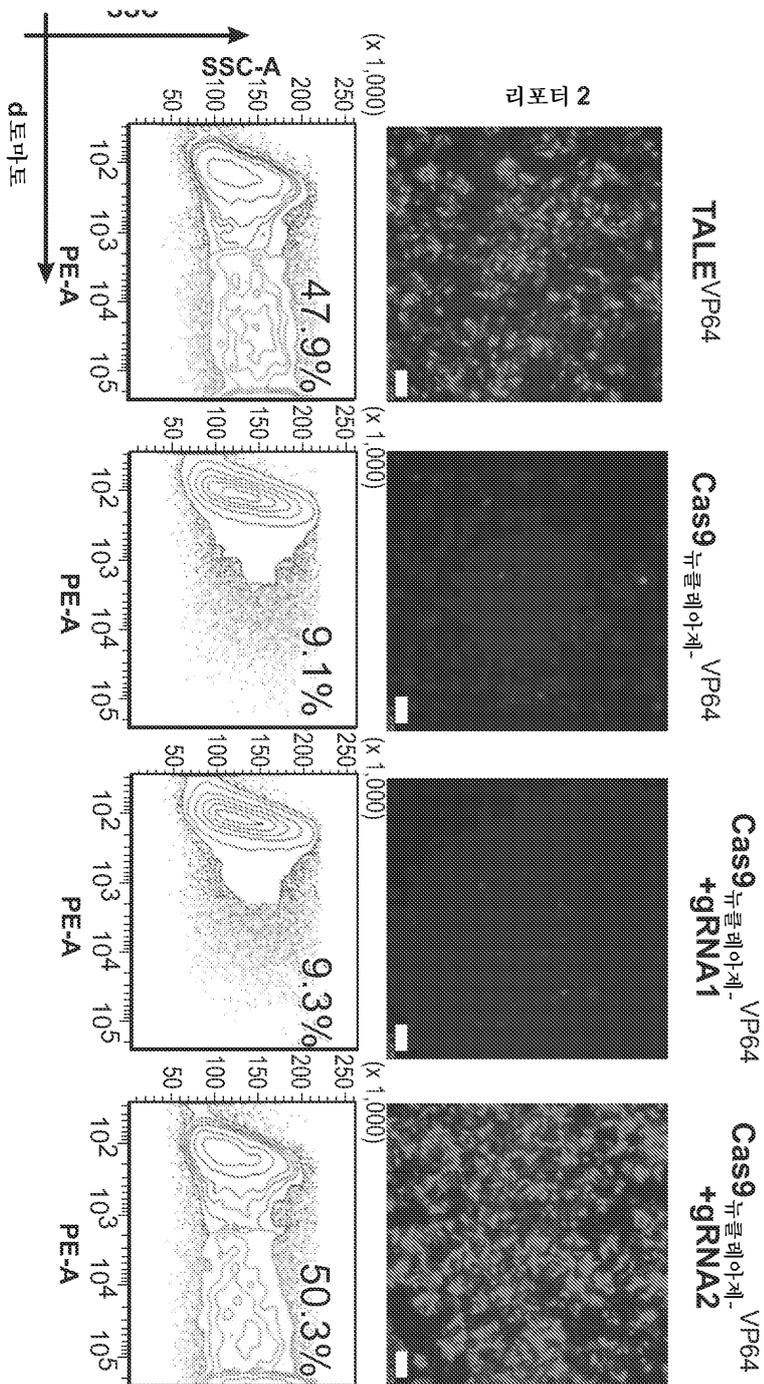
도면1c



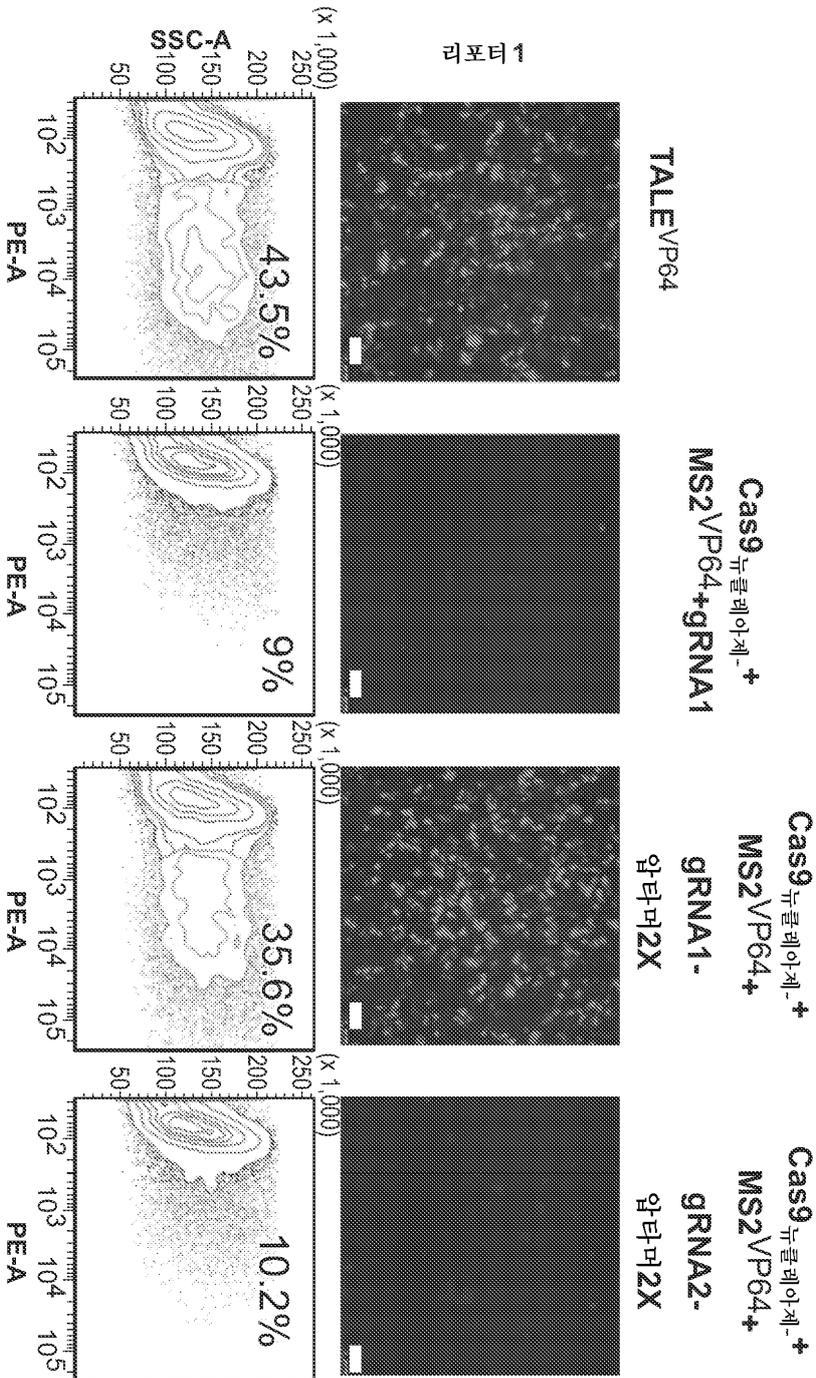
도면1da



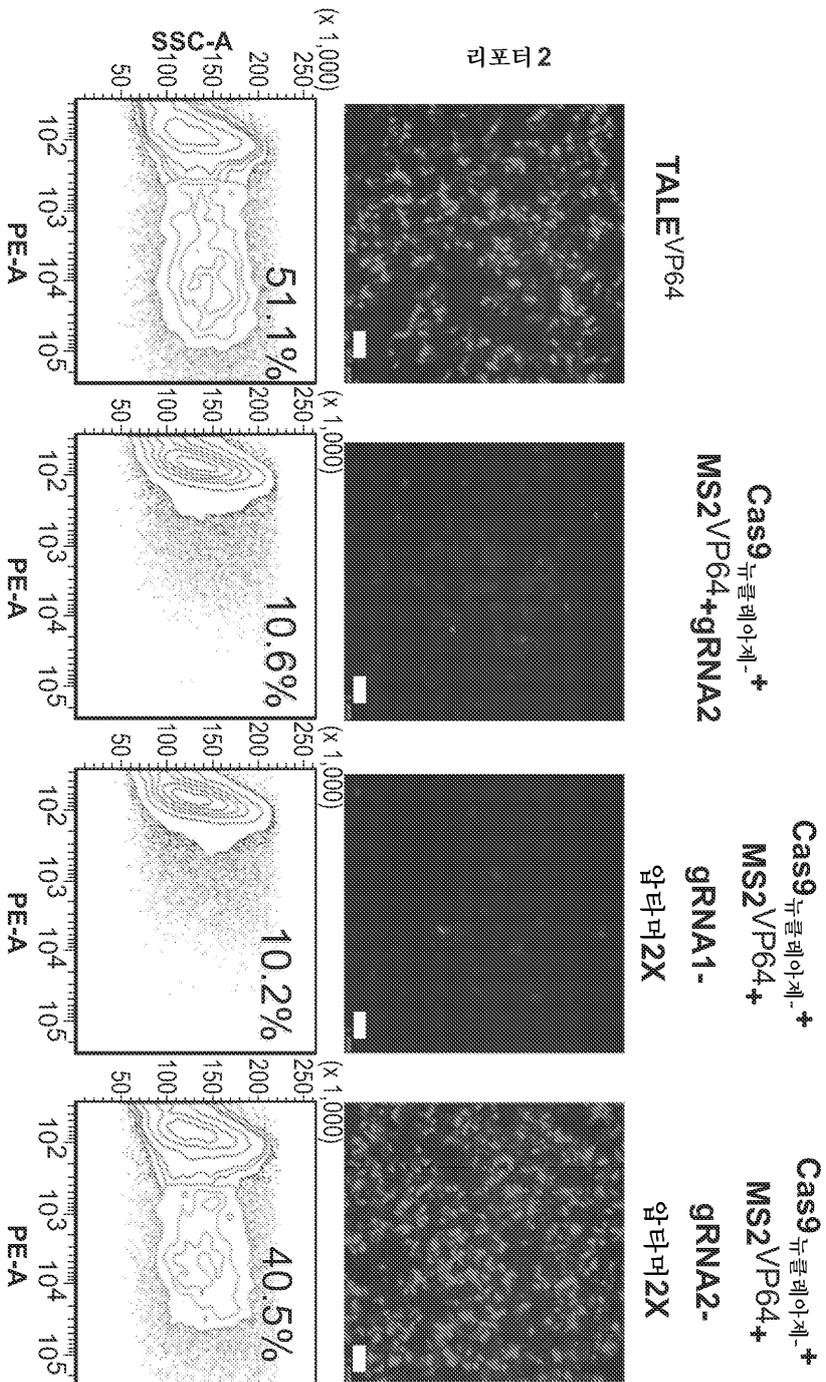
도면1db



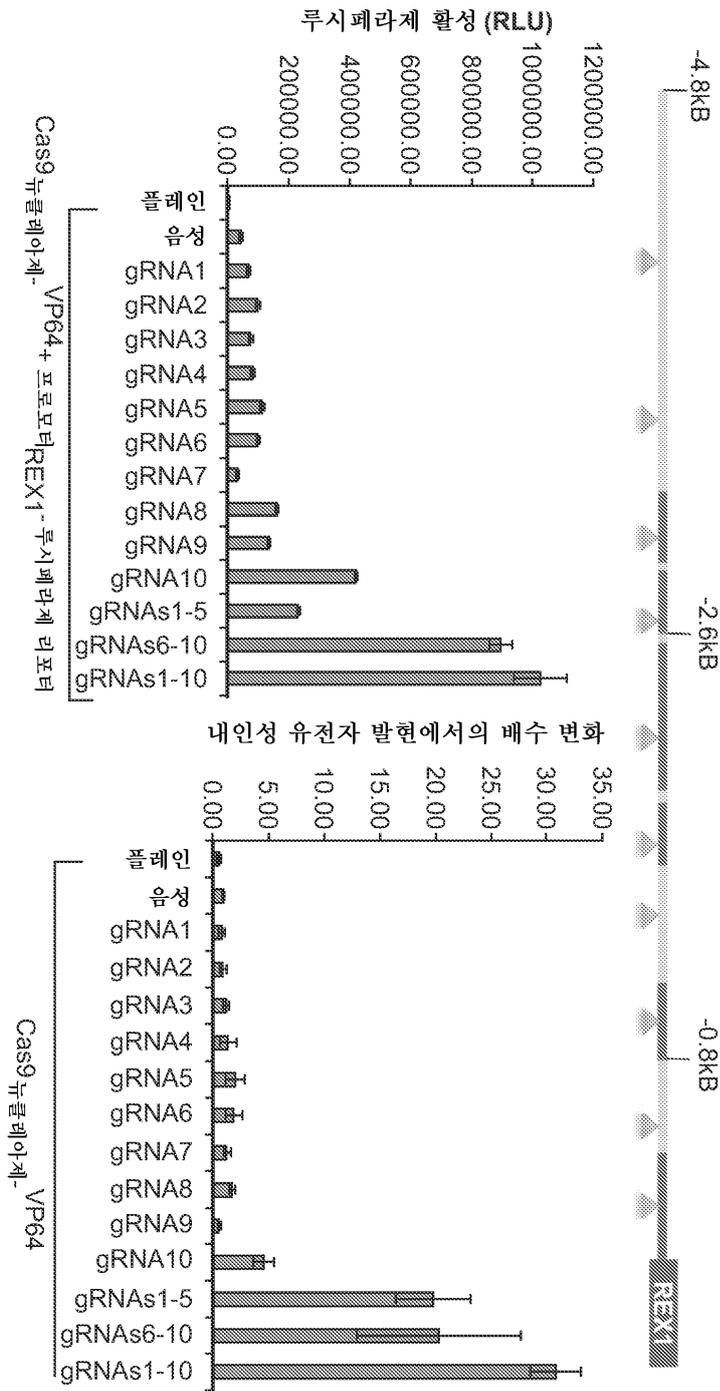
도면 1ea



도면 1eb

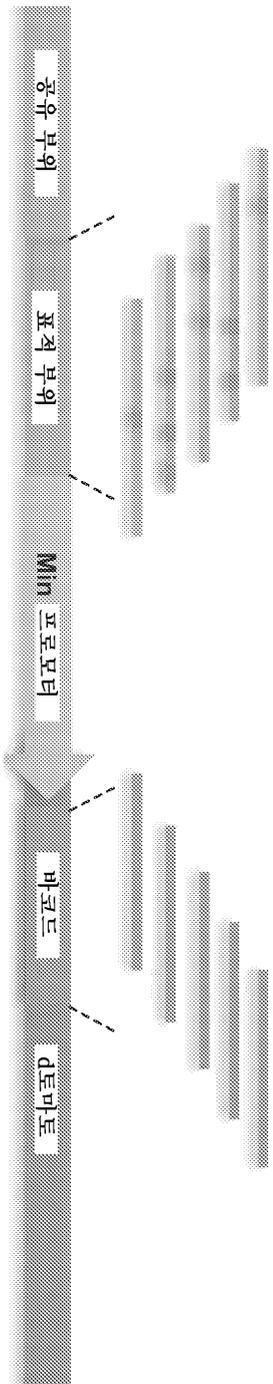


도면1f



결합 부위의 편제 라이브러리
(표적 뉴클레오티드: 79%, 나머지: 21%)

24bp 바코드
(A/C/G)



단계 1: 라이브러리 내의 상응하는 표적 부위에 대한 바코드 맵핑

단계 2: 1) 공유 부위에 결합하는 대조-TF; 또는

2) 표적 부위에 결합하는 TALE-TF/gRNA+ Cas9-TF (표적-TF)
에 의한 라이브러리의 자극

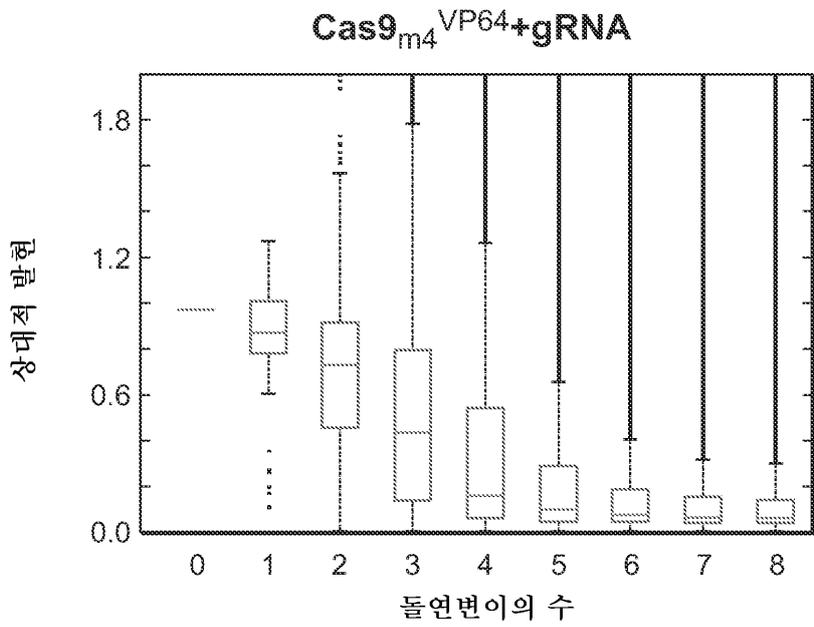
단계 3: RNaseq 수행 및 각각에 대한 발현 바코드 결정

단계 4: 상응하는 결합 부위에 대한 발현 바코드 재맵핑

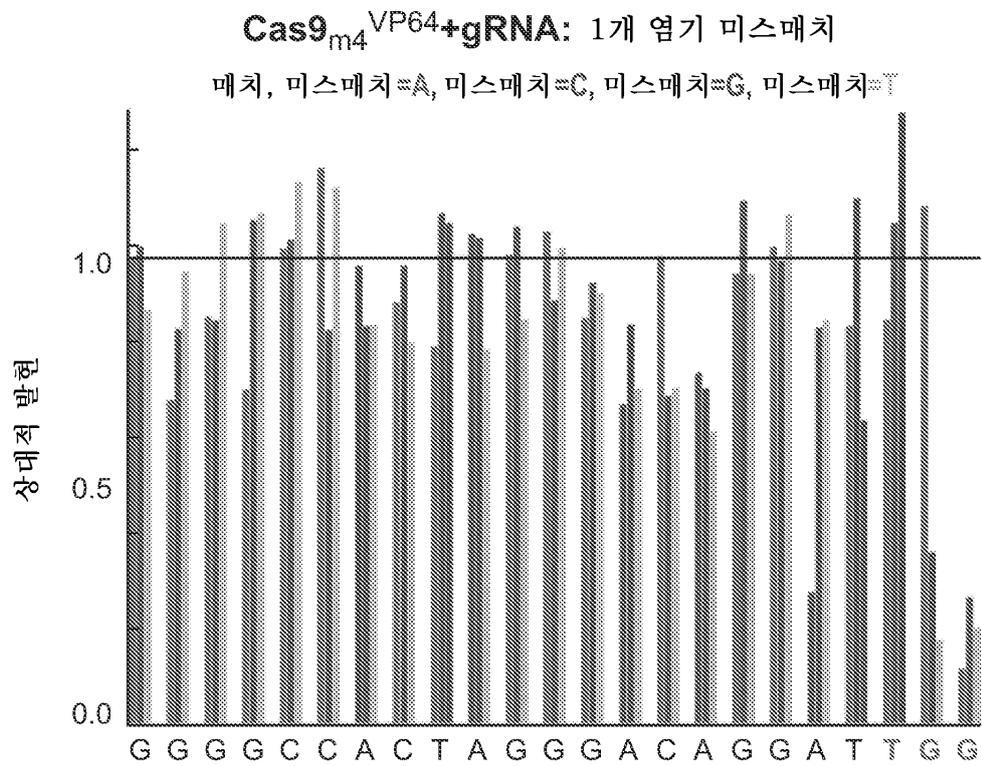
단계 5: 표적-TF 대 대조-TF 바코드의 상대적 풍부화 계산

도면2a

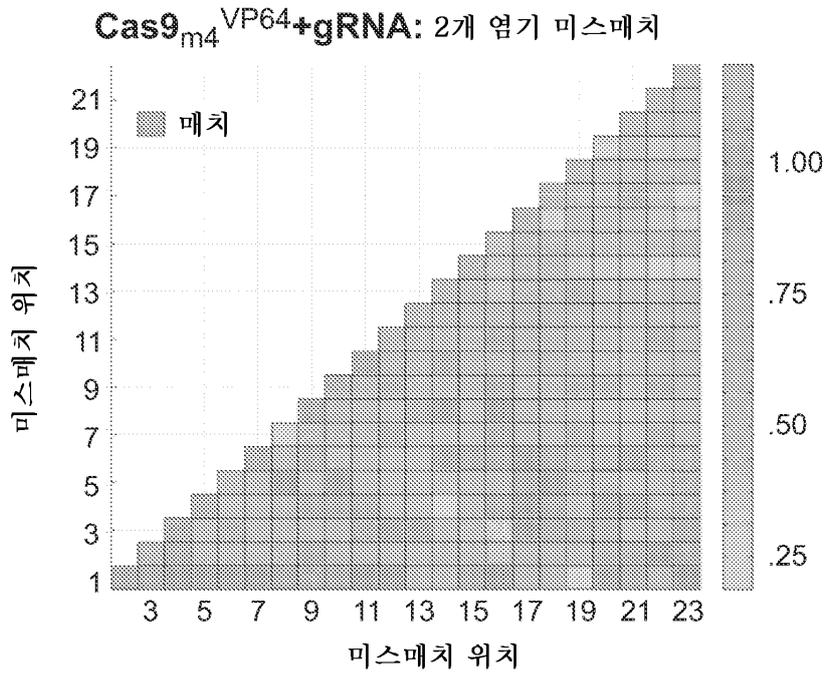
도면2b



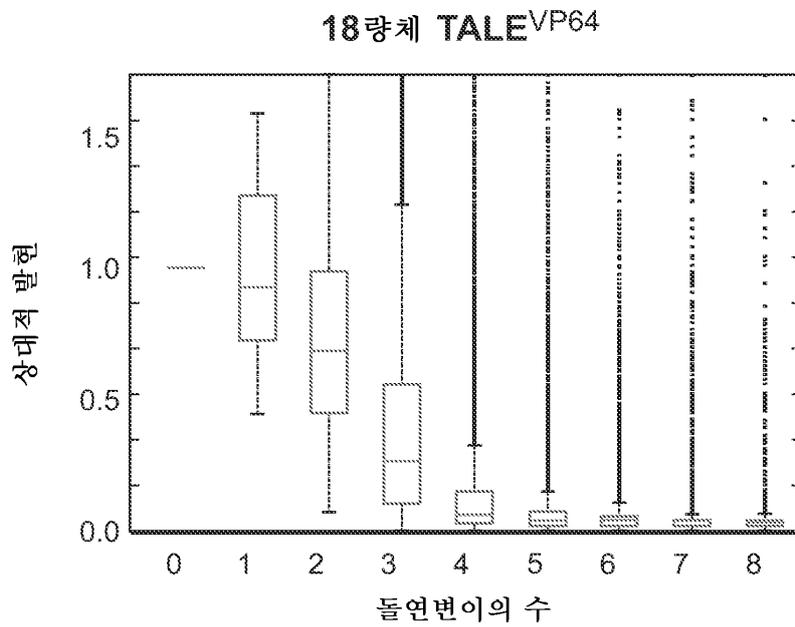
도면2c



도면2d

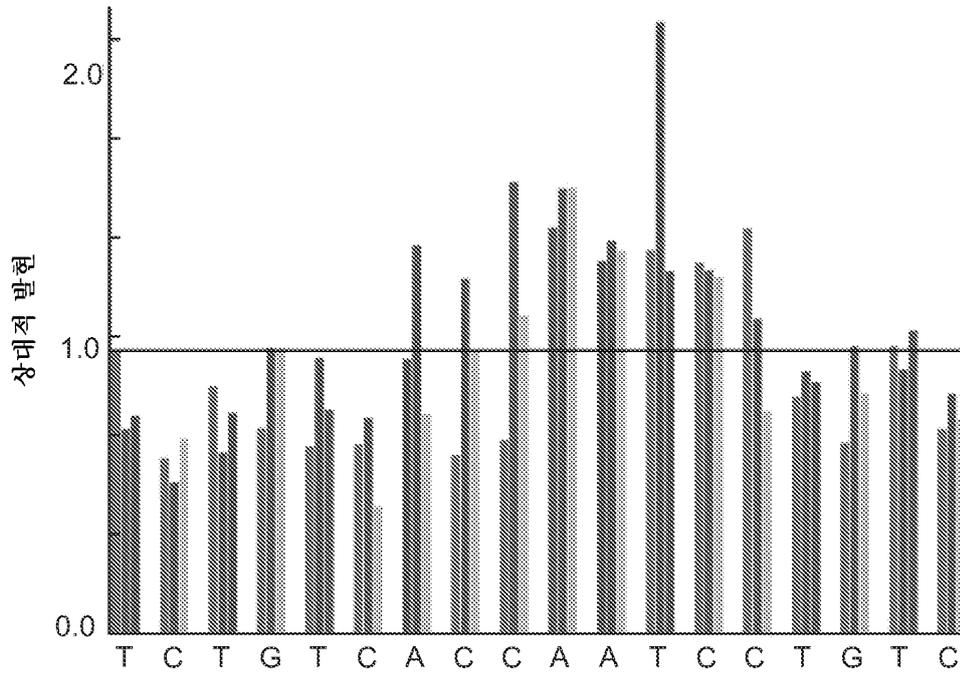


도면2e

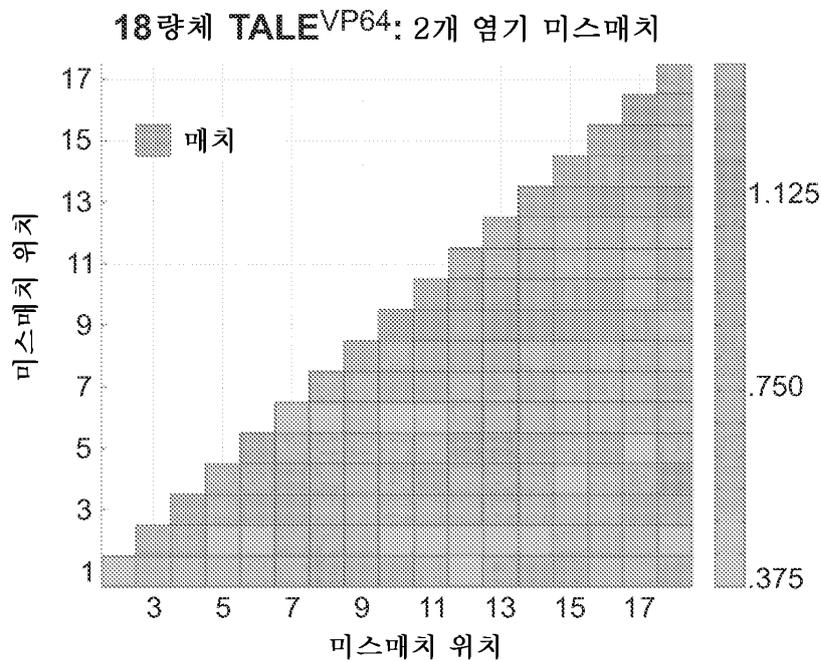


도면2f

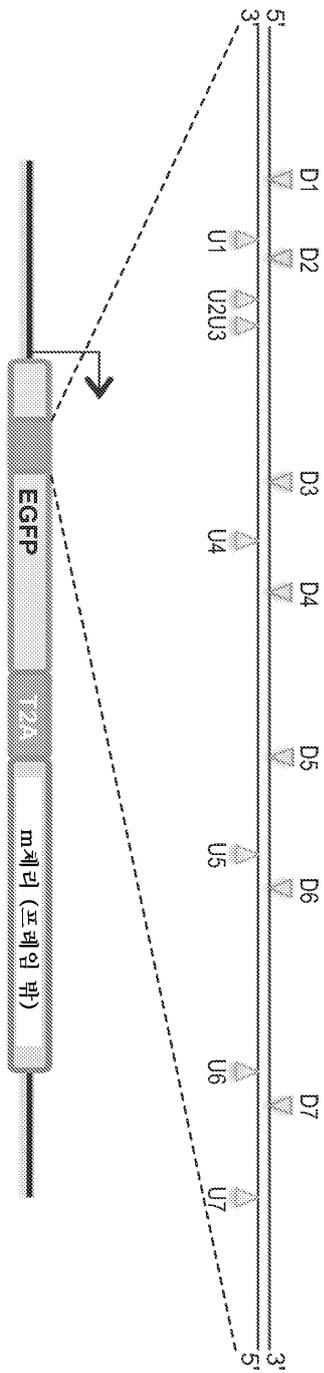
18량체 TALE^{VP64}: 1개 염기 미스매치
 매치, 미스매치=A, 미스매치=C, 미스매치=G, 미스매치=T



