



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2015-0101446
(43) 공개일자 2015년09월03일

- (51) 국제특허분류(Int. Cl.)
C12Q 1/68 (2006.01) C12N 15/113 (2010.01)
C12N 15/82 (2006.01) C12N 15/85 (2006.01)
C12N 9/16 (2006.01)
- (52) CPC특허분류(Coo. Cl.)
C12Q 1/683 (2013.01)
C12N 15/113 (2013.01)
- (21) 출원번호 10-2015-7013868
- (22) 출원일자(국제) 2013년10월23일
심사청구일자 없음
- (85) 번역문제출일자 2015년05월26일
- (86) 국제출원번호 PCT/KR2013/009488
- (87) 국제공개번호 WO 2014/065596
국제공개일자 2014년05월01일
- (30) 우선권주장
61/717,324 2012년10월23일 미국(US)
(뒷면에 계속)

- (71) 출원인
주식회사 틀젠
서울특별시 금천구 가산디지털1로 219, 제1204호,
제1205호, 제1206호 (가산동, 벽산디지털밸리6차)
- (72) 발명자
김진수
서울특별시 관악구 관악로1 (신림동) 자연과학대
학 화학부
조승우
서울특별시 관악구 관악로1 (신림동) 자연과학대
학 화학부
(뒷면에 계속)
- (74) 대리인
손민

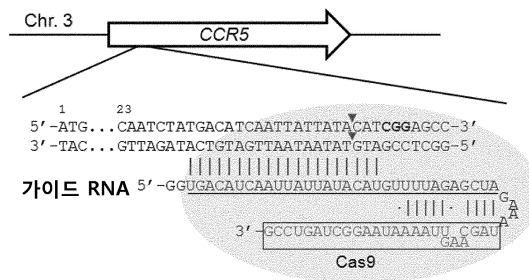
전체 청구항 수 : 총 57 항

(54) 발명의 명칭 표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 표적 DNA를 절단하기 위한 조성물 및 이의 용도

(57) 요약

본 발명은 진핵 세포 또는 유기체에서의 표적화된 유전체 교정에 관한 것이다. 보다 구체적으로, 본 발명은 표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 조성물, 및 그의 용도에 관한 것이다.

대표도



- (52) CPC특허분류(Coo. C1.)
C12N 15/8216 (2013.01)
C12N 15/8509 (2013.01)
C12N 9/16 (2013.01)
- (72) 발명자
김소정
서울특별시 관악구 관악로1 (신림동) 자연과학대학
화학부
김중민
부산광역시 사상구 백양대로979번길 63(모라동)
김석중
서울특별시 서초구 서초대로74길 29 1301호(
서초동, 서초파라곤)
- (30) 우선권주장
61/803,599 2013년03월20일 미국(US)
61/837,481 2013년06월20일 미국(US)
-

특허청구의 범위

청구항 1

표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 조성물.

청구항 2

제1항에 있어서, 상기 표적 DNA는 내재적 표적 DNA인, 조성물.

청구항 3

제1항에 있어서, 상기 가이드 RNA는 crRNA 및 tracrRNA를 포함하는 이중RNA (dualRNA)인, 조성물.

청구항 4

제1항에 있어서, 상기 가이드 RNA는 단일-사슬 가이드 RNA (sgRNA)인, 조성물.

청구항 5

제4항에 있어서, 상기 단일-사슬 가이드 RNA는 crRNA 및 tracrRNA의 부분을 포함하는, 조성물.

청구항 6

제1항에 있어서, 상기 가이드 RNA는 단일-사슬 가이드 RNA 또는 이중RNA의 crRNA의 5' 말단에 하나 이상의 추가적인 뉴클레오타이드를 더 포함하는, 조성물.

청구항 7

제1항에 있어서, 상기 가이드 RNA는 단일-사슬 가이드 RNA 또는 이중RNA의 crRNA의 5' 말단에 두 개의 추가적인 구아닌 뉴클레오타이드를 더 포함하는, 조성물.

청구항 8

제1항에 있어서, 진핵 세포 또는 유기체에서 표적화된 돌연변이 (targeted mutagenesis)을 유도하는, 조성물.

청구항 9

제1항에 있어서, 인 비트로(*in vitro*)에서 진핵 세포 또는 유기체의 유전체에 대한 유전형질 분석 (genotyping)에 사용하기 위한, 조성물.

청구항 10

제1항에 있어서, 상기 가이드 RNA 및 Cas 단백질은 쌍으로 작용하고, 상기 쌍은 서로 다른 가닥에 두 개의 틸을

유도하는 두 개의 가이드 RNA를 포함하는, 조성물.

청구항 11

제1항에 있어서, 상기 가이드 RNA는 분리된 RNA의 형태거나, 또는 벡터에 암호화되어 있고, 상기 벡터는 바이러스 벡터, 플라스미드 벡터, 또는 아그로박테리움 벡터인, 조성물.

청구항 12

제1항에 있어서, 표적 DNA에 특이적인 가이드 RNA 및 Cas9 단백질을 포함하는, 조성물.

청구항 13

제1항에 있어서, 엑스 비보 (*ex vivo*) 또는 인 비보 (*in vivo*)에서 진핵 세포 또는 유기체의 표적 DNA를 절단하기 위한, 조성물.

청구항 14

제1항에 있어서, 상기 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질은 스트렙토코커스 속 (genus *Streptococcus*)으로부터 유래되는, 조성물.

청구항 15

제14항에 있어서, 상기 스트렙토코커스 속은 스트렙토코커스 피요젠스 (*streptococcus pyogenes*)인, 조성물.

청구항 16

제15항에 있어서, 상기 Cas 단백질은 NGG 트리뉴클레오타이드 (trinucleotide)를 인식하는, 조성물.

청구항 17

제1항에 있어서, 상기 Cas 단백질은 Cas9 단백질 또는 이의 변이체인, 조성물.

청구항 18

제1항에 있어서, 상기 Cas 단백질은 단백질 전달 도메인 (protein transduction domain)에 연결되어 있는, 조성물.

청구항 19

제17항에 있어서, 상기 Cas9 단백질의 변이체는 측매 아스파라긴산 (aspartate) 잔기가 다른 아미노산으로 교체된, Cas9의 돌연변이 형태인, 조성물.

청구항 20

제19항에 있어서, 상기 아미노산은 알라닌(alanine)인, 조성물.

청구항 21

제1항에 있어서, 상기 Cas 단백질을 암호화하는 핵산은 서열번호: 1의 뉴클레오타이드 서열 또는 서열번호: 1과 적어도 50%의 상동성을 갖는 뉴클레오타이드 서열을 포함하는, 조성물.

청구항 22

진핵 세포 또는 유기체에서의 표적화된 돌연변이를 위한 제1항 내지 제21항 중 어느 한 항의 조성물의 용도.

청구항 23

제1항 내지 제21항 중 어느 한 항에 따른 조성물을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한, 키트.

청구항 24

Cas 단백질을 암호화하는 핵산 또는 Cas 단백질, 및 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA를 진핵 세포 또는 유기체에 공동-형질주입 (co-transfecting) 또는 단계적 형질주입 (serial-transfecting)하는 단계를 포함하는, Cas 단백질 및 가이드 RNA를 포함하는 진핵 세포 또는 유기체를 제조하는 방법.

청구항 25

제24항에 있어서, 상기 가이드 RNA는 crRNA 및 tracrRNA를 포함하는 이중RNA (dualRNA)인, 방법.

청구항 26

제24항에 있어서, 상기 가이드 RNA는 단일-사슬 가이드 RNA인, 방법.

청구항 27

제26항에 있어서, 상기 단일-사슬 가이드 RNA는 crRNA 및 tracrRNA의 부분을 포함하는, 방법.

청구항 28

제24항에 있어서, 상기 가이드 RNA는 단일-사슬 가이드 RNA 또는 이중RNA의 crRNA의 5' 말단에 하나 이상의 추가적인 뉴클레오타이드를 더 포함하는, 방법.

청구항 29

제24항에 있어서, 상기 가이드 RNA는 단일-사슬 가이드 RNA 또는 이중RNA의 crRNA의 5' 말단에 두 개의 추가적인 구아닌 뉴클레오타이드를 더 포함하는, 방법.

청구항 30

제24항에 있어서, 상기 진핵 세포 및 유기체는 Cas9 단백질 및 가이드 RNA가 공동-형질주입 (co-transfecting) 또는 단계적-형질주입 (serial-transfecting)된, 방법.

청구항 31

제24항에 있어서, 상기 단계적-형질주입은 Cas 단백질을 암호화하는 핵산을 먼저 형질 주입한 다음, 네이키드 가이드 RNA (naked guide RNA)를 두 번째 형질주입하여 수행되는, 방법.

청구항 32

제24항에 있어서, 상기 Cas 단백질은 Cas9 단백질 또는 이의 변이체인, 방법.

청구항 33

제24항에 있어서, 상기 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질은 스트렙토코커스 속 (genus Streptococcus)으로부터 유래된, 방법.

청구항 34

제33항에 있어서, 상기 스트렙토코커스 속은 스트렙토코커스 피요젠스 (streptococcus pyogenes)인, 방법.

청구항 35

제32항에 있어서, 상기 Cas9 단백질의 변이체는 측매 아스파라긴산(aspartate) 잔기가 다른 아미노산으로 교체된 Cas9의 돌연변이 형태인, 방법.

청구항 36

제35항에 있어서, 상기 아미노산은 알라닌(alanine)인, 방법.

청구항 37

제24항에 있어서, 상기 가이드 RNA 및 Cas 단백질은 쌍으로 작용하고, 상기 쌍은 서로 다른 가닥에 두 개의 틈을 유도하는 두 개의 가이드 RNAs를 포함하는, 방법.

청구항 38

제37항에 있어서, 상기 두 틈은 적어도 100 bps 떨어진 것인 방법.

청구항 39

제24항에 있어서, 상기 형질주입은 미세주입법, 전기천공법, DEAE-텍스트란 처리, 리포펙션, 나노파티클-매개 형질주입, 단백질 전달 도메인 매개 도입, 바이러스-매개 유전자 전달, 및 원생동물에서 PEG-매개 형질주입으로 구성된 군으로부터 선택된 방법으로 수행되는, 방법.

청구항 40

제24항 내지 제39항 중 어느 한 항에 따른 방법으로 제조된, Cas 단백질 및 가이드 RNA를 포함하는 진핵 세포 또는 유기체.

청구항 41

제1항 내지 제21항 중 어느 한 항에 따른 조성물을 표적 DNA를 포함하는 진핵 세포 또는 유기체에 형질주입하는 단계를 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하는 방법.

청구항 42

제41항에 있어서, 상기 진핵 유기체는 포유동물 또는 식물인, 방법.

청구항 43

제41항에 있어서, 상기 가이드 RNA 및 Cas 단백질은 쌍으로 작용하고, 상기 쌍은 서로 다른 가닥에 두 개의 틸을 유도하는 두 개의 가이드 RNAs를 포함하는, 방법.

청구항 44

제43항에 있어서, 상기 두 틸은 적어도 100 bps 떨어진 것인 방법.

청구항 45

제41항에 있어서, 상기 형질주입은 공동-형질주입 또는 단계적-형질주입인 방법.

청구항 46

제45항에 있어서, 상기 단계적-형질주입은 Cas 단백질을 암호화하는 핵산을 먼저 형질 주입한 다음, 네이키드 가이드 RNA (naked guide RNA)를 두 번째 형질주입하여 수행되는, 방법.

청구항 47

제41항에 있어서, 절단 패턴을 분석하는 단계를 더 포함하며, 상기 패턴은 유전체의 돌연변이 또는 변이의 감지를 나타내는, 방법.

청구항 48

제1항 내지 제21항 중 어느 한 항에 따른 조성물에 의해 교정된 유전체를 포함하는, 배아.

청구항 49

제48항의 배아를 난관에 이식하여 얻은 유전체-변형 동물.

청구항 50

제24항 내지 제39항 중 어느 한 항의 방법에 의해 제조된 유전체-변형 원생동물로부터 재생된 식물.

청구항 51

제1항 내지 제21항 중 어느 한 항에 따른 조성물을 동물의 배아에 도입하는 단계; 및 상기 배아를 가임신 위탁 모의 난관에 이식하여 유전체-변형 동물을 생산하는 단계를 포함하는, 유전체-변형 동물의 제조 방법.

청구항 52

표적 DNA 서열에 특이적인 가이드 RNA, Cas 단백질을 포함하는, 분리된 생물학적 시료에서 돌연변이 또는 변이 (variation)를 유전형질 분석(genotyping) 하기 위한 조성물.

청구항 53

제52항에 있어서, 상기 돌연변이 또는 변이는 세포에서 유전자 가위 (engineered nuclease)에 의해 유도된, 조성물.

청구항 54

제52항에 있어서, 상기 돌연변이 또는 변이는 자연 발생 돌연변이 또는 변이인, 조성물.

청구항 55

표적 DNA 서열에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는, 분리된 생물학적 시료에서 병원성 미생물의 핵산 서열을 유전형질 분석하기 위한 조성물.

청구항 56

제52항 내지 제55항 중 어느 한 항의 조성물을 포함하는, 분리된 생물학적 시료에서 돌연변이 또는 변이를 유전형질 분석 하기 위한 키트.

청구항 57

제52항 내지 제55항 중 어느 한 항의 조성물을 사용하여 분리된 생물학적 시료에서 돌연변이 또는 변이를 유전형질 분석하는 방법.

명세서

기술분야

본 발명은 진핵 세포 또는 유기체에서의 표적화된 유전체 교정에 관한 것이다. 보다 구체적으로, 본 발명은 표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 조성물, 및 그의 용도에 관한 것이다.

[0001]

배경 기술

- [0002] CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats)는 유전자 서열이 밝혀진 박테리아의 대략 40% 및 유전자 서열이 밝혀진 고세균의 90%의 유전체에서 발견되는 여러 짧은 직접 반복을 포함하는 좌위이다. 플라스미드 및 파지 등의 외인성 유전적 요소에 저항성을 부여한다는 점에서, CRISPR는 원핵 면역 시스템으로서 기능한다. CRISPR 시스템은 획득 면역의 한 형태를 제공한다. 스페이서(spacers)라고 불리는 외인성 DNA의 짧은 부분은 CRISPR 반복 사이의 게놈에 편입되고, 과거 노출을 기억하는 역할을 한다. 그때 CRISPR 스페이서는 진핵 유기체에서 RNAi와 유사한 방식으로 외인성 유전적 요소를 인지하고 묵살(silence)하는데 사용된다.
- [0003] Type II CRISPR/Cas 시스템에서 필수적인 단백질 요소인 Cas9은, CRISPR RNA (crRNA) 및 trans-activating crRNA(tracrRNA)로 명명된 두 개의 RNA와 복합체를 형성했을 때, 활성 엔도뉴클레아제(endonuclease)를 형성하고, 그렇게 함으로써 파지 또는 플라스미드의 침입에서 외부 유전적 요소를 묵살하여 숙주 세포를 보호한다. crRNA는 전달에 외부 침입자로부터 점유되었던 숙주 유전체의 CRISPR 요소로부터 전사된다. 최근, Jinek et al. (1)은 crRNA 및 tracrRNA에서 필수적인 부분의 융합에 의해 생산된 단일 사슬 키메라 RNA(chimeric RNA)가 Cas9/RNA 복합체에서 두 개의 RNA를 대체할 수 있어서 기능 엔도뉴클레아제를 형성한다는 것을 입증하였다.
- [0004] 뉴클레오타이드 결합 CRISPR-Cas 단백질의 위치 특이성(site specificity)은 디자인 및 합성하기에 더 까다로울 수 있는 DNA-결합 단백질 대신 RNA 분자에 의해 통제되기 때문에, CRISPR/Cas 시스템은 징크 핑거 (zinc finger) 및 전사 활성화자-유사 반응기 DNA-결합 단백질 (transcription activator-like effector DNA binding protein)에 이점을 제공한다.
- [0005] 하지만, 지금까지 CRISPR/Cas 시스템을 기반으로 RNA-가이드 엔도뉴클레아제 (RGEN)를 사용한 유전체 교정 방법은 고안되지 않았다.
- [0006] 한편, 제한효소 단편 길이 다형성(Restriction fragment length polymorphism, RFLP)은 가장 오래되고, 가장 편리하고, 최소의 비용이 드는 유전형질 분석(genotyping) 방법 중 하나이며, 분자 생물학 및 유전학 분야에 현재까지 널리 사용되지만, 종종 제한효소에 의해 인식되는 적절한 위치가 결여되는 한계가 있다.
- [0007] 유전자 가위(engineered nuclease)에 의한 돌연변이는 불일치-민감성 T7 엔도뉴클레아제 I (T7E1) 또는 Surveyor 뉴클레아제 어세이, RFLP, 형광 PCR 산물의 모세관 전기영동, 디데옥시 시퀀싱(Dideoxy sequencing) 및 딥 시퀀싱(deep sequencing)을 포함한 다양한 방법에 의해 탐지된다. T7E1 및 Surveyor 어세이는 널리 사용되지만 번거롭다.
- [0008] 더욱이, 돌연변이 서열이 서로 호모듀플렉스(homoduplexes)를 형성할 수 있고, 야생형 세포의 동형접합 이중 대립 유전자 돌연변이 클론(clone)을 구별하지 못하기 때문에, 상기 효소는 돌연변이 빈도를 과소 평가하는 경향이 있다. RFLP는 상기 제한효소를 가지고 있지 않으므로 선택의 한 방법이다. 사실, RFLP는 세포 및 동물에서 유전자 가위-매개 돌연변이를 탐지하는 가장 첫 번째 방법 중 하나였다. 하지만, 불행히도 RFLP는 적절한 제한 부위의 가용성에 의해 제한된다. 관심 있는 표적 위치에 제한 부위가 없다면 사용이 가능하다.

선행기술문헌

비특허문헌

- [0009] (비특허문헌 0001) 1. M. Jinek et al., Science 337, 816 (Aug 17, 2012).
- (비특허문헌 0002) 2. H. Kim, E. Um, S. R. Cho, C. Jung, J. S. Kim, Nat Methods 8, 941 (Nov, 2011).
- (비특허문헌 0003) 3. H. J. Kim, H. J. Lee, H. Kim, S. W. Cho, J. S. Kim, Genome Res 19, 1279 (Jul, 2009).
- (비특허문헌 0004) 4. E. E. Perez et al., Nat Biotechnol 26, 808 (Jul, 2008).
- (비특허문헌 0005) 5. J. C. Miller et al., Nat Biotechnol 29, 143 (Feb, 2011).

(비특허문헌 0006) 6. C. Mussolino et al., Nucleic Acids Res 39, 9283 (Nov, 2011).
 (비특허문헌 0007) 7. J. Cohen, Science 332, 784 (May 13, 2011).
 (비특허문헌 0008) 8. V. Pattanayak, C. L. Ramirez, J. K. Joung, D. R. Liu, Nat Methods 8, 765 (Sep, 2011).
 (비특허문헌 0009) 9. R. Gabriel et al., Nat Biotechnol 29, 816 (Sep, 2011).
 (비특허문헌 0010) 10. E. Kim et al., Genome Res, (Apr 20, 2012).
 (비특허문헌 0011) 11. H. J. Lee, J. Kweon, E. Kim, S. Kim, J. S. Kim, Genome Res 22, 539 (Mar, 2012).
 (비특허문헌 0012) 12. H. J. Lee, E. Kim, J. S. Kim, Genome Res 20, 81 (Jan, 2010).
 (비특허문헌 0013) 13. Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nat Biotech advance online publication (2013)

발명의 내용

해결하려는 과제

- [0010] 지금까지 CRISPR/Cas 시스템을 기반으로 RNA-가이드 엔도뉴클레아제(RGEN)을 사용한 유전체 교정 및 유전형질 분석 방법은 개발되지 않았다.
- [0011] 이러한 상황에서, 본 발명자는 CRISPR/Cas 시스템을 기반으로 한 유전체 교정 방법을 개발하고자 예의 노력하였고, 마침내 진핵 세포 및 유기체에서 표적화된 방법으로 DNA를 절단하는 프로그램화된 RNA-가이드 엔도뉴클레아제를 확립하였다.
- [0012] 또한, 본 발명자들은 RFLP 분석에서 RNA-가이드 엔도뉴클레아제(RGENs)을 사용하는 새로운 방법을 개발하고자 예의 노력하였다. 그들은 RGEN을 사용하여 암에서 발견되는 반복 돌연변이 (recurrent mutation) 및 RGEN을 포함하는 유전자 가위 자체에 의해 세포 및 유기체에서 유도되는 반복 돌연변이에 대한 유전형질을 분석하였고, 이로써 본 발명을 완성하였다.

과제의 해결 수단

- [0013] 본 발명의 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 조성물을 제공하는 것이다.
- [0014] 본 발명의 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적화된 돌연변이를 유도하기 위한 조성물을 제공하는 것이다.
- [0015] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 키트를 제공하는 것이다.
- [0016] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적화된 돌연변이를 유도하기 위한 키트를 제공하는 것이다.

- [0017] 본 발명의 또 다른 목적은 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질, 및 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA를 진핵 세포 및 유기체에 공동-형질주입 (co-transfecting) 또는 단계적 형질주입 (serial-transfecting)하는 단계를 포함하는, Cas 단백질 및 가이드 RNA를 포함하는 진핵 세포 또는 유기체를 제조하는 방법을 제공하는 것이다.
- [0018] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 진핵 세포 또는 유기체를 제공하는 것이다.
- [0019] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물을, 표적 DNA를 포함하는 진핵 세포 또는 유기체에 형질주입하는 단계를 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하는 방법을 제공하는 것이다.
- [0020] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물을 진핵 세포 또는 유기체에 처리하는 단계를 포함하는, 진핵 세포 또는 유기체에서 표적화된 돌연변이를 유도하는 방법을 제공하는 것이다.
- [0021] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물에 의해 교정된 유전체를 포함하는 배아, 유전체-변형 동물, 또는 유전체-변형 식물을 제공하는 것이다.
- [0022] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물을 동물의 배아에 도입하는 단계; 및 상기 배아를 가입된 위탁모(pseudopregnant foster mother)의 난관에 이식하여 유전체-변형 동물을 생산하는 단계를 포함하는, 유전체-변형 동물을 제조하는 방법을 제공하는 것이다.
- [0023] 본 발명의 또 다른 목적은 표적 DNA 서열에 특이적인 가이드 RNA, Cas 단백질을 포함하는, 분리된 생물학적 시료에서 돌연변이 또는 변이(variation)를 유전형질 분석(genotyping) 하기 위한 조성물을 제공하는 것이다.
- [0024] 본 발명의 또 다른 목적은 RNA-가이드 엔도뉴클레아제 (RGEN)을 사용하여 유전자 가위에 의해 세포에서 유도된 돌연변이 또는 자연 발생 돌연변이 또는 변이를 유전형질 분석하는 방법으로, 여기서 상기 RGEN은 표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는 방법을 제공하는 것이다.
- [0025] 본 발명의 또 다른 목적은 RNA-가이드 엔도뉴클레아제 (RGEN)을 포함하는, 유전자 가위에 의해 세포에서 유도된 돌연변이 또는 자연 발생 돌연변이 또는 변이를 유전형질 분석하기 위한 키트로서, 여기서 상기 RGEN은 표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는 키트를 제공하는 것이다.
- [0026] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 조성물을 제공하는 것이다.
- [0027] 본 발명의 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백

질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적화된 돌연변이를 유도하기 위한 조성물을 제공하는 것이다.

- [0028] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 키트를 제공하는 것이다.
- [0029] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적화된 돌연변이를 유도하기 위한 키트를 제공하는 것이다.
- [0030] 본 발명의 또 다른 목적은 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질, 및 가이드 RNA 또는 가이드 RNA를 암호화하는 DNA를 진핵 세포 및 유기체에 공동-형질주입(co-transfecting) 또는 단계적-형질주입(serial-transfecting)하는 단계를 포함하는, Cas 단백질 및 가이드 RNA를 포함하는 진핵 세포 또는 유기체를 제조하는 방법을 제공하는 것이다.
- [0031] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 진핵 세포 또는 유기체를 제공하는 것이다.
- [0032] 본 발명의 또 다른 목적은 표적 DNA를 포함하는 진핵 세포 또는 유기체에 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물을 형질주입하는 단계를 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하는 방법을 제공하는 것이다.
- [0033] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물을 진핵 세포 또는 유기체에 처리하는 단계를 포함하는, 진핵 세포 또는 유기체에서 표적화된 돌연변이를 유도하는 방법을 제공하는 것이다.
- [0034] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물에 의해 교정된 유전체를 포함하는 배아, 유전체-변형 동물, 또는 유전체-변형 식물을 제공하는 것이다.
- [0035] 본 발명의 또 다른 목적은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물을 동물의 배아에 도입하는 단계; 및 상기 배아를 가임신 위탁모(pseudopregnant foster mother)의 난관에 이식하여 유전체-변형 동물을 생산하는 단계를 포함하는, 유전체-변형 동물을 제조하는 방법을 제공하는 것이다.
- [0036] 본 발명의 또 다른 목적은 표적 DNA 서열에 특이적인 가이드 RNA, Cas 단백질을 포함하는, 분리된 생물학적 시료에서 돌연변이 또는 변이(variation)를 유전형질 분석(genotyping) 하기 위한 조성물을 제공하는 것이다.
- [0037] 본 발명의 또 다른 목적은 표적 DNA 서열에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는, 분리된 생물학적 시료에서 병원성 미생물의 핵산 서열을 유전형질 분석하기 위한 조성물을 제공하는 것이다.

[0038] 본 발명의 또 다른 목적은 RNA-가이드 엔도뉴클레아제 (RGEN)을 특별히 포함하는, 조성물을 포함하는, 분리된 생물학적 시료에서 돌연변이 또는 변이 (variation)를 유전형질 분석 (genotyping) 하기 위한 키트로서, 여기서 상기 RGEN은 표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는 키트를 제공하는 것이다.

[0039] 본 발명의 또 다른 목적은 특별히 RNA-가이드 엔도뉴클레아제 (RGEN)을 포함하는 상기 조성물을 사용하여, 분리된 생물학적 시료에서 돌연변이 또는 변이 (variation)를 유전형질 분석 (genotyping)하는 방법으로서, 여기서 상기 RGEN은 표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는 방법을 제공하는 것이다.

발명의 효과

[0040] 표적 DNA에 특이적인 가이드 RNA 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단 또는 표적화된 돌연변이를 유도하기 위한 본 발명의 조성물, 상기 조성물을 포함하는 키트, 및 표적화된 돌연변이를 유도하는 방법은 새롭고 편리한 유전체 교정 수단을 제공한다. 또한, 커스텀 RGENs (custom RGENs)은 어떤 DNA 서열도 표적화되도록 설계될 수 있으므로, 거의 모든 단일 염기 다형성 (single nucleotide polymorphism) 또는 작은 삽입/결실 (indel)은 RGEN-매개 RFLP를 통해 분석될 수 있다. 그러므로, 본 발명의 조성물 및 방법은 자연 발생 변이 및 돌연변이를 탐지하고 절단하는데 사용될 수 있다.

도면의 간단한 설명

[0041] 도 1은 인 비트로 (*in vitro*)에서 플라스미드 DNA의 Cas9-촉매 절단을 보여준다. (a) 표적 DNA 및 키메라 RNA 서열의 도식 표현. 적색 삼각형은 절단 부위를 나타낸다. Cas9에 의해 인식되는 PAM 서열은 굵은 글씨로 표시된다. crRNA 및 tracrRNA로부터 유래한 가이드 RNA의 서열은 각각 네모칸 (box) 및 밑줄로 나타낸다. (b) Cas9에 의한 플라스미드 DNA의 인 비보 (*in vivo*) 절단. 온전한 원형 플라스미드 또는 ApaLI-절단된 플라스미드는 Cas9 및 가이드 RNA와 함께 배양하였다.

도 2는 에피솜 표적 부위(episomal target site)에서의 Cas9-유도 돌연변이를 보여준다. (a) RFP-GFP 리포터를 사용한 세포-기반 어세이의 개요도. GFP 서열은 RFP 서열과 out-of-frame으로 융합되었기 때문에 GFP는 상기 리포터로부터 발현하지 않는다. RFP-GFP 융합 단백질은 두 서열 사이의 표적 부위가 위치-특이적 뉴클레아제에 의해 절단되었을 때만 발현한다. (b) Cas9을 형질주입한 세포의 유세포 분석(flow cytometry). RFP-GFP 융합 단백질을 발현하는 세포의 퍼센트가 표시된다.

도 3은 내재적 염색체 위치 (endogenous chromosomal site)에서의 RGEN에 의한 돌연변이를 보여준다. (a) CCR5 좌위. (b) C4BPB 좌위. (위) T7E1 어세이를 사용하여 RGEN에 의한 돌연변이를 탐지하였다. 화살표는 T7E1에 의해 절단된 DNA 밴드의 예상 위치를 나타낸다. 돌연변이 빈도 (Indels (%))는 밴드의 세기를 측정하여 계산하였다. (아래) CCR5 및 C4BPB 야생형 (WT) 및 돌연변이 클론의 DNA 서열. 가이드 RNA에 상보적인 표적 서열의 부분은 **bold**로 보여진다. PAM 서열은 굵은 글씨로 보여진다. 삼각형은 절단 부위를 나타낸다. 마이크로상동 (microhomology)에 상응하는 염기는 밑줄을 그었다. 오른쪽의 열은 삽입 또는 결실된 염기의 수를 나타낸다.

도 4는 RGEN에 의한 오프-타겟 (off-target) 돌연변이는 탐지되지 않는다는 것을 보여준다. (a) 온-타겟 (on-target) 및 잠재적 오프-타겟 서열. 잠재적 오프-타겟 위치에 대해 인 실리코 (*in silico*)에서 인간 유전체를 검색하였다. 네 개의 위치를 밝혀내었고, 각각의 위치는 CCR5 온-타겟 위치와 3-염기 불일치 (3-base mismatch)를 가져왔다. 불일치된 염기는 밑줄로 나타내었다. (b) T7E1 어세이를 사용하여 Cas9/RNA 복합체가 형질주입된 세포에서 상기 위치가 돌연변이 되었는지 여부를 조사하였다. 상기 위치에서 돌연변이는 탐지되지 않았다. N/A (적용할 수 없음), 유전자 간 위치 (intergenic site). (c) Cas9은 오프-타겟-연관 염색체 결실을 유도하지 않았다. CCR5-특이적 RGEN 및 ZFN을 인간 세포에서 발현하였다. PCR을 사용하여 상기 세포에서 15-kb 염색체 결실의 유도를 탐지하였다.

도 5는 마우스에서 RGEN-유도 Foxn1 유전자 타겟팅을 보여준다. (a) 마우스 Foxn1 유전자의 엑손 2에 특이적인 sgRNA를 묘사하는 개략도. 엑손 2에서의 PAM을 적색으로 표시되어 있고, 엑손 2와 상보적인 sgRNA의 서열이 밑줄로 표시되어 있다. 삼각형은 절단 부위를 나타낸다. (b) 1 세포 단계의 마우스 배아에 세포질 내 주입을 통해 전달된, Foxn1-특이적 sgRNA 및 Cas9 mRNA의 유전자 타겟팅 효율을 보여주는 대표적인 T7E1 어세이. 숫자는 가

장 높은 용량으로부터 만들어진 독립적인 파운더(founder) 마우스를 나타낸다. 화살표는 T7E1에 의해 절단된 밴드를 나타낸다. (c) b에서 규명된 세 개의 Foxn1 돌연변이 파운더에서 관찰되는 돌연변이 대립유전자의 DNA 서열. 발생 수는 괄호 안에 나타나있다. (d) Foxn1 파운더 #108 및 야생형 FVB/NTac와 교배하여 유래된 F1 자손의 PCR 유전형질 분석. Foxn1 파운더 #108의 자손에서 발견된 돌연변이 대립유전자 (mutant alleles)의 분리가 나타나있다.

도 6은 Cas9 mRNA 및 Foxn1-sgRNA의 세포질 내 주입에 의한 마우스 배아에서의 Foxn1 유전자 타겟팅을 보여준다. (a) 가장 높은 용량을 주입한 후 돌연변이율을 관찰한 T7E1 어세이의 대표적인 결과. 화살표는 T7E1에 의해 절단된 밴드를 나타낸다. (b) T7E1 어세이 결과의 요약. 표시된 RGEN 용량의 세포질 내 주입 후 획득한 인 비트로에서 배양된 배아 중 돌연변이 비율을 나타낸다. (c) T7E1-양성 돌연변이가 배아의 부분 집합(subset)으로부터 식별된 Foxn1 돌연변이 대립유전자의 DNA 서열. 야생형 대립유전자의 표적 서열은 상자 안에 표시되어 있다.

도 7은 재조합 Cas9 단백질: Foxn1-sgRNA 복합체를 이용한, 마우스 배아에서의 Foxn1 유전자 타겟팅을 보여준다. (a) 및 (b)는 대표적인 T7E1 어세이의 결과 및 이들의 요약이다. 배아를 (a) 전핵 주입(pronuclear injection) 또는 (b) 세포질 내 주입한 후 인 비트로에서 배양하였다. (b) 적색 숫자는 T7E1-양성 돌연변이 파운더 마우스를 나타낸다. (c) 가장 높은 용량의 재조합 Cas9 단백질: Foxn1-sgRNA 복합체의 전핵 주입에 의해 수득된 배아를 인 비트로에서 배양하고, 이로부터 식별된 Foxn1 돌연변이 대립유전자의 DNA 서열. 야생형 대립유전자의 표적 서열은 상자 안에 표시되어 있다.

도 8은 Foxn1 돌연변이 파운더 #12에서 발견되는 돌연변이 대립유전자의 생식선 이동(germ-line transmission)을 보여준다. (a) fPCR 분석. (b) 야생형 FVB/NTac, 파운더 마우스 및 그의 F1 자손의 PCR 유전형질 분석.

도 9는 Prkdc 돌연변이 파운더와 교배하여 발생시킨 배아의 유전자형을 보여준다. Prkdc 돌연변이 파운더 ♂25 및 ♀15를 교배하였고, E13.5 배아를 분리하였다. (a) 야생형, 파운더 ♂25 및 파운더 ♀15의 fPCR 분석. fPCR의 기술적 한계 때문에, 상기 결과들은 돌연변이 대립 유전자의 정확한 서열로부터 작은 차이를 보였다; 예를 들어, 서열 분석에서 $\Delta 269/\Delta 61/WT$ 및 $\Delta 5+1/+7/+12/WT$ 가 각각 파운더 ♂25 및 파운더 ♀15로부터 식별되었다. (b) 발생된 배아의 유전자형.

도 10은 Cas9 단백질/sgRNA 복합체가 표적화된 돌연변이를 유도하였음을 보여준다.

도 11은 애기장대 원형질체(Arabidopsis protoplast)에서 재조합 Cas9 단백질-유도 돌연변이를 보여준다.

도 12은 애기장대 BRI1 유전자에서 재조합 Cas9 단백질-유도 돌연변이 서열을 보여준다.

도 13은 Cas9-mal-9R4L 및 sgRNA/C9R4LC 복합체의 처리에 의해 293 세포의 내재적 CCR5 유전자 파괴를 보여주는 T7E1 어세이를 보여준다.

도 14 (a, b)는 Fu et al. (2013)에서 보고된 RGENs의 온-타겟 및 오프-타겟에서의 돌연변이 빈도를 보여준다. 각각 60 μg 및 120 μg 의 인 비트로에서 전사된 GX19 crRNA 및 tracrRNA, 및 20 μg 의 Cas9-암호화 플라스미드를 단계별 형질주입한 K562 세포 (1×10^5 세포), 또는 (d) 1 μg 의 Cas9-암호화 플라스미드 및 1 μg 의 GX19 sgRNA 발현 플라스미드를 공동-형질주입한 K562 세포 (2×10^5 세포)의 유전체 DNA를 분석한 T7E1 어세이.

도 15 (a, b)는 가이드 RNA 구조의 비교를 보여준다. Fu et al. (2013)에서 보고된 RGENs의 돌연변이 빈도를 T7E1 어세이를 이용하여 온-타겟 및 오프-타겟에서 측정하였다. K562 세포를 Cas9-암호화 플라스미드 및 GX19 sgRNA 또는 GGX20 sgRNA를 암호화하는 플라스미드로 공동-형질도입하였다. 오프-타겟 위치(OT1-3 등)는 Fu et al. (2013)에 나타나 있는 바와 같이 표시되어 있다.

도 16은 Cas9 니카아제(nickases)에 의한 인 비트로 DNA 절단을 보여준다. (a) Cas9 뉴클레아제 및 쌍을 이룬 Cas9 니카아제 (paired Cas9 nickase)의 개요도. PAM 서열 및 절단 위치는 상자 안에 표시되어 있다. (b) 인간 AAVS1 좌위에서의 표적 위치. 각 표적 부위의 위치는 삼각형 안에 표시되어 있다. (c) DNA 절단 반응의 개요도. FAM 염료 (상자 안에 표시됨)를 DNA 기질의 양 5' 말단에 연결하였다. (d) 형광 모세관 전기영동을 사용하여 분석한 DSBs 및 SSBs. 형광 표지된 DNA 기질을 전기영동 전에 Cas9 뉴클레아제 및 니카아제와 함께 배양하였다.

도 17은 Cas9 뉴클레아제 및 니카아제 작용 비교를 보여준다. (a) Cas9 뉴클레아제 (WT), 니카아제 (D10A), 및 니카아제 쌍 (paired nickase)과 관련된 온-타겟 돌연변이 빈도. 5' 오버행(overhang) 또는 3' 오버행을 만들 수 있는 니카아제 쌍이 나타나있다. (b) Cas9 뉴클레아제 및 니카아제 쌍의 오프-타겟 효과의 분석. 세 sgRNA의

7개의 잠재적 오프-타겟 위치의 전체를 분석하였다.

도 18은 다른 내재적 인간 좌위에서 시험한 Cas9 니카아제 쌍을 보여준다. (a,c) 인간 CCR5 및 BRCA2 좌위에서 sgRNA 표적 위치. PAM 서열은 적색으로 표시되어 있다. (b,d) 각 표적 위치에서의 유전체 교정 활성을 T7E1 어세이로 탐지하였다. 5' 오버행을 만들 수 있는 두 닉 (nick)의 수선 (repair)은 3' 오버행을 만드는 것보다 훨씬 더 자주 인델 (indel)의 형성으로 이어졌다.

도 19는 Cas9 니카아제가 상동 재조합을 매개함을 보여준다. (a) 상동 재조합을 탐지하는 전략. 공여체 DNA (donor DNA)는 두 상동 암(two homology arms) 사이에 XbaI 제한 효소 위치를 포함하였던 반면, 내재적 표적 위치는 그 위치가 결여되었다. PCR 어세이를 사용하여 상동 재조합이 일어난 서열을 탐지하였다. 오염된 공여체 DNA의 증폭을 막기 위해, 유전체 DNA에 특이적인 프라이머를 사용하였다. (b) 상동 재조합의 효율. 상동 재조합이 일어났던 영역의 앰플리콘 (amplicon)만이 XbaI에 의해 절단될 수 있다; 절단된 밴드의 강도로 이 방법의 효율을 측정하였다.

도 20은 Cas9 니카아제 쌍에 의해 유도된 DNA 스플라이싱 (splicing)을 보여준다. (a) 인간 AAVS1 좌위에서 니카아제 쌍의 표적 위치. AS2 위치와 각각의 다른 위치 사이의 거리를 보여준다. 화살표는 PCR 프라이머를 나타낸다. (b) PCR을 사용하여 탐지한 유전체 결실. 별표는 결실-특이적 PCR 산물을 나타낸다. (c) AS2 및 L1 sgRNA를 사용하여 얻은 결실-특이적 PCR 산물의 DNA 서열. 표적 위치 PAM 서열은 상자 안에 표시되어 있고, sgRNA-매칭 서열은 대문자로 표시되어 있다. 온전한 sgRNA-매칭 서열은 밑줄로 표시되어 있다. (d) Cas9 니카아제 쌍-매개 염색체 결실의 도식 모델. 새로 합성된 DNA 가닥은 상자 안에 표시되어 있다.

도 21은 Cas9 니카아제 쌍은 전좌 (translocation)를 유도하지 않는 것을 보여준다. (a) 온-타겟 및 오프-타겟 위치 사이의 염색체 전좌의 도식 개요. (b) 염색체 전좌를 탐지하기 위한 PCR 증폭. (c) 니카아제 쌍이 아닌 Cas9 뉴클레아제에 의해 유도된 전좌.

도 22는 T7E1 및 RFLP 어세이의 개념도를 보여준다. (a) 이배체 세포에 유전자 가위 처리 후, 네 가지 가능한 시나리오에서 어세이 절단 반응의 비교: (A) 야생형, (B) 단일 대립유전자성 돌연변이 (monoallelic mutation), (C) 서로 다른 이종대립유전자성 돌연변이, 이형 (different biallelic mutations, hetero), 및 (D) 동일 이종대립유전자성 돌연변이, 동형 (identical biallelic mutations, homo). 검정색 줄은 각 대립유전자로부터 유래한 PCR 산물을 나타내고; 대시 (dashed) 및 점선 (dotted)의 상자는 NHEJ에 의해 생성된 삽입/결실 돌연변이를 나타낸다. (b) 전기영동에 의해 분석된 T7E1 및 RGEN 절단의 예상된 결과.

도 23은 인델 (indel)을 지닌 C4BPB 표적 위치를 포함하는 선형화된 플라스미드의 인 비트로 절단 어세이를 보여준다. 개별적인 플라스미드 기질의 DNA 서열(위 패널). PAM 서열은 밑줄로 표시되어 있다. 삽입된 염기는 상자 안에 표시되어 있다. 화살표 (아래 패널)는 전기영동 후 야생형-특이적 RGEN에 의해 절단된 DNA 밴드의 예상된 위치를 나타낸다.

도 24는 RGEN-매개 RFLP를 통해 세포에서 유전자 가위에 의해 유도된 돌연변이의 유전형질 분석을 보여준다. (a) C4BPB 돌연변이 K562 세포 클론의 유전형질. (b) 불일치-민감 T7E1 어세이 (mismatch-sensitive T7E1 assay)의 RGEN-매개 RFLP 분석과의 비교. 검정색 화살표는 T7E1 효소 또는 RGENs의 처리에 의한 절단 산물을 나타낸다.

도 25는 RGEN-RFLP 기술을 통한 RGEN-유도 돌연변이의 유전형질 분석을 보여준다. (a) RGEN-RFLP 및 T7E1 어세이를 이용한 C4BPB-과피 클론의 분석. 화살표는 RGEN 또는 T7E1에 의해 절단되는 DNA 밴드의 예상 위치를 나타낸다. (b) T7E1 어세이와 RGEN-RFLP 분석의 정량적 비교. 야생형 및 C4BPB-과피 K562 세포에서 얻은 유전체 DNA (genomic DNA) 시료를 다양한 비율로 혼합하고, PCR 증폭하였다. (c) RFLP 및 T7E1 분석을 통한 HeLa 세포에서 HLA-B 유전자의 RGEN-유도 돌연변이에 대한 유전형질 분석.

도 26은 유기체에서 RGEN-매개 RFLP를 통한 유전자 가위에 의해 유도된 돌연변이의 유전형질 분석을 보여준다. (a) Pibf1 돌연변이 파운더 파우스의 유전형질. (b) 불일치-민감 T7E1 어세이 (mismatch-sensitive T7E1 assay)의 RGEN-매개 RFLP 분석과의 비교. 검정색 화살표는 T7E1 효소 또는 RGENs의 처리에 의한 절단 산물을 나타낸다.

