



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2023년05월10일
(11) 등록번호 10-2531016
(24) 등록일자 2023년05월04일

(51) 국제특허분류(Int. Cl.)
C12N 15/85 (2006.01) A01K 67/027 (2006.01)
C12N 15/10 (2017.01) C12N 15/113 (2010.01)
C12N 15/90 (2006.01) C12N 9/22 (2006.01)
C12Q 1/6888 (2018.01)
(52) CPC특허분류
C12N 15/8509 (2013.01)
A01K 67/0275 (2013.01)
(21) 출원번호 10-2022-7021859(분할)
(22) 출원일자(국제) 2015년11월20일
심사청구일자 2022년06월27일
(85) 번역문제출일자 2022년06월27일
(65) 공개번호 10-2022-0093013
(43) 공개일자 2022년07월04일
(62) 원출원 특허 10-2017-7016481
원출원일자(국제) 2015년11월20일
심사청구일자 2020년10월23일
(86) 국제출원번호 PCT/US2015/062023
(87) 국제공개번호 WO 2016/081923
국제공개일자 2016년05월26일
(30) 우선권주장
62/083,005 2014년11월21일 미국(US)
(뒷면에 계속)
(56) 선행기술조사문헌
WO2014104878 A1
(뒷면에 계속)

(73) 특허권자
리제너론 파마슈티칼스 인코포레이티드
미합중국 뉴욕주 10591 타리타운 올드 소우 밀 리버 로드 777
(72) 발명자
가글리아르디 앤써니
미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드 777 리제너론 파마슈티칼스 인코포레이티드 내
드로켓 구스타보
미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드 777 리제너론 파마슈티칼스 인코포레이티드 내
(뒷면에 계속)
(74) 대리인
특허법인와이에스장

전체 청구항 수 : 총 44 항

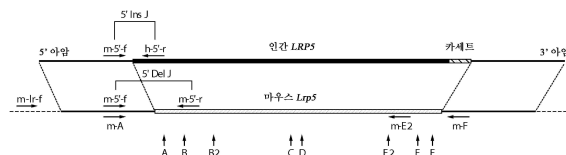
심사관 : 최성호

(54) 발명의 명칭 쌍 형성된 가이드 RNA를 사용하는 표적화된 유전자 변형을 위한 방법 및 조성물

(57) 요약

세포 내의 게놈에 대한 이중대립유전자 표적화된 변형을 생성하고 촉진하기 위한, 그리고 변형된 게놈을 포함하는 비인간 동물을 생성하기 위한 조성물 및 방법이 제공된다. 또한, 대립유전자에 대해 이형접합성인 세포 내의 게놈을 세포가 그러한 대립유전자에 대해 동형접합성이 되도록 변형시키기 위한 조성물 및 방법이 제공된다. 본 방법은 상기 게놈 표적 유전자와 내의 상이한 위치를 표적화하는 2개 이상의 가이드 RNA 및 Cas 단백질을 사용한다. 또한, 변형된 게놈을 갖는 세포를 확인하는 방법이 제공된다.

대표도 - 도1



(52) CPC특허분류

C12N 15/102 (2013.01)
C12N 15/113 (2013.01)
C12N 15/907 (2013.01)
C12N 9/22 (2013.01)
C12Q 1/6888 (2018.05)
A01K 2227/105 (2013.01)
C12N 2310/20 (2017.05)
C12Q 2537/16 (2013.01)

(72) 발명자

라이 카-만 비너스

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드
 777 리제너론 파마슈티칼스 인코포레이티드 내

맥도날드 린

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드
 777 리제너론 파마슈티칼스 인코포레이티드 내

머피 앤드류 제이.

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드
 777 리제너론 파마슈티칼스 인코포레이티드 내

발렌주엘라 데이비드 엠.

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드
 777 리제너론 파마슈티칼스 인코포레이티드 내

보로니나 베라

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드
 777 리제너론 파마슈티칼스 인코포레이티드 내

아우어바흐 보이테크

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드
 777 리제너론 파마슈티칼스 인코포레이티드 내

양코폴로스 조지 디.

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드
 777 리제너론 파마슈티칼스 인코포레이티드 내

프렌듀이 데이비드

미국 뉴욕 10591 테리타운 올드 소우 밀 리버 로드
 777 리제너론 파마슈티칼스 인코포레이티드 내

(56) 선행기술조사문헌

Nat Commun.,doi:10.1038/ncomms5240(2014.6.26.)
 Nucleic Acids Res.,43(3):e21.doi:
 10.1093/nar/gku1246(2014.11.20.)
 US 200500144655 A1
 BBRC.,445(4):791-794(2014.1.)

(30) 우선권주장

62/182,314 2015년06월19일 미국(US)
 62/211,421 2015년08월28일 미국(US)

명세서

청구범위

청구항 1

제1 대립유전자(allele)에 대해 이형접합성인 세포 내의 게놈을 변형시키는 방법으로서, 상기 세포는 시험관내(*in vitro*) 세포 또는 생체내(*in vivo*) 비인간 세포이고,

(I) 상기 게놈을

(a) Cas 단백질;

(b) 제1 비-대립유전자-특이적 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA 및 제1 tracrRNA를 포함하는 제1 가이드 RNA와 접촉시키는 단계로서,

상기 제1 대립유전자는 제1 상동 염색체 상에 있고, 상기 CRISPR RNA 인식 서열은 제2 상동 염색체 상의 상기 제1 대립유전자에 상응하는 유전자좌에 대해 동원체 측에 있고(centromeric),

상기 Cas 단백질과 상기 제1 가이드 RNA는 자연적으로 함께 발생되지 않고,

상기 Cas 단백질은 상기 제1 CRISPR RNA 인식 서열을 절단하여 이중-가닥 브레이크를 생성하고 상기 세포는 변형되어 상기 제1 대립유전자에 대해 동형접합성이 되는, 단계; 및

(II) 제1 대립유전자에 대해 동형접합성인 변형된 세포를 확인하는 단계

를 포함하는 방법.

청구항 2

제1 항에 있어서, 단계 (I)은 (a) Cas 단백질 또는 Cas 단백질을 인코딩하는 핵산; 및 (b) 제1 가이드 RNA 또는 제1 가이드 RNA를 인코딩하는 DNA를 세포로 도입하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 3

제2 항에 있어서, 단계 (I)은 Cas 단백질을 인코딩하는 핵산을 세포 내로 도입하고, 제1 가이드 RNA를 세포 내로 도입하는 것을 포함하고, 상기 Cas 단백질을 인코딩하는 핵산은 RNA인 것을 특징으로 하는 방법.

청구항 4

제2 항에 있어서, 단계 (I)은 Cas 단백질을 인코딩하는 핵산을 세포 내로 도입하고, 제1 가이드 RNA를 인코딩하는 DNA를 세포 내로 도입하는 것을 포함하고, 상기 Cas 단백질을 인코딩하는 핵산은 DNA인 것을 특징으로 하는 방법.

청구항 5

제2 항에 있어서, Cas 단백질 및 제1 가이드 RNA는 제1 단백질-RNA 복합체로서 세포 내로 도입되는 것을 특징으로 하는 방법.

청구항 6

제4 항에 있어서, 제1 가이드 RNA, 제1 CRISPR RNA, 및 제1 tracrRNA는 별개의 RNA 분자이고,

(a) 상기 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나;

(b) 상기 제1 CRISPR RNA를 인코딩하는 DNA는 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되고/되거나;

(c) 상기 제1 tracrRNA를 인코딩하는 DNA는 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되며, 상기 제1, 제2, 및 제3 프로모터는 상기 세포 내에서 활성인 것을 특징으로 하는 방법.

청구항 7

제6 항에 있어서, 제1, 제2 및 제3 발현 작제물은 단일 핵산 분자의 성분인 것을 특징으로 하는 방법.

청구항 8

제4 항에 있어서,

- (1) 상기 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나;
- (b) 상기 제1 CRISPR RNA를 인코딩하는 DNA는 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되며, 상기 제1 및 제2 프로모터는 상기 세포 내에서 활성화된 것을 특징으로 하는 방법.

청구항 9

제8 항에 있어서, 제1 및 제2 발현 작제물은 단일 핵산 분자의 성분인 것을 특징으로 하는 방법.

청구항 10

제1 항 내지 제9 항 중 어느 한 항에 있어서, 이형접합성의 상실이 상기 이중-가닥 브레이크의 말단소체 측에서 (telomeric) 일어나는 것을 특징으로 하는 방법.

청구항 11

제1 항 내지 제9 항 중 어느 한 항에 있어서, 단계 (I)은 게놈을 (c) 상기 제1 대립유전자에 상응하는 유전자좌에 대해 동원체 측에 있는 제2 비-대립유전자-특이적 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA 및 제2 tracrRNA를 포함하는 제2 가이드 RNA와 접촉시키는 단계를 추가로 포함하고,

상기 Cas 단백질은 상기 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여 적어도 하나의 이중-가닥 브레이크를 생성하는 것을 특징으로 하는 방법.

청구항 12

제11 항에 있어서, 단계 (I)은 제2 가이드 RNA 또는 제2 가이드 RNA를 인코딩하는 DNA를 세포 내로 도입하는 것을 특징으로 하는 방법.

청구항 13

제11 항에 있어서, Cas 단백질은 상기 제1 CRISPR RNA 인식 서열 및 상기 제2 CRISPR RNA 인식 서열을 절단하는 것을 특징으로 하는 방법.

청구항 14

제11 항에 있어서, 상기 제1 및 상기 제2 CRISPR RNA 인식 서열은 상기 제1 상동 염색체 상은 아니고 상기 제2 상동 염색체 상에 위치하는 것을 특징으로 하는 방법.

청구항 15

제11 항에 있어서, 상기 Cas 단백질은 니카제(nickase)인 것을 특징으로 하는 방법.

청구항 16

제15 항에 있어서, 상기 Cas 단백질은 니카제(nickase)이고,

상기 단계 (I)은 상기 게놈을

- (d) 상기 제1 대립유전자에 상응하는 제2 상동 염색체 상의 유전자좌에 대해 동원체 측에 있는 제3 CRISPR RNA 인식 서열에 혼성화되는 제3 CRISPR RNA 및 제3 tracrRNA를 포함하는 제3 가이드 RNA; 및
- (e) 상기 제1 대립유전자에 상응하는 제2 상동 염색체 상의 유전자좌에 대해 동원체 측에 있는 제4 CRISPR RNA 인식 서열에 혼성화되는 제4 CRISPR RNA 및 제4 tracrRNA를 포함하는 제4 가이드 RNA와 접촉시키는 단계를 추가로 포함하고,

상기 Cas 단백질은 상기 제1 및 제2 CRISPR RNA 인식 서열 내의 게놈 DNA의 제1 가닥을 절단하고, 상기 Cas 단백질은 상기 제3 및 제4 CRISPR RNA 인식 서열 내의 게놈 DNA의 제2 가닥을 절단하고,

상기 제1 및 제3 CRISPR RNA 인식 서열 내의 절단은 엇갈린 말단(staggered end)을 갖는 제1 이중-가닥 브레이크를 생성하기 위한 오프-셋(off-set) Nicking, 상기 제2 및 제4 CRISPR RNA 인식 서열 내의 절단은 엇갈린 말단(staggered end)을 갖는 제2 이중-가닥 브레이크를 생성하기 위한 오프-셋(off-set) Nicking 것을 특징으로 하는 방법.

청구항 17

제16 항에 있어서, 단계 (I)은 (d) 제3 가이드 RNA 또는 제3 가이드 RNA를 인코딩하는 DNA; 및 (e) 제4 가이드 RNA 또는 제4 가이드 RNA를 인코딩하는 DNA를 세포 내로 도입하는 것을 특징으로 하는 방법.

청구항 18

제1 항 내지 제9 항 중 어느 한 항에 있어서, Cas 단백질은 이중-가닥 DNA의 양쪽 가닥에 대해 뉴클레아제 활성을 갖는 것을 특징으로 하는 방법.

청구항 19

제1 항 내지 제9 항 중 어느 한 항에 있어서, 상기 제1 CRISPR RNA 인식 서열은 동원체로부터 100 bp 내지 1 kb, 1 kb 내지 10 kb, 10 kb 내지 100 kb, 100 kb 내지 1 Mb, 1 Mb 내지 10 Mb, 10 Mb 내지 20 Mb, 20 Mb 내지 30 Mb, 30 Mb 내지 40 Mb, 40 Mb 내지 50 Mb, 50 Mb 내지 60 Mb, 60 Mb 내지 70 Mb, 70 Mb 내지 80 Mb, 80 Mb 내지 90 Mb, 또는 90 Mb 내지 100 Mb만큼 떨어져 있는 것을 특징으로 하는 방법.

청구항 20

제1 항 내지 제9 항 중 어느 한 항에 있어서, 상기 제1 대립유전자는 상기 제1 CRISPR RNA 인식 서열로부터 100 bp 내지 1 kb, 1 kb 내지 10 kb, 10 kb 내지 100 kb, 100 kb 내지 1 Mb, 1 Mb 내지 10 Mb, 10 Mb 내지 20 Mb, 20 Mb 내지 30 Mb, 30 Mb 내지 40 Mb, 40 Mb 내지 50 Mb, 50 Mb 내지 60 Mb, 60 Mb 내지 70 Mb, 70 Mb 내지 80 Mb, 80 Mb 내지 90 Mb, 또는 90 Mb 내지 100 Mb만큼 떨어져 있는 것을 특징으로 하는 방법.

청구항 21

제1 항 내지 제9 항 중 어느 한 항에 있어서, 상기 제1 대립유전자는 상기 제1 CRISPR RNA 인식 서열로부터 적어도 100 bp, 적어도 1 kb, 적어도 10 kb, 적어도 100 kb, 적어도 1 Mb, 적어도 10 Mb, 적어도 20 Mb, 적어도 30 Mb, 적어도 40 Mb, 적어도 50 Mb, 적어도 60 Mb, 적어도 70 Mb, 적어도 80 Mb, 적어도 90 Mb, 또는 적어도 100 Mb만큼 떨어져 있는 것을 특징으로 하는 방법.

청구항 22

제1 항 내지 제9 항 중 어느 한 항에 있어서, 이형접합성의 상실에 의해 대체되는 제2 상동 염색체의 영역이 100 bp 내지 1 kb, 1 kb 내지 10 kb, 10 kb 내지 100 kb, 100 kb 내지 1 Mb, 1 Mb 내지 10 Mb, 10 Mb 내지 20 Mb, 20 Mb 내지 30 Mb, 30 Mb 내지 40 Mb, 40 Mb 내지 50 Mb, 50 Mb 내지 60 Mb, 60 Mb 내지 70 Mb, 70 Mb 내지 80 Mb, 80 Mb 내지 90 Mb, 또는 90 Mb 내지 100 Mb인 것을 특징으로 하는 방법.

청구항 23

제1 항 내지 제9 항 중 어느 한 항에 있어서, 이형접합성의 상실에 의해 대체되는 제2 상동 염색체의 영역이 적어도 100 bp, 적어도 1 kb, 적어도 10 kb, 적어도 100 kb, 적어도 1 Mb, 적어도 10 Mb, 적어도 20 Mb, 적어도 30 Mb, 적어도 40 Mb, 적어도 50 Mb, 적어도 60 Mb, 적어도 70 Mb, 적어도 80 Mb, 적어도 90 Mb, 또는 적어도 100 Mb인 것을 특징으로 하는 방법.

청구항 24

제1 항 내지 제9 항 중 어느 한 항에 있어서,

(1) 제1 대립유전자는 돌연변이를 포함하거나; 또는

(2) 제1 대립유전자는 야생형 대립유전자이고, 제2 상동 염색체 상의 상응하는 유전자좌는 돌연변이를 포함하는 것을 특징으로 하는 방법.

청구항 25

제24 항에 있어서, 제1 대립유전자는 돌연변이를 포함하고, 상기 돌연변이는 표적화 변형인 것을 특징으로 하는 방법.

청구항 26

제1 항 내지 제9 항 중 어느 한 항에 있어서, 상기 세포는 진핵세포인 것을 특징으로 하는 방법.

청구항 27

제26 항에 있어서, 상기 진핵 세포는 포유류 세포, 인간 세포, 비인간 세포, 설치류 세포, 마우스 세포, 래트 세포, 다능성(pluripotent) 세포, 비다능성 세포, 비인간 다능성 세포, 인간 다능성 세포, 설치류 다능성 세포, 마우스 다능성 세포, 래트 다능성 세포, 마우스 배아 줄기(ES) 세포, 래트 ES 세포, 비인간 포유류 ES 세포, 인간 성인 줄기 세포, 발생적으로 제한된 인간 선조 세포, 인간 유도 다능성 줄기(iPS) 세포, 비인간 포유류 1-세포 배아, 또는 마우스 1-세포 배아인 것을 특징으로 하는 방법.

청구항 28

제27 항에 있어서, 상기 진핵 세포는 포유류 세포인 것을 특징으로 하는 방법.

청구항 29

제27 항에 있어서, 상기 진핵 세포는 인간 iPS 세포인 것을 특징으로 하는 방법.

청구항 30

제27 항에 있어서, 상기 진핵 세포는 마우스 ES 세포 또는 래트 ES 세포인 것을 특징으로 하는 방법.

청구항 31

제30 항에 있어서, 상기 진핵 세포는 마우스 ES 세포인 것을 특징으로 하는 방법.

청구항 32

제1 항 내지 제9 항 중 어느 한 항에 있어서, Cas 단백질은 Cas9인 것을 특징으로 하는 방법.

청구항 33

제1 항 내지 제9 항 중 어느 한 항에 있어서, 제1 가이드 RNA 내의 제 CRISPR RNA 및 제1 tracr RNA는 함께 융합된 것을 특징으로 하는 방법.

청구항 34

제1 항 내지 제9 항 중 어느 한 항에 있어서, 제 CRISPR RNA 및 제1 tracr RNA는 별개의 RNA 분자인 것을 특징으로 하는 방법.

청구항 35

제1 항 내지 제9 항 중 어느 한 항에 있어서, 상기 세포는 비상동 말단 결합(non-homologous end joining)(NHEJ)을 감소시키고/시키거나 유전자 변환 또는 상동성-유도 수복(homology-directed repair)(HDR)을 증가시키도록 변형된 것을 특징으로 하는 방법.

청구항 36

제35 항에 있어서, 상기 세포는 하기 중 하나 이상의 발현 또는 활성을 감소시키도록 변형된 것을 특징으로 하는 방법: DNA-PK, PARP1, 및 리가제 IV.

청구항 37

제36 항에 있어서, 상기 발현 또는 활성의 감소는 유도성, 가역적, 시간 특이적, 및/또는 공간 특이적인 것을 특징으로 하는 방법.

청구항 38

제1 항 내지 제9 항 중 어느 한 항에 있어서, 단계 (II)의 확인하는 단계는 대립유전자의 변형 검정을 포함하는 것을 특징으로 하는 방법.

청구항 39

제1 항 내지 제9 항 중 어느 한 항에 있어서, 단계 (II)의 확인하는 단계는 이형접합성의 상실에 대해 스크리닝하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 40

제39 항에 있어서, 스크리닝은 제1 CRISPR RNA 인식 서열의 말단소체 측에서의 단일 뉴클레오티드 변이체 대립유전자 식별 검정 및/또는 구조적 변이체 다형성 PCR 검정을 수행하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 41

제40 항에 있어서, 스크리닝은 제1 CRISPR RNA 인식 서열의 말단소체 측 상이한 위치에서의 다수의 단일 뉴클레오티드 변이체 대립유전자 식별 검정 및/또는 다수의 구조적 변이체 다형성 PCR 검정을 포함하는 것을 특징으로 하는 방법.

청구항 42

제1 항 내지 제9 항 중 어느 한 항에 있어서, 상기 세포는 하나 이상의 추가 대립유전자에 대해 이형접합성이고, 제1 CRISPR RNA 인식 서열은 하나 이상의 추가 대립유전자에 상응하는 제2 상동 염색체 상의 유전자좌에 대해 동원체 측에 있고, 상기 세포가 하나 이상의 추가 대립유전자에 대해 동형 접합성이 되도록 변형되고, 단계 (II)는 하나 이상의 추가 대립유전자에 대해 동형 접합성인 변형된 세포를 확인하는 것을 포함하는 것을 특징으로 하는 방법.

청구항 43

제1 항 내지 제9 항 중 어느 한 항에 있어서, 상기 제1 대립유전자는 상기 제1 CRISPR RNA 인식 서열로부터 적어도 10kb 또는 적어도 100 kb만큼 떨어져 있는 것을 특징으로 하는 방법.

청구항 44

제43 항에 있어서, 상기 제1 대립유전자는 상기 제1 CRISPR RNA 인식 서열로부터 적어도 100 kb만큼 떨어져 있는 것을 특징으로 하는 방법.