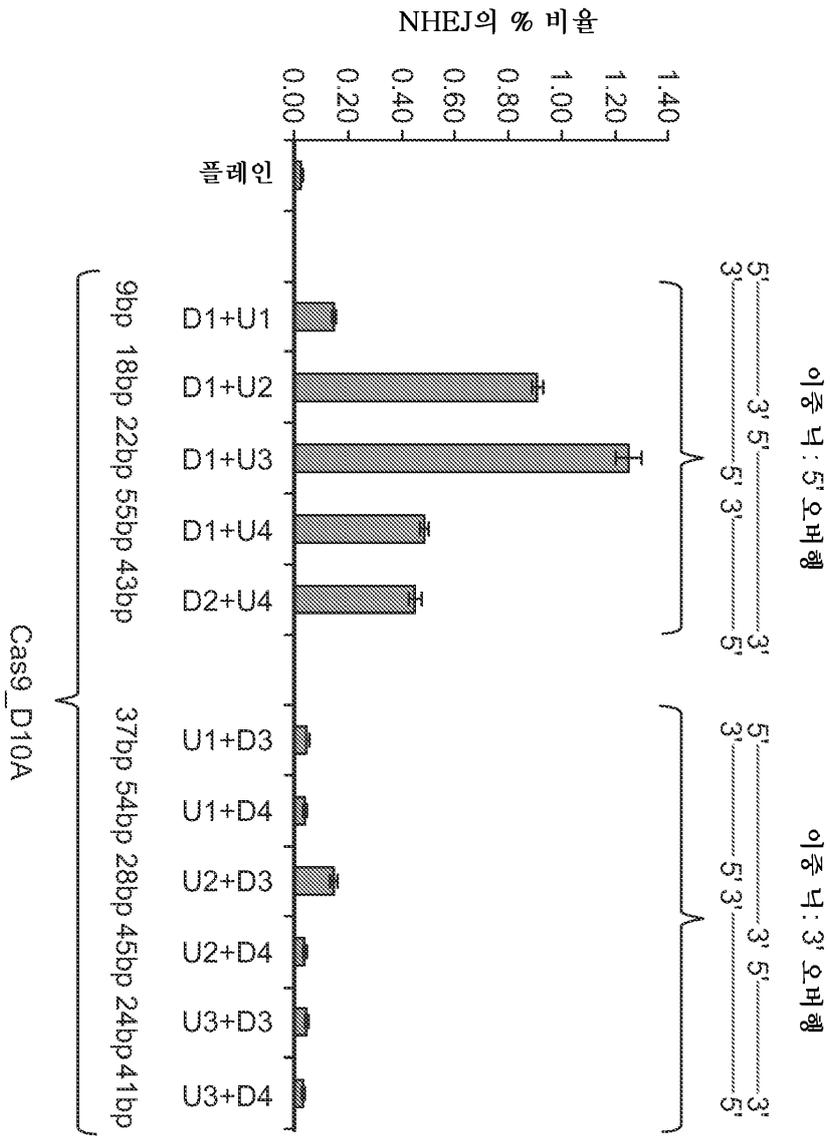
도면2g



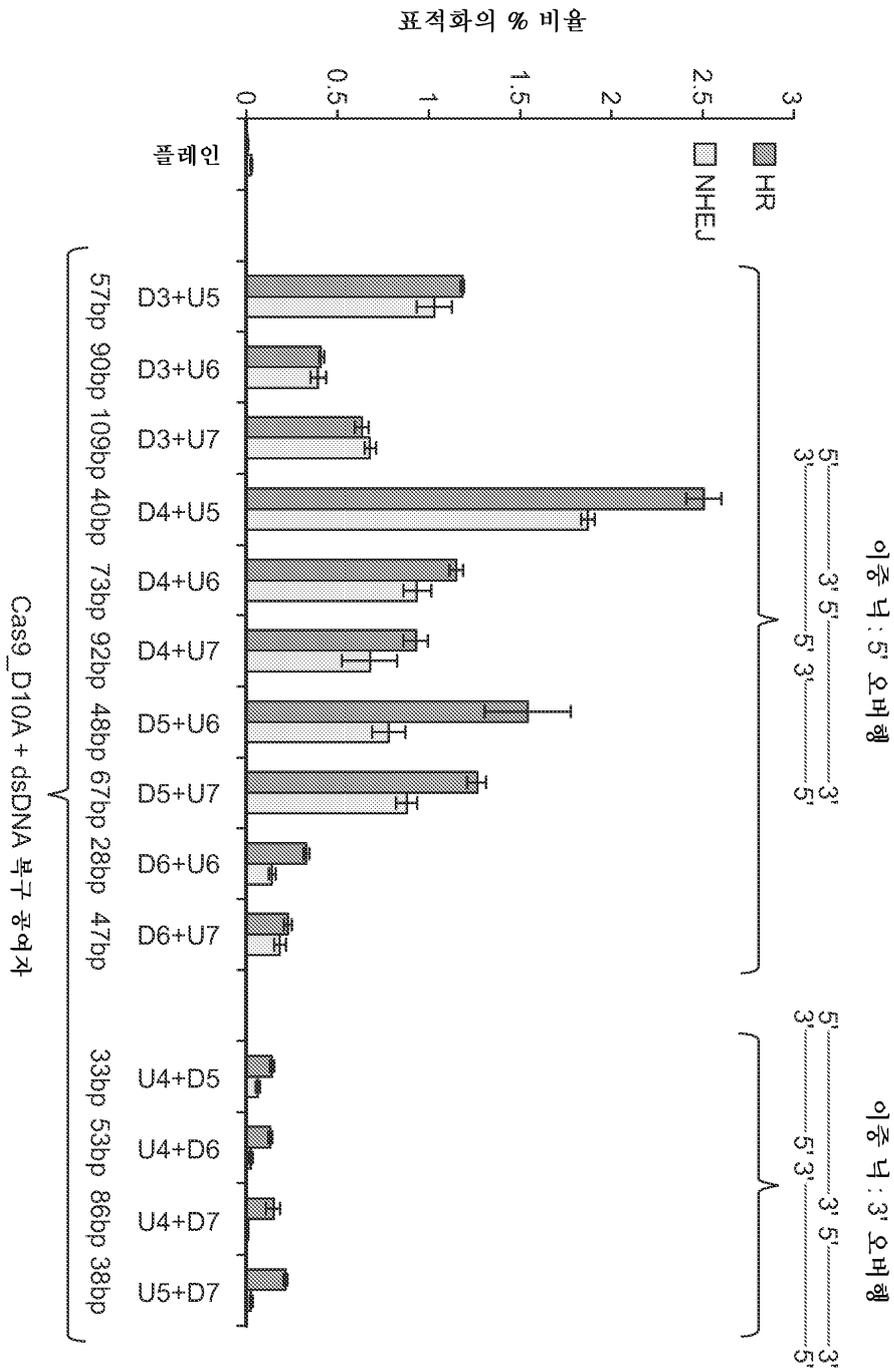
도면3a



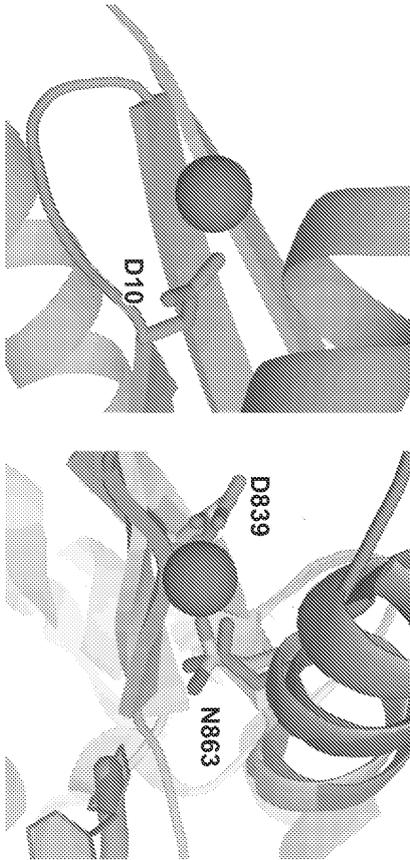
도면3b



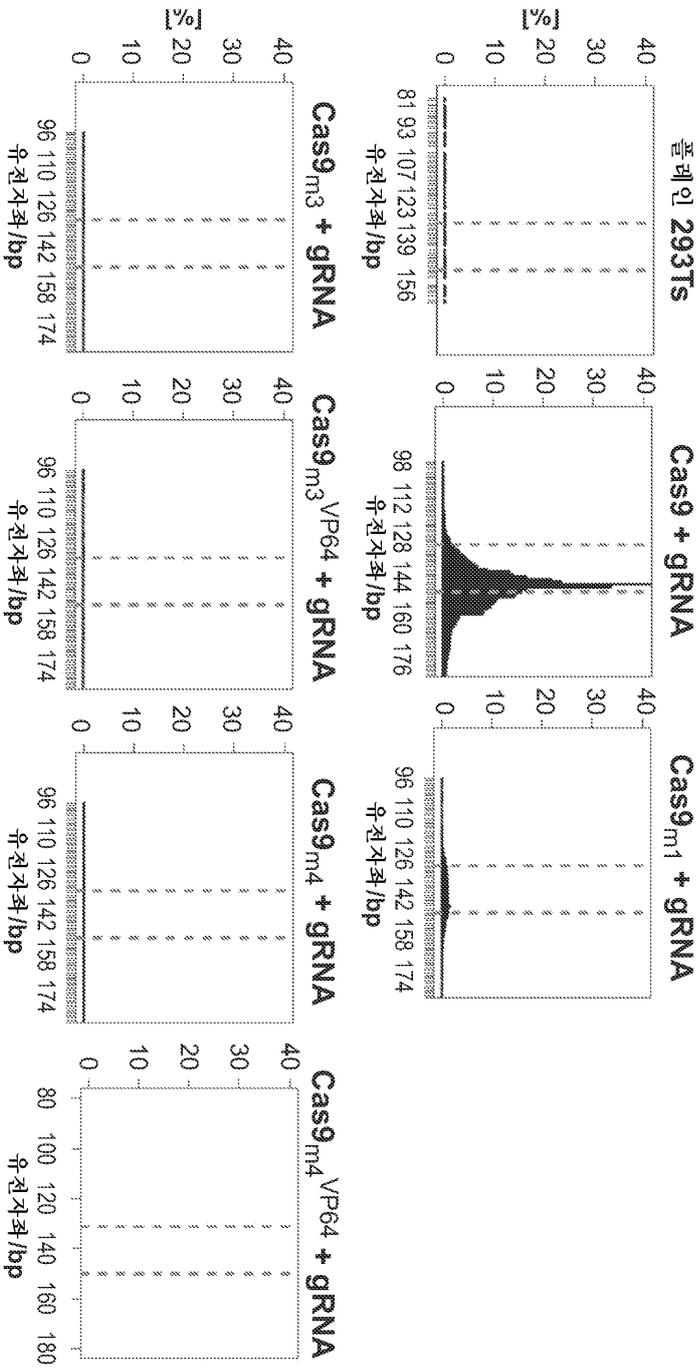
도면3c



도면4a

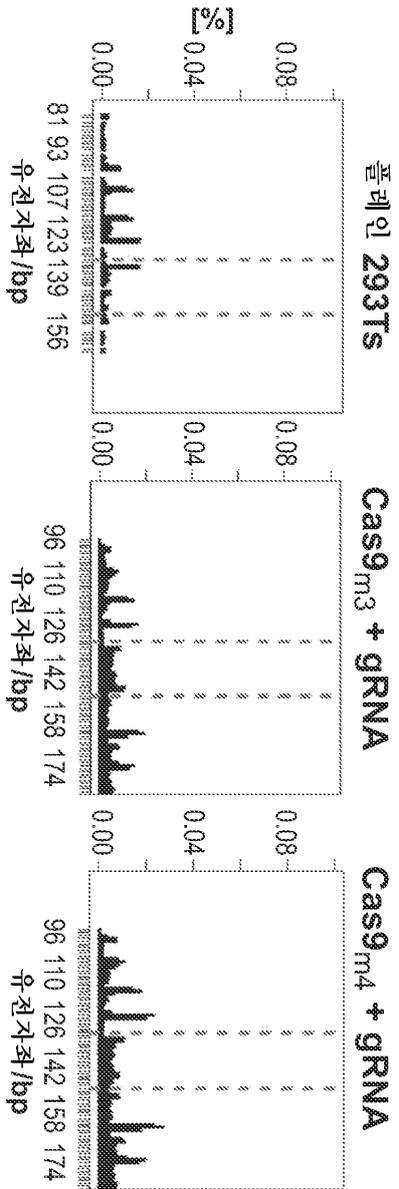


명칭	돌연변이
Cas9	야생형
Cas9 _{m1}	D10A
Cas9 _{m2}	D10A+H840A
Cas9 _{m3}	D10A+D839A+H840A
Cas9 _{m4}	D10A+D839A+H840A+N863A