도 27은 ZFN-유도 돌연변이의 RGEN-매개 유전형질 분석을 보여준다. ZFN 표적 위치는 상자 안에 표시된다. 검정색 화살표는 T7E1에 의해 절단된 DNA 밴드를 나타낸다.

도 28은 인간 HLA-B 유전자의 영역에서 다형성 위치를 보여준다. RGEN 표적 위치를 둘러싸는 서열은 HeLa 세포

로부터의 PCR 앰플리콘의 서열이다. 다형성 위치는 상자 안에 표시된다. RGEN 표적 위치 및 PAM 서열을 각각 대시 (dashed) 및 굵은 글씨 (bolded)의 상자 안에 표시되어 있다. 프라이머 서열을 밑줄로 표시하였다.

도 29는 RGEN-RFLP 분석을 통한 발암성 돌연변이의 유전형질 분석을 보여준다. (a) HCT116 세포에서 인간 CTNNB1 유전자에서의 반복 돌연변이 (recurrent mutation) (TCT의 c.133-135 결실)를 RGENs로 탐지하였다. HeLa 세포를 음성 대조군으로 사용하였다. (b) 불일치 가이드 RNA (mismatched guide RNA)를 포함한 RGENs으로 A549 암세포에서 KRAS 치환 돌연변이 (c.34 G>A)의 유전형질 분석. 불일치 뉴클레오타이드 (mismatched nucleotide)가 상자 안에 표시되어 있다. HeLa 세포를 음성 대조군으로 사용하였다. 화살표는 RGENs에 의해 절단된 DNA 밴드를 나타낸다. Sanger 시퀀싱에 의해 확인된 DNA 서열이 표시되어 있다.

도 30은 RGEN-RFLP 분석을 통한 HEK293T 세포에서 CCR5 delta32 대립유전자의 유전형질 분석을 보여준다. (a) 세포주의 RGEN-RFLP 어레이. K562, SKBR3, 및 HeLa 세포를 야생형 대조군으로 사용하였다. 화살표는 RGENs에 의해 절단된 DNA 밴드를 나타낸다. (b) 야생형 및 delta32 CCR5 대립유전자의 DNA 서열. RFLP 분석에 사용된 RGENs의 온-타겟 위치 및 오프-타겟 위치 모두를 밑줄로 표시하였다. 두 위치 간의 단일-뉴클레오타이드 불일치는 상자 안에 표시되어 있다. PAM 서열은 밑줄로 표시되어 있다. (c) 야생형-특이적 RGENs을 이용한 야생형 또는 del132 CCR5 대립유전자를 갖고 있는 플라스미드의 인 비트로 절단. (d) CCR5 좌위에서 CCR5-delta32-특이적 RGEN의 오프-타겟 위치의 존재 확인. del132-특이적 RGENs의 다양한 양을 이용한 온-타겟 또는 오프-타겟 서열 중 어느 하나를 가지고 있는 플라스미드의 인 비트로 절단 어레이.

도 31은 KRAS 점 돌연변이 (c.34 G>A)의 유전형질 분석을 보여준다. (a) 암 세포주에서 KRAS 돌연변이 (c.34 G>A)의 RGEN-RFLP 분석. 점 돌연변이에 대해 동형인 HeLa 세포 (야생형 대조군으로 사용됨) 또는 A549 세포의 PCR 산물을, 야생형 서열 또는 돌연변이 서열에 특이적이며, 완벽하게 일치하는 crRNAs (perfectly matched crRNAs)와 함께 RGENs으로 절단하였다. 상기 세포의 KRAS 유전형질은 Sanger 시퀀싱으로 확인하였다. (b) 야생형 또는 돌연변이 KRAS 서열 중 어느 하나를 가지는 플라스미드를, 완벽하게 일치하는 crRNAs (perfectly matched crRNAs) 또는 약화된, 하나의 염기가 불일치된 crRNAs와 함께 RGENs을 사용하여 절단하였다. 유전형질 분석을 위해 선택된, 약화된 crRNAs가 젤 위의 상자 안에 표시되어 있다.

도 32는 PIK3CA 점 돌연변이 (c.3140 A>G)의 유전형질 분석을 보여준다. (a) 암 세포주에서 PIK3CA 돌연변이 (c.3140 A>G)의 RGEN-RFLP 분석. 점 돌연변이가 이형접합인 HeLa 세포 (야생형 대조군으로 사용됨) 또는 HCT116 세포의 PCR 산물을, 야생형 서열 또는 돌연변이 서열에 특이적이며, 완벽하게 일치하는 crRNA와 함께 RGENs으로 절단하였다. 상기 세포의 PIK3CA 유전형질을 Sanger 시퀀싱으로 확인하였다. (b) 야생형 또는 돌연변이 PIK3CA 서열 중 어느 하나를 갖는 플라스미드를, 완벽하게 일치하는 crRNAs, 또는 약화된, 하나의 염기가 불일치하는 crRNAs와 함께 RGENs을 사용하여 절단하였다. 유전형질 분석을 위해 선택된, 약화된 crRNAs를 젤 위의 상자 안에 표시되어 있다.

도 33은 암 세포주에서 반복 점 돌연변이 (recurrent point mutation)의 유전형질 분석을 보여준다. (a) IDH에서 반복 발암 점 돌연변이 (c.394c>T)의 RGEN-RFLP 어레이, (b) PIK3CA (c.394A>T), (c) NRAS (c.181C>A), (d) 및 BRAF 유전자 (c.1799T>A). Sanger 시퀀싱에 의해 확인된 각 세포주의 유전형질이 표시되어 있다. 불일치 뉴클레오타이드 (mismatched nucleotide)가 상자 안에 표시되어 있다. 검정색 화살표는 RGENs에 의해 절단된 DNA 밴드를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0042]

본 발명의 하나의 측면에 따르면, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 조성물을 제공한다. 또한, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 조성물의 용도를 제공한다.

[0043]

본 발명에서, 조성물은 또한 RNA-가이드 뉴클레아제 (RNA-guided nuclease, RGEN) 조성물로도 지칭된다.

[0044]

ZFNs 및 TALENs는 포유동물, 모델 유기체, 식물 및 가축에서 표적화된 돌연변이 (targeted mutagenesis)를 가능하게 하지만, 개별적인 뉴클레아제 (individual nuclease)로 획득된 돌연변이 빈도는 서로 매우 다르다. 더욱

이, 몇몇 ZFNs 및 TALENs는 어떠한 유전체 교정 활성을 보여주지 못한다. DNA 메틸화는 표적 위치에 대한 상가 유전자 가위 (engineered nuclease)의 결합을 제한할 수 있다. 게다가, 개개인의 요구에 맞춘 뉴클레아제 (customized nuclease)를 만드는 것은 기술적으로 까다롭고 시간이 많이 걸린다.

[0045] 본 발명자들은 Cas 단백질 기반의 새로운 RNA-가이드 엔도뉴클레아제 조성물을 개발하여 ZFNs 및 TALENs의 단점을 극복하였다.

[0046] 본 발명에 앞서, Cas 단백질의 엔도뉴클레아제 활성은 밝혀져 있었다. 하지만, 진핵 유전체의 복잡성 때문에 Cas 단백질의 엔도뉴클레아제 활성이 진핵 세포에서도 기능을 하는지 여부는 알려지지 않았다. 추가로, 지금까지 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한 Cas 단백질 또는 Cas 단백질을 암호화하는 핵산 및 표적 DNA에 특이적인 가이드 RNA를 포함하는 조성물은 개발되지 않았다.

[0047] ZFNs 및 TALENs와 비교하여, 오로지 합성된 가이드 RNA 구성요소 (synthetic guide RNA component)가 새로운 유전체-교정 뉴클레아제 (new genome-editing nuclease)를 만들기 위해 대체되기 때문에, Cas 단백질을 기반으로 한 본 발명의 RGEN 조성물은 더욱 쉽게 커스텀화(customized) 될 수 있을 것이다. 커스텀화 RNA 가이드 엔도뉴클레아제를 만드는 것에 있어 서브-클로닝 (sub-cloning) 단계는 관여하지 않는다. 더욱이, TALEN 유전자(~6 kbp)의 쌍과 비교했을 때, Cas 유전자의 상대적으로 작은 크기 (예를 들어, Cas9는 4.2 kbp)는 바이러스-매개 유전자 전달 같은 몇몇 적용 분야에서 RNA-가이드 엔도뉴클레아제 조성물에 이점을 제공한다. 추가로, 이러한 RNA-가이드 엔도뉴클레아제는 오프-타겟 (off-target) 효과를 갖지 않고, 이에 따라 원하지 않는 돌연변이, 결실, 반전 및 중복을 야기하지 않는다. 이러한 특징은 본 발명의 RNA-가이드 엔도뉴클레아제 조성물이 진핵 세포 및 유기체에서의 유전체 공학 (genome engineering)에 대한 확장 가능하며 (scalable), 다목적적으로 쓰이며 (versatile), 편리한 (convenient) 수단이 될 수 있도록 한다. 게다가, RGEN은 어떠한 DNA 서열도 표적화하도록 설계될 수 있고, 거의 모든 단일 뉴클레오타이드 다형성 (single nucleotide polymorphism) 또는 작은 삽입/결실 (indel)이 RGEN-매개 RFLP에 의해 분석될 수 있다. RGENs의 특이성은 20 염기쌍(bp)까지의 길이의 표적 DNA 서열과 혼성화되는 RNA 요소 및 프로토스페이스어-인접 모티프(protospacer-adjacent motif, PAM)를 인식하는 Cas9 단백질에 의해 결정된다. RGENs는 RNA 구성요소를 대체함으로써 쉽게 리프로그래밍된다. 그러므로, RGENs는 다양한 서열 변이에 대해 간단하고 강력한 RFLP 분석을 사용하는 플랫폼을 제공한다.

[0048] 표적 DNA는 내재적 DNA (endogenous DNA), 또는 인위적인 DNA (artificial DNA)일 수 있고, 바람직하게는, 내재적 DNA이다.

[0049] 본원에서 사용된, 용어 "Cas 단백질"은 CRISPR/Cas 시스템에서 필수적인 단백질 요소를 의미하고, CRISPR RNA (crRNA) 및 트랜스-활성화 crRNA (trans-activating crRNA, tracrRNA)로 불리는 두 RNA와 복합체를 형성할 때, 활성 엔도뉴클레아제 또는 니카아제 (nickase)를 형성한다.

[0050] Cas 유전자 및 단백질의 정보는 국립생명공학정보센터 (national center for biotechnology information, NCBI)의 GenBank에서 구할 수 있으나, 이에 제한되지 않는다.

[0051] Cas 단백질을 암호화하는 CRISPR-연관 (CRISPR-associated, cas) 유전자는 종종 CRISPR-반복 스페이스 배열 (CRISPR repeat-spacer array)과 관련된다. 40개 이상의 서로 다른 Cas 단백질 패밀리가 기재되어 왔다. 이러한 단백질 패밀리 중, Cas1은 서로 다른 CRISPR/Cas 시스템 중에서 아주 흔한 (ubiquitous) 것으로 보인다. CRISPR-Cas 시스템은 세 종류가 있다. 이들 중에서, Cas9 단백질 및 crRNA 및 tracrRNA를 수반하는 타입 II CRISPR/Cas 시스템이 대표적이며, 잘 알려져 있다. cas 유전자 및 반복 구조 (repeat structure)의 특정 조합은 8개의 CRISPR 하위 유형 (Ecoli, Ypest, Nmeni, Dvulg, Tneap, Hmari, Apern, 및 Mtube)을 정의하는데 사용되어 왔다.

[0052] Cas 단백질은 단백질 전달 도메인 (protein transduction domain)과 연결될 수 있다. 상기 단백질 전달 도메인은 폴리-아르기닌(poly-arginine) 도메인 또는 HIV로부터 유래한 TAT 단백질일 수 있지만, 이에 한정되는 것은

아니다.

- [0053] 본 발명의 조성물은 단백질의 형태 또는 Cas 단백질을 암호화하는 핵산의 형태로 Cas 요소를 포함할 수 있다.
- [0054] 본 발명에서, Cas 단백질은 가이드 RNA와 복합체를 형성할 때 엔도뉴클레아제 또는 니카아제 활성을 갖는다면, 어떠한 Cas 단백질일 수 있다.
- [0055] 바람직하게, Cas 단백질은 Cas9 단백질 또는 이의 변이체이다.
- [0056] Cas9 단백질의 변이체는 촉매적 아스파라긴산 잔기 (catalytic aspartate residue)가 임의의 다른 아미노산으로 변경된 Cas9의 돌연변이 형태일 수 있다. 바람직하게, 다른 아미노산은 알라닌(alanine)일 수 있지만, 이에 제한되지 않는다.
- [0057] 추가로, Cas9 단백질은 스트렙토코커스 sp. (Streptococcus sp.), 바람직하게는 스트렙토코커스 피요젠스 (Streptococcus pyogens)와 같은 유기체로부터 분리된 것 또는 재조합 단백질일 수 있으나, 이에 제한되지 않는다.
- [0058] 스트렙토코커스 피요젠스로부터 유래한 Cas 단백질은 NGG 트리뉴클레오타이드(trinucleotide)를 인식할 수 있다. 상기 Cas 단백질은 서열번호: 109의 아미노산 서열을 포함할 수 있으나, 이에 제한되지 않는다.
- [0059] 상기 용어 "재조합"은, 예컨대 세포, 핵산, 단백질 또는 벡터 등을 언급하며 사용될 때, 이종 (heterologous) 핵산 또는 단백질의 도입 또는 천연형 (native) 핵산 또는 단백질의 변경, 또는 변형된 세포로부터 유래한 세포에 의해 변형된 세포, 핵산, 단백질, 또는 벡터를 나타낸다. 따라서, 예컨대, 재조합 Cas 단백질은 인간 코돈 표 (human codon table)를 이용하여 Cas 단백질을 암호화하는 서열을 재구성함으로써 만들 수 있다.
- [0060] 본 발명에 관해서, Cas 단백질-암호화 핵산은 CMV 또는 CAG와 같은 프로모터 하에서 Cas-암호화 서열을 포함하는 플라스미드 같은 벡터의 형태일 수 있다. Cas 단백질이 Cas9일 때, Cas9 암호화 서열은 스트렙토코커스 sp.로부터 유래한 것일 수 있고, 바람직하게 스트렙토코커스 피요젠스로부터 유래한 것일 수 있다. 예를 들어, Cas9 암호화 핵산은 서열번호: 1의 뉴클레오타이드 서열을 포함할 수 있다. 더욱이, Cas9 암호화 핵산은 서열번호: 1의 서열과 적어도 50%의 상동성을 갖는 뉴클레오타이드 서열을 포함할 수 있고, 바람직하게는 서열번호: 1의 서열과 적어도 60, 70, 80, 90, 95, 97, 98, 또는 99%의 상동성을 가지는 뉴클레오타이드 서열을 포함할 수 있으나, 이에 제한되는 것은 아니다. Cas9 암호화 핵산은 서열번호 108, 서열번호 110, 서열번호 106, 또는 서열번호 107의 뉴클레오타이드 서열을 포함할 수 있다.
- [0061] 본원에서 사용된, 용어 "가이드 RNA" 는 표적 DNA에 특이적인 RNA로, Cas 단백질과 복합체를 형성할 수 있고, Cas 단백질을 표적 DNA에 가져오는 RNA를 말한다.
- [0062] 본 발명에서, 상기 가이드 RNA는 두 개의 RNA, 즉, CRISPR RNA (crRNA) 및 트랜스활성화 crRNA (transactivating crRNA, tracrRNA)로 이루어져 있는 것일 수 있으며, 또는 crRNA 및 tracrRNA의 필수적 부분의 융합에 의해 생성된 단일 사슬 RNA (single-chain RNA, sgRNA)일 수 있다.
- [0063] 상기 가이드 RNA는 crRNA 및 tracrRNA를 포함하는 이중RNA (dual RNA)일 수 있다.
- [0064] 만약 상기 가이드 RNA가 crRNA 및 tracrRNA의 필수적인 부분 및 표적과 상보적인 부분을 포함한다면, 어떠한 가이드 RNA라도 본 발명에 사용될 수 있다.
- [0065] 상기 crRNA는 표적 DNA와 혼성화될 수 있다.
- [0066] 상기 RGEN은 Cas 단백질 및 이중RNA (불변의 tracrRNA 및 표적-특이적 crRNA), 또는 Cas 단백질 및 sgRNA (불변의 tracrRNA 및 표적-특이적 crRNA의 필수적 부분의 융합)으로 구성될 수 있고, crRNA를 대체하여 쉽게 리프로그래밍될 수 있다.

- [0067] 상기 가이드 RNA는 단일-사슬 가이드 RNA 또는 이중RNA의 crRNA의 5' 말단에서 하나 또는 그 이상의 추가적인 뉴클레오타이드를 더 포함할 수 있다.
- [0068] 바람직하게, 상기 가이드 RNA는 단일-사슬 가이드 RNA 또는 이중RNA의 crRNA의 5' 말단에 2개의 추가적인 구아닌(guanine) 뉴클레오타이드를 더 포함할 수 있다.
- [0069] 가이드 RNA는 RNA의 형태 또는 가이드 RNA를 암호화하는 DNA의 형태로 세포 또는 유기체에 전달될 수 있다. 가이드 RNA는 분리된 RNA의 형태, 바이러스 벡터에 포함되어 있는 RNA, 또는 벡터에 암호화되어 있는 형태일 수도 있다. 바람직하게, 상기 벡터는 바이러스 벡터, 플라스미드 벡터, 또는 아그로박테리움 (agrobacterium) 벡터일 수 있지만, 이에 제한되는 것은 아니다.
- [0070] 가이드 RNA를 암호화하는 DNA는 가이드 RNA를 암호화하는 서열을 포함하는 벡터일 수 있다. 예를 들어, 분리된 가이드 RNA 또는 가이드 RNA를 암호화하는 서열 및 프로모터를 포함하는 플라스미드 DNA를 세포 또는 유기체에 형질주입하여, 세포 또는 유기체에 가이드 RNA를 전달할 수 있다.
- [0071] 다른 방법으로, 바이러스-매개 유전자 전달을 이용하여 가이드 RNA를 세포 또는 유기체에 전달할 수 있다.
- [0072] 가이드 RNA가 분리된 RNA의 형태로 세포 또는 유기체에 형질주입될 때, 당업계에 알려진 임의의 인 비트로 전사 시스템을 사용하여 인 비트로 전사함으로써 가이드 RNA를 제조할 수 있다. 가이드 RNA는, 바람직하게, 가이드 RNA를 암호화하는 서열을 포함하는 플라스미드의 형태보다 분리된 RNA의 형태로 세포에 전달된다. 본원에 사용된, 용어 "분리된 RNA"는 "네이키드 RNA (naked RNA)"와 교체하여 사용할 수 있다. 이는 클로닝 단계를 필요로 하지 않기 때문에 비용 및 시간을 절약할 수 있다. 하지만, 가이드 RNA의 형질주입을 위한 플라스미드 DNA 또는 바이러스-매개 유전자 전달의 사용이 배제되는 것은 아니다.
- [0073] Cas 단백질 또는 Cas 단백질-암호화 핵산 및 가이드 RNA를 포함하는 본 발명의 RGEN 조성물은, 표적에 대한 가이드 RNA의 특이성 및 Cas 단백질의 엔도뉴클레아제 또는 니카아제 활성 때문에 표적 DNA를 특이적으로 절단할 수 있다.
- [0074] 본원에 사용된, 용어 "절단"은 뉴클레오타이드 분자의 공유 결합 백본 (covalent backbone)의 파손 (breakage)을 말한다.
- [0075] 본 발명에서, 가이드 RNA는 절단하고자 하는 어떠한 표적에 특이적이 되도록 제조될 수 있다. 따라서, 본 발명의 RGEN 조성물은 가이드 RNA의 표적-특이적 부분을 조작하거나 유전형질 분석(genotyping)함으로써 어떠한 표적 DNA도 절단할 수 있다.
- [0076] 가이드 RNA 및 Cas 단백질은 한 쌍으로서 작용할 수 있다. 본원에 사용된, 용어 "Cas 니카아제 쌍 (paired Cas nickage)"은 쌍으로서 기능하는 가이드 RNA 및 Cas 단백질을 의미한다. 한 쌍은 두 개의 가이드 RNA를 포함한다. 가이드 RNA 및 Cas 단백질은 쌍으로서 작용할 수 있고, 서로 다른 DNA 가닥에 두 개의 틈 (nick)을 유도할 수 있다. 두 개의 틈은 적어도 100 bps 분리되어 있을 수 있지만, 이에 제한되는 것은 아니다.
- [0077] 실시예에서, 본 발명자들은 Cas 니카아제 쌍이 인간 세포의 유전체에서 표적화된 돌연변이 및 1-kbp 염색체 부분까지 큰 결실을 야기한다는 것을 확인하였다. 중요하게도, 니카아제 쌍은 이들의 상응하는 뉴클레아제가 돌연변이를 유발하는 오프-타겟 위치에서 인델 (indel)을 유도하지 않았다. 더욱이, 뉴클레아제와 다르게, 니카아제 쌍은 오프-타겟 DNA 절단과 관련된 원하지 않는 전좌를 유도하지 않았다. 원칙적으로, 니카아제 쌍은 Cas9-매개 돌연변이의 특이성을 두 배로 하고, 유전자 및 세포 치료 같은 정확한 유전체 교정을 요구하는 적용 분야에서 RNA-가이드 효소의 효용성을 넓힐 것이다.
- [0078] 본 발명에서, 상기 조성물은 인 비트로에서 진핵 세포 또는 유기체의 유전체의 유전형질 분석에 사용될 수

있다.

- [0079] 하나의 특정 양태에서, 상기 가이드 RNA는 서열번호 1의 뉴클레오타이드 서열을 포함할 수 있고, 여기서 뉴클레오타이드 위치 3 ~ 22의 부분은 표적-특이적 부분이고, 그리고 그 부분의 서열은 표적에 따라 바뀔 수 있다.
- [0080] 본 발명에서 사용된, 진핵 세포 또는 유기체는 효모, 곰팡이, 원생동물 (protozoa), 식물, 고등 식물 및 곤충, 또는 양서류의 세포, 또는 CHO, HeLa, HEK293, 및 COS-1과 같은 포유 동물의 세포일 수 있고, 예를 들어, 당업계에서 일반적으로 사용되는, 배양된 세포 (인 비트로), 이식된 세포 (graft cell) 및 일차 세포 배양 (인 비트로 및 엑스 비보(*ex vivo*)), 및 인 비보(*in vivo*) 세포, 또한 인간을 포함하는 포유동물의 세포 (mammalian cell)일 수 있으나, 이에 제한되지 않는다.
- [0081] 하나의 특정 양태에서, Cas9 단백질/단일 사슬 가이드 RNA는 인 비트로 및 높은 빈도로 표적화된 유전체 돌연변이를 유도하는 자발적인 수선을 하는 포유 동물의 세포에서 위치-특이적 DNA 이중 나선의 절단을 생성할 수 있다는 것을 밝혔다.
- [0082] 더욱이, 유전자-녹아웃(knockout) 마우스는 Cas9 단백질/가이드 RNA 복합체 또는 Cas9 mRNA/가이드 RNA를 1 세포 단계 (one-cell stage)의 배아에 주입하여 유도할 수 있고, 생식선 유전성 돌연변이 (germ-line transmittable mutation)는 Cas9/가이드 RNA 시스템에 의해 생성될 수 있다는 것을 밝혔다.
- [0083] 외인성 DNA가 유기체로 도입되지 않기 때문에 표적화된 돌연변이를 유도하기 위해서는 Cas 단백질을 암호화하는 핵산보다 Cas 단백질을 사용하는 것이 더 유리하다. 그러므로, Cas 단백질 및 가이드 RNA를 포함하는 조성물을 치료제 또는 부가가치 작물, 가축, 가금류, 생선, 애완동물 등을 개발하는데 사용할 수 있다.
- [0084] 본 발명의 다른 측면에 따르면, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적화된 돌연변이를 유도하기 위한 조성물을 제공한다. 또한, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적화된 돌연변이를 유도하기 위한 상기 조성물의 용도를 제공한다.
- [0085] 가이드 RNA, Cas 단백질을 암호화하는 핵산 또는 Cas 단백질은 상기 기술된 바와 같다.
- [0086] 본 발명의 다른 측면에 따르면, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는, 진핵 세포 또는 유기체에서 표적 DNA를 절단하기 위한, 또는 표적화된 돌연변이를 유도하기 위한 키트를 제공한다.
- [0087] 가이드 RNA, Cas 단백질을 암호화하는 핵산 또는 Cas 단백질은 상기 기술된 바와 같다.
- [0088] 상기 키트는 가이드 RNA 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 별도의 구성요소 또는 하나의 조성물로서 포함할 수 있다.
- [0089] 본 발명의 키트는 가이드 RNA 및 Cas 요소를 세포 또는 유기체에 전달하는데 필요한 몇몇 추가적인 요소를 포함할 수 있다. 예를 들어, 상기 키트는 DEPC-처리된 주입 버퍼와 같은 주입 버퍼 (injection buffer), 및 표적 DNA의 돌연변이를 분석하는데 필요한 재료를 포함할 수 있지만, 이에 제한되는 것은 아니다.

- [0090] 또 다른 측면에 따르면, 본 발명은 진핵 세포 또는 유기체에 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질, 및 가이드 RNA 또는 가이드 RNA를 암호화하는 DNA를 공동-형질주입(co-transfecting) 또는 단계적-형질주입(serial-transfecting)하는 단계를 포함하는, Cas 단백질 및 가이드 RNA를 포함하는 진핵 세포 또는 유기체를 제조하는 방법을 제공한다.
- [0091] 가이드 RNA, Cas 단백질을 암호화하는 핵산 또는 Cas 단백질은 상기 기술된 바와 같다.
- [0092] 본 발명에서, Cas 단백질을 암호화하는 핵산 또는 Cas 단백질 및 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA는 미세주입법 (microinjection), 전기천공법 (electroporation), DEAE-덱스트란 처리 (DEAE-dextran treatment), 리포펙션 (lipofection), 나노파티클-매개 형질주입, 단백질 전달 도메인 매개 도입, 바이러스-매개 유전자 전달, 및 원생동물에서 PEG-매개 형질주입 등과 같은 당업계의 다양한 방법에 의해 세포로 전달될 수 있지만, 이에 제한되는 것은 아니다. 또한, Cas 단백질을 암호화하는 핵산 또는 Cas 단백질 및 가이드 RNA는 주입 같은, 유전자 또는 단백질을 부여하는 당업계의 다양한 방법에 의해 유기체로 전달될 수 있다. Cas 단백질-암호화 핵산 또는 Cas 단백질은 가이드 RNA와의 복합체의 형태 또는 독립된 형태로 세포 안으로 전달될 수 있다. Tat와 같이 단백질 전달 도메인이 융합된 Cas 단백질은 세포 내로 효율적으로 전달될 수 있다.
- [0093] 바람직하게, 진핵 세포 또는 유기체는 Cas9 단백질 및 가이드 RNA와 함께 공동-형질주입 또는 단계적-형질주입될 수 있다.
- [0094] 단계적-형질주입은 처음에 Cas 단백질을 암호화하는 핵산, 이어서 두 번째는 네이키드 가이드 RNA (naked guide RNA)를 형질주입할 수 있다. 바람직하게, 두 번째 형질주입은 3, 6, 12, 18, 24 시간 후이지만, 이에 제한되는 것은 아니다.
- [0095] 다른 측면에 따르면, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질-암호화 핵산 또는 Cas 단백질을 포함하는 진핵 세포 또는 유기체를 제공한다.
- [0096] 상기 진핵 세포 또는 유기체는 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질-암호화 핵산 또는 Cas 단백질을 포함하는 조성물을 세포 또는 유기체에 전달함으로써 제조될 수 있다.
- [0097] 상기 진핵 세포는 효모, 곰팡이, 원생동물 (protozoa), 식물, 고등 식물 및 곤충, 또는 양서류의 세포, 또는 CHO, HeLa, HEK293, 및 COS-1과 같은 포유 동물의 세포일 수 있고, 예를 들어, 당업계에서 일반적으로 사용되는, 배양된 세포 (인 비트로), 이식된 세포 (graft cell) 및 일차 세포 배양 (인 비트로 및 엑스 비보(*ex vivo*)), 및 인 비보 (*in vivo*) 세포, 및 또한 인간을 포함하는 포유동물의 세포 (mammalian cell)일 수 있으나, 이에 제한되지 않는다. 또한, 상기 유기체는 효모, 곰팡이, 원생동물, 식물, 고등 식물 및 곤충, 양서류, 또는 포유 동물일 수 있다.
- [0098] 발명의 다른 측면에 따르면, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 조성물로 표적 DNA를 포함하는 세포 또는 유기체를 처리하는 단계를 포함하는, 진핵 세포 또는 유기체에서 표적 DNA 절단 또는 표적화된 돌연변이를 유도하는 방법을 제공한다.
- [0099] 세포 또는 유기체에 조성물을 처리하는 단계는, 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 본 발명의 조성물을 세포 또는 유기

체에 전달함으로써 수행될 수 있다.

- [0100] 상기 기술한 바와 같이, 상기 전달 (transfer)은 미세주입법 (microinjection), 형질주입 (transfection), 전기 천공법 (electroporation) 등에 의해 수행될 수 있다.
- [0101] 발명의 다른 측면에 따르면, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 본 발명의 RGEN 조성물에 의해 교정된 유전체를 포함하는 배아를 제공한다.
- [0102] 어느 배아도 본 발명에서 사용될 수 있고, 본 발명을 위하여, 상기 배아는 마우스의 배아일 수 있다. 상기 배아는 PMSG (Pregnant Mare Serum Gonadotropin) 및 hCG (human Chorionic Gonadotropin)를 4 내지 7주령의 암컷 마우스에 주입하여 생산할 수 있고, 과배란된 암컷 마우스(super-ovulated female mouse)가 수컷과 교배될 수 있고, 수정된 배아를 난관 (oviduct)으로부터 모을 수 있다.
- [0103] 배아에 도입된 본 발명의 RGEN 조성물은 Cas 단백질의 활동에 의해 가이드 RNA에 상보적인 표적 DNA를 절단할 수 있고, 표적 DNA에서 돌연변이를 야기할 수 있다. 그러므로, 본 발명의 RGEN 조성물이 도입된 배아는 교정된 유전체를 갖는다.
- [0104] 하나의 특정한 양태에서, 본 발명의 RGEN 조성물은 마우스 배아에서 돌연변이를 야기할 수 있고, 돌연변이는 자손에게 전달될 수 있다.
- [0105] RGEN 조성물을 배아에 도입하는 방법은 미세주입법, 줄기세포 삽입 (stem cell insertion), 레트로바이러스 삽입 (retrovirus insertion) 등과 같은 당업계에 알려진 어떠한 방법일 수 있다. 바람직하게, 미세주입법 기술이 사용될 수 있다.
- [0106] 다른 측면에 따르면, 본 발명은 본 발명의 RGEN 조성물에 의해 교정된 유전체를 포함하는 배아를 동물의 난관에 이식하여 수득한 유전체-변형 동물을 제공한다.
- [0107] 본 발명에서, 용어 "유전체-변형 동물"은 본 발명의 RGEN 조성물에 의해 배아 단계에서 유전체가 변형된 동물을 말하고, 동물의 종류는 제한되지 않는다.
- [0108] 상기 유전체-변형 동물은 본 발명의 RGEN 조성물에 기반한 표적화된 돌연변이에 의해 야기된 돌연변이를 갖는다. 상기 돌연변이는 결실, 삽입, 전좌, 반전 중 어느 하나일 수 있다. 돌연변이의 위치는 RGEN 조성물의 가이드 RNA의 서열에 의존한다.
- [0109] 유전자에 돌연변이를 갖는 유전체-변형 동물은 유전자 기능을 확인하기 위해 사용될 수 있다.
- [0110] 발명의 다른 측면에 따르면, 본 발명은 표적 DNA에 특이적인 가이드 RNA 또는 상기 가이드 RNA를 암호화하는 DNA, 및 Cas 단백질을 암호화하는 핵산 또는 Cas 단백질을 포함하는 본 발명의 RGEN 조성물을 동물의 배아에 도입하는 단계; 및 상기 배아를 가임신 위탁모(pseudopregnant foster mother)의 난관에 이식하여 유전체-변형 동물을 생산하는 단계를 포함하는, 유전체-변형 동물을 제조하는 방법을 제공한다.

- [0111] 본 발명의 RGEN 조성물을 도입하는 단계는 미세주입법, 줄기세포 삽입, 레트로바이러스 삽입 등과 같은 당업계에 알려진 어떠한 방법에 의해 달성될 수 있다.
- [0112] 발명의 다른 측면에 따르면, 본 발명은 RGEN 조성물을 포함하는 원핵 세포를 위한 방법에 의해 제조된, 유전체-변형 원생동물로부터 재생된 식물을 제공한다.
- [0113] 발명의 다른 측면에 따르면, 본 발명은 표적 DNA 서열에 특이적인 가이드 RNA, Cas 단백질을 포함하는, 분리된 생물학적 시료에서 돌연변이 또는 변이(variation)를 유전형질 분석(genotyping) 하기 위한 조성물을 제공한다. 또한, 본 발명은 표적 DNA 서열에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는, 분리된 생물학적 시료에서 병원성 미생물의 핵산 서열을 유전형질 분석하기 위한 조성물을 제공한다.
- [0114] 가이드 RNA, Cas 단백질-암호화하는 핵산 또는 Cas 단백질은 상기 기술한 바와 같다.
- [0115] 본원에서 사용된, 용어 "유전형질 분석 (genotyping)"은 "제한 단편 길이 다형성(RLFP) 어세이"를 지칭한다.
- [0116] RLFP는 1) 세포 또는 유기체에서 유전자 가위에 의해 유도된 인델 (indel)의 탐지, 2) 세포 또는 유기체에서 자연 발생 돌연변이 또는 변이의 유전형질 분석, 또는 3) 바이러스 또는 박테리아 등을 포함하는 감염된 병원성 미생물의 DNA의 유전형질 분석에 사용될 수 있다.
- [0117] 돌연변이 또는 변이는 유전자 가위에 의해 세포에 유도될 수 있다.
- [0118] 유전자 가위는 징크 핑거 뉴클레아제 (Zinc Finger Nuclease, ZFNs), 전사 활성화자-유사 반응기 뉴클레아제 (Transcription Activator-Like Effector Nucleases, TALENs), 또는 RGENs일 수 있지만, 이에 제한되는 것은 아니다.
- [0119] 본원에서 사용된 용어 "생물학적 시료"는 조직, 세포, 전혈, Semm, 혈장, 타액, 객담, 뇌척수액 또는 소변과 같은 분석을 위한 시료를 포함하지만, 이에 제한되는 것은 아니다.
- [0120] 돌연변이 또는 변이는 자연 발생 돌연변이 또는 변이일 수 있다.
- [0121] 돌연변이 또는 변이는 병원성 미생물에 의해 유도될 수 있다. 다시 말해, 병원성 미생물이 탐지되고, 생물학적 시료가 감염된 것이라고 판명될 때, 돌연변이 또는 변이는 병원성 미생물의 감염으로 인하여 발생한다.
- [0122] 병원성 미생물은 바이러스 또는 박테리아일 수 있지만, 이에 제한되는 것은 아니다.
- [0123] 유전자 가위-유도 돌연변이는 불일치-민감 Surveyor (mismatch-sensitive Surveyor) 또는 T7 엔도뉴클레아제 I (T7E1) 어세이, RFLP 분석, 형광 PCR, DNA 멜팅 (melting) 분석, 및 Sanger 및 deep 시퀀싱을 포함하는 다양한 방법에 의해 검출된다. T7E1 및 Surveyor 어세이는 널리 사용되지만, 이형이중가닥 (heteroduplexes) (돌연변이와 야생형 서열 또는 두 개의 다른 돌연변이 서열의 혼성화에 의해 형성됨)을 탐지하기 때문에 종종 돌연변이 빈도를 감소한다; 상기 어세이는 동일한 돌연변이 서열의 혼성화에 의해 형성된 동형이중가닥은 탐지하지 못한다. 그러므로, 이러한 어세이는 야생형 세포에서 동형접합 이중대립유전자 돌연변이 클론 (homozygous biallelic mutant clone)과 이형접합 단일대립유전자 돌연변이체 (heterozygous monoallelic mutant)로부터 이형접합 이중대립유전자 돌연변이체 (heterozygous biallelic mutant) 중 어느 것도 구별하지 못한다 (도 22). 또한, 상기 효소가 이러한 서로 다른 야생형 대립 유전자의 혼성화에 의해 형성되는 이형이중가닥을 절단할 수 있기 때문에, 뉴클레아제 표적 서열 근처 서열 다형성(sequence polymorphism)은 혼란스러운 결과를 생산할 수 있다.

RFLP 분석은 상기 한계가 없어서, 선택의 한 방법이다. 정말로, RFLP 분석은 유전자 가위-매개 돌연변이를 탐지하기 위해 사용되는 첫 번째 방법 중에 하나다. 하지만, 불행히도 적절한 제한효소 위치의 유용성이 제한되어 있다.

[0124] 발명의 다른 측면에 따르면, 본 발명은 분리된 생물학적 시료에서 돌연변이 또는 변이(variation)를 유전형질 분석(genotyping) 하기 위한 조성물을 포함하는, 분리된 생물학적 시료에서 돌연변이 또는 변이(variation)를 유전형질 분석(genotyping) 하기 위한 키트를 제공한다. 또한, 본 발명은 표적 DNA 서열에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는, 분리된 생물학적 시료에서 병원성 미생물의 핵산 서열을 유전형질 분석하기 위한 키트를 제공한다.

[0125] 가이드 RNA, Cas 단백질을 암호화하는 핵산 또는 Cas 단백질은 상기에서 기술한 바와 같다.

[0126] 발명의 다른 측면에 따르면, 본 발명은 분리된 생물학적 시료에서 돌연변이 또는 변이를 유전형질 분석하기 위한 조성물을 사용하여 분리된 생물학적 시료에서 돌연변이 또는 변이를 유전형질 분석하는 방법을 제공한다. 또한, 본 발명은 표적 DNA 서열에 특이적인 가이드 RNA 및 Cas 단백질을 포함하는, 분리된 생물학적 시료에서 병원성 미생물의 핵산 서열을 유전형질 분석하는 방법을 제공한다.

[0127] 가이드 RNA, Cas 단백질을 암호화하는 핵산 또는 Cas 단백질은 상기에서 기술한 바와 같다.

[0128] **실시예**

[0129] 이하, 본 발명은 실시예를 참고하여 보다 상세히 기술될 것이다. 그러나, 이들 실시예는 단지 예시적인 목적이며, 본 발명을 이들 실시예에 의해 제한하고자 하는 의도가 아니다.

[0130] **실시예 1: 유전체 교정 어세이**

[0131] **1-1. Cas9 단백질의 DNA 절단 활성**

[0132] 먼저, 키메라 가이드 RNA (chimeric guide RNA)의 존재 또는 부재 상태에서 스트렙토코커스 피요젠스 (*Streptococcus pyogenes*)로부터 유래된 Cas9 단백질의 DNA 절단 활성을 인 비트로에서 시험하였다.

[0133] 이를 위해, 대장균에서 발현하고 정제한 재조합 Cas9 단백질을 사용하여 23-염기쌍(bp)의 인간 CCR5 표적 서열을 포함하는, 평이한 형태 (predigested) 또는 원형 플라스미드 DNA를 절단하였다. Cas9 표적 서열은 crRNA 또는 키메라 가이드 RNA에 상보적인 20bp DNA 서열 및 Cas9 자체에 의해 인식되는 트리뉴클레오타이드 (trinucleotide) (5'-NGG-3') 프로토스페이스 인접 모티프 (protospacer adjacent motif, PAM)로 구성되어 있다 (도 1A).

[0134] 구체적으로, 스트렙토코커스 피요젠스 균주 M1 GAS (NC_002737.1)에서 유래한, Cas9-암호화 서열 (4,104bp)을 인간 코돈 사용표를 이용하여 재구성하였고, 올리고뉴클레오타이드를 이용하여 합성하였다. 먼저, 중복되는 ~35-머 올리고뉴클레오타이드 (overlapping 35-mer oligonucleotide) 및 Phusion 폴리머라제 (New England Biolabs)를 이용하여 1-kb DNA 단편을 조립하였고, T-벡터 내로 클로닝하였다 (SolGent). 전장 Cas9 서열 (full-length Cas9 sequence)을 네 개의 1-kbp DNA 단편을 이용하여 중복 PCR (overlap PCR)로 조립하였다. Cas9-암호화 DNA 단편을 pcDNA3.1에서 유래한 p3s (Invitrogen)에 서브클로닝하였다. 상기 벡터에서 HA 항원결

정부위 및 핵 위치 신호 (nuclear localization signal, NLS)를 포함하는 펩타이드 태그 (NH₂-GGSGPPKKRKYVPYDVPDYA-COOH, 서열번호: 2)를 Cas9의 C-말단에 덧붙였다. HEK 293T 세포에서 Cas9 단백질의 발현 및 핵 위치를 항-HA 항체 (Santa Cruz)를 사용한 웨스턴 블롯팅 (western blotting)으로 확인하였다.

[0135] 그리고, Cas9 카세트를 pET28-b(+)에 서브클로닝하였고, BL21(DE)에 형질전환하였다. 25°C에서 4시간 동안 0.5 mM IPTG를 이용하여 Cas9의 발현을 유도하였다. C-말단에 His-태그를 포함하는 Cas9 단백질을 Ni-NTA 아가로스 레진 (Qiagen)을 이용하여 정제하였고, 20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM DTT, 및 10% 글리세롤 (1)로 투석하였다. 정제된 Cas9 (50 nM)을 초나선 (super-coiled) 또는 평이한 (pre-digested) 플라스미드 DNA (300 ng) 및 키메라 RNA (50 nM)와 함께 37°C에서 1시간 동안 NEB 버퍼 3의 20 µl의 반응 부피에서 반응시켰다. 절단된 DNA를 0.8% 아가로스 젤을 이용한 전기영동으로 분리하였다.

[0136] Cas9는 합성 RNA가 존재할 때만 예상된 위치에서 플라스미드 DNA를 효율적으로 절단하였고, 표적 서열이 결여된 대조군 플라스미드는 절단하지 않았다 (도 1B).

[0137] **1-2. 인간 세포에서 Cas9/가이드 RNA 복합체에 의한 DNA 절단**

[0138] RFP-GFP 리포터를 사용하여 포유동물 세포에서 RFP 및 GFP 서열 사이에 삽입된 표적 서열을 Cas9/가이드 RNA 복합체가 절단할 수 있는지를 조사하였다.

[0139] 이 리포터에서, GFP 서열을 out-of-frame으로 RFP 서열에 융합하였다 (2). 표적 서열이 위치-특이적 뉴클레아제에 의해 절단되었을 때만, 활성 GFP가 발현되었고, 이것은 이중 나선 절단 (double strand break, DSB)의 오류 유발 비-상동 말단-결합 (non-homologous end-joining, NHEJ) 수선을 통해 표적 서열 주변의 프레임 이동 작은 삽입 또는 결실(indels)을 야기한다 (도 2).