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

청구항 60

삭제

청구항 61

삭제

청구항 62

삭제

청구항 63

삭제

청구항 64

삭제

청구항 65

삭제

청구항 66

삭제

청구항 67

삭제

청구항 68

삭제

청구항 69

삭제

청구항 70

삭제

청구항 71

삭제

청구항 72

삭제

청구항 73

삭제

청구항 74

삭제

청구항 75

삭제

청구항 76

삭제

청구항 77

삭제

청구항 78

삭제

청구항 79

삭제

청구항 80

삭제

청구항 81

삭제

발명의 설명

기술 분야

[0001] 관련 출원과의 상호 참조

[0002] 본 출원은 2014년 11월 21일에 출원된 미국 특허 출원 제62/083,005호, 2015년 6월 19일에 출원된 미국 특허 출원 제62/182,314호 및 2015년 8월 28일에 출원된 미국 특허 출원 제62/211,421호의 이익을 주장하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0003] EFS 웹을 통해 텍스트 파일로서 제시된 서열 목록에 대한 참조

[0004] 파일 472225SEQLIST.txt에 기재된 서열 목록은 32.7 kb이고, 2015년 11월 20일에 생성되었으며, 본 명세서에 참고로 포함된다.

배경 기술

[0005] 다양한 게놈 유전자좌를 표적화하는 데 있어서 진보가 이루어져 왔지만, 효율적으로 표적화될 수 없는 많은 유형의 게놈 유전자좌 또는 종래의 표적화 전략으로 적절하게 또는 효율적으로 달성될 수 없는 게놈 변형이 여전히 남아 있다. 예를 들어, 큰 표적화된 게놈 결실 또는 다른 큰 표적화된 유전자 변형을 생성하고자 시도할 때 어려움이 발생되는데, 이는 진행 세포 및 유기체에서 특히 그러하다.

[0006] 특히, 종래의 표적화 전략을 사용하는 경우 큰 표적화된 게놈 결실 또는 다른 게놈 변형에 대해 동형접합성(homozygous) 또는 복합 이형접합성(compound heterozygous)(예를 들어, 반접합성(hemizygous))인 세포 또는 동물을 효율적으로 생성하는 것이 어렵다. 예를 들어, 큰 표적화된 게놈 결실에 대해 이형접합성인 F0 세대 마우스가 종래의 표적화 전략을 통해 얻어질 수 있지만, 후속으로, 결실에 대해 동형접합성인 F2 세대 마우스를 생성하기 위해 이러한 이형접합성 마우스의 교배가 요구된다. 이들 추가의 교배 단계는 비용이 많이 들고 시간 소모적이다.

발명의 내용

[0007] 세포 내에서 게놈을 변형시키는 방법 및 조성물이 제공된다. 일 태양에서, 본 발명은 세포 내의 게놈에 대한 변형을 생성하는 방법을 제공하며, 본 방법은 상기 게놈을 (a) 제1 Cas 단백질; (b) 게놈 표적 유전자좌 내의 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA; (c) 상기 게놈 표적 유전자좌 내의 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA; (d) tracrRNA; 및 (e) 5' 표적 서열에 혼성화되는 5' 상동성 아암(homology arm) 및 3' 표적 서열에 혼성화되는 3' 상동성 아암에 의해 플랭킹된(flanked) 핵산 삽입물을 포함하는 표적화 벡터 - 단, 상기 세포가 1-세포기 배아인 경우, 상기 표적화 벡터는 5 kb 길이 이하임 - 와 접촉시키는 단계를 포함하며, 상기 게놈은 상기 게놈 표적 유전자좌를 포함하는 한 쌍의 제1 및 제2 상동 염색체를 포함하고; 상기 제1 Cas 단백질은 상기 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여, 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 적어도 하나의 이중-가닥 브레이크를 생성한다. 일 태양에서, 본 발명은 세포 내의 게놈에 대한 이중대립유전자(biallelic) 변형을 생성하는 방법을 제공하며, 본 방법은 상기 게놈을 (a) 제1 Cas 단백질; (b) 게놈 표적 유전자좌 내의 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA; (c) 상기 게놈 표적 유전자좌 내의 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA; (d) tracrRNA; 및 (e) 5' 표적 서열에 혼성화되는 5' 상동성 아암 및 3' 표적 서열에 혼성화되는 3' 상동성 아암에 의해 플랭킹된 핵산 삽입물을 포함하는 표적화 벡터 - 단, 상기 세포가 1-세포기 배아인 경우, 상기 표적화 벡터는 5 kb 길이 이하임 - 와 접촉시키는 단계를 포함하며, 상기 게놈은 상기 게놈 표적 유전자좌를 포함하는 한 쌍의 제1 및 제2 상동 염색체를 포함하고; 상기 제1 Cas 단백질은 상기 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여, 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 적어도 하나의 이중-가닥 브레이크를 생성한다.

[0008] 본 방법은 상기 변형된 게놈을 포함하는 세포를 확인하는 단계를 추가로 포함할 수 있다. 일부 방법에서, 상기 핵산 삽입물은 제1 표적 서열에 혼성화되는 제1 상동성 아암에 인접한 선택 카세트를 포함하며, 상기 제1 상동성 아암은 상기 5' 상동성 아암이고 상기 제1 표적 서열은 상기 5' 표적 서열이거나, 또는 상기 제1 상동성 아암은 상기 3' 상동성 아암이고 상기 제1 표적 서열은 상기 3' 표적 서열이고, 상기 확인하는 단계는 (a) 상기 세포로부터 DNA를 얻는 단계; (b) 상기 세포의 DNA를 상기 제1 표적 서열 내에서 결합하는 프로브(probe), 상기

핵산 삽입물 내에서 결합하는 프로브, 및 기지의 카피수를 갖는 참조 유전자 내에서 결합하는 프로브에 노출시키는 단계 - 여기서, 각각의 프로브는 결합 시에 검출가능한 신호를 발생시킴 -; (c) 상기 프로브들 각각의 결합으로부터 신호를 검출하는 단계; 및 (d) 상기 참조 유전자 프로브로부터의 신호를 상기 제1 표적 서열 프로브로부터의 신호와 비교하여 상기 제1 표적 서열에 대한 카피수를 결정하고, 상기 참조 유전자 프로브로부터의 신호를 상기 핵산 삽입물 프로브로부터의 신호와 비교하여 상기 핵산 삽입물에 대한 카피수를 결정하는 단계를 포함하며, 1개 또는 2개의 핵산 삽입물 카피수 및 2개의 제1 표적 서열 카피수는 상기 게놈 표적 유전자좌에서의 상기 핵산 삽입물의 표적화된 삽입을 나타내고, 1개 이상의 핵산 삽입물 카피수 및 3개 이상의 제1 표적 서열 카피수는 상기 게놈 표적 유전자좌 이외의 게놈 유전자좌에서의 상기 핵산 삽입물의 무작위 삽입을 나타낸다.

[0009] 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1 및 제2 상동 염색체 각각에서 상기 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여, 상기 제1 및 제2 상동 염색체 각각에서 적어도 하나의 이중-가닥 브레이크를 생성한다. 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 상기 제1 및 제2 CRISPR RNA 인식 서열을 절단하여, 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 적어도 2개의 이중-가닥 브레이크를 생성한다.

[0010] 일부 방법은 상기 게놈을 상기 게놈 표적 유전자좌 내의 제3 CRISPR RNA 인식 서열에 혼성화되는 제3 CRISPR RNA; 및 상기 게놈 표적 유전자좌 내의 제4 CRISPR RNA 인식 서열에 혼성화되는 제4 CRISPR RNA와 접촉시키는 단계를 추가로 포함한다. 선택적으로, 상기 제1 CRISPR RNA 인식 서열과 상기 제3 CRISPR RNA 인식 서열은 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 약 900 bp 내지 약 1 kb, 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 3 kb, 약 3 kb 내지 약 4 kb, 약 4 kb 내지 약 5 kb, 약 5 kb 내지 약 6 kb, 약 6 kb 내지 약 7 kb, 약 7 kb 내지 약 8 kb, 약 8 kb 내지 약 9 kb, 약 9 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 또는 약 90 kb 내지 약 100 kb만큼 분리된다. 선택적으로, 상기 제2 CRISPR RNA 인식 서열과 제4 CRISPR RNA 인식 서열은 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 약 900 bp 내지 약 1 kb, 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 3 kb, 약 3 kb 내지 약 4 kb, 약 4 kb 내지 약 5 kb, 약 5 kb 내지 약 6 kb, 약 6 kb 내지 약 7 kb, 약 7 kb 내지 약 8 kb, 약 8 kb 내지 약 9 kb, 약 9 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 또는 약 90 kb 내지 약 100 kb만큼 분리된다. 선택적으로, 상기 제1 및 제3 CRISPR RNA 인식 서열이 제1 쌍의 CRISPR RNA 인식 서열이고, 상기 제2 및 제4 CRISPR RNA 인식 서열이 제2 쌍의 CRISPR RNA 인식 서열이며, 상기 제1 쌍과 제2 쌍은 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 약 900 bp 내지 약 1 kb, 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 약 2.5 Mb 내지 약 3 Mb, 약 3 Mb 내지 약 4 Mb, 약 4 Mb 내지 약 5 Mb, 약 5 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb만큼 분리된다.

[0011] 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1, 제2, 제3 및 제4 CRISPR RNA 인식 서열 중 적어도 2개를 절단하여, 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 적어도 2개의 이중-가닥 브레이크를 생성한다. 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1, 제2, 제3 및 제4 CRISPR RNA 인식 서열 중 적어도 2개를 절단하여,

상기 제1 및 제2 상동 염색체 둘 모두에서 적어도 2개의 이중-가닥 브레이크를 생성한다.

[0012] 일부 방법에서, 상기 핵산 삽입물은 상기 5' 표적 서열과 상기 3' 표적 서열 사이에 삽입된다. 선택적으로, 상기 5' 및 3' 표적 서열은 상기 게놈 표적 유전자좌 내에 있다. 선택적으로, 상기 세포는 1-세포기 배아가 아니고, 상기 표적화 벡터는 적어도 10 kb인 대형 표적화 벡터(large targeting vector)(LTVEC)이다.

[0013] 일부 방법에서, 상기 게놈을 상기 제1 및 제2 CRISPR RNA 둘 모두와 접촉시키는 것은 상기 게놈을 단독으로서의 상기 제1 CRISPR RNA 또는 제2 CRISPR RNA와 접촉시키는 것과 비교하여 증가된 이중대립유전자 변형 효율을 가져온다. 일부 방법에서, 상기 세포는 이배체이고, 상기 이중대립유전자 변형은 상기 게놈 표적 유전자좌에서 동형접합성(homozygosity) 또는 복합 이형접합성(compound heterozygosity)을 가져온다. 선택적으로, 상기 복합 이형접합성은 반접합성이다. 일부 방법에서, 상기 이중대립유전자 변형은 상기 제1 상동 염색체에서의 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열 사이의 결실을 포함한다. 일부 방법에서, 상기 이중대립유전자 변형은 상기 제1 및 제2 상동 염색체 둘 모두에서의 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열 사이의 결실을 포함한다. 일부 방법에서, 상기 이중대립유전자 변형은 상기 제1 및 제2 상동 염색체 둘 모두에서의 상기 5' 표적 서열과 상기 3' 표적 서열 사이의 상기 핵산 삽입물의 삽입을 추가로 포함한다. 일부 방법에서, 이중대립유전자 변형은 (1) 상기 제1 및 제2 상동 염색체 둘 모두에서의 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열 사이의 결실; 및 (2) 상기 제2 상동 염색체에서는 아니고 상기 제1 상동 염색체에서의 상기 5' 표적 서열과 상기 3' 표적 서열 사이의 상기 핵산 삽입물의 삽입을 포함한다. 일부 방법에서, 이중대립유전자 변형은 (1) 상기 제1 상동 염색체에서의 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열 사이의 결실; 및 (2) 상기 제2 상동 염색체에서의 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열 사이의 유전자좌의 파괴를 포함한다. 일부 방법에서, 이중대립유전자 변형은 (1) 상기 제1 상동 염색체에서의 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열 사이의 결실; (2) 상기 제1 상동 염색체에서의 상기 5' 표적 서열과 상기 3' 표적 서열 사이의 상기 핵산 삽입물의 삽입; 및 (3) 상기 제2 상동 염색체에서의 상기 5' 표적 서열과 상기 3' 표적 서열 사이의 유전자좌의 파괴를 포함한다. 일부 방법에서, 상기 이중대립유전자 변형은 (1) 상기 제1 상동 염색체에서의 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열 사이의 결실; 및 (2) 상기 제1 상동 염색체에서의 상기 5' 표적 서열과 상기 3' 표적 서열 사이의 상기 핵산 삽입물의 삽입을 포함하며, 상기 핵산 삽입물 서열은 상기 결실된 서열과 상동성 또는 오르토로그성(orthologous)이다.

[0014] 일부 방법에서, 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열은 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb만큼 분리된다. 일부 방법에서, 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열은 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb만큼 분리된다. 일부 방법에서, 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열은 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 또는 약 900 bp 내지 약 1 kb만큼 분리된다. 일부 방법에서, 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열은 25 bp 미만, 50 bp 미만, 100 bp 미만, 150 bp 미만, 200 bp 미만, 250 bp 미만, 300 bp 미만, 350 bp 미만, 400 bp 미만, 450 bp 미만, 500 bp 미만, 600 bp 미만, 700 bp 미만, 800 bp 미만, 900 bp 미만, 1 kb 미만, 2 kb 미만, 3 kb 미만, 4 kb 미만, 5 kb 미만, 또는 10 kb 미만만큼 분리된다.

[0015] 일부 방법에서, 상기 제1 및 제2 CRISPR RNA 인식 서열은 각각 상기 5' 및 3' 표적 서열 둘 모두로부터 적어도 50 bp, 적어도 100 bp, 적어도 200 bp, 적어도 300 bp, 적어도 400 bp, 적어도 500 bp, 적어도 600 bp, 적어도 700 bp, 적어도 800 bp, 적어도 900 bp, 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 6 kb, 적어도 7 kb, 적어도 8 kb, 적어도 9 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40

kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 또는 적어도 100 kb만큼 떨어져서 위치된다. 일부 방법에서, 상기 제1 및 제2 CRISPR RNA 인식 서열은 각각 상기 5' 및 3' 표적 서열 둘 모두로부터 약 50 bp 내지 약 100 bp, 약 200 bp 내지 약 300 bp, 약 300 bp 내지 약 400 bp, 약 400 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 약 900 bp 내지 약 1 kb, 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 3 kb, 약 3 kb 내지 약 4 kb, 약 4 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 또는 약 50 kb 내지 약 100 kb만큼 떨어져서 위치된다. 일부 방법에서, 상기 제1 및 제2 CRISPR RNA 인식 서열은 각각 상기 5' 및 3' 표적 서열 둘 모두로부터 50 bp 초과, 100 bp 초과, 200 bp 초과, 300 bp 초과, 400 bp 초과, 500 bp 초과, 600 bp 초과, 700 bp 초과, 800 bp 초과, 900 bp 초과, 1 kb 초과, 2 kb 초과, 3 kb 초과, 4 kb 초과, 5 kb 초과, 6 kb 초과, 7 kb 초과, 8 kb 초과, 9 kb 초과, 10 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과, 또는 100 kb 초과만큼 떨어져서 위치된다.

[0016] 일부 방법에서, 상기 결실된 핵산은 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb이다. 일부 방법에서, 상기 결실된 핵산은 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb이다. 선택적으로, 상기 결실된 핵산은 적어도 550 kb, 적어도 600 kb, 적어도 650 kb, 적어도 700 kb, 적어도 750 kb, 적어도 800 kb, 적어도 850 kb, 적어도 900 kb, 적어도 950 kb, 적어도 1 Mb, 적어도 1.5 Mb, 또는 적어도 2 Mb이다.

[0017] 일부 방법에서, 상기 표적화 벡터는 선형 형태이다. 선택적으로, 상기 표적화 벡터는 단일-가닥 또는 이중-가닥이다. 일부 방법에서, 상기 세포는 1-세포기 배아가 아니고, 상기 표적화 벡터는 적어도 10 kb인 대형 표적화 벡터(LTVEC)이다. 일부 방법에서, 상기 세포는 1-세포기 배아가 아니고, 상기 표적화 벡터는 5' 및 3' 상동성 아암의 총합이 적어도 10 kb인 대형 표적화 벡터(LTVEC)이다. 선택적으로, 상기 LTVEC는 약 50 kb 내지 약 300 kb, 약 50 kb 내지 약 75 kb, 약 75 kb 내지 약 100 kb, 약 100 kb 내지 125 kb, 약 125 kb 내지 약 150 kb, 약 150 kb 내지 약 175 kb, 약 175 kb 내지 약 200 kb, 약 200 kb 내지 약 225 kb, 약 225 kb 내지 약 250 kb, 약 250 kb 내지 약 275 kb 또는 약 275 kb 내지 약 300 kb이다. 선택적으로, 상기 LTVEC의 5' 및 3' 상동성 아암의 총합은 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 약 90 kb 내지 약 100 kb, 약 100 kb 내지 약 110 kb, 약 110 kb 내지 약 120 kb, 약 120 kb 내지 약 130 kb, 약 130 kb 내지 약 140 kb, 약 140 kb 내지 약 150 kb, 약 150 kb 내지 약 160 kb, 약 160 kb 내지 약 170 kb, 약 170 kb 내지 약 180 kb, 약 180 kb 내지 약 190 kb, 또는 약 190 kb 내지 약 200 kb이다.

[0018] 일부 방법에서, 상기 세포는 진핵 세포이다. 선택적으로, 상기 진핵 세포는 포유류 세포, 인간 세포, 비인간 세포, 설치류 세포, 마우스 세포, 및 래트 세포이다. 선택적으로, 상기 진핵 세포는 다능성(pluripotent) 세포, 비다능성 세포, 비인간 다능성 세포, 인간 다능성 세포, 설치류 다능성 세포, 마우스 다능성 세포, 래트 다능성 세포, 마우스 배아 줄기(ES) 세포, 래트 ES 세포, 인간 ES 세포, 인간 성인 줄기 세포, 발생적으로 제한된 인간 선조 세포, 또는 인간 유도 다능성 줄기(iPS) 세포이다. 선택적으로, 상기 진핵 세포는 1-세포기 배아이다. 선택적으로, 상기 진핵 세포는 1-세포기 배아이고, 상기 표적화 벡터는 약 50개 뉴클레오티드 내지 약 5 kb 길이이다. 선택적으로, 상기 진핵 세포는 1-세포기 배아이고, 상기 표적화 벡터는 단일-가닥 DNA이고, 약 60 내지 약 200개 뉴클레오티드 길이이다.

[0019] 일부 방법에서, 상기 제1 Cas 단백질은 Cas9이다. 일부 방법에서, 상기 제1 Cas 단백질은 이중-가닥 DNA의 양쪽 가닥에 대해 뉴클레아제 활성을 갖는다.

[0020] 일부 방법에서, 상기 제1 Cas 단백질은 니카제(nickase)이다. 일부 방법은 상기 게놈을 (f) 니카제인 제2 Cas 단백질; (g) 제3 CRISPR RNA 인식 서열에 혼성화되는 제3 CRISPR RNA; 및 (h) 제4 CRISPR RNA 인식 서열에 혼성화되는 제4 CRISPR RNA와 접촉시키는 단계를 추가로 포함하며, 상기 제1 Cas 단백질은 상기 제1 CRISPR RNA

인식 서열 내의 그리고 상기 제2 CRISPR RNA 인식 서열 내의 게놈 DNA의 제1 가닥을 절단하고, 상기 제2 Cas 단백질을 상기 제3 CRISPR RNA 인식 서열 내의 그리고 상기 제4 CRISPR RNA 인식 서열 내의 게놈 DNA의 제2 가닥을 절단한다.

[0021] 일부 방법에서, 상기 제1 CRISPR RNA와 상기 tracrRNA는 제1 가이드 RNA(gRNA)로서 함께 융합되고/되거나, 상기 제2 CRISPR RNA와 상기 tracrRNA는 제2 gRNA로서 함께 융합된다. 일부 방법에서, 상기 제1 CRISPR RNA와 상기 tracrRNA는 별개의 RNA 분자이고/이거나, 상기 제2 CRISPR RNA와 상기 tracrRNA는 별개의 RNA 분자이다.

[0022] 일부 방법에서, 상기 접촉하는 단계는 상기 제1 Cas 단백질, 상기 제1 및 제2 CRISPR RNA, 및 상기 tracrRNA를 상기 세포 내로 도입하는 단계를 포함한다. 일부 방법에서, (a) 상기 제1 Cas 단백질은 단백질, 상기 제1 Cas 단백질을 인코딩하는 메신저 RNA(mRNA), 또는 상기 제1 Cas 단백질을 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (b) 상기 제1 CRISPR RNA는 RNA의 형태로 또는 상기 제1 CRISPR RNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (c) 상기 제2 CRISPR RNA는 RNA의 형태로 또는 상기 제2 CRISPR RNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (d) 상기 tracrRNA는 RNA의 형태로 또는 상기 tracrRNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입된다. 일부 실시 형태에서, 상기 제1 Cas 단백질, 상기 제1 CRISPR RNA, 및 상기 tracrRNA는 제1 단백질-RNA 복합체로서 상기 세포 내로 도입되고/되거나, 상기 제1 Cas 단백질, 상기 제2 CRISPR RNA, 및 상기 tracrRNA는 제2 단백질-RNA 복합체로서 상기 세포 내로 도입된다. 일부 방법에서, (a) 상기 제1 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나; (b) 상기 제1 CRISPR RNA를 인코딩하는 DNA는 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되고/되거나; (c) 상기 제2 CRISPR RNA를 인코딩하는 DNA는 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되고/되거나; (d) 상기 tracrRNA를 인코딩하는 DNA는 제4 발현 작제물 내의 제4 프로모터에 작동가능하게 연결되며, 상기 제1, 제2, 제3, 및 제4 프로모터는 상기 세포 내에서 활성이다. 선택적으로, 상기 제1, 제2, 제3, 및/또는 제4 발현 작제물은 단일 핵산 분자의 성분들이다. 일부 방법에서, (a) 상기 제1 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나; (b) 상기 제1 CRISPR RNA를 인코딩하는 DNA와 상기 tracrRNA를 인코딩하는 DNA는 제1 가이드 RNA(gRNA)를 인코딩하는 DNA에서 함께 융합되고, 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되고/되거나; (c) 상기 제2 CRISPR RNA를 인코딩하는 DNA와 상기 tracrRNA를 인코딩하는 DNA는 제2 gRNA를 인코딩하는 DNA에서 함께 융합되고, 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되며, 상기 제1, 제2, 및 제3 프로모터는 상기 세포 내에서 활성이다. 선택적으로, 상기 제1, 제2, 및/또는 제3 발현 작제물은 단일 핵산 분자의 성분들이다.