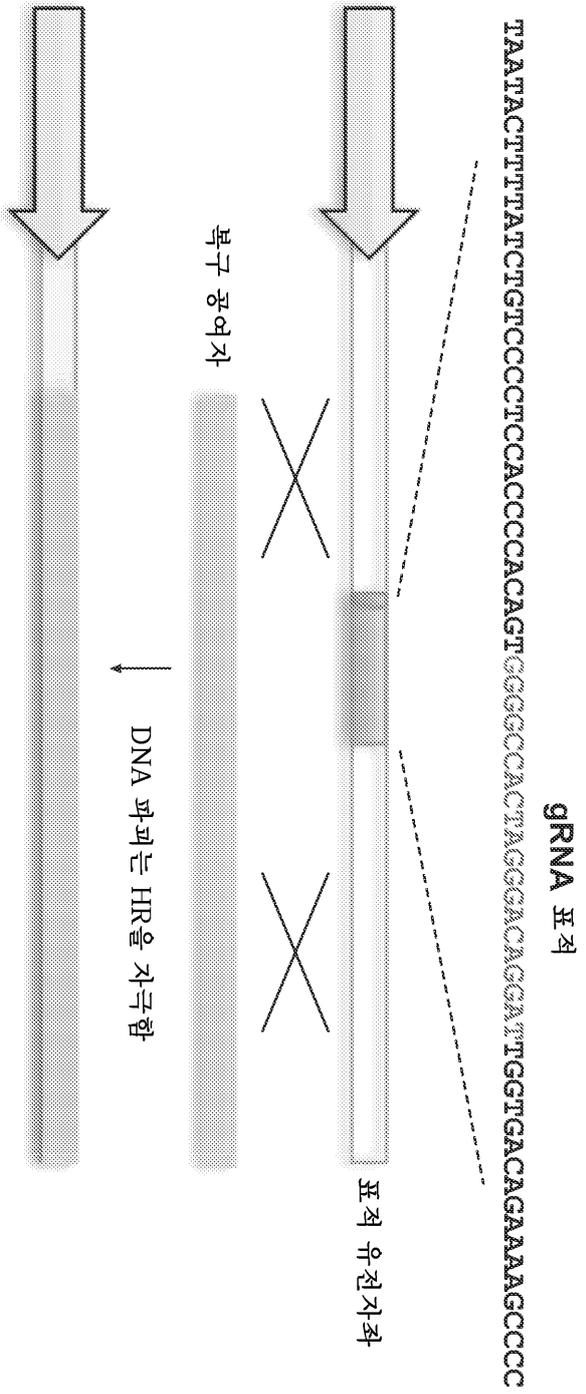


도면4b

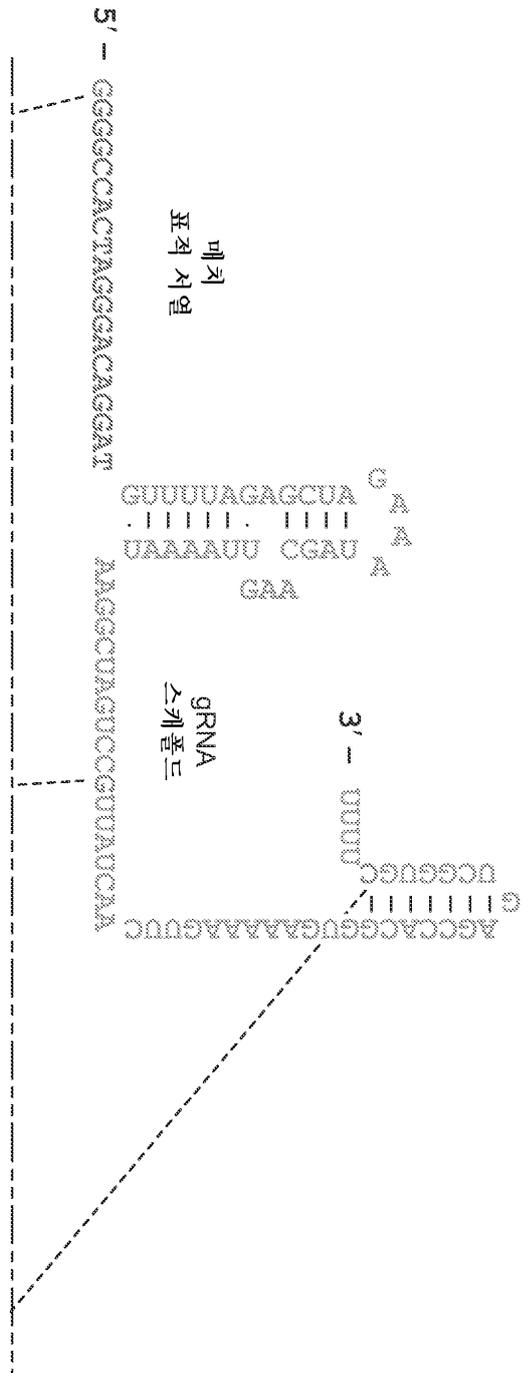
도면4c



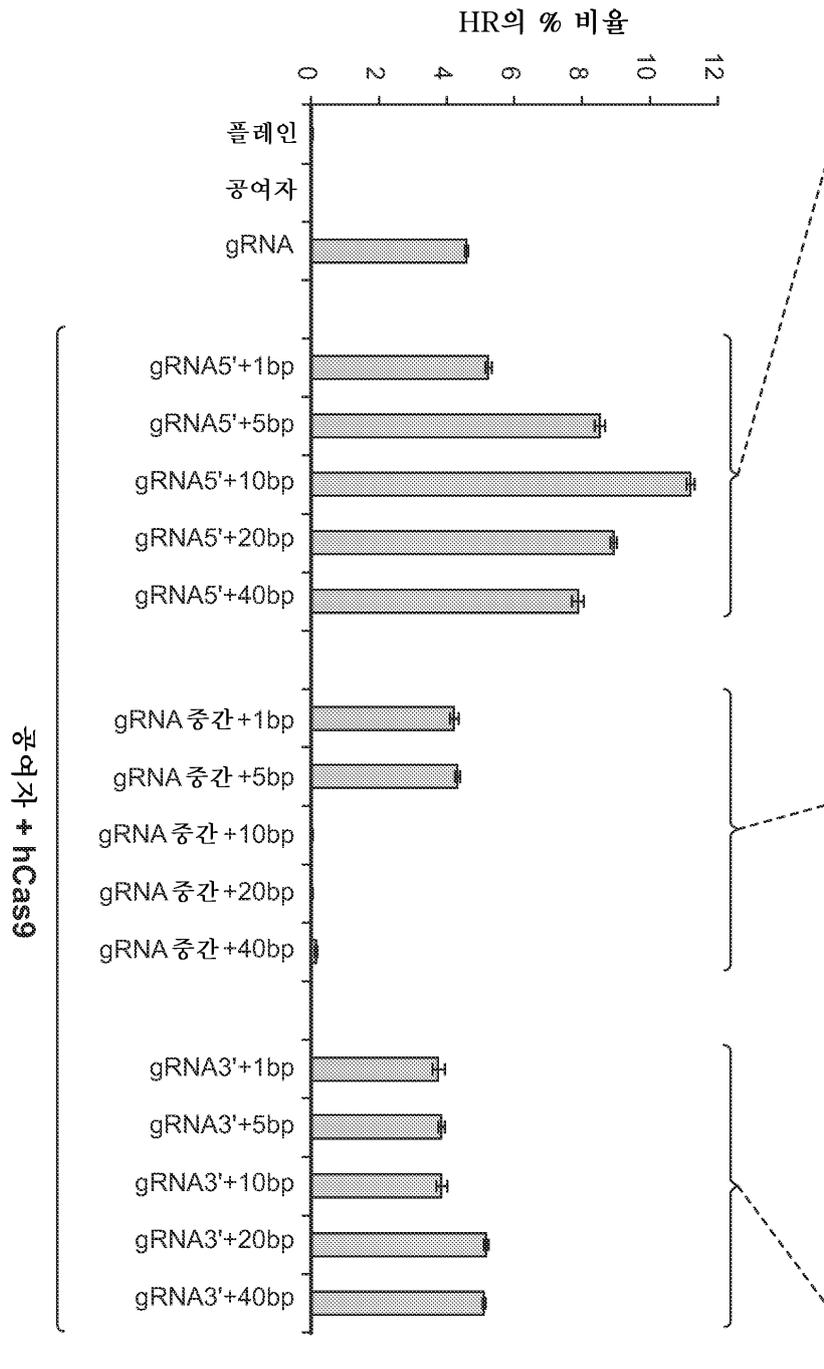
도면5a



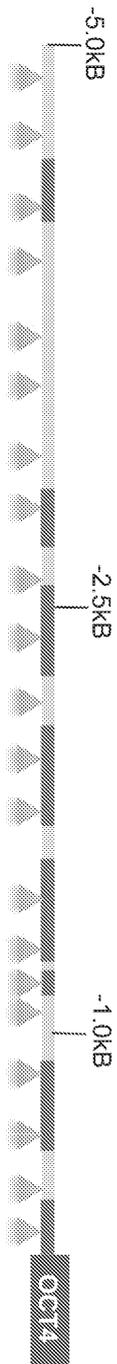
도면5ba



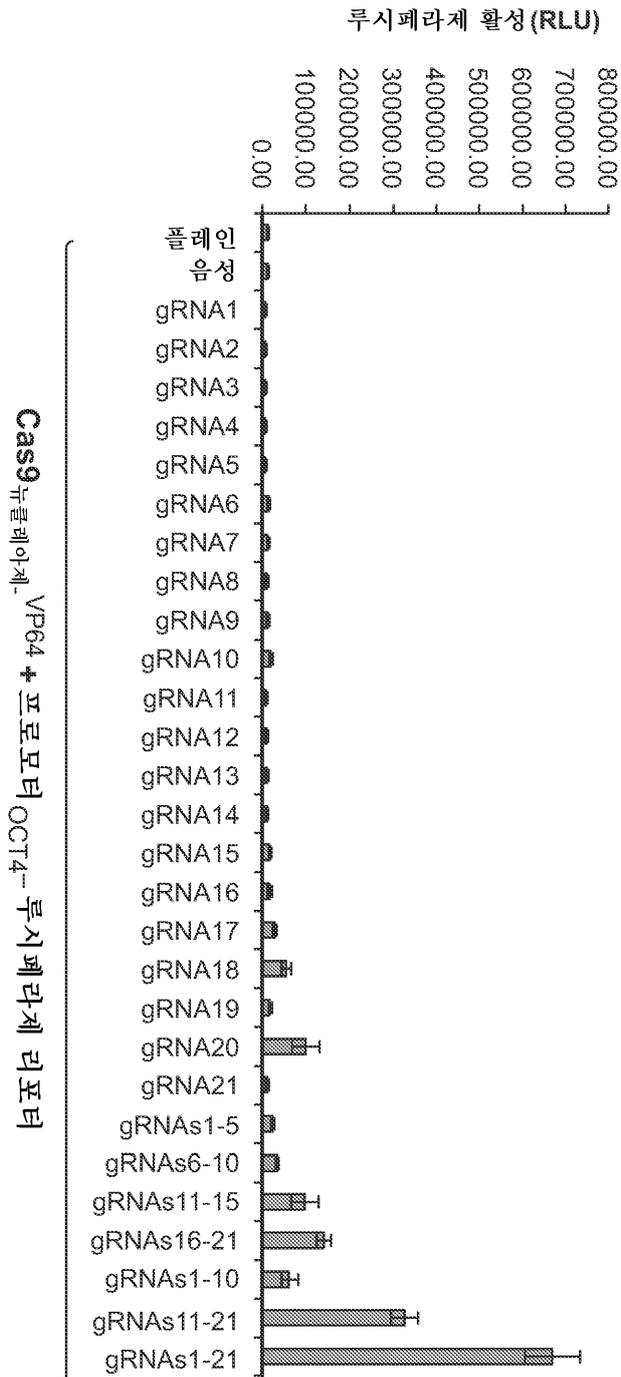
도면5bb



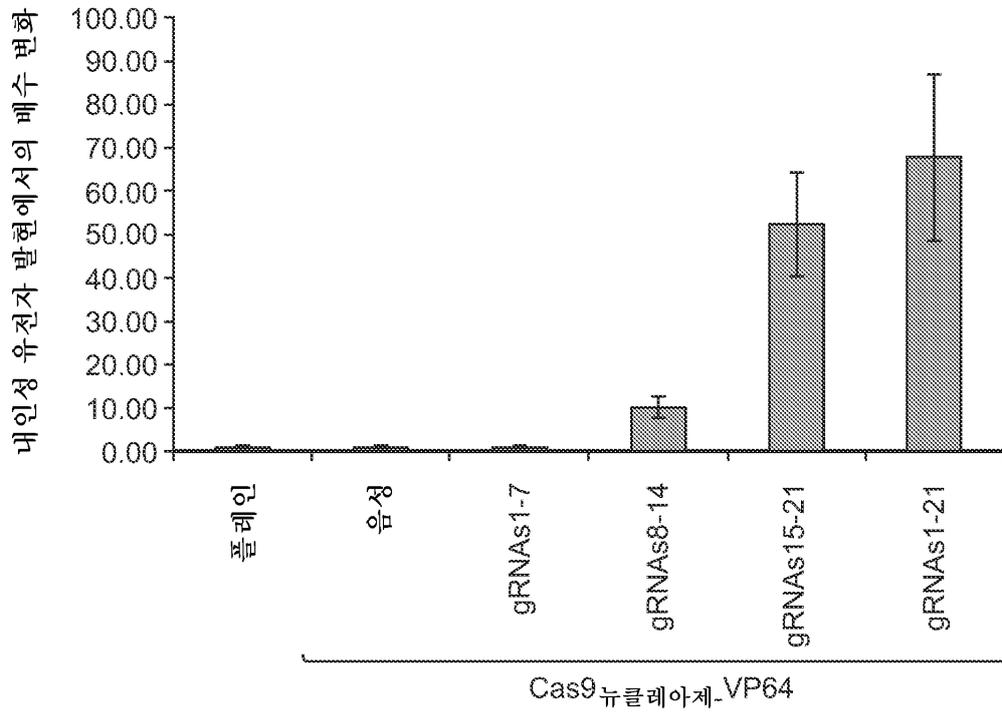
도면6a



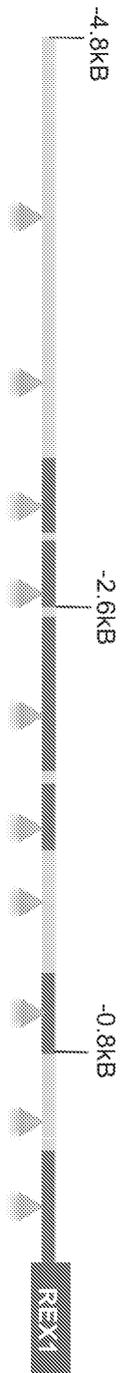
도면6b



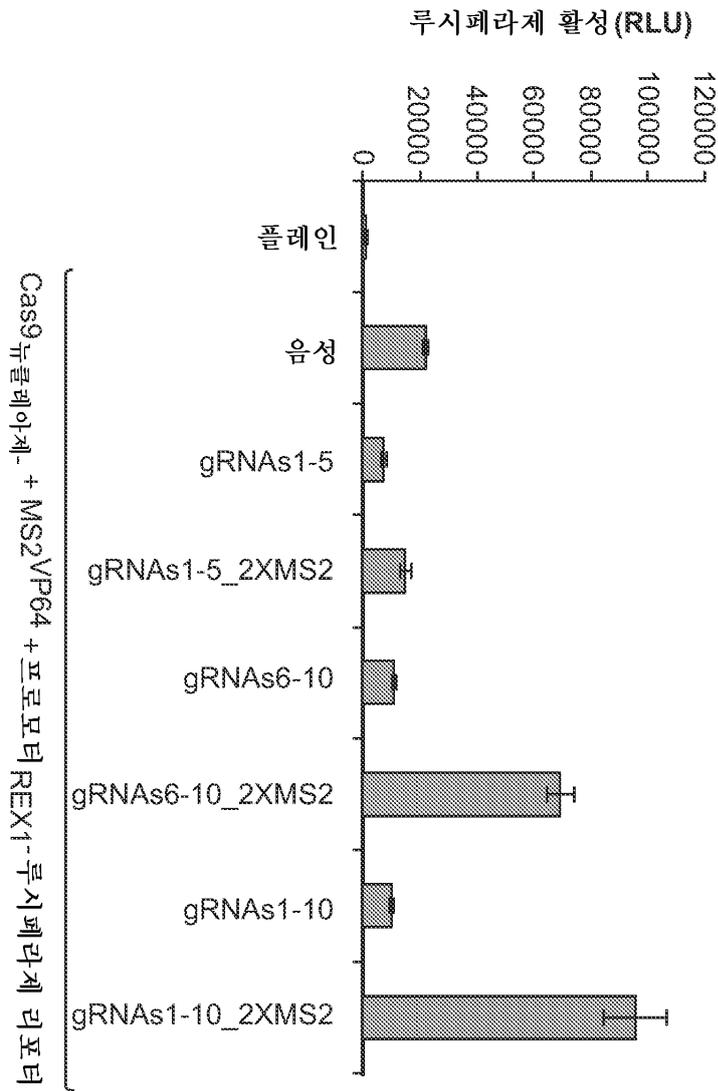
도면6c



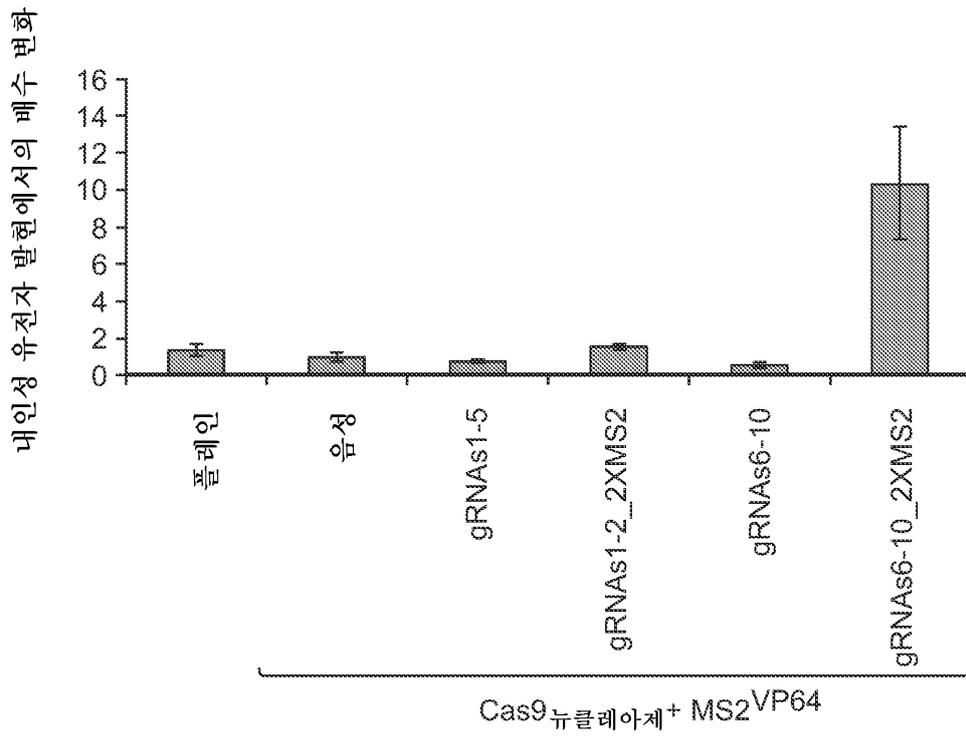
도면7a



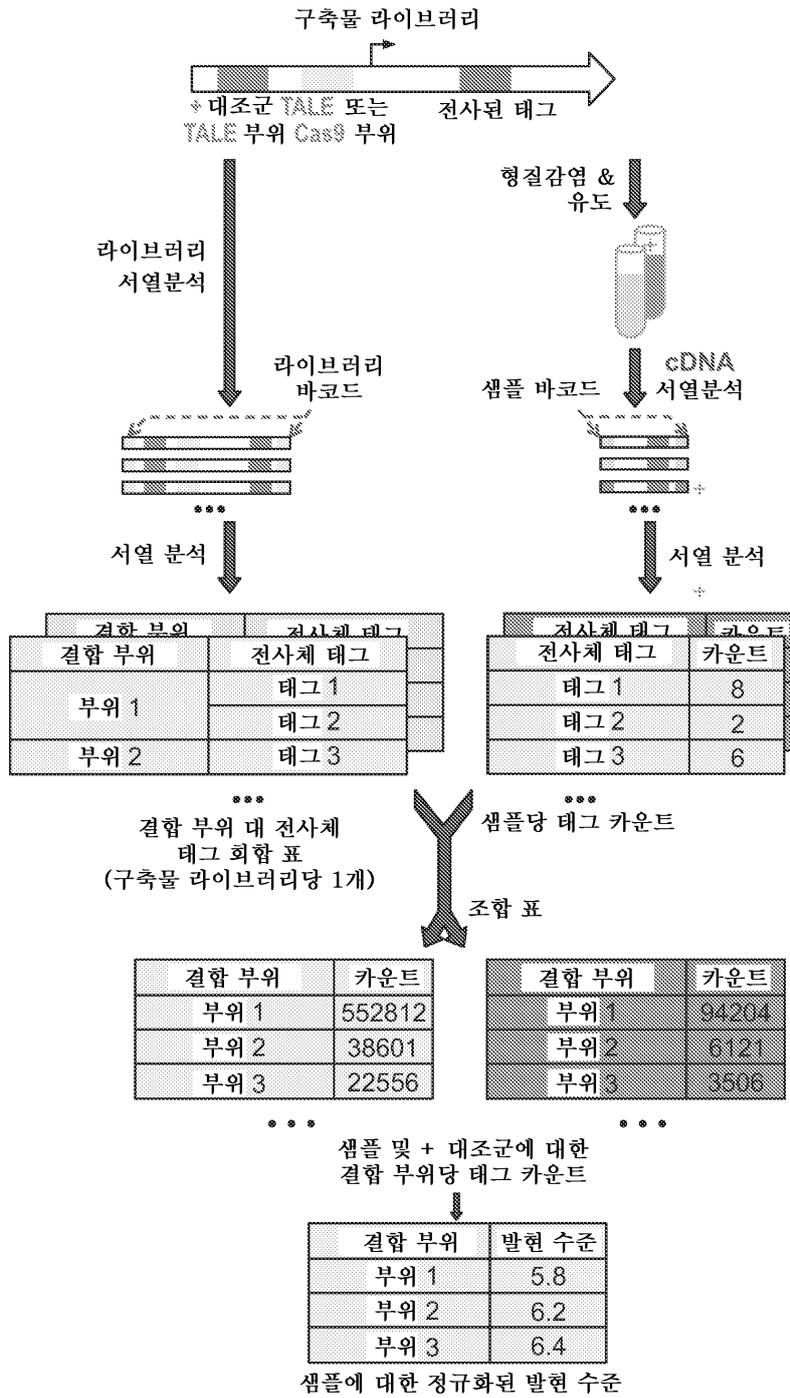
도면7b



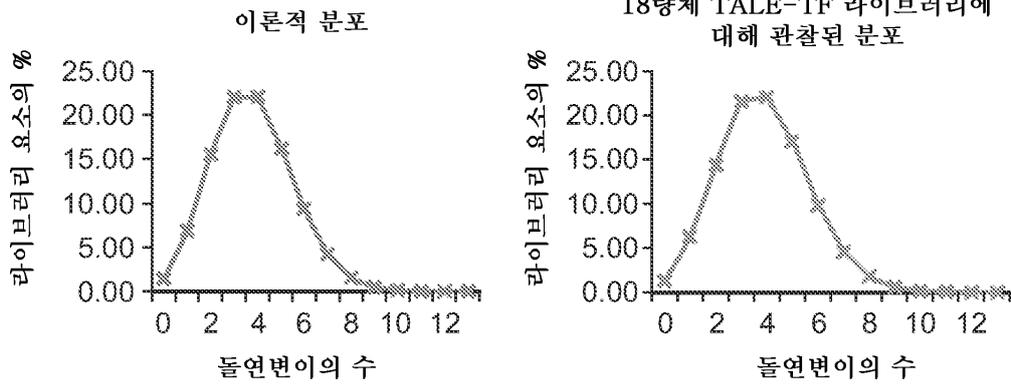
도면7c



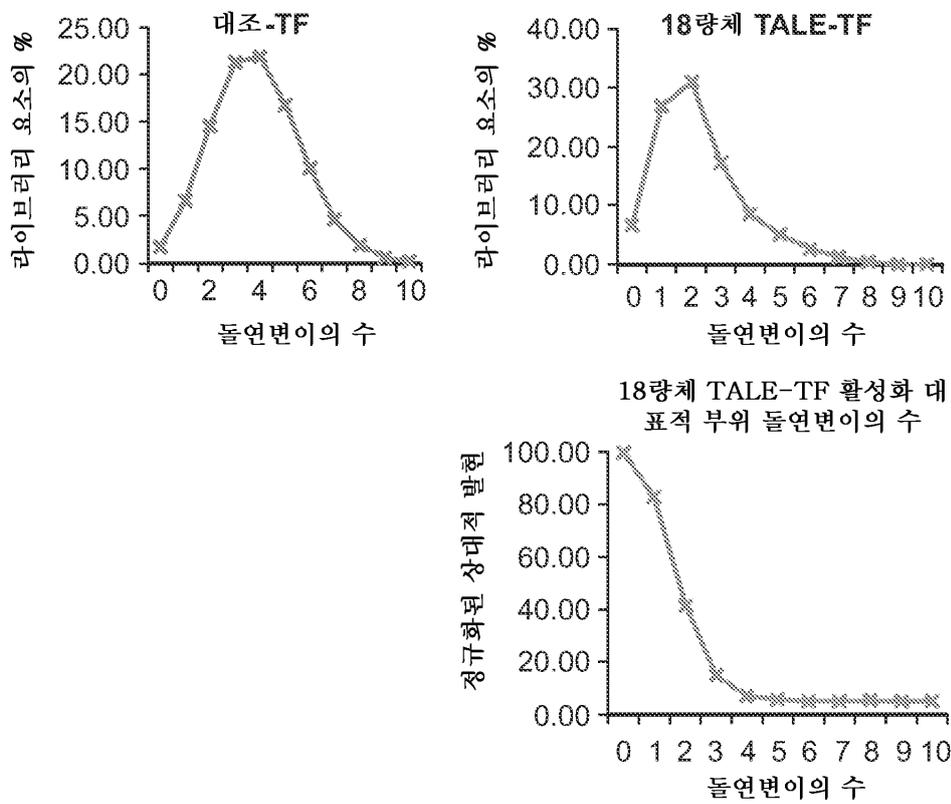
도면8a



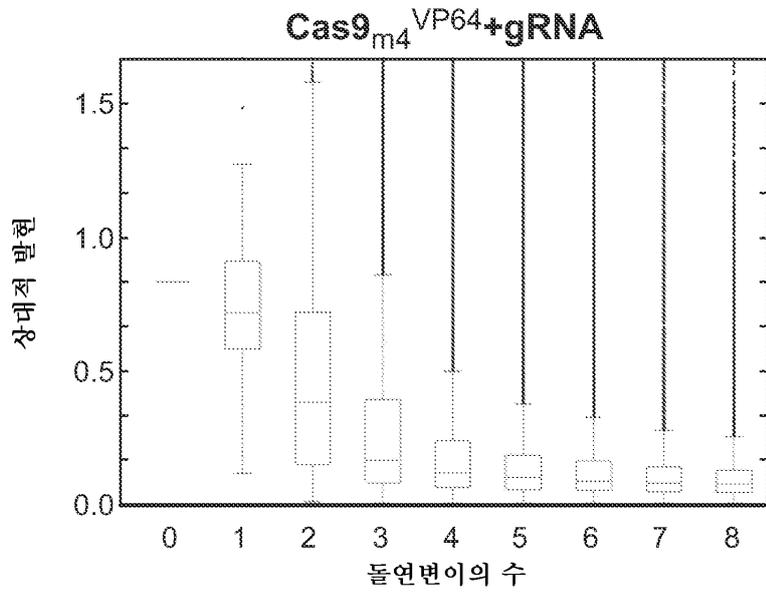
도면8b



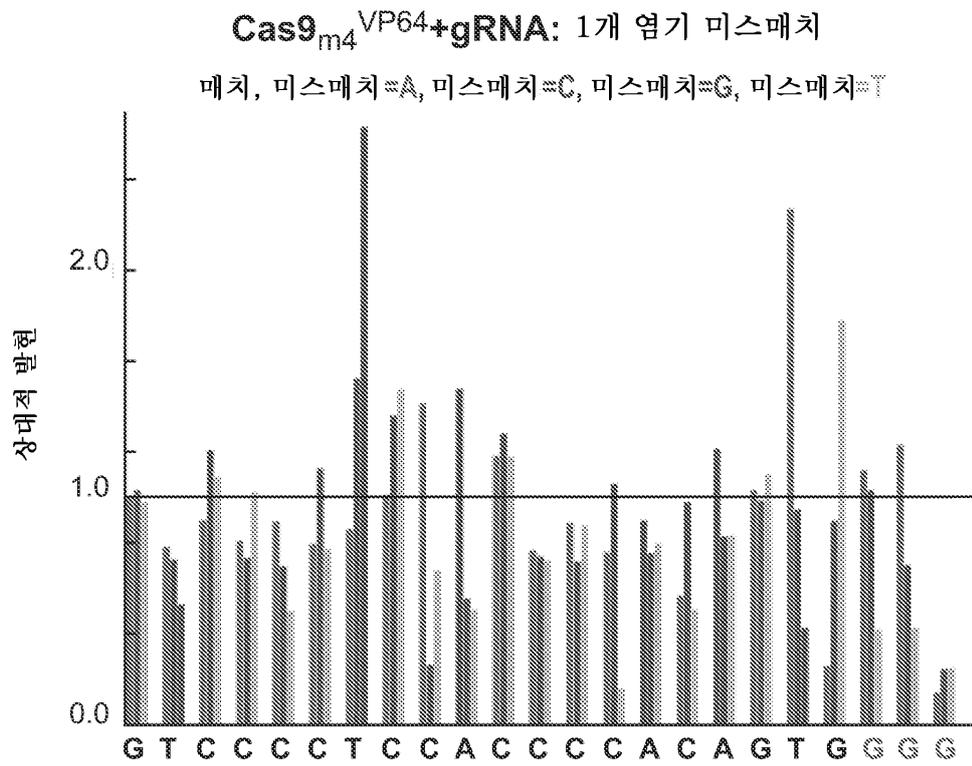
도면8c



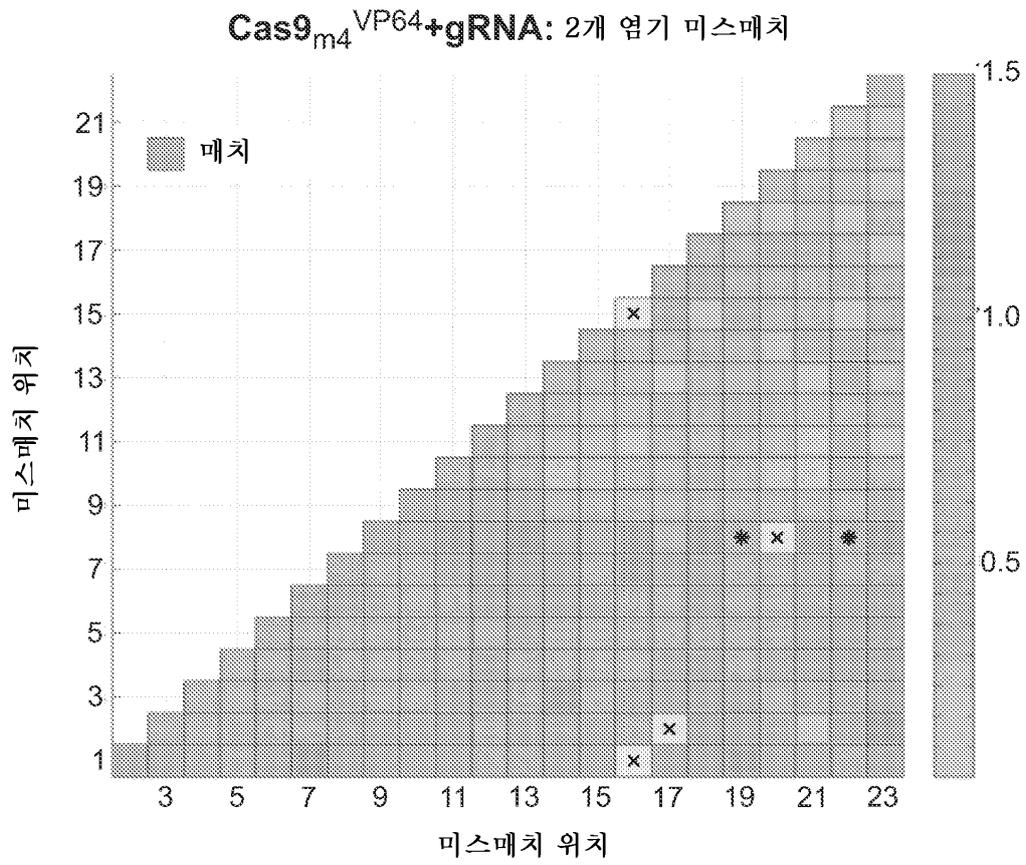
도면9a

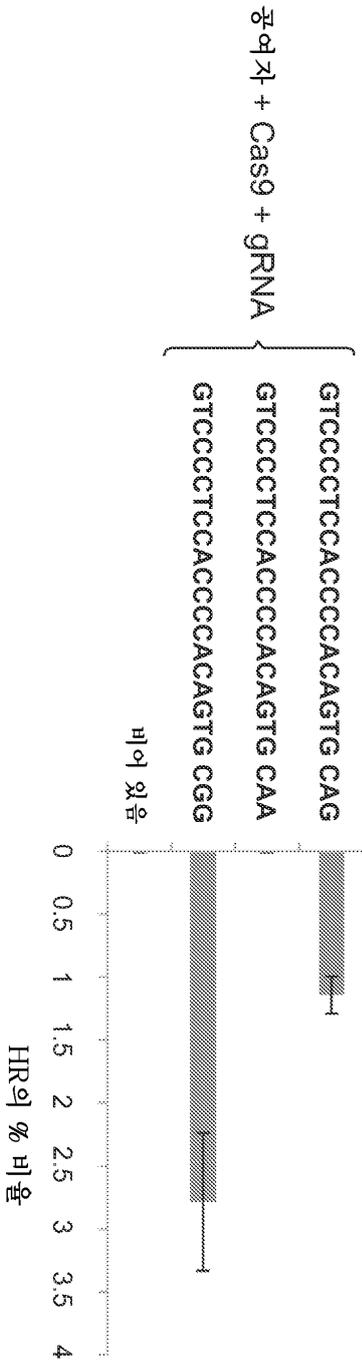
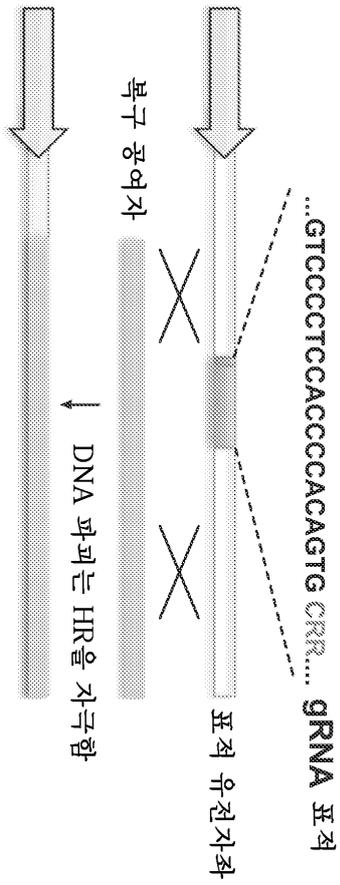