[0140] 본 발명에서 사용된 RFP-GFP 리포터 플라스미드는 이전 (2)에 기술되어 있는 바와 같이 구성하였다. 표적 위치에 상응하는 올리고뉴클레오타이드 (표 1)를 합성하였고 (Macrogen), 어닐링 (annealing)하였다. 어닐링된 올리고뉴클레오타이드는 EcoRI 및 BamHI 으로 절단된 리포터 벡터에 연결하였다.

[0141] 24-웰 플레이트에서 리포펙타민 2000 (Invitrogen)을 이용하여 HEK 293T 세포에 Cas9-암호화 플라스미드 (0.8 µg) 및 RFP-GFP 리포터 플라스미드 (0.2 µg)를 공동-형질주입하였다.

[0142] 한편, 인 비트로에서 전사된 키메라 RNA는 다음과 같이 준비하였다. RNA는 제조자의 매뉴얼에 따라 MEGAscript T7 키트 (Ambion)를 이용하여 run-off 반응을 통해 인 비트로 전사하였다. RNA 인 비트로 전사를 위한 주형은 두 상보적인 단일 가닥 DNA의 어닐링 또는 PCR 증폭으로 생성하였다 (표 1). 전사된 RNA를 8% 변성 urea-PAGE 젤에서 분리하였다. RNA를 포함하는 젤 단편을 잘라내었고, 프로브 용출 버퍼 (probe elution buffer)에 옮겼다. RNA를 뉴클레아제가 없는 물 (nuclease-free water)에서 회수한 다음에, 페놀:클로로포름 추출, 클로로포름 추출 및 에탄올 침전하였다. 정제된 RNAs를 분광계로 정량하였다.

[0143] 형질주입 12시간 후, 인 비트로 전사로 제조한 키메라 RNA (1 µg)를 리포펙타민 2000을 이용하여 형질주입하였다.

[0144] 형질주입 3일 후, 형질주입된 세포를 유세포 분석기에 적용하고, RFP 및 GRP 모두를 발현하는 세포의 수를 계수하였다.

[0145] Cas9 플라스미드를 먼저 형질주입하고, 그 다음 12시간 후에 가이드 RNA를 형질주입하였을 때만, GFP-발현 세포를 수득하였음을 발견하였고 (도 2), 이는 RGEN이 배양된 인간 세포에서 표적 DNA 서열을 인식 및 절단할 수 있다는 것을 의미한다. 이에, GFP-발현 세포는 공동-형질주입보다 Cas9 플라스미드 및 가이드 RNA의 단계적-형질주입에 의해 얻을 수 있었다.

표 1

[0146]

유전자		서열 (5' to 3')	서열번호
리포터 플라스미드의 제작에 사용한 올리고뉴클레오타이드			
CCR5	F	AATTCATGACATCAATTATTATACATCGGAGGAG	3
	R	GATCCTCCTCCGATGTATAATAATTGATGTCATG	4
T7E1 어세이에 사용한 프라이머			

CCR5	F1	CTCCATGGTGTATAGAGCA	5
	F2	GAGCCAAGCTCTCCATCTAGT	6
	R	GCCCTGTCAAGAGTTGACAC	7
C4BPB	F1	TATTTGGCTGGTTGAAAGGG	8
	R1	AAAGTCATGAAATAAACACACCCA	9
	F2	CTGCATTGATATGGTAGTACCATG	10
	R2	GCTGTTCATTGCAATGGAATG	11
오프-타겟 사이트의 증폭에 사용한 프라이머			
ADCY5	F1	GCTCCACCTTAGTGCTCTG	12
	R1	GGTGGCAGGAACCTGTATGT	13
	F2	GTCATTGGCCAGAGATGTGGA	14
	R2	GTCCCATGACAGGCGTGTAT	15
KCNJ6	F	GCCTGGCCAAGTTTCAGTTA	16
	R1	TGGAGCCATTGGTTTGCATC	17
	R2	CCAGAACTAAGCCGTTTCTGAC	18
CNTNAP2	F1	ATCACCGACAACCAGTTTCC	19
	F2	TGCAGTGCAGACTCTTTCCA	20
	R	AAGGACACAGGGCAACTGAA	21
N/A Chr. 5	F1	TGTGGAACGAGTGGTGACAG	22
	R1	GCTGGATTAGGAGGCAGGATTC	23
	F2	GTGCTGAGAACGCTTCATAGAG	24
	R2	GGACCAAACCACATTCTCTCAC	25
염색체 결실의 탐지에 사용한 프라이머			
결실	F	CCACATCTCGTTCGCGTTT	26
	R	TCACAAGCCACAGATATTT	27

[0147] 실시예 1-3. 포유동물 세포에서 RGEN에 의한 내재적 유전자의 표적화된 분해

[0148] RGENs이 포유동물의 내재적 유전자의 표적화된 분해에 사용될 수 있는지 여부를 테스트하기 위해, T7 엔도뉴클레아제 1 (T7E1), 야생형 및 돌연변이 DNA 서열의 혼성화에 의해 형성된 이형이중가닥(heteroduplex)을 특이적으로 인지 및 절단하는 불일치-민감 엔도뉴클레아제 (mismatch-sensitive endonuclease)를 사용하여형질주입된 세포로부터 분리된 유전체 DNA에 대해 분석하였다 (3).

[0149] RGENs을 이용하여 포유동물의 세포에 DSBs를 도입하기 위해, 2×10^6 K562 세포를 제조자의 프로토콜에 따라 4D-Nucleofector, SF Cell Line 4D-Nucleofector X Kit, Program FF-120 (Lonza)를 이용하여 Cas9-암호화 플라스미드 20 μg 을 형질주입하였다. 본 실험을 위해, K562 (ATCC, CCL-243) 세포를 10% FBS 및 페니실린/스트렙토마이신 혼합액 (각각 100 U/ml 및 100 $\mu\text{g}/\text{ml}$)을 첨가한 RPMI-1640 배지에서 배양하였다.

[0150] 24시간 후, 인 비트로에서 전사한 키메라 RNA의 10 - 40 μg 을 1×10^6 K562 세포에 핵 내로 도입하였다. 인 비트로 전사된 키메라 RNA는 실시예 1-2에 따라 제조하였다.

[0151] RNA 형질주입 이틀 후, 세포를 모아서 유전체 DNA를 분리하였다. 표적 위치가 포함된 부위를 표 1에 명시된 프라이머를 이용하여 PCR-증폭하였다. (3)에 기술된 바와 같이 T7E1 어세이에 앰플리콘 (amplicon)을 적용하였다. 서열 분석을 위해 유전체 변형에 상응하는 PCR 산물을 정제하고, T-Blunt PCR 클로닝 키트 (SolGent)를 이용하여 T-Blunt 벡터에 클로닝하였다. 클로닝된 산물을 M13 프라이머를 이용하여 서열 분석하였다.

[0152] 세포에 단계적으로 Cas9-암호화 플라스미드를 형질주입하고, 그 다음 가이드 RNA를 형질 주입하였을 때만, 돌연변이가 유도된다는 것을 확인하였다 (도 3). 상대적인 DNA 밴드의 강도로 추산된 돌연변이 빈도 (도 3A의 Indels (%))는 RNA-용량 의존적이었고, 그 범위는 1.3%에서 5.1%이었다. PCR 앰플리콘의 DNA 서열 분석으로 내재적 위치에서 RGEN-매개 돌연변이의 유도임을 확증하였다. 오류 유발 NHEJ의 특징인 Indels 및 마이크로상동

(microhomology)가 표적 위치에서 관찰되었다. 다이렉트 시퀀싱 (direct sequencing)으로 측정된 돌연변이 빈도는 7.3% (= 7 돌연변이 클론 / 96 클론)이었고, 이는 징크 핑거 뉴클레아제 (zinc finger nucleases, ZFNs) 또는 전사 활성화자-유사 반응기 뉴클레아제 (transcription-activator-like effector nucleases, TALENs)에서 얻은 빈도와 비슷하였다.

[0153] Cas9 플라스미드 및 가이드 RNA의 단계적 형질주입 (serial-transfection)은 세포에서 돌연변이를 유도하는데 필요하였다. 그러나, 가이드 RNA를 암호화하는 플라스미드일 때, 단계적 형질주입은 필요하지 않고, Cas9 플라스미드 및 가이드 RNA-암호화 플라스미드로 공동-형질주입하였다.

[0154] 한편, ZFNs 및 TALENs 둘 모두는, HIV 감염에 필수적인 공동 수용체인 G-단백질 연관 케모카인 수용체 (G-protein coupled chemokine receptor)를 암호화하는 인간 CCR5 유전자를 파괴하기 위한 것으로 성공적으로 고안되었다 (3-6). 현재, CCR5-특이적 ZFN은 미국에서 AIDS의 치료를 위한 임상 시험 중이다 (7). 그러나, 이러한 ZFNs 및 TALENs는, 서열이 온-타겟 서열에 상동성을 갖는 위치에서의 로컬 돌연 변이 (6, 8-10) 및 온-타겟 및 오프-타겟 위치에서 유발된 두 개의 동시 (concurrent) DSBs 수선으로부터 발생한 유전체 재배열 (11-12)을 모두 유발하는 오프-타겟 효과를 가진다. 이러한 CCR5-특이적 유전자 가위와 관련된 가장 현저한 오프-타겟 위치는, CCR5의 15-kbp 업스트림 (upstream)에 위치한 CCR5의 가까운 동족체 (close homolog of CCR5)인 CCR2 좌위에 위치한다. CCR2 유전자에서 오프-타겟 돌연변이를 피하고, CCR5 온-타겟과 CCR2 오프-타겟 위치 사이의 15-kbp 염색체 부분 (chromosomal segment)의 원치 않는 결실 (deletion), 반전 (inversion), 및 중복 (duplication)을 피하기 위해, 본 발명자들은 의도적으로 CCR2 서열과 명백한 상동성을 갖지 않는 CCR 5 서열 내의 부위를 인지하는 우리의 CCR5-특이적 RGEN의 표적 위치를 선택하였다.

[0155] 본 발명자들은 CCR5-특이적 RGEN이 오프-타겟 효과를 갖는지 여부를 조사하였다. 이를 위해, 본 발명자들은 의도된 23-bp 타겟 서열과 가장 상동성이 높은 위치를 알아냄으로써 인간 유전체에서 잠재적 오프-타겟 위치를 조사하였다. 예상한 대로, CCR2 유전자에서는 그러한 위치가 발견되지 않았다. 대신에, 각 위치가 온-타겟 위치에서 3-염기 불일치 (3-base mismatches)를 갖는 네 개의 위치를 발견하였다(도 4A). T7E1 어세이는 이러한 위치에서 돌연변이를 감지하지 않았고 (어세이 감도, ~0.5%), 이는 RGENs의 정교한 특이성을 나타낸다 (도 4B). 또한, PCR을 사용하여 CCR5에 특이적인 ZFN 및 RGEN을 암호화하는 플라스미드를 각기 형질주입한 세포에서 염색체 결실의 유도를 감지하였다. ZFN은 결실을 유도한 반면, RGEN은 결실을 유도하지 않았다 (도 4C).

[0156] 그 다음, CCR5-특이적 가이드 RNA를, 전사인자인 C4b-결합 단백질의 베타 사슬을 암호화하는 인간 C4BPB 유전자를 표적화하도록 설계한 새로 합성한 CCR5-특이적 가이드 RNA로 대체하여 RGEN을 리프로그래밍하였다. 상기 RGEN은 K562 세포의 염색체 표적 위치에서 높은 빈도로 돌연변이를 유도하였다 (도 3B). T7E1 어세이 및 다이렉트 시퀀싱에 의해 측정된 돌연변이 빈도는 각각 14% 및 8.3% (= 4 돌연변이 클론 / 48 클론)이었다. 네 개의 돌연변이 서열 중, 두 개의 클론은 CCR5 표적 위치에서 관찰되는 패턴인 절단 위치에 하나의 염기 또는 두 개의 염기 삽입을 정확하게 포함하였다. 상기 결과는 RGENs이 세포의 예상된 위치에서 염색체 표적 DNA를 절단한다는 것을 의미한다.

[0157] **실시에 2: 단백질성 RGEN-매개 유전체 교정 (proteinaceous RGEN-mediated genome editing)**

[0158] RGENs은 많은 다른 형태로 세포 안에 전달될 수 있다. RGENs은 Cas9 단백질, crRNA 및 tracrRNA로 구성된다. 상기 두 RNA는 단일사슬 가이드 RNA (sgRNA)를 형성하기 위해 융합될 수 있다. CMV 또는 CAG와 같은 프로모터 하에서 Cas9 단백질을 암호화하는 플라스미드는 세포 안으로 형질주입될 수 있다. crRNA, tracrRNA, 또는 sgRNA는 상기 RNA들을 암호화하고 있는 플라스미드를 이용하여 세포 안에서 또한 발현될 수 있다. 그러나 플라스미드의 사용은 때때로 숙주의 유전체 안에서 전체 또는 일부분의 플라스미드가 통합되는 결과를 낳는다. 플라스미드 DNA에 통합된 박테리아 서열은 인 비보 (*in vivo*)에서 원치 않는 면역반응을 야기할 수 있다. 세포 치료를 위한 플라스미드가 형질주입된 세포 또는 DNA-형질주입된 세포로부터 유래한 동물 및 식물은 대부분 선진국의 시장 승인 전에, 고가이며 오랜 규제 절차를 통과해야만 한다. 또한, 플라스미드 DNA는 형질 주입 후 며칠 동안 세포 내에 지속할 수 있어서, RGEN의 오프-타겟 효과를 악화시킬 수 있다.

- [0159] 여기에서, 본 발명자들은 인 비트로 전사된 가이드 RNA와 복합체를 형성한 재조합 Cas9 단백질을 사용하여 인간 세포에서 내재적 유전자의 표적화된 파괴 (targeted disruption)를 유도하였다. 헥사-히스티딘 (hexahistidine) 태그와 융합된 재조합 Cas9 단백질을 대장균에서 발현하고, 표준 Ni 이온 친화성 크로마토그래피 및 겔 여과 (gel filtration)를 이용하여 정제하였다. 정제된 재조합 Cas9 단백질을 저장 버퍼 (20 mM HEPES pH 7.5, 150 mM KCl, 1 mM DTT, 및 10% 글리세롤)에서 농축하였다. Cas9 단백질/sgRNA 복합체를 뉴클레오펙션 (nucleofection)으로 K562 세포로 직접적으로 도입하였다: 100 μ l 용액에서 인 비트로 전사된 sgRNA 100 μ g (또는 crRNA 40 μ g 및 tracrRNA 80 μ g)과 혼합된, 22.5-225 (1.4-14 μ M)의 Cas9 단백질 혼합물을 1×10^6 K562 세포에 제조자의 프로토콜에 따라 4D-Nucleofector, SF Cell Line 4D-Nucleofector X Kit, 프로그램 FF-120 (Lonza)를 이용하여 형질주입하였다. 뉴클레오펙션 후, 6-웰 플레이트에서 성장 배지에 세포를 위치하도록 하고, 48시간 동안 배양하였다. 2×10^5 K562 세포를 1/5로 규모가 다운된 프로토콜로 형질주입하였을 때, 6 내지 60 μ g의 인 비트로 전사된 sgRNA (또는 crRNA 8 μ g 및 tracrRNA 16 μ g)과 혼합된, 4.5-45 μ g의 Cas9 단백질을 사용하여 20 μ l 용액에서 뉴클레오펙션하였다. 이후, 뉴클레오펙션된 세포를 48-웰 플레이트에서 성장 배지에 두었다. 48시간 후, 세포를 모으고 유전체 DNA를 분리하였다. 표적 위치에 걸친 (spanning) 유전체 DNA 부분을 PCR로 증폭하였고, T7E1 어세이에 적용하였다.
- [0160] 도 10에서 볼 수 있듯이, Cas9 단백질/sgRNA 복합체는 sgRNA 또는 Cas9 단백질의 용량-의존적인 방식으로 CCR5 좌위에서 4.8 내지 38% 범위의 빈도로 표적화된 돌연변이를 유도하였고, 이는 Cas9 플라스미드 형질주입에서 얻은 빈도 (45%)와 같았다. Cas9 단백질/crRNA/tracrRNA 복합체는 9.4%의 빈도로 돌연변이를 유도할 수 있었다. Cas9 단백질 단독은 돌연변이를 유도하지 못했다. 2×10^5 K562 세포에 1/5로 규모가 다운된 용량으로 Cas9 단백질 및 sgRNA를 형질주입하였을 때, CCR5 좌위에서의 돌연변이 빈도는 용량-의존적인 방식으로 2.7 내지 57% 범위였고, 이는 Cas9 플라스미드 및 sgRNA 플라스미드의 공동-형질주입으로 얻은 빈도 (32%)보다 더 높았다.
- [0161] 본 발명자들은 또한, ABCC11 유전자를 표적하는 Cas9 단백질/sgRNA 복합체를 시험하였고, 상기 복합체는 35%의 빈도로 인델 (indel)을 유도하여, 이 방법의 일반 공용성을 나타내었다.

표 2

가이드 RNA의 서열

표적	RNA 타입	RNA 서열 (5' 에서 3')	길이	서열번호
CCR5	sgRNA	GGUGACAUCAAUUUUAUACAUGUUUAGAGCUAGAAAUAAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU	104bp	28
	crRNA	GGUGACAUCAAUUUUAUACAUGUUUAGAGCUAUGCUGUUUUG	44bp	29
	tracrRNA	GGAACCAUUCAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUUUU	86bp	30

[0163] **실시예 3: 마우스에서의 RNA-가이드 유전체 교정 (RNA-guided genome editing in mouse)**

- [0164] 전핵 단계 (pronuclear (PN)-stage)의 마우스 배아에서 RGENs의 유전자-타겟팅 능력을 알아보기 위해, 흥선 발달 및 케라틴 세포 분화에 중요한 forkhead box N1 (Foxn1) 유전자 (Nehls et al., 1996), 및 DNA DSB 수선 및 재조합에 중요한 효소를 암호화하는 the protein kinase, DNA activated, catalytic polypeptide (Prkdc) 유전자 (Taccioli et al., 1998)를 사용하였다.
- [0165] Foxn1-RGEN의 유전체 교정 활성을 평가하기 위하여, 본 발명자들은 다양한 농도의 sgRNA (도 5a)와 함께 Cas9 mRNA (10 ng/ μ l용액)를 PN-단계 마우스 배아의 세포질에 주입하였고, 인 비트로 배양한 배아에서 얻은 유전체 DNA를 이용하여 T7 엔도뉴클레아제 I (T7E1) 어세이 (Kim et al. 2009)를 수행하였다 (도 6a).
- [0166] 다른 방법으로, 본 발명자들은 두 배 초과한 몰수의 Foxn1-특이적 sgRNA (0.14 내지 14 ng/ μ l)와 복합체를 형성한 재조합 Cas9 단백질 (0.3 내지 30 ng/ μ l)의 형태로 RGEN을 1 세포 마우스 배아의 세포질 또는 전핵에 바로 주입하였고, 인 비트로 배양된 배아를 이용하여 Foxn1 유전자의 돌연변이를 분석하였다 (도 7).

[0167] 특히, Cas9 mRNA 및 sgRNA를 각각 mMMESSAGE mMACHINE T7 울트라 키트 (Ambion) 및 MEGAscript T7 키트 (Ambion)를 이용하여 선형 DNA 주형으로부터 제조자의 지시에 따라 인 비트로 합성하였고, 적당한 양의 디에틸 피로카보네이트 (DEPC, Sigma)-처리된 주입 버퍼 (0.25 mM EDTA, 10 mM Tris, pH 7.4)에 희석하였다. sgRNA 합성의 주형은 표 3에 나열된 올리고뉴클레오타이드를 이용하여 생성하였다. 재조합 Cas9 단백질은 ToolGen, Inc.에서 획득하였다.

표 3

RNA 이름	방향 (Direction)	서열 (5' 에서 3')	서열번호
Foxn1 #1 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>CAGTCTGACGTCACACTTCCG</i> <u>TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	31
Foxn1 #2 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>ACTTCCAGGCTCCACCCGACG</i> <u>TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	32
Foxn1 #3 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>CCAGGCTCCACCCGACTGGAG</i> <u>TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	33
Foxn1 #4 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>ACTGGAGGGCGAACCCCAAGG</i> <u>TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	34
Foxn1 #5 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>ACCCCAAGGGACCTCATGCG</i> <u>TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	35
Prkdc #1 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>TTAGTTTTTCCAGAGACTTGT</i> <u>TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	36
Prkdc #2 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>TTGGTTTGCTGTGTTTATCG</i> <u>TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	37
Prkdc #3 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>CACAAGCAAACCAAGTCTG</i> <u>TTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	38
Prkdc #4 sgRNA	F	<u>GAAATTAATACGACTCACTATAGG</u> <i>CCTCAATGCCTAAGCGACTT</i> <u>CGTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG</u>	39

[0169] 모든 동물 실험은 한국식품의약품안전처 (KFDA)의 기준에 따라 수행하였다. 프로토콜은 연세대학교 실험동물연구센터의 기관동물보호 및 사용위원회 (IACUC)에 의해 검토받고 승인받았다 (승인번호: 2013-0099). 모든 마우스는 연세 실험동물 연구센터의 특정병원체-부재 시설 (specific pathogen-free facility)에서 유지하였다. FVB/NTac (Taconic) 및 ICR 마우스 종을 각각 배아 기증자 (embryo donor) 및 위탁모 (foster mother)로 사용하였다. 48시간 간격으로 암컷 FVB/NTac 마우스 (7 내지 8주령)에 5 IU 임신 암컷 말 혈청 생식선자극호르몬 (PMSG, Sigma) 및 5 IU 인간 융모성 생식선 자극 호르몬 (hCG, Sigma)을 복강 내 주입하여 과배란하였다. 과배란된 암컷 마우스를 FVB/NTac 스타드 (stud) 수컷과 교배하였고, 난관 (oviduct)으로부터 수정란을 수집하였다.

[0170] Piezo-driven 미세조작기 (Prime Tech)를 이용하여 M2 배지의 Cas9 mRNA 및 sgRNA (Sigma)를 잘 알려진 전핵 (well-known pronuclei)을 가지는 수정란의 세포질에 주입하였다.

[0171] 재조합 Cas9 단백질 주입의 경우, 재조합 Cas9 단백질:Foxn1-sgRNA 복합체를 DEPC-처리된 주입 버퍼 (0.25 mM EDTA, 10 mM Tris, pH 7.4)에 희석하였고, TransferMan NK2 미세조작기 및 FemtoJet 미세주입기 (Eppendorf)를 이용하여 수컷 전핵에 주입하였다.

[0172] 조작된 배아는 가임신 위탁모의 난관에 이식하여 살아있는 동물을 생산하거나, 또는 추가의 분석을 위해 인 비트로에서 배양하였다.

[0173] RGEN-유도 돌연변이를 갖는 F0 마우스 및 인 비트로 배양된 마우스 배아를 스크리닝하기 위해, 꼬리 바이옴시 (biopsy) 및 전체 배아의 용해물로부터 얻은 유전체 DNA 시료를 이용하여 이전에 기술된 바와 같이 (Cho et al., 2013), T7E1 어세이를 수행하였다.

[0174] 요약하면, RGEN 표적 부위를 포함하는 유전체 부분을 PCR-증폭 (PCR-amplified), 용융 (melted), 및 재-어닐링 (re-annealed)하여, T7 엔도뉴클레아제 I (New England Biolabs)으로 처리된, 이중이중가닥 DNA (heteroduplex DNA)를 형성한 다음, 아가로스 젤 전기영동으로 분리하였다. 잠재적 오프-타겟 위치를 bowtie 0.12.9로 검색하여 규명하였고, 또한 T7E1 어세이로 유사하게 모니터링하였다. 상기 어세이에 사용한 프라이머 쌍을 표 4 및 표 5에 나열하였다.

표 4

T7E1 어세이에 사용한 프라이머

유전자	방향 (Direction)	서열 (5' 에서 3')	서열번호
Foxn1	F1	GTCTGTCTATCATCTCTCCCTTCTCTCC	40
	F2	TCCCTAATCCGATGGCTAGCTCCAG	41
	R1	ACGAGCAGCTGAAGTTAGCATGC	42
	R2	CTACTCAATGCTCTTAGAGCTACCAGGCTTGC	43
Prkdc	F	GACTGTTGTGGGAGGGCCG	44
	F2	GGGAGGGCCGAAAGTCTTATTTTG	45
	R1	CCTGAAGACTGAAGTTGCCAGAAGTGAG	46
	R2	CTTTAGGGCTTCTCTCTACAATCAG	47

표 5

오프-타겟 위치의 증폭에 사용된 프라이머

유전자	표기 (Notation)	방향 (Direction)	서열(5'에서 3')	서열번호
Foxn1	off 1	F	CTCGGTGTGTAGCCCTGAC	48
		R	AGACTGGCCTGGAATCACAG	49
	off 2	F	CACTAAAGCCTGTCAGGAAGCCG	50
		R	CTGTGGAGAGCACACAGCAGC	51
	off 3	F	GCTGCGACCTGAGACCATG	52
		R	CTTCAATGGCTTCCTGCTTAGGCTAC	53
	off 4	F	GGTTCAGATGAGGCCATCCTTC	54
		R	CCTGATCTGCAGGCTTAACCCCTTG	55
Prkdc	off 1	F	CTCACCTGCACATCACATGTGG	56
		R	GGCATCCACCCTATGGGGTC	57
	off 2	F	GCCTTGACCTAGAGCTTAAAGAGCC	58
		R	GGTCTTGTAGCAGGAAGGACACTG	59
	off 3	F	AAAACCTCTGCTTGATGGGATATGTGGG	60
		R	CTCTACTGGTTATCTGTGCTCCTTC	61
	off 4	F	GGATCAATAGGTGTGGGGATG	62
		R	GTGAATGACACAATGTGACAGCTTCAG	63
	off 5	F	CACAAGACAGACCTCTCAACATTCAGTC	64
		R	GTGCATGCATATAATCCATTCTGATTGCTCTC	65
	off 6	F1	GGGAGGCAGAGGCAGGT	66
		F2	GGATCTCTGTGAGTTGAGGCCA	67
R1		GCTCCAGAACTACTCTTAGGCTC	68	

[0177]

T7E1 어세이로 밝힌 돌연변이 파운더 (founder)를 fPCR로 추가적으로 분석하였다. 유전체 DNA의 적절한 부위를 이전에 기술된 바에 따라 서열을 분석하였다 (Sung et al., 2013). F1 자손을 위한 루틴(routine) PCR 유전형질 분석의 경우, 야생형 및 돌연변이 대립유전자 모두에 대하여 다음의 프라이머 쌍을 사용하였다:

[0178]

Foxn1 유전자를 위한 5'-CTACTCCCTCCGAGTCTGA-3' (서열번호 69) 및 5'-CCAGGCCTAGGTTCCAGGTA-3' (서열번호 70),

[0179]

Prkdc 유전자를 위한 5'-CCCCAGCATTGCAGATTCC-3' (서열번호 71) 및 5'-AGGGCTTCTCTCTACAATCAG-3' (서열번호 72).

[0180]

Cas9 mRNA 주입의 경우, 돌연변이 비율 (돌연변이 배아의 수 / 전체 배아의 수)은 용량-의존적이었고, 범위는

33% (1 ng/ μ l sgRNA) 내지 91% (100 ng/ μ l)이었다 (도 6b). 서열 분석으로 Foxn1 유전자의 돌연변이를 확인하였다; 대부분의 돌연변이는 ZFNs 및 TALENs에 의한 것이라는 것을 암시하는 (Kim et al., 2013), 작은 결실이었다 (도 6c).

[0181] Cas9 단백질 주입의 경우, 주입 용량 및 방법은 인 비트로에서의 마우스 배아의 생존 및 발달에 최소한의 영향을 미쳤다: RGEN-주입 배아의 70% 이상이 모든 실험에서 정상적으로 부화하였다. 다시, Cas9 단백질 주입으로 얻어진 돌연변이 비율은 용량 의존적이었으며, 전핵 주입을 통한 가장 높은 용량에서 88%까지 달했고, 세포질 내 주입을 통해서 71%까지 도달했다 (도 7a 및 7b). sgRNA 더하기 Cas9 mRNA에 의해 유도된 돌연변이 패턴과 비슷하게 (도 6c), Cas9 단백질-sgRNA 복합체에 의해 유도된 상기 돌연변이는 대부분 작은 결실이었다 (도 7c). 상기 결과는 마우스 배아에서 RGENs이 높은 유전자-타겟팅 활성을 갖는다는 것을 분명히 보여준다.

[0182] RGENs에 의해 유도된 높은 돌연변이 빈도와 낮은 세포독성에 힘입어, 본 발명자들은 가임신 위탁모의 난관에 마우스 배아를 이식함으로써 살아있는 동물을 생산하였다.

[0183] 특히, 출생 비율은 58% 내지 73%의 범위로 매우 높았고, Foxn1-sgRNA의 증가하는 용량에도 영향을 받지 않았다 (표 6).

표 6

[0184] FVB/NTac 마우스에서 RGEN-매개 유전자 타겟팅

표적 유전자	Cas9 mRNA + sgRNA (ng/ μ l)	주입된 배아 (Injected embryos)	이식된 배아 (Transferred embryos) (%)	전체 새로 태어난 마우스 (Total newborns) (%)	살아있는 새로 태어난 마우스* (Live newborns*) (%)	파운더† (Founders†) (%)
Foxn1	10 + 1	76	62 (82)	45 (73)	31 (50)	12 (39)
	10 + 10	104	90 (87)	52 (58)	58 (64)	33 (57)
	10 + 100	100	90 (90)	62 (69)	58 (64)	54 (93)
	Total	280	242 (86)	159 (66)	147 (61)	99 (67)
Prkdc	50 + 50	73	58 (79)	35 (60)	33 (57)	11 (33)
	50 + 100	79	59 (75)	22 (37)	21 (36)	7 (33)
	50 + 250	94	73 (78)	37 (51)	37 (51)	21 (57)
	Total	246	190 (77)	94 (49)	91 (48)	39 (43)

[0185] 147 마리의 새로 태어난 마우스 중, 본 발명자들은 99 마리의 돌연변이 파운더 마우스를 획득하였다. 배양된 배아에서 관찰되는 결과와 부합하여(도 6b), 돌연변이 비율은 Foxn1-sgRNA의 용량에 비례하였고, 93% (100 ng/ μ l Foxn1-sgRNA)까지 도달하였다 (표 6 및 표 7, 도 5b).

표 7

[0186] T7E1-양성 돌연변이 파운더의 부분집합 (subset)으로부터 확인된 Foxn1 돌연변이 대립유전자의 DNA 서열

Sequence	del+ins	#	Founder mice
ACTTCCAGGCTCCACCCGACTGGAGGGCGAACCCCAAGGGGACCTCATGCAGG			
ACTTCCAGGC-----AACCCCAAGGGGACCTCATGCAGG	Δ 19	1	20
ACTTCCAGGC-----GAACCCCAAGGGGACCTCATGCAGG	Δ 18	1	115
ACTTCCAGGCTCC-----	Δ 60	1	19
ACTTCCAGGCTCC-----	Δ 44	1	108
ACTTCCAGGCTCC-----CAAGGGGACCTCATGCAGG	Δ 21	1	64
ACTTCCAGGCTCC-----TTAGGAGGCGAACCCCAAGGGGACCTCA	Δ 12+6	1	126
ACTTCCAGGCTCCACC-----TCATGCAGG	Δ 28	1	5
ACTTCCAGGCTCCACCC-----CCAAGGGACCTCATG	Δ 21+4	1	61
ACTTCCAGGCTCCACCC-----AAGGGGACCTCATGCAGG	Δ 18	2	95, 29

ACTTCCAGGCTCCACCC-----CAAGGGGACCTCATGCAGG	Δ 17	7	12, 14, 27, 66, 108, 114, 126
ACTTCCAGGCTCCACCC-----ACCCAAGGGGACCTCATGCAG	Δ 15+1	1	32
ACTTCCAGGCTCCACCC-----CACCCAAGGGGACCTCATGCA	Δ 15+2	1	124
ACTTCCAGGCTCCACCC-----ACCCAAGGGGACCTCATGCAGG	Δ 13	1	32
ACTTCCAGGCTCCACCC-----GGCGAACCCCAAGGGGACCTCATGCAGG	Δ 8	1	110
ACTTCCAGGCTCCACCCT-----GGGGACCTCATGCAGG	Δ 20+1	1	29
ACTTCCAGGCTCCACCCG-----AACCCCAAGGGGACCTCATGCAGG	Δ 11	1	111
ACTTCCAGGCTCCACCCGA-----ACCTCATGCAGG	Δ 22	1	79
ACTTCCAGGCTCCACCCGA-----GGGGACCTCATGCAGG	Δ 18	2	13, 127
ACTTCCAGGCTCCACCCCA-----AGGGGACCTCATGCAGG	Δ 17	1	24
ACTTCCAGGCTCCACCCGA-----ACCCAAGGGGACCTCATGCAGG	Δ 11	5	14, 53, 58, 69, 124
ACTTCCAGGCTCCACCCGA-----GACCCCAAGGGGACCTCATGCAGG	Δ 10	1	14
ACTTCCAGGCTCCACCCGA-----GGGCGAACCCCAAGGGGACCTCATGCAGG	Δ 5	3	53, 79, 115
ACTTCCAGGCTCCACCCGAC-----CTCATGCAGG	Δ 23	1	108
ACTTCCAGGCTCCACCCGAC-----CCCCAAGGGGACCTCATGCAGG	Δ 11	1	3
ACTTCCAGGCTCCACCCGAC-----GAAGGGCCCCAAGGGGACCTCA	Δ 11+6	1	66
ACTTCCAGGCTCCACCCGAC-----GAACCCCAAGGGGACCTCATGCAGG	Δ 8	2	3, 66
ACTTCCAGGCTCCACCCGAC-----GGCGAACCCCAAGGGGACCTCATGCAGG	Δ 5	1	27
ACTTCCAGGCTCCACCCGAC--GTGCTTGAGGGCGAACCCCAAGGGGACCTCA	Δ 2+6	2	5
ACTTCCAGGCTCCACCCGACT-----CACTATCTTCTGGGCTCCTCCATGTC	Δ 6+25	2	21, 114
ACTTCCAGGCTCCACCCGACT---TGCGGAACCCCAAGGGGACCTCATGCAG	Δ 4+1	1	53
ACTTCCAGGCTCCACCCGACT--TGCAGGGCGAACCCCAAGGGGACCTCATGC	Δ 2+3	1	126
ACTTCCAGGCTCCACCCGACTTGGAGGGCGAACCCCAAGGGGACCTCATGCAG	+1	15	3, 5, 12, 19, 29, 55, 56, 61, 66, 68, 81, 108, 111, 124, 127
ACTTCCAGGCTCCACCCGACTTGGAGGGCGAACCCCAAGGGGACCTCATGCA	+2	2	79, 120
ACTTCCAGGCTCCACCCGACTGTTGGAGGGCGAACCCCAAGGGGACCTCATGC	+3	1	55
ACTTCCAGGCTCCACCCGACTGGAG(+455)GGCGAACCCCAAGGGGACCTCC	+455	1	13

[0187]

Pkrdc-표적 마우스를 생산하기 위해, 증가하는 용량의 Pkrdc-sgRNA (50, 100, 및 250 ng/μl)과 함께 5배 높은 농도의 Cas9 mRNA (50 ng/μl)를 적용하였다. 다시, 출생 비율은 51% 내지 60%의 범위로 매우 높았고, 분석을 위한 충분한 수의 새로운 마우스를 생산하기에 충분하였다 (표 6). Pkrdc-sgRNA의 최대 용량에서 돌연변이 비율은 57% (37 마리의 새로 태어난 마우스 중 21마리의 돌연변이 파운더)이었다. RGENs으로 얻은 상기 출생률은 본 발명자의 이전 연구에서 보고한 TALENs으로 얻은 것 (Sung et al., 2013)보다 대략 2 내지 10배 더 높았다. 상기 결과는 RGENs이 최소의 독성을 갖는 잠재적 유전자-타겟팅 시약이라는 것을 설명한다.

[0188]

돌연변이 대립 유전자의 생식선 이동 (germ-line transmission)을 시험하기 위해, 네 개의 서로 다른 대립유전자의 모자이크를 갖는 Foxn1 돌연변이 파운더 #108 (도 5c 및 표 8)를 야생형 마우스와 교배하였고, F1 자손의 유전자형을 관찰하였다.

표 8

Foxn1 돌연변이 마우스의 유전자형

[0189]

파운더 NO.	sgRNA (ng/ml)	유전형질 분석 요약 (Genotyping Summary)	탐지된 대립 유전자 (Detected alleles)
58*	1	not determined	Δ 11
19	100	이중 대립 형질 (bi-allelic)	Δ 60/+1
20	100	이중 대립 형질 (bi-allelic)	Δ 67/ Δ 19
13	100	이중 대립 형질 (bi-allelic)	Δ 18/+455

32	10	이중 대립 형질 (bi-allelic), (이형접합, heterozygote)	$\Delta 13 / \Delta 15+1$
115	10	이중 대립 형질 (bi-allelic), (이형접합, heterozygote)	$\Delta 18 / \Delta 5$
111	10	이중 대립 형질 (bi-allelic), (이형접합, heterozygote)	$\Delta 11/+1$
110	10	이중 대립 형질 (bi-allelic), 동형접합, homozygote)	$\Delta 8 / \Delta 8$
120	10	이중 대립 형질 (bi-allelic), 동형접합, homozygote)	$+2/+2$
81	100	이형접합 (heterozygote)	$+1/WT$
69	100	동형접합 (homozygote)	$\Delta 11 / \Delta 11$
55	1	모자이크 (mosaic)	$\Delta 18 / \Delta 1/+1/+3$
56	1	모자이크 (mosaic)	$\Delta 127 / \Delta 41 / \Delta 2/+1$
127	1	모자이크 (mosaic)	$\Delta 18/+1/WT$
53	1	모자이크 (mosaic)	$\Delta 11 / \Delta 5 / \Delta 4+1/WT$
27	10	모자이크 (mosaic)	$\Delta 17 / \Delta 5/WT$
29	10	모자이크 (mosaic)	$\Delta 18 / \Delta 20+1/+1$
95	10	모자이크 (mosaic)	$\Delta 18 / \Delta 14 / \Delta 8 / \Delta 4$
108	10	모자이크 (mosaic)	$+1 / \Delta 17 / \Delta 23 / \Delta 44$
114	10	모자이크 (mosaic)	$\Delta 17 / \Delta 8 / \Delta 6+25$
124	10	모자이크 (mosaic)	$\Delta 11 / \Delta 15+2/+1$
126	10	모자이크 (mosaic)	$\Delta 17 / \Delta 2+3 / \Delta 12+6$
12	100	모자이크 (mosaic)	$\Delta 30 / \Delta 28 / \Delta 17/+1$
5	100	모자이크 (mosaic)	$\Delta 28 / \Delta 11 / \Delta 2+6/+1$
14	100	모자이크 (mosaic)	$\Delta 17 / \Delta 11 / \Delta 10$
21	100	모자이크 (mosaic)	$\Delta 127 / \Delta 41 / \Delta 2 / \Delta 6+25$
24	100	모자이크 (mosaic)	$\Delta 17/+1/WT$
64	100	모자이크 (mosaic)	$\Delta 31 / \Delta 21/+1/WT$
68	100	모자이크 (mosaic)	$\Delta 17 / \Delta 11/+1/WT$
79	100	모자이크 (mosaic)	$\Delta 22 / \Delta 5/+2/WT$
61	100	모자이크 (mosaic)	$\Delta 21+4 / \Delta 6/+1/+9$
66**	100	모자이크 (mosaic)	$\Delta 17 / \Delta 8 / \Delta 11+6/+1/WT$
3	100	모자이크 (mosaic)	$\Delta 11 / \Delta 8/+1$

[0190]

밀줄 그은 대립유전자의 서열을 분석하였다.

[0191]

적색으로 표시된 대립유전자를 fPCR이 아닌 시퀀싱에 의해 분석하였다.

[0192]

*오직 하나의 클론만 서열을 분석하였다.

[0193]

**fPCR에 의해 결정되지 않았다.

[0194]

예상한 대로, 모든 자손들은 야생형 대립 유전자 및 돌연변이 대립유전자 중 하나를 포함하는 이형 접합성 돌연변이였다 (도 5d). 본 발명자들은 또한 독립적인 파운더 마우스에서 Foxn1 (도 8) 및 Prkdc (도 9)의 생식선 이동을 확인하였다. 우리가 아는 한에서, 상기 결과는 동물에서 RGEN-유도 돌연변이 대립유전자가 안정적으로 F1 자손에게 전달된다는 첫 번째 증거를 제공한다.

[0195] 실시예 4: 식물에서의 RNA-가이드 유전체 교정

[0196] 4-1. Cas9 단백질의 생산

[0197] 스트렙토코커스 피요젠스 균주 M1 GAS (NC_002737.1)에서 유래한 Cas9 암호화 서열 (4104bp)을 pET28-b(+) 플라스미드로 클로닝하였다. 핵 표적 서열 (nuclear targeting sequence, NLS)를 단백질 N 말단에 포함시켜 상기 단백질이 핵에 위치할 수 있도록 하였다. Cas9 ORF를 포함하는 pET28-b(+) 플라스미드를 BL21(DE3)에 형질전환시켰다. 0.2 mM IPTG를 이용하여 16시간 동안 18°C에서 Cas9을 유도하였고, 제조사의 지시에 따라 Ni-NTA 아가로스 비드 (Qiagen)를 이용하여 정제하였다. 정제된 Cas9 단백질을 Ultracel - 100K (Millipore)를 이용하여 농축하였다.

[0198] 4-2. 가이드 RNA의 생산

[0199] Cas9 타겟팅에 필요한 엑손에서 프로토스페이서(protospacer) 인접 모티프 (PAM)이라고 불리는 NGG 모티프의 존재 여부에 대해 BRI1을 암호화하는 아기장대 유전자의 유전체 서열을 스크리닝하였다. 아기장대의 BRI1 유전자를 파괴하고자, 본 발명자들은 NGG 모티프를 포함하는 엑손에서 두 RGEN 표적 위치를 규명하였다. 주형 DNA를 사용하여 인 비트로에서 sgRNA를 생산하였다. 두 개의 부분적으로 중첩되는 올리고뉴클레오타이드 (two partially overlapped oligonucleotides) (Macrogen, 표 X1)의 연장 및 다음의 조건을 가지는 Phusion 폴리머라제 (Thermo Scientific)을 사용하여 각 주형 DNA를 생산하였다 - 98°C 30초 {98°C 10초, 54°C 20초, 72°C 2분}x20, 72°C 5분.

표 9

[0200] 인 비트로 전사를 위한 주형 DNA의 생산을 위한 올리고뉴클레오타이드

올리고뉴클레오타이드	서열 (5'-3')	서열번호
BRI1 target 1 (정방향)	GAAATTAATACGACTCACTATAGGTTTGAAAGATGGAAGCGGGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	73
BRI1 target 2 (정방향)	GAAATTAATACGACTCACTATAGGTGAACTAAACTGGTCCACAGTTTTAGA GCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCG	74
Universal (역방향)	AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTT ATTTTAACTTGC	75

[0201] 연장된 DNA를 정제하여 MEGashortscript T7 키트 (Life Technologies)를 사용하여 가이드 RNA를 인 비트로 생산하기 위한 주형으로 사용하였다. 그 다음, 가이드 RNA를 페놀/클로로포름 추출 및 에탄올 침전으로 정제하였다. Cas9/sgRNA 복합체를 제조하기 위해, 10 µl의 정제한 Cas9 단백질 (12 µg/µl) 및 각 두 srRNAs의 4 µl (11 µg/µl)를 NEB3 버퍼 (New England Biolabs) 20 µl에 혼합하고, 37°C에서 10분 동안 반응시켰다.