[0023] 일부 방법에서, 상기 세포는 비상동 말단 결합(non-homologous end joining)(NHEJ)을 감소시키고/시키거나 유전자 변환 또는 상동성-유도 수복(homology-directed repair)(HDR)을 증가시키도록 변형되었다. 선택적으로, 상기 세포는 DNA-PK 발현 또는 활성을 감소시키고/시키거나 PARP1 발현 또는 활성을 감소시키도록 변형되었다. 선택적으로, 상기 세포는 리가제 IV 발현 또는 활성을 감소시키도록 변형되었다. 선택적으로, 상기 발현 또는 활성의 감소는 유도성, 가역적, 시간 특이적, 및/또는 공간 특이적이다.

[0024] 일부 방법에서, (1) 상기 세포는 1-세포기 배아가 아니고, 상기 표적화 벡터는 상기 5' 및 3' 상동성 아암의 총합이 적어도 10 kb인 대형 표적화 벡터이고; (2) 상기 제1 및 제2 CRISPR RNA 인식 서열은 각각 상기 5' 및 3' 표적 서열 둘 모두로부터 200 bp 초과, 300 bp 초과, 400 bp 초과, 500 bp 초과, 600 bp 초과, 700 bp 초과, 800 bp 초과, 900 bp 초과, 1 kb 초과, 2 kb 초과, 3 kb 초과, 4 kb 초과, 5 kb 초과, 6 kb 초과, 7 kb 초과, 8 kb 초과, 9 kb 초과, 10 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과, 또는 100 kb 초과만큼 떨어져서 위치되고; (3) 상기 제1 Cas 단백질은 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 상기 제1 및 제2 CRISPR RNA 인식 서열을 절단하여, 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 적어도 2개의 이중-가닥 브레이크를 생성하고; (4) 상기 이중대립유전자 변형은 상기 제1 상동 염색체에서의 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열 사이의 결실, 및 상기 제1 상동 염색체에서의 상기 5' 표적 서열과 상기 3' 표적 서열 사이의 상기 핵산 삽입물의 삽입을 포함하며, 상기 핵산 삽입물 서열은 상기 결실된 서열과 상동성 또는 오르토로그성이다.

[0025] 본 발명은 또한 F0 세대 비인간 동물을 생성하는 방법을 제공하며, 본 방법은 (a) 상기 방법들 중 어느 하나에 의해 생성된 비인간 ES 세포를 비인간 숙주 배아 내로 도입하는 단계; 및 (b) 대리모(surrogate mother)에서 상기 비인간 숙주 배아를 잉태시키는 단계를 포함하며, 상기 대리모는 상기 이중대립유전자 변형을 포함하는 상기 F0 세대 비인간 동물을 생성한다. 일부 방법은 (a) 상기 비인간 ES 세포 내의 게놈을 (i) 제1 Cas 단백질; (ii) 게놈 표적 유전자좌 내의 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA; (iii) 상기 게놈 표적 유전자좌 내의 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA; (iv) tracrRNA; 및 (v) 5' 상동성 아

암 및 3' 상동성 아암에 의해 플랭킹된 핵산 삽입물을 포함하는 표적화 벡터와 접촉시키는 단계 - 여기서, 상기 게놈은 상기 게놈 표적 유전자좌를 포함하는 한 쌍의 제1 및 제2 상동 염색체를 포함하고; 상기 제1 Cas 단백질은 상기 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여, 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 적어도 하나의 이중-가닥 브레이크를 생성함 -; (b) 상기 이중대립유전자 변형을 포함하는 비인간 ES 세포를 확인하는 단계; (c) 상기 이중대립유전자 변형을 포함하는 상기 비인간 ES 세포를 비인간 숙주 배아 내로 도입하는 단계; 및 (d) 대리모에서 상기 비인간 숙주 배아를 잉태시키는 단계를 포함하며, 상기 대리모는 상기 이중대립유전자 변형을 포함하는 상기 F0 세대 비인간 동물을 생성한다.

[0026] 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1 및 제2 상동 염색체 각각에서 상기 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여, 상기 제1 및 제2 상동 염색체 각각에서 적어도 하나의 이중-가닥 브레이크를 생성한다. 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 상기 제1 및 제2 CRISPR RNA 인식 서열을 절단하여, 상기 제1 및 제2 상동 염색체 중 적어도 하나에서 적어도 2개의 이중-가닥 브레이크를 생성한다.

[0027] 일부 방법에서, 상기 비인간 동물은 마우스이고, 상기 비인간 ES 세포는 마우스 ES 세포이고, 상기 비인간 숙주 배아는 마우스 숙주 배아이다. 일부 방법에서, 상기 비인간 동물은 래트이고, 상기 비인간 ES 세포는 래트 ES 세포이고, 상기 비인간 숙주 배아는 래트 숙주 배아이다.

[0028] 일부 방법에서, 상기 이중대립유전자 변형은 상기 게놈 표적 유전자좌에서 동형접합성 또는 복합 이형접합성을 가져온다. 선택적으로, 상기 복합 이형접합성은 반접합성이다.

[0029] 일부 방법에서, 상기 제1 CRISPR RNA와 상기 tracrRNA는 제1 가이드 RNA(gRNA)로서 함께 융합되고/되거나, 상기 제2 CRISPR RNA와 상기 tracrRNA는 제2 gRNA로서 함께 융합된다. 일부 방법에서, 상기 제1 CRISPR RNA와 상기 tracrRNA는 별개의 RNA 분자이고/이거나, 상기 제2 CRISPR RNA와 상기 tracrRNA는 별개의 RNA 분자이다.

[0030] 일부 방법에서, 상기 접촉하는 단계는 상기 제1 Cas 단백질, 상기 제1 및 제2 CRISPR RNA, 및 상기 tracrRNA를 상기 세포 내로 도입하는 단계를 포함한다. 일부 방법에서, (a) 상기 제1 Cas 단백질은 단백질, 상기 제1 Cas 단백질을 인코딩하는 메신저 RNA(mRNA), 또는 상기 제1 Cas 단백질을 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (b) 상기 제1 CRISPR RNA는 RNA의 형태로 또는 상기 제1 CRISPR RNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (c) 상기 제2 CRISPR RNA는 RNA의 형태로 또는 상기 제2 CRISPR RNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (d) 상기 tracrRNA는 RNA의 형태로 또는 상기 tracrRNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입된다. 일부 실시 형태에서, 상기 제1 Cas 단백질, 상기 제1 CRISPR RNA, 및 상기 tracrRNA는 제1 단백질-RNA 복합체로서 상기 세포 내로 도입되고/되거나, 상기 제1 Cas 단백질, 상기 제2 CRISPR RNA, 및 상기 tracrRNA는 제2 단백질-RNA 복합체로서 상기 세포 내로 도입된다. 일부 방법에서, (a) 상기 제1 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나; (b) 상기 제1 CRISPR RNA를 인코딩하는 DNA는 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되고/되거나; (c) 상기 제2 CRISPR RNA를 인코딩하는 DNA는 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되고/되거나; (d) 상기 tracrRNA를 인코딩하는 DNA는 제4 발현 작제물 내의 제4 프로모터에 작동가능하게 연결되며, 상기 제1, 제2, 제3, 및 제4 프로모터는 상기 세포 내에서 활성이다. 선택적으로, 상기 제1, 제2, 제3, 및/또는 제4 발현 작제물은 단일 핵산 분자의 성분들이다. 일부 방법에서, (a) 상기 제1 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나; (b) 상기 제1 CRISPR RNA를 인코딩하는 DNA와 상기 tracrRNA를 인코딩하는 DNA는 제1 가이드 RNA(gRNA)를 인코딩하는 DNA에서 함께 융합되고, 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되고/되거나; (c) 상기 제2 CRISPR RNA를 인코딩하는 DNA와 상기 tracrRNA를 인코딩하는 DNA는 제2 gRNA를 인코딩하는 DNA에서 함께 융합되고, 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되며, 상기 제1, 제2, 및 제3 프로모터는 상기 세포 내에서 활성이다. 선택적으로, 상기 제1, 제2, 및/또는 제3 발현 작제물은 단일 핵산 분자의 성분들이다.

[0031] 일부 방법에서, 상기 제1 Cas 단백질은 Cas9이다. 일부 방법에서, 상기 제1 Cas 단백질은 이중-가닥 DNA의 양쪽 가닥에 대해 뉴클레아제 활성을 갖는다.

[0032] 일부 방법에서, 상기 제1 Cas 단백질은 니카제이다. 일부 방법은 상기 게놈을 (f) 니카제인 제2 Cas 단백질; (g) 제3 CRISPR RNA 인식 서열에 혼성화되는 제3 CRISPR RNA; 및 (h) 제4 CRISPR RNA 인식 서열에 혼성화되는 제4 CRISPR RNA와 접촉시키는 단계를 추가로 포함하며, 상기 제1 Cas 단백질은 상기 제1 CRISPR RNA 인식 서열 내의 그리고 상기 제2 CRISPR RNA 인식 서열 내의 게놈 DNA의 제1 가닥을 절단하고, 상기 제2 Cas 단백질은 상기 제3 CRISPR RNA 인식 서열 내의 그리고 상기 제4 CRISPR RNA 인식 서열 내의 게놈 DNA의 제2 가닥을 절단한

다.

- [0033] 일부 방법에서, 상기 세포는 비상동 말단 결합(NHEJ)을 감소시키고/시키거나 유전자 변환 또는 상동성-유도 수복(HDR)을 증가시키도록 변형되었다. 선택적으로, 상기 세포는 DNA-PK 발현 또는 활성을 감소시키고/시키거나 PARP1 발현 또는 활성을 감소시키도록 변형되었다. 선택적으로, 상기 세포는 리가제 IV 발현 또는 활성을 감소시키도록 변형되었다. 선택적으로, 상기 발현 또는 활성의 감소는 유도성, 가역적, 시간 특이적, 및/또는 공간 특이적이다.
- [0034] 본 발명은 또한 F0 세대 비인간 동물을 생성하는 방법을 제공하며, 본 방법은 상기 방법들 중 어느 하나에 의해 생성된 유전자 변형된 1-세포기 배아를 대리모 내로 이식하는 단계를 포함하며, 상기 대리모는 상기 이중대립유전자 변형을 포함하는 상기 F0 세대 비인간 동물을 생성한다.
- [0035] 본 발명은 또한 제1 대립유전자에 대해 이형접합성인 세포 내의 게놈을 변형시키는 방법을 제공하며, 본 방법은 상기 게놈을 (a) 제1 Cas 단백질; (b) tracrRNA; (c) 제2 대립유전자 내의 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA - 여기서, 상기 제1 대립유전자는 제1 상동 염색체 상에 있고, 상기 제2 대립유전자는 제2 상동 염색체 상의 상응하는 유전자좌에 있음 -; 및 (d) 상기 제2 대립유전자 내의 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA와 접촉시키는 단계를 포함하며, 상기 제1 Cas 단백질은 상기 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여 적어도 하나의 이중-가닥 브레이크 및 말단 서열들을 생성하고, 상기 말단 서열들은 재조합을 거치며, 상기 재조합은 상기 제1 대립유전자와 상기 제2 대립유전자 사이에서 이루어져서, 상기 제1 대립유전자에 대해 동형접합성인 변형된 게놈을 형성한다. 일부 방법은 상기 제1 대립유전자에 대해 동형접합성인 세포를 확인하는 단계를 추가로 포함한다.
- [0036] 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1 CRISPR RNA 인식 서열 및 상기 제2 CRISPR RNA 인식 서열을 절단한다. 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1 CRISPR RNA 인식 서열 및 상기 제2 CRISPR RNA 인식 서열을 절단하여, 적어도 2개의 이중-가닥 브레이크 및 말단 서열들을 생성하고, 상기 말단 서열은 재조합을 거친다. 일부 방법에서, 상기 제1 및 제2 CRISPR RNA 인식 서열은, 상기 제1 대립유전자 내에는 아니고 상기 제2 대립유전자 내에 위치된다. 일부 방법에서, 상기 Cas 단백질과 상기 제1 CRISPR RNA는 자연적으로 함께 발생되지 않는다.
- [0037] 일부 방법에서, 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열은 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb만큼 분리된다. 일부 방법에서, 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열은 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb만큼 분리된다. 일부 방법에서, 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열은 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 또는 약 900 bp 내지 약 1 kb만큼 분리된다. 일부 방법에서, 상기 제1 CRISPR RNA 인식 서열과 상기 제2 CRISPR RNA 인식 서열은 25 bp 미만, 50 bp 미만, 100 bp 미만, 150 bp 미만, 200 bp 미만, 250 bp 미만, 300 bp 미만, 350 bp 미만, 400 bp 미만, 450 bp 미만, 500 bp 미만, 600 bp 미만, 700 bp 미만, 800 bp 미만, 900 bp 미만, 1 kb 미만, 2 kb 미만, 3 kb 미만, 4 kb 미만, 5 kb 미만, 또는 10 kb 미만만큼 분리된다.
- [0038] 일부 방법에서, 상기 제1 대립유전자와 상기 제2 대립유전자 사이의 서열 차이가 약 100 bp 내지 약 200 bp, 약 200 bp 내지 약 400 bp, 약 400 bp 내지 약 600 bp, 약 600 bp 내지 약 800 bp, 약 800 bp 내지 약 1 kb, 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 3 kb, 약 4 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약

400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb에 이른다. 일부 방법에서, 상기 제1 대립유전자와 상기 제2 대립유전자 사이의 서열 차이가 적어도 100 bp, 적어도 200 bp, 적어도 300 bp, 적어도 400 bp, 적어도 500 bp, 적어도 600 bp, 적어도 700 bp, 적어도 800 bp, 적어도 800 bp, 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 6 kb, 적어도 7 kb, 적어도 8 kb, 적어도 9 kb, 적어도 10 kb, 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb에 이른다.

[0039] 일부 방법에서, 상기 제1 대립유전자는 표적화된 변형을 포함하고, 상기 제2 대립유전자는 야생형 대립유전자이다. 일부 방법에서, 상기 제1 대립유전자는 야생형 대립유전자이고, 상기 제2 대립유전자는 질병-유발 돌연변이를 포함한다.

[0040] 일부 방법에서, 상기 재조합은 유전자 변환을 포함한다. 일부 방법에서, 상기 재조합은 이형접합성의 상실(loss of heterozygosity)(LOH)을 포함한다.

[0041] 일부 방법에서, 상기 세포는 진핵 세포이다. 선택적으로, 상기 진핵 세포는 포유류 세포, 인간 세포, 비인간 세포, 설치류 세포, 마우스 세포, 또는 래트 세포이다. 선택적으로, 상기 진핵 세포는 다능성 세포, 비다능성 세포, 비인간 다능성 세포, 인간 다능성 세포, 설치류 다능성 세포, 마우스 다능성 세포, 래트 다능성 세포, 마우스 배아 줄기(ES) 세포, 래트 ES 세포, 인간 ES 세포, 인간 성인 줄기 세포, 발생적으로 제한된 인간 선조 세포, 또는 인간 유도 다능성 줄기(iPS) 세포이다.

[0042] 일부 방법에서, 상기 제1 Cas 단백질은 Cas9이다. 일부 방법에서, 상기 제1 Cas 단백질은 이중-가닥 DNA의 양쪽 가닥에 대해 뉴클레아제 활성을 갖는다.

[0043] 일부 방법에서, 상기 제1 Cas 단백질은 니카제이다. 일부 방법은 상기 게놈(f) 니카제인 제2 Cas 단백질; (g) 제3 CRISPR RNA 인식 서열에 혼성화되는 제3 CRISPR RNA; 및 (h) 제4 CRISPR RNA 인식 서열에 혼성화되는 제4 CRISPR RNA와 접촉시키는 단계를 추가로 포함하며, 상기 제1 Cas 단백질은 상기 제1 CRISPR RNA 인식 서열 내의 그리고 상기 제2 CRISPR RNA 인식 서열 내의 게놈 DNA의 제1 가닥을 절단하고, 상기 제2 Cas 단백질은 상기 제3 CRISPR RNA 인식 서열 내의 그리고 상기 제4 CRISPR RNA 인식 서열 내의 게놈 DNA의 제2 가닥을 절단한다.

[0044] 일부 방법에서, 상기 제1 CRISPR RNA와 상기 tracrRNA는 제1 가이드 RNA(gRNA)로서 함께 융합되고/되거나, 상기 제2 CRISPR RNA와 상기 tracrRNA는 제2 gRNA로서 함께 융합된다. 일부 방법에서, 상기 제1 CRISPR RNA와 상기 tracrRNA는 별개의 RNA 분자이고/이거나, 상기 제2 CRISPR RNA와 상기 tracrRNA는 별개의 RNA 분자이다.

[0045] 일부 방법에서, 상기 접촉하는 단계는 상기 제1 Cas 단백질, 상기 제1 및 제2 CRISPR RNA, 및 상기 tracrRNA를 상기 세포 내로 도입하는 단계를 포함한다. 일부 방법에서, (a) 상기 제1 Cas 단백질은 단백질, 상기 제1 Cas 단백질을 인코딩하는 메신저 RNA(mRNA), 또는 상기 제1 Cas 단백질을 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (b) 상기 제1 CRISPR RNA는 RNA의 형태로 또는 상기 제1 CRISPR RNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (c) 상기 제2 CRISPR RNA는 RNA의 형태로 또는 상기 제2 CRISPR RNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입되고/되거나; (d) 상기 tracrRNA는 RNA의 형태로 또는 상기 tracrRNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입된다. 일부 실시 형태에서, 상기 제1 Cas 단백질, 상기 제1 CRISPR RNA, 및 상기 tracrRNA는 제1 단백질-RNA 복합체로서 상기 세포 내로 도입되고/되거나, 상기 제1 Cas 단백질, 상기 제2 CRISPR RNA, 및 상기 tracrRNA는 제2 단백질-RNA 복합체로서 상기 세포 내로 도입된다. 일부 방법에서, (a) 상기 제1 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나; (b) 상기 제1 CRISPR RNA를 인코딩하는 DNA는 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되고/되거나; (c) 상기 제2 CRISPR RNA를 인코딩하는 DNA는 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되고/되거나; (d) 상기 tracrRNA를 인코딩하는 DNA는 제4 발현 작제물 내의 제4 프로모터에 작동가능하게 연결되며, 상기 제1, 제2, 제3, 및 제4 프로모터는 상기 세포 내에서 활성이다. 선택적으로, 상기 제1, 제2, 제3, 및/또는 제4 발현 작제물은 단일 핵산 분자의 성분들이다. 일부 방법에서, (a) 상기 제1 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나; (b) 상기 제1 CRISPR RNA를 인코딩하는 DNA와 상기 tracrRNA를 인코딩하는 DNA는 제1 가이드 RNA(gRNA)를 인코딩하는 DNA에서 함께 융합되고, 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되고/되거나; (c) 상기 제2

CRISPR RNA를 인코딩하는 DNA와 상기 tracrRNA를 인코딩하는 DNA는 제2 gRNA를 인코딩하는 DNA에서 함께 융합되고, 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되며, 상기 제1, 제2, 및 제3 프로모터는 상기 세포 내에서 활성이다. 선택적으로, 상기 제1, 제2, 및/또는 제3 발현 작제물은 단일 핵산 분자의 성분들이다.

[0046] 일부 방법에서, 상기 세포는 비상동 말단 결합(NHEJ)을 감소시키고/시키거나 유전자 변환 또는 상동성-유도 수복(HDR)을 증가시키도록 변형되었다. 선택적으로, 상기 세포는 DNA-PK 발현 또는 활성을 감소시키고/시키거나 PARP1 발현 또는 활성을 감소시키도록 변형되었다. 선택적으로, 상기 세포는 리가제 IV 발현 또는 활성을 감소시키도록 변형되었다. 선택적으로, 상기 발현 또는 활성의 감소는 유도성, 가역적, 시간 특이적, 및/또는 공간 특이적이다.