도면9b



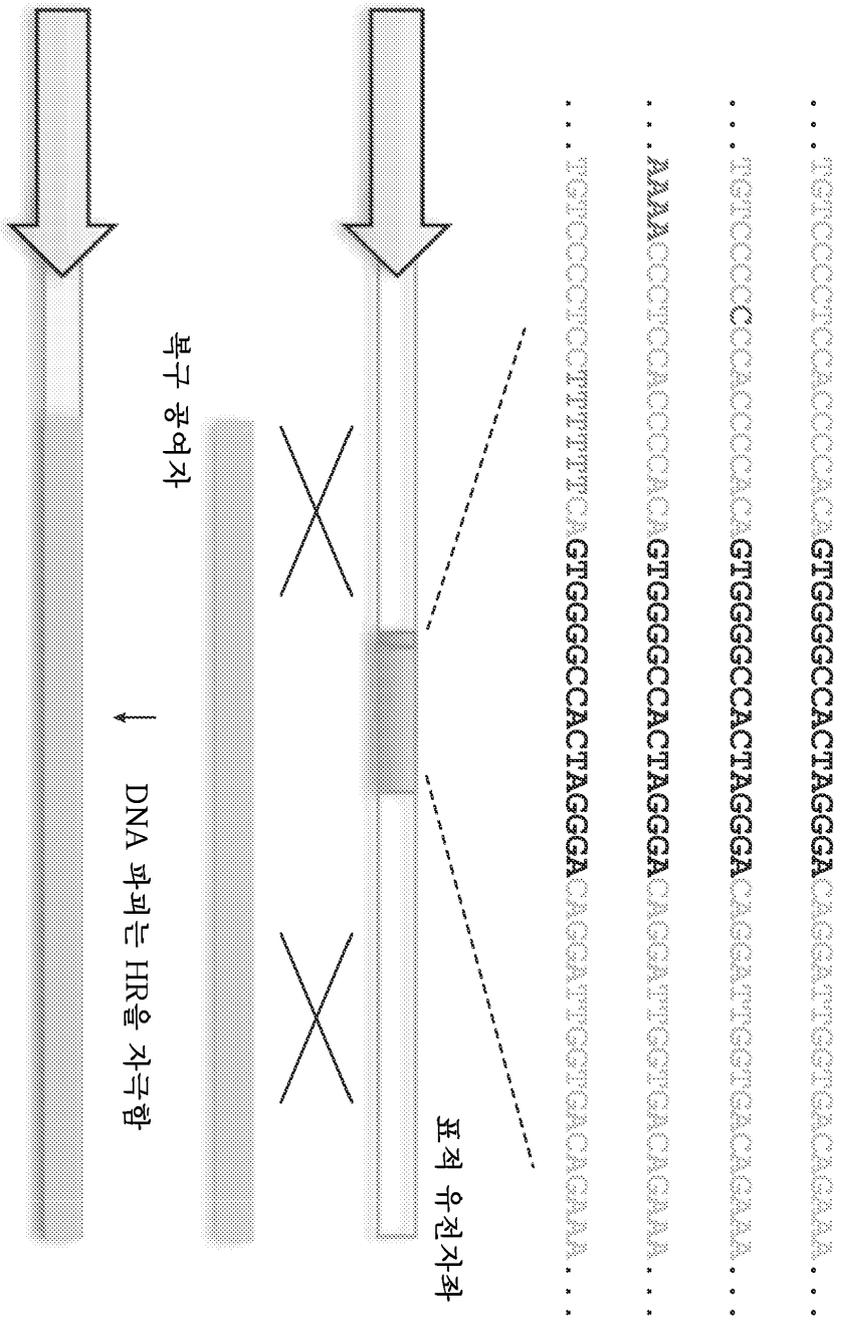
도면9c



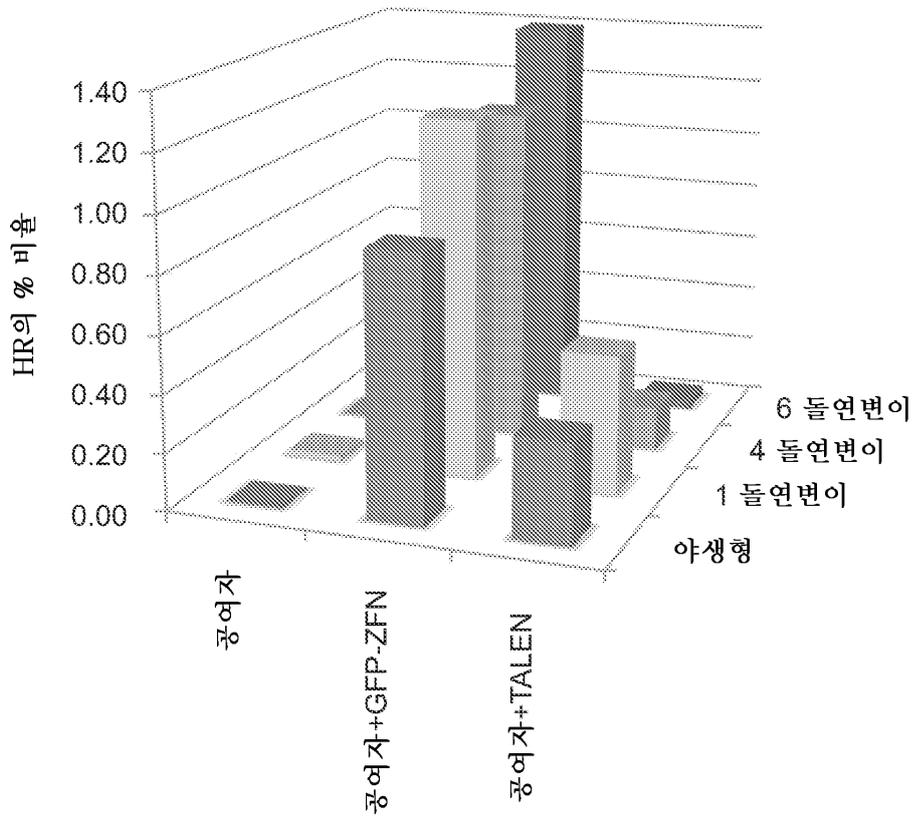


도면9d

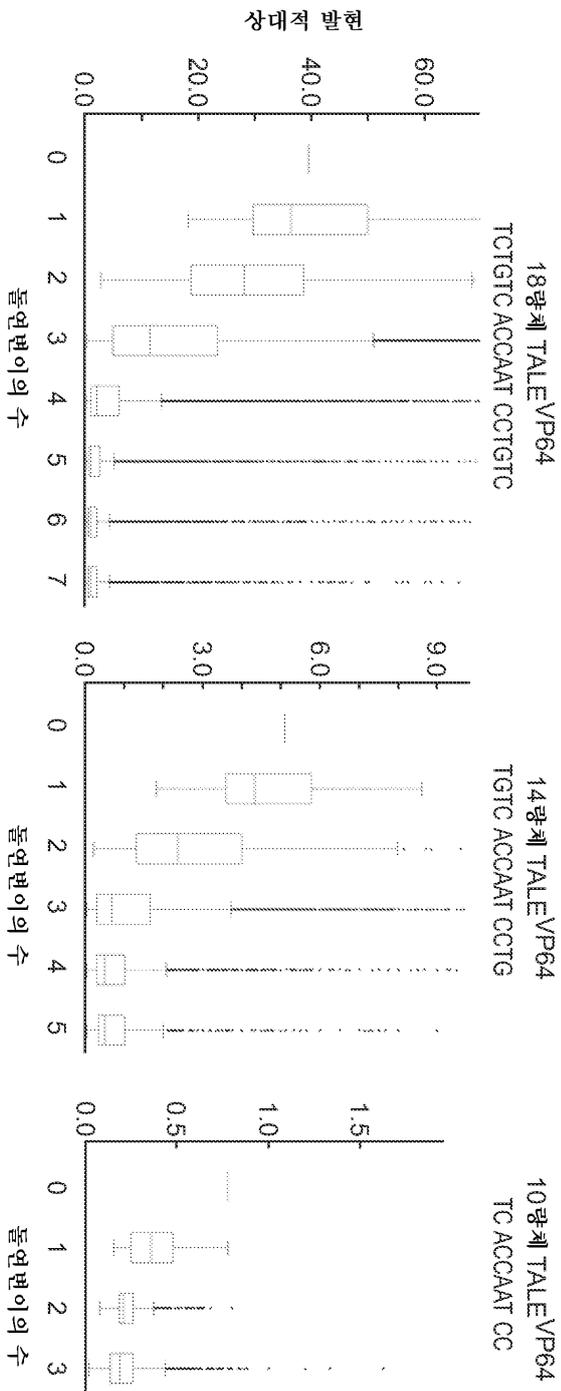
도면10aa



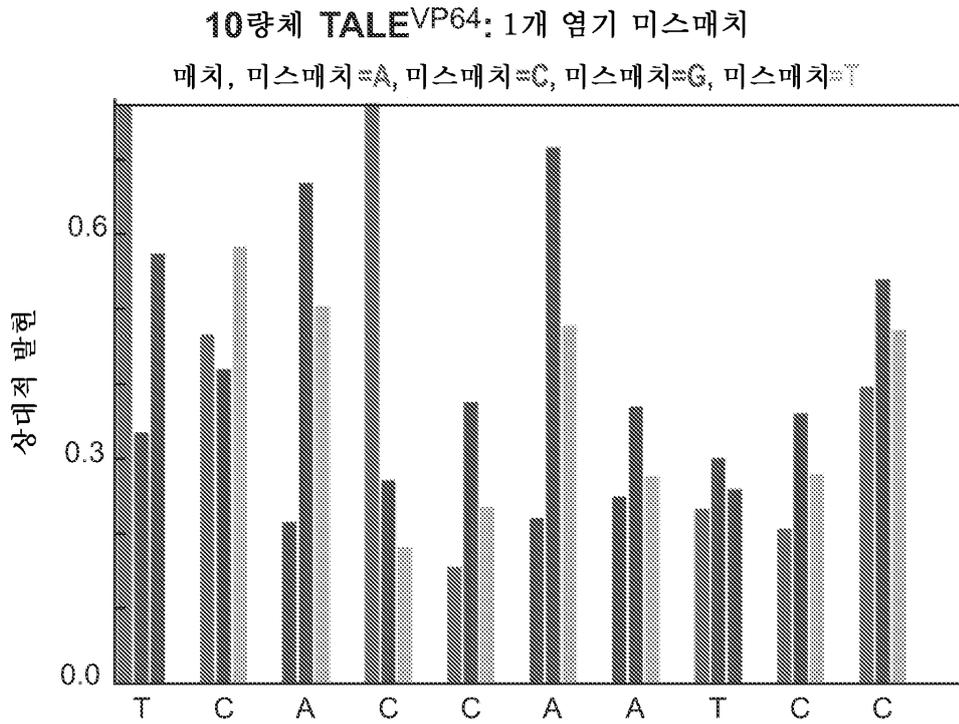
도면10ab



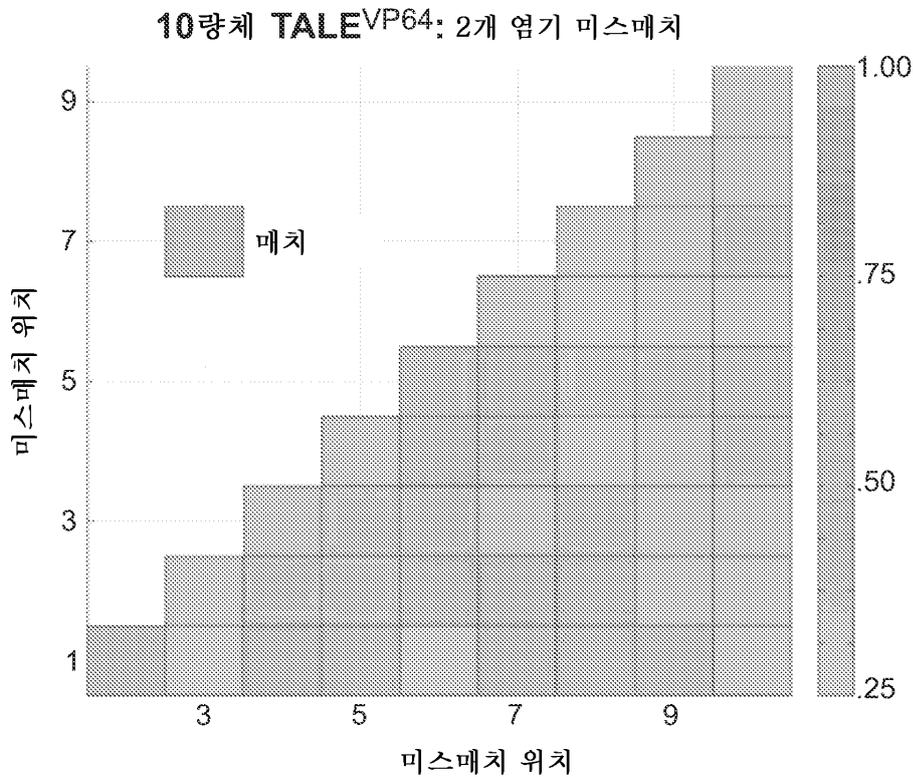
도면10b



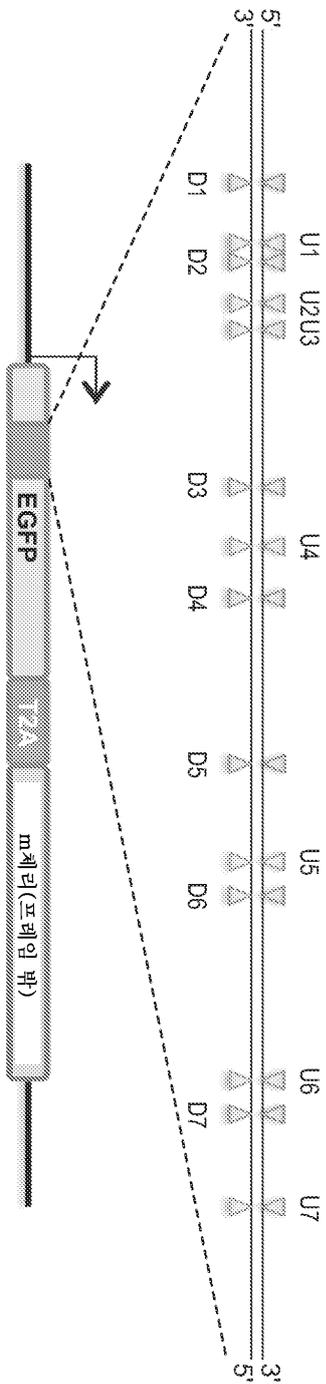
도면10c

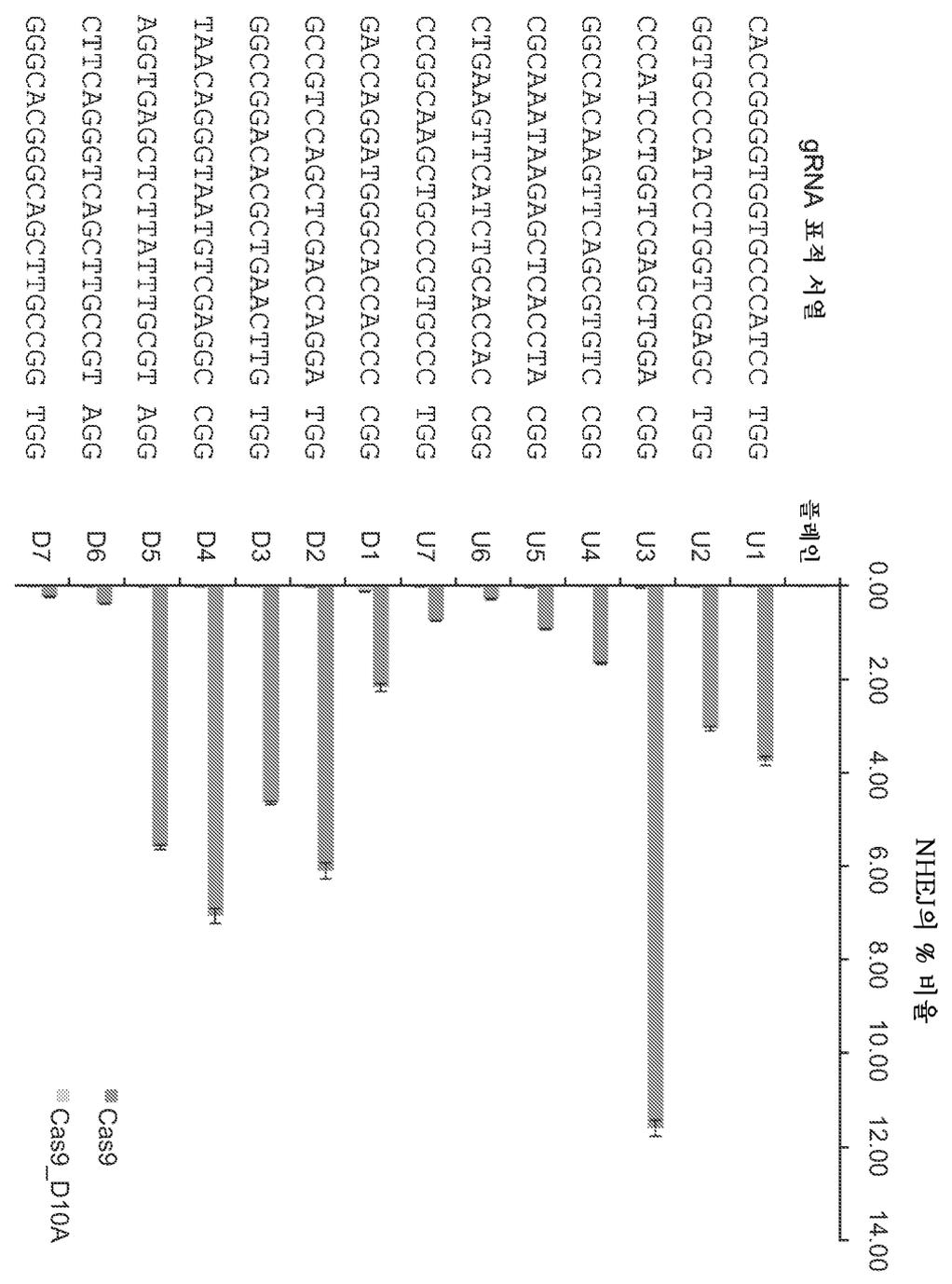


도면10d



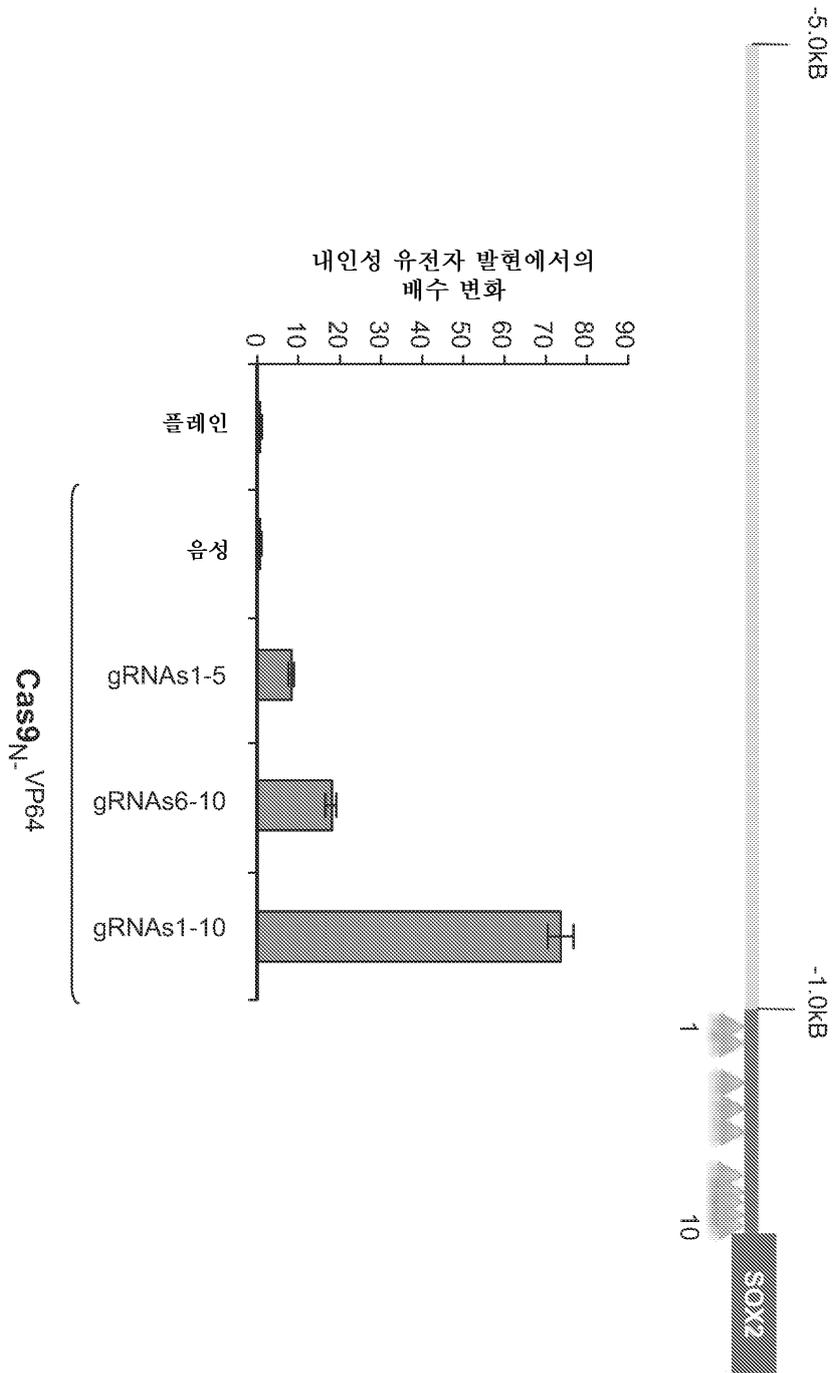
도면11a



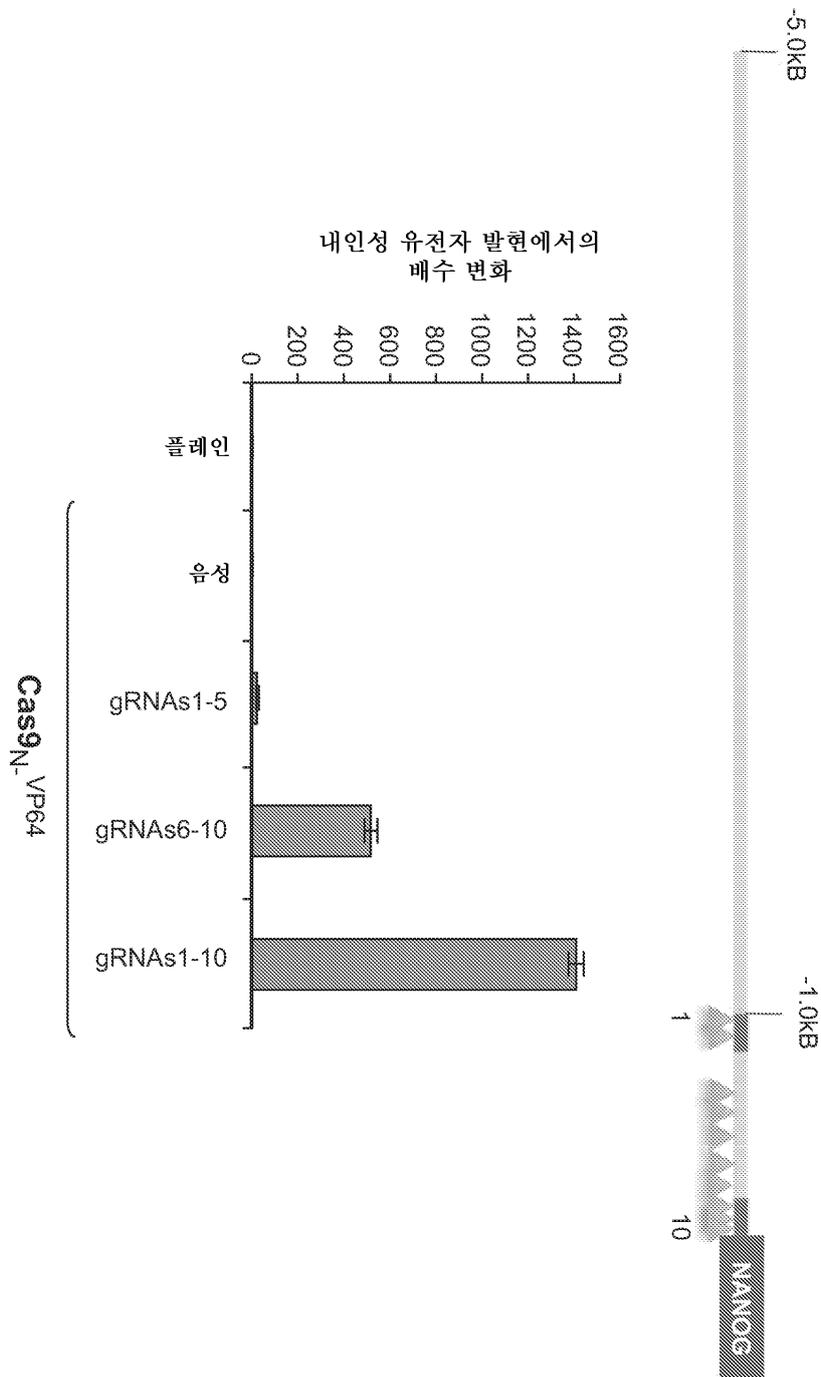


도면11b

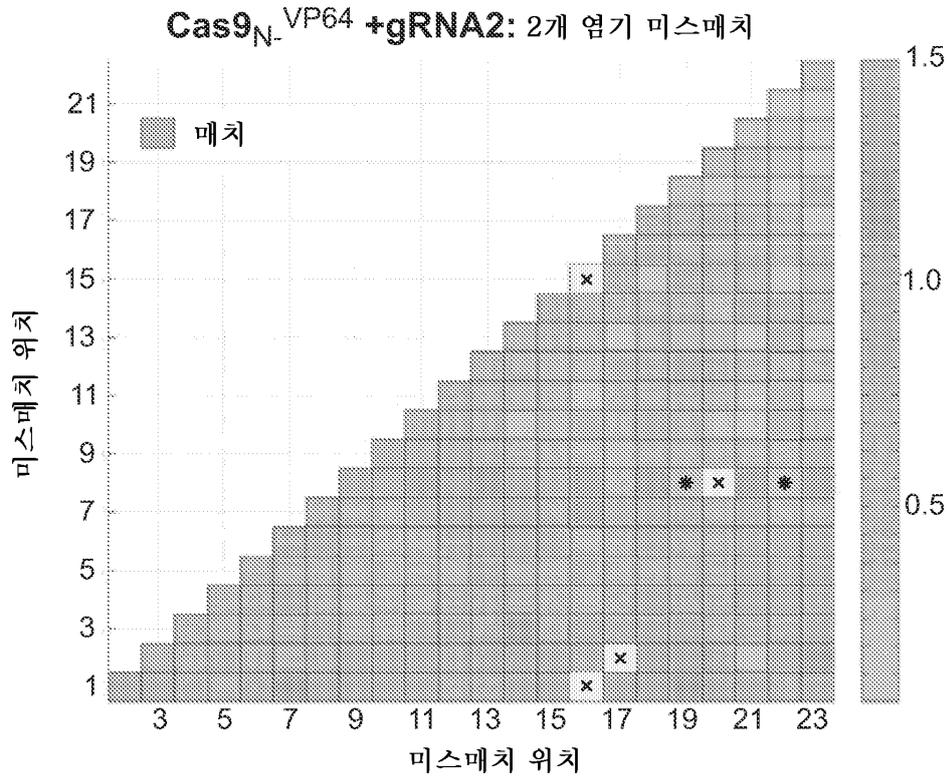
도면12a



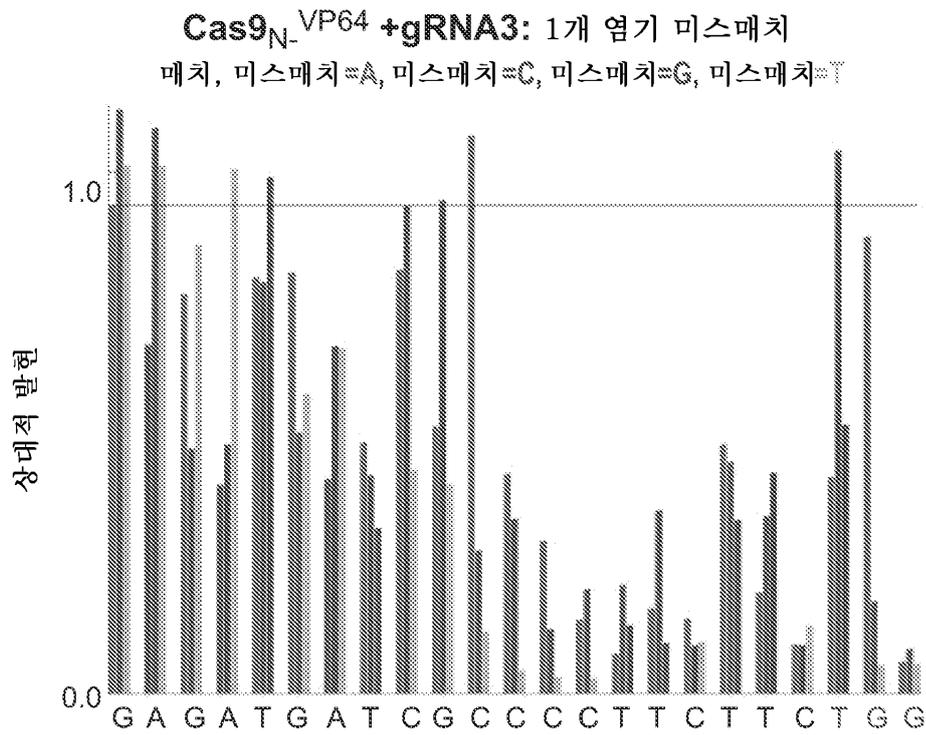
도면12b



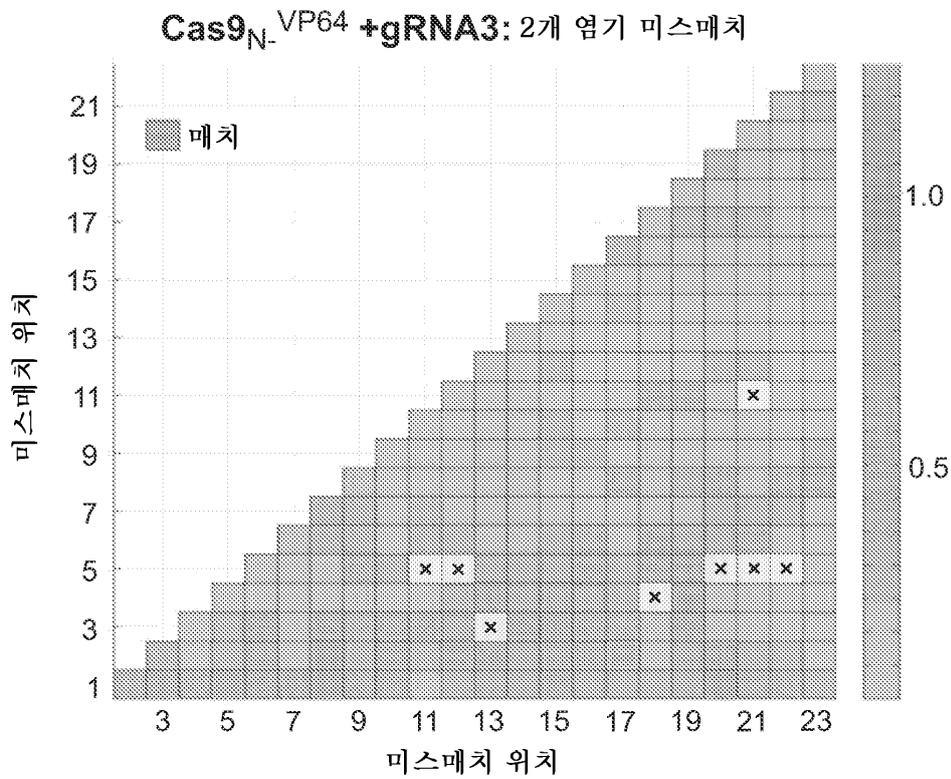
도면13c



도면13e



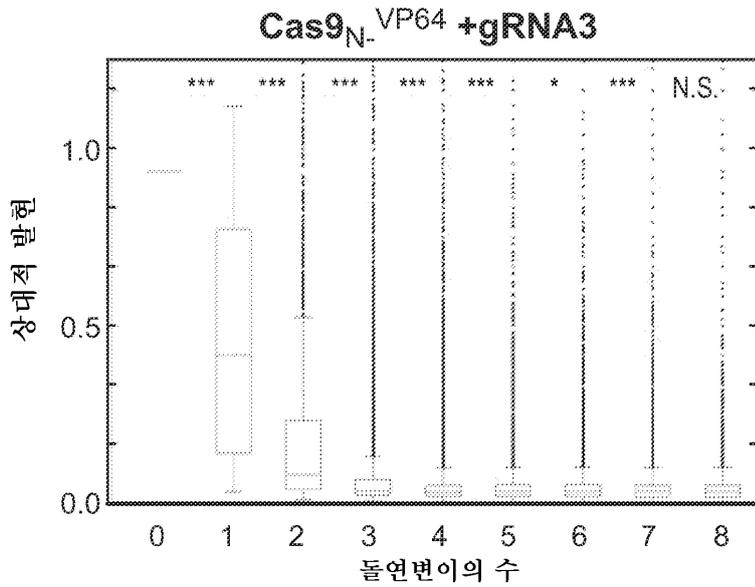
도면13f



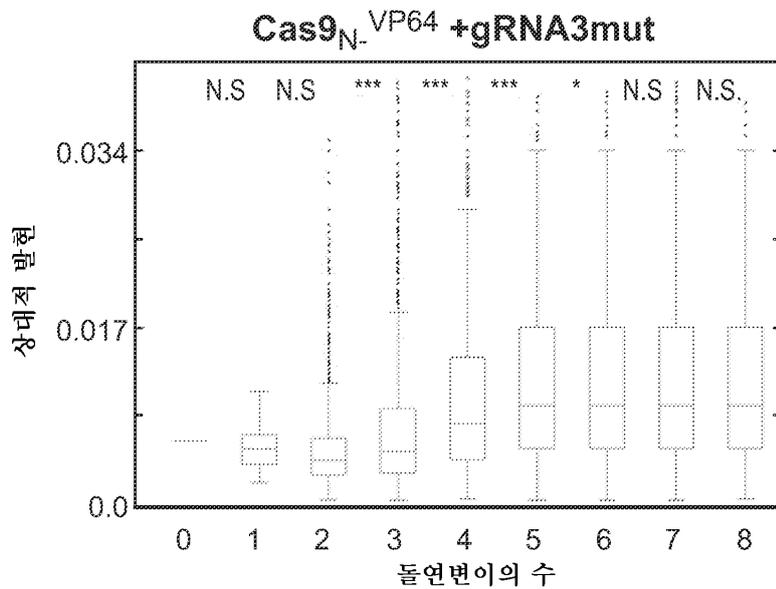
도면14a