[0202] 4-3. Cas9/sgRNA 복합체의 원생동물로의 형질주입 (transfection of Cas9/sgRNA complex to protoplast)

[0203] 페트리 접시에서 무균 배양한 4주된 아기장대의 잎을 효소 용액 (1% 셀룰로스 R10, 0.5% 마세로자임 (macerozyme) R10, 450 mM 만니톨, 20mM MES pH 5.7 및 CPW 염)에서 25°C 및 8 내지 16시간 동안 암 상태에서 40 rpm으로 교반하여 분해하였다. 효소/원생동물 용액을 여과하여 100 x g에서 3 내지 5분 동안 원심분리하였다. 혈구계수기 (hemacytometer)를 이용하여 현미경 (X100) 하에서 세포를 계수한 다음, 원생동물을 CPW 용액에 재현탁하였다. 끝으로, 원생동물을 MMG 용액 (4mM HEPES pH 5.7, 400 mM 만니톨 및 15 mM MgCl2)에서 1X10⁶/ml의 농도로 재현탁하였다. Cas9/sgRNA 복합체를 원생동물에 형질주입하기 위해, 200 µl의 원생동물 현탁액 (200,000 원생동물)을 3.3 µl 또는 10 µl의 Cas9/sgRNA 복합체 [Cas9 단백질(6 µg/µl) 및 두

sgRNAs (각 2.2 $\mu\text{g}/\mu\text{l}$) 및 200 μl 의 40% 폴리에틸렌글리콜 형질주입 버퍼 (40% PEG4000, 200 mM 만니톨 및 100 mM CaCl₂)와 함께 2 ml 튜브에서 부드럽게 혼합하였다. 상온에서 5 내지 20분간 반응한 후에, W5 용액 (2 mM MES pH 5.7, 154 mM NaCl, 125 mM CaCl₂ 및 5 mM KCl)과 함께 세척 버퍼를 첨가하여 형질주입을 중단하였다. 그 다음, 원생동물을 100 x g에서 5분 동안 원심분리하여 모았고, W5 용액 1 ml로 세척하고, 100 x g에서 5분 동안 더 원심분리 하였다. 원생동물의 밀도를 1×10^5 /ml로 조정하였고, 이를 400 mM 글루코스가 포함된 변형된 KM 8p 액체 배지에서 배양하였다.

[0204]

[0205]

4-4. 애기장대 원생동물 및 식물에서 돌연변이의 감지

[0206]

형질주입 24시간 또는 72시간 후에, 원생동물을 모으고, 유전체 DNA를 분리하였다. 두 표적 위치를 감싸는 (spanning) 유전체 DNA 부위를 PCR-증폭하였고, T7E1 어세이에 적용하였다. 도 11에 나타난 것처럼, 인델 (indels)은 RGENs에 의해 50 내지 70%의 범위의 높은 비율로 유도되었다. 놀랍게도, 돌연변이는 형질주입 후 24 시간째에 유도되었다. 명백한 것은 Cas9 단백질은 형질주입 후 즉시 작용한다. PCR 산물을 정제하였고, T-Blunt PCR 클로닝 키트 (Solgent)로 클로닝하였다. 플라스미드를 정제하였고, M13F 프라이머와 함께 Sanger 시퀀싱에 적용하였다. 하나의 돌연변이 서열은 한 위치에 7-bp 결실을 가졌다 (도 12). 다른 세 돌연변이 서열은 두 RGEN 위치 사이에서 ~220-bp의 DNA 부위의 결실을 가졌다.

[0207]

실시에 5: 세포-침투 펩타이드 또는 단백질 전달 도메인을 이용한 Cas9 단백질 전달 (Cas9 protein transduction using a cell-penetrating peptide or protein transduction domain)

[0208]

5-1. His-Cas9-암호화 플라스미드의 구축

[0209]

이전에 기술된 Cas9 플라스미드 {Cho, 2013 #166}를 주형으로 이용하여 C-말단에 시스테인 (cysteine)을 갖는 Cas9을 PCR 증폭으로 제조하였고, N-말단에 His-태그를 포함하는 pET28-(a) 벡터 (Novagen, Merk Millipore, Germany)에 클로닝하였다.

[0210]

5-2. 세포배양

[0211]

293T (인간 배아 신장 세포주) 및 HeLa (인간 난소암 세포주)를 10% FBS 및 1% 페니실린 및 스트렙토마이신을 보충한 DMEM (GIBCO-BRL Rockville)에서 배양하였다.

[0212]

5-3. Cas9 단백질의 발현 및 정제

[0213]

Cas9 단백질을 발현하기 위해, 대장균 BL21 세포를 Cas9을 암호화하는 pET28-(a) 벡터에 형질전환하였고, 50 μg /mL 카나마이신 (Amresco, Solon, OH)을 포함하는 루리아-버타니 (LB) 아가 배지에 플레이팅하였다. 다음날, 단일 콜로니를 선택하여 50 $\mu\text{g}/\text{mL}$ 카나마이신을 포함하는 LB 배양액에서 37°C에서 밤새 배양하였다. 그 다음날, 0.1 OD600에서 시작한 배양액을 50 $\mu\text{g}/\text{mL}$ 카나마이신을 포함하는 루리아 배양액에 접종하였고, OD600이 0.6-0.8에 도달할 때까지 37°C에서 2시간 동안 배양하였다. Cas9 단백질의 발현을 유도하기 위해, 이소프로필- β -D-티오갈락토피라노사이드 (IPTG)(Promega, Madison, WI)를 최종 농도 0.5mM가 되도록 첨가한 다음, 세포를 30°C에서 밤새 배양하였다.

[0214]

세포를 4000rpm에서 15 내지 20분 동안 원심분리하여 모았고, 용출 버퍼 (20mM Tris-Cl pH8.0, 300mM NaCl, 20mM 이미다졸, 1X 프로테아제 억제제 카테일, 1 mg/ml 라이소자임)에 재현탁하고, 음파 처리 (40% duty, 10

sec pulse, 30 sec rest, for 10 mins on ice)로 용해하였다. 수용성 분획을 4°C 및 15,000rpm에서 20분 동안 원심분리하여 상층액으로서 분리하였다. Cas9 단백질을 Ni-NTA 아가로스 레진 (QIAGEN)을 포함하는 컬럼 및 AKTA prime 기기 (AKTA prime, GE Healthcare, UK)를 이용하여 4°C에서 정제하였다. 상기 크로마토그래피 단계 동안, 수용성 단백질 분획을 1 ml/분의 유속으로 Ni-NTA 아가로스 레진 (GE Healthcare, UK)에 로딩하였다. 상기 컬럼을 세척 버퍼 (20mM Tris-Cl pH8.0, 300mM NaCl, 20mM 이미다졸, 1X 프로테아제 억제제 카테일)로 세척하였고, 결합된 단백질을 0.5 ml/분의 유속으로, 용출 버퍼 (20mM Tris-Cl pH8.0, 300mM NaCl, 250mM 이미다졸, 1X 프로테아제 억제제 카테일)로 용출하였다. 모은 용출된 분획을 농축하였고, 저장 버퍼 (50 mM Tris-HCl, pH8.0, 200 mM KCl, 0.1 mM EDTA, 1 mM DTT, 0.5 mM PMSF, 20% 글리세롤)로 투석하였다. 단백질 농도는 브래드포드 (Bradford) 어세이 (Biorad, Hercules, CA)로 정량하였고, 순도는 소 혈청 알부민을 대조군으로 사용한 SDS-PAGE로 분석하였다.

[0215] **5-4. 9R4L에 Cas9의 접합 (conjugation of Cas9 to 9R4L)**

[0216] 1mg/ml의 농도로 PBS에 희석한 1mg Cas9 단백질과 25 µl DW에 있는 50 µg의 말레이미드-9R4L 펩타이드를 2시간 동안 상온 및 그 후 4°C에서 밤새 로터를 이용하여 부드럽게 혼합하였다. 접합하지 않은 maleimide-9R4L를 제거하기 위해 상기 시료를 50kDa 분자량 컷오프 막을 이용하여 4°C에서 24시간 동안 DPBS (pH 7.4)에 대해 투석하였다. Cas9-9R4L 단백질을 투석막으로부터 모았고, 단백질 양을 브래드포드 어세이를 이용하여 측정하였다.

[0217] **5-5. sgRNA-9R4L의 제조**

[0218] sgRNA (1 µg)을 100 µl DPBS (pH 7.4)에서 다양한 양의 C9R4LC 펩타이드 (1 내지 40 무게 비율의 범위)에 부드럽게 첨가하였다. 상기 혼합물을 30분 동안 상온에서 반응시켰고, RNAase가 없는 탈이온화된 물을 이용하여 10배 희석하였다. 형성된 나노입자의 유체역학적 직경 및 z-전위를 동적 광산란 (dynamic light scattering) (Zetasizer-nano analyzer ZS; Malvern instruments, Worcestershire, UK)을 이용하여 측정하였다.

[0219] **5-6. Cas9 단백질 및 sgRNA의 처리**

[0220] Cas9-9R4L 및 sgRNA-C9R4LC를 다음과 같이 세포에 처리하였다: 1 µg의 sgRNA 및 15 µg의 C9R4LC 펩타이드를 250 ml의 OPTIMEM 배지에 처리하였고, 상온에서 30분 동안 반응시켰다. 분주 (seeding) 후 24시간 시점에, 세포를 OPTIMEM 배지로 세척하였고, sgRNA-C9R4LC 복합체로 37°C에서 4시간 동안 처리하였다. 세포를 다시 OPTIMEM 배지로 세척하였고, Cas9-C9R4LC로 37°C에서 2시간 동안 처리하였다. 처리 후, 배양 배지를 혈청이 포함된 완전 배지에 교체하였고, 다음 처리 전에 37°C에서 24시간 동안 배양하였다. 동일한 방법으로 연속적인 삼일 동안 Cas9 및 sgRNA를 여러 번 처리하였다.

[0221] **5-7. Cas9-9R4L 및 sgRNA-C9R4LC는 추가의 전달 수단의 사용 없이 배양된 포유동물 세포에서 내재적 유전자를 교정 (edit)할 수 있다.**

[0222] Cas9-9R4L 및 sgRNA-9R4L이 추가의 전달 수단의 사용 없이 배양된 포유동물 세포에서 내재적 유전자를 교정할 수 있는지 여부를 확인하기 위해, 본 발명자들은 CCR5 유전자를 타겟팅하는 Cas9-9R4L 및 sgRNA-9R4L을 239 세포에 처리하였고, 유전체 DNA를 분석하였다. T7E1 어세이에서 Cas9-9R4L 및 sgRNA-9R4L 둘 다 처리한 세포는 9%의 CCR5 유전자가 파괴되었음을 보였고, CCR5 유전자 파괴는 상기 Cas9-9R4L 및 sgRNA-9R4L을 처리하지 않거나, Cas9-9R 또는 sgRNA-9R4L 중 어느 하나를 처리하거나, 또는 변형하지 않은 Cas-9 또는 sgRNA 모두를 처리한 것을 포함하여 대조군 세포에서 관찰되지 않았고(도 13), 이는 변형하지 않은 Cas9 또는 sgRNA가 아닌, Cas9-9R4L 단백질 및 9R4L과 접합한 sgRNA의 처리가 포유동물 세포에서 효율적인 유전체 교정을 가져올 수 있음을 제

안하는 것이다.

[0223] **실시예 6: 가이드 RNA 구조에 따른 오프-타겟 돌연변이의 조절**

[0224] 최근, 세 그룹은 RGENs이 인간 세포에서 오프-타겟 효과를 갖는다는 것을 보고하였다. 놀랍게도, RGENs은 온-타겟 위치와 3 내지 5 뉴클레오타이드가 다른 오프-타겟 위치에서 돌연변이를 효율적으로 유도하였다. 하지만, 우리는 본 발명자의 RGENs과 다른 발명자에 의해 사용된 RGENs 간에는 여러 다른 점이 있는 것을 발견하였다. 첫 번째, 본 발명자들은 crRNA 및 tracrRNA의 필수적 부분을 구성하는 단일-가이드 RNA (sgRNA) 대신, crRNA 더하기 tracrRNA인 이중RNA (dualRNA)를 사용하였다. 두 번째, 본 발명자들은 crRNA를 암호화하는 플라스미드 대신에 합성한 crRNA를 (HeLa 세포가 아닌) K562 세포에 형질주입하였다. HeLa 세포는 crRNA-암호화 플라스미드를 형질주입하였다. 다른 발명자는 sgRNA-암호화 플라스미드를 사용하였다. 세 번째, 본 발명자의 가이드 RNA는 인 비트로에서 T7 폴리머라제에 의한 효율적 전사에 필요한, 5' 말단에 두 개의 추가적 구아닌(guanine) 뉴클레오타이드를 가졌다. 상기 추가적 뉴클레오타이드는 다른 발명자에 의해 사용된 sgRNA에 포함되지 않았다. 그러므로, 본 발명자의 가이드 RNA의 RNA서열은 5'-GGX₂₀로 나타낼 수 있고, 반면 X₂₀ 또는 GX₁₉가 20-bp 표적 서열과 대응되는 5'-GX₁₉는 다른 발명자에 의해 사용된 서열을 나타낸다. 첫 번째 구아닌 뉴클레오타이드는 세포에서 RNA 폴리머라제로 전사하는데 필요하다. 오프-타겟 RGEN 효과가 상기 차이에 기여할 수 있는지 여부를 평가하기 위해, 인간 세포에서 높은 비율로 오프-타겟 돌연변이를 유도하는 네 개의 RGENs을 선택하였다 (13). 우선, 본 발명자들은 인 비트로 전사된 이중RNA를 사용한 우리의 방법과 K562 세포에서 sgRNA-암호화 플라스미드를 형질주입하는 방법을 비교하였고, T7E1 어세이를 통해 온-타겟 및 오프-타겟 위치에서의 돌연변이 빈도를 측정하였다. 세 RGENs은 가이드 RNA의 조성에 관계없이 온-타겟 및 오프-타겟 위치에서 비슷한 돌연변이 빈도를 보였다. 흥미롭게도, 합성된 이중RNA를 사용하였을 때, 하나의 RGEN (VEFGA 위치 1)은 온-타겟 위치 (용어 OT1-11, 도 14)에서 세 개의 뉴클레오타이드가 다른, 하나의 유효한 오프-타겟 위치 (one validated off-target site)에서 인델(indels)을 유도하지 않았다. 하지만 합성된 이중RNA는 온-타겟 위치에서 두 개의 뉴클레오타이드가 다른, 나머지 유효한 오프-타겟 위치 (OT1-3)를 구별하지 않았다.

[0225] 다음으로, 본 발명자들은 5'-GGX₂₀ (또는 5'-GGGX₁₉) sgRNA와 5'-GX₁₉ sgRNA를 비교하여, sgRNA의 5' 말단에 첨가된 두 개의 구아닌 뉴클레오타이드가 RGENs을 보다 특이적으로 만드는지 여부를 시험하였다. Cas9과 복합체를 형성한 네 개의 GX₁₉ sgRNA는 네 개까지 뉴클레오타이드의 불일치를 용인하며, 인델(indel)을 온-타겟 및 오프-타겟 위치에서 동등한 효율로 유도하였다. 날카롭게 대조하면, GGX₂₀ sgRNAs는 효율적으로 오프-타겟 위치를 구별하였다. 사실, 본 발명자들이 4개의 GGX₂₀ sgRNAs를 사용하였을 때, T7E1 어세이는 7개의 유효한 오프-타겟 위치 중 6개에서 RGEN-유도 인델을 거의 감지하지 않았다 (도 15). 하지만, 본 발명자들은 두 GGX₂₀ sgRNAs (VEGFA 위치 1 및 3)는 GX₁₉ sgRNA에 대응되는 온-타겟 위치에서 활성이 더 적었다. 상기 결과는, 5' 말단에서 추가적 뉴클레오타이드가 아마 가이드 RNA 안정성, 농도 또는 이차 구조의 변화에 의해 온-타겟 및 오프-타겟 위치에서 돌연변이 빈도에 영향을 미칠 수 있다는 것을 보여준다.

[0226]

[0227] 상기 결과는 세 개의 요소 -가이드 RNA-암호화 플라스미드보다 합성 가이드 RNA의 사용, sgRNA보다는 이중RNA의 사용, 및 GX₁₉ sgRNA보다 GGX₂₀ sgRNAs의 사용 - 가 오프-타겟 위치의 구별에 있어 누적 효과를 갖는다는 것을 암시한다.

[0228] **실시예 7: Cas9 니카아제 쌍 (Paired Cas9 nickases)**

[0229] 원칙적으로, 단일-사슬 절단 (single-strand break, SSBs)은 오류 유발 NHEJ에 의해 수선될 수 없지만, 높은 정확도의 상동성-인도 수리 (homology-directed repair, HDR) 또는 염기 절단 수선 (base excision repair)을 촉진한다. 그러나 HDR을 통한 니카아제(nickase)-유도 표적화된 돌연변이는 뉴클레아제-유도 돌연변이에 비해 덜 효율적이다. 본 발명자들은 Cas9 니카아제 쌍이 NHEJ 또는 HDR을 통해 DNA 수선을 야기하는 복합 DSBs

(composite DSBs)를 생성하여, 효율적인 돌연변이를 유도할 것이라 추론하였다 (도 16A). 더욱이, 니카아제 쌍은 Cas9-기반 유전체 교정의 특이성을 두 배로 만들 수 있다.

[0230]

본 발명자들은 먼저 AAVS1 좌위에서 표적 위치에 대해 설계된 여러 Cas9 뉴클레아제 및 니카아제를 인 비트로에서 형광 모세관 전기영동을 통해 시험하였다(도 16B). DNA 기질의 양 가닥을 절단하는 Cas9 뉴클레아제와 다르게, 가이드 RNA 및 촉매 아스파라긴산 (aspartate) 잔기가 알라닌 (alanine)으로 변경된 Cas9의 돌연변이 형태 (D10A Cas9)로 구성된 Cas9 니카아제는 오직 하나의 가닥만 분해하였고, 위치-특이적 틈 (nick)을 만들었다 (도 16 C,D). 그러나 흥미롭게도, 몇몇 니카아제 (도 17A에서 AS1, AS2, AS3 및 S6)는 인간 세포의 표적 위치에서 인텔을 유도하였고, 비록 비효율적이기는 하지만 틈이 인 비보에서 DSBs로 전환될 수 있다는 것을 제안하였다. 반대의 DNA 가닥 (opposite DNA strand)에서 인접한 두 틈을 만드는 Cas9 니카아제 쌍은 뉴클레아제 쌍에 의한 효과와 비교했을 때, 14 내지 91%의 범위의 빈도로 인텔을 만들었다 (도 17A). 5' 오버행 (overhang)을 만드는 두 틈의 수선은 세 유전체 좌위에서 3' 오버행을 만드는 것보다 더 빈번하게 인텔이 형성을 가져왔다 (도 17A 및 도 18). 또한, 니카아제 쌍은 단일 니카아제에 의한 것보다 더 효율적으로 상동-인도 수선을 통한 표적 유전체 교정을 가능하게 하였다 (도 19).

[0231]

다음, 딥 시퀀싱을 이용하여 오프-타겟 위치에서의 니카아제 쌍 및 뉴클레아제의 돌연변이 빈도를 측정하였다. 세 개의 sgRNAs와 복합체를 형성한 Cas9 니카아제는, 상응하는 온-타겟 위치와 하나 또는 두 개의 뉴클레오타이드가 다른 여섯 개의 위치에서 오프-타겟 돌연변이를 0.5% 내지 10%의 범위의 빈도로 유도하였다 (도 17B). 대조적으로, Cas9 니카아제 쌍은 여섯 개의 위치 중 어느 곳에서도 0.1%의 탐지 한도 (detection limit)를 넘는 인텔을 생산하지 않았다. 온-타겟 위치에서 PAM의 첫 번째에 위치하는 단일 뉴클레오타이드 (즉, NGG에서 N)가 다른 S2 오프-1 위치는 또 다른 온-타겟 위치로서 간주될 수 있다. 예상한 대로, S2 sgRNA와 복합체를 형성한 Cas9 뉴클레아제는 상기 위치 및 온-타겟 위치에서 동일한 효율을 보였다. 날카롭게 대조하면, S2 및 AS2 sgRNAs와 복합체를 형성한 D10A Cas9는 270배의 인수로 온-타겟 위치로부터 상기 위치를 구별하였다. 이러한 니카아제 쌍은 또한 각각 160배 및 990배의 인수로 온-타겟 위치로부터 AS2 오프-타겟 위치 (도 17B에서 Off-1 및 Off-9)를 구별하였다.

[0232]

실시예 8: Cas9 니카아제 쌍에 의해 유도된 염색체 DNA 스플라이싱 (chromosomal DNA splicing induced by paired Cas9 nickases)

[0233]

ZFNs 및 TALENs와 같은 유전자 가위에 의해 생산된 두 동시의 DSBs가, 개입한 유전체 부분 (intervening chromosomal segment)의 큰 결실을 촉진할 수 있다는 것이 보고되었다. 본 발명자들은 Cas9 니카아제 쌍에 의해 유도된 두 SSBs가 또한 인간 세포에서 결실을 생산할 수 있는지를 시험하였다. 본 발명자들은 PCR을 사용하여 결실 발생을 탐지하였고, 일곱 개의 니카아제 쌍이 Cas9 뉴클레아제 쌍만큼 효과적으로 1.1-kbp 염색체 부분까지 결실을 유도하였음을 확인하였다 (도 20A,B). PCR 산물의 DNA 서열로 결실을 확인하였다 (도 20C). 흥미롭게도, sgRNA-매칭 서열은 일곱 개의 결실-특이적 PCR 앰플리콘 중 2개에서 그대로 남아있었다 (도 20C에서 밑줄). 대조적으로, Cas9 뉴클레아제 쌍은 온전한 표적 위치를 포함하는 서열을 만들지 않았다. 이러한 발견은 두 개의 떨어진 틈은 개입된 염색체 부분의 결실을 촉진하는 두 개의 분리된 DSBs로 전환되지 않는다는 것을 암시한다. 또한, 100 bp보다 더 떨어진 두 개의 틈은, 녹는점 (melting temperature)이 매우 높기 때문에 생리학적 조건 하에서 큰 오버행과 함께 복합 DSBs를 생성할 수 있다.

[0234]

본 발명자들은 두 개의 떨어진 틈이 머리-머리(head-to-head) 방향에서 가닥 변위 (strand displacement)에 의해 수선되고, 중간 (middle)에 DSB의 형성을 야기하며, NHEJ를 통한 이의 수선은 작은 결실을 야기한다는 것을 제시한다 (도 20D). 상기 과정 동안 두 표적 위치는 그대로 남아있기 때문에 니카아제는 SSBs를 다시 유도할 수 있고, 표적 위치가 결실될 때까지 사이클을 반복적으로 유도한다. 상기 메커니즘은 왜 3' 오버행을 생산하는 것이 아닌 5' 오버행을 생산하는 두 오프셋 틈 (two offset nicks)이 세 좌위에서 인텔을 효율적으로 유도하는지 설명한다.

[0235] 그 다음으로 본 발명자들은 온-타겟 및 오프-타겟 DNA 절단의 NHEJ 수선에 의한 결과인 원치않는 유전체 전좌를 Cas9 뉴클레아제 및 니카아제가 유도할 수 있는지 여부를 조사하였다 (도 21A). 본 발명자들은 PCR을 이용하여 Cas9 뉴클레아제에 의해 유도된 전좌를 탐지할 수 있었다 (도 21 B,C). 어떠한 PCR 산물도 AS2+S3 Cas9 니카아제 쌍을 암호화하는 플라스미드를 형질주입한 세포로부터 분리한 유전체 DNA를 이용하여 증폭되지 않았다. 상기 결과는 AS2 및 S3 니카아제 모두가, 이들의 상응하는 뉴클레아제와는 달리 오프-타겟 위치에서 indels를 생산하지 못했다는 사실과 일치한다 (도 17B).

[0236] 이러한 결과는 Cas9 니카아제 쌍이 인간 세포에서 표적화된 돌연변이 및 1-kbp 염색체 단편에 달하는 큰 결실이 일어나는 것을 허용한다는 것을 암시한다. 중요한 것은, 니카아제 쌍은, 이들의 상응하는 뉴클레아제가 돌연변이를 유도하는 오프-타겟 위치에서 인델을 유도하지 않았다. 또한, 뉴클레아제와 다르게, 니카아제 쌍은 오프-타겟 DNA 절단과 관련된 원하지 않는 전좌를 촉진하지 않았다. 원칙적으로, 니카아제 쌍은 Cas-매개 돌연변이의 특이성을 두 배로 하고, 유전자 및 세포 치료제 같은 정확한 유전체 교정을 요구하는 응용에서 RNA-가이드 효소의 효용성을 높일 것이다. 이러한 접근에서 한가지 주의할 점은 두 개의 매우 높은 활성을 갖는 sgRNA가 표적이 될 수 있는 위치를 제한하면서 효율적인 니카아제 쌍을 만드는 데 필요하다는 것이다. 본 발명 및 다른 연구에서 볼 수 있듯이, 모든 sgRNAs의 활성이 동일한 것은 아니다. 세포 집단 보다 단일 클론이 후속 연구 또는 응용될 때, 유전체에서 독특한 서열을 나타내는 가이드 RNAs의 선택 및 최적화된 가이드 RNA의 사용으로 Cas9 뉴클레아제와 관련된 오프-타겟 돌연변이를 막는데 충분할 것이다. 본 발명자들은 Cas9 뉴클레아제 및 니카아제 쌍 모두가 세포 및 유기체에서 정확한 유전체 교정을 촉진할 수 있는 강력한 선택임을 제안한다.

[0237] **실시예 9: CRISPR/Cas-유래 RNA-가이드 엔도뉴클레아제를 이용한 유전형질 분석 (genotyping with CRISPR/Cas-derived RNA-guided endonucleases)**

[0238] 다음으로, 본 발명자들은 통상적인 제한 효소를 대체하며, RGENs이 제한 단편 길이 다형성 (RFLP) 분석에 사용될 수 있을 것임을 추론하였다. 뉴클레아제에 의해 야기된 DSBs가 오류 유발 비상동 말단 결합 (NHEJ) 시스템에 의해 수선될 때, RGENs을 포함하는 유전자 가위는 표적 위치에서 인델을 유도한다. 표적 서열을 인식하도록 설계된 RGENs은 인델을 가진 돌연변이 서열을 절단하지 못하지만, 야생형 타겟 서열은 효율적으로 절단할 수 있을 것이다.

[0239] **9-1. RGEN 요소 (RGEN component)**

[0240] crRNA 및 tracrRNA를 제조자의 지시에 따라 MEGAscript T7 키트 (Ambion)를 이용한 인 비트로 전사로 제조하였다. 전사된 RNAs를 8% 변성 요소-PAGE 겔에서 분리하였다. RNA가 포함된 겔 단편을 잘라내어 용출 버퍼 (elution buffer)에 옮겼다. RNA를 뉴클레아제가 없는 물에서 회수한 다음, 페놀:클로로포름 추출, 클로로포름 추출 및 에탄올 침전을 수행하였다. 정제된 RNAs를 분광계로 정량하였다. X₂₀이 타겟 서열인 5'-GAAATTAATACGACTCACTATAGG_{X20}GTTTGTAGAGCTATGCTGTTTG-3' (서열번호: 76)로 나타낸 서열 및 그것의 상보적인 서열을 갖는 올리고뉴클레오타이드로 어닐링하여 crRNA에 대한 주형을 제조하였다. tracrRNA의 주형을 Phusion 폴리머라제(New England Biolabs)를 이용하여 정방향 및 역방향 올리고뉴클레오타이드 5'-GAAATTAATACGACTCACTATAGGAACCAATCAAAACAGCATAGCAAGTAAAATAAGGCTAGTCCG-3' (서열번호: 77) 및 5'-AAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTATTTTAACTTGCTATG-3' (서열번호: 78)의 연장으로 합성하였다.

[0241] **9-2. 재조합 Cas9 단백질 정제**

[0242] C-말단에 His6-태그와 융합된 Cas9을 암호화하고 있는, 본 발명자들의 앞선 실시예에서 사용된 Cas9 DNA 작제물

을 pET-28a 발현 벡터에 삽입하였다. 재조합 Cas9 단백질을 1mM IPTG로 유도한 후에 4시간 동안 25°C의 LB 배지에서 배양시킨 대장균 균주 BL21 (DE3)에서 발현시켰다. 세포를 수확하고, 20 mM Tris PH 8.0, 500 mM NaCl, 5 mM 이미다졸, 및 1 mM PMSF가 포함된 버퍼에 재현탁하였다. 세포를 액체 질소에 얼리고, 4°C에서 녹인 후, 음과 처리하였다. 원심분리 후, 용해물에 있는 Cas9 단백질을 Ni-NTA 아가로스 레진 (Qiagen)에 결합시켰고, 20 mM Tris pH 8.0, 500 mM NaCl, 및 20 mM 이미다졸이 포함된 버퍼로 세척한 후, 20 mM Tris pH 8.0, 500 mM NaCl, 및 250 mM 이미다졸이 포함된 버퍼로 용출하였다. 정제된 Cas9 단백질을 20 mM HEPES (pH 7.5), 150 mM KCl, 1 mM DTT, 및 10% 글리세롤로 투석하였고, SDS-PAGE를 이용하여 분석하였다.

[0243] **9-3. T7 엔도뉴클레아제 I 어세이**

[0244] T7E1 어세이를 다음과 같이 수행하였다. 요약하면, 유전체 DNA를 이용하여 증폭한 PCR 산물을 95°C에서 변성시켰고, 16°C에서 재어닐링하여 5 유닛(unit)의 T7 엔도뉴클레아제 I (New England BioLabs)과 함께 20분 동안 37°C에서 반응시켰다. 반응 산물을 2 내지 2.5%의 아가로스 젤 전기영동을 이용하여 분리하였다.

[0245] **9-4. RGEN-RFLP 어세이**

[0246] PCR 산물 (100-150 ng)을 10 μ l의 NEB 버퍼 3 (1X)에서 Cas9 단백질, tracrRNA, crRNA의 최적화된 농도 (표 10)와 함께 60분 동안 37°C에서 반응시켰다. 절단 반응 후, RNase A (4 μ g)를 첨가하여 반응 혼합물을 30분 동안 37°C에서 반응시켜 RNA를 제거하였다. 반응을 30% 글리세롤, 1.2% SDS, 및 100 mM EDTA가 포함된 6X 중단용액 버퍼 (stop solution buffer)로 중단시켰다. 산물을 1 내지 2.5% 아가로스 젤 전기영동을 이용하여 분리하였고, EtBr 염색으로 가시화하였다.

표 10

RGEN-RFLP 어세이에서 RGEN 요소의 농도

[0247]

표적명	Cas9 (ng/ μ l)	crRNA (ng/ μ l)	tracrRNA (ng/ μ l)
<i>C4BPB</i>	100	25	60
<i>PIBF</i> -NGG-RGEN	100	25	60
<i>HLA-B</i>	1.2	0.3	0.7
<i>CCR5</i> -ZFN	100	25	60
<i>CTNNB1</i> Wild type specific	30	10	20
<i>CTNNB1</i> mutant specific	30	10	20
<i>CCR5</i> WT-specific	100	25	60
<i>CCR5</i> 32-specific	10	2.5	6
<i>KRAS</i> WT specific(wt)	30	10	20
<i>KRAS</i> mutant specific(m8)	30	10	20
<i>KRAS</i> WT specific (m6)	30	10	20
<i>KRAS</i> mutant specific (m6,8)	30	10	20
<i>PIK3CA</i> WT specific (wt)	100	25	60
<i>PIK3CA</i> mutant specific(m4)	30	10	20
<i>PIK3CA</i> WT specific (m7)	100	25	60
<i>PIK3CA</i> mutant specific(m4,7)	30	10	20
<i>BRAF</i> WT-specific	30	10	20
<i>BRAF</i> mutant-specific	100	25	60
<i>NRAS</i> WT-specific	100	25	60
<i>NRAS</i> mutant-specific	30	10	20
<i>IDH</i> WT-specific	30	10	20

<i>IDH</i> mutant-specific	30	10	20
<i>PIBF</i> -NAG-RGEN	30	10	60

표 11

프라이머

유전자 (위치)	방향	서열 (5'에서 3')	서열번호
CCR5(RGEN)	F1	CTCCATGGTGCTATAGAGCA	79
	F2	GAGCCAAGCTCTCCATCTAGT	80
	R	GCCCTGTCAAGAGTTGACAC	81
CCR5(ZFN)	F	GCACAGGGTGAACAAGATGGA	82
	R	GCCAGGTACCTATCGATTGTCAGG	83
CCR5(de132)	F	GAGCCAAGCTCTCCATCTAGT	84
	R	ACTCTGACTG GGTCCACCAGC	85
<i>C4BPB</i>	F1	TATTTGGCTGGTTGAAAGGG	86
	R1	AAAGTCATGAAATAAACACACCCA	87
	F2	CTGCATTGATATGGTAGTACCATG	88
	R2	GCTGTTCAATGCAATGGAATG	89
CTNNB1	F	ATGGAGTTGGACATGGCCATGG	90
	R	ACTCACTATCCACAGTTCAGCATTTACC	91
<i>KRAS</i>	F	TGGAGATAGCTGTCAGCAACTTT	92
	R	CAACAA AGCAAAGGTAAAGTTGGTAATAG	93
<i>PIK3CA</i>	F	GGTTTCAGGAGATGTTACAAGGC	94
	R	GATTGTGCAATTCCATATGCAATCGGTC	95
<i>NRAS</i>	F	CACTGGGTACTTAATCTGTAGCCTC	96
	R	GGTTCCAAGTCATTCCCAGTAGC	97
<i>IDH1</i>	F	CATCACTGCAGTTGTAGGTTATAACTATCC	98
	R	TTGAAAACCACAGATCTGGTTGAACC	99
<i>BRAF</i>	F	GGAGTGCCAAGAGAATATCTGG	100
	R	CTGAAACTGGTTTCAAATATTCGTTTTAAGG	101
<i>PIBF</i>	F	GCTCTGTATGCCCTGTAGTAGG	102
	R	TTTGCATCTGACCTTACCTTTG	103

[0248]

[0249]

9-5. 플라스미드 절단 어세이

[0250]

제한 효소가 처리된 선형 플라스미드 (100 ng)를 10 μ l의 NEB 3 버퍼 (1X)에서 Cas9 단백질(0.1 μ g), tracrRNA (60 ng), 및 crRNA (25 ng)과 함께 60분 동안 37 $^{\circ}$ C에서 배양하였다. 반응을 30% 글리세롤, 1.2% SDS, 및 100 mM EDTA를 포함하는 6X 중단 용액으로 중단시켰다. 산물을 1% 아가로스 젤 전기영동을 이용하여 분리하였고, EtBr 염색으로 가시화하였다.

[0251]

9-6. RFLP의 전략

[0252]

원하는 DNA 특이성을 갖는 새로운 RGENs은 crRNA를 대체하여 쉽게 만들어질 수 있다; 한 번 재조합 Cas9 단백질이 이용가능하면, 커스텀 단백질 (custom)의 디노보 (de novo) 정제는 필요 없다. 뉴클레아제에 의해 야기된 DSBs가 오류 유발 비 상동 말단 결합 (NHEJ)에 의해 수선될 때, RGENs을 포함한 유전자 가위는 표적 위치에서 작은 삽입 또는 결실 (indels)을 유도한다. 표적 서열을 인식하도록 설계된 RGEN은 야생형 서열을 효과적으로 절단하나, 인델을 가진 돌연변이 서열은 절단할 수 없다 (도 22).

[0253]

본 발명자들은 먼저 야생형 *C4BPB* 표적 서열 또는 절단 위치에서 1- 에서 3- 염기 indel을 갖는 변형된 *C4BPB*

표적 서열을 포함하는 플라스미드를 RGENs이 서로 다르게 절단할 수 있는지 여부를 시험하였다. 상기 indel을 갖는 여섯 개의 플라스미드 중 어느 것도 표적-특이적 crRNA, tracrRNA, 및 재조합 Cas9 단백질로 구성된 C4BPB-특이적 RGEN5에 의해 절단되지 않았다 (도 23). 대조적으로, 온전한 표적 서열을 갖는 플라스미드는 상기 RGEN에 의해 효율적으로 절단되었다.

[0254] **9-7. RGEN-매개 RFLP를 이용한 동일한 RGENs에 의해 유도된 돌연변이의 탐지 (detection of mutations induced by the same RGENs using RGEN-mediated RFLP)**

[0255] 다음으로, 동일한 RGENs에 의해 유도된 돌연변이의 탐지를 위한 RGEN-매개 RFLP의 실행가능성을 시험하기 위해, 본 발명자들은 RGEN 타겟팅 C4BPB 유전자를 이용하여 확립한 유전자-변형 K562 인간 암 세포 클론을 이용하였다 (표 12).

표 12

[0256] 본 발명에서 사용된 RGENs의 표적 서열

유전자	표적 서열	서열번호
human C4BPB	AATGACCACTACATCCTCAAGGG	104
mouse Pibf1	AGATGATGTCTCATCAGAGG	105

[0257] 본 발명에서 사용된 C4BPB 돌연변이 클론은 94 bp 결실에서 67 bp 삽입의 범위에 이르는 다양한 돌연변이를 갖는다 (도 24A). 중요하게도, 돌연변이 클론에서 발생하는 모든 돌연변이는 RGEN 표적 위치의 손실을 가져온다. 분석한 6개의 C4BPB 클론 중, 4개의 클론이 야생형 및 돌연변이 대립유전자 모두를 가졌고 (+/-), 2개의 클론이 돌연변이 대립유전자만을 가졌다 (-/-).

[0258] 야생형 K562 유전체 DNA로부터 증폭된 RGEN 타겟 위치를 감싸는 PCR 산물은, 표적-특이적 crRNA, tracrRNA, 및 대장균에서 발현되고 정제된 재조합 Cas9 단백질로 이루어진 RGEN에 의해 완전히 분해되었다 (도 24B / Lane 1). C4BPB 돌연변이 클론을 RGEN을 이용한 RFLP에 적용하였을 때, 야생형 및 돌연변이 대립유전자 모두를 포함하는 +/- 클론의 PCR 앰플리콘이 부분적으로 분해되었고, 야생형 대립유전자를 포함하지 않는 -/- 클론의 PCR 앰플리콘은 완전히 분해되지 않아, 야생형 서열에 상응하는 절단 산물을 생성하지 않았다 (도 24B). 표적 위치에서의 단일 염기의 삽입조차도 C4BPB RGEN에 의해 증폭된 돌연변이 대립유전자의 분해를 막았고 (#12 및 #28 클론), RGEN-매개 RFLP의 높은 특이성을 보여주었다. 본 발명자들은 PCR 앰플리콘을 불일치-민감 T7E1 어세이에 동일하게 적용하였다 (도 24B). 특히, T7E1 어세이는 +/- 클론으로부터 -/- 클론을 구별하지 못하였다. 설상가상으로, 동일한 돌연변이 서열의 어닐링은 동형이중가닥 (homoduplex)을 형성할 것이기 때문에 T7E1 어세이는 동일한 돌연변이 서열을 포함하는 동형접합 돌연변이 클론을 야생형 클론으로부터 구별할 수 없다. 그러므로, RGEN-매개 RFLP는 ZFNs, TALENs 및 RGENs를 포함하는 유전자 가위에 의해 유도된 돌연변이 클론의 분석에 있어 일반적인 불일치-민감 뉴클레아제 어세이보다 더 중요한 이점을 갖는다.

[0259] **9-8. RGEN-RFLP 분석을 위한 정량적 어세이**

[0260] 본 발명자들은 또한 RGEN-RFLP 분석이 정량적인 방법인지 여부를 조사하였다. C4BPB null 클론 및 야생형 세포로부터 분리한 유전체 DNA 시료를 다양한 비율로 혼합하고, PCR 증폭에 사용하였다. PCR 산물은 RGEN 유전형질 분석 및 T7E1 어세이에 동일하게 적용하였다 (도 25b). 예상한 대로, RGEN에 의한 DNA 절단은 야생형 대 돌연변이 비율과 비례하였다. 대조적으로, T7E1 어세이의 결과는 상기 비율에서 추론한 돌연변이 빈도와 저조하게 연관되었고, 상보적인 돌연변이 서열들이 서로 혼상화하여 동형이중가닥을 형성할 수 있는 상황에서, 특히 높은 돌연변이 %에서, 부정확하였다.

[0261] **9-9. RGEN-매개 RFLP 유전형질 분석을 이용한 돌연변이 마우스 파운더의 분석**

[0262] 본 발명자들은 RGEN-매개 RFLP 유전형질 분석 (줄여서 RGEN 유전형질 분석)을, 마우스 1 세포 배아에 TALENs를 주입하여 확립한 돌연변이 마우스 파운더의 분석에 적용하였다 (도 26A). 본 발명자들은 Pibf1 유전자에서 TALEN 표적 위치를 인식하는 RGEN을 설계하고 사용하였다 (표 10). 야생형 마우스 및 돌연변이 마우스에서 유전체 DNA를 분리하였고, PCR 증폭한 후에 RGEN 유전형질 분석에 적용하였다. RGEN 유전형질 분석은 1 내지 27-bp 결실의 범위로 다양한 돌연변이를 성공적으로 탐지하였다 (도 26B). T7E1 어세이와 다르게, RGEN 유전형질 분석은 +/- 및 -/- 파운더의 구별적인 탐지를 가능하게 하였다.

[0263] **9-10. 인간 세포에서 RGENs을 사용한 CCR5-특이적 ZFN으로 유도된 돌연변이의 탐지**

[0264] 또한, 본 발명자들은 RGEN을 사용하여 또 다른 클래스의 유전자 가위를 대표하는 CCR5-특이적 ZFN으로 인간세포에서 유도된 돌연변이를 탐지하였다 (도 27). 이러한 결과는 RGENs이 RGEN과 다른 뉴클레아제에 의해 유도된 돌연변이를 탐지할 수 있다는 것을 보여준다. 사실 본 발명자들은 RGENs이 비록 모두는 아닐지라도 대부분의 유전자 가위에 의해 유도되는 돌연변이를 탐지하도록 고안될 수 있을 것이라 기대한다. RGEN 유전형질 분석 어세이의 설계에 있어 제한점은 단지 평균적으로 4bp 당 한번 발생하는, Cas9 단백질에 의해 인식되는 PAM 서열에서 GG 또는 AG (상보적 가닥에서는 CC 또는 CT) 다이뉴클레오타이드 (dinucleotide)의 요구이다. crRNA 및 PAM 뉴클레오타이드에서 여러 염기의 시드 부위 (seed region) 내의 어디서라도 유도되는 인델 (indel)은 RGEN-촉매 DNA 절단을 방해할 것으로 예상된다. 확실히, 본 발명자들은 ZFN 또는 TALEN 위치의 대부분 (98%)에서 적어도 하나의 RGEN 위치를 밝혀내었다.