[0047] 본 발명은 또한 제1 대립유전자에 대해 이형접합성인 세포 내의 게놈을 변형시키는 방법을 제공하며, 본 방법은 상기 게놈을 (a) 제1 Cas 단백질; (b) tracrRNA; 및 (c) 제1 비-대립유전자-특이적 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA와 접촉시키는 단계를 포함하며, 상기 제1 대립유전자는 제1 상동 염색체 상에 있고, 상기 CRISPR RNA 인식 서열은 제2 상동 염색체 상의 상기 제1 대립유전자에 상응하는 유전자좌에 대해 동원체 측에 있고(centromeric); 상기 제1 Cas 단백질은 상기 제1 CRISPR RNA 인식 서열을 절단하여 이중-가닥 브레이크를 생성하고 상기 세포는 변형되어 상기 제1 대립유전자에 대해 동형접합성이 된다. 일부 방법은 상기 제1 대립유전자에 대해 동형접합성인 세포를 확인하는 단계를 추가로 포함한다. 일부 방법에서, 상기 Cas 단백질과 상기 제1 CRISPR RNA는 자연적으로 함께 발생되지 않는다.

[0048] 그러한 방법은 상기 게놈을 제2 상동 염색체 상의 상기 제1 대립유전자에 상응하는 유전자좌에 대해 동원체 측에 있는 제2 비-대립유전자-특이적 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA와 접촉시키는 단계를 추가로 포함할 수 있으며, 상기 제1 Cas 단백질은 상기 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여 적어도 하나의 이중-가닥 브레이크를 생성한다. 일부 방법에서, 상기 제1 Cas 단백질은 상기 제1 CRISPR RNA 인식 서열 및 상기 제2 CRISPR RNA 인식 서열을 절단한다.

[0049] 일부 방법에서, 이형접합성의 상실이 상기 이중-가닥 브레이크의 말단소체 측에서(telomeric) 일어난다.

[0050] 일부 방법에서, 상기 제1 및 제2 CRISPR RNA 인식 서열은, 상기 제1 상동 염색체 상은 아니고 상기 제2 상동 염색체 상에 위치된다. 일부 방법에서, 상기 제1 CRISPR RNA 인식 부위는 상기 동원체로부터 약 100 bp 내지 약 1 kb, 약 1 kb 내지 약 10 kb, 약 10 kb 내지 약 100 kb, 약 100 kb 내지 약 1 Mb, 약 1 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb만큼 떨어져 있다. 일부 방법에서, 상기 제1 대립유전자는 상기 제1 CRISPR RNA 인식 부위로부터 약 100 bp 내지 약 1 kb, 약 1 kb 내지 약 10 kb, 약 10 kb 내지 약 100 kb, 약 100 kb 내지 약 1 Mb, 약 1 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb만큼 떨어져 있다. 일부 방법에서, 이형접합성의 상실에 의해 대체되는 상기 제2 상동 염색체의 영역이 약 100 bp 내지 약 1 kb, 약 1 kb 내지 약 10 kb, 약 10 kb 내지 약 100 kb, 약 100 kb 내지 약 1 Mb, 약 1 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb이다.

[0051] 일부 방법에서, 상기 세포는 진핵 세포이다. 선택적으로, 상기 진핵 세포는 포유류 세포, 인간 세포, 비인간 세포, 설치류 세포, 마우스 세포, 래트 세포, 다능성 세포, 비다능성 세포, 비인간 다능성 세포, 인간 다능성 세포, 설치류 다능성 세포, 마우스 다능성 세포, 래트 다능성 세포, 마우스 배아 줄기(ES) 세포, 래트 ES 세포, 인간 ES 세포, 인간 성인 줄기 세포, 발생적으로 제한된 인간 선조 세포, 인간 유도 다능성 줄기(iPS) 세포, 또는 1-세포기 배아이다.

[0052] 일부 방법에서, 상기 제1 Cas 단백질은 Cas9이다. 일부 방법에서, 상기 제1 Cas 단백질은 이중-가닥 DNA의 양쪽 가닥에 대해 뉴클레아제 활성을 갖는다. 일부 방법에서, 상기 제1 Cas 단백질은 니카제이다. 선택적으로, 상기 제1 Cas 단백질은 니카제이고, 상기 방법은 상기 게놈을 니카제인 제2 Cas 단백질; 제3 CRISPR RNA 인식 서열에 혼성화되는 제3 CRISPR RNA; 및 제4 CRISPR RNA 인식 서열에 혼성화되는 제4 CRISPR RNA와 접촉시키는 단계를 추가로 포함하며, 상기 제1 Cas 단백질은 상기 제1 CRISPR RNA 인식 서열 내의 그리고 상기 제2 CRISPR RNA 인식 서열 내의 게놈 DNA의 제1 가닥을 절단하고, 상기 제2 Cas 단백질은 상기 제3 CRISPR RNA 인식 서열 내의 그리고 상기 제4 CRISPR RNA 인식 서열 내의 게놈 DNA의 제2 가닥을 절단한다.

- [0053] 일부 방법에서, 상기 제1 CRISPR RNA와 상기 tracrRNA는 제1 가이드 RNA(gRNA)로서 함께 융합되고/되거나, 상기 제2 CRISPR RNA와 상기 tracrRNA는 제2 gRNA로서 함께 융합된다. 일부 방법에서, 상기 제1 CRISPR RNA와 상기 tracrRNA는 별개의 RNA 분자이고/이거나, 상기 제2 CRISPR RNA와 상기 tracrRNA는 별개의 RNA 분자이다.
- [0054] 일부 방법에서, 상기 접촉하는 단계는 상기 제1 Cas 단백질, 상기 제1 및 제2 CRISPR RNA, 및 상기 tracrRNA를 상기 세포 내로 도입하는 단계를 포함한다. 일부 방법에서, 상기 제1 Cas 단백질은 단백질, 상기 제1 Cas 단백질을 인코딩하는 메신저 RNA(mRNA), 또는 상기 제1 Cas 단백질을 인코딩하는 DNA의 형태로 상기 세포 내로 도입된다. 선택적으로, 상기 제1 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되며, 상기 제1 프로모터는 상기 세포에서 활성이다. 일부 방법에서, 상기 제1 CRISPR RNA는 RNA의 형태로 또는 상기 제1 CRISPR RNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입된다. 선택적으로, 상기 제1 CRISPR RNA를 인코딩하는 DNA는 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되며, 상기 제2 프로모터는 상기 세포에서 활성이다. 일부 방법에서, 상기 제2 CRISPR RNA는 RNA의 형태로 또는 상기 제2 CRISPR RNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입된다. 선택적으로, 상기 제2 CRISPR RNA를 인코딩하는 DNA는 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되며, 상기 제3 프로모터는 상기 세포에서 활성이다. 일부 방법에서, 상기 tracrRNA는 RNA의 형태로 또는 상기 tracrRNA를 인코딩하는 DNA의 형태로 상기 세포 내로 도입된다. 선택적으로, 상기 tracrRNA를 인코딩하는 DNA는 제4 발현 작제물 내의 제4 프로모터에 작동가능하게 연결되며, 상기 제4 프로모터는 상기 세포에서 활성이다. 선택적으로, 상기 제1, 제2, 제3, 및/또는 제4 발현 작제물은 단일 핵산 분자의 성분들이다.
- [0055] 선택적으로, 상기 제1 Cas 단백질을 인코딩하는 DNA는 제1 발현 작제물 내의 제1 프로모터에 작동가능하게 연결되고/되거나; 상기 제1 CRISPR RNA를 인코딩하는 DNA와 상기 tracrRNA를 인코딩하는 DNA는 제1 가이드 RNA(gRNA)를 인코딩하는 DNA에서 함께 융합되고, 제2 발현 작제물 내의 제2 프로모터에 작동가능하게 연결되고/되거나; 상기 제2 CRISPR RNA를 인코딩하는 DNA와 상기 tracrRNA를 인코딩하는 DNA는 제2 gRNA를 인코딩하는 DNA에서 함께 융합되고, 제3 발현 작제물 내의 제3 프로모터에 작동가능하게 연결되며, 상기 제1, 제2, 및 제3 프로모터는 상기 세포 내에서 활성이다. 선택적으로, 상기 제1, 제2, 및/또는 제3 발현 작제물은 단일 핵산 분자의 성분들이다.
- [0056] 선택적으로, 상기 제1 Cas 단백질, 상기 제1 CRISPR RNA, 및 상기 tracrRNA는 제1 단백질-RNA 복합체로서 상기 세포 내로 도입되고/되거나, 상기 제1 Cas 단백질, 상기 제2 CRISPR RNA, 및 상기 tracrRNA는 제2 단백질-RNA 복합체로서 상기 세포 내로 도입된다.
- [0057] 일부 방법에서, 상기 세포는 비상동 말단 결합(NHEJ)을 감소시키고/시키거나 유전자 변환 또는 상동성-유도 수복(HDR)을 증가시키도록 변형되었다. 선택적으로, 상기 세포는 하기 중 하나 이상의 발현 또는 활성을 감소시키도록 변형되었다: DNA-PK, PARP1, 및 리가제 IV. 선택적으로, 상기 발현 또는 활성의 감소는 유도성, 가역적, 시간 특이적, 및/또는 공간 특이적이다.
- [0058] 일부 방법에서, 상기 제1 대립유전자는 돌연변이를 포함한다. 선택적으로, 상기 돌연변이는 표적화된 변형이다. 일부 실시 형태에서, 상기 제1 대립유전자는 야생형 대립유전자이고, 상기 제2 상동 염색체 상의 상응하는 유전자좌가 돌연변이를 포함한다.
- [0059] 본 발명은 또한 1-세포기 배아가 아닌 이배체 세포 내의 표적 게놈 유전자좌에서의 핵산 삽입물의 표적화된 삽입을 확인하는 방법을 제공하며, 본 방법은 (a) 상기 세포로부터 DNA를 얻는 단계 - 여기서, 상기 세포는 제1 표적 서열에 혼성화되는 제1 상동성 아암 및 제2 표적 서열에 혼성화되는 제2 상동성 아암에 의해 플랭킹된 핵산 삽입물을 포함하는 대형 표적화 벡터(LTVEC)와 접촉되었으며, 상기 핵산 삽입물은 상기 제1 상동성 아암에 인접한 선택 카세트를 포함함 -; (b) 상기 세포의 DNA를 상기 제1 표적 서열 내에서 결합하는 프로브, 상기 핵산 삽입물 내에서 결합하는 프로브, 및 기지의 카피수를 갖는 참조 유전자 내에서 결합하는 프로브에 노출시키는 단계 - 여기서, 각각의 프로브는 결합 시에 검출가능한 신호를 발생시킴 -; (c) 상기 프로브들 각각의 결합으로부터 신호를 검출하는 단계; 및 (d) 상기 참조 유전자 프로브로부터의 신호를 상기 제1 표적 서열 프로브로부터의 신호와 비교하여 상기 제1 표적 서열에 대한 카피수를 결정하고, 상기 참조 유전자 프로브로부터의 신호를 상기 핵산 삽입물 프로브로부터의 신호와 비교하여 상기 핵산 삽입물에 대한 카피수를 결정하는 단계를 포함하며, 1개 또는 2개의 핵산 삽입물 카피수 및 2개의 제1 표적 서열 카피수는 상기 표적 게놈 유전자좌에서의 상기 핵산 삽입물의 표적화된 삽입을 나타내고, 1개 이상의 핵산 삽입물 카피수 및 3개 이상의 제1 표적 서열 카피수는 상기 표적 게놈 유전자좌 이외의 게놈 유전자좌에서의 상기 핵산 삽입물의 무작위 삽입을 나타낸다.
- [0060] 일부 방법에서, 상기 제1 표적 서열 프로브의 결합으로부터의 신호는 상기 제1 표적 서열에 대한 역치 주기

(threshold cycle)(Ct) 값을 결정하는 데 사용되고, 상기 참조 유전자 프로브의 결합으로부터의 신호는 상기 참조 유전자에 대한 역치 주기(Ct) 값을 결정하는 데 사용되고, 상기 제1 표적 서열의 카피수는 상기 제1 표적 서열 Ct 값과 상기 참조 유전자 Ct 값을 비교함으로써 결정된다. 일부 방법에서, 상기 핵산 삽입물 프로브의 결합으로부터의 신호는 상기 핵산 삽입물에 대한 역치 주기(Ct) 값을 결정하는 데 사용되고, 상기 핵산 삽입물의 카피수는 상기 제1 표적 서열 Ct 값과 상기 참조 유전자 Ct 값을 비교함으로써 결정된다.

[0061] 일부 방법에서, 상기 선택 카세트는 약물 저항성 유전자를 포함한다.

[0062] 일부 방법에서, 상기 핵산 삽입물은 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 150 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb이다. 일부 방법에서, 상기 제1 표적 서열과 상기 선택 카세트에서 상기 프로브들이 결합하는 서열들 사이의 거리가 100개 뉴클레오타이드 이하, 200개 뉴클레오타이드 이하, 300개 뉴클레오타이드 이하, 400개 뉴클레오타이드 이하, 500개 뉴클레오타이드 이하, 600개 뉴클레오타이드 이하, 700개 뉴클레오타이드 이하, 800개 뉴클레오타이드 이하, 900개 뉴클레오타이드 이하, 1 kb 이하, 1.5 kb 이하, 2 kb 이하, 2.5 kb 이하, 3 kb 이하, 3.5 kb 이하, 4 kb 이하, 4.5 kb 이하, 또는 5 kb 이하이다.

[0063] 일부 방법은 상기 제2 표적 서열의 카피수를 결정하는 단계를 추가로 포함한다. 선택적으로, 단계 (b)는 상기 세포의 DNA를 상기 제2 표적 서열과 결합하는 프로브에 노출시키는 단계를 추가로 포함하고, 단계 (c)는 제2 표적 서열 프로브의 결합으로부터 신호를 검출하는 단계를 추가로 포함하고, 단계 (d)는 상기 참조 유전자 프로브로부터의 신호를 상기 제2 표적 서열 프로브로부터의 신호와 비교하여 상기 제2 표적 서열에 대한 카피수를 결정하는 단계를 추가로 포함한다.

[0064] 일부 방법은 상기 핵산 삽입물 내의 하나 이상의 추가 서열의 카피수를 결정하는 단계를 추가로 포함한다. 선택적으로, 단계 (b)는 상기 세포의 DNA를 상기 핵산 삽입물과 결합하는 하나 이상의 추가 프로브에 노출시키는 단계를 추가로 포함하고, 단계 (c)는 상기 하나 이상의 추가 프로브의 결합으로부터 신호를 검출하는 단계를 추가로 포함하고, 단계 (d)는 참조 유전자 프로브로부터의 신호를 상기 하나 이상의 추가 핵산 삽입물 프로브로부터의 신호와 비교하여, 상기 핵산 삽입물 내의 상기 하나 이상의 추가 서열에 대한 카피수를 결정하는 단계를 추가로 포함한다. 선택적으로, 상기 핵산 삽입물 내의 상기 하나 이상의 추가 서열은 상기 제2 표적 서열에 인접한 서열을 포함한다.

[0065] 일부 방법에서, 상기 LTVEC는 상기 표적 게놈 유전자좌로부터의 내인성 서열을 결실시키도록 설계되거나, 또는 상기 세포는 Cas 단백질, 표적 게놈 유전자좌 내의 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA, 상기 표적 게놈 유전자좌 내의 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA, 및 tracrRNA와 추가로 접촉된다. 선택적으로, 그러한 방법은 상기 표적 게놈 유전자좌에서의 내인성 서열의 카피수를 결정하는 단계를 추가로 포함한다. 선택적으로, 단계 (b)는 상기 세포의 DNA를 상기 표적 게놈 유전자좌에서의 내인성 서열과 결합하는 프로브에 노출시키는 단계를 추가로 포함하고, 단계 (c)는 상기 내인성 서열 프로브의 결합으로부터 신호를 검출하는 단계를 추가로 포함하고, 단계 (d)는 상기 참조 유전자 프로브로부터의 신호를 상기 내인성 서열 프로브로부터의 신호와 비교하여 상기 내인성 서열에 대한 카피수를 결정하는 단계를 추가로 포함한다.

도면의 간단한 설명

[0066] 도 1은 LTVEC 및 2개 중 어느 하나 또는 2개의 5' 영역(A, B, B2), 중간 영역(C, D), 및 3' 영역(E2, E, F) gRNA를 사용한 마우스 *Lrp5* 엑토도메인의 결실 및 상응하는 인간 *LRP5* 버전에 의한 대체의 동시 수행에 대한 개략도를 나타낸다. LTVEC는 도면의 상부 부분에 나타나 있고, 마우스 *Lrp5* 유전자좌는 도면의 하부 부분에 나타나 있다. 8개의 가이드 RNA에 의해 가이드된 Cas9 절단 부위의 위치들은 마우스 유전자 서열 아래에 수직 화살표로 표시되어 있다. 수평 화살표는 마우스 및 인간 서열에 대한 PCR 프라이머를 나타낸다.

도 2a는 LTVEC 및 2개의 가이드 RNA(가이드 RNA A 및 B)를 사용하는 마우스 유전자의 결실 및 상응하는 인간 버전에 의한 대체의 동시 수행에 대한 일반적인 개략도를 나타낸다. LTVEC는 도 2a의 상부 부분에 나타나 있고, 마우스 유전자좌는 도 2a의 하부 부분에 나타나 있다. 2개의 가이드 RNA에 의해 가이드된 Cas9 절단 부위의 위치들은 마우스 유전자 서열 아래에 화살표로 표시되어 있다.

도 2b 내지 도 2e는 2개의 가이드 RNA가 사용될 때 더 큰 빈도로 일어나는 특유의 이중대립유전자 변형(대립유전자 유형)을 나타낸다. 대각선 빗금이 그려진 두꺼운 선은 마우스 유전자를 나타내고, 점선은 마우스 유전자에서의 결실을 나타내고, 두꺼운 흑색선은 인간 유전자의 삽입을 나타낸다. 도 2b는 동형접합성 붕괴된 대립유

전자들(큰 CRISPR-유도 결실)을 나타낸다. 도 2c는 동형접합성 표적화된 대립유전자들을 나타낸다. 도 2d는 반접합성 표적화된 대립유전자들을 나타낸다. 도 2e는 복합 이형접합성 대립유전자들을 나타낸다.

도 3a 및 도 3b는 선택된 클론의 유전자형을 확인하는 PCR 검정을 나타낸다. 도 3a는 프라이머 m-lr-f 및 m-5'-r을 사용하는 선택된 ES 세포 클론에 대한 긴 범위 PCR 검정의 결과를 나타내는데, 이들 프라이머는 인간 삽입물과, 5' 상동성 아암과 상동성인 서열의 외부에 있는 서열 사이의 결합을 확립하여 올바른 표적화를 제공한다. 도 3b는 5' Del J, 5' Ins J, Del A + F, 및 Del A + E2 PCR 검정으로부터의 결과를 나타낸다. 5' Del J는 m-5'-f 및 m-5-r 프라이머를 사용한 PCR 산물을 나타내며, 이는 gRNA A 절단 부위 주위의 야생형 서열을 증폭시켜 이 서열의 보유 또는 상실을 확립한다. 5' Ins J는 m-5'-f 및 h-5'-r 프라이머를 사용한 PCR 산물을 나타내며, 이는 인간 삽입물과 마우스 게놈 사이의 결합을 확립한다. 이 검정은 표적화된 클론 및 무작위 통합된 클론 둘 모두에 있어서 양성 결과를 제공할 것이다. Del A + F는 클론 BO-F10 및 AW-A8에서 이중 gRNA A 및 F 절단에 의해 매개된 큰 결실에 대한 예측된 앰플리콘(amplicon) 크기(359 bp) 및 실제의 밴드를 나타낸다. Del A + E2는 클론 BA-A7에 대한 동일한 개념을 나타낸다. NT는 주형 없음을 나타내고, +/-는 모체 VGF1 혼성 ES 세포 야생형 대조군을 나타내고, H/+는 이형접합성 인간화 유전자형을 나타내고, H/ Δ 는 반접합성 인간화 유전자형을 나타내고, H/H는 동형접합성 인간화 유전자형을 나타내고, Δ / Δ 는 동형접합성 결실된 유전자형을 나타낸다.

도 4a 내지 도 4c는, Cas9 및 2개의 gRNA와 조합된 *Lrp5* 인간화 LTVEC로 표적화된, 마우스 ES 세포 클론 AW-D9(도 4a) 및 BA-D5(도 4c), 및 LTVEC를 단독으로 사용하여 표적화된 클론 BS-C4(도 4b)의 형광 동소 혼성화(fluorescence *in situ* hybridization)(FISH) 분석을 나타낸다. 화살표는 염색체 19의 밴드 B에 대한 혼성화 신호의 위치들을 나타낸다. 적색 신호는 단지 마우스 프로브와의 혼성화를 나타낸다(파선 화살표, 도 4b). 황색이 혼합된 색상 신호는 적색 마우스 프로브 및 녹색 인간 프로브 둘 모두와의 혼성화를 나타낸다. 적색 신호를 갖는 하나의 염색체 19 밴드 B(파선 화살표) 및 황색 신호를 갖는 다른 하나의 염색체 19 밴드 B(실선 화살표)는 BS-C4 클론에 대한 올바른 유전자좌 및 이형접합성 유전자형에 대해 표적화하였음을 확인시켜 주었다(도 4b). 황색 신호를 갖는 양쪽 염색체 19의 B 밴드(실선 화살표, 도 4a 및 도 4c)는 AW-D9 및 BS-C4 클론에 대한 올바른 유전자좌 및 동형접합성 유전자형에 대해 표적화하였음을 확인시켜 주었다.