표적 : GAGATGATCGCCCCTTCTTC TGG
 gRNA3 : GAGATGATCGCCCCTTCTTC
 gRNA3mut : GTGATGACCGGCCGTTCTTC

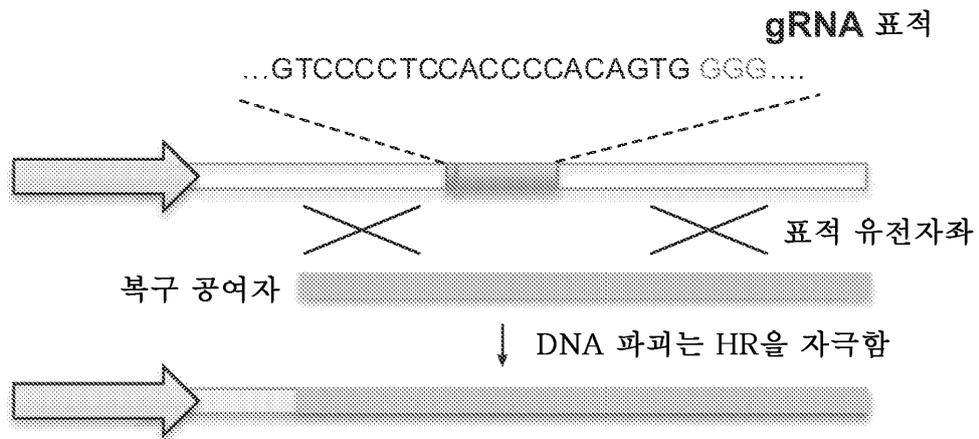
도면14b



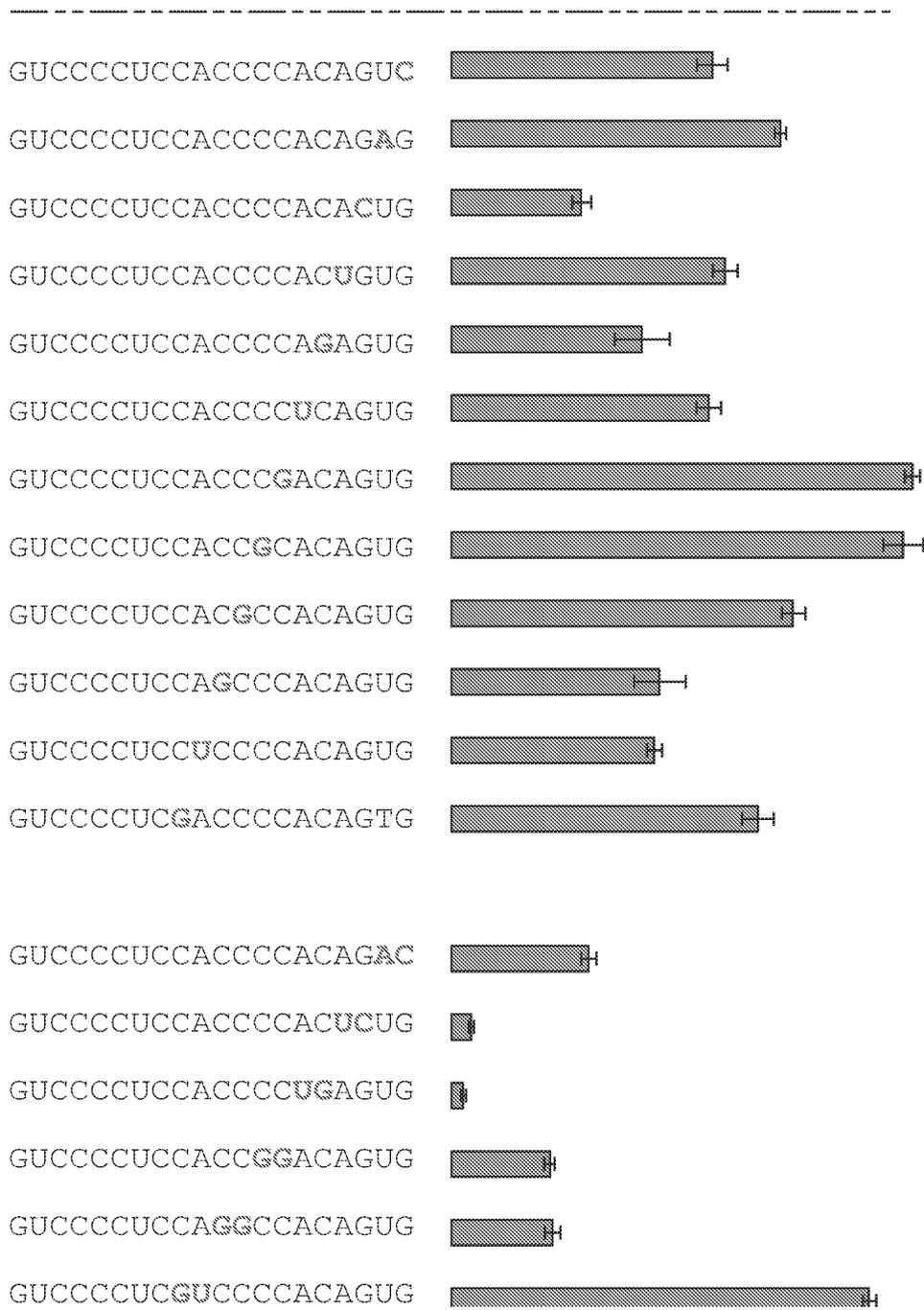
도면14c



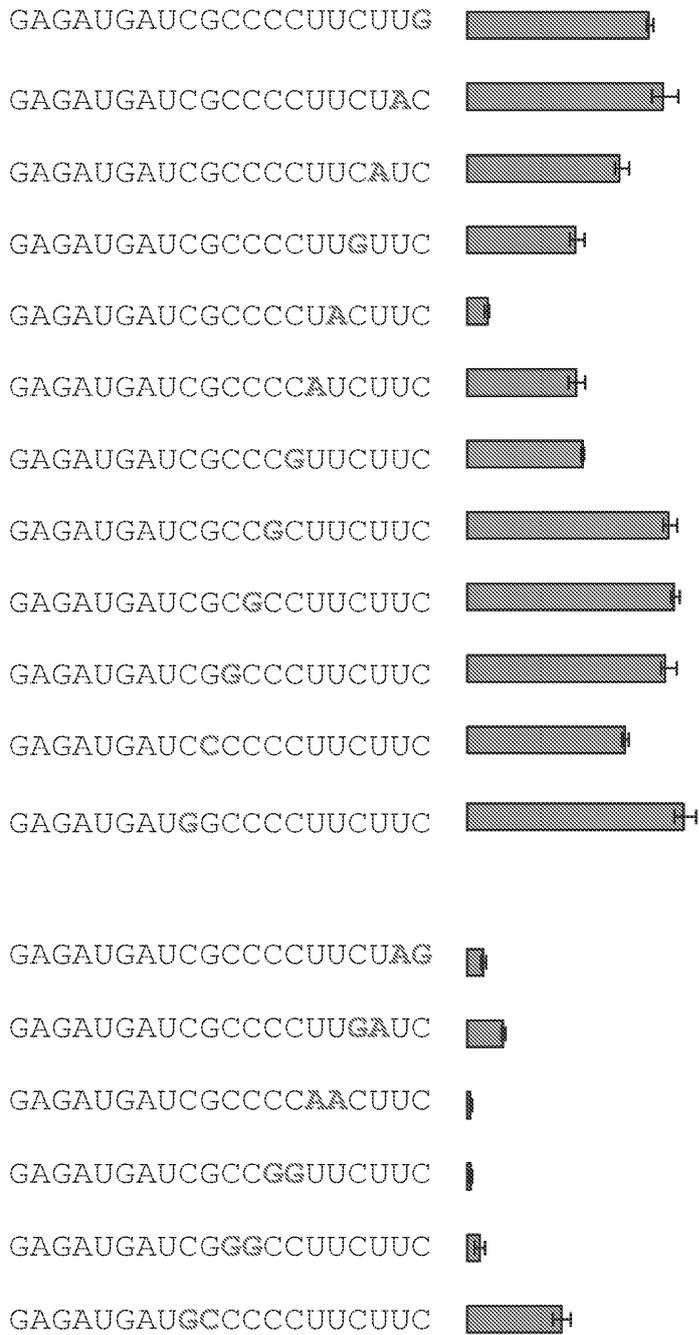
도면15a



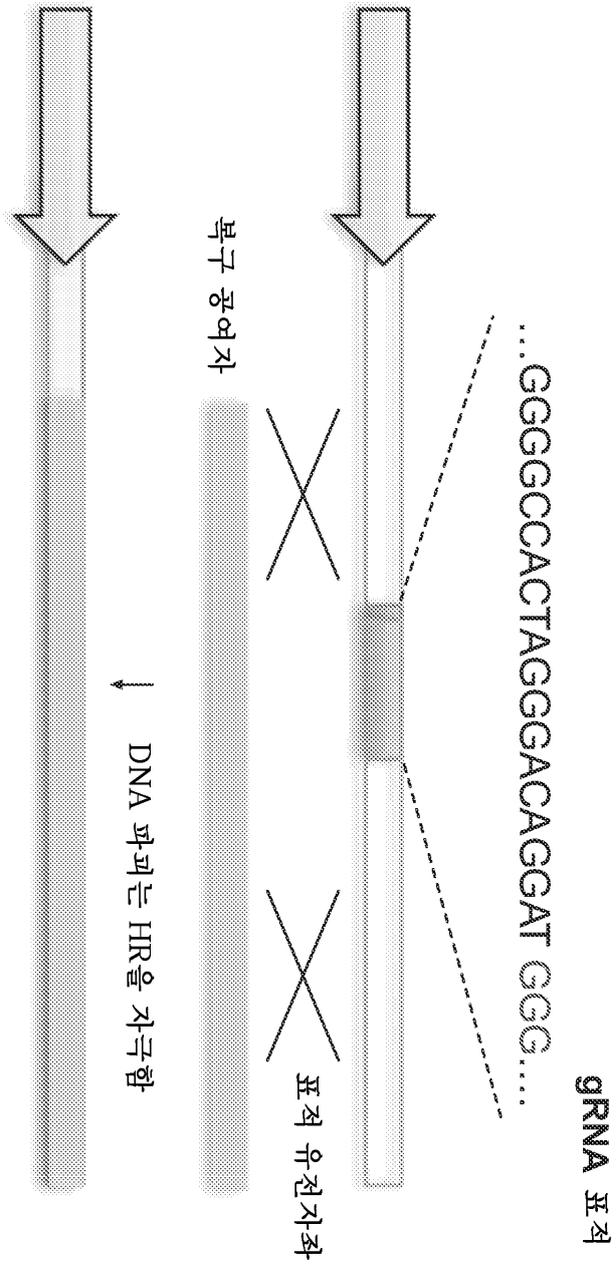
도면 15bb



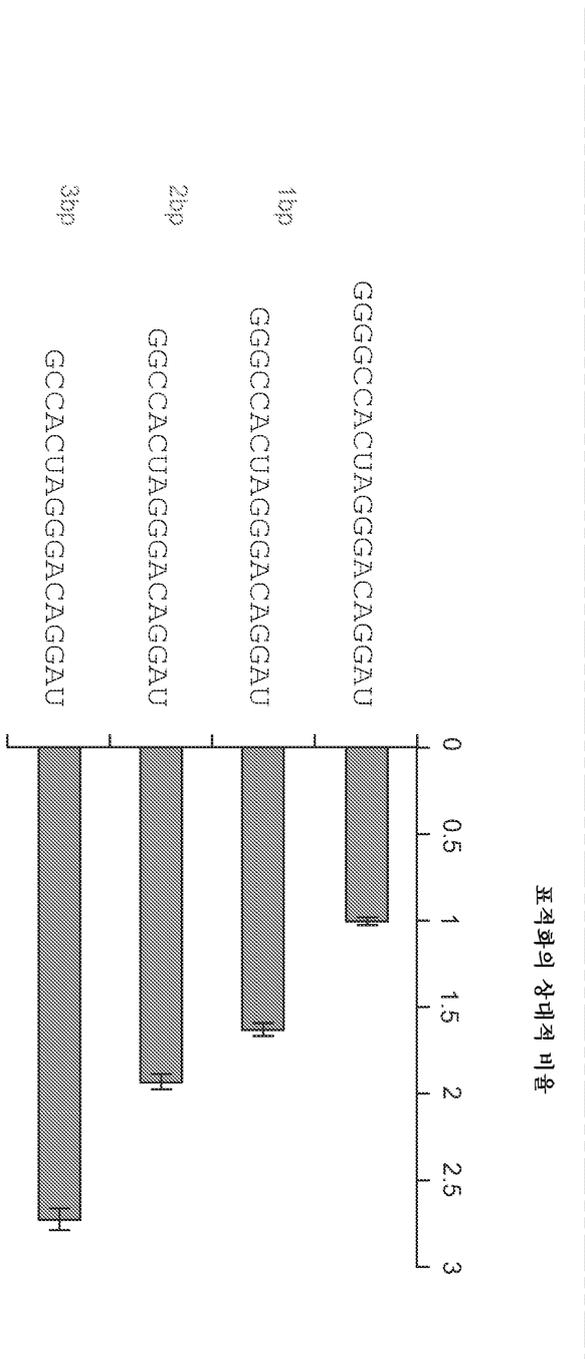
도면15db



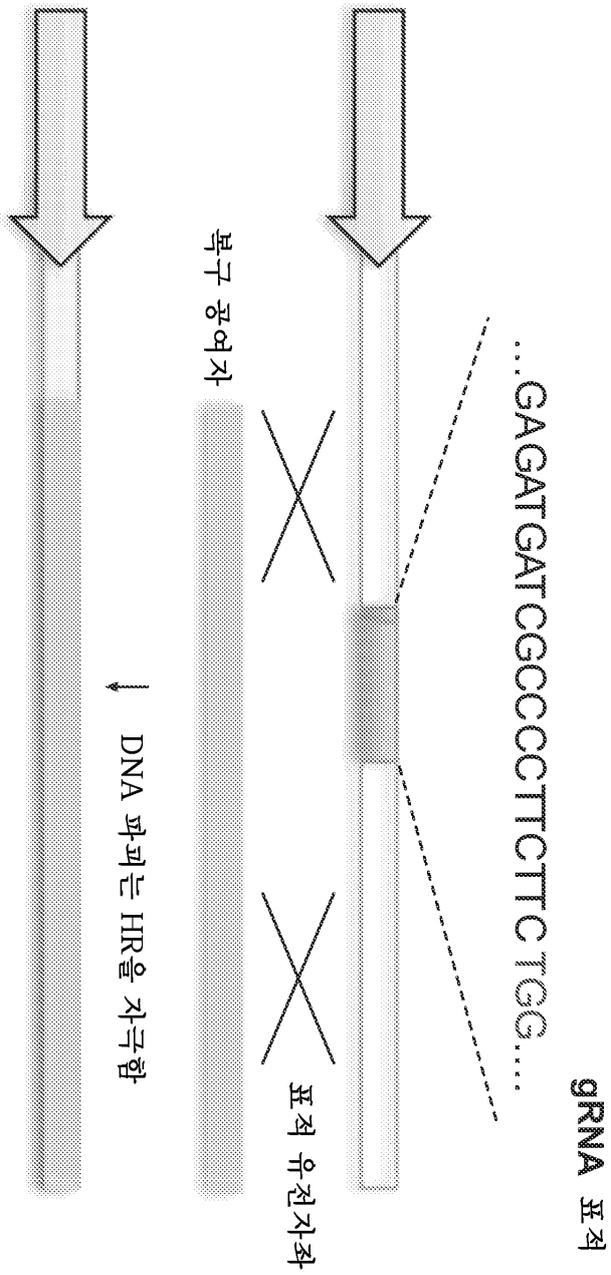
도면16a



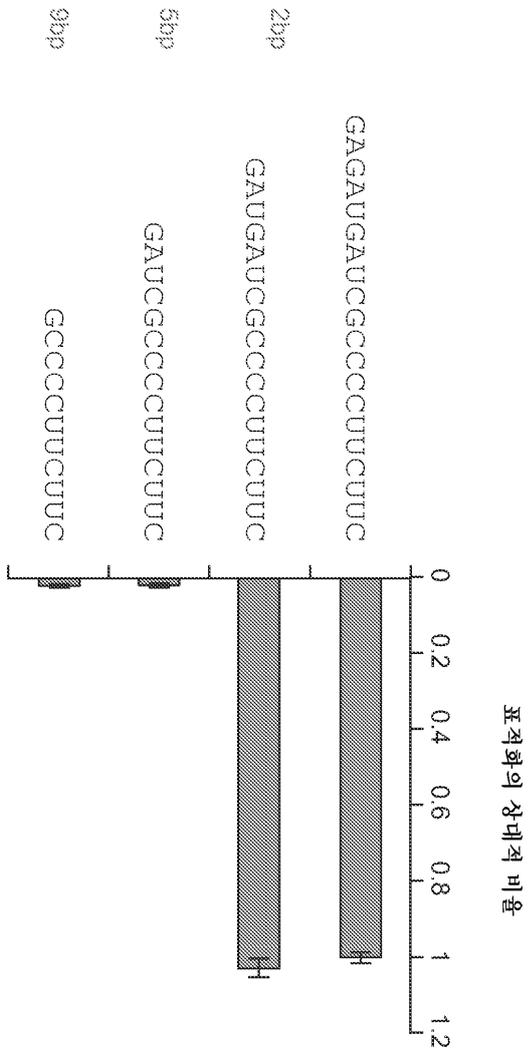
도면16bb



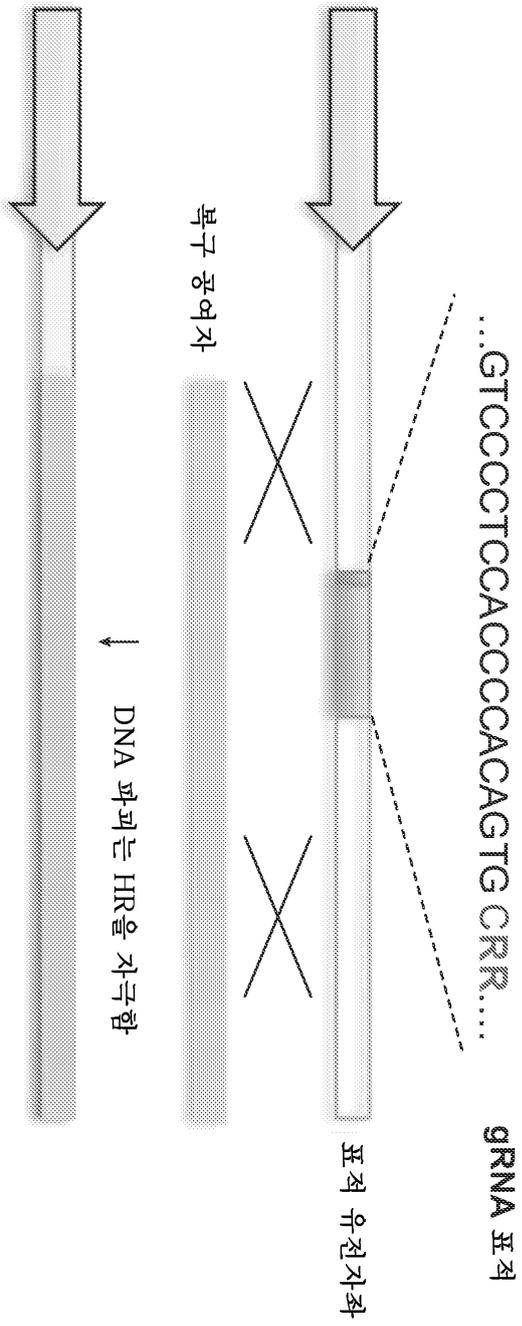
도면16c



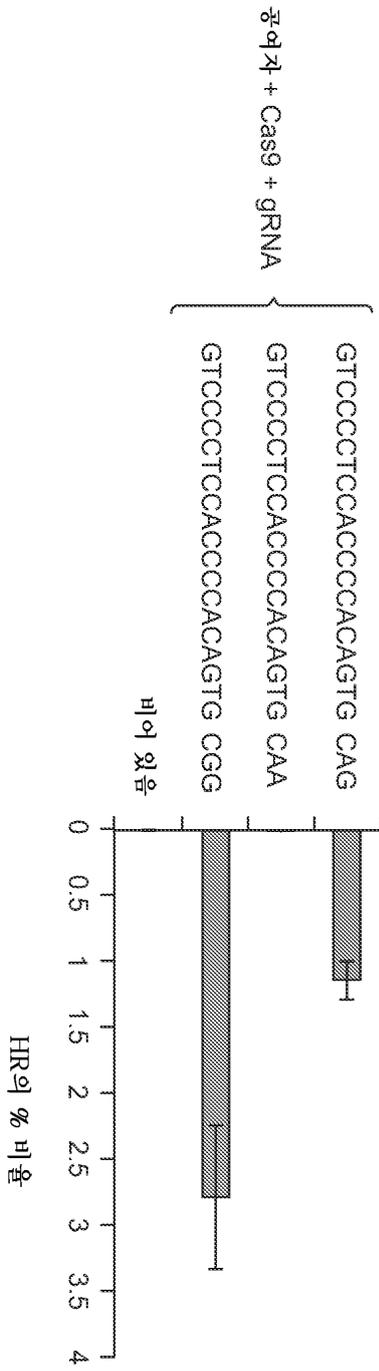
도면16db



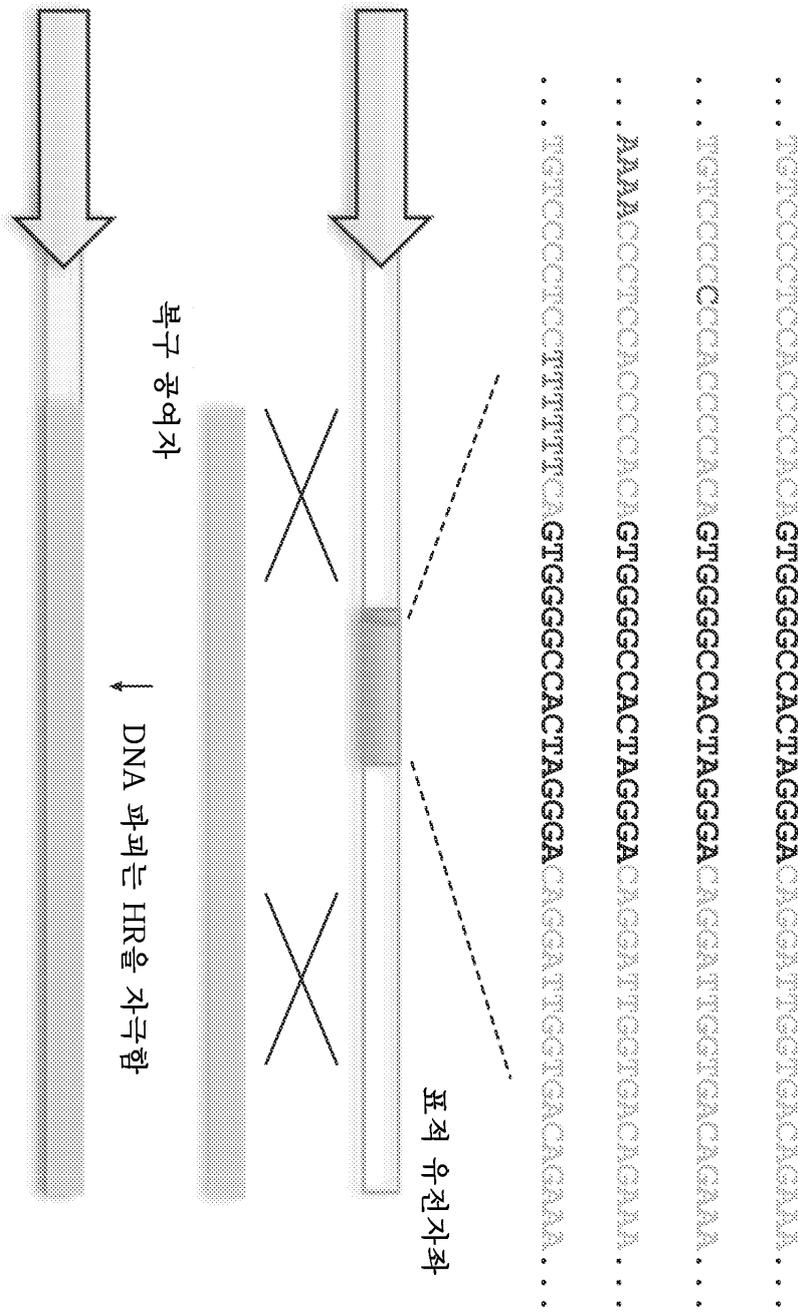
도면17a



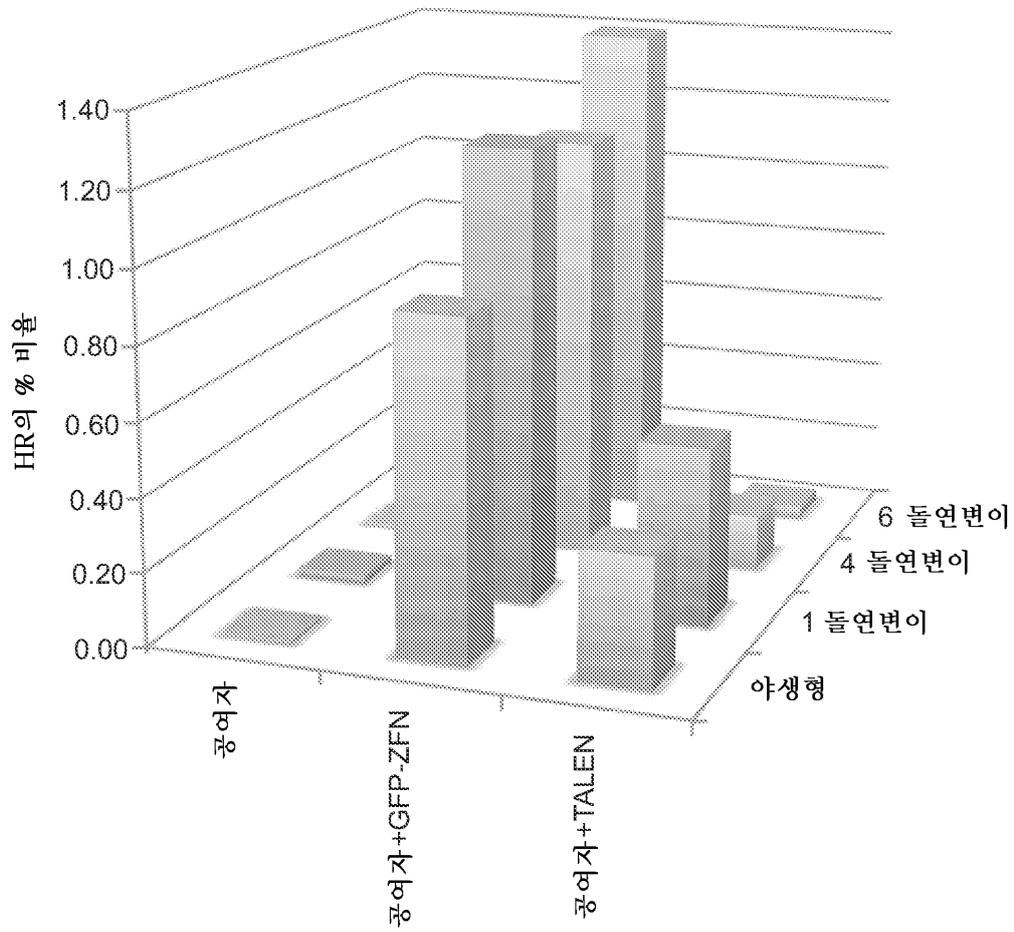
도면17b



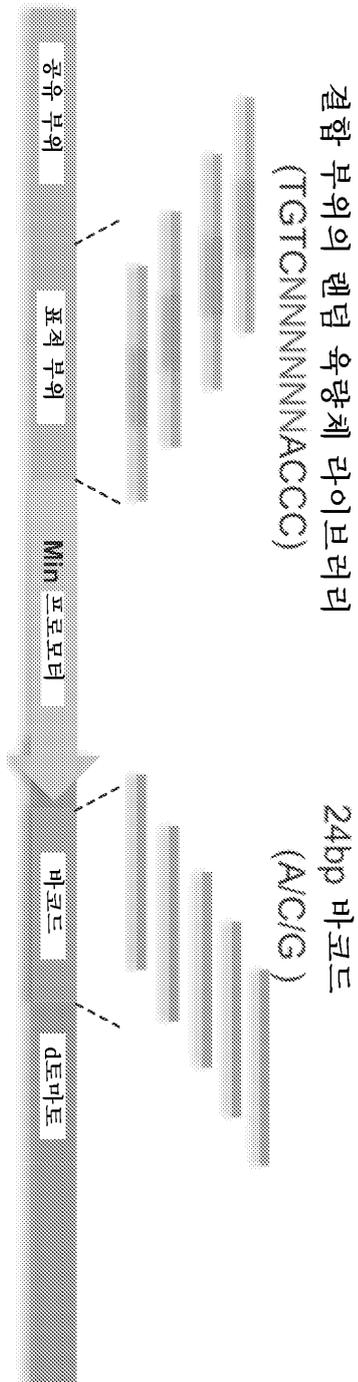
도면18a



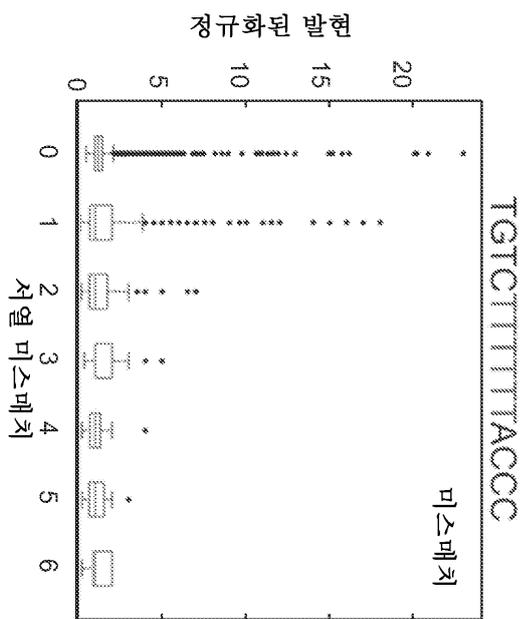