[0265] **9-11. RGEN을 이용한 다형성 또는 변이의 탐지 (detection of polymorphisms or variations using RGEN)**

[0266] 다음으로, 본 발명자들은 인간 백혈구 항원 B (a.k.a. MHC 클래스 I 단백질)를 암호화하는, 고 다형성 좌위 (highly polymorphic locus)인 HLA-B를 표적하는 새로운 RGEN을 설계하고 시험하였다 (도 28). HeLa 세포에 RGEN 플라스미드를 형질주입하였고, 유전체 DNA를 T7E1 및 RGEN-RFLP 분석에 동일하게 적용하였다. T7E1은 표적 위치와 인접한 서열 다형성에 기인한 위양성 밴드 (false positive band)를 만들었다 (도 25c). 하지만 예상한 대로, 유전자 파괴에 사용한 동일한 RGEN은 야생형의 PCR 산물을 완전히 분해하였지만, RGEN-형질주입 세포의 PCR 산물은 부분적으로 분해하여, 표적 위치에 RGEN-유도 indels의 존재를 암시하였다. 이러한 결과는 특히 관심 있는 세포에서 표적 유전자가 다형성 또는 변이를 갖는지 여부가 알 수 없을 때, RGEN-RFLP 분석이 T7E1 어세이에 대해 분명한 이점을 갖는다는 것을 보여준다.

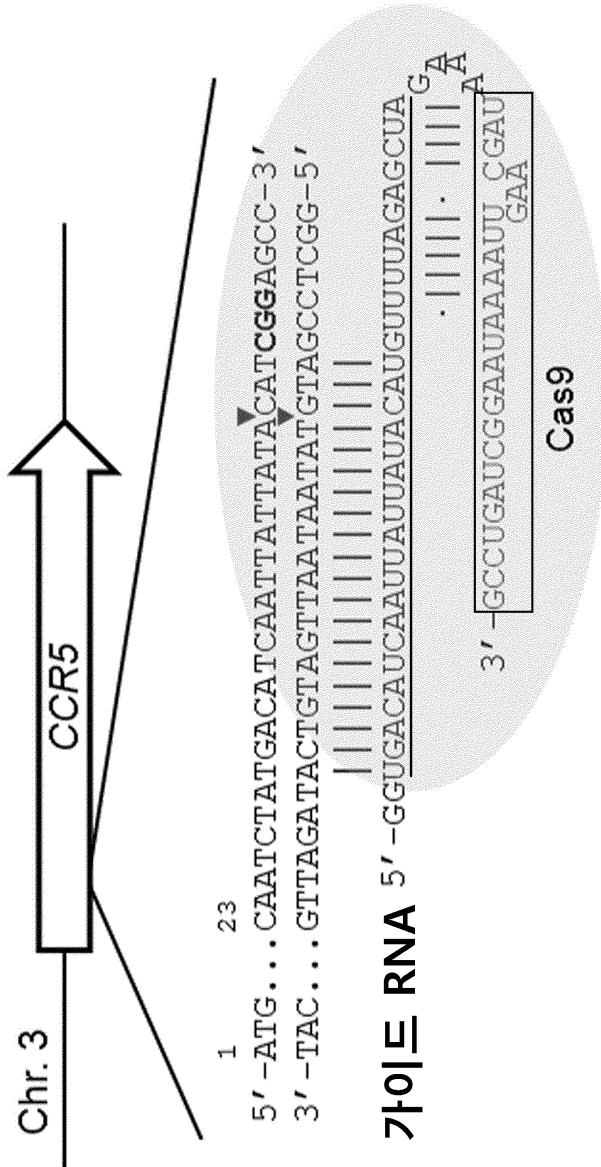
[0267] **9-12. RGEN-RFLP 분석을 통한 암에서 발견되는 반복 돌연변이 및 자연 발생 다형성의 탐지 (detection of recurrent mutations found in cancer and naturally-occurring polymorphisms through RGEN-RFLP analysis)**

[0268] RGEN-RFLP 분석은 유전자 가위-유도 돌연변이의 유전형질 분석을 뛰어넘는 응용분야를 갖는다. 본 발명자들은 RGEN 유전형질 분석을 사용하여 암에서 발견되는 반복 돌연변이 및 자연 발생 다형성을 탐지하고자 하였다. 본 발명자들은 베타-카테닌 (beta-catenin)을 암호화하는 발암 CTNNB1 유전자에서 기능 획득 3-bp 결실 (gain-of-function 3-bp deletion)을 가지는 인간 대장암 세포주, HCT116를 선택하였다. HCT116 세포에서 이형접합 유전형질과 비슷하게, HCT116 유전체 DNA로부터 증폭된 PCR 산물을 야생형-특이적 및 돌연변이-특이적 RGENs 모두를 이용하여 부분적으로 절단하였다 (도 29a). 날카롭게 대조하면, 오직 야생형 대립유전자만 갖는 HeLa 세포로부터 유래한 DNA로부터 증폭된 PCR 산물이 야생형-특이적 RGEN으로 완전히 분해되었고, 돌연변이-특이적 RGEN으로는 완전히 분해되지 않았다.

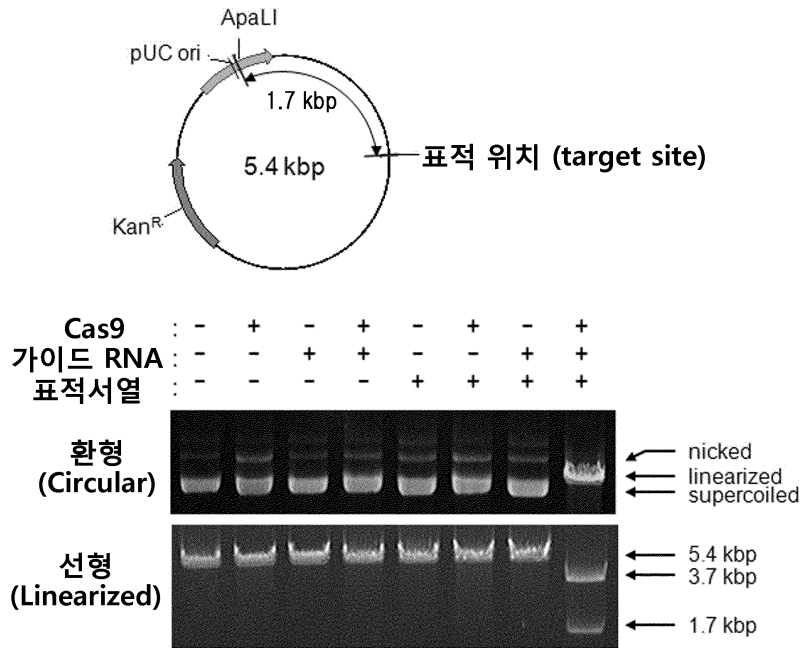
- [0269] 본 발명자들은 HEK293 세포가 HIV 감염의 필수적 공동-수용체를 암호화하는 CCR5 유전자에서 32-bp 결실 (de132)을 갖는다는 것을 주목하였다: 동형접합 de132 CCR5 캐리어는 HIV 감염에 면역성이 있다. 본 발명자들은 de132 대립유전자에 특이적인 하나의 RGEN 및 야생형 대립유전자에 특이적인 다른 RGEN을 설계하였다. 예상한 대로, 야생형-특이적 RGEN은 K562, SKBR3, 또는 HeLa 세포 (야생형 대조군으로 사용됨)로부터 수득한 PCR 산물을 완전히 분해하였지만, HEK293 세포로부터 수득한 PCR 산물은 부분적으로 분해하여 (도 30a), HEK293 세포에서 절단되지 않은 de132 대립유전자의 존재를 확인하였다. 그러나 예상치못하게, de132-특이적 RGEN은 HEK293 세포로부터의 PCR 산물과 같이 효과적으로 야생형 세포로부터 유래한 PCR 산물을 절단하였다. 흥미롭게도, 이러한 RGEN은 온-타겟 위치에서 바로 옆 다운스트림 (downstream)에 위치하는 단일 염기 불일치 (single-base mismatch)를 가지는 오프-타겟을 가지고 있었다 (도 30). 상기 결과는 RGENs을 자연 발생 indels의 탐지에 사용할 수 있지만, 단일 뉴클레오타이드 다형성 또는 오프-타겟 효과에 따른 점 돌연변이를 갖는 서열은 구별할 수 없다는 것을 제시한다.
- [0270] RGENs을 이용하여 발암 단일 뉴클레오타이드 변이 (oncogenic single-nucleotide variant)를 유전형질 분석하기 위해, 본 발명자들은 완벽하게 일치하는 RNA 대신 단일 염기가 불일치하는 가이드 RNA를 이용하여 RGEN 활성을 약화시켰다. 야생형 서열 또는 돌연변이 서열에 완벽하게 일치하는 가이드 RNA를 갖는 RGENs은 두 서열을 모두 절단하였다 (도 31a 및 32a). 대조적으로, 단일 염기가 불일치하는 가이드 RNA를 포함하는 RGENs은 두 서열을 구별하였고, 인간 암 세포주에서 KRAS, PIK3CA, 및 IDH1 유전자에 있는 세 개의 반복 발암 점 돌연변이의 유전형질 분석을 가능케 하였다 (도 29b 및 도 33a, b). 또한, 본 발명자들은 NAG PAM 서열을 인식하는 RGENs을 사용하여 BRAF 및 NRAS 유전자에서 점 돌연변이를 탐지할 수 있었다 (도 33c, d). 본 발명자들은 RGEN-RFLP를 사용하여 인간 및 다른 유전체에서 전부는 아니나 거의 모든 돌연변이 또는 다형성에 대한 유전형질 분석할 수 있다고 믿는다.
- [0271] 상기 데이터는 RGEN이 다양한 서열 변이에서 간단하고 강력한 RFLP 분석을 사용하기 위한 플랫폼을 제공함을 제시한다. 리프로그래밍 표적 서열의 높은 유연성으로, RGEN을 사용하여 질병 연관 반복 돌연변이, 환자의 약물 반응 관련 유전자형과 또한 세포에서의 유전자 가위에 의해 유도된 돌연변이와 같은 다양한 유전적 변이 (단일 뉴클레오타이드 변이, 작은 삽입/결실, 구조적 변이)를 검출할 수 있다. 여기서, 본 발명자들은 RGEN 유전형질 분석을 사용하여 세포 및 동물에서 유전자 가위에 의해 유도되는 돌연변이를 검출하였다. 원칙적으로, 자연 발생 변이 및 돌연변이를 특이적으로 탐지하고 절단하는 RGENs을 또한 사용할 수 있다.
- [0272] 상기 설명에 기초하여, 다음 청구항에 정의된 발명의 기술적 사상 또는 본질적 특징을 벗어남이 없이 본 발명을 수행하는데 여기에 기술된 발명의 양태에 대한 다양한 대안이 사용될 수 있다는 것을 당업자는 이해해야 한다. 이와 관련하여, 전술한 실시예는 단지 예시의 목적이며, 본 발명은 이들 실시예에 의해 한정되는 것이 아니다. 본 발명의 범주는 다음 청구항의 의미 및 범위 또는 그와 동등한 개념으로부터 유래한 변형 또는 변형된 형태를 모두 포함하는 것으로 이해되어야 한다.

도면

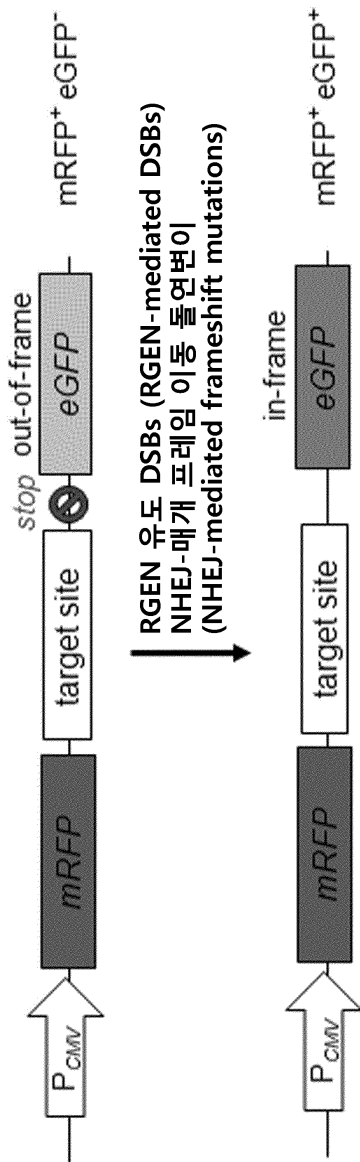
도면1a



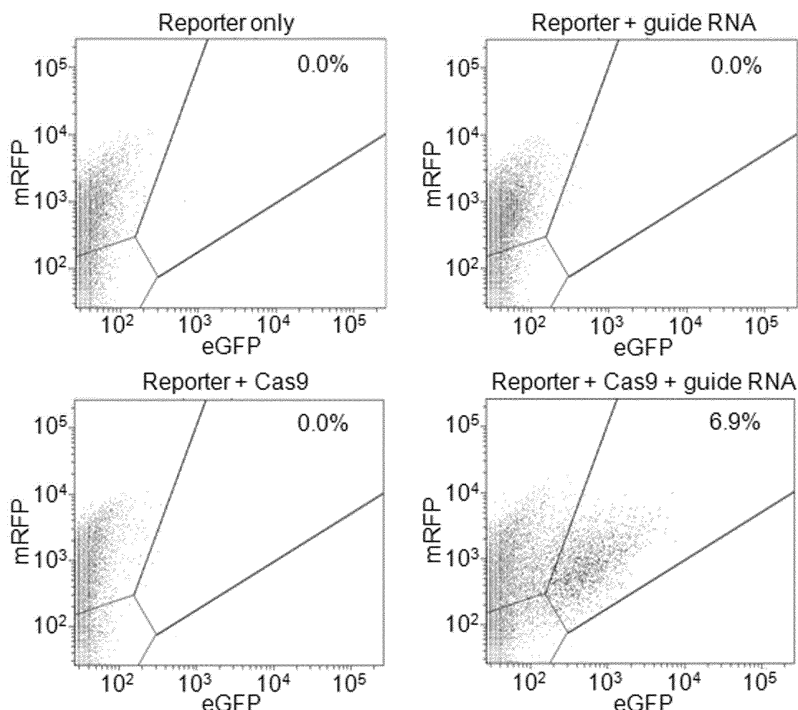
도면1b



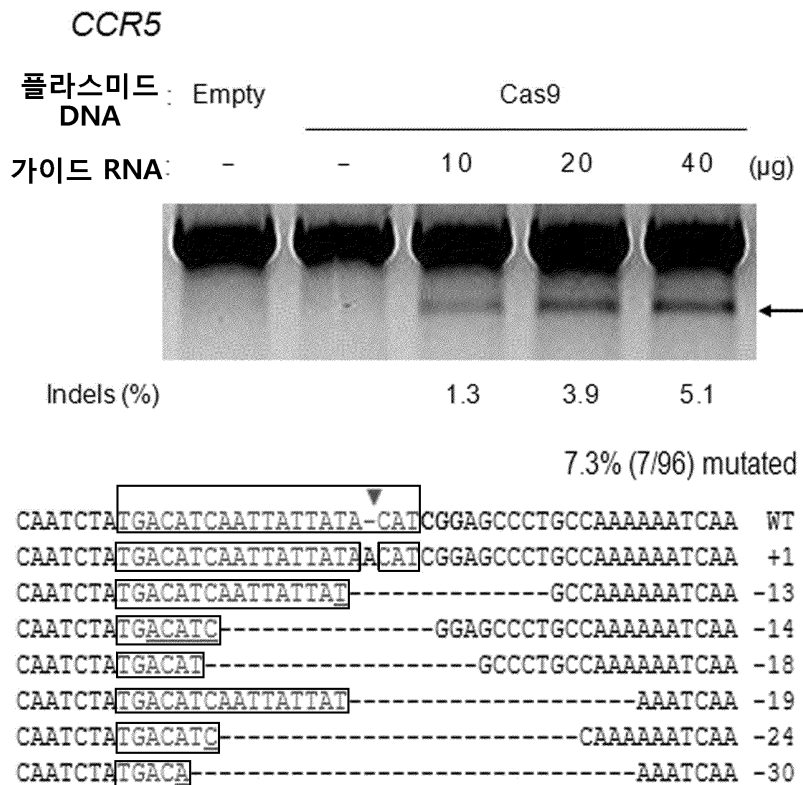
도면2a



도면2b



도면3a



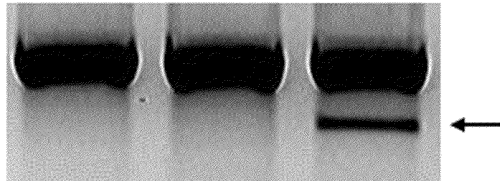
도면3b

C4BPB

플라스미드 DNA : Empty

Cas9

가이드 RNA : - - +



Indels (%)

14

8.3% (4/48) mutated

```

TATGTGC AATGACCACTACATCCT--CAAGGGCAGCAATCGGAGCCAG  WT
TATGTGC AATGACCACTACATCCTT-CAAGGGCAGCAATCGGAGCCAG  +1
TATGTGC AATGACCACTACATCCTCTCAAGGGCAGCAATCGGAGCCAG  +2
TATGTGC AATG-----GAGCCAG  -30
TATGTGC AATGAC-----180 bp----- -180
    
```

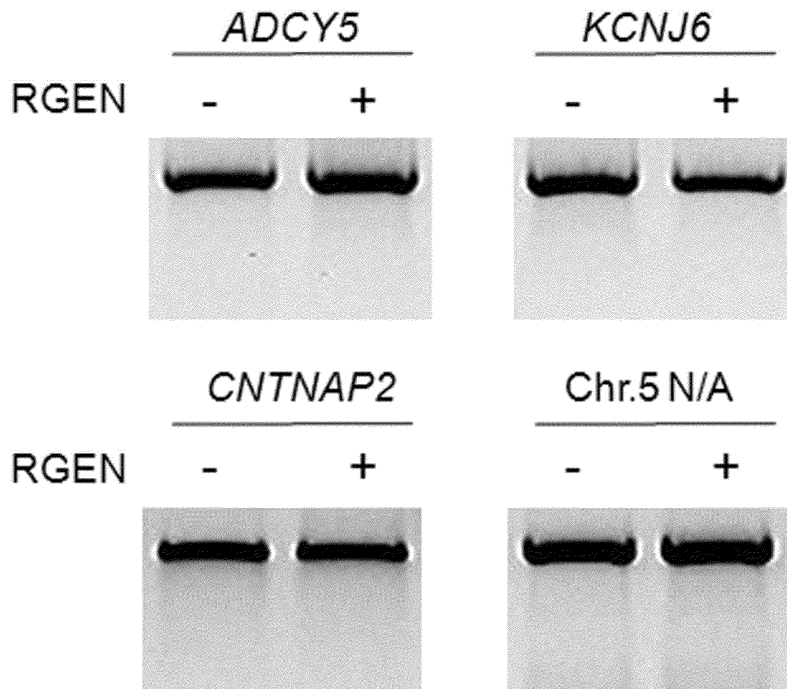
도면4a

```

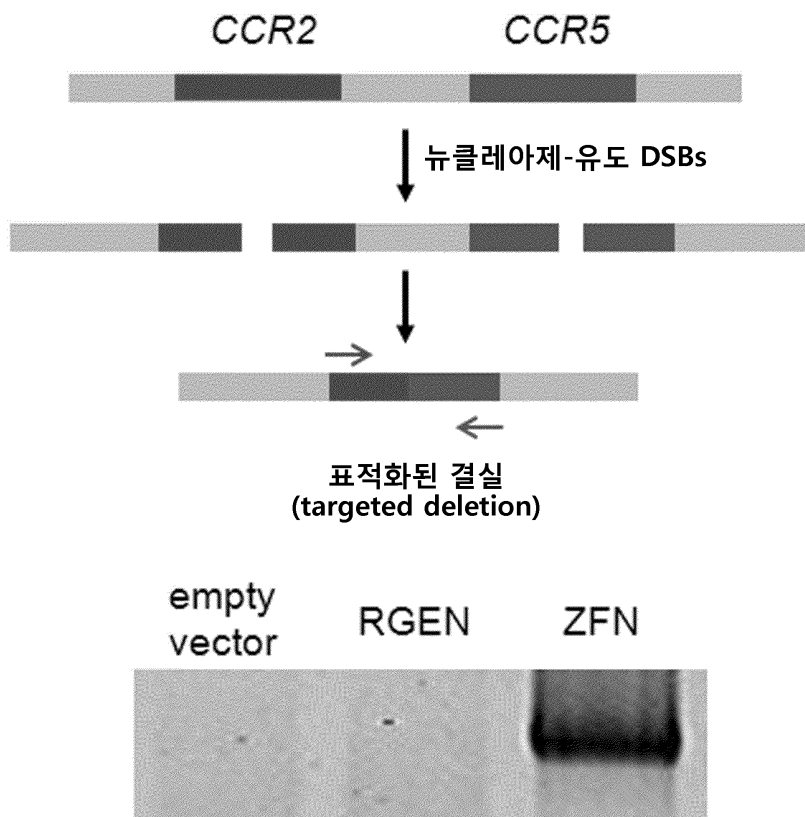
TGACATCAATTATTATACATCGG  CCR5

TGACATCAATTATTATAgATgGa  ADCY5
TGACATCAcTTATTATgCATgGG  KCNJ6
TGACATaAATTATTcTACATgGG  CNTNAP2
TGAaATCAATTATcATAgATCGG  Chr. 5 N/A
    
```

도면4b

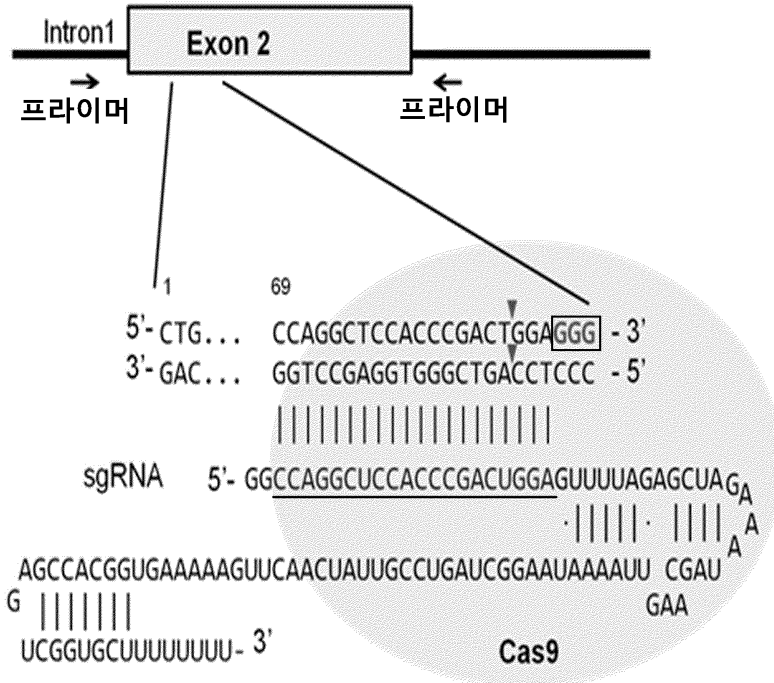


도면4c

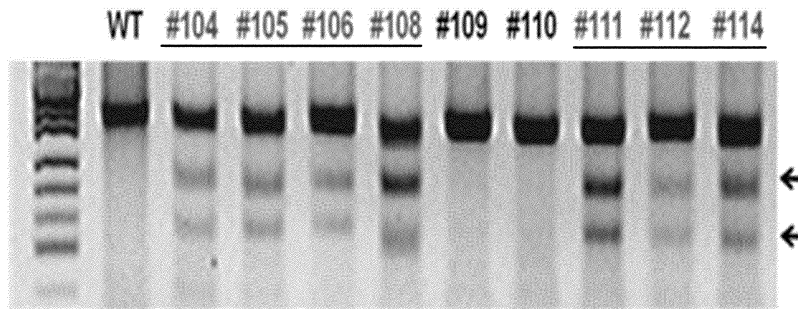


도면5a

Foxn1



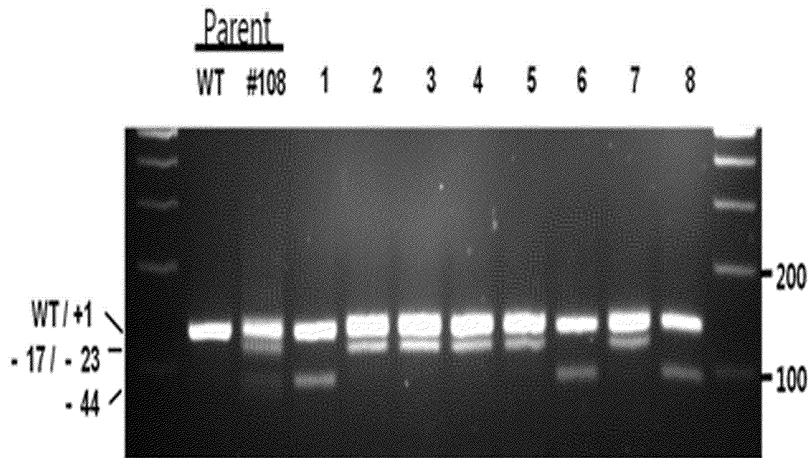
도면5b



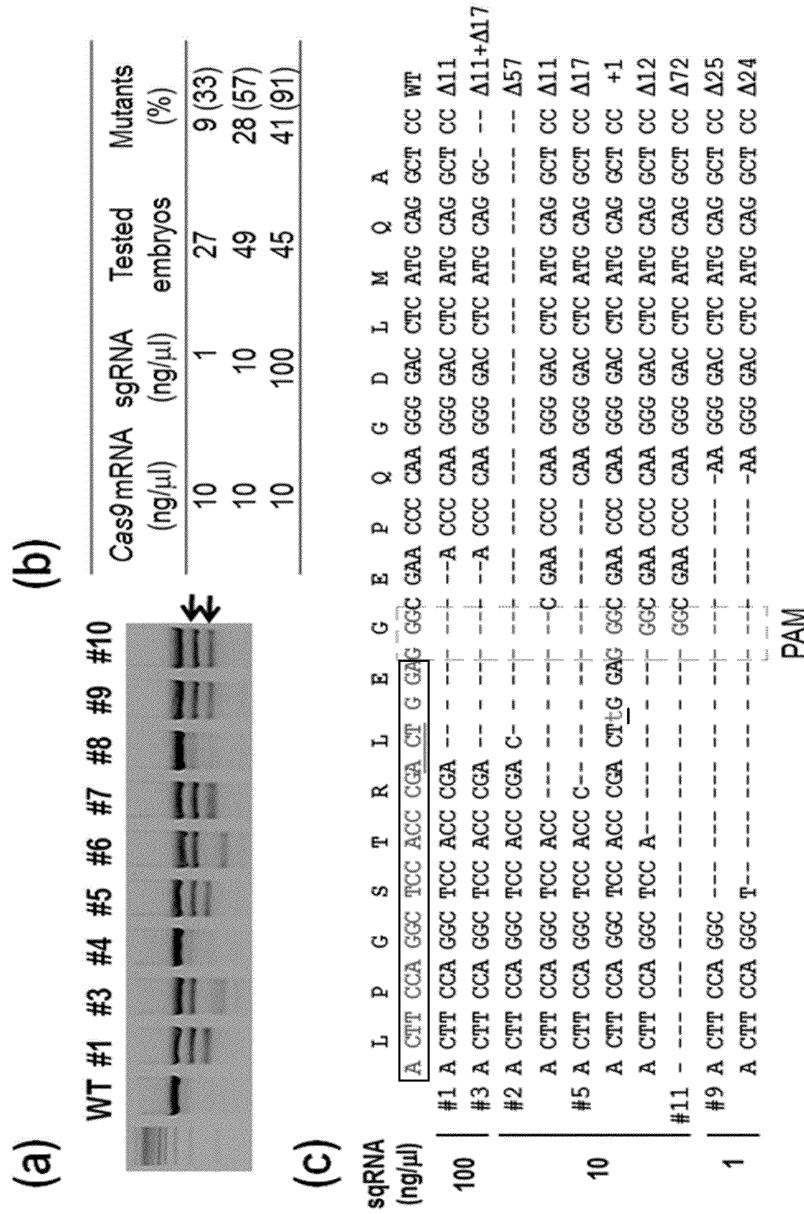
도면5c

WT				
ACTTCCAGGCTCCACCCGACT	▼	GGAGGGCGAACCCCAAGGGGACCTCATGCAGG		
#108				
ACTTCCAGGCTCC-----	▼	-----44bp-----	-44	(X6)
ACTTCCAGGCTCCACCCGAC-		-----CTCATGCAGG	-23	(X2)
ACTTCCAGGCTCCACCC-		-----CAAGGGGACCTCATGCAGG	-17	(X1)
ACTTCCAGGCTCCACCCGACT	▼	GGAGGGCGAACCCCAAGGGGACCTCATGCAGG	+1	(X1)
#111				
ACTTCCAGGCTCCACCCGACT	▼	GGAGGGCGAACCCCAAGGGGACCTCATGCAGG	+1	(X2)
ACTTCCAGGCTCCACCCG-		-----AACCCCAAGGGGACCTCATGCAGGG	-11	(X6)
#114				
ACTTCCAGGCTCCACCCGACT	▼	-----CACTATCTTCTGGGCTCCTCCATGTC	-6	+25 (X3)
ACTTCCAGGCTCCACCC-		-----CAAGGGGACCTCATGCAGG	-17	(X6)
ACTTCCAGGCTCCACCCGAC-		-----GAACCCCAAGGGGACCTCATGCAGG	-8	(X1)

도면5d

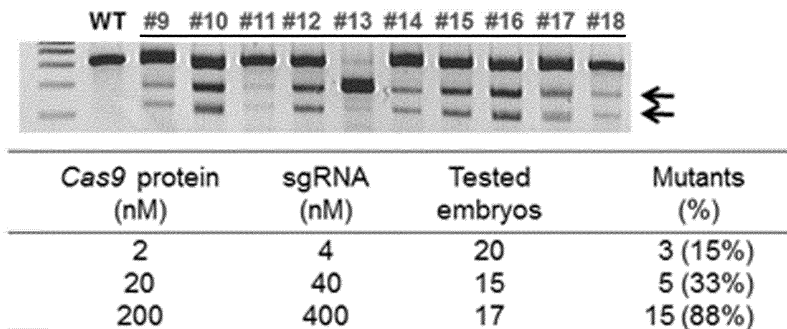


도면6



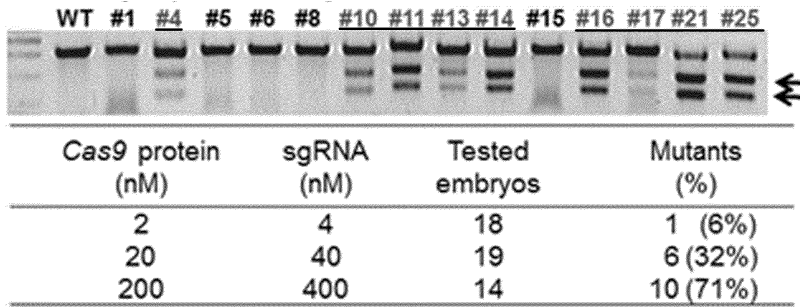
도면7a

전핵 주입 (pronuclear injection)



도면7b

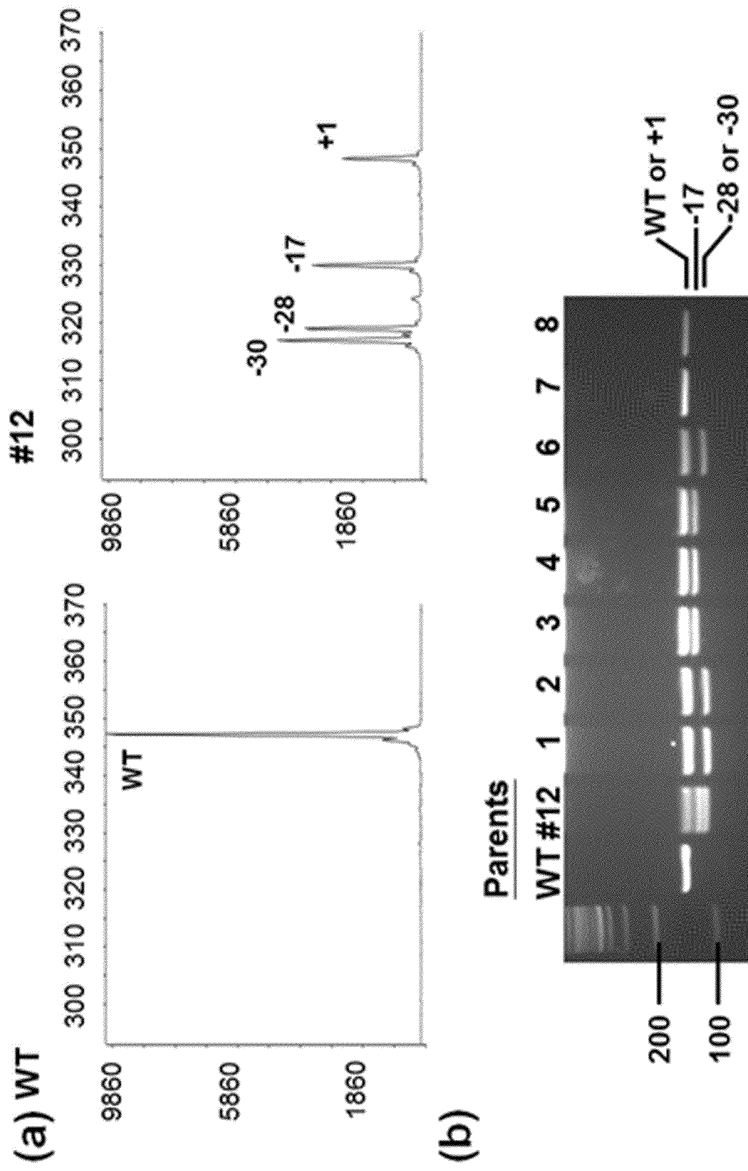
세포질 내 주입 (intra-cytoplasmic injection)



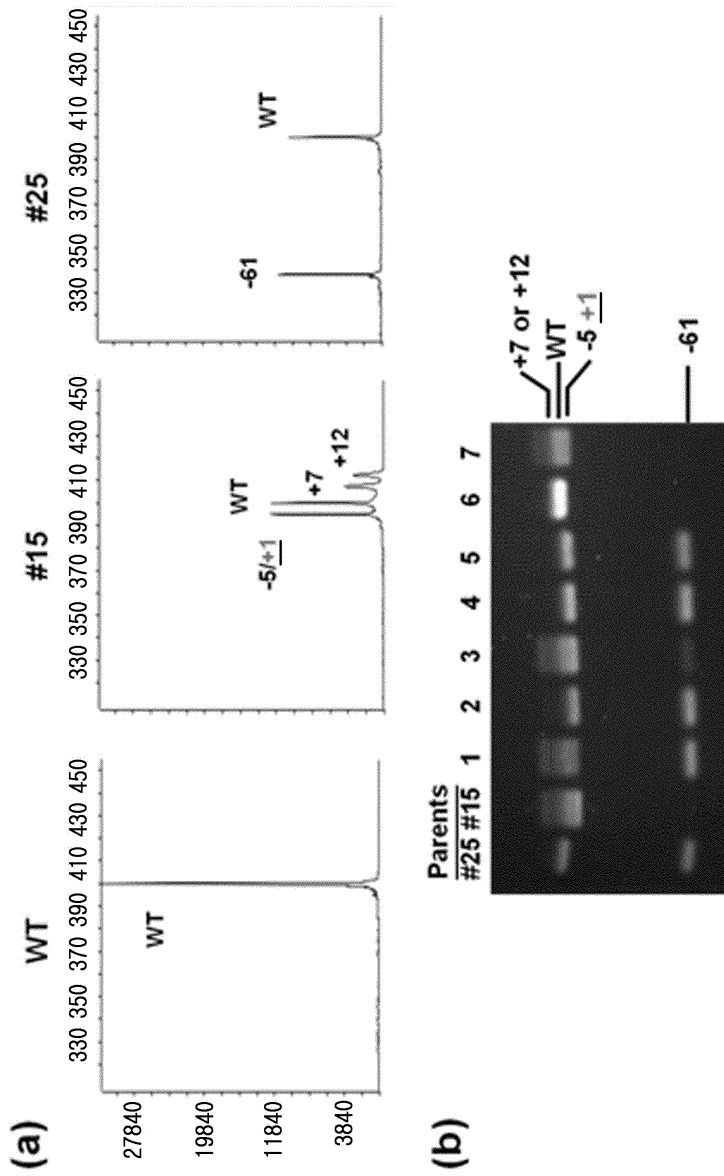
도면7c

서열	Indels	배아 번호 (Embryo No.)
ACTTCCAGGCTCCACCCGACTGGA <u>GGGCGA</u> ACCCCAAGGGGACCTCATGCAG	WT	
ACTTCCAGGCGAACCCC-----AAGGGGACCTCATGCAG	Δ18	2
ACTTCCAGGCTCCAC-----AAGGGGACCTCATGCAG	Δ20	1
ACTTCCAGGCTCCACCC-----AAGGGGACCTCATGCCC	Δ19	1
ACTTCCAGGCTCCACCC-----CAAGGGGACCTCATGCAG	Δ17	1
ACTTCCAGGCTCCACCCGA-----ACCCCAAGGGGACCTCATGCAG	Δ11	3
ACTTCCAGGCTCCACCCGA <u>AA</u> --GGAGGGCGAACCCCAAGGGGACCTCATGCA	Δ3+1	1
ACTTCCAGGCTCCACCCGACT-- <u>AGGGCGA</u> ACCCCAAGGGGACCTCATGCAG	Δ2	1
ACTTCCAGGCTCCACCCGACTGG <u>GAGGGCGA</u> ACCCCAAGGGGACCTCATGCA	+1	1
ACTTCCAGGCTCCACCCGACT <u>TGGAGGGCGA</u> ACCCCAAGGGGACCTCATGCA	+1	10
ACTTCCAGGCTCCACCCGA----- <u>GGCGA</u> ACCCCAAGGGGACCTCATGCAG	Δ6	1
ACTTCCAGGCTCCACCCGA----- <u>GGGCGA</u> ACCCCAAGGGGACCTCATGCAG	Δ5	2
ACTTCCAGGCTCCACC-----TCATGCAG	Δ28	1
----- <u>AGGGCGA</u> ACCCCAAGGGGACCTCATGCAG	Δ126	1
Total		26

도면8



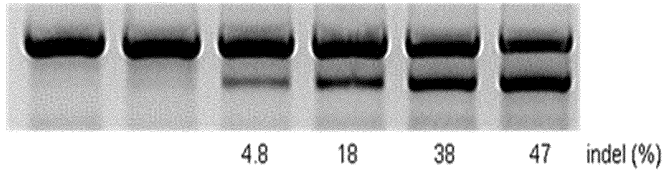
도면9



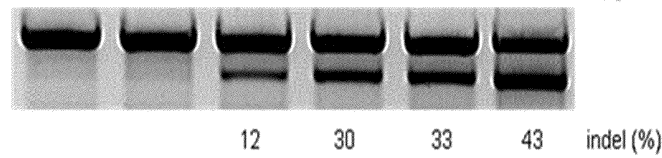
도면10a

CCR5#4

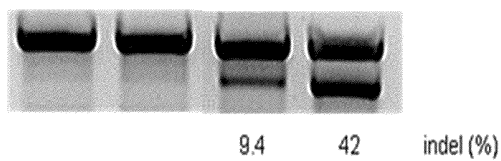
sgRNA	-	100 (29)	10 (2.9)	30 (8.8)	100 (29)	50 (15)	µg (µM)
Cas9 단백질	225 (14)	-	22.5 (1.4)	75 (4.5)	225 (14)	-	µg (µM)
Cas9 플라스미드	-	-	-	-	-	5	µg



sgRNA	-	100 (29)			50 (15)	µg (µM)	
Cas9 단백질	225 (14)	-	22.5 (1.4)	75 (4.5)	225 (14)	-	µg (µM)
Cas9 플라스미드	-	-	-	-	-	5	µg



crRNA	-	40 (29)	20 (15)	µg (µM)	
tracrRNA	-	80 (29)	40 (15)	µg (µM)	
Cas9 단백질	225 (14)	-	225 (14)	-	µg (µM)
Cas9 플라스미드	-	-	-	5	µg



도면10b

CCR5#4

Cas9 단백질	45 (14)	-	15 (4.5)					-	µg (µM)
sgRNA	-	60 (87)	1.3 (1.9)	3.2 (4.6)	8 (12)	20 (29)	50 (73)	-	µg (µM)
sgRNA 플라스미드	-	-	-	-	-	-	-	1	µg
Cas9 플라스미드	-	-	-	-	-	-	-	1	µg



Indel (%) 2.7 8.3 33 51 57 32

Cas9 단백질	45 (14)	-	4.5 (1.4)	15 (4.5)	µg (µM)
sgRNA	-	60 (87)	6 (8.7)	20 (29)	µg (µM)



Indel (%) 27 51

도면10c

CCR5

```

CAATCTATGACATCAATTATTATA-CATCGGAGCCCTGCCAAAAAATCAA  WT
CAATCTATGACATCAATTATTAT-----CGGAGCCCTGCCAAAAAATCAA  -4
CAATCTATGACATCAATTAT-----CATCGGAGCCCTGCCAAAAAATCAA  -4
CAATCTATGACATCAATTAT-----CGGAGCCCTGCCAAAAAATCAA  -7
CAATCTATGACATCAATTATTAT--CATCGGAGCCCTGCCAAAAAATCAA  -1
CAATCTATGACATCAATTATTATAACATCGGAGCCCTGCCAAAAAATCAA  +1
CAATCTATGACAA-----GAGCCCTGCCAAAAAATCAA  -17,+1
    
```

도면10d

ABCC11

Cas9 단백질	-	15 (4.5)	µg (µM)
sgRNA	-	20 (29)	µg (µM)