도 5는 VGF1 혼성 ES 세포에서 이형접합성의 상실(LOH)을 분석함으로써 2개의 가이드 RNA에 의해 매개된 유전자 변환 또는 유사분열 재조합 이벤트를 조사하도록 설계된 검정에 의한 염색체 19의 개략도를 나타낸다. TaqMan® qPCR 염색체 카피수(CCN) 검정의 대략적인 위치들이 화살표로 나타나 있다. 구조적 변이체(SV) 다형성(polymorphism) PCR 검정의 대략적인 위치들이 V형 모양으로 나타나 있으며, 이와 함께 *Lrp5* 유전자좌로부터의 그들의 거리(단위: Mb)가 그 위에 주어져 있다. 단일 뉴클레오타이드 변이체(SNV) TaqMan® 대립유전자 식별 검정의 대략적인 위치들이 화살표 모양으로 나타나 있으며, 이와 함께 *Lrp5* 유전자좌로부터의 그들의 거리(단위: Mb)가 그 아래에 주어져 있다. F, E2, D, B2, 및 A에 대한 gRNA 인식 부위들의 위치들이 *Lrp5* 유전자의 도면 위에 대각선 화살표로 나타나 있다.

도 6은 LTVEC 및 2개 중 어느 하나 또는 2개의 5' 영역(A, B), 중간 영역(C, D), 및 3' 영역(E, E2) gRNA를 사용한 마우스 *C5(Hc)* 유전자의 엑손 2로부터 정지 코돈까지의 영역의 결실 및 상응하는 인간 *C5* 버전에 의한 대체의 동시 수행에 대한 개략도를 나타낸다. LTVEC는 도면의 상부 부분에 나타나 있고, 마우스 *C5(Hc)* 유전자좌는 도면의 하부 부분에 나타나 있다. 6개의 가이드 RNA에 의해 가이드된 Cas9 절단 부위의 위치들은 마우스 유전자 서열 아래에 화살표로 표시되어 있다.

도 7a 및 도 7b는, Cas9 및 2개의 gRNA와 조합된 *Hc* 인간화 LTVEC로 표적화된, 마우스 ES 세포 클론 Q-E9(도 7a) 및 O-E3(도 7b)의 형광 동소 혼성화(FISH) 분석을 나타낸다. 화살표는 염색체 2의 밴드 B에 대한 혼성화 신호의 위치들을 나타낸다. 적색 신호는 단지 마우스 프로브와의 혼성화를 나타낸다(파선 화살표, 도 7a). 황색이 혼합된 색상 신호는 적색 마우스 프로브 및 녹색 인간 프로브 둘 모두와의 혼성화를 나타낸다(실선 화살표). 적색 신호를 갖는 하나의 염색체 2 밴드 B(파선 화살표) 및 황색 신호를 갖는 다른 하나의 염색체 2 밴드 B(실선 화살표)는 Q-E9 클론에 대한 올바른 유전자좌 및 이형접합성 유전자형에 대해 표적화하였음을 확인시켜 주었다(도 7a). 황색 신호를 갖는 양쪽 염색체 2의 B 밴드(실선 화살표, 도 7b)는 O-E3 클론에 대한 올바른 유전자좌 및 동형접합성 유전자형에 대해 표적화하였음을 확인시켜 주었다.

도 8은 LTVEC 및 2개 중 어느 하나 또는 2개의 5' 영역(A, B), 중간 영역(D, C), 및 3' 영역(E, F) gRNA를 사용한 마우스 *Ror1* 유전자의 결실 및 상응하는 인간 *ROR1* 버전에 의한 대체의 동시 수행에 대한 개략도를 나타낸다. LTVEC는 도면의 상부 부분에 나타나 있고, 마우스 *Ror1* 유전자좌는 도면의 하부 부분에 나타나 있다. 6개

의 가이드 RNA에 의해 가이드된 Cas9 절단 부위의 위치들은 마우스 유전자 서열 아래에 화살표로 표시되어 있다.

도 9는 LTVEC 및 2개 중 어느 하나 또는 2개의 5' 영역(A, A2, B), 중간 영역(C, D), 및 3' 영역(E2, E, F) gRNA를 사용한 마우스 *Trpa1* 유전자의 결실 및 상응하는 인간 *TRPA1* 버전에 의한 대체의 동시 수행에 대한 개략도를 나타낸다. LTVEC는 도면의 상부 부분에 나타나 있고, 마우스 *Trpa1* 유전자좌는 도면의 하부 부분에 나타나 있다. 8개의 가이드 RNA에 의해 가이드된 Cas9 절단 부위의 위치들은 마우스 유전자 서열 아래에 화살표로 표시되어 있다.

도 10a 내지 도 10e는 클론 BR-B4, BP-G7, B0-G11, B0-F10, B0-A8, 및 BC-H9의 구조적 변이(SV) 검정의 결과를 나타내며, 이때 VGF1(FIH4), 129, 및 B6 DNA가 대조군으로서 사용된다. 이들 검정은 *Lrp5* 유전자좌에 대해 말단소체 측으로 하기의 거리에서 행해진다: 13.7 Mb(도 10a), 20.0 Mb(도 10b), 36.9 Mb(도 10c), 48.3 Mb(도 10d), 및 56.7 Mb(도 10e). B6 및 129 대립유전자에 대한 PCR 산물의 위치들은 화살표로 나타나 있다.

도 11a 내지 도 11c는 *Lrp5*에 대해 동원체 측으로 0.32 Mb(도 11a), *Lrp5*에 대해 동원체 측으로 1.2 Mb(도 11b), 및 *Lrp5*에 대해 동원체 측으로 57.2 Mb(도 11c)에 대한 대립유전자 식별 도표를 나타낸다. 각각의 축 상의 값은 상대 형광 세기를 나타낸다. 이들 도표는 각각의 샘플에 대한 4회의 반복 시험을 나타내는데, 이는 채워진 도트(B6 대립유전자), 채워지지 않은 도트(129 대립유전자), 및 대각선을 갖는 도트(B6/129 대립유전자 둘 모두)로 나타나 있다.

도 12a 내지 도 12c는 이형접합성의 상실에 의해 검출된 폭넓은 유전자 변환 및 동형접합성 이벤트를 일으킬 수 있는 세포 주기의 G2 기 동안 유사분열 재조합에 대해 가능한 기전을 나타낸 개략도이다. 도 12a는 129 상동체 상에서의 표적화된 인간화에 대해 이형접합성인 혼성 129/B6 ES 세포에서 2개의 염색분체를 보여주는 복제된 상동 염색체들을 나타낸다. 양방향 화살표는 이중 gRNA-유도 Cas9 절단에 의해 생성된 잠재적 이중 가닥 브레이크를 나타내는데, 이러한 절단은 상동 염색체들 상의 염색분체들 사이의 상동 재조합에 의한 상호 교환을 촉진시키며, 이는 표적화된 대립유전자의 동원체 측에서의 교차로서 나타나 있으며, 그 결과 도 12b에 나타난 혼성 염색분체가 생성된다. 도 12c는 유사분열 및 세포 분열 후에, 딸 세포로의 4가지 유형의 염색체 분리가 가능함을 나타낸다. 이형접합성의 보유를 갖는 2개, 즉 모체 유형 이형접합체(Hum/+, 좌측 상단) 및 동일한 교환에 의한 이형접합체(Hum/+, 우측 상단)는 LOH 검정에 의해 구별될 수 없다. 다른 2개는 이형접합성의 상실, 즉 말단소체 측 B6 대립유전자의 상실을 갖는 인간화 동형접합체(Hum/Hum, 예를 들어 클론 B0-A8, 좌측 하단) 및 말단소체 측 129 대립유전자의 상실을 갖는 야생형 동형접합체(+/, 우측 하단)를 나타낸다. 이 후자의 유형은 그것이 인간화된 대립유전자의 약물 저항성 카세트를 보유하지 않기 때문에 상실될 것이다.

도 13은 35 kb 및 31 kb의 상동성 아암 크기를 갖는 표적화 벡터(LTVEC) 또는 각각 5 kb의 상동성 아암 크기를 갖는 표적화 벡터(sTVEC) 및 2개 중 어느 하나 또는 2개의 5' 영역(A, B), 중간 영역(C, D), 및 3' 영역(E, E2) gRNA를 사용한 마우스 *C5(Hc)* 유전자의 엑손 2로부터 정지 코돈까지의 영역의 결실 및 상응하는 인간 *C5* 버전에 의한 대체의 동시 수행에 대한 개략도를 나타낸다. 2개의 표적화 벡터는 도면의 상부 부분에 나타나 있고, 마우스 *C5(Hc)* 유전자좌는 도면의 하부 부분에 나타나 있다. 6개의 가이드 RNA에 의해 가이드된 Cas9 절단 부위의 위치들은 마우스 유전자 서열 아래에 수직 화살표로 표시되어 있고, 스크리닝에 사용된 프라이머들은 수평 화살표로 표시되어 있다. 삽입물 카피수를 정량하는 대립유전자의 획득(gain of allele)(GOA) 검정 및 결실에 대해 표적화된 마우스 서열을 정량하는 대립유전자의 상실(loss of allele)(LOA) 검정의 위치들은 삼각형으로 표시되어 있다.

도 14는 LTVEC 및 2개의 5' 영역(A, B) gRNA를 사용한 마우스 *Cmah* 유전자의 처음 5개의 엑손의 결실 및 *lacZ* 리포터 및 하이그로마이신 저항성 선택 카세트에 의한 대체의 동시 수행에 대한 개략도를 나타낸다. LTVEC는 도면의 상부 부분에 나타나 있고, 마우스 *Cmah* 유전자좌는 도면의 하부 부분에 나타나 있다. 2개의 가이드 RNA에 의해 가이드된 Cas9 절단 부위의 위치들은 마우스 유전자 서열 아래에 수직 화살표로 표시되어 있고, 삽입물 카피수를 정량하는 GOA 검정 및 결실에 대해 표적화된 마우스 서열을 정량하는 LOA 검정의 위치들은 삼각형으로 표시되어 있다.

도 15는 마우스 *Cmah* 유전자좌(서열 번호 109)가 2개의 5' 영역 gRNA(A 및 B; 각각 서열 번호 107 및 108)로 표적화될 때의 절단 이벤트 및 생성된 절체 산물(서열 번호 112)의 개략도를 나타낸다. *Cmah* 유전자좌에 혼성화된 gRNA 서열은 볼드체이고, Cas9 단백질은 작은 점들이 있는 타원형으로 표시되어 있고, Cas9 절단 부위는 수직 화살표로 나타나 있고, 프로토스페이스 인접 모티프(protospacer adjacent motif)(PAM)는 박스로 표시되어 있다. TaqMan® LOA 검정 정방향 프라이머, 프로브, 및 역방향 프라이머의 대략적인 위치들은 도면의 상단에

수평 바 및 화살표로 표시되어 있다. 절단 및 절제 후에 생성된 5' 및 3' 단편은 각각 서열 번호 110 및 111이다.

도 16a 내지 도 16e는 129S6/SvEvTac 마우스 계통으로부터 유래된 하나의 반수체의 염색체 상보체 및 C57BL/6NTac(B6) 마우스 계통으로부터 유래된 하나의 반수체의 염색체 상보체를 갖는 F1 혼성 마우스 ES 세포에서의 CRISPR/Cas9-보조 인간화 실험에서, 이형접합성의 상실(LOH)을 포함한, 관찰된 결과를 설명하는 가능한 기전을 나타낸다. 도 16a는, 이형접합성 변형이 게놈 복제 전에 또는 게놈 복제 후에 129 염색체 상에서 일어난 후 자매 염색분체들 사이에 유전자 변환이 행해지는, 유사분열 교차에 의한 상호 염색분체 교환을 나타낸다. 도 16b는, 단일 129 염색분체가 게놈 복제 후에 변형되는, 유사분열 교차에 의한 상호 염색분체 교환을 나타낸다. 도 16c는, LTVEC 표적화는 일어나지 않았지만, Cas9 절단이 129 또는 B6 염색체 상에서 일어난(B6 절단이 도시되어 있음), 유사분열 교차에 의한 상호 염색분체 교환을 나타낸다. 도 16d는, 이형접합성 변형이 게놈 복제 전에 또는 게놈 복제 후에 129 염색체 상에서 일어난 후 자매 염색분체들 사이에 유전자 변환이 행해지는, 브레이크-유도 복제에 의한 염색분체 카피를 나타낸다. 도 16e는, 단일 129 염색분체가 게놈 복제 후에 변형되는, 브레이크-유도 복제에 의한 염색체 카피를 나타낸다. 도 16f는, LTVEC 표적화는 일어나지 않았지만, Cas9 절단이 129 또는 B6 염색체 상에서 일어난(B6 절단이 도시되어 있음), 브레이크-유도 복제에 의한 염색분체 카피를 나타낸다.

도 17a 내지 도 17c는 표적화된 변형에 대한 스크리닝 전략을 나타낸다. 도 17a는, 마우스 염색체 내의 내인성 서열이 결실되고 Neo-SDC 삽입물로 교체되는 대형 표적화 벡터(LTVEC)에 의한 이형접합성 표적화를 검출하기 위한 표준적인 대립유전자의 변형(MOA) 스크리닝 전략을 나타낸다. 이 전략은 결실에 대해 표적화된 내인성 서열의 상류 및 하류 영역에 대해 TaqMan® 프로브 mTU 및 mTD를 사용한다. 도 17b는 대립유전자의 변형(MOA) 검정(대립유전자의 상실(LOA) 검정에 대해서는 mTGU, mTM, 및 mTGD 프로브, 및 대립유전자의 획득(GOA) 검정에 대해서는 hTU 및 hTD 프로브)과 조합하여 TaqMan® 보유 검정(retention assay)(retU 및 retD 프로브)을 사용하여 CRISPR/Cas9-보조 인간화에 대해 스크리닝하는 것을 나타낸다. 도 17c는, 대립유전자의 상실(LOA) 검정(mTGU, mTM, 및 mTGD) 프로브)과 조합하여 TaqMan® 보유 검정(retU 및 retD 프로브)을 사용하여, 쌍 형성된 가이드 RNA들(gU 및 gD)을 사용한 CRISPR/Cas9-보조 결실에 대해 스크리닝하는 것을 나타낸다.

도 18은, loxP 부위들에 의해 플랭킹된 Pkg-Neo 삽입물(네오마이신 포스포트랜스퍼라제 유전자에 작동가능하게 연결된 포스포글리세레이트 키나제 I 프로모터)을 갖는 표적화 벡터 및 인간 대응부(삼각형)로 대체된 가변 영역 유전자 세그먼트를 갖는 마우스 면역글로불린 중쇄 유전자좌의 대략 900 kb 영역의 개략도(축적대로 그려지지 않음)를 나타낸다. 2개의 gRNA를 사용하여 5' 말단에서 마우스 면역글로불린 중쇄 유전자좌를 절단하고, 2개의 gRNA를 사용하여 3' 말단에서 유전자좌를 절단하고, 표적화 벡터는 마우스 면역글로불린 중쇄 유전자좌를 결실시키고 이를 Pkg-Neo 삽입물로 대체한다. 4개의 가이드 RNA에 의해 가이드된 Cas9 절단 부위의 위치들은 표적 유전자좌 아래에 수직 화살표로 표시되어 있다. 원으로 둘러싸인 수평선은 대립유전자의 변형(MOA) 검정(hIGH31, hIGH1, mIGHA1, mIGHA7, 및 hIGH9) 및 보유 검정(5' IgH Arm 1, 5' IgH Arm 2, mIGM-398, 및 mIGM-1045)에 대한 TaqMan® 프로브를 나타낸다.

정의

본 명세서에서 상호교환적으로 사용되는 용어 "단백질", "폴리펩티드", 및 "펩티드"는 코딩 및 비-코딩된 아미노산, 및 화학적으로 또는 생화학적으로 변형되거나 유도체화된 아미노산을 비롯한 임의의 길이의 아미노산의 중합체 형태를 포함한다. 이들 용어는 또한 변형된 중합체, 예컨대 변형된 펩티드 골격을 갖는 폴리펩티드를 포함한다.

본 명세서에서 상호교환적으로 사용되는 용어 "핵산" 및 "폴리뉴클레오타이드"는 리보뉴클레오타이드, 데옥시리보뉴클레오타이드, 또는 이들의 유사체 또는 변형된 버전을 비롯한 임의의 길이의 뉴클레오타이드의 중합체 형태를 포함한다. 이들은 단일-가닥, 이중-가닥, 및 다중-가닥 DNA 또는 RNA, 게놈 DNA, cDNA, DNA-RNA 혼성체, 및 푸린 염기, 피리미딘 염기, 또는 다른 천연, 화학적으로 변형된, 생화학적으로 변형된, 비천연, 또는 유도체화된 뉴클레오타이드 염기를 포함하는 중합체를 포함한다.

"코돈 최적화"는 일반적으로 천연 아미노산 서열을 유지하면서 천연 서열의 적어도 하나의 코돈을 숙주 세포의 유전자에서 더 빈번하게 또는 가장 빈번하게 사용되는 코돈으로 치환함으로써 특정 숙주 세포에서의 향상된 발현을 위해 핵산 서열을 변형시키는 방법을 포함한다. 예를 들어, Cas 단백질을 인코딩하는 핵산은 자연적으로 발생한 핵산 서열과 비교하여, 세균 세포, 효모 세포, 인간 세포, 비인간 세포, 포유류 세포, 설치류 세포, 마우스 세포, 래트 세포, 햄스터 세포, 또는 임의의 다른 숙주 세포를 비롯한 주어진 원핵 또는 진핵 세포에서 더

높은 사용 빈도를 갖는 코돈으로 치환되도록 변형될 수 있다. 코돈 사용 빈도 표는, 예를 들어 "코돈 사용 빈도 데이터베이스(Codon Usage Database)"에서 용이하게 이용할 수 있다. 이러한 표는 다수의 방법으로 조정될 수 있다. 문헌[Nakamura *et al.* (2000) *Nucleic Acids Research* 28:292]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 특정 숙주에서의 발현을 위해 특정 서열의 코돈 최적화를 위한 컴퓨터 알고리즘이 또한 이용가능하다(예를 들어, 문헌[Gene Forge] 참조).

"작동가능한 연결" 또는 "작동가능하게 연결된"은 2개 이상의 성분(예를 들어, 프로모터 및 다른 서열 요소)의 병치(juxtaposition)를 포함하는데, 이때 이들 두 성분은 정상적으로 기능하고, 이들 성분 중 적어도 하나가 나머지 다른 성분들 중 적어도 하나에 미치는 기능을 매개할 수 있는 가능성을 허용하도록 한다. 예를 들어, 프로모터가 하나 이상의 전사 조절 인자의 존재 또는 부재에 따라 코딩 서열의 전사 수준을 제어하는 경우에, 프로모터는 코딩 서열에 작동 가능하게 연결될 수 있다.

핵산의 "상보성"은 하나의 핵산 가닥의 뉴클레오티드 서열이 이의 핵산 염기 그룹의 배향으로 인해, 대향하는 핵산 가닥 상의 다른 서열과 수소 결합을 형성한다는 것을 의미한다. DNA의 상보적 염기는 전형적으로 A와 T 및 C와 G이다. RNA에서, 이들은 전형적으로 C와 G 및 U와 A이다. 상보성은 완전하거나 상당/충분할 수 있다. 2개의 핵산들 사이의 완전한 상보성은 2개의 핵산이, 듀플렉스의 모든 염기가 왓슨-크릭 염기쌍 형성(Watson-Crick pairing)에 의해 상보적 염기에 결합되는 듀플렉스를 형성할 수 있음을 의미한다. "상당한" 또는 "충분한" 상보성은 한쪽 가닥의 서열이 대향하는 가닥의 서열에 전적으로 맞/또는 완전히 상보적이지는 않지만, 일련의 혼성화 조건(예를 들어, 염 농도 및 온도)에서 안정한 혼성 복합체를 형성하도록 두 가닥 상의 염기들 사이에 충분한 결합이 일어남을 의미한다. 그러한 조건은 혼성화된 가닥의 T_m(용융 온도)을 예측하는 서열 및 표준 수학적 계산을 이용하여 예측될 수 있거나, 일상적인 방법을 사용하여 T_m의 경험적 결정에 의해 예측될 수 있다. T_m은 2개의 핵산 가닥들 사이에 형성된 혼성화 복합체 집단이 50% 변성되는 온도를 포함한다. T_m보다 낮은 온도에서는 혼성화 복합체의 형성이 촉진되는 반면에, T_m보다 높은 온도에서는 혼성화 복합체의 가닥들의 용융 또는 분리가 촉진된다. T_m은, 예를 들어 T_m=81.5+0.41(% G+C)을 사용하여 1 M NaCl 수용액 중의 기지의 G+C 함량을 갖는 핵산에 대하여 추정될 수 있지만, 다른 공지된 T_m 계산은 핵산 구조적 특성을 고려한다.