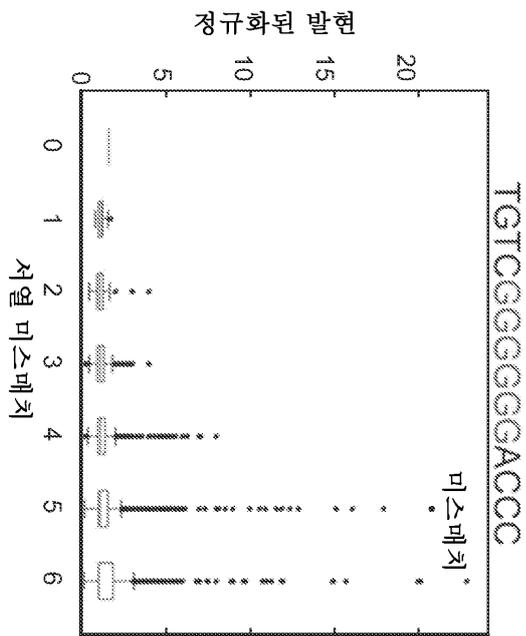
도면18b



도면19a

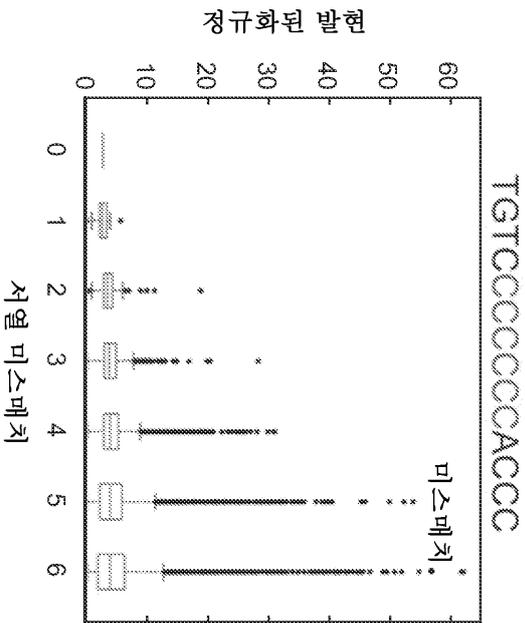
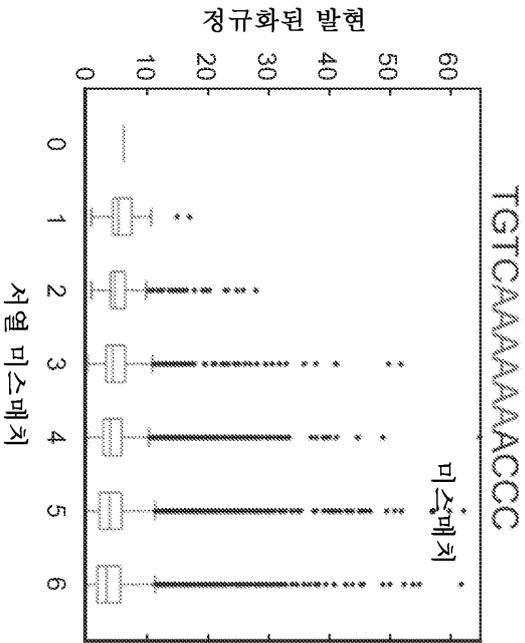


도면19bb

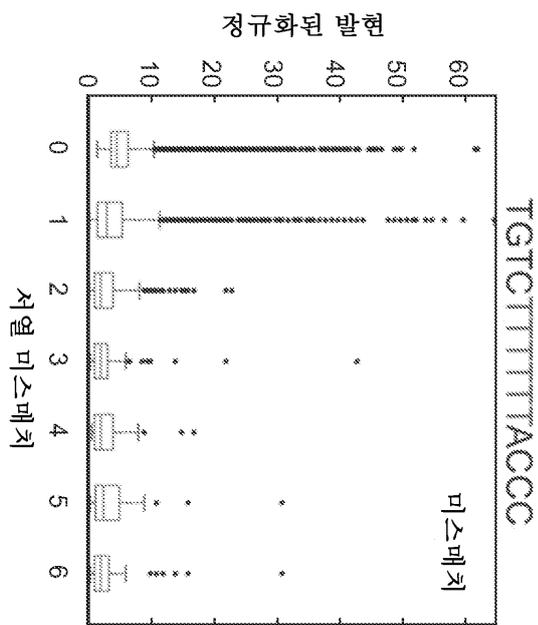
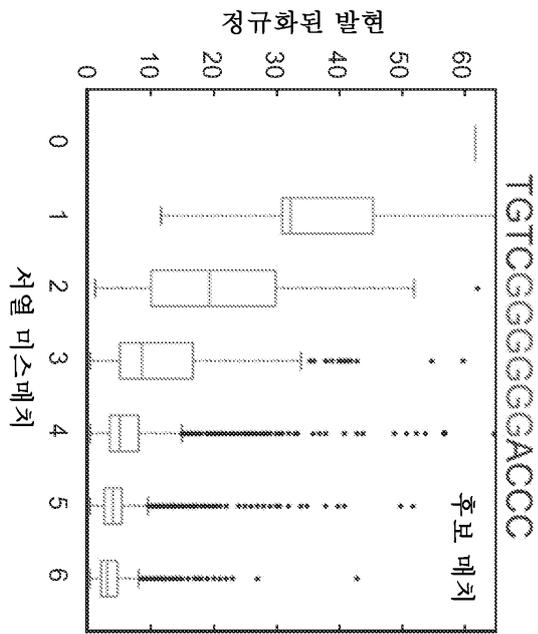


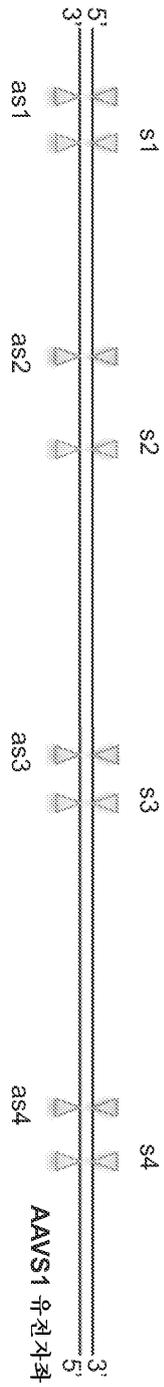
도면19ca

표적 TALE-TF: NG NN NG HD NH NH NH NH NH NH NH NH NH NI HD HD HD



도면19cb



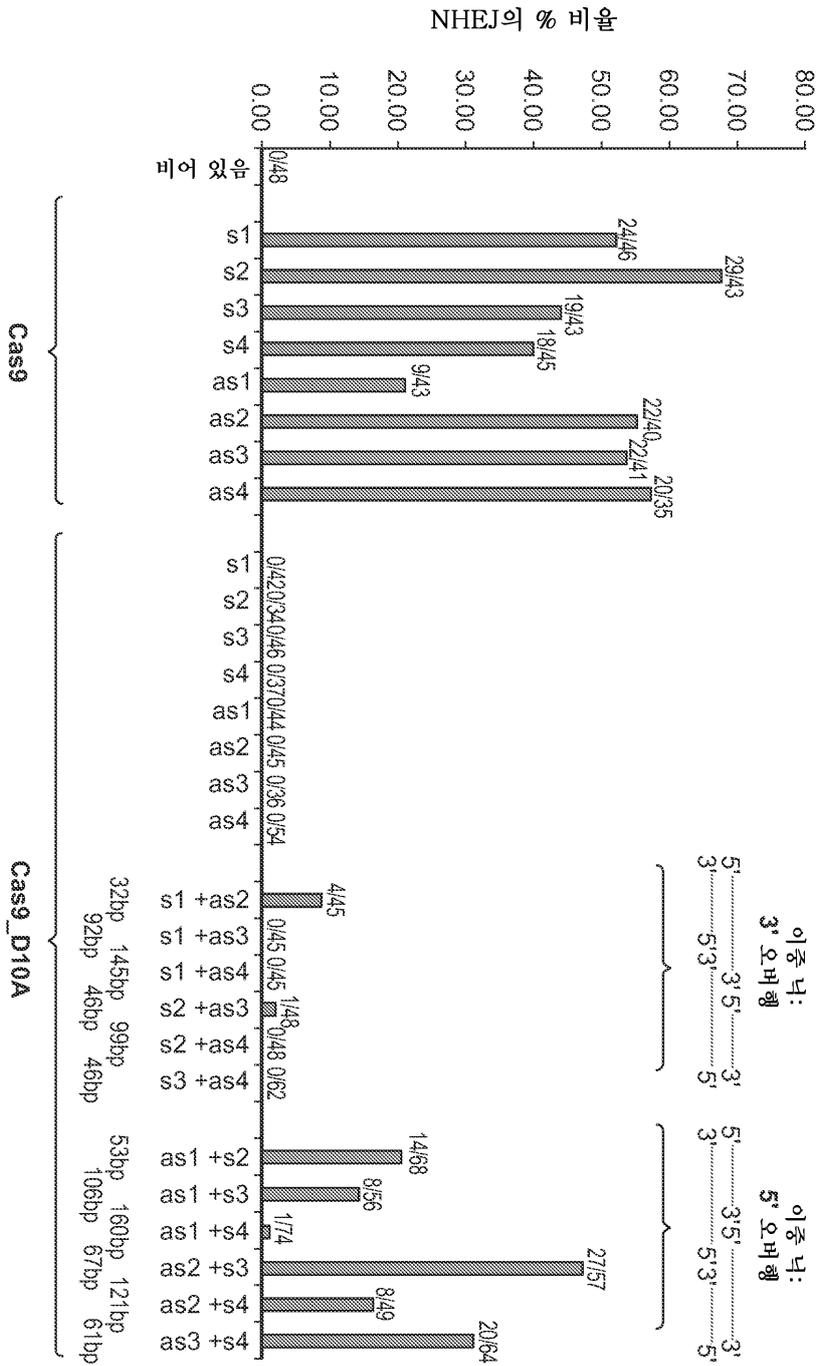


gRNA	서열
------	----

AAVS1__s1	GGATCCCTGTGTCCCCGAGCT GGG
AAVS1__s2	GTTAATGTGGCTCTGGTTCT GGG
AAVS1__s3	GGGCCCACTAGGGACAGGAT TGG
AAVS1__s4	CTTCTAGTCTCTCCTGATATT GGG
AAVS1__as1	TGGTCCCAGCTCGGGGACAC AGG
AAVS1__as2	AGAACCAGAGCCACATTAAC CCG
AAVS1__as3	GTCACCAATCCTGTCCCTAG TGG
AAVS1__as4	AGACCCAATATCAGGAGACT AGG