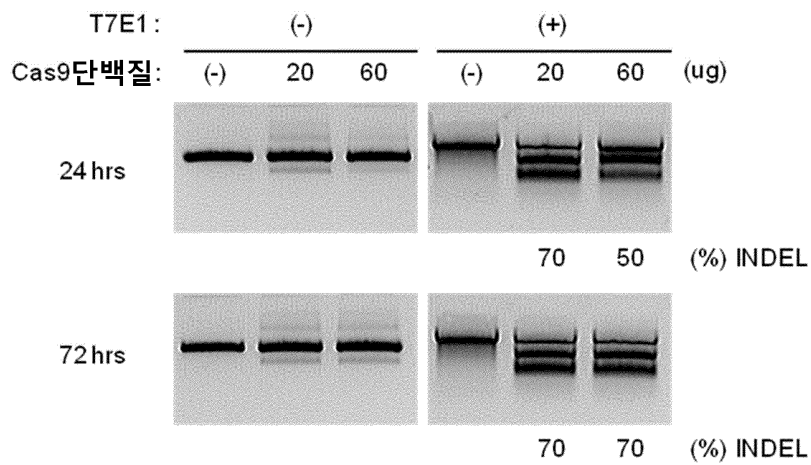
Indel (%) 35

도면10e

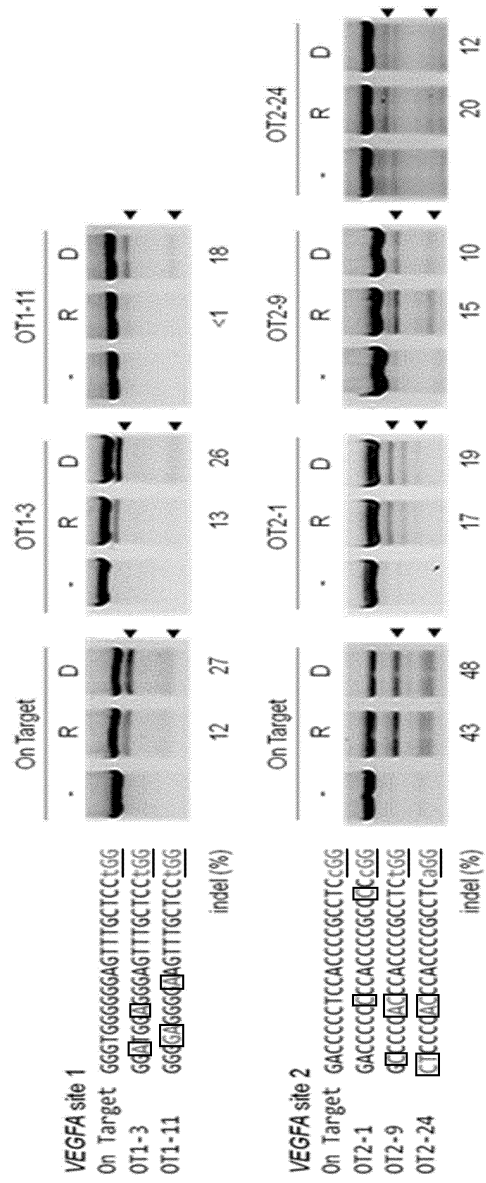
ABCC11

TTCTCAAGGCAGCATCATACTTCC[▼]CCACGGTGGGACAGCTGCCCTCCCTGG WT
 TTCTCAAGGCAGCATCATACTTCC-----CTGGGACAGCTGCCCTCCCTGG -6
 TTCTCAAGGCAGCATCATACTTC---CACGGTGGGACAGCTGCCCTCCCTGG -3
 TTCTCAAGGCAGC-----TGCCCTCCCTGG -29
 TTCTCAAGGCAGCATCATACTT-----CCCTCCCTGG -20
 TTCTCAAGGCAGCATCATACTT-----CCCTCCCTGG -20
 TTCTC----- -256

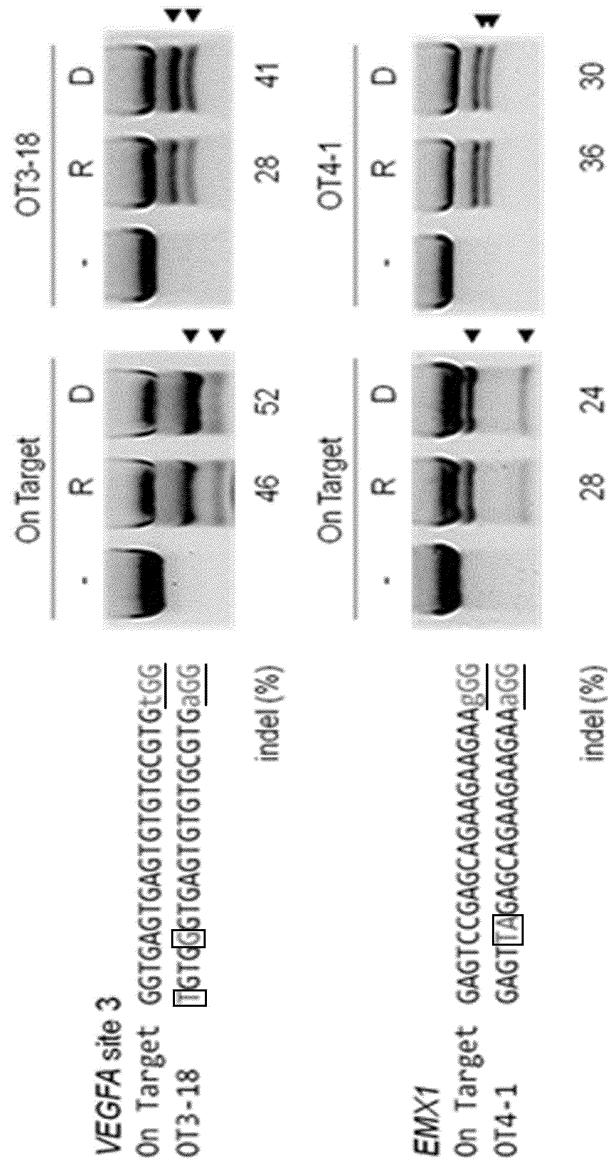
도면11



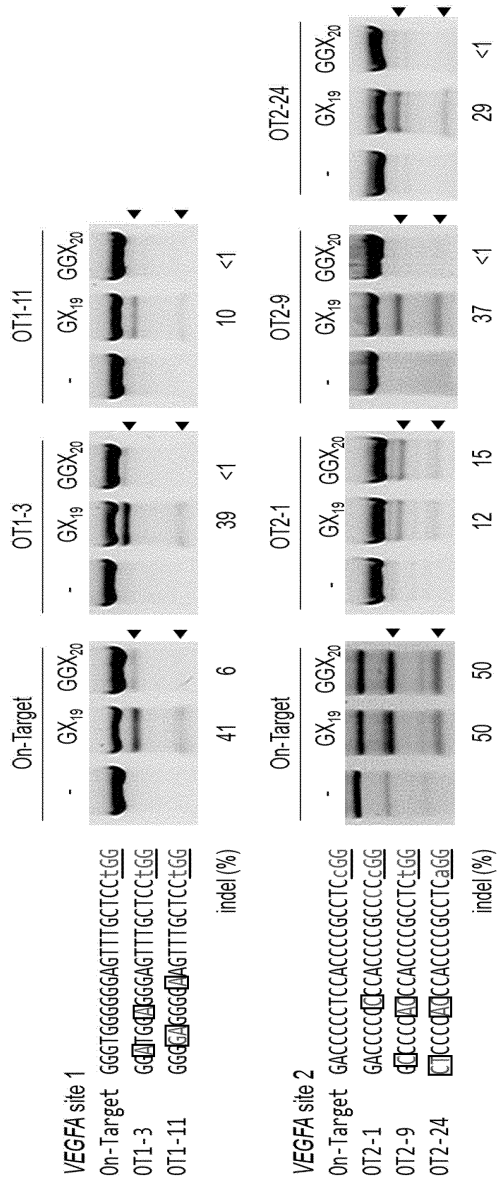
도면14a



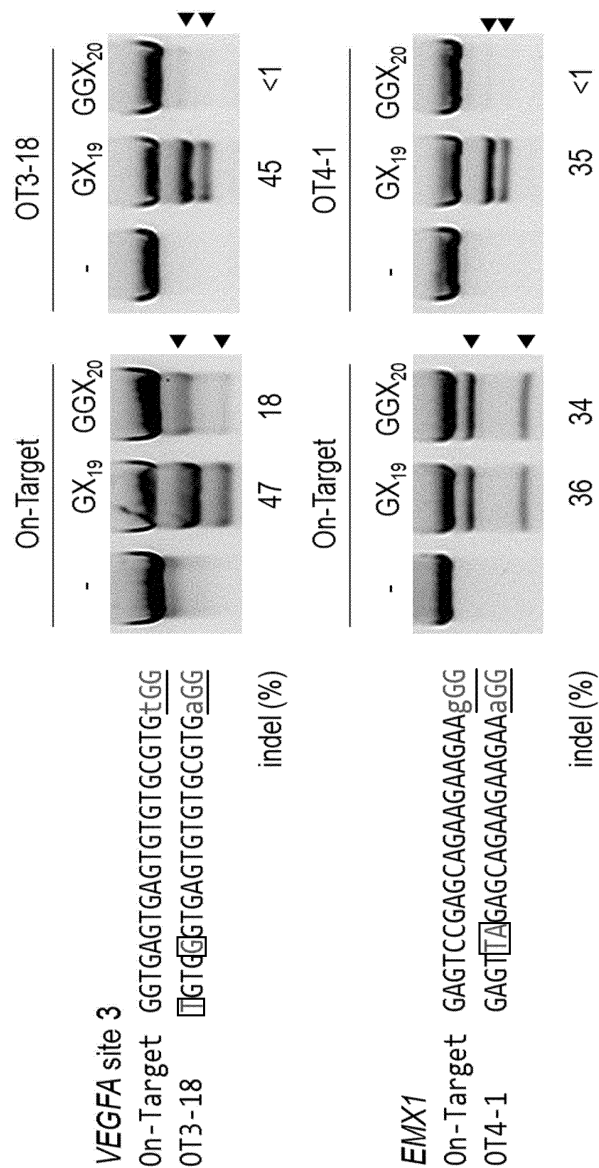
도면14b



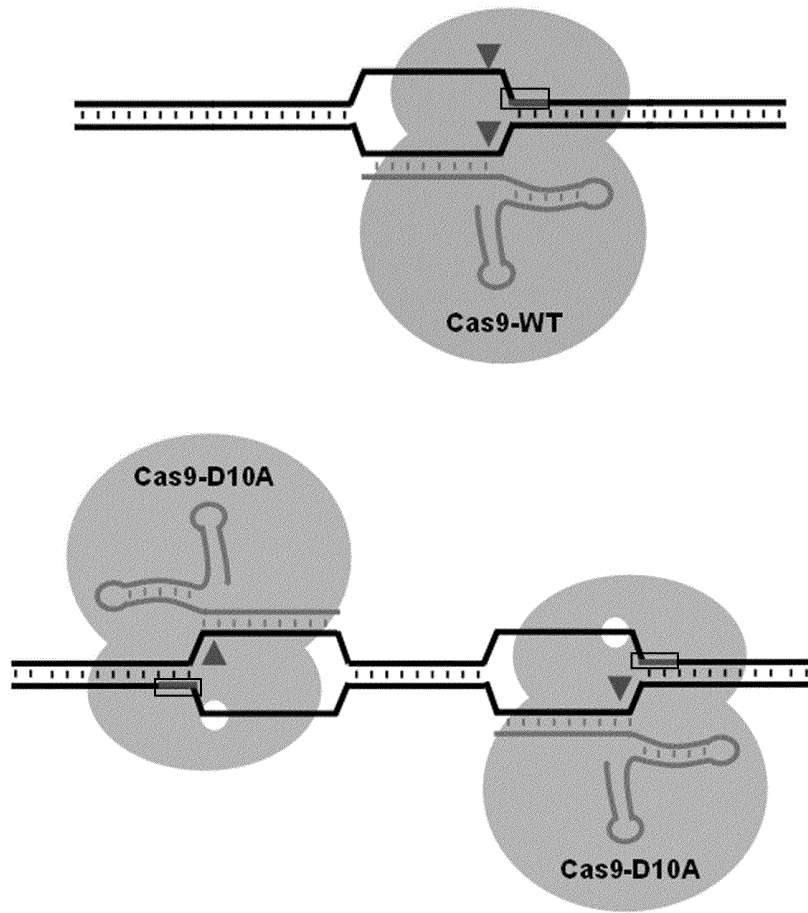
도면15a



도면15b

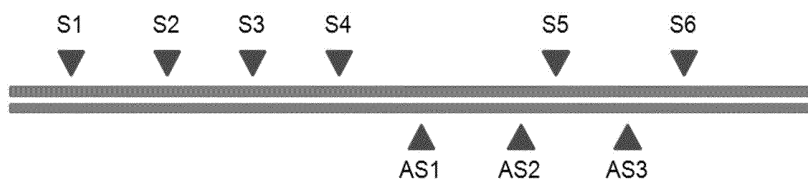


도면16a

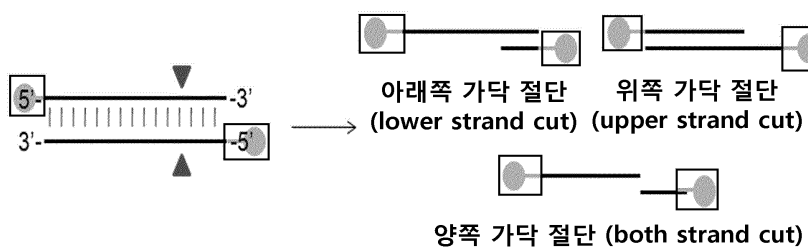


도면16b

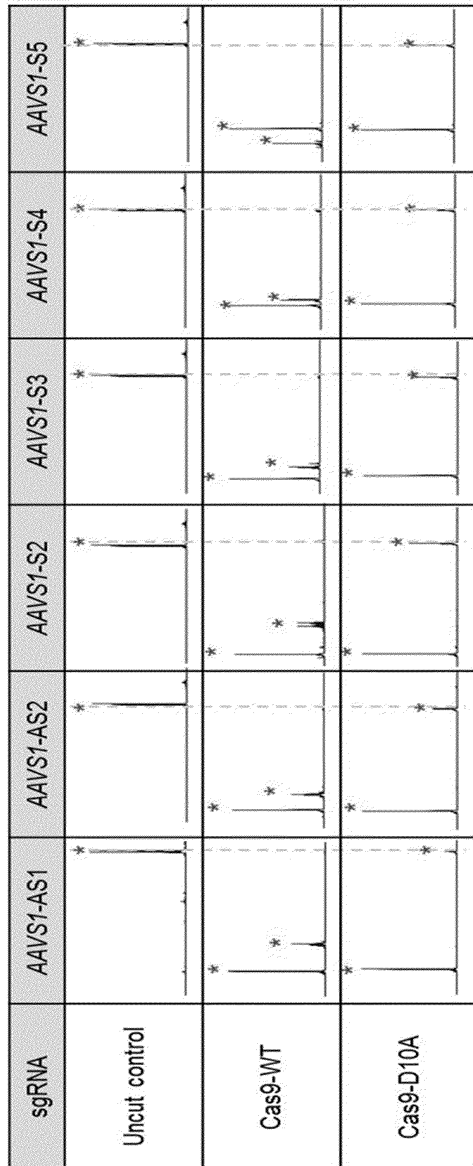
AAVS1 locus



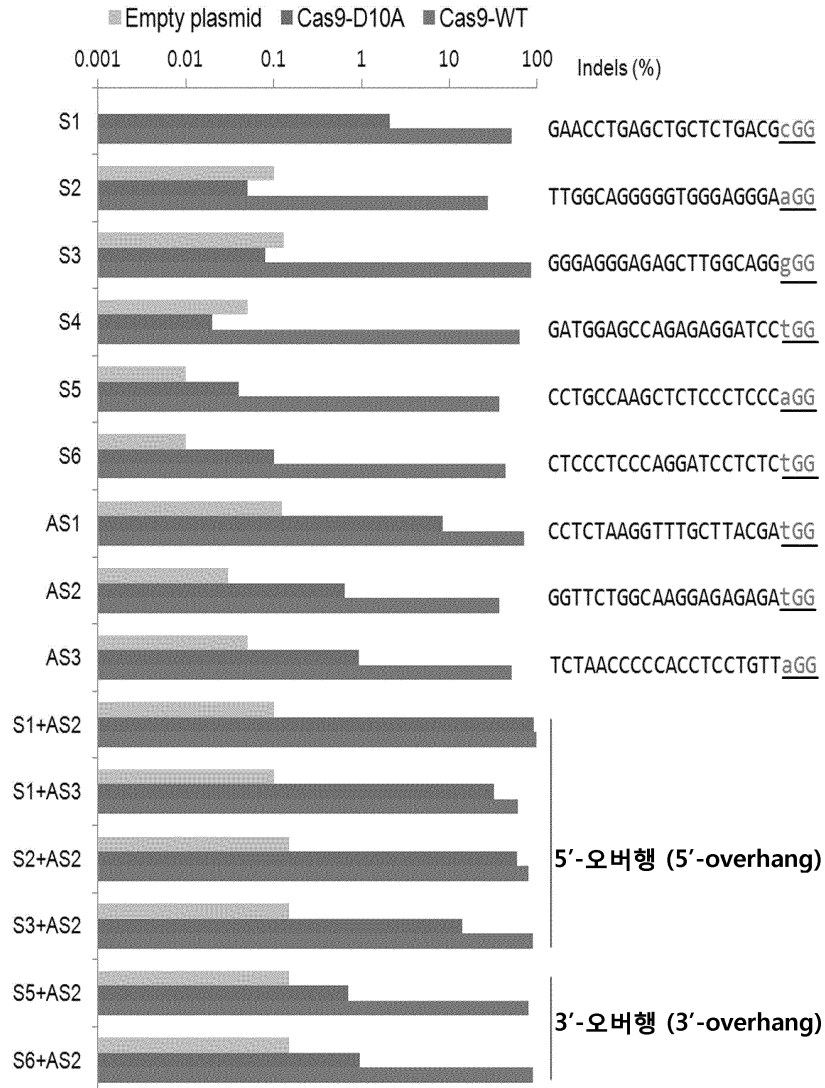
도면16c



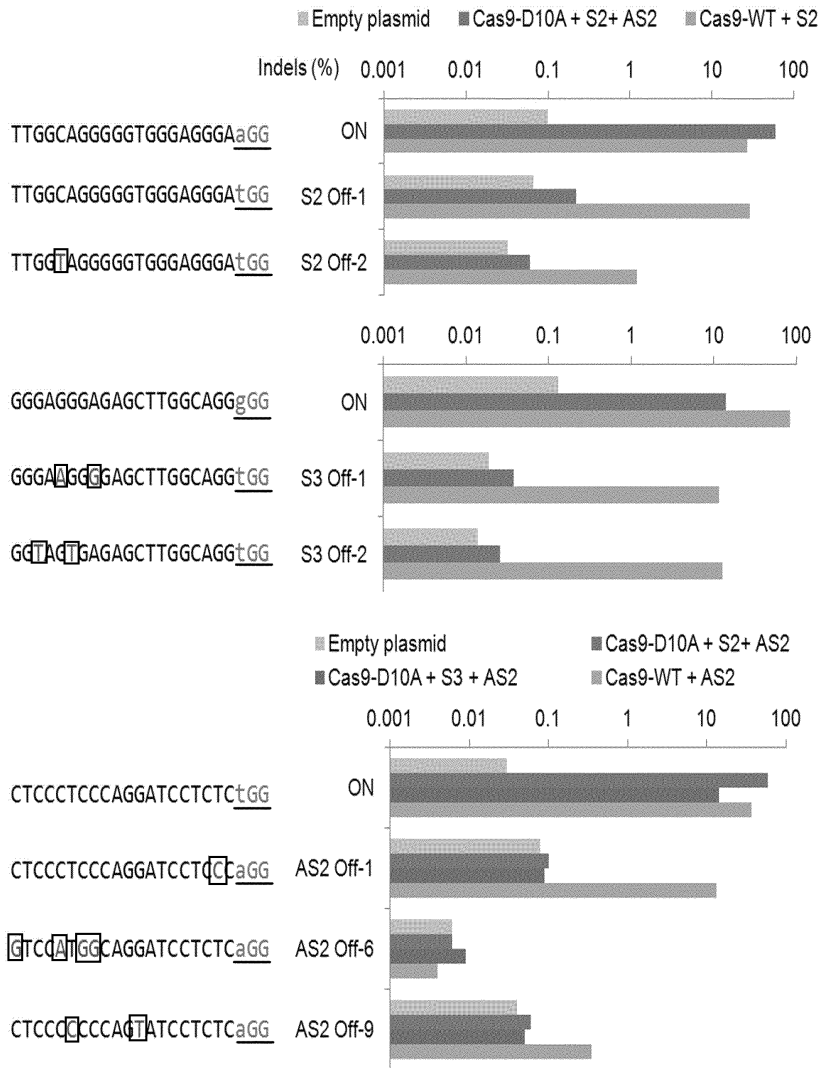
도면16d



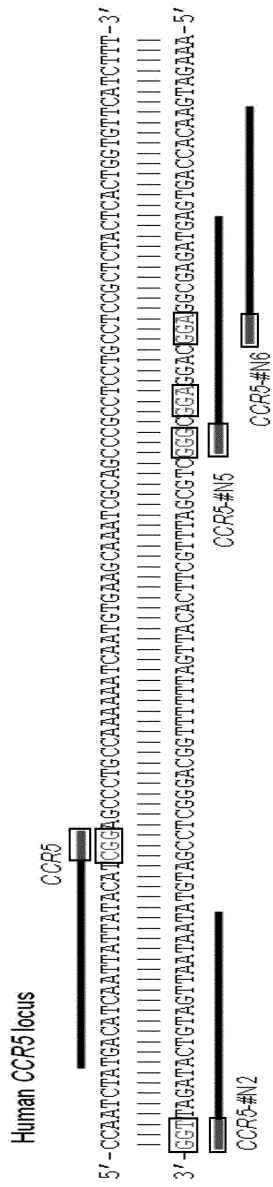
도면17a



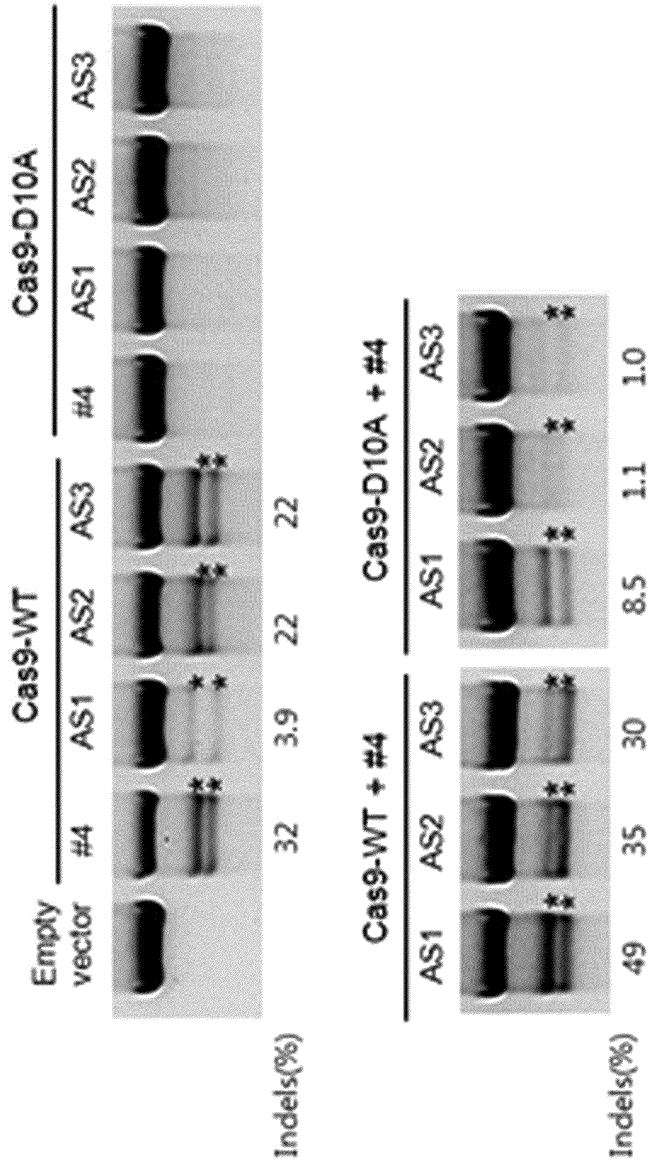
도면17b



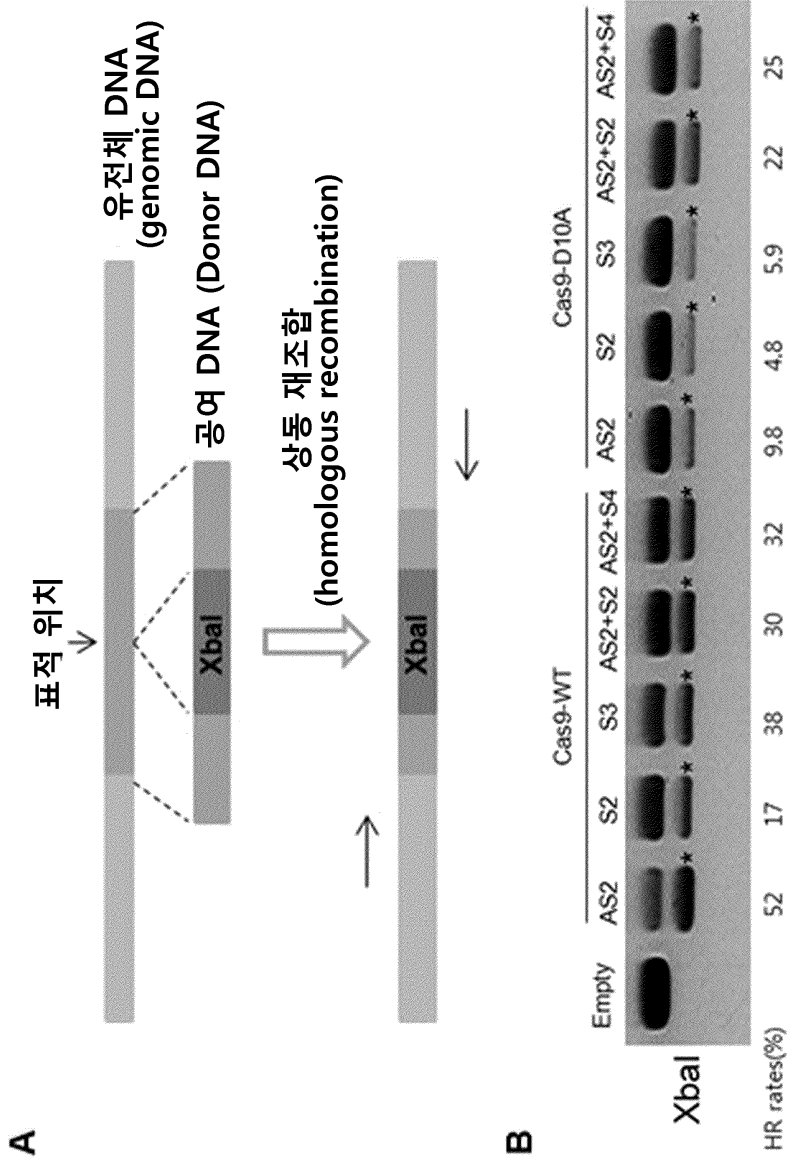
도면18a



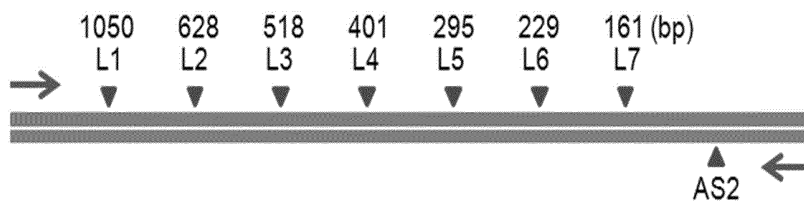
도면18b



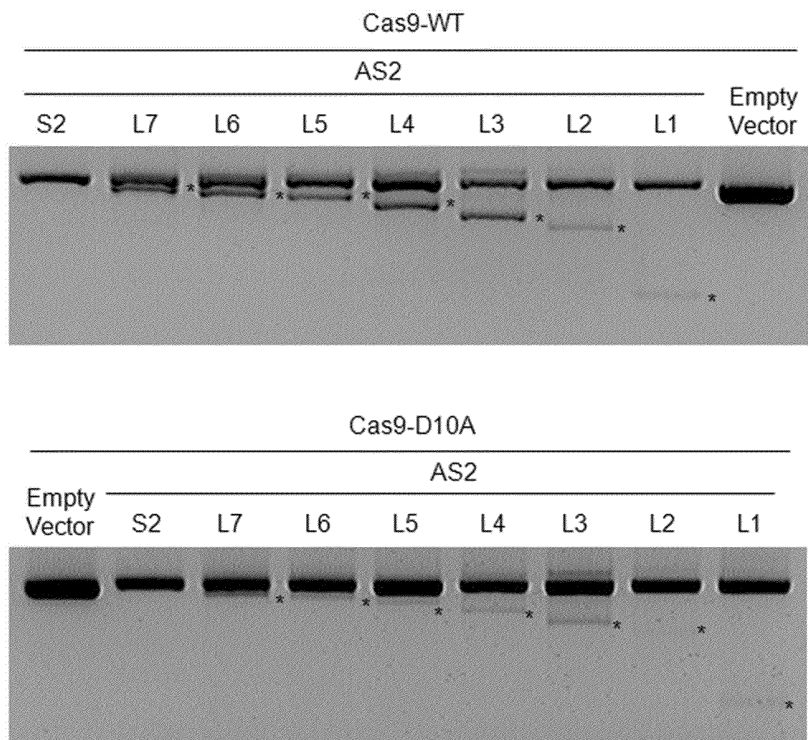
도면19



도면20a



도면20b



도면20c

Cas9-WT AS2 + L1 (~1050bp 결실)

```

ggccggggaatcaagagtcacccagagcagacagtgaccaaccattccctggttt...//...agctctccctccacaggaatccctctccatcgtaagcaaaccttagaagttctggcaaggagagagatg WT
ggccggggaatcaagagtcacccag-----TGACC AACCATCctt-----gtaagcaaaccttagaagttctggcaaggagagagatg
ggccggggaatcaagagtcacccag-----gaa
ggccggggaatcaagagtcacccagA-----CCCTCTATGcttccatcgtaagcaaaccttagaagttctggcaaggagagagatg
ggccggggaatcaagagtcaccc-----taacag
ggccggggaatcaaga-----cgtTGcttccatcgtaagcaaaccttagaagttctggcaaggagagagatg
ggc-----tccatcgtaagcaaaccttagaagttctggcaaggagagagatg
ggccggggaatcaagagtcacccagA-----CTCTATGcttccatcgtaagcaaaccttagaagttctggcaaggagagagatg

```

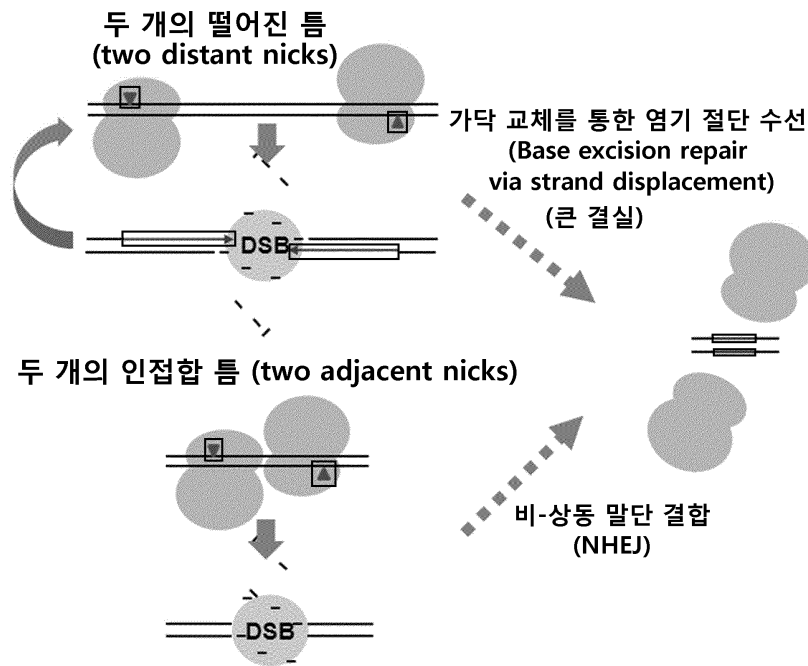
Cas9-D10A AS2 + L1 (~1050bp 결실)

```

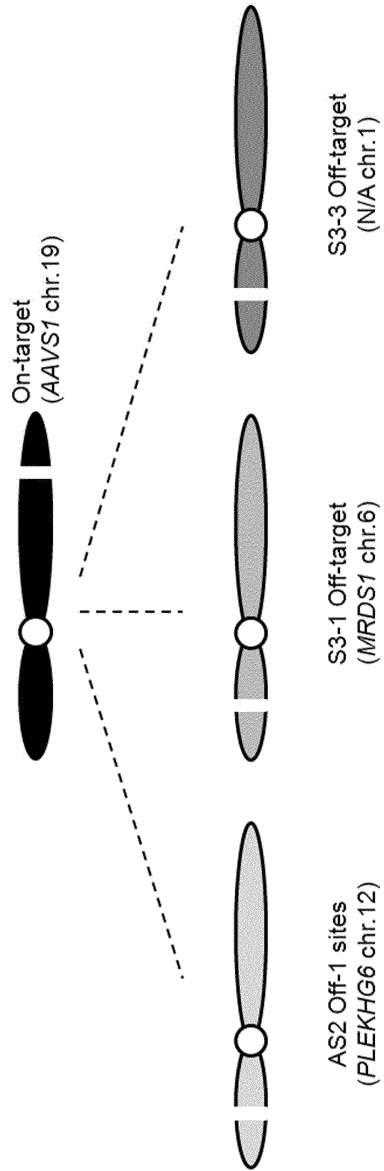
ggccggggaatcaagagtcacccagagcagacagtgaccaaccattccctggttt...//...agctctccctccacaggaatccctctccatcgtaagcaaaccttagaagttctggcaaggagagagatg WT
ggccggggaatcaagagtcacccagagcagacagtgaccaaccattccctggttt...//...gtaagcaaaccttagaagttctggcaaggagagagatg
gg-----tccatcgtaagcaaaccttagaagttctggcaaggagagagatg
ggccggggaatcaagagtcacccA-----tccatcgtaagcaaaccttagaagttctggcaaggagagagatg
ggccggggaatcaagagtcacccagA-----tctccatcgtaagcaaaccttagaagttctggcaaggagagagatg
gg-----CTATGcttccatcgtaagcaaaccttagaagttctggcaaggagagagatg
ggccggggaatcaagagtcacccagagcagacagtgaccaaccattcc-----atca
ggccggggaatcaagagtc-----tcgtaagcaaaccttagaagttctggcaaggagagagatg XZ

```

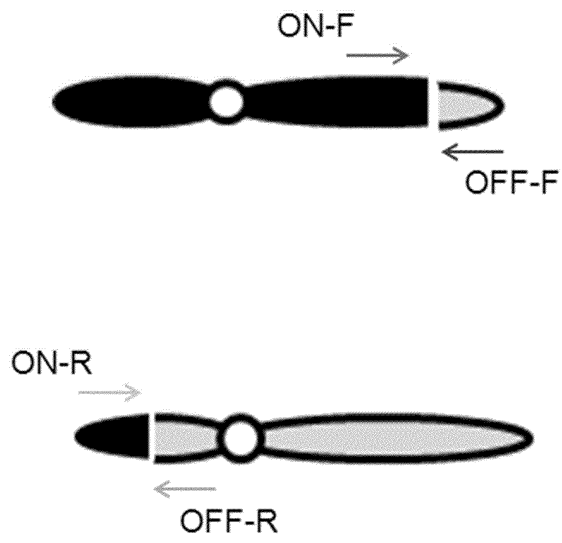
도면20d



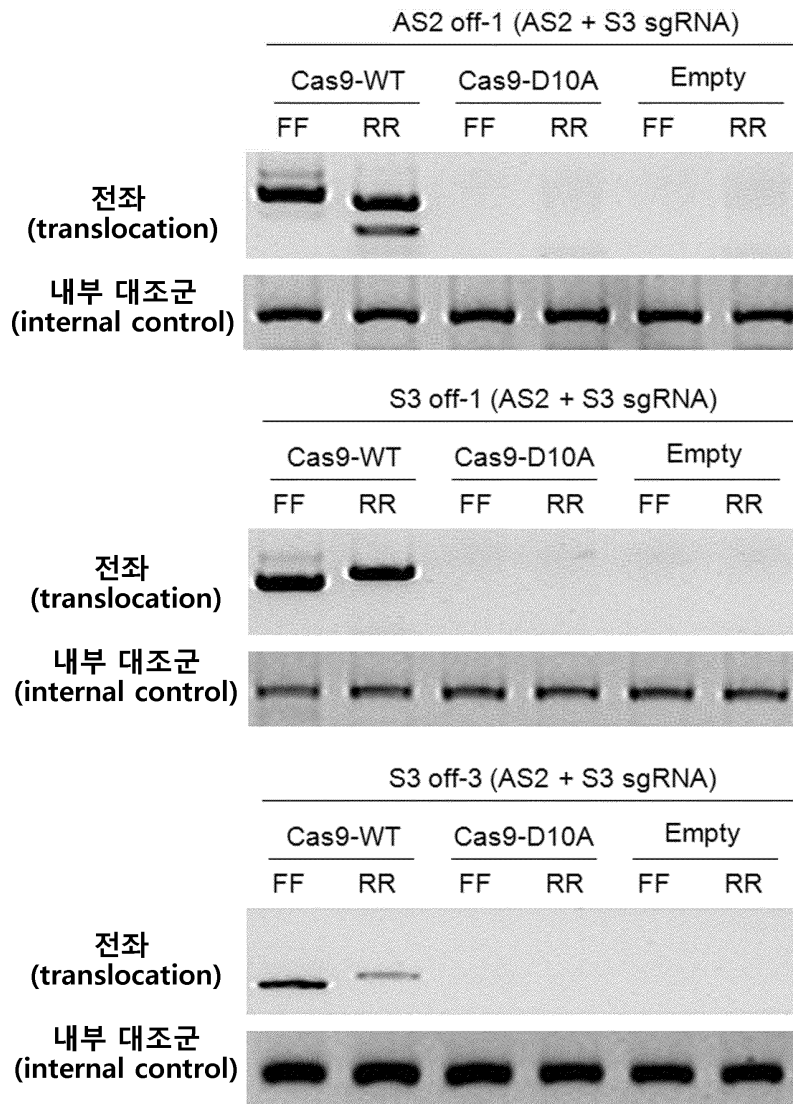
도면21a



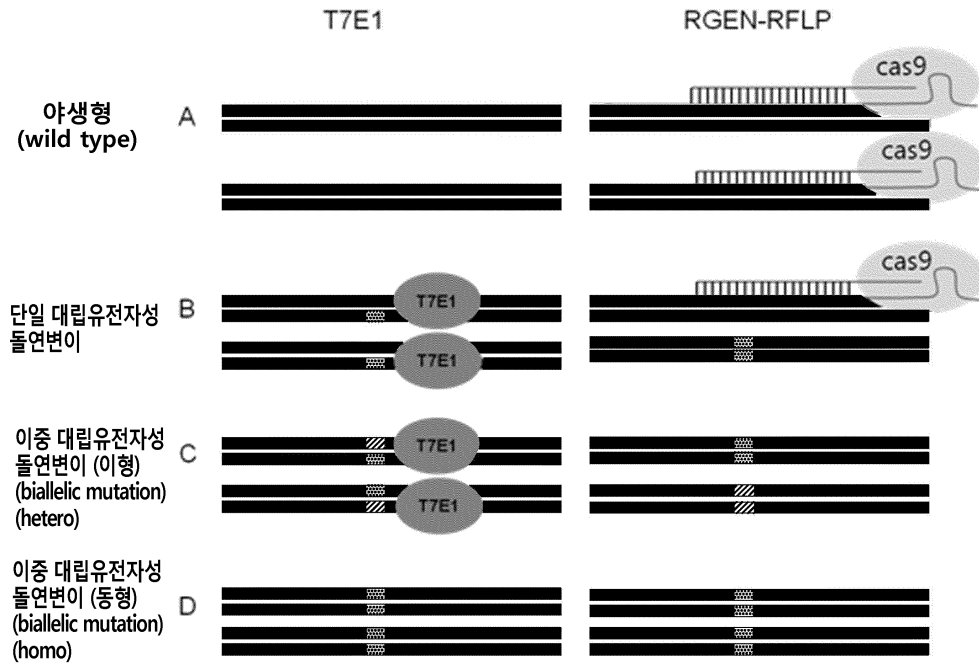
도면21b



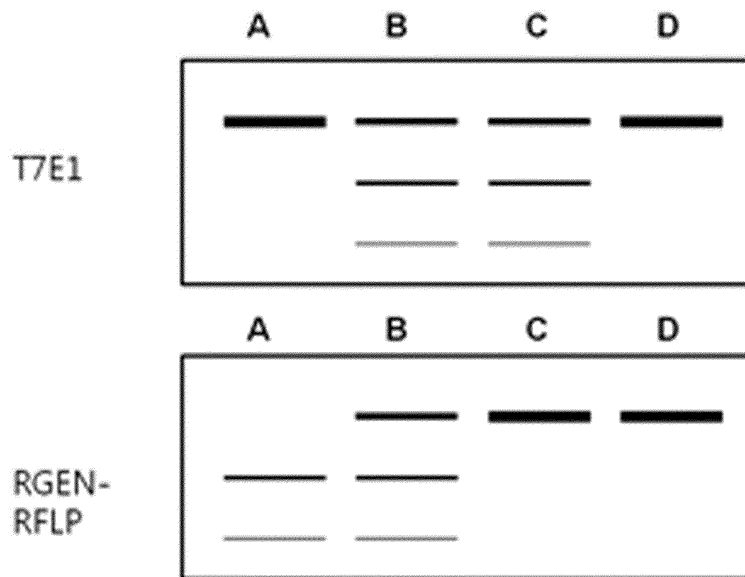
도면21c



도면22a



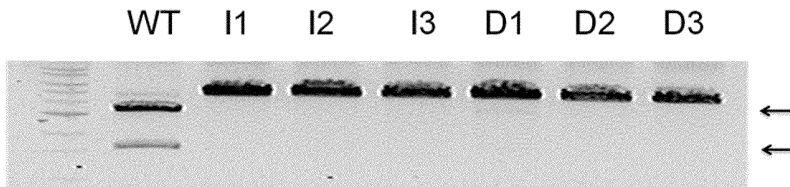
도면22b



도면23

플라스미드 표적 서열

AATGACCACTACATCCT---CAAGGG WT
 AATGACCACTACATCCTT- CAAGGG I1
 AATGACCACTACATCCTTT- CAAGGG I2
 AATGACCACTACATCCTTTT CAAGGG I3
 AATGACCACTACATCCT----AAGGG D1
 AATGACCACTACATCCT-----AGGG D2
 AATGACCACTACATCCT-----GGG D3



도면24a

#1 (+/-)
 TATGTGC AATGACCACTACATCCT---CAAGGGCAGCAATCGGAG WT
 TATGTGC AATGACCACTACATCCTCCTCAAGGGCAGCAATCGGAG +3

#2 (+/-)
 TATGTGC AATGACCACTACATCCTCAAGGGCAGCAATCGGAG WT
 TATGTGC AATGACCACTACATC-----AATCGGAG -12

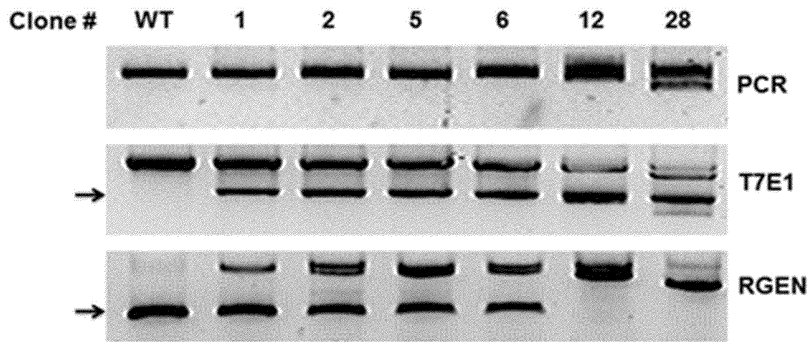
#5 (+/-)
 TATGTGC AATGACCACTACATCCTCAAGGGCAGCAATCGGAG WT
 TATGTGC AATGACCACTACAT-----CAGCAATCGGAG -9