"혼성화 조건"은 하나의 핵산 가닥이 상보적 가닥 상호작용 및 수소 결합에 의해 제2 핵산 가닥에 결합하여 혼성화 복합체를 생성하는 누적 환경을 포함한다. 그러한 조건은 핵산을 함유하는 수용액 또는 유기 용액의 화학 성분 및 이의 농도(예를 들어, 염, 킬레이트화제, 포름아미드), 및 혼합물의 온도를 포함한다. 다른 인자, 예컨대 배양 기간 또는 반응 챔버 치수가 환경에 기여할 수 있다. 예를 들어, 문헌[Sambrook *et al.*, Molecular Cloning, A Laboratory Manual, 2.sup.nd ed., pp. 1.90-1.91, 9.47-9.51, 11.47-11.57 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989)]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

혼성화는 2개의 핵산이 상보적 서열을 함유하는 것을 필요로 하지만, 염기 간의 불일치(mismatch)가 가능하다. 2개의 핵산들 사이의 혼성화에 적절한 조건은 당업계에서 익히 공지된 변수인 핵산 길이 및 상보성 정도에 좌우된다. 2개의 뉴클레오티드 서열들 사이의 상보성 정도가 클수록, 그러한 서열들을 갖는 핵산들의 혼성체에 대한 용융 온도(T_m)의 값이 커진다. 짧은 스트레치(stretch)의 상보성(예를 들어, 35개 이하, 30개 이하, 25개 이하, 22개 이하, 20개 이하 또는 18개 이하의 뉴클레오티드에 걸친 상보성)을 갖는 핵산들 사이의 혼성화의 경우, 불일치 위치가 중요해진다(상기 문헌[Sambrook *et al.*, 11.7-11.8] 참조). 전형적으로, 혼성화가능한 핵산 길이는 적어도 약 10개의 뉴클레오티드이다. 혼성화가능한 핵산의 예시적인 최소 길이는 적어도 약 15개의 뉴클레오티드, 적어도 약 20개의 뉴클레오티드, 적어도 약 22개의 뉴클레오티드, 적어도 약 25개의 뉴클레오티드, 및 적어도 약 30개의 뉴클레오티드를 포함한다. 더욱이, 온도 및 세척 용액 염 농도는 필요에 따라 상보성 영역의 길이 및 상보성 정도와 같은 인자에 따라 조정될 수 있다.

폴리뉴클레오티드의 서열은 특이적으로 혼성화가능하게 되는 이의 표적 핵산의 서열에 대하여 100% 상보적일 필요는 없다. 더욱이, 개재 또는 인접 세그먼트가 혼성화 이벤트(예를 들어, 루프 구조 또는 헤어핀 구조)에 관여하지 않도록 하나 이상의 세그먼트에 걸쳐 폴리뉴클레오티드가 혼성화될 수 있다. 폴리뉴클레오티드(예를 들어, gRNA)는, 이것이 표적화하는 표적 핵산 서열 내의 표적 영역에 대하여 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 또는 100% 서열 상보성을 포함할 수 있다. 예를 들어, 20개의 뉴클레오티드 중 18개가 표적 영역에 상보적이며, 이에 따라 특이적으로 혼성화되는 gRNA는 90% 상보성을 나타낼 것이다. 이러한 예에서, 나머지 비상보적 뉴클레오티드들은 상보적 뉴클레오티드와 클러스터를 이루거나 그들 사이에 상보적 뉴클레오티드가 산재될 수 있으며, 서로끼리 또는 상보적 뉴클레오티드에 인접할 필요는 없다.

핵산 내의 핵산 서열의 특정 스트레치들 사이의 %상보성은 당업계에 공지된 BLAST 프로그램(기본 논리 정렬 검색 툴) 및 PowerBLAST 프로그램을 사용하거나(문헌[Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410]; 문헌[Zhang and Madden (1997) *Genome Res.* 7:649-656]), 또는 스미스-워터맨(Smith and Waterman) 알고리즘을 사용하는, 디폴트 설정(default setting)을 사용한 Gap 프로그램(유닉스용 위스콘신 서열 분석 패키지(Wisconsin Sequence Analysis Package), 버전 8, 미국 위스콘신주 매디슨 소재의 University Research Park의 Genetics Computer Group)을 사용함으로써(문헌[Adv. Appl. Math., 1981, 2, 482-489]) 일상적으로 결정될 수 있다.

본 명세서에 제공된 방법 및 조성물은 다양한 상이한 성분을 사용한다. 일부 성분들이 활성 변이체 및 단편을 가질 수 있다는 것이 상세한 설명 전반에 걸쳐 인식된다. 그러한 성분은, 예를 들어 Cas 단백질, CRISPR RNA, tracrRNA, 및 가이드 RNA를 포함한다. 이들 성분 각각에 대한 생물학적 활성은 본 명세서의 어딘가 다른 곳에서 기재되어 있다.

2개의 폴리뉴클레오타이드 또는 폴리펩티드 서열과 관련하여 "서열 동일성" 또는 "동일성"은 2개의 서열에서 잔기들이 명시된 비교 창(window)에 걸쳐 최대로 상응하도록 정렬되는 경우 동일하다는 것을 언급한다. 서열 동일성의 백분율이 단백질과 관련하여 사용될 때, 동일하지 않은 잔기 위치들이 흔히 보존적 아미노산 치환에 의해 차이가 있는 것으로 인지되고, 여기서 아미노산 잔기들은 유사한 화학적 성질(예를 들어, 전하 또는 소수성)을 갖는 다른 아미노산 잔기들로 치환되므로 분자의 기능적 성질들을 변화시키지 않는다. 서열이 보존적 치환으로 차이가 있을 때, %서열 동일성은 치환의 보존적 성질에 대해 보정하기 위해 상향 조정될 수 있다. 그러한 보존적 치환에 의해 차이가 있는 서열은 "서열 유사성" 또는 "유사성"을 갖는다고 한다. 이러한 조정을 하기 위한 수단은 당업자에게 널리 공지되어 있다. 전형적으로 이는 보존적 치환을 완전 불일치라기보다는 부분 불일치로서 스코어링하여, 백분율 서열 동일성을 증가시키는 것을 포함한다. 따라서, 예를 들어, 동일한 아미노산에는 1의 스코어가 주어지고 비-보존적 치환에는 0의 스코어가 주어지고, 보존적 치환에는 0과 1 사이의 스코어가 주어진다. 보존적 치환의 스코어링은, 예를 들어, 프로그램 PC/GENE(미국 캘리포니아주 마운틴 뷰 소재의 IntelliGenetics)에서 구현되는 바와 같이 계산된다.

"서열 동일성의 백분율"은 비교 창에 걸쳐 2개의 최적으로 정렬된 서열을 비교함으로써 결정된 값을 포함하고, 여기서 비교 창에서 폴리뉴클레오타이드 서열의 일부는 2개의 서열의 최적의 정렬에 대해 참조 서열(이는 부가 또는 결실을 포함하지 않음)과 비교하여 부가 또는 결실(즉, 갭)을 포함할 수 있다. 백분율은 동일한 핵산 염기 또는 아미노산 잔기들이 양 서열에 존재하는 위치의 수를 결정하여 매칭된 위치의 수를 산출하고, 매칭된 위치의 수를 비교 창 내의 위치의 총수로 나누고, 결과에 100을 곱하여 서열 동일성의 백분율을 산출함으로써 계산된다.

달리 언급되지 않는 한, 서열 동일성/유사성 값은 하기의 파라미터를 사용하여 GAP 버전 10을 사용하여 얻은 값을 포함한다: 50의 GAP 중량 및 3의 길이 중량, 및 nwsgapdna.cmp 스코어링 매트릭스를 사용한 뉴클레오타이드 서열에 대한 %동일성 및 %유사성; 8의 GAP 중량 및 2의 길이 중량, 및 BLOSUM62 스코어링 매트릭스를 사용한 아미노산 서열에 대한 %동일성 및 %유사성; 또는 이와 동등한 임의의 프로그램. "동등한 프로그램"은 대상이 되는 임의의 2개의 서열에 대하여 GAP 버전 10에 의해 생성된 상응하는 정렬과 비교할 때 동일한 뉴클레오타이드 또는 아미노산 잔기 매치 및 동일한 %서열 동일성을 갖는 정렬을 생성하는 임의의 서열 비교 프로그램을 포함한다.

용어 "시험관내(*in vitro*)"는 인공 환경을 포함하고, 인공 환경(예를 들어, 시험관) 내에서 일어나는 공정 또는 반응을 지칭한다. 용어 "생체내(*in vivo*)"는 천연 환경(예를 들어, 세포 또는 유기체 또는 신체)을 포함하고, 천연 환경 내에서 일어나는 공정 또는 반응을 지칭한다. 용어 "생체외(*ex vivo*)"는 개체의 신체로부터 적출된 세포를 포함하고, 그러한 세포 내에서 일어나는 공정 또는 반응을 지칭한다.

하나 이상의 열거된 요소를 "포함하는" 또는 "함유하는" 조성물 또는 방법은 구체적으로 열거되지 않은 다른 요소를 포함할 수 있다. 예를 들어, 단백질을 "포함하는" 또는 "함유하는" 조성물은 단백질을 단독으로 함유하거나 다른 성분과 조합하여 함유할 수 있다.

일정 범위의 값의 지정은 범위 내 또는 그 범위를 한정하는 모든 정수, 및 그 범위 내의 정수로 한정되는 모든 하위범위를 포함한다.

문맥상 달리 명백하지 않는 한, 용어 "약"은 명시된 값의 측정 표준 오차(예를 들어, SEM) 내의 값을 포함한다.

문맥상 달리 명확히 지시하지 않는 한, 관사의 단수형("a", "an" 및 "the")은 복수형 지시 대상을 포함한다. 예를 들어, 용어 "Cas 단백질" 또는 "적어도 하나의 Cas 단백질"은 이들의 혼합물을 비롯하여, 복수의 Cas 단백질을 포함할 수 있다.

발명을 실시하기 위한 구체적인 내용

I. 개요

세포 내에서 게놈을 변형시키는 방법 및 조성물이 제공된다. 본 방법 및 조성물은 단일 게놈 표적 유전자좌 내의 상이한 부위를 표적화하는 2개의 가이드 RNA(gRNA)를 사용하는 CRISPR/Cas 시스템을 사용한다. 예를 들어, 본 방법 및 조성물은, 이러한 2개의 가이드 RNA(gRNA)를 사용하는 CRISPR/Cas 시스템을 사용하여, 단일 게놈 표적 유전자좌 내의 상이한 부위에서 쌍 형성된 이중-가닥 브레이크를 생성할 수 있다. 대안적으로, 본 방법 및 조성물은, 이러한 2개의 가이드 RNA(gRNA)를 사용하는 CRISPR/Cas 시스템을 사용하여, 단일 게놈 표적 유전자좌 내의 상이한 부위에서 쌍 형성된 단일-가닥 브레이크를 생성할 수 있다. 일부 방법에서, 2개 이상(예를 들어, 3개 또는 4개)의 가이드 RNA가 사용되어, 예를 들어 단일 게놈 표적 유전자좌 내의 상이한 부위에서 2개 이상의 단일-가닥 브레이크 또는 이중-가닥 브레이크를 생성할 수 있다.

일부 방법은 이중대립유전자 유전자 변형을 촉진하고 게놈 붕괴를 포함하는데, 이때 게놈 붕괴에 의해서는 긴 핵산 서열이 염색체로부터 2개의 절단 부위들 사이에서 결실된다. 다른 방법은 이중대립유전자 유전자 변형을 촉진하고, 세포 내의 핵산 서열의 결실과 외인성 핵산 서열로의 대체의 동시 수행을 포함한다. 하기에 추가로 상세히 설명된 바와 같이, 2개의 gRNA를 사용하는 이들 방법은 단일 표적화 단계로 이중대립유전자 표적화된 유전자 변형을 갖는 세포 또는 동물의 생성을 촉진함으로써 그러한 세포 또는 동물의 생성 효율을 증가시킨다. 결과적으로, 이중대립유전자 표적화된 유전자 변형을 갖는 동물을 생성하는 데 필요한 동물 및 교배의 수가 감소된다.

다른 방법은 유전자 변환 또는 이형접합성의 상실을 포함하는데, 이에 의해서는, 대립유전자에 대해 이형접합성인 게놈이 상응하는 상동 염색체 상의 상응하는 대립유전자에서의 2개의 gRNA에 의해 결정된 부위들에서의 절단을 통해 상기 대립유전자에 대해 동형접합성이 되도록 변형된다. 하기에 추가로 상세히 설명된 바와 같이, 이들 방법에서의 2개의 gRNA의 사용은 유전자 변환의 빈도를 증가시키고 큰 트랙(track)의 염색체 DNA에 걸친 유전자 변환을 가능하게 한다.

II. CRISPR/Cas 시스템

본 명세서에 개시된 방법 및 조성물은 클러스터링된 규칙적으로 사이 공간을 둔 짧은 회문성 반복부(CRISPR)/CRISPR-관련(Cas) 시스템 또는 그러한 시스템의 성분들을 사용하여 세포 내의 게놈을 변형시킬 수 있다. CRISPR/Cas 시스템은 전사체 및 Cas 유전자의 발현에 관여하거나, 이의 활성을 유도하는 다른 요소들을 포함한다. CRISPR/Cas 시스템은 타입 I, 타입 II, 또는 타입 III 시스템일 수 있다. 본 명세서에 개시된 방법 및 조성물은 핵산의 부위-지정 절단에 대한 CRISPR 복합체(Cas 단백질과 복합체화된 가이드 RNA(gRNA)를 포함)를 사용함으로써 CRISPR/Cas 시스템을 사용한다.

본 명세서에 개시된 방법에서 사용된 일부 CRISPR/Cas 시스템은 비자연적으로 발생한다. "비자연적으로 발생한" 시스템은, 시스템의 하나 이상의 성분들이 그들의 자연적으로 발생한 상태로부터 변경 또는 돌연변이되거나, 그들이 자연에서 자연적으로 관련된 적어도 하나의 다른 성분이 적어도 실질적으로 없거나, 그들이 자연적으로 관련되지 않은 적어도 하나의 다른 성분과 관련되는 것과 같이 사람의 손의 개입을 나타내는 임의의 것을 포함한다. 예를 들어, 일부 CRISPR/Cas 시스템은 자연적으로는 함께 발생하지 않는 gRNA 및 Cas 단백질을 포함하는 비자연 발생 CRISPR 복합체를 사용한다.

A. Cas RNA-가이드된 엔도뉴클레아제

Cas 단백질은 일반적으로 적어도 하나의 RNA 인식 또는 결합 도메인을 포함한다. 그러한 도메인은 가이드 RNA(gRNA, 아래에 더욱 자세하게 설명됨)와 상호작용할 수 있다. Cas 단백질은 또한 뉴클레아제 도메인(예를 들어, DNase 또는 RNase 도메인), DNA 결합 도메인, 헬리카제 도메인, 단백질-단백질 상호작용 도메인, 이량체화 도메인, 및 다른 도메인을 포함할 수 있다. 뉴클레아제 도메인은 핵산 절단을 위한 촉매 활성을 갖는다. 절단은 핵산 분자의 공유 결합의 파괴를 포함한다. 절단은 평활 말단(blunt end) 또는 엇갈린 말단(staggered end)을 생성할 수 있고, 이는 단일-가닥 또는 이중-가닥이 되게 할 수 있다.

Cas 단백질의 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e(CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9(Csn1 또는 Csx12), Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1(CasA), Cse2(CasB), Cse3(CasE), Cse4(CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15,

Csf1, Csf2, Csf3, Csf4, 및 Cu1966, 및 이들의 상동체 또는 변형된 버전을 포함한다.

[0077] 일부 경우에, Cas 단백질은 타입 II CRISPR/Cas 시스템으로부터 유래된다. 예를 들어, Cas 단백질은 Cas9 단백질일 수 있거나 Cas9 단백질로부터 유래될 수 있다. 전형적으로, Cas9 단백질은 보존된 구조와 함께 4개의 주요 모티프를 공유한다. 모티프 1, 2, 및 4는 RuvC-유사 모티프이고, 모티프 3은 HNH 모티프이다. Cas9 단백질은, 예를 들어, 스트렙토코쿠스 피오게네스(*Streptococcus pyogenes*), 스트렙토코쿠스 테르모필루스(*Streptococcus thermophilus*), 스트렙토코쿠스 종(*Streptococcus* sp.), 스태필로코쿠스 아우레우스(*Staphylococcus aureus*), 노카르디옵시스 다손빌레이(*Nocardia opsis dassonvillei*), 스트렙토미세스 프리스티나이스피랄리스(*Streptomyces pristinaespiralis*), 스트렙토미세스 비리도크로모게네스(*Streptomyces viridochromogenes*), 스트렙토미세스 비리도크로모게네스, 스트렙토스포랑기움 로세움(*Streptosporangium roseum*), 스트렙토스포랑기움 로세움, 알리시클로바실루스 아시도칼다리우스(*Alicyclobacillus acidocaldarius*), 바실루스 슈도미코이데스(*Bacillus pseudomycoides*), 바실루스 셀레니티레두센스(*Bacillus selenitireducens*), 엑시구오박테리움 시비리쿰(*Exiguobacterium sibiricum*), 락토바실루스 델브루에키(*Lactobacillus delbrueckii*), 락토바실루스 살리바리우스(*Lactobacillus salivarius*), 마이크로실라 마리나(*Microscilla marina*), 부르크홀데리아레스 박테리움(*Burkholderiales bacterium*), 폴라로모나스 나프탈레니보란스(*Polaromonas naphthalenivorans*), 폴라로모나스 종(*Polaromonas* sp.), 크로코스피아라 왓소니(*Crocospira watsonii*), 시아노테세 종(*Cyanotheca* sp.), 마이크로시스티스 아이루기노사(*Microcystis aeruginosa*), 시넨코코쿠스 종(*Synechococcus* sp.), 아세토할로비움 아라바티쿰(*Acetohalobium arabaticum*), 암모니펙스 데겐시(*Ammonifex degensii*), 칼디세룰로시럽토르 벅스키(*Caldicelulosiruptor beccsii*), 칸디다투스 데술포르디스(*Candidatus Desulforudis*), 클로스트리디움 보툴리눔(*Clostridium botulinum*) 클로스트리디움 디피실레(*Clostridium difficile*), 피네골디아 마그나(*Fingoldia magna*), 나트라나에로비우스 테르모필루스(*Natranaerobius thermophilus*), 펠로토마쿨룸 테르모프로피오니쿰(*Pelotomaculum thermopropionicum*), 아시디티오바실루스 칼두스(*Acidithiobacillus caldus*), 아시디티오바실루스 페로옥시단스(*Acidithiobacillus ferrooxidans*), 알로크로마티움 비노숨(*Allochromatium vinosum*), 마리노박테르 종(*Marinobacter* sp.), 니트로소코쿠스 할로필루스(*Nitrosococcus halophilus*), 니트로소코쿠스 왓소니(*Nitrosococcus watsoni*), 슈도알테로모나스 할로플란크티스(*Pseudoalteromonas haloplanktis*), 크테노도박테르 라세미페르(*Ktedonobacter racemifer*), 메타노할로비움 에베스티가툼(*Methanohalobium evestigatum*), 아나바이나 바리아빌리스(*Anabaena variabilis*), 노둘라리아 스푼미게나(*Nodularia spumigena*), 노스톡 종(*Nostoc* sp.), 아르트로스피라 막시마(*Arthrospira maxima*), 아르트로스피라 플라텐시스(*Arthrospira platensis*), 아르트로스피라 종(*Arthrospira* sp.), 링비아 종(*Lyngbya* sp.), 마이크로콜레우스 크토노플라스테스(*Microcoleus chthonoplastes*), 오스킬라토리아 종(*Oscillatoria* sp.), 페트로토가 모빌리스(*Petrotoxa mobilis*), 테르모시포 아프리카누스(*Thermosiphon africanus*) 또는 아카리오크로리스 마리나(*Acaryochloris marina*)로부터 유래될 수 있다. Cas9 패밀리 구성원의 추가적인 예는 전체적으로 본 명세서에 참고로 포함되는 국제특허 공개 WO 2014/131833호에 기재된 것들을 포함한다. 구체적인 예에서, Cas9 단백질은 S. 피오게네스로부터 유래된 Cas9 단백질이거나 그로부터 유래된다. S. 피오게네스로부터 유래된 Cas9 단백질의 아미노산 서열은, 예를 들어 SwissProt 데이터베이스에서 수탁 번호 Q99ZW2로 찾을 수 있다.

[0078] Cas 단백질은 야생형 단백질(즉, 자연에서 발생한 것), 변형된 Cas 단백질(즉, Cas 단백질 변이체), 또는 야생형 또는 변형된 Cas 단백질의 단편일 수 있다. Cas 단백질은 또한 야생형 또는 변형된 Cas 단백질의 활성 변이체 또는 단편일 수 있다. 활성 변이체 또는 단편은 야생형 또는 변형된 Cas 단백질 또는 이들의 일부에 대해 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 포함할 수 있고, 여기서 활성 변이체는 원하는 절단 부위를 자르는 능력을 보유하고 있어서 Nick-유도 또는 이중-가닥-브레이크-유도 활성을 보유한다. Nick-유도 또는 이중-가닥-브레이크-유도 활성에 대한 검정법은 공지되어 있고, 일반적으로 절단 부위를 함유하는 DNA 기질 상에서 Cas 단백질의 전체 활성 및 특이성을 측정한다.