도면20a

도면20b



Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
 180 185 190
 Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
 195 200 205
 Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
 210 215 220
 Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
 225 230 235 240
 Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
 245 250 255
 Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
 260 265 270
 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

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
 770 775 780
 Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 785 790 795 800
 Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu
 805 810 815
 Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg
 820 825 830
 Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys
 835 840 845
 Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
 850 855 860
 Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 865 870 875 880
 Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 885 890 895
 Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr

915 920 925
 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960
 Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

 980 985 990
 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020
 Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
 1025 1030 1035
 Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

 1040 1045 1050
 Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
 1055 1060 1065
 Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
 1070 1075 1080
 Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
 1085 1090 1095
 Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys

 1100 1105 1110
 Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
 1115 1120 1125
 Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
 1130 1135 1140
 Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
 1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

1160 1165 1170

Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys

1175 1180 1185

Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190 1195 1200

Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly

1205 1210 1215

Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val

1220 1225 1230

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

1250 1255 1260

His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys

1265 1270 1275

Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala

1280 1285 1290

Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn

1295 1300 1305

Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala

1310 1315 1320

Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser

1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr

1340 1345 1350

Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp

1355 1360 1365

<210> 2

<211> 4332

<212> DNA

<213> Artificial

<220><223> VP64-activator construct

<400> 2

gccacatgg acaagaagta ctccattggg ctcgctatcg gcacaaacag cgtcggctgg 60
 gccgtcatta cggacgagta caaggtgccg agcaaaaaat tcaaagttct gggcaatacc 120
 gatcgccaca gcataaagaa gaacctcatt ggcgccctcc tgttcgactc cggggagacg 180

gccgaagcca cgcggctcaa aagaacagca cggcgcagat ataccgcag aaagaatcgg 240
 atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc 300
 cataggctgg aggagtcttt tttggtggag gaggataaaa agcacgagcg ccaccaatc 360
 tttggcaata tcgtggacga ggtggcgtac catgaaaagt acccaacat atatcatctg 420
 aggaagaagc ttgtagacag tactgataag gctgacttgc ggttgatcta tctcgcgctg 480
 gcgcatatga tcaaatttcg gggacacttc ctcatcgagg gggacctgaa ccagacaac 540
 agcgaatgctg acaaactctt tatccaactg gttcagactt acaatcagct tttcgaagag 600

aaccgatca acgcatcgg agttgacgcc aaagcaatcc tgagcgtag getgtccaaa 660
 tcccggcggc tcgaaaact catcgcacag ctccctgggg agaagaagaa cggcctgttt 720
 ggtaatctta tcgccctgtc actcgggctg acccccaact ttaaactaa cttcgacctg 780
 gccgaagatg ccaagcttca actgagcaaa gacacctag atgatgatct cgacaatctg 840
 ctggcccaga tcggcgacca gtacgcagac ctttttttgg cggcaaagaa cctgtcagac 900
 gccattctgc tgagtgatat tctcgcgagt aacacggaga tcaccaaagc tccgctgagc 960
 gctagtatga tcaagcgcta tgatgagcac caccaagact tgactttgct gaaggccctt 1020

gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc 1080
 tacgccggat acattgacgg cggagcaagc caggaggaat tttaaaaatt tattaagccc 1140
 atcttggaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg 1200
 ttgcgcaaac agcgacttt cgacaatgga agcatcccc accagattca cctgggcgaa 1260
 ctgcacgcta tectcaggcg gcaagaggat ttctaccct tttgaaaga taacagggaa 1320
 aagattgaga aaatctcac atttcggata ccctactatg taggccccct cggccgggga 1380
 aattccagat tcgcgtggat gactcgcaaa tcagaagaga ccatcactcc ctggaacttc 1440

gaggaagtgc tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt 1500
 gataaaaate tgcctaacga aaagtgctt ctaaacact ctctgctgta cgagtacttc 1560
 acagtttata acgagctcac caagtcaaa tacgtcacag aagggatgag aaagccagca 1620
 ttctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa 1680

gttacctga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt 1740
gaaatcagcg gagtggagga tgcgttcaac gcatccctgg gaactatca cgatctcctg 1800
aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac 1860

attgtcctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact 1920
tacgtcctc tcttcgacga caaagtcacg aaacagctca agaggcgccg atatacagga 1980
tggggcgccg tgcacaagaa actgatcaat gggatccgag acaagcagag tggaaagaca 2040
atcctggatt ttcttaagtc cgatggattt gccaacgga acttcatgca gttgatccat 2100
gatgactctc tcaccttaa ggaggacatc cagaaaagcac aagtttctgg ccagggggac 2160
agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaaactg 2220
cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccgagaat 2280

atcgttatcg agatggcccc agagaaccaa actaccaga agggacagaa gaacagtagg 2340
gaaaggatga agaggattga agagggtata aaagaactgg ggtcccaat ccttaaggaa 2400
caccagttg aaaacaccca gcttcagaat gagaagctct acctgacta cctgcagaac 2460
ggcagggaca tgcactgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg 2520
gctgctatcg tgccccagtc ttttctcaaa gatgattcta ttgataataa agtgttgaca 2580
agatccgata aagctagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa 2640
atgaaaaatt attggcggca gctgctgaac gccaaactga tcacacaacg gaagttcgat 2700

aatctgacta agctgaacg aggtggcctg tctgagtgg ataaagccgg cttcatcaaa 2760
aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaaattct cgattcacgc 2820
atgaacacca agtacgatga aaatgacaaa ctgattcgag aggtgaaagt tattactctg 2880
aagtctaage tggctcaga tttcagaaag gactttcagt tttataaggt gagagagatc 2940
aacaattacc accatgcgca tgatgcctac ctgaatgcag tggtaggcac tgcacttatc 3000
aaaaaatatc ccaagcttga atctgaattt gtttacggag actataaagt gtacgatgtt 3060
aggaaaatga tcgcaaagtc tgagcaggaa ataggcaagg ccaccgctaa gtacttcttt 3120

tacagcaata ttatgaattt tttcaagacc gagattacac tggccaatgg agagattcgg 3180
aagcgaccac ttatcgaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg 3240
gatttcgca cagtcaggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc 3300
gaagtacaga ccggaggctt ctccaaggaa agtatcctcc cgaagaggaa cagcgacaag 3360
ctgatcgac gcaaaaaaga ttgggacccc aagaaatagc gcggattcga ttctctaca 3420
gtcgttaca gtgtactggt tgtggccaaa gtggagaaag ggaagtctaa aaaactcaaa 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 cctctaaat acgttaattt cttgtatctg 3720
 gccagccact atgaaaagct caaaggtctt cccgaagata atgagcagaa gcagctgttc 3780
 gtggaacaac acaaacacta ccttgatgag atcatcgagc aaataagcga attctccaaa 3840
 agagtgatcc tcgccgacgc taacctcgat aaggtgcttt ctgcttacia taagcacagg 3900
 gataagccca tcagggagca ggcagaaaac attatccact tgtttactct gaccaacttg 3960

 ggcgcgctg cagccttcaa gtacttcgac accacatag acagaaagcg gtacacctct 4020
 acaaaggagg tcctggacgc cacactgatt catcagtcaa ttacggggct ctatgaaaca 4080
 agaatcgacc tctctcagct cggtggagac agcagggtcg accccaagaa gaagaggaag 4140
 gtggaggcca gcggttcgag acgggctgac gcattggacg atttgatct ggatatgctg 4200
 ggaagtgacg cctcctgatg ttttgacctt gacatgcttg gttcggatgc ccttgatgac 4260
 tttgacctcg acatgctcgg cagtgcgccc cttgatgatt tcgacctgga catgctgatt 4320
 aactctagat ga 4332

 <210> 3
 <211> 4365
 <212> DNA
 <213> Artificial
 <220><223> VP64-activator construct
 <400> 3

 gccaccatgc ccaagaagaa gaggaaggtg ggaaggggga tggacaagaa gtactccatt 60
 gggctcgcta tcggcacaaa cagcgtcggc tgggccgtca ttacggacga gtacaaggtg 120
 ccgagcaaaa aattcaaagt tctgggcaat accgatcgcc acagcataaa gaagaacctc 180
 attggcgccc tectgttcga ctccggggag acggccgaag ccacgcgct caaagaaca 240
 gcacggcgca gatatacccg cagaaagaat cggatctgct acctgcagga gatcttagt 300
 aatgagatgg ctaaggtgga tgactcttcc ttccataggc tggaggagtc ctttttgggtg 360

 gaggaggata aaaagcacga gcgccacca atctttggca atatcgtgga cgaggaggcg 420
 taccatgaaa agtaccacac catatatcat ctgaggaaga agctttaga cagtactgat 480
 aaggtgact tcggtttgat ctatctcgcg ctggcgcata tgatcaaatt tcggggacac 540
 ttctcatcg agggggacct gaaccagac aacagcgatg tcgacaaact ctttatccaa 600

ctggttcaga cttacaatca gcttttcgaa gagaaccgga tcaacgcata cggagttgac 660
 gccaaagcaa tcctgagcgc taggctgtcc aaatcccggc ggctcgaaaa cctcatcgca 720
 cagctccctg gggagaagaa gaacggcctg tttggtaatc ttatcgccct gtcactcggg 780

 ctgaccccca acittaaatc taacttcgac ctggccgaag atgccaagct tcaactgagc 840
 aaagacacct acgatgatga tctcgacaat ctgctggccc agatcggcga ccagtacgca 900
 gacctttttt tggcggcaaa gaacctgtca gacgccattc tgctgagtga tattctcgca 960
 gtgaacacgg agatcaccaa agctccgctg agcgcctagta tgatcaagcg ctatgatgag 1020
 caccaccaag acttgacttt gctgaaggcc ctgttcagac agcaactgcc tgagaagtac 1080
 aaggaaatth tcttcgatca gtctaaaaat ggctacgccg gatacattga cggcggagca 1140
 agccaggagg aatthtaca atthattaag cccatcttgg aaaaaatgga cggcaccgag 1200

 gagctgctgg taaagcttaa cagagaagat ctgttgcgca aacagcgcac tttcgacaat 1260
 ggaagcatcc cccaccagat tcacctgggc gaactgcacg ctatcctcag gcggcaagag 1320
 gatttctacc ctttttgaa agataacagg gaaaagattg agaaaatcct cacatttcgg 1380
 atacctact atgtaggccc cctcgcccg ggaattcca gattcgctg gatgactcgc 1440
 aatcagaag agaccatcac tccttggaaac ttcgaggaag tcgtggataa gggggcctct 1500
 gccagtcct tcatgaaag gatgactaac tttgataaaa atctgcctaa cgaaaaggtg 1560
 ctctctaacc actctctgct gtacgagtac ttcacagttt ataacgagct caccaaggtc 1620

 aaatacgtca cagaagggat gagaaagcca gcattcctgt ctggagagca gaagaaagct 1680
 atcgtggacc tctcttcaa gacgaaccgg aaagttaccg tgaaacagct caaagaagac 1740
 tatttcaaaa agattgaatg tttcgactct gttgaaatca gcggagtgga ggatcgcttc 1800
 aacgcatccc tgggaacgta tcacgatctc ctgaaaatca ttaaagaca ggacttctg 1860
 gacaatgagg agaacgagga cattcttgag gacattgtcc tcacccttac gttgtttgaa 1920
 gatagggaga tgattgaaga acgcttgaaa acttacgctc atctcttga cgacaaagtc 1980
 atgaaacagc tcaagaggcg ccgatataca ggatgggggc ggctgtcaag aaaactgatc 2040

 aatgggatcc gagacaagca gagtggaaag acaatcctgg atthtcttaa gtccgatgga 2100
 tttgccaacc ggaacttcat gcagttgatc catgatgact ctctcacctt taaggaggac 2160
 atccagaaag cacaagtttc tggccagggg gacagtcttc acgagcacat cgctaactt 2220
 gcaggtagcc cagctatcaa aaagggaata ctgcagaccg ttaaggtcgt ggatgaactc 2280
 gtcaaagtaa tgggaaggca taagcccag aatatcgtha tcgagatgac ccgagagaac 2340
 caaactacce agaagggaca gaagaacagt agggaaaagga tgaagaggat tgaagagggt 2400

ataaaagaac tggggtccca aatccttaag gaacacccag ttgaaaacac ccagcttcag 2460

 aatgagaagc tctacctgta ctacctgcag aacggcaggg acatgtacgt ggatcaggaa 2520
 ctggacatca atcggctctc cgactacgac gtggctgcta tcgtgccca gtcttttctc 2580
 aaagatgatt ctattgataa taaagtgttg acaagatccg ataaagctag agggaagagt 2640
 gataacgtcc cctcagaaga agttgtcaag aaaatgaaaa attattggcg gcagctgctg 2700
 aacgccaaac tgatcacaca acggaagttc gataatctga ctaaggctga acgaggtggc 2760
 ctgtctgagt tggataaagc cggcttcac taaaaggcagc ttgttgagac acgccagatc 2820
 accaagcacg tggcccaaat tctcgattca cgcataaaca ccaagtacga tgaaaatgac 2880

 aaactgattc gagaggigaa agttattact ctgaagtcta agctggcttc agatttcaga 2940
 aaggactttc agttttataa ggtgagagag atcaacaatt accacatgc gcatgatgcc 3000
 tactgaatg cagtggtagg cactgcactt atcaaaaaat atccaagct tgaatctgaa 3060
 tttgtttacg gagactataa agtgtacgat gttaggaaaa tgatcgaaa gtctgagcag 3120
 gaaataggca aggccaccgc taagtacttc ttttacagca atattatgaa tttttcaag 3180
 accgagatta cactggccaa tggagagatt cggaaagcgc cacttatcga aacaaacgga 3240
 gaaacaggag aaatcgtgtg ggacaagggt agggatttcg cgacagtccg gaaggtcctg 3300

 tccatgccgc aggtgaacat cgttaaaaag accgaagtac agaccggagg ctctccaag 3360
 gaaagtatcc tcccgaaaag gaacagcgcac aagctgatcg cacgcaaaaa agattgggac 3420
 cccaagaaat acggcggatt cgattctcct acagtcgctt acagtgact ggttgtggcc 3480
 aaagtggaga aagggaagtc taaaaaactc aaaagcgtca aggaactgct gggcatcaca 3540
 atcatggagc gatcaagctt cgaaaaaac cccatcgact ttctcgagc gaaaggatat 3600
 aaagaggtca aaaaagacct catcattaag ctcccaagt actctctctt tgagcttgaa 3660
 aacggccgga aacgaatgct cgctagtgcg ggcgagctgc agaaaggtaa cgagctggca 3720

 ctgccctcta aatagttaa tttctgtat ctggccagcc actatgaaaa gctcaaaggg 3780
 tctcccgaag ataatagca gaagcagctg ttcgtggaac aacacaaaca ctacctgat 3840
 gagatcatcg agcaataag cgaattctcc aaaagagtga tcctcgccga cgtaacctc 3900
 gataaggtgc tttctgctta caataagcac agggataagc ccatcagga gcaggcagaa 3960
 aacattatcc acttgttac tctgaccaac ttgggcgcgc ctgcagcctt caagtacttc 4020
 gacaccacca tagacagaaa gcggtacacc tctacaaagg aggtcctgga cgccacactg 4080
 attcatcagt caattacggg gctctatgaa acaagaatcg acctctctca gctcgggtgga 4140

gacagcaggg ctgaccccaa gaagaagagg aaggtggagg ccagcggttc cggacgggct 4200
gacgcattgg acgattttga tctggatatg ctgggaagtg acgccctcga tgattttgac 4260
cttgacatgc ttggttcgga tgcctttgat gactttgacc tcgacatgct cggcagtgac 4320
gcccttgatg atttcgacct ggacatgctg attaactcta gatga 4365
<210> 4
<211> 4425
<212> DNA
<213> Artificial
<220><223> VP64-activator construct
<400> 4
gccaccatgg acaagaagta ctccattggg ctgctatcg gcacaaacag cgtcggctgg 60
gccgtcatta cggacgagta caaggtgccg agcaaaaaat tcaaagttct gggcaatacc 120
gatcgccaca gcataaagaa gaacctcatt ggcgcctcc tgttcgactc cggggagacg 180
gccgaagcca cgcggctcaa aagaacagca cggcgcagat ataccgcgag aaagaatcgg 240
atctgctacc tgcaggagat ctttagtaat gagatggcta aggtggatga ctctttcttc 300
cataggctgg aggagtcctt tttggtggag gaggataaaa agcacgagcg ccaccaatc 360
tttggcaata tcgtggacga ggtggcgtac 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 tgagcgttag gctgtccaaa 660
tcccggcggc tcgaaaacct catcgcacag ctccctgggg agaagaagaa cggcctgttt 720
ggtaatctta tcgccctgtc actcgggctg acccccaact ttaaatctaa 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 gaaggcctt 1020
gtcagacagc aactgcctga gaagtacaag gaaattttct tcgatcagtc taaaaatggc 1080
tacgccgatg acattgacgg cggagcaagc caggaggaat ttacaaatt tattaagccc 1140
atcttgaaa aaatggacgg caccgaggag ctgctggtaa agcttaacag agaagatctg 1200
ttgcgcaaac agcgcacttt cgacaatgga agcatcccc accagattca cctgggcgaa 1260
ctgcacgcta tcctcaggcg gcaagaggat ttctaccct ttttgaaaga taacagggaa 1320

aagattgaga aaatcctcac atttcggata ccctactatg taggccccct cgcccgggga 1380

aattccagat tcgctggat gactcgc aaa tcagaagaga ccatcactcc ctggaacttc 1440
gaggaagtcg tggataaggg ggcctctgcc cagtccttca tcgaaaggat gactaacttt 1500
gataaaaatc tgctaacga aaagtgctt cctaaacact ctctgctgta cgagtacttc 1560
acagtttata acgagctcac caaggtcaaa tacgtcacag aagggatgag aaagccagca 1620
ttcctgtctg gagagcagaa gaaagctatc gtggacctcc tcttcaagac gaaccggaaa 1680
gttaccgtga aacagctcaa agaagactat ttcaaaaaga ttgaatgttt cgactctgtt 1740
gaaatcagcg gagtggagga tcgcttcaac gcatccctgg gaacgtatca cgatctctg 1800

aaaatcatta aagacaagga cttcctggac aatgaggaga acgaggacat tcttgaggac 1860
attgtcctca cccttacgtt gtttgaagat agggagatga ttgaagaacg cttgaaaact 1920
tacgtcctc tcttcagca caaagtcatg aaacagctca agaggcgccg atatacagga 1980
tggggcggc tgc aagaaa actgatcaat gggatccgag acaagcagag tggaaagaca 2040
atcctggatt ttcttaagtc cgatggattt gccaacgga acttcatgca gttgatccat 2100
gatgactctc tcaccttaa ggaggacatc cagaaagcac aagtttctgg ccagggggac 2160
agtcttcacg agcacatcgc taatcttgca ggtagcccag ctatcaaaaa gggaatctg 2220

cagaccgtta aggtcgtgga tgaactcgtc aaagtaatgg gaaggcataa gcccgagaat 2280
atcgttatcg agatggcccc agagaaccaa actaccaga agggacagaa gaacagtagg 2340
gaaaggatga agaggattga agaggtata aaagaactgg ggtcccaaat ccttaaggaa 2400
caccagttg aaaacacca gcttcagaat gagaagctct acctgacta cctgcagaac 2460
ggcagggaca tgiacgtgga tcaggaactg gacatcaatc ggctctccga ctacgacgtg 2520
gctgctatcg tgccccagtc ttttctcaaa gatgattcta ttgataataa agtgttgaca 2580
agatccgata aagctagagg gaagagtgat aacgtcccct cagaagaagt tgtcaagaaa 2640

atgaaaaatt attggcgga gctgctgaac gccaaactga tcacacaacg gaagttcgat 2700
aatctgacta aggtgaacg aggtggcctg tctgagtgg ataaagccgg cttcatcaa 2760
aggcagcttg ttgagacacg ccagatcacc aagcacgtgg cccaattct cgattcacgc 2820
atgaacacca agtacgatga aatgacaaa ctgattcgag aggtgaaagt tattactctg 2880
aagtctaagc tggctcaga tttcagaaaag gactttcagt tttataaggt 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 ttatcgaaac aaacggagaa acaggagaaa tcgtgtggga caagggtagg 3240
 gatttcgca cagtccggaa ggtcctgtcc atgccgcagg tgaacatcgt taaaaagacc 3300
 gaagtacaga ccggaggctt ctccaaggaa agtatcctcc cgaaggaa cagcgacaag 3360
 ctgatcgac gcaaaaaaga ttgggacccc aagaaatcgc gcggattcga ttctctaca 3420
 gtcgcttaca gtgtactggt tgtggccaaa gtggagaaag ggaagtctaa aaaactcaaa 3480

agcgtcaagg aactgctggg catcacaatc atggagcgat caagcttcga aaaaaacccc 3540
 atcgacttcc tcgaggcgaa aggatataaa gaggtcaaaa aagacctcat cattaagctt 3600
 cccaagtact ctctcttga gcttgaaaac ggccggaaac gaatgctcgc tagtgcgggc 3660
 gagctgcaga aagtaacga gctggcactg cctctaaat acgttaattt cttgtatctg 3720
 gccagccact atgaaaagct caaagggtct cccgaagata atgagcagaa gcagctgttc 3780
 gtggaacaac acaaacacta cttgatgag atcatcgagc aaataagcga attctcaaaa 3840
 agagtgatec tcgccgacgc taacctgat aaggtgcttt ctgcttaca taagcacagg 3900

gataagccca tcaggagca ggcagaaaac attatccact tgtttactct gaccaacttg 3960
 ggcgcgctg cagccttcaa gtacttcgac accacatag acagaaagcg gtacacctct 4020
 acaaaggagg tctggacgc cacactgatt catcagtcaa ttacgggct ctatgaaaca 4080
 agaatcgacc tctctcagct cgggtggagac agcagggtc accccaagaa gaagaggaag 4140
 gtggaggcca gcggttccgg acgggctgac gcattggacg atttgatct ggatatgctg 4200
 ggaagtgacg cctcgtatg ttttgacctt gacatgcttg gttcggatgc cttgatgac 4260
 tttgacctg acatgctcgg cagtgcgcc cttgatgatt tcgacctgga catgctgatt 4320

aactctagag cgcccgaga tccaaaaaag aagagaaagg tagatccaaa aaagaagaga 4380
 aaggtagatc caaaaaagaa gagaaggta gatacggccg catag 4425

<210> 5
 <211> 587
 <212> DNA
 <213> Artificial
 <220><223> MS2-activator construct
 <400> 5

ccaccatggg acctaagaaa aagaggaagg tggcggccgc ttctagaatg gtttctaact 60
 ttactcagtt cgttctcgtc gacaatggcg gaactggcga cgtgactgic gcccgaagca 120
 acttcgctaa cgggatcgtc gaatggatca gctctaactc gcgttcacag gcttacaag 180

taacctgtag cgttcgtcag agctctgcgc agaatcgcaa atacaccatc aaagtcgagg 240

 tgcctaaagg cgctggcgt tcgtacttaa atatggaact aaccattcca attttcgcca 300
 cgaattccga ctgcgagctt attgttaagg caatgcaagg tctcctaaaa gatggaaacc 360
 cgattccctc agcaatcgca gcaaaactccg gcatctacga ggccagcggg tccggacggg 420
 ctgacgcatt ggacgatttt gatctggata tgcctgggaag tgacgccctc gatgattttg 480
 accttgacat gcttggttcg gatgcccttg atgactttga cctcgacatg ctccggcagtg 540
 agcccttga tgatttcgac ctggacatgc tgattaactc tagatga 587

 <210> 6
 <211> 681
 <212> DNA
 <213> Artificial

 <220><223> MS2-activator construct
 <400> 6

 gccaccatgg gacctaagaa aaagaggaag gtggcggcgg cttctagaat ggcttctaac 60
 ttactcagt tcgttctcgt cgacaatggc ggaactggcg acgtgactgt cgccccaagc 120
 aatttcgcta acgggatcgc tgaatggatc agctctaact cgcgttcaca ggcttacaaa 180
 gtaacctgta gegtctgca gagctctcgc cagaatcgca aataccatc caaagtcgag 240
 gtgcctaaag gcgcctggcg ttcgtactta aatatggaac taaccattcc aattttcgcc 300
 acgaattccg actgcgagct tattgttaag gcaatgcaag gtctcctaaa agatggaaac 360
 ccgattccct cagcaatcgc agcaaaactcc ggcatctacg aggccagcgg ttccggacgg 420

 gctgacgcat tggacgattt tgatctggat atgctgggaa gtgacgccct cgatgatttt 480
 gaccttgaca tgettggttc ggatgccctt gatgactttg acctcgacat getcggcagt 540
 gacgcccttg atgatttcga cctggacatg ctgattaact ctagagcggc cgcagatcca 600
 aaaaagaaga gaaaggtaga tccaaaaaag aagagaaagg tagatccaaa aaagaagaga 660
 aaggtagata cggccgcata g 681