#6 (+/-)
 TATGTGC AATGACCACTACATCCTCAAGGGCAGCAATCGGAG WT
 TATGTGC AATGACCACTACATCC-----AGCAATCGGAG -8

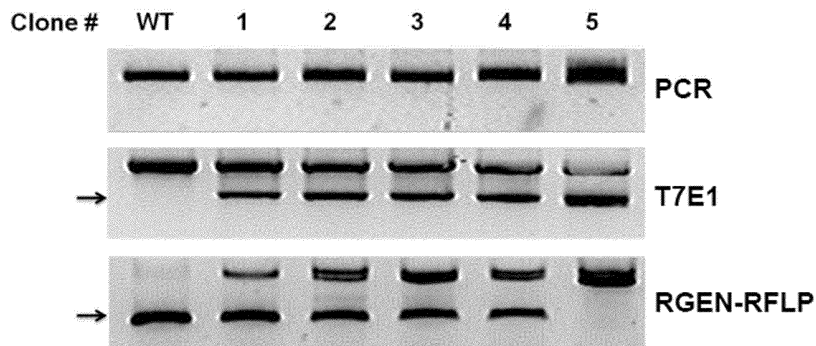
#12 (-/-)
 -----CAGCAATCGG -36
 TATGTGC AATGACCACTACATCCT-----CAAGGGCAGCAATCGG +1
 TATGTGC AATGACCACTACATCCT-----CAAGGGCAGCAATCGG +1
 TATGTGC AATGACCACTACATCCT/67bp/CAAGGGCAGCAATCGG +67

#28 (-/-)
 TATGTGC AATGACCACTACATCCTTCAAGGGCAGCAATCGG +1
 TATGTGC AATGACCACTACAT--T-----GGCAGCAATCGG -7,+1
 TATGTGC AATGACCACTACAT----- -94

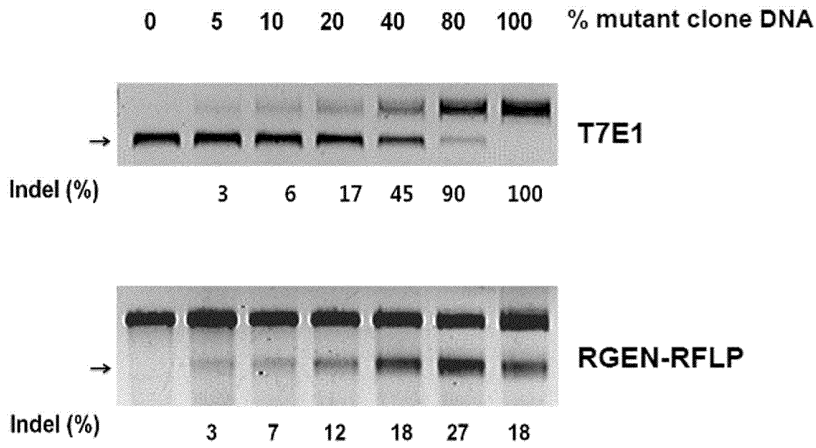
도면24b



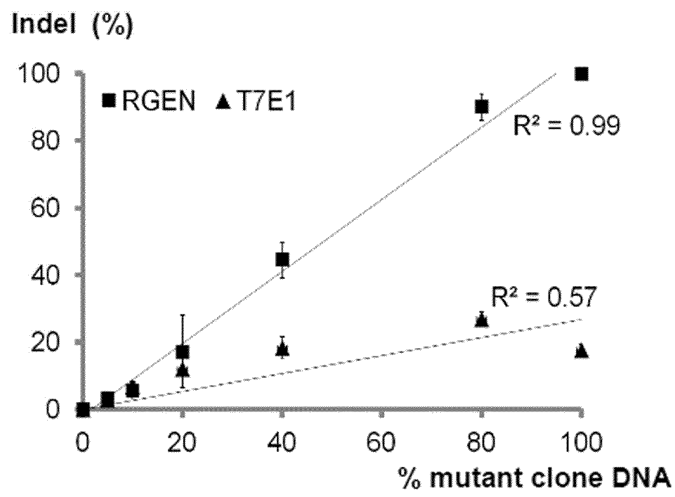
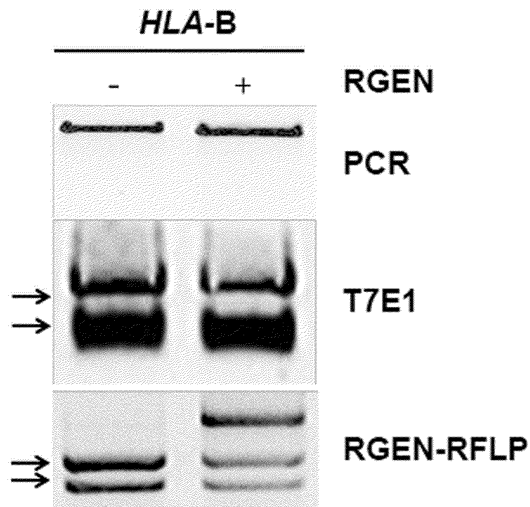
도면25a



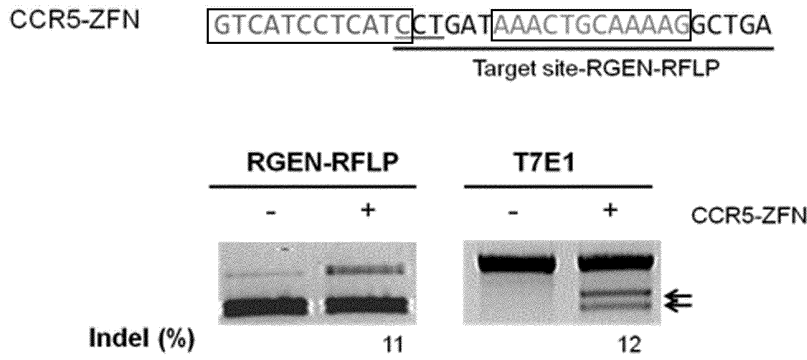
도면25b



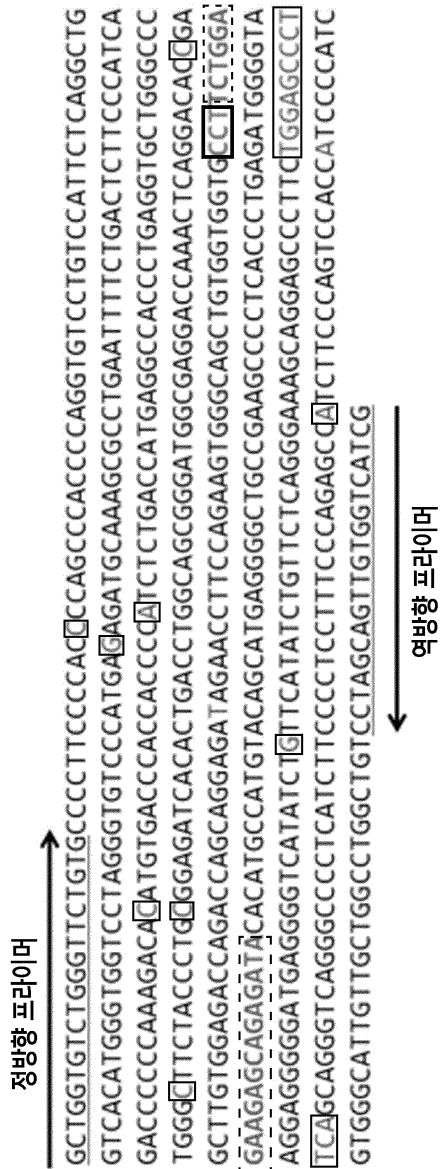
도면25c



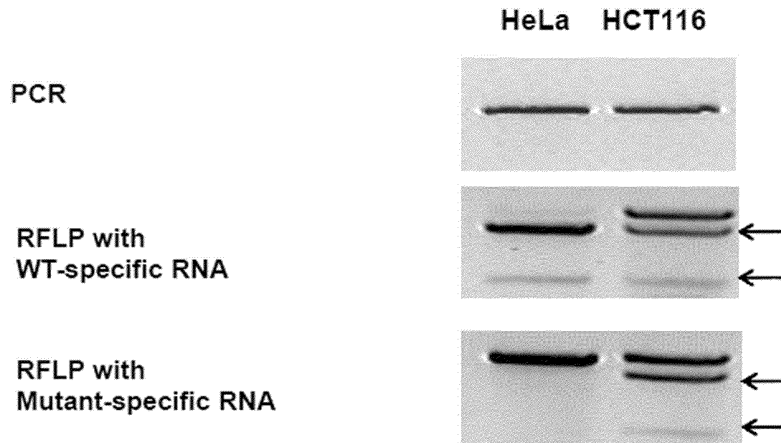
도면27



도면28



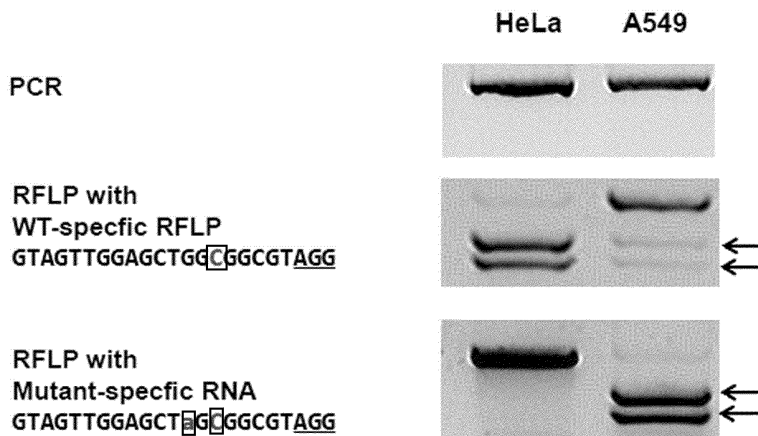
도면29a



HeLa
ACTACCACAGCTCCTTCTCTGAGTGG wild-type

HCT116
ACTACCACAGCTCCTTCTCTGAGTGG wild-type
ACTACCACAGCTCCT---CTGAGTGG c.133-135 del TCT

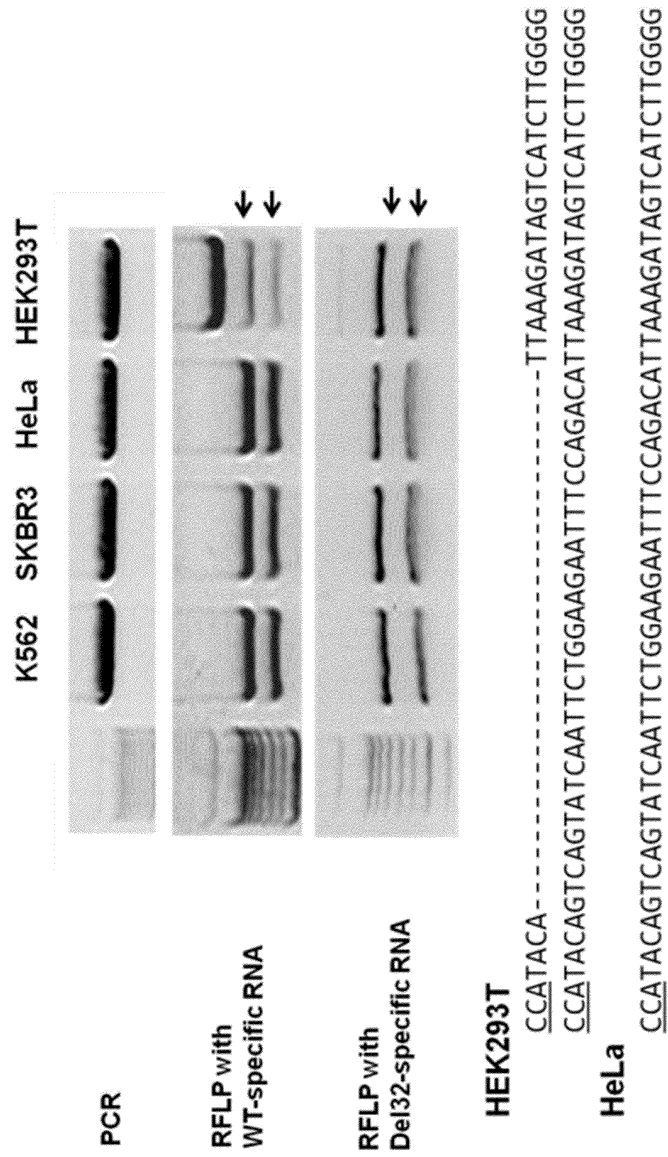
도면29b



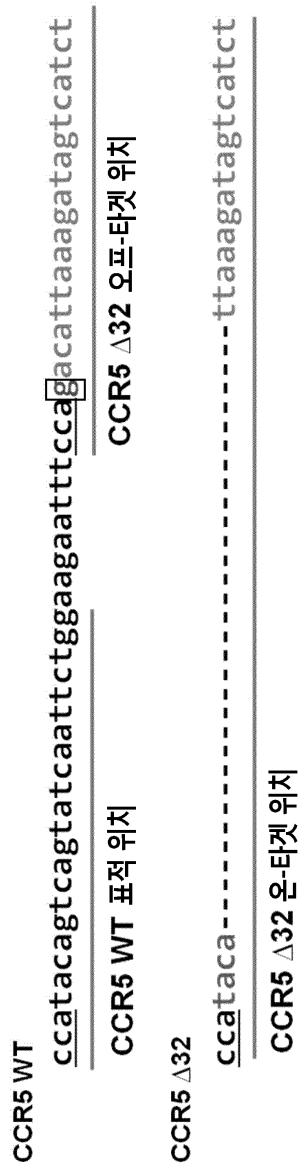
HeLa
GTAGTTGGAGCTGGTGGCGTAGG wild-type

A549
GTAGTTGGAGCTGGTGGCGTAGG c.34G>A

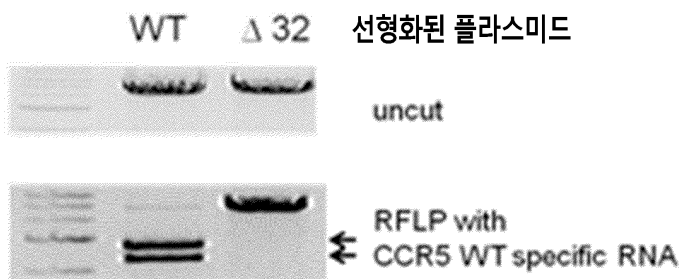
도면30a



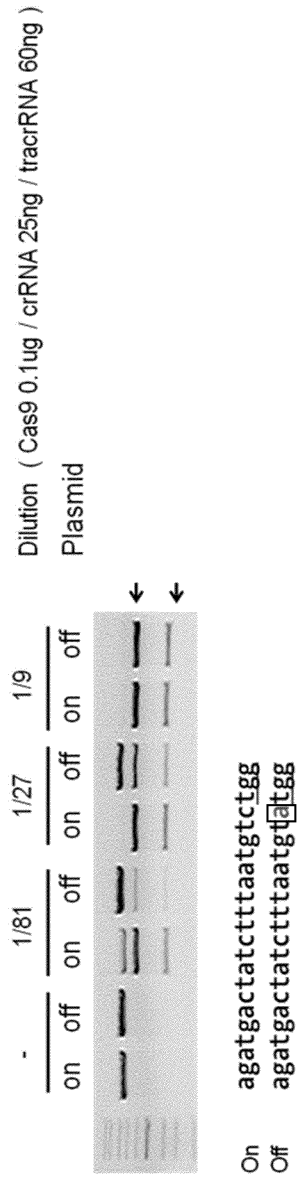
도면30b



도면30c



도면30d



도면31a

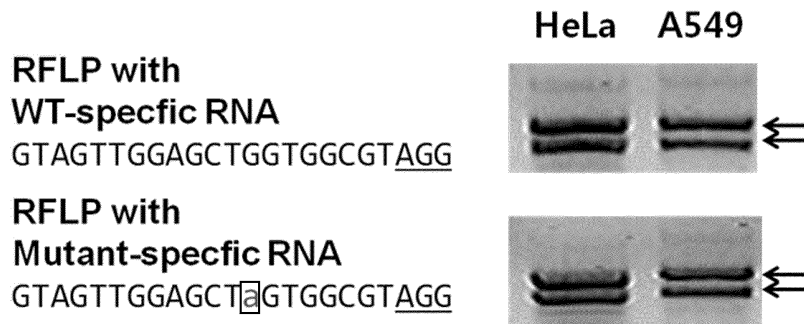
KRAS

HeLa

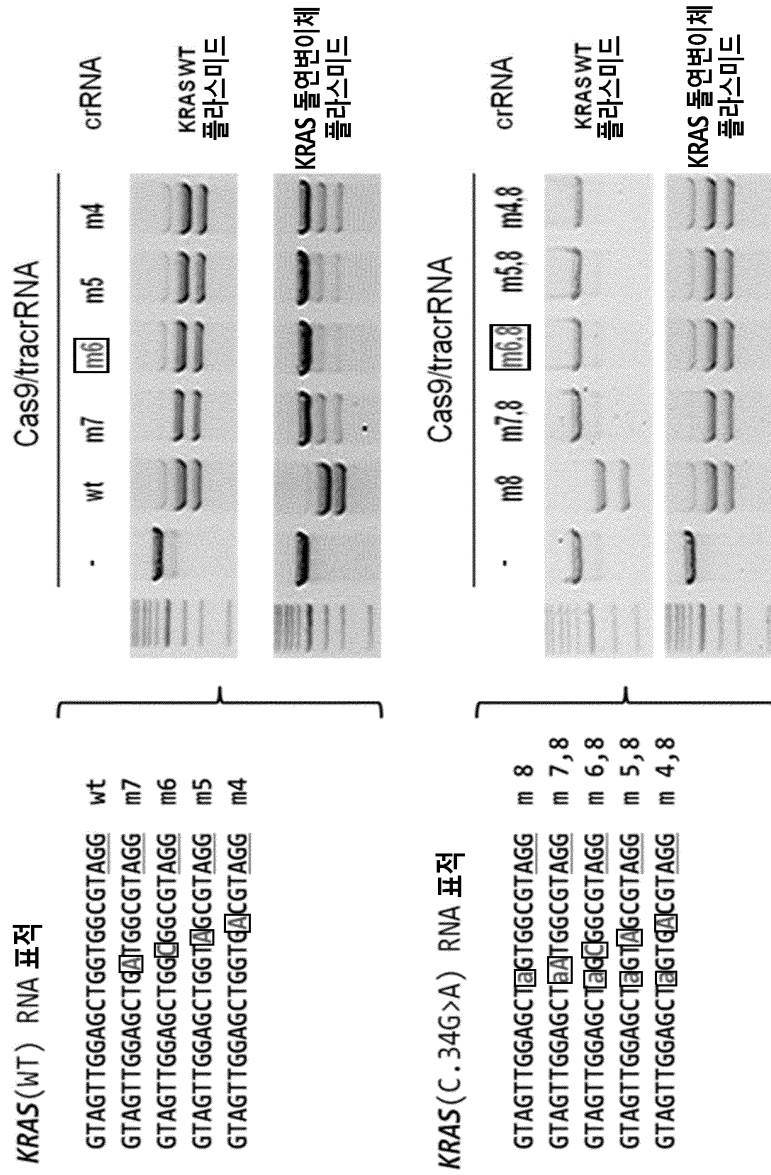
GTAGTTGGAGCTGGTGGCGTAGG 야생형

A549

GTAGTTGGAGCT[a]GTGGCGTAGG c.34G>A



도면31b



도면32a

PIK3CA

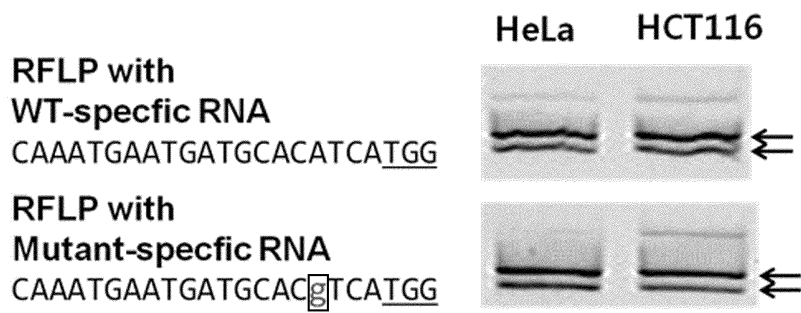
HeLa

CAAATGAATGATGCACATCATGG 아생형

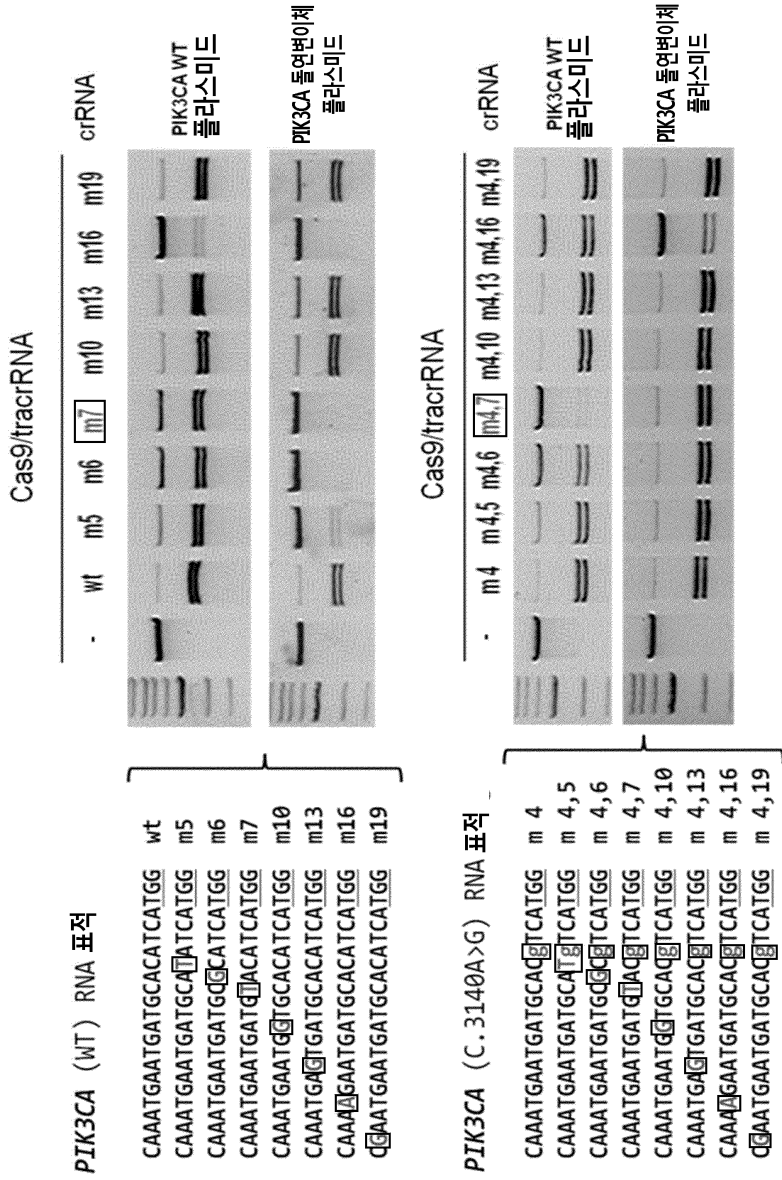
HCT116

CAAATGAATGATGCACATCATGG Wild-type

CAAATGAATGATGCAC[g]TCATGG C.3140A>G



도면32b



도면33a

IDH1

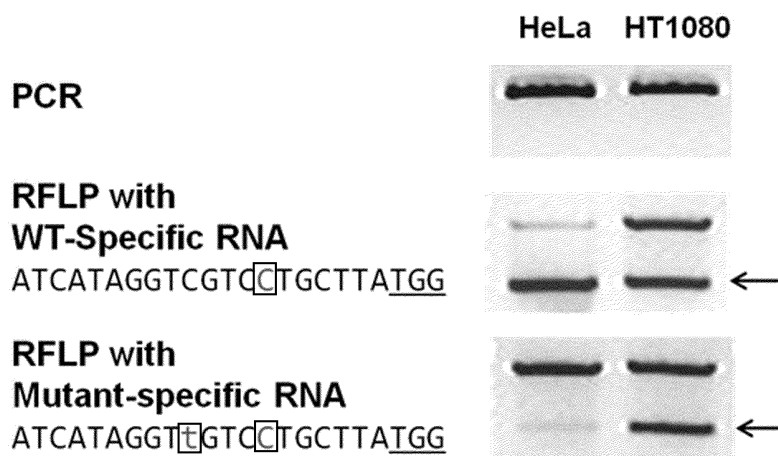
HeLa

ATCATAGGTCGTCATGCTTATGG 야생형

HT1080

ATCATAGGTCGTCATGCTTATGG 야생형

ATCATAGGTCTGTCATGCTTATGG c.394C>T



도면33b

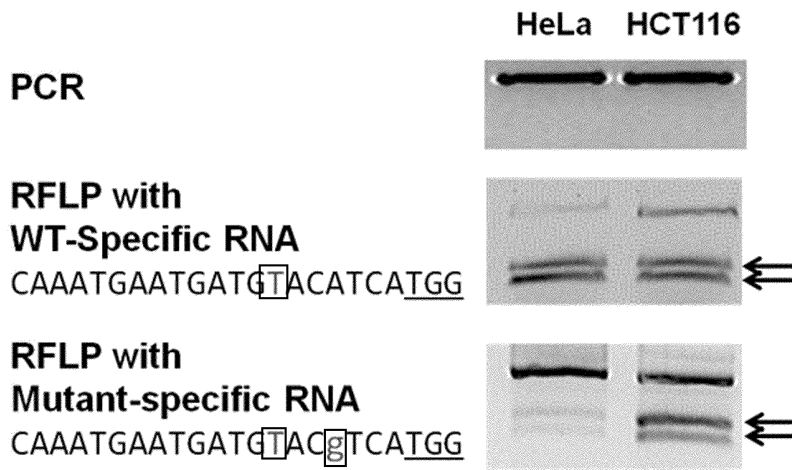
PIK3CA

HeLa

CAAATGAATGATGCACATCATGG 야생형

HCT116

CAAATGAATGATGCACATCATGG 야생형
CAAATGAATGATGCACgTCATGG C.3140A>G



도면33c

NRAS

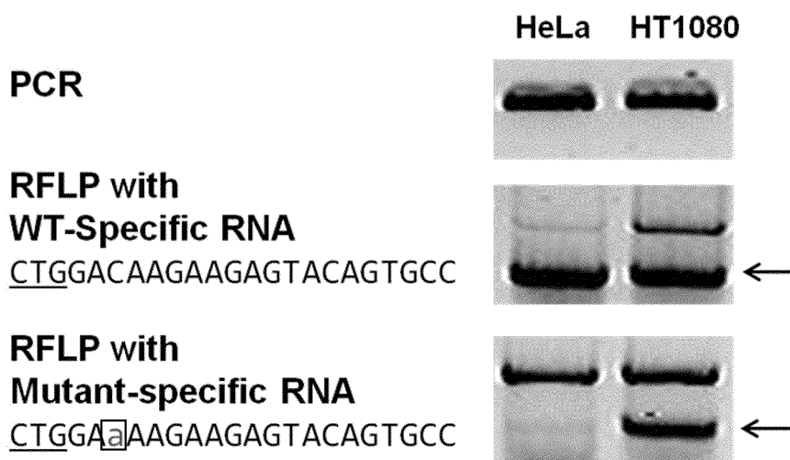
HeLa

CTGGACAAGAAGAGTACAGTGCC 야생형

HT1080

CTGGACAAGAAGAGTACAGTGCC 야생형

CTGGAAAAGAAGAGTACAGTGCC c.181C>A



도면33d

BRAF

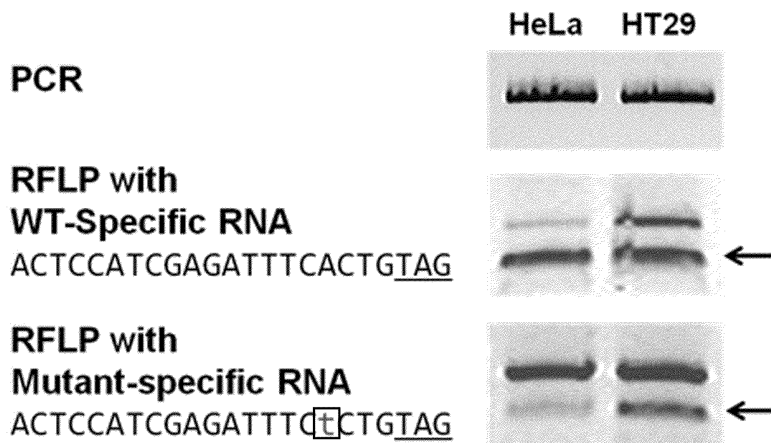
HeLa

ACTCCATCGAGATTTCACTGTAG 야생형

HT29

ACTCCATCGAGATTTCACTGTAG 야생형

ACTCCATCGAGATTTCTTCTGTAG (c.1799T>A)



서열목록

- <110> TOOLGEN INCORPORATED
- <120> Composition for cleaving a target DNA comprising a guide RNA specific for the target DNA and Cas protein-encoding nucleic acid or Cas protein, and use thereof
- <130> IKPA141187KR
- <150> US 61/717,324
- <151> 2012-10-23
- <150> US 61/803,599
- <151> 2013-03-20
- <150> US 61/837,481
- <151> 2013-06-20
- <160> 111
- <170> Kopatent In 2.0
- <210> 1
- <211> 4107

<212> DNA

<213> Artificial Sequence

<220><223

> Cas9-coding sequence

<400> 1

```

atggacaaga agtacagcat cggcctggac atcggtagca acagcgtggg ctgggccgtg      60
atcaccgacg agtacaaggt gcccagcaag aagtcaagg tgctgggcaa caccgaccgc      120
cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgccgag      180
gccacccgcc tgaagcgcac cgcccgcgcg cgctacacc gcccaagaa ccgcatctgc      240
tacctgcagg agatcttcag caacgagatg gccaaggtgg acgacagctt cttccaccgc      300
ctggaggaga gcttcctggt ggaggaggac aagaagcacg agcgccacc catcttcggc      360

aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgcgcaag      420
aagctggtgg acagcaccga caaggccgac ctgcgctga tctacctggc cctggcccac      480
atgatcaagt tccgcggeca cttcctgacg gagggcgacc tgaacccga caacagcgac      540
gtggacaagc tgttcattca gctggtgcag acctacaacc agctgttcga ggagaacccc      600
atcaacgcca gcgcgctgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc      660
cgcctggaga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggcaac      720
ctgatcgcgc tgagcctggg cctgaccccc aactcaaga gcaacttca cctggccgag      780

gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc      840
cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc      900
ctgctgagcg acatcctcgc cgtgaacacc gagatcacca aggcccccct gagcgccagc      960
atgatcaagc gctacgacga gcaccaccag gacctgacce tgctgaagge cctggtgcgc      1020
cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc      1080
ggctacatcg acggcggcgc cagccaggag gaggttctaca agttcatcaa gcccatcctg      1140
gagaagatgg acggcaccga ggagctgctg gtgaagctga accgcgagga cctgctgcgc      1200

aagcagcgca ccttcgacaa cggcagcadc ccccaccaga tccacctggg cgagctgcac      1260
gccatcctgc gccccagga ggaacttctac cccttcctga aggacaaccg cgagaagatc      1320
gagaagatcc tgaccttccg catcccctac tacgtgggcc ccctggcccg cggcaacagc      1380
cgcttcgctt ggatgaccgg caagagcgag gagaccatca cccctggaa cttcgaggag      1440
gtggtggaca agggcgccag cgcccagagc ttcacgagc gcatgaccaa cttcgacaag      1500
aacctgcccc acgagaaggt gctgccccaa cacagcctgc tgtacgagta cttcaccgtg      1560

```

tacaacgagc tgaccaaggt gaagtacgtg accgagggca tgcgcaagcc cgccttcctg 1620

agcggcgagc agaagaaggc catcgtggac ctgctgttca agaccaaccg caagtgacc 1680

gtgaagcagc tgaaggagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc 1740

agcggcgtgg aggaccgctt caacgccagc ctgggcacct accacgacct gctgaagatc 1800

atcaaggaca aggacttctt ggacaacgag gagaacgagg acatcctgga ggacatcgtg 1860

ctgacctga ccttgttcga ggaccgcgag atgatcgagg agcgcctgaa gacctacgcc 1920

cacctgttcg acgacaaggt gatgaagcag ctgaagcgc gccgctacac cggctggggc 1980

cgctgagcc gcaagcttat caacggcatc cgcgacaagc agagcggcaa gaccatcctg 2040

gacttctga agagcgacgg cttcgccaac cgcaacttca tgcagctgat ccacgacgac 2100

agcctgacct tcaaggagga catccagaag gcccaggtga gcggccaggg cgacagcctg 2160

cacgagcaca tcgccaacct ggccggcagc cccgccatca agaagggcat cctgcagacc 2220

gtgaaggtgg tggacgagct ggtgaaggtg atgggccgcc acaagcccga gaacatcgtg 2280

atcgagatgg cccgcgagaa ccagaccacc cagaagggcc agaagaacag ccgcgagcgc 2340

atgaagcga tcgaggaggg catcaaggag ctgggcagcc agatcctgaa ggagcacccc 2400

gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccgc 2460

gacatgtacg tggaccagga gctggacatc aaccgcctga gcgactacga cgtggaccac 2520

atcgtgcccc agagcttctt gaaggacgac agcatcgaca acaaggtgct gacccgcagc 2580

gacaagaacc gcggcaagag cgacaacgtg cccagcgagg aggtggtgaa gaagatgaag 2640

aactactggc gccagctgct gaacccaag ctgatcacc agcgaagt t cgacaacctg 2700

accaaggccg agcggcgagg cctgagcgag ctggacaagg cggcttcat caagcggcag 2760

ctggtggaga cccgccagat caccaagcac gtggcccaga tcctggacag ccgcatgaac 2820

accaagtacg acgagaacga caagctgatc cgcgaggtga aggtgatcac cctgaagagc 2880

aagctggtga gcgacttccg caaggacttc cagttctaca aggtgcgca gatcaacaac 2940

taccaccagc cccacgacgc ctacctgaac gccgtggtgg gcaccgcctt gatcaagaag 3000

taccccaagc tggagagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcgcaag 3060

atgatgccca agagcgagca ggagatcggc aaggccaccg ccaagtactt cttctacagc 3120

aacatcatga acttcttcaa gaccgagatc acctggcca acggcgagat ccgcaagcgc 3180

cccctgatcg agaccaacgg cgagaccggc gagatcgtgt gggacaaggg ccgcgacttc 3240

gccaccgtgc gcaaggtgct gagcatgccc caggtgaaca tcgtgaagaa gaccgaggtg 3300

cagaccggcg gcttcagcaa ggagagcatc ctgcccaagc gcaacagcga caagctgatc 3360
 gcccgaaga aggactggga cccaagaag tacggcggct tcgacagccc cacctggcc 3420
 tacagcgtgc tgggtggggc caaggtggag aaggcaaga gcaagaagct gaagagcgtg 3480
 aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac 3540
 ttctggagg ccaagggcta caaggagtg aagaaggacc tgatcatcaa gctgcccaag 3600
 tacagcctgt tcgagctgga gaacggccgc aagcgcagc tggccagcgc cggcagctg 3660
 cagaagggca acgagctggc cctgccccagc aagtacgtga acttctgta cctggccagc 3720

cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag 3780
 cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg 3840
 atcctggcgg acgccaacct ggacaaggtg ctgagcgcct acaacaagca ccgcgacaag 3900
 cccatccgcg agcaggccga gaacatcatc cacctgttca ccctgaccaa cctgggcgcc 3960
 cccgccgct tcaagtactt cgacaccacc atcgaccgca agcgtacac cagcaccaag 4020
 gagggtgctgg acgccacct gatccaccag agcatcaccg gtctgtacga gaccgcagc 4080
 gacctgagcc agctggggcg cgactaa 4107

<210> 2

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> peptide tag

<400> 2

Gly Gly Ser Gly Pro Pro Lys Lys Lys Arg Lys Val Tyr Pro Tyr Asp

1 5 10 15

Val Pro Asp Tyr Ala

20

<210> 3

<211> 34

<212> DNA

<213> Artificial Sequence

<220><223> F primer for CCR5

<400> 3

aattcatgac atcaattatt atacatcgga ggag 34

<210> 4

<211> 34

<

212> DNA

<213> Artificial Sequence

<220><223> R primer for CCR5

<400> 4

gatcctcctc cgatgtataa taattgatgt catg 34

<210> 5

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> F1 primer for CCR5

<400> 5

ctccatgggtg ctatagagca 20

<210> 6

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> F2 primer for CCR5

<400> 6

gagccaagct ctccatctag t 21

<210> 7

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> R primer for CCR5

<400> 7

gccctgtcaa gattgacac 20

<210> 8

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> F1 primer for C4BPB

<400> 8

tatttggtg gttgaaaggg 20

<210> 9
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> R1 primer for C4BPB
 <400>
 > 9
 aaagtcatga aataaacaca ccca 24
 <210> 10
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> F2 primer for C4BPB
 <400> 10
 ctgcattgat atggtagtac catg 24
 <210> 11
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> R2 primer for C4BPB
 <400> 11
 gctgttcatt gcaatggaat g 21
 <210> 12
 <211>
 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> F1 primer for ADCY5
 <400> 12
 gctccacct tagtgctctg 20
 <210> 13
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> R1 primer for ADCY5

<400> 13
 ggtggcagga acctgtatgt 20
 <210> 14
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> F2 primer for ADCY5
 <400> 14
 gtcattggcc agagatgtgg a 21

<210> 15
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> R2 primer for ADCY5
 <400> 15
 gtcccatgac aggcgtgtat 20

<210> 16
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for KCNJ6
 <400> 16
 gcctggccaa gtttcagtta 20

<210> 17
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> R1 primer for KCNJ6

<400> 17
 tggagccatt ggtttgcatc 20
 <210> 18
 <211> 22
 <212> DNA

<213> Artificial Sequence

<220><223> R2 primer for KCNJ6

<400> 18

ccagaactaa gccgtttctg ac 22

<210> 19

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> F1 primer for CNTNAP2

<400> 19

atcaccgaca accagtttcc 20

<210> 20

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> F2 primer for CNTNAP2

<400> 20

tgcagtgcag actctttcca 20

<210> 21

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> R primer for CNTNAP2

<400> 21

aaggacacag ggcaactgaa 20

<210> 22

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> F1 primer for N/A Chr. 5

<400>

> 22

tgtggaacga gtggtgacag 20

<210> 23

<211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> R1 primer for N/A Chr. 5
 <400> 23
 gctggattag gaggcaggat tc 22
 <210> 24
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> F2 primer for N/A Chr. 5
 <400> 24
 gtgctgagaa cgcttcatag ag 22
 <210> 25

 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> R2 primer for N/A Chr. 5
 <400> 25
 ggaccaaacc acattcttct cac 23
 <210> 26
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for deletion
 <400> 26
 ccacatctcg ttctcggttt 20
 <210> 27
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for deletion
 <400>
 > 27

tcacaagccc acagatattt	20
<210> 28	
<211> 105	
<212> RNA	
<213> Artificial Sequence	
<220><223> sgRNA for CCR5	
<400> 28	
ggugacauca auuauuuac auguuuuaga gcuagaaaua gcaaguuaaa auaaggcuag	60
uccguuauca acuugaaaaa guggcaccga gucggugcuu uuuuu	105
<210> 29	
<211> 44	
<212> RNA	
<213> Artificial Sequence	
<220><223> crRNA for CCR5	
<400> 29	
ggugacauca auuauuuac auguuuuaga gcuavgcugu uuug	44
<210> 30	
<211> 86	
<212> RNA	
<213> Artificial Sequence	
<220><223> tracrRNA for CCR5	
<400> 30	
ggaaccauuc aaaacagcau agcaaguuaa auaaggcua guccguuauc aacuugaaaa	60
aguggcaccg agucggugcu uuuuuu	86
<210> 31	
<211> 86	
<212> DNA	
<213> Artificial Sequence	
<220><223> Foxn1 #1 sgRNA	
<400> 31	
gaaattaata cgactcacta taggcagtct gacgtcacac ttccgtttta gagctagaaa	60
tagcaagtta aaataaggct agtccg	86
<210> 32	

<211> 86
 <212> DNA
 <213> Artificial Sequence
 <220><223> Foxn1 #2 sgRNA
 <400> 32
 gaaattaata cgactcacta taggacttcc aggctccacc cgacgtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86
 <210> 33
 <211> 86
 <212> DNA
 <213> Artificial Sequence
 <220><223> Foxn1 #3 sgRNA
 <400> 33
 gaaattaata cgactcacta taggccaggc tccacccgac tggagtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86
 <210> 34
 <211> 86
 <212> DNA
 <213> Artificial Sequence
 <220><223> Foxn1 #4 sgRNA
 <400> 34
 gaaattaata cgactcacta taggactgga gggcgaacc caaggtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86
 <210> 35
 <211> 86
 <212> DNA
 <213> Artificial Sequence
 <220><223> Foxn1 #5 sgRNA
 <400> 35
 gaaattaata cgactcacta taggacccca aggggacctc atgcgtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86
 <210> 36
 <211> 86

<212> DNA
 <213> Artificial Sequence
 <220><223> Prkdc #1 sgRNA
 <400> 36
 gaaattaata cgactcacta taggttagtt tttccagag acttgtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86
 <210> 37
 <211> 86
 <212> DNA
 <213> Artificial Sequence
 <220><223> Prkdc #2 sgRNA
 <400> 37
 gaaattaata cgactcacta taggttggtt tgcttggtt tatcgtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86

 <210> 38
 <211> 86
 <212> DNA
 <213> Artificial Sequence
 <220><223> Prkdc #3 sgRNA
 <400> 38
 gaaattaata cgactcacta taggcacaag caaaccaag tctcgtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86
 <210> 39
 <211> 86
 <212> DNA
 <213> Artificial Sequence
 <220><223> Prkdc #4 sgRNA
 <400> 39
 gaaattaata cgactcacta taggcctcaa tgctaagcga cttcgtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86

 <210> 40
 <211> 29
 <212> DNA

<213> Artificial Sequence
 <220><223> F1 primer for Foxn1
 <400> 40
 gtctgtctat catctcttcc cttctctcc 29
 <210> 41
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> F2 primer for Foxn1
 <400> 41
 tccctaattcc gatggctagc tccag 25
 <210> 42
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> R1 primer for Foxn1
 <400> 42
 acgagcagct gaagtttagca tgc 23
 <210> 43
 <211> 32
 <212> DNA
 <213> Artificial Sequence
 <220><223> R2 primer for Foxn1
 <400> 43
 ctactcaatg ctcttagagc taccaggett gc 32
 <210> 44
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Prkdc
 <400> 44
 gactgttggtg gggaggccg 20
 <210> 45
 <

211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> F2 primer for Prkdc
 <400> 45
 gggagggccg aaagtcttat ttg 24
 <210> 46
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> R1 primer for Prkdc
 <400> 46
 cctgaagact gaagttggca gaagtgag 28
 <210> 47
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220><223> R2 primer for Prkdc
 <400> 47

 ctttagggct tcttctctac aatcacg 27
 <210> 48
 <211> 38
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Foxn1
 <400> 48
 ctcggtgtgt agccctgacc tcggtgtgta gccctgac 38
 <210> 49
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for Foxn1
 <400> 49
 agactggcct ggaactcaca g 21