[0079] Cas 단백질은 핵산 결합 친화성, 핵산 결합 특이성, 및/또는 효소 활성을 증가 또는 감소시키기 위해 변형될 수 있다. Cas 단백질은 또한 단백질의 임의의 다른 활성 또는 성질, 예컨대 안정성을 변화시키기 위해 변형될 수 있다. 예를 들어, Cas 단백질의 하나 이상의 뉴클레아제 도메인은 변형되거나, 결실되거나, 불활성화될 수 있거나, 또는 Cas 단백질은 단백질의 기능에 본질적이지 않은 도메인을 제거하거나 Cas 단백질의 활성을 최적화(예를 들어, 향상 또는 감소)시키기 위해 절단될 수 있다.

[0080] 일부 Cas 단백질은 적어도 2개의 뉴클레아제 도메인, 예컨대 DNase 도메인을 포함한다. 예를 들어, Cas9 단백질은 RuvC-유사 뉴클레아제 도메인 및 HNH-유사 뉴클레아제 도메인을 포함할 수 있다. RuvC 및 HNH 도메인 각

각은 이중-가닥 DNA의 상이한 가닥을 잘라서 DNA에 이중-가닥 브레이크를 만들 수 있다. 예를 들어, 문헌 [Jinek *et al.* (2012) *Science* 337:816-821]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0081] 뉴클레아제 도메인들 중 하나 또는 둘 모두는 더 이상 기능적이지 않거나 뉴클레아제 활성이 감소되도록 결실되거나 돌연변이될 수 있다. 뉴클레아제 도메인들 중 하나가 결실되거나 돌연변이된다면, 생성되는 Cas 단백질 (예를 들어, Cas9)은 Nickase로서 지칭될 수 있고 이중-가닥 DNA 내 CRISPR RNA 인식 서열에 이중-가닥 브레이크가 아니라 단일-가닥 브레이크를 생성할 수 있다(즉, 이것은 상보적 가닥과 비-상보적 가닥 둘 모두가 아니라 상보적 가닥 또는 비-상보적 가닥을 절단할 수 있다). 뉴클레아제 도메인 둘 모두가 결실되거나 돌연변이된다면, 생성되는 Cas 단백질(예를 들어, Cas9)은 이중-가닥 DNA의 가닥 둘 모두를 절단하는 능력이 감소될 것이다. Cas9을 Nickase로 변환시키는 돌연변이의 예는 S. 피오게네스 유래의 Cas9의 RuvC 도메인에서의 D10A(Cas9의 위치 10에서 아스파르트산을 알라닌으로) 돌연변이이다. 마찬가지로, S. 피오게네스 유래의 Cas9의 HNH 도메인에서의 H939A(아미노산 위치 839에서 히스티딘을 알라닌으로) 또는 H840A(아미노산 위치 840에서 히스티딘을 알라닌으로)가 Cas9을 Nickase로 변환시킬 수 있다. Cas9을 Nickase로 변환시키는 돌연변이의 다른 예는 S. 테르모필루스 유래의 Cas9에 상응하는 돌연변이를 포함한다. 예를 들어, 문헌[Sapranaukas *et al.* (2011) *Nucleic Acids Research* 39:9275-9282] 및 국제특허 공개 WO 2013/141680호를 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 그러한 돌연변이는 부위-지정 돌연변이생성, PCR-매개 돌연변이생성, 또는 전체 유전자 합성과 같은 익히 공지된 방법을 사용하여 생성될 수 있다. Nickase를 생성하는 다른 돌연변이의 예는, 예를 들어, 국제특허 공개 WO/2013/176772A1호 및 WO/2013/142578A1호에서 찾을 수 있으며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0082] Cas 단백질은 또한 융합 단백질일 수 있다. 예를 들어, Cas 단백질은 절단 도메인, 후생적 변형 도메인, 전사 활성화 도메인, 또는 전사 억제인자 도메인에 융합될 수 있다. 국제특허 공개 WO 2014/089290호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. Cas 단백질은 또한 증가된 또는 감소된 안정성을 제공하는 이중 폴리펩티드에 융합될 수 있다. 융합된 도메인 또는 이중 폴리펩티드는 N-말단에, C-말단에, 또는 내부적으로 Cas 단백질 내에 위치될 수 있다.

[0083] Cas 융합 단백질의 일례는 세포내 국재화(subcellular localization)를 제공하는 이중 폴리펩티드에 융합된 Cas 단백질이다. 그러한 서열은, 예를 들어, 핵 국재화 신호(nuclear localization signal)(NLS), 예컨대 핵을 표적화하기 위한 SV40 NLS, 미토콘드리아를 표적화하기 위한 미토콘드리아 국재화 신호(mitochondrial localization signal), ER 보류 신호 등을 포함할 수 있다. 예를 들어, 문헌[Lange *et al.* (2007) *J. Biol. Chem.* 282:5101-5105]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 예를 들어, Cas 단백질은 하나 이상의 핵 국재화 신호(예를 들어, 2개 또는 3개의 핵 국재화 신호)에 융합될 수 있다. 그러한 세포내 국재화 신호는 N-말단에, C-말단에, 또는 Cas 단백질 내의 어디에도 위치될 수 있다. NLS는 염기성 아미노산들의 스트레치(stretch)를 포함할 수 있고, 단립형(monopartite) 서열 또는 양립형(bipartite) 서열일 수 있다.

[0084] Cas 단백질은 또한 세포-투과 도메인을 포함할 수 있다. 예를 들어, 세포-투과 도메인은 HIV-1 TAT 단백질, 인간 B형 간염 바이러스로부터의 TLM 세포-투과 모티프, MPG, Pep-1, VP22, 단순 포진 바이러스로부터의 세포 투과 펩티드, 또는 폴리아르기닌 펩티드 서열로부터 유래될 수 있다. 예를 들어, 국제특허 공개 WO 2014/089290호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 세포-투과 도메인은 N-말단에, C-말단에, 또는 Cas 단백질 내의 어디에도 위치될 수 있다.

[0085] Cas 단백질은 또한 추적 또는 정제를 용이하게 하기 위한 이중 폴리펩티드, 예컨대 형광 단백질, 정제 태그, 또는 에피토프 태그를 포함할 수 있다. 형광 단백질의 예는 녹색 형광 단백질(예를 들어, GFP, GFP-2, tagGFP, turboGFP, eGFP, 에메랄드(Emerald), 아자미 그린(Azami Green), 단량체성(Monomeric) 아자미 그린, CopGFP, AceGFP, ZsGreen1), 황색 형광 단백질(예를 들어, YFP, eYFP, 시트린(Citrine), 비너스(Venus), YPet, PhiYFP, ZsYellow1), 청색 형광 단백질(예를 들어, eBFP, eBFP2, 아주라이트(Azurite), mKalamal, GFPuv, 사파이어(Sapphire), T-사파이어), 시안 형광 단백질(예를 들어 eCFP, 세룰리안(Cerulean), CyPet, AmCyan1, 미도리시-시안(Midoriishi-Cyan)), 적색 형광 단백질(mKate, mKate2, mPlum, DsRed 단량체, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-단량체, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), 오렌지색 형광 단백질(mOrange, mKO, 쿠사비라-오렌지(Kusabira-Orange), 단량체성 쿠사비라-오렌지, mTangerine, tdTomato), 및 임의의 다른 적합한 형광 단백질을 포함한다. 태그의 예는 글루타티온-S-트랜스퍼라제(GST), 키틴 결합 단백질(CBP), 말토스 결합 단백질, 티오레독신(TRX), 폴리(NANP), 탠덤 친화성 정제(TAP)

태그, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, 혈구응집소(HA), nus, 소프트태그(Softag) 1, 소프트태그 3, 스트렙(Strep), SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 히스티딘(His), 비오틴 카르복실 캐리어 단백질(BCCP), 및 칼모둘린을 포함한다.

[0086] Cas 단백질은 임의의 형태로 제공될 수 있다. 예를 들어, Cas 단백질은 단백질의 형태, 예컨대 gRNA와 복합체화된 Cas 단백질의 형태로 제공될 수 있다. 대안적으로, Cas 단백질은 Cas 단백질을 인코딩하는 핵산의 형태, 예컨대 RNA(예를 들어, 메신저 RNA(mRNA)) 또는 DNA로 제공될 수 있다. 선택적으로, Cas 단백질을 인코딩하는 핵산은 특정 세포 또는 유기체에서 단백질로의 효율적인 번역에 최적화된 코돈일 수 있다. 예를 들어, Cas 단백질을 인코딩하는 핵산은 자연적으로 발생한 폴리뉴클레오타이드 서열과 비교하여, 세균 세포, 효모 세포, 인간 세포, 비인간 세포, 포유류 세포, 설치류 세포, 마우스 세포, 래트 세포, 또는 임의의 다른 관심 숙주 세포에서 더 높은 사용 빈도를 갖는 코돈으로 치환되도록 변형될 수 있다. Cas 단백질을 인코딩하는 핵산이 세포 내로 도입되는 경우, Cas 단백질은 세포 내에서 일시적으로, 조건적으로 또는 구성적으로 발현될 수 있다.

[0087] Cas 단백질을 인코딩하는 핵산은 세포의 게놈에 안정적으로 통합될 수 있고 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 대안적으로, Cas 단백질을 인코딩하는 핵산은 발현 작제물에서 프로모터에 작동가능하게 연결될 수 있다. 발현 작제물은, 관심 유전자 또는 다른 핵산 서열(예를 들어, Cas 유전자)의 발현을 유도할 수 있고 그런 관심 핵산 서열을 표적 세포로 전달할 수 있는 임의의 핵산 작제물을 포함한다. 예를 들어, Cas 단백질을 인코딩하는 핵산은 핵산 삽입물을 포함하는 표적화 벡터 및/또는 gRNA를 인코딩하는 DNA를 포함하는 벡터 내에 있을 수 있다. 대안적으로, 그것은 핵산 삽입물을 포함하는 표적화 벡터와는 별개의 그리고/또는 gRNA를 인코딩하는 DNA를 포함하는 벡터와는 별개의 벡터 또는 플라스미드 내에 있을 수 있다. 발현 작제물에 사용될 수 있는 프로모터는, 예를 들어, 다능성 래트 세포, 진행 세포, 포유류 세포, 비인간 포유류 세포, 인간 세포, 설치류 세포, 마우스 세포, 또는 햄스터 세포에서 활성인 프로모터를 포함한다. 1-세포기 배아에서 활성인 프로모터가 또한 사용될 수 있다. 그러한 프로모터는, 예를 들어, 조건적 프로모터, 유도성 프로모터, 구성적 프로모터, 또는 조직-특이적 프로모터일 수 있다. 다른 프로모터의 예가 본 명세서의 어딘가 다른 곳에 설명되어 있다.

[0088] B. 가이드 RNA(gRNA)

[0089] "가이드 RNA" 또는 "gRNA"는, Cas 단백질에 결합하고 표적 DNA 내의 특이적 위치로 Cas 단백질을 표적화하는 RNA 분자를 포함한다. 가이드 RNA는 다음 2개의 세그먼트를 포함할 수 있다: "DNA-표적화 세그먼트" 및 "단백질-결합 세그먼트". "세그먼트"는 분자의 세그먼트, 부분(section), 또는 영역, 예컨대 RNA에서 뉴클레오타이드들의 인접 스트레치를 포함한다. 일부 gRNA는 다음 2개의 별개의 RNA 분자들을 포함한다: "활성화인자-RNA" 및 "표적자(targeter)-RNA". 다른 gRNA는 단일 RNA 분자(단일 RNA 폴리뉴클레오타이드)이고, 이는 "단일-분자 gRNA", "단일-가이드 RNA", 또는 "sgRNA"로도 지칭된다. 예를 들어, 국제특허 공개 WO/2013/176772A1호, WO/2014/065596A1호, WO/2014/089290A1호, WO/2014/093622A2호, WO/2014/099750A2호, WO/2013/142578A1호, 및 WO 2014/131833A1호를 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 용어 "가이드 RNA" 및 "gRNA"는 포괄적이어서, 이중-분자 gRNA 및 단일-분자 gRNA 둘 모두를 포함한다.

[0090] 예시적인 2-분자 gRNA는 crRNA-유사("CRISPR RNA" 또는 "표적자-RNA" 또는 "crRNA" 또는 "crRNA 반복부") 분자 및 상응하는 tracrRNA-유사("트랜스-활성화 CRISPR RNA" 또는 "활성화인자-RNA" 또는 "tracrRNA" 또는 "스캐폴드") 분자를 포함한다. crRNA는 gRNA의 DNA-표적화 세그먼트(단일-가닥으로 됨) 및 gRNA의 단백질-결합 세그먼트의 dsRNA 듀플렉스(duplex)의 반쪽을 형성하는 뉴클레오타이드들의 스트레치 둘 모두를 포함한다.

[0091] 상응하는 tracrRNA(활성화인자-RNA)는 gRNA의 단백질-결합 세그먼트의 dsRNA 듀플렉스의 반쪽을 형성하는 뉴클레오타이드들의 스트레치를 포함한다. crRNA의 뉴클레오타이드들의 스트레치는 tracrRNA의 뉴클레오타이드의 스트레치에 상보적이고 이와 혼성화되어 gRNA의 단백질-결합 도메인의 dsRNA 듀플렉스를 형성한다. 그렇기 때문에, 각각의 crRNA는 상응하는 tracrRNA를 갖고 있다고 할 수 있다. tracrRNA는 임의의 형태(예를 들어, 전장(full-length) tracrRNA 또는 활성의 부분 tracrRNA)일 수 있고 길이가 다양할 수 있다. tracrRNA의 형태는 일차(primary) 전사체 또는 가공된 형태를 포함할 수 있다. 예를 들어, S. 피오게네스에서, tracrRNA의 상이한 형태는 171-뉴클레오타이드, 89-뉴클레오타이드, 75-뉴클레오타이드, 및 65-뉴클레오타이드 버전을 포함한다. 예를 들어, 문헌[Deltcheva *et al.* (2011) *Nature* 471:602-607] 및 국제특허 공개 WO 2014/093661호를 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0092] crRNA와 상응하는 tracrRNA는 혼성화되어 gRNA를 형성한다. crRNA는 추가적으로 CRISPR RNA 인식 서열에 혼성화되는 단일-가닥 DNA-표적화 세그먼트를 제공한다. 세포 내에서의 변형에 사용된다면, 주어진 crRNA 또는

tracrRNA 분자의 정확한 서열은 RNA 분자가 사용될 중에 특이적이 되도록 설계될 수 있다. 예를 들어, 문헌[Mali *et al.* (2013) *Science* 339:823-826]; 문헌[Jinek *et al.* (2012) *Science* 337:816-821]; 문헌[Hwang *et al.* (2013) *Nat. Biotechnol.* 31:227-229]; 문헌[Jiang *et al.* (2013) *Nat. Biotechnol.* 31:233-239]; 및 문헌[Cong *et al.* (2013) *Science* 339:819-823]을 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0093] 주어진 gRNA의 DNA-표적화 세그먼트(crRNA)는 표적 DNA에서의 서열에 상보적인 뉴클레오티드 서열을 포함한다. gRNA의 DNA-표적화 세그먼트는 혼성화(즉, 염기쌍 형성)을 통해 서열-특이적 방식으로 표적 DNA와 상호작용한다. 그렇기 때문에, DNA-표적화 세그먼트의 뉴클레오티드 서열은 다양할 수 있고, gRNA 및 표적 DNA와 상호작용할 표적 DNA 내의 위치를 결정한다. 대상 gRNA의 DNA-표적화 세그먼트는 표적 DNA 내의 임의의 원하는 서열에 혼성화되도록 변형될 수 있다. 자연적으로 발생한 crRNA는 Cas9 시스템 및 유기체에 따라 다르지만, 21 내지 46개 뉴클레오티드 길이의 2개의 직접 반복부(DR)에 의해 플랭킹된 21 내지 72개 뉴클레오티드 길이의 표적화 세그먼트를 종종 함유한다(예를 들어, 국제특허 공개 WO2014/131833호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함함). S. 피로게네스의 경우, DR은 36개 뉴클레오티드 길이이고 표적화 세그먼트는 30개 뉴클레오티드 길이이다. 3'에 위치한 DR은 상응하는 tracrRNA에 상보적이고 이와 혼성화되며, 이는 다시, Cas9 단백질에 결합한다.

[0094] DNA-표적화 세그먼트의 길이는 약 12개 뉴클레오티드 내지 약 100개 뉴클레오티드일 수 있다. 예를 들어, DNA-표적화 세그먼트의 길이는 약 12개 뉴클레오티드(nt) 내지 약 80 nt, 약 12 nt 내지 약 50 nt, 약 12 nt 내지 약 40 nt, 약 12 nt 내지 약 30 nt, 약 12 nt 내지 약 25 nt, 약 12 nt 내지 약 20 nt, 또는 약 12 nt 내지 약 19 nt일 수 있다. 대안적으로, DNA-표적화 세그먼트의 길이는 약 19 nt 내지 약 20 nt, 약 19 nt 내지 약 25 nt, 약 19 nt 내지 약 30 nt, 약 19 nt 내지 약 35 nt, 약 19 nt 내지 약 40 nt, 약 19 nt 내지 약 45 nt, 약 19 nt 내지 약 50 nt, 약 19 nt 내지 약 60 nt, 약 19 nt 내지 약 70 nt, 약 19 nt 내지 약 80 nt, 약 19 nt 내지 약 90 nt, 약 19 nt 내지 약 100 nt, 약 20 nt 내지 약 25 nt, 약 20 nt 내지 약 30 nt, 약 20 nt 내지 약 35 nt, 약 20 nt 내지 약 40 nt, 약 20 nt 내지 약 45 nt, 약 20 nt 내지 약 50 nt, 약 20 nt 내지 약 60 nt, 약 20 nt 내지 약 70 nt, 약 20 nt 내지 약 80 nt, 약 20 nt 내지 약 90 nt, 또는 약 20 nt 내지 약 100 nt일 수 있다.

[0095] 표적 DNA의 뉴클레오티드 서열(CRISPR RNA 인식 서열)에 상보적인 DNA-표적화 세그먼트의 뉴클레오티드 서열의 길이는 적어도 약 12 nt일 수 있다. 예를 들어, DNA-표적화 서열(즉, 표적 DNA 내의 CRISPR RNA 인식 서열에 상보적인, DNA-표적화 세그먼트 내의 서열)의 길이는 적어도 약 12 nt, 적어도 약 15 nt, 적어도 약 18 nt, 적어도 약 19 nt, 적어도 약 20 nt, 적어도 약 25 nt, 적어도 약 30 nt, 적어도 약 35 nt, 또는 적어도 약 40 nt일 수 있다. 대안적으로, DNA-표적화 서열의 길이는 약 12개 뉴클레오티드(nt) 내지 약 80 nt, 약 12 nt 내지 약 50 nt, 약 12 nt 내지 약 45 nt, 약 12 nt 내지 약 40 nt, 약 12 nt 내지 약 35 nt, 약 12 nt 내지 약 30 nt, 약 12 nt 내지 약 25 nt, 약 12 nt 내지 약 20 nt, 약 12 nt 내지 약 19 nt, 약 19 nt 내지 약 20 nt, 약 19 nt 내지 약 25 nt, 약 19 nt 내지 약 30 nt, 약 19 nt 내지 약 35 nt, 약 19 nt 내지 약 40 nt, 약 19 nt 내지 약 45 nt, 약 19 nt 내지 약 50 nt, 약 19 nt 내지 약 60 nt, 약 20 nt 내지 약 25 nt, 약 20 nt 내지 약 30 nt, 약 20 nt 내지 약 35 nt, 약 20 nt 내지 약 40 nt, 약 20 nt 내지 약 45 nt, 약 20 nt 내지 약 50 nt, 또는 약 20 nt 내지 약 60 nt일 수 있다. 일부 경우에, DNA-표적화 서열의 길이는 약 20 nt일 수 있다.

[0096] tracrRNA는 임의의 형태 (예를 들어, 전장 tracrRNA 또는 활성의 부분 tracrRNA)일 수 있고 길이가 다양할 수 있다. 이는 일차 전사체 또는 가공된 형태를 포함할 수 있다. 예를 들어, (단일-가이드 RNA의 일부로서의 또는 2-분자 gRNA의 일부로서의 별개 분자로서의) tracrRNA는 야생형 tracrRNA 서열의 전부 또는 일부(예를 들어, 야생형 tracrRNA 서열의 대략 하기 개수의 또는 하기 개수를 초과하는 뉴클레오티드: 약 20, 26, 32, 45, 48, 54, 63, 67, 85개, 또는 그 이상)를 포함하거나 이로 이루어질 수 있다. S. 피로게네스 유래의 야생형 tracrRNA 서열의 예는 171-뉴클레오티드, 89-뉴클레오티드, 75-뉴클레오티드, 및 65-뉴클레오티드 버전을 포함한다. 예를 들어, 문헌[Deltcheva *et al.* (2011) *Nature* 471:602-607]; 국제특허 공개 WO 2014/093661호를 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 단일-가이드 RNA(sgRNA) 내의 tracrRNA의 예는 sgRNA의 +48, +54, +67, 및 +85 버전 내에서 발견되는 tracrRNA 세그먼트를 포함하고, 여기서 "+n"은 야생형 tracrRNA의 최대 +n 뉴클레오티드가 sgRNA 내에 포함된다는 것을 나타낸다. 미국 특허 제8,697,359호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0097] DNA-표적화 서열과 표적 DNA 내의 CRISPR RNA 인식 서열 사이의 %상보성은 적어도 60%(예를 들어, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 적어도 98%, 적어도

99%, 또는 100%)일 수 있다. 일부 경우에, DNA-표적화 서열과 표적 DNA 내의 CRISPR RNA 인식 서열 사이의 %상보성은 약 20개의 인접 뉴클레오타이드들에 걸쳐 적어도 60%이다. 일례로서, DNA-표적화 서열과 표적 DNA 내의 CRISPR RNA 인식 서열 사이의 %상보성은 표적 DNA의 상보적 가닥 내의 CRISPR RNA 인식 서열의 5' 말단에서 14개의 인접 뉴클레오타이드에 걸쳐 100%이고 나머지에 걸쳐서는 0%만큼 낮다. 그러한 경우에, DNA-표적화 서열은 14개 뉴클레오타이드 길이인 것으로 간주될 수 있다. 다른 예에서, DNA-표적화 서열과 표적 DNA 내의 CRISPR RNA 인식 서열 사이의 %상보성은 표적 DNA의 상보적 가닥 내의 CRISPR RNA 인식 서열의 5' 말단에서 7개의 인접 뉴클레오타이드에 걸쳐 100%이고 나머지에 걸쳐서는 0%만큼 낮다. 그러한 경우에, DNA-표적화 서열은 7개 뉴클레오타이드 길이인 것으로 간주될 수 있다.