 <210> 7
 <211> 557
 <212> DNA
 <213> Artificial

 <220><223> MS2-activator construct
 <220><221> misc_feature

<222> (320)..(339)

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

<400> 7

```

tgtacaaaa agcaggcttt aaaggaacca attcagtcga ctggatccgg taccaaggct      60
gggcaggaag agggcctatt tcccatgatt ccttcatatt tgcatatacg atacaaggct      120
gttagagaga taattagaat taatttgact gtaaacacaa agatattagt acaaaatagc      180
tgacgtagaa agtaataatt tcttgggtag ttgacagttt taaaattatg ttttaaagt      240
gactatcata tgcttaccgt aacttgaaag tatttcgatt tcttggcttt atatatcttg      300
tgaaaaggac gaaacaccgn nnnnnnnnnn nnnnnnnng ttttagagct agaaatagca      360
agttaaata aggctagtcc gttatcaact tgaaaaagtg gcaccgagtc ggtgctctgc      420

aggctgactc tagaaaacat gaggatcacc catgtctgca gtattcccgg gttcattaga      480
tcctaaggta cctaattgcc tagaaaacat gaggatcacc catgtctgca ggtcgactct      540
agaaattttt tctagac                                                    557

```

<210> 8

<211> 882

<212> DNA

<213> Artificial

<220><223> Activation reporter construct

<400> 8

```

tagggataac agggtaatag tgtcccctcc accccacagt ggggcgaggt aggcgtgtac      60
ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc      120
gccaccatgg actacaagga tgacgacgat aaaacttccg gtggcggact gggttccacc      180

gtgagcaagg gcgaggaggt catcaaagag ttcattgcgt tcaaggtgcg catggagggc      240
tccatgaacg gccacgagtt cgagatcgag ggcgagggcg agggccgccc ctacgagggc      300
accagaccg ccaagctgaa ggtgaccaag ggcggceccc tgccttcgc ctgggacatc      360
ctgtceccc agttcatgta cggctccaag gcgtacgtga agcaccgcc cgacatcccc      420
gattacaaga agctgtcctt ccccgagggc ttcaagtggg agcgcgtgat gaacttcgag      480
gacggcggtc tggtgaccgt gaccaggac tcctccctgc aggacggcac gctgatctac      540
aaggtgaaga tgcgcggcac caacttccc cccgacggcc ccgtaatgca gaagaagacc      600

atgggctggg aggcctccac cgagcgcctg tacccecgcg acggcgtgct gaaggcgag      660
atccaccagg ccctgaagct gaaggacggc ggccactacc tggtgagtt caagaccatc      720

```

tacatggcca agaagcccgt gcaactgccc ggctactact acgtggacac caagctggac 780
 atcacctccc acaacgagga ctacaccatc gtggaacagt acgagcgctc cgagggccgc 840
 caccacctgt tctgtacgg catggacgag ctgtacaagt aa 882

<210> 9

<211> 882

<212> DNA

<213> Artificial

<220><223> Activation reporter construct

<400> 9

tagggataac agggtaatag tggggccact agggacagga ttggcgaggt aggcgtgtac 60

ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc 120

gccaccatgg actacaagga tgacgacgat aaaacttccg gtggcggact gggttccacc 180

gtgagcaagg gcgaggaggt catcaaagag ttcatgcgct tcaaggtgcg catggagggc 240

tccatgaacg gccacgagtt cgagatcgag ggcgagggcg agggccgccc ctacgagggc 300

accgagaccg ccaagctgaa ggtgaccaag ggcgcccccc tgccttcgc ctgggacatc 360

ctgtccccc agttcatgta cggctccaag gcgtacgtga agcaccgcc cgacatcccc 420

gattacaaga agctgtcctt ccccgagggc ttcaagtggg agcgcgtgat gaacttcgag 480

gacggcggtc tggtgaccgt gaccaggac tcctccctgc aggacggcac gctgatctac 540

aaggtgaaga tgcgcggcac caacttccc cccgacggcc ccgtaatgca gaagaagacc 600

atgggctggg aggcctccac cgagcgcctg taccctccgc acggcgtgct gaaggcgcgag 660

atccaccagg ccctgaagct gaaggacggc ggccactacc tgggtggagt caagaccatc 720

tacatggcca agaagcccgt gcaactgccc ggctactact acgtggacac caagctggac 780

atcacctccc acaacgagga ctacaccatc gtggaacagt acgagcgctc cgagggccgc 840

caccacctgt tctgtacgg catggacgag ctgtacaagt aa 882

<210> 10

<211> 912

<212> DNA

<213> Artificial

<220><223> Specificity reporter library

<220><221> misc_feature

<222> (22)..(44)

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

<220><221> misc_feature
 <222> (154)..(177)
 <223> wherein N is G, A, T or C
 <400> 10
 tagggataac aggtaatag tnnnnnnnnn nnnnnnnnnn nnnncgaggt aggcgtgtac 60
 ggtgggaggc ctatataagc agagctcgtt tagtgaaccg tcagatcgcc tggagaattc 120
 gccacatgg actacaagga tgacgacgat aaannnnnnn nnnnnnnnnn nnnnnnnact 180
 tccggtggcg gactgggttc caccgtgagc aagggcgagg aggtcatcaa agagttcatg 240

 cgcttcaagg tgcgcatgga gggctccatg aacggccacg agttcgagat cgagggcgag 300
 ggcgagggcc gccctacga gggcacccag accccaagc tgaaggtgac caagggcggc 360
 cccctgccct tcgctggga catcctgtcc ccccagtca tgtacggctc caaggcgtac 420
 gtgaagcacc ccgccacat ccccgattac aagaagctgt ccttccccga ggcttcaag 480
 tgggagcgcg tgatgaactt cgaggacggc ggtctggtga ccgtgacca ggactcctcc 540
 ctgcaggacg gcacgtgat ctacaagtg aagatgcgcg gcaccaactt ccccccgac 600
 ggccccgtaa tgcagaagaa gaccatgggc tgggaggcct ccaccgagcg cctgtacccc 660

 cgcgacggcg tgctgaaggg cgagatccac caggccctga agctgaagga cggcggccac 720
 tacttgggtg agttcaagac catctacatg gccaagaagc ccgtgcaact gcccgctac 780
 tactactggtg acaccaagct ggacatcacc tcccacaacg aggactacac catcgtggaa 840
 cagtacgagc gctccgaggc cgccaccac ctgttcctgt acggcatgga cgagctgtac 900
 aagtaagaat tc 912
 <210> 11
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 11
 ctggcggatc actcgcggtt agg 23

 <210> 12
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe

<400> 12
 cctcggcctc caaaagtgct agg 23
 <210> 13
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 13
 acgctgattc ctgcagatca ggg 23
 <210> 14
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 14
 ccaggaatac gtatccacca ggg 23

 <210> 15
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 15
 gccacacca agegatcaaa tgg 23
 <210> 16
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 16
 aaataataca ttctaaggta agg 23
 <210> 17
 <211> 23
 <212> DNA
 <213> Artificial

<220><223> Target probe
 <400> 17
 gctactgggg aggctgagc agg 23

<210> 18
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 18
 tagcaataca gtcacattaa tgg 23

<210> 19
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 19
 ctcatgtgat cccccgtct cgg 23

<210> 20
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 20
 ccgggcagag agtgaacgcg cgg 23

<210> 21
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 21
 ttcttccct ctcccgtgt tgg 23

<210> 22
 <211> 23

<212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 22
 tctctgcaaa gccctggag agg 23
 <210> 23
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 23
 aatgcagttg ccgagtgag tgg 23

 <210> 24
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 24
 cctcagcctc ctaaagtgt ggg 23
 <210> 25
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 25
 gagtccaat cctctttact agg 23
 <210> 26
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 26
 gagtgtctgg atttgggata agg 23

<210> 27
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 27
 cagcacctca tctcccagtg agg 23
 <210> 28
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 28
 tctaaaaccc agggaatcat ggg 23
 <210> 29
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 29
 cacaaggcag ccagggatcc agg 23

 <210> 30
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 30
 gatggcaagc tgagaaacac tgg 23
 <210> 31
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 31

tgaaatgcac gcatacaatt agg 23
 <210> 32
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 32
 ccagtccaga cctggccttc tgg 23

 <210> 33
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 33
 cccagaaaaa cagaccctga agg 23
 <210> 34
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 34
 aagggtgag cacttgttta ggg 23
 <210> 35
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 35
 atgtctgagt ttggttgag agg 23

 <210> 36
 <211> 23
 <212> DNA
 <213> Artificial

<220><223> Target probe
 <400> 36
 ggtcccttga aggggaagta ggg 23
 <210> 37
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 37
 tggcagtcta ctcttgaaga tgg 23
 <210> 38
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 38
 ggcacagtgc cagaggtctg tgg 23

 <210> 39
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 39
 taaaaataaa aaaactaaca ggg 23
 <210> 40
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 40
 tctgtggggg acctgcactg agg 23
 <210> 41
 <211> 23
 <212> DNA

<213> Artificial
 <220><223> Target probe
 <400> 41
 ggccagaggt caaggctagt ggg 23

<210> 42
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 42
 cagaccgaa acccttctta cgg 23

<210> 43
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 43
 gttgaatgaa gacagtctag tgg 23

<210> 44
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 44
 taagaacaga gcaagttacg tgg 23

<210> 45
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 45
 tgtaaggtaa gagaggagag cgg 23

<210> 46

<211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 46
 tgacacacca actcctgcac tgg 23
 <210> 47
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 47
 tttaccact tccttcgaaa agg 23

 <210> 48
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 48
 gtggctggca ggctggctct ggg 23
 <210> 49
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 49
 ctccccggc ctccccgcg cgg 23
 <210> 50
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 50

caaaaccgg cagcgaggct ggg 23

<210> 51
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 51

aggagccgcc gcgcgctgat tgg 23

<210> 52
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 52

cacacacacc cacacagat ggg 23

<210> 53
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 53

gaagaagcta aagagccaga ggg 23

<210> 54
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 54

atgagaattt caataacctc agg 23

<210> 55
 <211> 23
 <212> DNA
 <213> Artificial

<220><223> Target probe
 <400> 55
 tcccgtctg ttgccagc tgg 23
 <210> 56
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 56
 cagacacca ccacatgcg tgg 23

 <210> 57
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 57
 tccaattta ctgggattac agg 23
 <210> 58
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 58
 tgatttaaaa gttgaaacg tgg 23
 <210> 59
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 59
 tctagtccc cacctagtct ggg 23

 <210> 60
 <211> 23

<212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 60
 gattaactga gaattcaca ggg 23
 <210> 61
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target probe
 <400> 61
 cgccaggagg ggtgggtcta agg 23
 <210> 62
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Reporter construct
 <400> 62
 gtcccctcca cccacagtg ggg 23

 <210> 63
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Reporter construct
 <400> 63
 gggccacta gggacaggat tgg 23
 <210> 64
 <211> 71
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 64
 taatactttt atctgtcccc tccacccac agtggggcca ctaggacag gattggtgac 60
 agaaaagccc c 71

<210> 65
 <211> 20
 <212> DNA
 <213> Artificial
 <220><223>
 > Target oligonucleotide sequence
 <400> 65
 ggggccacta gggacaggat 20
 <210> 66
 <211> 80
 <212> RNA
 <213> Artificial
 <220><223> Guide RNA
 <400> 66
 guuuuagagc uagaaauagc aaguuaaaau aaggcuagcu uguuaucaac uugaaaaagu 60
 ggcaccgagu cggugcuuuu 80
 <210> 67
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 67
 gtcccctcca cccacagtg cag 23

 <210> 68
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 68
 gtcccctcca cccacagtg caa 23
 <210> 69
 <211> 23
 <212> DNA
 <213> Artificial

<220><223> Target oligonucleotide sequence
 <400> 69
 gtcccctcca cccacagtg cgg 23
 <210> 70
 <211> 52
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 70
 tgtcccctcc accccacagt ggggccacta gggacaggat tggtagacaga aa 52
 <210> 71
 <211> 52
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 71
 tgtcccccc accccacagt ggggccacta gggacaggat tggtagacaga aa 52
 <210> 72
 <211> 52
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 72
 aaaaccctcc accccacagt ggggccacta gggacaggat tggtagacaga aa 52
 <210> 73
 <211> 52
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 73
 tgtcccctcc ttttttcagt ggggccacta gggacaggat tggtagacaga aa 52
 <210> 74
 <211> 23

<212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 74
 caccggggtg gtgccatcc tgg 23
 <210> 75
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 75
 ggtgcccatc ctggtcgagc tgg 23
 <210> 76
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 76
 cccatcctgg tcgagctgga cgg 23

 <210> 77
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 77
 ggccacaagt tcagcgtgtc cgg 23
 <210> 78
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 78
 cgcaaataag agctcaccta cgg 23
 <210> 79

<211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 79
 ctgaagtcca tctgaccac cgg 23

<210> 80
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 80
 ccggcaagct gccctgccc tgg 23

<210> 81
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 81
 gaccaggatg ggcaccacc cgg 23

<210> 82
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 82
 gccgtccagc tcgaccagga tgg 23

<210> 83
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 83

ggccggacac gctgaacttg tgg 23

<210> 84

<211> 23

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 84

taacagggta atgtcgaggc cgg 23

<210> 85

<211> 23

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 85

aggtgagctc ttatttgcgt agg 23

<210> 86

<211> 23

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 86

cttcagggtc agcttgccgt agg 23

<210> 87

<211> 23

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 87

gggcacgggc agcttgccgg tgg 23

<210> 88

<211> 23

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 88
gagatgatcg ccccttcttc tgg 23

<210> 89
<211> 20
<212> DNA
<213> Artificial
<220><223> Target oligonucleotide sequence
<400> 89
gagatgatcg ccccttcttc 20
<210> 90
<211> 20
<212> DNA
<213> Artificial
<220><223> Target oligonucleotide sequence
<400> 90
gtgatgaccg gccgttcttc 20
<210> 91
<211> 23
<212> DNA
<213> Artificial
<220><223> Target oligonucleotide sequence
<400> 91
gtcccctcca cccacagtg ggg 23
<210> 92
<211> 23
<212> DNA
<213> Artificial
<220><223> Target oligonucleotide sequence
<400> 92
gagatgatcg cccgttcttc tgg 23
<210> 93
<211> 20
<212> RNA

<213> Artificial
 <220><223> RNA target sequence
 <400> 93
 guccuccu cccacagug 20
 <210> 94
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 94
 guccuccu cccacaguc 20

 <210> 95
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 95
 guccuccu cccacagag 20
 <210> 96
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 96
 guccuccu cccacaguc 20
 <210> 97
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 97
 guccuccu cccacagug 20

 <210> 98

<211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 98
 gucccccucca ccccagagug 20
 <210> 99
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 99
 gucccccucca ccccucagug 20
 <210> 100
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 100
 gucccccucca cccgacagug 20

 <210> 101
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 101
 gucccccucca ccgcacagug 20
 <210> 102
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 102
 gucccccucca cgccacagug 20

<210> 103
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 103
 guccccucca gccacagug 20

<210> 104
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 104
 gucccccuccu cccacagug 20

<210> 105
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 105
 gucccccucga cccacagug 20

<210> 106
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 106
 guccccucca cccacagac 20

<210> 107
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence

<400> 107	
gucccccucca ccccacucug	20
<210> 108	
<211> 20	
<212> RNA	
<213> Artificial	
<220><223> RNA target sequence	
<400> 108	
gucccccucca ccccugagug	20
<210> 109	
<211> 20	
<212> RNA	
<213> Artificial	
<220><223> RNA target sequence	
<400> 109	
gucccccucca ccggacagug	20
<210> 110	
<211> 20	
<212> RNA	
<213> Artificial	
<220><223> RNA target sequence	
<400> 110	
gucccccucca ggccacagug	20
<210> 111	
<211> 20	
<212> RNA	
<213> Artificial	
<220><223> RNA target sequence	
<400> 111	
gucccccucgu cccacagug	20
<210> 112	
<211> 23	
<212> DNA	
<213> Artificial	

<220><223> Target oligonucleotide sequence
 <400> 112
 ggggccacta gggacaggat ggg 23

<210> 113
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 113
 gagaugaucg ccccuucuc 20

<210> 114
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 114
 gagaugaucg ccccuucuug 20

<210> 115
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 115
 gagaugaucg ccccuucuac 20

<210> 116
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 116
 gagaugaucg ccccuucauc 20

<210> 117
 <211> 20

<212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 117
 gagaugaucg ccccuuguuc 20
 <210> 118
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 118
 gagaugaucg ccccuacuuc 20

 <210> 119
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 119
 gagaugaucg ccccaucuuc 20
 <210> 120
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 120
 gagaugaucg cccguucuuc 20
 <210> 121
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 121
 gagaugaucg ccgcuucuuc 20

<210> 122
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 122
 gagaugaucg cgccuucuuc 20
 <210> 123
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 123
 gagaugaucg gcccuucuuc 20
 <210> 124
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 124
 gagaugaucc ccccuucuuc 20

 <210> 125
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 125
 gagaugaugg ccccuucuuc 20
 <210> 126
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 126

gagaugaucg ccccuacuag 20

<210> 127

<211> 20

<212> RNA

<213> Artificial

<220><223> RNA target sequence

<400> 127

gagaugaucg ccccuugauc 20

<210> 128

<211> 20

<212> RNA

<213> Artificial

<220><223> RNA target sequence

<400> 128

gagaugaucg ccccaacuuc 20

<210> 129

<211> 20

<212> RNA

<213> Artificial

<220><223> RNA target sequence

<400> 129

gagaugaucg ccgguucuuc 20

<210> 130

<211> 20

<212> RNA

<213> Artificial

<220><223> RNA target sequence

<400> 130

gagaugaucg ggccuucuuc 20

<210> 131

<211> 20

<212> RNA

<213> Artificial

<220><223> RNA target sequence
 <400> 131
 gagaugaugc ccccuucuc 20
 <210> 132
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 132
 gagatgatcg cccttcttc tgg 23
 <210> 133
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 133
 ggggccacua gggacaggau 20
 <210> 134
 <211> 19
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 134
 gggccacuag ggacaggau 19
 <210> 135
 <211> 18
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 135
 ggccacuagg gacaggau 18
 <210> 136
 <211> 17
 <212> RNA

<213> Artificial
 <220><223> RNA target sequence
 <400> 136
 gccacuaggg acaggau 17

<210> 137
 <211> 20
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 137
 gagaugaucg ccccuucuc 20

<210> 138
 <211> 18
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 138
 gaugaucgcc ccuucuc 18

<210> 139
 <211> 15
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 139
 gaucgccccu ucuuc 15

<210> 140
 <211> 11
 <212> RNA
 <213> Artificial
 <220><223> RNA target sequence
 <400> 140
 gccccuucu c 11

<210> 141

<211> 21
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 141
 gtccctcca cccacagtg c 21
 <210> 142
 <211> 14
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <220><221> misc_feature
 <222> (5)..(10)

 <223> wherein N id G, A, T or C
 <400> 142
 tgtcnnnnn accc 14
 <210> 143
 <211> 14
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 143
 tgtcaaaaa accc 14
 <210> 144
 <211> 14
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 144
 tgtcggggg accc 14

 <210> 145
 <211> 14
 <212> DNA

<213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 145
 tgtcaaaaa accc 14
 <210> 146
 <211> 14
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 146
 tgtcgggggg accc 14
 <210> 147
 <211> 14
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 147
 tgtcccccc accc 14

 <210> 148
 <211> 14
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 148
 tgtcttttt accc 14
 <210> 149
 <211> 14
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 149
 tgtcccccc accc 14
 <210> 150
 <211> 14

<212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 150
 tgtctttttt accc 14

<210> 151
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 151
 ggatcctgtg tccccgagct ggg 23

<210> 152
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 152
 gttaatgtgg ctctggttct ggg 23

<210> 153
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 153
 ggggccacta gggacaggat tgg 23

<210> 154
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 154
 cttcctagtc tcctgatatt ggg 23

<210> 155
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 155
 tgggccagc tcggggacac agg 23
 <210> 156
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 156
 agaaccagag ccacattaac cgg 23

 <210> 157
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 157
 gtcaccaatc ctgtccctag tgg 23
 <210> 158
 <211> 23
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 158
 agaccaata tcaggagact agg 23
 <210> 159
 <211> 75
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 159

gggatcctgt gtccccgagc tgggaccacc ttatattccc agggccggtt aatgtggctc 60

tggttctggg tactt 75

<210> 160

<211> 69

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 160

gggatcctgt gtccccgagc tgggaccacc ttatattccc agggccggtt aatgtggctc 60

tgggtactt 69

<210> 161

<211> 113

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 161

gggatcctgt gtccccgagc tgggaccacc ttatattccc agggcagggc cggttggacc 60

accttatatt cccagggcag ggccggtaa tgtggctctg gttctgggta ctt 113

<210> 162

<211> 34

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 162

gggatcctgt gtccccgctc ggttctgggt actt 34

<210> 163

<211> 47

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 163

gggatcctgt gtccccgagc tgggaccacc ttatattctg ggtactt 47

<210> 164

<211> 17
 <212> DNA
 <213> Artificial

<220><223> Target oligonucleotide sequence
 <400> 164
 gggatcctgt ggtactt 17
 <210> 165
 <211> 93
 <212> DNA
 <213> Artificial

<220><223> Target oligonucleotide sequence
 <400> 165
 agggccggtt aatgtggctc tggttctggg tacttttatac tgtcccctcc accccacagt 60
 ggggccacta gggacaggat tggtagacaga aaa 93
 <210> 166
 <211> 83
 <212> DNA
 <213> Artificial

<220><223> Target oligonucleotide sequence
 <400>
 > 166
 agggccggtt aatgaatgtg gctctggttc tgggtacttt tatctgtccc ctccaccca 60
 cagtggggcc actagacaga aaa 83
 <210> 167
 <211> 76
 <212> DNA
 <213> Artificial

<220><223> Target oligonucleotide sequence
 <400> 167
 agggccggtt aatgtggctc tggttctggg tacttttatac tgtccccag tggggccact 60
 gattggtgac agaaaa 76
 <210> 168
 <211> 29
 <212> DNA

<213> Artificial
 <220><223> Target oligonucleotide sequence

<400> 168
 agggccggtt caggattggt gacagaaaa 29

<210> 169
 <211> 34
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence

<400> 169
 agggccggtt aatgtggcga ttggtgacag aaaa 34

<210> 170
 <211> 63
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence

<400> 170
 agggccggtt aatgtggctc tggttctggg tacttttatac tgtccccgat ttggtgacaga 60
 aaa 63

<210> 171
 <211> 84
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence

<400> 171
 agggccggtt aatgtggctc tggttctggg tacttttatac tgtcccctcc accccacagt 60
 ggggacagga ttggtgacag aaaa 84

<210> 172
 <211> 27
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence

<400> 172

agggccggtt aatgtggtga cagaaaa 27

<210> 173

<211> 105

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 173

agggccggtt aatgtggctc tggttctggg tacttttadc tgtcccctcc accccagggg 60

acagtctgtc ccctccacc caggacagg attggtgaca gaaaa 105

<210> 174

<211> 80

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 174

agggccggtt aatgtggctc tggttctggg tacttttadc tgtcccctcc accactaggg 60

acaggattgg tgacagaaaa 80

<210> 175

<211> 53

<212>

> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 175

cccacagtgg ggccactagg gacaggattg gtgacagaaa agcccatac ccc 53

<210> 176

<211> 22

<212> DNA

<213> Artificial

<220><223> Target oligonucleotide sequence

<400> 176

cccacagtgg ggccactacc cc 22

<210> 177

<211> 96

<212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 177
 cccacagtgg ggccactagt agaaaagccc catccttagg cctccccat ccttaggcct 60

 ctccttctct agtctctctga tattgggtct aacccc 96
 <210> 178
 <211> 94
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 178
 cccacagtgg ggccactagg gacaggattg gtgacagaaa agccccatcc ttaggcctcc 60
 tccttctctag tctctctgata ttgggtctaa cccc 94
 <210> 179
 <211> 62
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 179
 cccacagtgg ggccaccctt aggcctcttc cttctctagtc tctgatatt ggttctaacc 60

 cc 62
 <210> 180
 <211> 38
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 180
 cccacagtgg ggccactagt gatattgggt ctaacccc 38
 <210> 181
 <211> 94
 <212> DNA
 <213> Artificial

<220><223> target oligonucleotide sequence
 <400> 181
 cccacagtgg ggccactagg gacaggattg gtgacaaaaa agcccatcc ttacgcctcc 60
 tccttctag tctctgata ttgggtctaa cccc 94

<210> 182
 <211> 65
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 182
 cccacagtgg ggccactagg gacaggcctc ctcttctta gtctctgat attgggtcta 60
 acccc 65

<210> 183
 <211> 102
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 183
 cccacagtgg ggccactagg gacaggggga caggattggt gacagaaaag ccccatcctt 60
 aggcctctc ctctctagtc tctgatatt gggcttaacc cc 102

<210> 184
 <211> 76
 <212> DNA
 <213> Artificial
 <220><223> Target oligonucleotide sequence
 <400> 184
 cccacaggat tggtagacaga aaagcccat ccttaggcct cctccttct agtctctga 60
 tattgggtct aacccc 76