<210> 50
 <211> 23
 <212
 > DNA
 <213> Artificial Sequence
 <220><223> F primer for Foxn1
 <400> 50
 cactaaagcc tgtcaggaag ccg 23
 <210> 51
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for Foxn1
 <400> 51
 ctgtggagag cacacagcag c 21
 <210> 52
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Foxn1
 <400> 52
 gctgcgacct gagaccatg 19

 <210> 53
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for Foxn1
 <400> 53
 cttcaatggc ttctgctta ggctac 26
 <210> 54
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Foxn1

<400> 54
 ggttcagatg aggccatcct ttc 23
 <210> 55
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for Foxn1

<400> 55
 cctgatctgc aggettaacc cttg 24
 <210> 56
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Prkdc

<400> 56
 ctcacctgca catcacatgt gg 22
 <210> 57
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for Prkdc

<400> 57
 ggcatccacc ctatggggtc 20
 <210> 58
 <211
 > 25
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Prkdc

<400> 58
 gccttgacct agagcttaaa gagcc 25
 <210> 59
 <211> 25
 <212> DNA

<213> Artificial Sequence
 <220><223> R primer for Prkdc
 <400> 59
 ggtcttgta gcaggaagga cactg 25
 <210> 60
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Prkdc
 <400> 60
 aaaactctgc ttgatgggat atgtggg 27

 <210> 61
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for Prkdc
 <400> 61
 ctctcactgg ttatctgtgc tccttc 26
 <210> 62
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Prkdc
 <400> 62
 ggatcaatag gtggtggggg atg 23
 <210> 63
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for Prkdc

 <400> 63
 gtgaatgaca caatgtgaca gcttcag 27
 <210> 64

<211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> F primer for Prkdc
 <400> 64
 cacaagacag acctctcaac attcagtc 28
 <210> 65
 <211> 32
 <212> DNA
 <213> Artificial Sequence
 <220><223> R primer for Prkdc
 <400> 65
 gtgcatgcat ataatccatt ctgattgctc tc 32
 <210> 66
 <211>
 > 17
 <212> DNA
 <213> Artificial Sequence
 <220><223> F1 primer for Prkdc
 <400> 66
 gggaggcaga ggcaggt 17
 <210> 67
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> F2 primer for Prkdc
 <400> 67
 ggatctctgt gattttgagg cca 23
 <210> 68
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> R1 primer for Prkdc
 <400> 68

<p>gctccagaac tcactcttag gctc</p> <p><210> 69</p> <p><211> 20</p> <p><212> DNA</p> <p><213> Artificial Sequence</p> <p><220><223> Primer for Foxn1</p> <p><400> 69</p> <p>ctactccctc cgcagtctga</p> <p><210> 70</p> <p><211> 20</p> <p><212> DNA</p> <p><213> Artificial Sequence</p> <p><220><223> Primer for Foxn1</p> <p><400> 70</p> <p>ccaggcctag gttccaggta</p> <p><210> 71</p> <p><211> 20</p> <p><212> DNA</p> <p><213> Artificial Sequence</p> <p><220><223> Primer for Prkdc</p> <p><400> 71</p> <p>ccccagcatt gcagatttcc</p> <p><210> 72</p> <p><211> 23</p> <p><212> DNA</p> <p><213> Artificial Sequence</p> <p><220><223> Primer for Prkdc</p> <p><400> 72</p> <p>agggttctt ctctacaatc acg</p> <p><210> 73</p> <p><211> 86</p> <p><212> DNA</p> <p><213> Artificial Sequence</p>	<p>24</p> <p>20</p> <p>20</p> <p>20</p> <p>20</p> <p>23</p>
------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------

<220><223> BRI1 target 1
 <400> 73
 gaaattaata cgactcacta taggtttgaa agatggaagc gcggtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86

<210> 74
 <211> 86
 <212> DNA
 <213> Artificial Sequence

<220><223> BRI1 target 2
 <400> 74
 gaaattaata cgactcacta taggtgaaac taaactggtc cacagtttta gagctagaaa 60
 tagcaagtta aaataaggct agtccg 86

<210> 75
 <211> 64
 <212> DNA
 <213> Artificial Sequence

<220><223> Universal
 <400> 75
 aaaaaagcac cgactcgtg ccaacttttc aagttgataa cggactagcc ttattttaac 60
 ttgc 64

<210> 76
 <211> 65
 <212> DNA
 <213> Artificial Sequence

<220><223> Templates for crRNA
 <400> 76
 gaaattaata cgactcacta taggnnnnnn nnnnnnnnnn nnnngtttta gagctatgct 60
 gtttt 65

<210> 77
 <211> 67
 <212> DNA
 <213> Artificial Sequence

<220><223> tracrRNA

<400> 77
gaaattaata cgactcacta taggaacat tcaaacacgc atagcaagtt aaaataaggc 60
tagtccg 67

<210> 78
<211> 69
<212> DNA
<213> Artificial Sequence
<220><223> tracrRNA
<400> 78
aaaaaaagca ccgactcggg gccacttttt caagttgata acggactagc cttattttaa 60
cttgctatg 69

<210> 79
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Primer
<400> 79
ctccatggtg ctatagagca 20

<210> 80
<211> 21
<212> DNA
<213> Artificial Sequence
<220><223> Primer
<400> 80
gagccaagct ctccatctag t 21

<210> 81
<211> 20
<212> DNA
<213> Artificial Sequence
<220><223> Primer
<400> 81
gcctgtcaa gaggtagacac 20

<210> 82

<211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 82
 gcacaggtg gaacaagatg ga 22
 <210> 83
 <211
 > 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 83
 gccaggtacc tatcgattgt cagg 24
 <210> 84
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 84
 gagccaagct ctccatctag t 21
 <210> 85
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 85
 actctgactg ggtcaccagg 20

 <210> 86
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 86

tatttgctg gttgaaaggg	20
<210> 87	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	
<400> 87	
aaagtcatga aataaacaca ccca	24
<210> 88	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	
<400> 88	
ctgcattgat atggtagtac catg	24
<210> 89	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	
<400> 89	
gctgttcatt gcaatggaat g	21
<210> 90	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	
<400> 90	
atggagttgg acatggccat gg	22
<210> 91	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Primer	

<400>	91	
	actcactatc cacagttcag catttacc	28
<210>	92	
<211>	23	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Primer	
<400>	92	
	tggagatagc tgtcagcaac ttt	23
<210>	93	
<211>	29	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Primer	
<400>	93	
	caacaaagca aaggtaaagt tggtaatag	29
<210>	94	
<211>	25	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Primer	
<400>	94	
	ggtttcagga gatgtgttac aaggc	25
<210>	95	
<211>	27	
<212>	DNA	
<213>	Artificial Sequence	
<220><223>	Primer	
<400>	95	
	gattgtgcaa ttcctatgca atcggtc	27
<210>	96	
<211>	25	
<212>	DNA	

<213> Artificial Sequence
 <220><223> Primer
 <400> 96
 cactgggtac ttaatctgta gcctc 25
 <210> 97
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 97
 ggtccaagt cattcccagt agc 23

 <210> 98
 <211> 30
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 98
 catcactgca gttgtaggtt ataactatcc 30
 <210> 99
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 99
 ttgaaaacca cagatctggt tgaacc 26
 <210> 100
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 100
 ggagtgcaa gagaatatct gg 22

 <210> 101

<211> 32
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 101
 ctgaaactgg tttcaaaata ttcgttttaa gg 32

<210> 102
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 102
 gctctgtatg ccctgtagta gg 22

<210> 103
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Primer
 <400> 103
 tttgcatctg accttacctt tg 22

<210> 104
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Target sequence of RGEN
 <400> 104
 aatgaccact acatcctcaa ggg 23

<210> 105
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220><223> Target sequence of RGEN
 <400> 105
 agatgatgtc tcatcatcag agg 23

<210> 106
 <211> 4170
 <212> DNA
 <213> Artificial Sequence
 <220><223> Cas9-coding sequence in p3s-Cas9HC (humanized, C-term tagging,

human cell experiments)

<400> 106

atggacaaga agtacagcat cggcctggac atcggtagca acagcgtggg ctgggccgtg	60
atcaccgacg agtacaaggt gcccagcaag aagttcaagg tgctgggcaa caccgaccgc	120
cacagcatca agaagaacct gatcggcgcc ctgctgttcg acagcggcga gaccgcccag	180
gccacccgcc tgaagcgcac cgcccgcgcg cgctacacc gcccaagaa ccgcatctgc	240
tacctgcagg agatcttcag caacgagatg gccaaggtgg acgacagctt cttccaccgc	300
ctggaggaga gcttctgtgt ggaggaggac aagaagcacg agcggcacc catcttcggc	360
aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgcgcaag	420
aagctggtgg acagcaccga caaggccgac ctgcgcctga tctacctggc cctggcccac	480
atgatcaagt tcccgggcca cttctgatc gagggcgacc tgaaccccga caacagcgac	540
gtggacaagc tgttcatcca gctggtgcag acctacaacc agctgttcga ggagaacccc	600
atcaacgcca gcgcgctgga cgccaaggcc atcctgagcg cccgcctgag caagagccgc	660
cgcttgagga acctgatcgc ccagctgccc ggcgagaaga agaacggcct gttcggcaac	720
ctgatcgcgc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag	780
gacgccaagc tgcagctgag caaggacacc tacgacgacg acctggacaa cctgctggcc	840
cagatcggcg accagtacgc cgacctgttc ctggccgcca agaacctgag cgacgccatc	900
ctgctgagcg acatcctcgc cgtgaacacc gagatcacca aggccccct gagcgccagc	960
atgatcaagc gctacgacga gcaccaccag gacctgacc tgctgaaggc cctggtgcgc	1020
cagcagctgc ccgagaagta caaggagatc ttcttcgacc agagcaagaa cggctacgcc	1080
ggctacatcg acggcggcgc cagccaggag gaggttctaca agttcatcaa gcccacctg	1140
gagaagatgg acggcaccga ggagctgctg gtgaagctga accgcgagga cctgctgcgc	1200
aagcagcga ccttcgacaa cggcagcadc ccccaccaga tccacctggg cgagctgcac	1260
gcccacctgc gccgccagga ggacttctac cccttcctga aggacaaccg cgagaagatc	1320
gagaagatcc tgaccttcg catcccctac tacgtgggcc ccctggcccg cggcaacagc	1380
cgcttcgctt ggatgaccgc caagagcgag gagaccatca ccccctggaa cttcgaggag	1440

gtggtggaca agggcgccag cgcccagagc ttcacgagc gcatgaccaa cttcgacaag 1500
aacctgcccc acgagaaggt gctgccaag cacagcctgc tgtacgagta cttcacctg 1560
tacaacgagc tgaccaaggt gaagtacgtg accgagggca tgcgcaagcc cgccttctg 1620

agcggcgagc agaagaagc catcgtggac ctgctgttca agaccaaccg caaggtgacc 1680
gtgaagcagc tgaaggagga ctacttcaag aagatcgagt gcttcgacag cgtggagatc 1740
agcggcgtgg aggaccgctt caacgccagc ctgggcacct accacgacct gctgaagatc 1800
atcaaggaca aggacttctt ggacaacgag gagaacgagg acatcctgga ggacatcgtg 1860
ctgacctga cctgttctga ggaccgcgag atgatcgagg agcgcctgaa gacctacgcc 1920
cacctgttgc acgacaaggt gatgaagcag ctgaagcggc gccgctacac cggctggggc 1980
cgcctgagcc gcaagcttat caacggcatc cgcgacaagc agagcggcaa gaccatcctg 2040

gacttctga agagcgacgg cttcgccaac cgcaacttca tgcagctgat ccacgacgac 2100
agcctgacct tcaaggagga catccagaag gcccaggatga gcggccaggg cgacagcctg 2160
cacgagcaca tcgccaacct ggccggcagc cccgccatca agaaggcat cctgcagacc 2220
gtgaaggatg tggacgagct ggtgaaggtg atgggccc acaagcccga gaacatcgtg 2280
atcgagatgg cccgcgagaa ccagaccacc cagaagggcc agaagaacag ccgcgagcgc 2340
atgaagcga tcgaggaggg catcaaggag ctgggcagcc agatcctgaa ggagcacccc 2400
gtggagaaca cccagctgca gaacgagaag ctgtacctgt actacctgca gaacggccgc 2460

gacatgtacg tggaccagga gctggacatc aaccgctga gcgactacga cgtggaccac 2520
atcgtgcccc agagcttctt gaaggacgac agcatcgaca acaaggtgct gaccgcagc 2580
gacaagaacc gcggaagag cgacaacgtg cccagcgagg aggtggtgaa gaagatgaag 2640
aactactggc gccagctgct gaacgccaag ctgatcacc agcgaagtt cgacaacctg 2700
accaaggccg agcgcggcgg cctgagcgag ctggacaagg cggcttcat caagcgccag 2760
ctggtggaga cccgccagat caccaagcac gtggcccaga tcctggacag ccgcatgaac 2820
accaagtacg acgagaacga caagctgatc cgcgaggtga aggtgatcac cctgaagagc 2880

aagctggtga gcgacttccg caaggacttc cagtctaca aggtgcgca gatcaacaac 2940
taccaccagc cccacgacgc ctacctgaac gccgtggtgg gcaccgcct gatcaagaag 3000
taccccaagc tggagagcga gttcgtgtac ggcgactaca aggtgtacga cgtgcaag 3060
atgatcgcca agagcgagca ggagatcggc aaggccaccg ccaagtactt cttctacagc 3120
aacatcatga acttcttcaa gaccgagatc accctggcca acggcgagat ccgcaagcgc 3180
cccctgatcg agaccaacgg cgagaccggc gagatcgtgt gggacaaggg ccgcgacttc 3240

gccaccgtgc gcaaggtgct gagcatgccc caggtgaaca tcgtgaagaa gaccgaggtg 3300

cagaccggcg gcttcagcaa ggagagcatc ctgcccaagc gcaacagcga caagctgata 3360

gcccccaaga aggactggga ccccaagaag tacggcggct tcgacagccc caccgtggcc 3420

tacagcgtgc tgggtggggc caaggtggag aaggccaaga gcaagaagct gaagagcgtg 3480

aaggagctgc tgggcatcac catcatggag cgcagcagct tcgagaagaa ccccatcgac 3540

ttcctggagg ccaagggcta caaggaggtg aagaaggacc tgatcatcaa gctgcccaag 3600

tacagcctgt tcgagctgga gaacggccgc aagcgcagtc tggccagcgc cggcagctg 3660

cagaagggca acgagctggc cctgcccagc aagtacgtga acttctgta cctggccagc 3720

cactacgaga agctgaaggg cagccccgag gacaacgagc agaagcagct gttcgtggag 3780

cagcacaagc actacctgga cgagatcatc gagcagatca gcgagttcag caagcgcgtg 3840

atcctggcgg acgccaacct ggacaaggtg ctgagcgctt acaacaagca ccgcgacaag 3900

cccatccgag agcaggccga gaacatcatc cacctgttca ccctgaccaa cctgggcgcc 3960

cccgcgctt tcaagtactt cgacaccacc atcgaccgca agcgcctacac cagcaccaag 4020

gaggtgctgg acgccacct gatccaccag agcatcaccg gtctgtacga gaccgcatac 4080

gacctgagcc agctgggccc cgacggcggc tccggacctc caaagaaaa gagaaaagta 4140

taccctacg acgtgcccca ctacgcctaa 4170

<210> 107

<211> 4194

<212> DNA

<213> Artificial Sequence

<220><223> Cas9 coding sequence in p3s-Cas9HN (humanized codon, N-term
tagging (underlined), human cell experiments)

<400> 107

atggtgtacc cctacgacgt gcccgactac gccgaattgc ctccaaaaa gaagagaaag 60

gtagggatcc gaattcccgg ggaaaaaccg gacaagaagt acagcatcgg cctggacatc 120

ggtaccaaca gcgtgggctg ggccgtgata accgacgagt acaaggtgcc cagcaagaag 180

ttcaaggtgc tgggcaaac acgaccgcc agcatcaaga agaacctgat cggcgccctg 240

ctgttcgaca gccggcagac cgccgaggcc acccgctga agcgcaccgc ccgcccgcgc 300

tacaccgcc gcaagaaccg catctgtac ctgcaggaga tcttcagcaa cgagatggcc 360

aaggtggagc acagcttctt ccaccgctg gaggagagct tcctgggtga ggaggacaag 420

aagcagagc gccacccat cttcggaac atcgtggagc aggtggccta ccacgagaag 480

tacccacca tctaccact gcgcaagaag ctggtggaca gcaccgacaa ggccgacctg 540
 cgctgatct acctggccct ggcccacatg atcaagtcc gcggccactt cctgatcgag 600

ggcgacctga accccgacaa cagcgacctg gacaagctgt tcatccagct ggtgcagacc 660
 tacaaccagc tgttcgagga gaaccccatc aagccagcg gcgtggacgc caaggccatc 720
 ctgagcgccc gcctgagcaa gagccgccgc ctggagaacc tgatcgcca gctgcccggc 780
 gagaagaaga acggcctgtt cggcaactg atcgccctga gcctgggctt gacccccaac 840
 ttcaagacaa acttcgacct ggccgaggac gccaagctgc agctgagcaa ggacacctac 900
 gacgacgacc tggacaacct gctggcccag atcgccgacc agtacgccga cctgttcctg 960
 gccccaaga acctgagcga cgccatcctg ctgagcgaca tcctgcgcgt gaacaccgag 1020

atcaccaagg cccccctgag cgccagcatg atcaagcgt acgacgagca ccaccaggac 1080
 ctgacctgc tgaaggcctt ggtgcgccag cagctgcccg agaagtaca ggagatcttc 1140
 ttcgaccaga gcaagaacgg ctacccggc tacatcgacg gcggcgccag ccaggaggag 1200
 ttctacaagt tcatcaagcc catcctggag aagatggacg gcaccgagga gctgctggtg 1260
 aagctgaacc gcgaggacct gctgcgcaag cagcgacct tcgacaacgg cagcatcccc 1320
 caccagatcc acctgggcca gctgcacgcc atcctgcgcc gccaggagga cttctacccc 1380
 ttctgaagg acaaccgca gaagatcgag aagatcctga ccttccgat cccctactac 1440

gtgggcccc tggcccgcgg caacagccgc ttcgctgga tgaccgcaa gagcgaggag 1500
 accatcacc cctggaactt cgaggagggtg gtggacaagg gcgccagcgc ccagagcttc 1560
 atcgagcgca tgaccaactt cgacaagaac ctgcccaacg agaaggtgct gccaagcac 1620
 agcctgctgt acgagtactt caccgtgtac aacgagctga ccaaggtgaa gtacgtgacc 1680
 gagggcctgc gcaagcccgc cttcctgagc ggcgagcaga agaaggccat cgtggacctg 1740
 ctgttcaaga ccaaccgcaa ggtgaccgtg aagcagctga aggaggacta cttcaagaag 1800
 atcgagtct tcgacagcgt ggagatcagc ggcgtggagg accgcttcaa cgccagcctg 1860

ggcacctacc acgacctgct gaagatcatc aaggacaagg acttcctgga caacgaggag 1920
 aacgaggaca tcctggagga catcgtgctg accctgacct tgttcgagga ccgcgagatg 1980
 atcgaggagc gcctgaagac ctacgccac ctgttcgacg acaaggtgat gaagcagctg 2040
 aagcggccgc gctacaccgg ctggggccgc ctgagccgca agcttatcaa cggcacccgc 2100
 gacaagcaga gcggaagac catcctggac ttctgaaga gcgacggctt cgccaaccgc 2160
 aacttcatgc agctgatcca cgacgacagc ctgaccttca aggaggacat ccagaaggcc 2220

caggtgagcg gccagggcga cagcctgcac gagcacatcg ccaacctggc cggcagcccc 2280

 gccatcaaga agggcatcct gcagaccgtg aaggtggtgg acgagctggt gaagtgatg 2340
 ggccgccaca agcccagaaa catcgtgatc gagatggccc gcgagaacca gaccaccag 2400
 aagggccaga agaacagccg cgagcgcgatg aagcgcgatc aggaggcat caaggagctg 2460
 ggagccaga tcctgaagga gcaccccggtg gagaacacc agctgcagaa cgagaagctg 2520
 tacctgtact acctgcagaa cggccgcgac atgtacgtgg accaggagct ggacatcaac 2580
 cgctgagcg actacgacgt ggaccacatc gtgcccaga gcttctgaa ggacgacgc 2640
 atcgacaaca aggtgctgac ccgcagcgc aagaaccgcg gcaagagcga caactgccc 2700

 agcgaggagg tggtaagaa gatgaagaac tactggcgc agctgctgaa cgccaagctg 2760
 atcacccagc gcaagttcga caacctgacc aaggccgagc gcggcggcct gagcgagctg 2820
 gacaaggccg gcttcatcaa gcgccagctg gtggagacc gccagatcac caagcacgtg 2880
 gccagatcc tggacagccg catgaacacc aagtacgacg agaacgaaa gctgatccgc 2940
 gaggtgaagg tgatcacct gaagagcaag ctggtgagcg acttccgcaa ggacttccag 3000
 ttctacaagg tgcgcgagat caacaactac caccacgccc acgacgcta cctgaacgcc 3060
 gtggtgggca ccgcctgat caagaagtac cccaagctgg agagcgagtt cgtgtacggc 3120

 gactacaagg tgtacgacgt gcgcaagatg atcgccaaga gcgagcagga gatcggcaag 3180
 gccaccgcca agtacttctt ctacagcaac atcatgaact tcttcaagac cgagatcacc 3240
 ctggccaacg gcgagatccg caagcgcgcc ctgatcgaga ccaacggcga gaccggcgag 3300
 atcgtgtggg acaaggccg cgacttcgcc accgtgcgca aggtgctgag catgcccag 3360
 gtgaacatcg tgaagaagac cgaggtgcag accggcggct tcagcaagga gagcatcctg 3420
 cccaagcga acagcgaaa gctgatgcc cgcaagaagg actgggacc caagaagtac 3480
 ggccgcttcg acagccccc cgtggcctac agcgtgctgg tggggcga ggtggagaag 3540

 ggcaagagca agaagctgaa gagcgtgaag gagctgctgg gcatcacat catggagcg 3600
 agcagcttcg agaagaacc catcgacttc ctggaggcca agggctaaa ggaggtgaag 3660
 aaggacctga tcatcaagct gcccaagtac agcctgttcg agctggagaa cggccgcaag 3720
 cgcatgctgg ccagcgcgg cgagctgcag aaggcaacg agctggcct gccagcaag 3780
 tacgtgaact tcctgtacct ggccagccac tacgagaagc tgaaggcag ccccaggac 3840
 aacgagcaga agcagctgtt cgtggagcag cacaagcact acctggacga gatcatcgag 3900
 cagatcagcg agttcagcaa gcgcgtgatc ctggccgacg ccaacctgga caagtgctg 3960

agcgctaca acaagcaccg cgacaagccc atccgcgagc aggccgagaa catcatccac 4020
 ctgttcaccc tgaccaacct gggcgcccc gccgccttca agtacttcga caccaccatc 4080
 gaccgcaagc gctacaccag caccaaggag gtgctggacg ccaccctgat ccaccagagc 4140
 atcaccggtc tgtacgagac ccgcatcgac ctgagccagc tggcgggcga ctaa 4194
 <210> 108
 <211> 4107
 <212> DNA
 <213> Artificial Sequence
 <220><223> Cas9-coding sequence in Streptococcus pyogenes
 <400> 108
 atggataaga aatactcaat aggcttagat atcggcacia atagcgtcgg atggcggtg 60

 atcactgatg aatataaggt tccgtctaaa aagtcaagg ttctgggaaa tacagaccgc 120
 cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcgga 180
 gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt 240
 tatctacagg agatttttct aatgagatg gcgaaagtag atgatagttt ctttcacga 300
 cttgaagagt ctttttttgg ggaagaagac aagaagcatg aacgtcatcc tatttttga 360
 aatatagtag atgaagttgc ttatcatgag aaatatccaa ctatctatca tctgcgaaaa 420
 aaattgtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgc 480

 atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtgat 540
 gtggacaaac tatttatcca gttggtacaa acctacaate aattatttga agaaaacct 600
 attaacgcaa gtggagtaga tgctaaagcg attctttctg cacgattgag taaatcaaga 660
 cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat 720
 ctcatgtctt tgtcattggg tttgaccctt aattttaat caaattttga tttggcagaa 780
 gatgctaat tacagcttcc aaaagatact tacgatgatg atttagataa tttattggcg 840
 caaattggag atcaatatgc tgatttgttt ttggcagcta agaatttatc agatgctatt 900

 ttactttcag atacctaag agtaataact gaaataacta aggctccctc atcagcttca 960
 atgattaaac gctacgatga acatcatcaa gacttgactc ttttaaaagc tttagttcga 1020
 caacaacttc cagaaaagta taaagaaatc ttttttgatc aatcaaaaaa cggatatgca 1080
 ggttatattg atgggggagc tagccaagaa gaattttata aatttatcaa accaatttta 1140
 gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc 1200
 aagcaacgga cctttgacaa cggtcttatt ccccatcaaa ttcacttggg tgagctgcat 1260

gctatfffga gaagacaaga agactffftat ccatffftaa aagacaatcg tgagaagatt 1320

gaaaaaatct tgactffftcg aattccttat tatgttggtc cattggcgcg tggcaatagt 1380

cgffftgcat ggatgactcg gaagtctgaa gaaacaatta ccccatggaa tffftgaagaa 1440

gttgcgata aagggtcttc agctcaatca tffftattgaac gcatgacaaa cffftgataaa 1500

aatcttccaa atgaaaaagt actacccaaa catagffftgc tffftatgagta tffftacggtt 1560

tataacgaat tgacaaaagg caaatatgtt actgaaggaa tgcgaaaacc agcatttctt 1620

tcaggatgaa agaagaaaag cattgttgat ttactcttca aaacaaatcg aaaagtaacc 1680

gttaagcaat taaaaaga tffftttcaaa aaaatagaat gffftttgatag tgttgaaatt 1740

tcaggagtg aagatagatt taatgcttca ttaggtacct accatgattt gctaaaaatt 1800

attaaagata aagattffftt ggataatgaa gaaaaatgaag atatcttaga ggatattgtt 1860

ttaacattga ccttatttga agataggag atgattgagg aaagacttaa aacatatgct 1920

cacctctttg atgataagg gatgaaacag cttaaacgtc gccgttatac tggttgggga 1980

cgffftgtctc gaaaattgat taatggtatt agggataagc aatctggcaa aacaatatta 2040

gattfffttga aatcagatgg tffftgccaat cgcaatttca tgcagctgat ccatgatgat 2100

agffftgacat ttaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagfffta 2160

catgaacata ttgcaattt agctggtagc cctgctatta aaaaaggat tffftacagact 2220

gtaaaagtg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatctgctt 2280

attgaaatgg cacgtgaaaa tcagacaact caaaaggcc agaaaaattc gcgagagcgt 2340

atgaaacgaa tcgaagaagg tatcaagaa ttaggaagtc agattcttaa agagcatcct 2400

gttgaataa ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga 2460

gacatgatg tggaccaaga attagatatt aatcgffftaa gtgattatga tgtcgatcac 2520

attgtccac aaagfffttct taaagacgat tcaatagaca ataaggtctt aacgcgttct 2580

gataaaaatc gtggtaaaat ggataaacgtt ccaagtgaag aagtagtcaa aaagatgaaa 2640

aactattgga gacaacttct aaacccaag ttaatcactc aacgtaagtt tgataattta 2700

acgaaagctg aacgtggagg tffftgagtga cttgataaag ctggtffftat caaacccaa 2760

ttggttgaat ctgccaat cactaagcat gtggcacaaa tffftggatag tcgcatgaat 2820

actaaatcag atgaaaatga taaacttatt cgagaggffa aagtgattac cttaaaatct 2880

aaattagffft ctgacttccg aaaagatttc caattctata aagtagtga gattaacaat 2940

taccatcatg cccatgatgc gtatctaaat gccgtcgttg gaactgcttt gattaagaaa 3000

tatccaaaac ttgaatcgga gtttgctat ggtgattata aagtttatga tgttcgtaaa 3060
 atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt cttttactct 3120
 aatatcatga acttcttcaa aacagaaatt acacttgcaa atggagagat tcgcaaacgc 3180
 cctctaactg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagatttt 3240
 gccacagtgc gcaaagtatt gtccatgccc caagtcaata ttgtcaagaa aacagaagta 3300
 cagacaggcg gattctcaa ggagtcaatt ttacaaaaa gaaattcgga caagcttatt 3360
 gctcgtaaaa aagactggga tccaaaaaaa tatggtggtt ttgatagtcc aacggtagct 3420

tattcagtcc tagtgggtgc taagtgga aaagggaaat cgaagaagtt aaaatccgtt 3480
 aaagagttac tagggatcac aattatgaa agaagttcct ttgaaaaaaa tccgattgac 3540
 tttttagaag ctaaaggata taaggaagtt aaaaaagact taatcattaa actacctaaa 3600
 tatagtcttt ttgagttaga aaacggctgc aaacggatgc tggctagtgc cggagaatta 3660
 caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttata tttagctagt 3720
 cattatgaaa agttgaaggg tagtccagaa gataacgaac aaaaacaatt gtttgggag 3780
 cagcataagc attatttaga tgagattatt gagcaaatca gtgaattttc taagcgtgtt 3840

at tttagcag atgccaattt agataaagtt cttagtgc atatacaaca tagagacaaa 3900
 ccaatacgtg aacaagcaga aaatattatt cattatttta cgttgacgaa tcttggagct 3960
 cccgctgctt ttaaatattt tgatacaaca attgatcgta aacgatatac gtctacaaaa 4020
 gaagttttag atgccactct tatccatcaa tccatcactg gtctttatga aacacgcatt 4080
 gatttgagtc agctaggagg tgactaa 4107

<210> 109

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Amino acid sequence of Cas9 from *S.pyogenes*

<400> 109

Met Asp Lys Lys Tyr Ser Ile Gly Leu Asp Ile Gly Thr Asn Ser Val

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575

Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
 705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
 740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
 755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 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

1060 1065 1070
 Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg Lys Val Leu Ser

 1075 1080 1085
 Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu Val Gln Thr Gly Gly
 1090 1095 1100
 Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg Asn Ser Asp Lys Leu Ile
 1105 1110 1115 1120
 Ala Arg Lys Lys Asp Trp Asp Pro Lys Lys Tyr Gly Gly Phe Asp Ser
 1125 1130 1135
 Pro Thr Val Ala Tyr Ser Val Leu Val Val Ala Lys Val Glu Lys Gly
 1140 1145 1150

 Lys Ser Lys Lys Leu Lys Ser Val Lys Glu Leu Leu Gly Ile Thr Ile
 1155 1160 1165
 Met Glu Arg Ser Ser Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala
 1170 1175 1180
 Lys Gly Tyr Lys Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys
 1185 1190 1195 1200
 Tyr Ser Leu Phe Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser
 1205 1210 1215
 Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr

 1220 1225 1230
 Val Asn Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser
 1235 1240 1245
 Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
 1250 1255 1260
 Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg Val
 1265 1270 1275 1280
 Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr Asn Lys
 1285 1290 1295

 His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile Ile His Leu
 1300 1305 1310

Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe Lys Tyr Phe Asp
 1315 1320 1325
 Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr Lys Glu Val Leu Asp
 1330 1335 1340
 Ala Thr Leu Ile His Gln Ser Ile Thr Gly Leu Tyr Glu Thr Arg Ile
 1345 1350 1355 1360
 Asp Leu Ser Gln Leu Gly Gly Asp

1365

<210> 110

<211> 4221

<212> DNA

<213> Artificial Sequence

<220><223> Cas9-coding sequence in pET-Cas9N3T for the production of
 recombinant Cas9 protein in E. coli (humanized codon;
 hexa-His-tag and a nuclear localization signal at the N terminus)

<400> 110

atgggcagca gccatcatca tcatcatcat gtgtaccct acgacgtgcc cgactacgcc 60
 gaattgcctc caaaaaagaa gaaaaggta gggatcgaga acctgtactt ccaggcgac 120
 aagaagtaca gcatcggcct ggacatcggg accaacagcg tgggctgggc cgtgatcacc 180

 gacgagtaca aggtgcccag caagaagttc aaggtgctgg gcaacaccga ccgccacagc 240
 atcaagaaga acctgatcgg cgcctgctg ttcgacagcg gcgagaccgc cgaggccacc 300
 cgcctgaagc gcaccgcccg ccgcccgtac acccgccgca agaaccgcat ctgctacctg 360
 caggagatct tcagcaacga gatggccaag gtggacgaca gcttcttcca ccgctggag 420
 gagagcttcc tgggtgagga ggacaagaag cacgagcgcc accccatctt cggcaacatc 480
 gtggacgagg tggcctacca cgagaagtac cccaccatct accacctgcg caagaagctg 540
 gtggacagca ccgacaaggc cgacctgcgc ctgatctacc tggcctggc cccatgatc 600

 aagtccgcg gccacttct gatcgagggc gacctgaacc ccgacaacag cgactggac 660
 aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggagaa ccccatcaac 720
 gccagcggcg tggacgcaa ggccatcctg agcggccgcc tgagcaagag ccgcccctg 780
 gagaacctga tcgccagct gcccggcgag aagaagaacg gcctgttcgg caacctgatc 840
 gccctgagcc tgggctgac cccaacttc aagagcaact tcgacctggc cgaggacgcc 900
 aagctgcagc tgagcaagga cacctacgac gacgacctgg acaacctgct ggcccagatc 960

ggcgaccagt acgccgacct gttcctggcc gccaaagaacc tgagcgacgc catcctgctg 1020

agcgacatcc tgcgcgtgaa caccgagatc accaaggccc ccctgagcgc cagcatgac 1080

aagcgctacg acgagcacca ccaggacctg acctgctga aggccttggc gcgccagcag 1140

ctgcccgaga agtacaagga gatcttcttc gaccagagca agaacggcta cgccggctac 1200

atcgacggcg gcgccagcca ggaggagttc tacaagtta tcaagccat cctggagaag 1260

atggacggca ccgaggagct gctggtgaag ctgaaccgcg aggacctgct gcgcaagcag 1320

cgcaccttcg acaacggcag catccccac cagatccacc tggcgagct gcacccatc 1380

ctgcccgcgc aggaggactt ctacccttc ctgaaggaca accgcgagaa gatcgagaag 1440

atcctgacct tccgcatccc ctactacgtg ggccccctgg cccgcggcaa cagccgcttc 1500

gcctggatga ccccaagag cgaggagacc atcacccctt ggaacttcga ggagtggtg 1560

gacaagggcg ccagcgccea gagcttcatc gagcgatga ccaacttcga caagaacctg 1620

cccaacgaga aggtgctgcc caagcacagc ctgctgtacg agtacttcac cgtgtacaac 1680

gagctgacca aggtgaagta cgtgaccgag ggcatgcga agcccgcctt cctgagcggc 1740

gagcagaaga aggccatcgt ggacctgctg ttcaagacca acccaaggt gaccgtgaag 1800

cagctgaagg aggactactt caagaagatc gagtgcttcg acagcgtgga gatcagcggc 1860

gtggaggacc gcttcaacgc cagcctgggc acctaccagc acctgctgaa gatcatcaag 1920

gacaaggact tcctggacaa cgaggagaac gaggacatcc tggaggacat cgtgctgacc 1980

ctgacctgt tgcaggaccg cgagatgatc gaggagcgc tgaagaccta cgcccactg 2040

ttcgacgaca aggtgatgaa gcagctgaag cgccgccgt acaccggctg gggccgctg 2100

agccgcaagc ttatcaacgg catccgcgac aagcagagcg gcaagacat cctggacttc 2160

ctgaagagcg acggcttcgc caaccgcaac tteatgcagc tgatccacga cgacagcctg 2220

accttcaagg aggacatcca gaaggcccag gtgagcggcc agggcgacag cctgcacgag 2280

cacatcgcca acctggccgg cagcccccc atcaagaagg gcatcctgca gaccgtgaag 2340

gtggtggacg agctggtgaa ggtgatgggc cgccacaagc ccgagaacat cgtgatcgag 2400

atggcccgcg agaaccagac caccagaag ggccagaaga acagccgca gcgcatgaag 2460

cgcatcagg agggcatcaa ggagctgggc agccagatcc tgaaggagca ccccgaggag 2520

aacaccagc tgcagaacga gaagctgtac ctgtactacc tgcagaacgg ccgcgacatg 2580

tacgtggacc aggagctgga catcaaccgc ctgagcgact acgacgtgga ccacatcgtg 2640

ccccagagct tcctgaagga cgacagatc gacaacaagg tgctgaccgg cagcgacaag 2700

aaccgcgca agagcgacaa cgtgcccagc gaggaggtgg tgaagaagat gaagaactac	2760
tggcgccagc tgctgaacgc caagctgac acccagcgca agttcgacaa cctgaccaag	2820
gccgagcgcg gggcctgag cgagctggac aaggccggct tcatcaagcg ccagctggtg	2880
gagacccgcc agatcaccaa gcacgtggcc cagatcctgg acagccgat gaacaccaag	2940
tacgacgaga acgacaagct gatccgcgag gtgaaggtga tcaccctgaa gagcaagctg	3000
gtgagcgact tcccaagga cttccagttc tacaaggtgc gcgagatcaa caactaccac	3060
cacgcccacg acgcctacct gaacgccgtg gtgggcaccg ccctgatcaa gaagtacccc	3120
aagctggaga gcgagttcgt gtacggcgac tacaaggtgt acgacgtcgc caagatgac	3180
gccaaagagcg agcaggagat cggcaaggcc accgccaagt acttcttcta cagcaacatc	3240
atgaacttct tcaagaccga gatcacctg gccaacggcg agatccgcaa gcgccccctg	3300
atcgagacca acggcgagac cggcgagatc gtgtgggaca agggccgca cttcgccacc	3360
gtgcaaggtg tgctgagcat gccccaggtg aacatcgtga agaagaccga ggtgcagacc	3420
ggcggttca gcaaggagag catcctgccc aagcgaaca gcgacaagct gatcgcccgc	3480
aagaaggact gggaccccaa gaagtacggc ggcttcgaca gccccaccgt ggcctacagc	3540
gtgctggtgg tggccaaggt ggagaagggc aagagcaaga agctgaagag cgtgaaggag	3600
ctgctgggca tcaccatcat ggagcgcagc agcttcgaga agaaccctat cgacttctg	3660
gaggccaagg gctacaagga ggtgaagaag gacctgatca tcaagctgcc caagtacagc	3720
ctgttcgagc tggagaacgg ccgcaagcgc atgtggcca gcgccggcga gctgcagaag	3780
ggcaacgagc tggccctgcc cagcaagtac gtgaacttcc tgtacctggc cagccactac	3840
gagaagctga agggcagccc cgaggacaac gagcagaagc agctgttcgt ggagcagcac	3900
aagcactacc tggacgagat catcgagcag atcagcgagt tcagcaagcg cgtgatcctg	3960
gccgacgcca acctggacaa ggtgctgagc gcctacaaca agcaccgca caagccatc	4020
cgcgagcagg ccgagaacat catccactg ttaccctga ccaacctggg cgccccgcc	4080
gcctcaagt acttcgacac caccatgac cgcaagcgt acaccagcac caaggagtg	4140
ctggacgcca ccctgatcca ccagacatc accggtctgt acgagaccg catcgacctg	4200
agccagctgg gcgcgacta a	4221
<210> 111	
<211> 1406	
<212> PRT	
<213> Artificial Sequence	
<220><223> Amino acid sequence of Cas9 (pET-Cas9N3T)	

<400> 111

Met Gly Ser Ser His His His His His His Val Tyr Pro Tyr Asp Val

1 5 10 15

Pro Asp Tyr Ala Glu Leu Pro Pro Lys Lys Lys Arg Lys Val Gly Ile

20 25 30

Glu Asn Leu Tyr Phe Gln Gly Asp Lys Lys Tyr Ser Ile Gly Leu Asp

35 40 45

Ile Gly Thr Asn Ser Val Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys

50 55 60

Val Pro Ser Lys Lys Phe Lys Val Leu Gly Asn Thr Asp Arg His Ser

65 70 75 80

Ile Lys Lys Asn Leu Ile Gly Ala Leu Leu Phe Asp Ser Gly Glu Thr

85 90 95

Ala Glu Ala Thr Arg Leu Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg

100 105 110

Arg Lys Asn Arg Ile Cys Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met

115 120 125

Ala Lys Val Asp Asp Ser Phe Phe His Arg Leu Glu Glu Ser Phe Leu

130 135 140

Val Glu Glu Asp Lys Lys His Glu Arg His Pro Ile Phe Gly Asn Ile

145 150 155 160

Val Asp Glu Val Ala Tyr His Glu Lys Tyr Pro Thr Ile Tyr His Leu

165 170 175

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

180 185 190

Tyr Leu Ala Leu Ala His Met Ile Lys Phe Arg Gly His Phe Leu Ile

195 200 205

Glu Gly Asp Leu Asn Pro Asp Asn Ser Asp Val Asp Lys Leu Phe Ile

210 215 220

Gln Leu Val Gln Thr Tyr Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn

225 230 235 240

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

Asn Ser Arg Phe Ala Trp Met Thr Arg Lys Ser Glu Glu Thr Ile Thr

500 505 510

Pro Trp Asn Phe Glu Glu Val Val Asp Lys Gly Ala Ser Ala Gln Ser

515 520 525

Phe Ile Glu Arg Met Thr Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys

530 535 540

Val Leu Pro Lys His Ser Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn

545 550 555 560

Glu Leu Thr Lys Val Lys Tyr Val Thr Glu Gly Met Arg Lys Pro Ala

565 570 575

Phe Leu Ser Gly Glu Gln Lys Lys Ala Ile Val Asp Leu Leu Phe Lys

580 585 590

Thr Asn Arg Lys Val Thr Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys

595 600 605

Lys Ile Glu Cys Phe Asp Ser Val Glu Ile Ser Gly Val Glu Asp Arg

610 615 620

Phe Asn Ala Ser Leu Gly Thr Tyr His Asp Leu Leu Lys Ile Ile Lys

625 630 635 640

Asp Lys Asp Phe Leu Asp Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp

645 650 655

Ile Val Leu Thr Leu Thr Leu Phe Glu Asp Arg Glu Met Ile Glu Glu

660 665 670

Arg Leu Lys Thr Tyr Ala His Leu Phe Asp Asp Lys Val Met Lys Gln

675 680 685

Leu Lys Arg Arg Arg Tyr Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu

690 695 700

Ile Asn Gly Ile Arg Asp Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe

705 710 715 720

Leu Lys Ser Asp Gly Phe Ala Asn Arg Asn Phe Met Gln Leu Ile His

725 730 735

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

Lys Arg Met Leu Ala Ser Ala Gly Glu Leu Gln Lys Gly Asn Glu Leu
1250 1255 1260

Ala Leu Pro Ser Lys Tyr Val Asn Phe Leu Tyr Leu Ala Ser His Tyr
1265 1270 1275 1280

Glu Lys Leu Lys Gly Ser Pro Glu Asp Asn Glu Gln Lys Gln Leu Phe
1285 1290 1295

Val Glu Gln His Lys His Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser
1300 1305 1310

Glu Phe Ser Lys Arg Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val
1315 1320 1325

Leu Ser Ala Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala
1330 1335 1340

Glu Asn Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala
1345 1350 1355 1360

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

1365 1370 1375

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
1380 1385 1390

Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1395 1400 1405