[0098] gRNA의 단백질-결합 세그먼트는 서로 상보적인 뉴클레오타이드들의 2개의 스트레치를 포함할 수 있다. 단백질-결합 세그먼트의 상보적 뉴클레오타이드는 혼성화되어 이중-가닥 RNA 듀플렉스(dsRNA)를 형성한다. 대상 gRNA의 단백질-결합 세그먼트는 Cas 단백질과 상호작용하고, gRNA는 결합된 Cas 단백질을 DNA-표적화 세그먼트를 통해 표적 DNA 내의 특정 뉴클레오타이드 서열로 유도한다.

[0099] 가이드 RNA는 추가적인 바람직한 특성(예를 들어, 변형된 또는 조절된 안정성; 세포내 표적화; 형광 표지를 사용한 추적; 단백질 또는 단백질 복합체에 대한 결합 부위 등)을 제공하는 변형 또는 서열을 포함할 수 있다. 그러한 변형의 예는, 예를 들어, 5' 캡(예를 들어, 7-메틸구아닐레이트 캡(m7G)); 3' 폴리아데닐화된 꼬리(즉, 3' 폴리(A) 꼬리); (예를 들어, 단백질 및/또는 단백질 복합체에 의한 조절된 안정성 및/또는 조절된 접근성을 허용하기 위한 리보스위치(riboswitch) 서열; 안정성 제어 서열; dsRNA 듀플렉스(즉, 헤어핀(hairpin))를 형성하는 서열; RNA를 하위세포 위치(예를 들어, 핵, 미토콘드리아, 엽록체 등)로 표적화시키는 변형 또는 서열; 추적(예를 들어, 형광 분자에 대한 직접 접합(conjugation), 형광 검출을 촉진하는 모이어티(moiety)에 대한 접합, 형광 검출을 허용하는 서열 등)을 제공하는 변형 또는 서열; 단백질(예를 들어, 전사 활성화인자, 전사 억제인자, DNA 메틸트랜스퍼라제, DNA 데메틸라제, 히스톤 아세틸트랜스퍼라제, 히스톤 데아세틸라제 등을 포함하는, DNA 상에서 작용하는 단백질)에 대한 결합 부위를 제공하는 변형 또는 서열; 및 이들의 조합을 포함한다.

[0100] gRNA는 crRNA와 tracrRNA를 인코딩하는 핵산 서열을 포함할 수 있다. 예를 들어, gRNA는 (a) 핵산 서열 5'-GUUUUAGAGCUAGAAUAGCAAGUUAUAAUAGGCUAGUCCGUUAUACAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (서열 번호 1)을 갖는 키메라(chimeric) RNA; 또는 (b) 핵산 서열 5'-GUUUUAGAGCUAGAAUAGCAAGUUAUAAUAGGCUAGUCCG-3' (서열 번호 2)을 갖는 키메라 RNA를 포함할 수 있다.

[0101] 일부 경우에, crRNA는 5'-GUUUUAGAGCUAGAAUAGCAAGUUAUAAU-3' (서열 번호 3); 5'-GUUUUAGAGCUAGAAUAGCAAGUUAUAAUAG (서열 번호 4); 또는 5'-GAGUCCGAGCAGAAGAAGUUAU-3' (서열 번호 5)을 포함한다.

[0102] 일부 경우에, tracrRNA는 5'-AAGGCUAGUCCG-3' (서열 번호 6) 또는 5'-AAGGCUAGUCCGUUAUACAACUUGAAAAAGUGGCACCGAGUCGGUGCUUUU-3' (서열 번호 7)을 포함한다.

[0103] 가이드 RNA는 임의의 형태로 제공될 수 있다. 예를 들어, gRNA는 2개의 분자(별개의 crRNA 및 tracrRNA)로서 또는 1개의 분자(sgRNA)로서 RNA의 형태로, 그리고 선택적으로는 Cas 단백질과의 복합체의 형태로 제공될 수 있다. gRNA는 또한 RNA를 인코딩하는 DNA의 형태로 제공될 수 있다. gRNA를 인코딩하는 DNA는 단일 RNA 분자(sgRNA) 또는 별개의 RNA 분자들(예를 들어, 별개의 crRNA 및 tracrRNA)을 인코딩할 수 있다. 후자의 경우, gRNA를 인코딩하는 DNA는 crRNA 및 tracrRNA를 각각 인코딩하는 별개의 DNA 분자로서 제공될 수 있다.

[0104] gRNA를 인코딩하는 DNA가 세포 내로 도입될 때, gRNA는 세포 내에서 일시적으로, 조건적으로, 또는 구성적으로 발현될 수 있다. gRNA를 인코딩하는 DNA는 세포의 게놈에 안정적으로 통합될 수 있고 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 대안적으로, gRNA를 인코딩하는 DNA는 발현 작제물에서 프로모터에 작동가능하게 연결될 수 있다. 예를 들어, gRNA를 인코딩하는 DNA는 핵산 삽입물을 포함하는 표적화 벡터 및/또는 Cas 단백질을 인코딩하는 핵산을 포함하는 벡터 내에 있을 수 있다. 대안적으로, 그것은 핵산 삽입물을 포함하는 표적화 벡터와는 별개의 그리고/또는 Cas 단백질을 인코딩하는 핵산을 포함하는 벡터와는 별개의 벡터 또는 플라스미드 내에 있을 수 있다. 그러한 발현 작제물에 사용될 수 있는 프로모터는, 예를 들어, 다능성 래트 세포, 진핵 세포, 포유류 세포, 비인간 포유류 세포, 인간 세포, 설치류 세포, 마우스 세포, 또는 햄스터 세포에서 활성인 프로모터를 포함한다. 1-세포기 배아에서 활성인 프로모터가 또한 사용될 수 있다. 그러한 프로모터는, 예를 들어, 조건적 프로모터, 유도성 프로모터, 구성적 프로모터, 또는 조직-특이적 프로모터일 수 있다. 일부 경우에, 프로모터는 RNA 폴리머라제 III 프로모터, 예컨대 인간 U6 프로모터, 래트 U6 폴리머라제 III 프로모터, 또는 마우스 U6 폴리머라제 III 프로모터이다. 다른 프로모터의 예가 본 명세서의 어딘가 다른 곳에

설명되어 있다.

[0105] 대안적으로, gRNA는 다양한 다른 방법으로 제조될 수 있다. 예를 들어, gRNA는, 예를 들어, T7 RNA 폴리머라제를 사용하여, 시험관내 전사에 의해 제조될 수 있다(예를 들어, 국제특허 공개 WO 2014/089290호 및 WO 2014/065596호를 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함됨). 가이드 RNA는 또한 화학적 합성에 의해 제조된, 합성적으로 생산된 분자일 수 있다.

C. CRISPR RNA 인식 서열

[0107] 용어 "CRISPR RNA 인식 서열"은, 결합에 대한 충분한 조건이 존재하기만 한다면, gRNA의 DNA-표적화 세그먼트가 결합할, 표적 DNA에 존재하는 핵산 서열을 포함한다. 예를 들어, CRISPR RNA 인식 서열은 가이드 RNA가 상보성을 가지도록 설계된 서열을 포함하고, 여기서 CRISPR RNA 인식 서열과 DNA 표적화 서열 사이의 혼성화는 CRISPR 복합체의 형성을 촉진한다. 완전한 상보성이 반드시 필요한 것은 아니고, 혼성화를 야기하고 CRISPR 복합체의 형성을 촉진하기에 충분한 상보성이 있기만 하면 된다. CRISPR RNA 인식 서열은 또한 Cas 단백질에 대한 절단 부위를 포함하며, 이는 아래에 더욱 상세하게 설명되어 있다. CRISPR RNA 인식 서열은, 예를 들어, 세포의 핵 또는 세포질에 또는 세포의 세포소기관, 예컨대 미토콘드리아 또는 엽록체 내에 위치될 수 있는 임의의 폴리뉴클레오티드를 포함할 수 있다.

[0108] 표적 DNA 내의 CRISPR RNA 인식 서열은 Cas 단백질 또는 gRNA에 의해 표적화될 수 있다(즉, 이에 의해 결합되거나, 이와 혼성화되거나, 이에 상보적일 수 있다). 적합한 DNA/RNA 결합 조건은 세포에 정상적으로 존재하는 생리적 조건을 포함한다. 다른 적합한 DNA/RNA 결합 조건(예를 들어, 무세포계에서의 조건)이 당업계에 알려져 있다(예를 들어, 문헌[Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook *et al.*, Harbor Laboratory Press 2001)]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함됨). Cas 단백질 또는 gRNA에 상보적이고 이와 혼성화되는 표적 DNA의 가닥은 "상보적 가닥"으로 불릴 수 있고, "상보적 가닥"에 상보적인(따라서 Cas 단백질 또는 gRNA에 상보적이지 않은) 표적 DNA의 가닥은 "비상보적 가닥" 또는 "주형 가닥"으로 불릴 수 있다.

[0109] Cas 단백질은 gRNA의 DNA-표적화 세그먼트가 결합할 표적 DNA에 존재하는 핵산 서열 내의 또는 그 외부의 부위에서 핵산을 절단할 수 있다. "절단 부위"는 Cas 단백질이 단일-가닥 브레이크 또는 이중-가닥 브레이크를 생성하는 핵산의 위치를 포함한다. 예를 들어, (CRISPR RNA 인식 서열에 혼성화되고 Cas 단백질과 복합체화된 gRNA를 포함하는) CRISPR 복합체의 형성은 gRNA의 DNA-표적화 세그먼트가 결합할 표적 DNA에 존재하는 핵산 서열 내에서 또는 그 부근에서 (예를 들어, 이로부터 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50개, 또는 그 이상의 염기쌍 이내에서) 한쪽 또는 양쪽 가닥의 절단을 야기할 수 있다. 절단 부위가 gRNA의 DNA-표적화 세그먼트가 결합할 핵산 서열의 외부에 있다면, 절단 부위는 "CRISPR RNA 인식 서열" 내에 있는 것으로 여전히 간주된다. 절단 부위는 핵산의 단지 한쪽 가닥 상에만 있거나 양쪽 가닥 상에 있을 수 있다. 절단 부위는 핵산의 양쪽 가닥 상의 동일한 위치에 있을 수 있거나(평행 말단을 생성), 또는 각 가닥 상의 상이한 부위에 있을 수 있다(엇갈린 말단을 생성). 엇갈린 말단은, 예를 들어, 2개의 Cas 단백질을 사용하여 생성될 수 있고, 이들 각각은 각 가닥 상의 상이한 절단 부위에서 단일-가닥 브레이크 생성하여 이중-가닥 브레이크를 생성한다. 예를 들어, 제1 Nickase는 이중-가닥 DNA(dsDNA)의 제1 가닥 상에 단일-가닥 브레이크를 만들 수 있고, 제2 Nickase는 dsDNA의 제2 가닥 상에 단일-가닥 브레이크를 만들 수 있어서 돌출된(overhanging) 서열이 만들어지게 된다. 일부 경우에, 제1 가닥 상의 Nickase의 CRISPR RNA 인식 서열은 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500, 또는 1,000개의 염기쌍에 의해 제2 가닥 상의 Nickase의 CRISPR RNA 인식 서열로부터 분리된다.

[0110] Cas9에 의한 표적 DNA의 부위-특이적 절단은 표적 DNA에서 (i) gRNA와 표적 DNA 사이의 염기쌍 형성 상보성 및 (ii) 프로토스페이서 인접 모티프(PAM)로 불리는, 짧은 모티프 둘 모두에 의해 결정되는 위치에서 일어날 수 있다. PAM은 CRISPR RNA 인식 서열을 플랭킹할 수 있다. 선택적으로, CRISPR RNA 인식 서열은 3' 말단 상에서 PAM에 의해 플랭킹될 수 있다. 예를 들어, Cas9의 절단 부위는 PAM 서열의 상류 또는 하류의 약 1 내지 약 10 개 또는 약 2 내지 약 5개의 염기쌍(예를 들어, 3개의 염기쌍)일 수 있다. 일부 경우에(예를 들어, S. 피오케네스로부터의 Cas9 또는 근연(closely related) Cas9가 사용될 때), 비-상보적 가닥의 PAM 서열은 5'-N₁GG-3'일 수 있고, 여기서 N₁은 임의의 DNA 뉴클레오티드이고 표적 DNA의 비-상보적 가닥의 CRISPR RNA 인식 서열의 3'에 바로 있다. 그렇기 때문에, 상보적 가닥의 PAM 서열은 5'-CCN₂-3'일 것이고, 여기서 N₂는 임의의 DNA 뉴클레오티드이고 표적 DNA의 상보적 가닥의 CRISPR RNA 인식 서열의 5'에 바로 있다. 일부 그러한 경우에, N₁과 N₂는

상보적일 수 있고 N_1 - N_2 염기쌍은 임의의 염기쌍(예를 들어, $N_1=C$ 및 $N_2=G$; $N_1=G$ 및 $N_2=C$; $N_1=A$ 및 $N_2=T$; 또는 $N_1=T$, 및 $N_2=A$)일 수 있다.

[0111] CRISPR RNA 인식 서열의 예는 grRNA의 DNA-표적화 세그먼트에 상보적인 DNA 서열, 또는 PAM 서열에 덧붙여진 그러한 DNA 서열을 포함한다. 예를 들어, 표적 모티프는 Cas 단백질에 의해 인식되는 NGG 모티프 직전의 20-뉴클레오타이드 DNA 서열, 예컨대 $GN_{19}NGG$ (서열 번호 8) 또는 $N_{20}NGG$ (서열 번호 9)일 수 있다(예를 들어, 국제특허 공개 WO 2014/165825호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함됨). 5' 말단에서의 구아닌은 세포에서 RNA 폴리머라제에 의한 전사를 촉진할 수 있다. CRISPR RNA 인식 서열의 다른 예는 시험관내에서의 T7 폴리머라제에 의한 효율적인 전사를 용이하게 하기 위해 5' 말단에 2개의 구아닌 뉴클레오타이드(예를 들어, $GGN_{20}NGG$; 서열 번호 10)를 포함할 수 있다. 예를 들어, 국제특허 공개 WO 2014/065596호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 다른 CRISPR RNA 인식 서열은 5' G 또는 GG 및 3' GG 또는 NGG를 포함하는 서열 번호 8 내지 10의 길이가 4 내지 22개인 뉴클레오타이드를 가질 수 있다. 또 다른 CRISPR RNA 인식 서열은 서열 번호 8 내지 10의 길이가 14 내지 22개인 뉴클레오타이드를 가질 수 있다. CRISPR RNA 인식 서열의 구체적인 예는 서열 번호 11 내지 38 중 어느 하나를 포함하는 핵산에 상보적인 DNA 서열을 포함한다.

[0112] CRISPR RNA 인식 서열은 세포에 내인성 또는 외인성인 임의의 핵산 서열일 수 있다. CRISPR RNA 인식 서열은 유전자 산물(예를 들어, 단백질)을 코딩하는 서열 또는 비-코딩 서열(예를 들어, 조절 서열)일 수 있거나 둘 모두를 포함할 수 있다. 일부 경우에, CRISPR RNA 인식 서열은 질병-관련 유전자 또는 핵산 내에 그리고/또는 신호전달 경로-관련 유전자 또는 핵산 내에 있을 수 있다. 질병-관련 유전자 또는 핵산은 비질병 대조군의 조직 또는 세포와 비교하여 질병-영향(disease-affected) 조직으로부터 유래된 세포에서 비정상 수준으로 또는 비정상 형태로 전사 또는 번역 산물을 산출하는 임의의 유전자 또는 핵산을 포함한다. 예를 들어, 질병-관련 유전자는 질병의 병인의 직접적인 원인이 되거나 질병의 병인의 원인이 되는 하나 이상의 유전자와 연관 불균형(linkage disequilibrium) 상태에 있는 하나 이상의 돌연변이 또는 유전자 변이를 가질 수 있다. 전사된 또는 번역된 산물은 알려져 있거나 알려져 있지 않을 수 있고, 정상 또는 비정상 수준으로 존재할 수 있다. 질병-관련 유전자 및 핵산의 예는 존스 홉킨스 대학(Johns Hopkins University)의 맥쿠식-네이션스 유전의학연구소(McKusick-Nathans Institute of Genetic Medicine)(미국 메릴랜드주 발티모어 소재) 및 미국 국립의학도서관(National Library of Medicine)의 미국 국립생물공학정보센터(National Center for Biotechnology Information)(미국 메릴랜드주 베데스다 소재)로부터 입수가능하며, 이들은 월드 와이드 웹 상에서 이용가능하다. 질병-관련 유전자 및 핵산의 추가의 예에 대해서는, 미국 특허 제8,697,359호를 참조하며, 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0113] 질병-유발 유전자에서의 돌연변이는 열성 돌연변이 또는 우성 돌연변이일 수 있다. 이배체 유기체(즉, 각각의 염색체의 2개의 카피를 갖는 유기체)는 전형적으로 각각의 유전자의 2개의 카피를 보유한다. 개체에서의 2개의 카피가 동일하다면, 개체는 유전자에 대해 동형접합성이다. 이들 카피가 상이한 대립유전자라면, 개체는 유전자에 대해 이형접합성이다. 용어 유전자형(genotype)은 개체가 단일 유전자(또는 유전자들) 내에 돌연변이를 보유하는지의 여부를 포함하고, 용어 표현형(phenotype)은 그러한 유전자형의 물리적 및 기능적 결과를 포함한다. 열성 돌연변이는 양쪽 대립 유전자가, 돌연변이 표현형이 관찰되기 위하여 돌연변이되어야 하는 돌연변이를 포함한다(즉, 유기체는 돌연변이 표현형을 나타내기 위해서 돌연변이 대립유전자에 대해 동형접합성이어야 함). 열성 돌연변이는, 예를 들어, 영향을 받은 유전자를 불활성화시키고 기능의 상실을 초래할 수 있다. 예를 들어, 열성 돌연변이는 염색체로부터 유전자의 전부 또는 일부를 제거하거나, 유전자의 발현을 파괴하거나, 인코딩된 단백질의 구조를 변경시켜 그의 기능을 변경시킬 수 있다. 대조적으로, 우성 돌연변이는 돌연변이 표현형이 돌연변이에 대해 이형접합성인 유기체에서 관찰되는 돌연변이를 포함한다(즉, 유기체는 하나의 돌연변이 대립유전자와 하나의 야생형 대립유전자를 보유함). 우성 돌연변이는, 예를 들어 기능의 획득으로 이어질 수 있다. 예를 들어, 우성 돌연변이는 주어진 유전자 산물의 활성을 증가시키거나, 유전자 산물에 대해 새로운 활성을 부여하거나, 또는 그의 부적절한 공간적 및 시간적 발현으로 이어질 수 있다. 우성 돌연변이는 또한 기능의 상실과 관련될 수 있다. 일부 경우에, 유전자의 2개의 카피가 정상 기능에 필요한 경우, 단일 카피를 제거하는 것은 돌연변이 표현형을 야기할 수 있다. 그러한 유전자는 반수체 기능부전(haplo-insufficient)을 나타낸다. 다른 경우에, 한쪽 대립유전자 내의 돌연변이는 다른 한쪽 대립유전자에 의해 인코딩되는 야생형 단백질의 기능을 방해하는 단백질의 구조적 변화로 이어질 수 있다. 그러한 돌연변이는 우성 음성 돌연변이이다. 일부 대립유전자는 열성 표현형 및 우성 표현형 둘 모두와 관련될 수 있다.

- [0114] 일부 CRISPR RNA 인식 서열은 돌연변이를 포함하는 유전자 또는 핵산 내에 있다. 돌연변이는, 예를 들어 우성 돌연변이 또는 열성 돌연변이일 수 있다. 일부 경우에, 우성 돌연변이는 우성 돌연변이에 대해 이형접합성인 세포 내에 있다(즉, 세포는 야생형 대립유전자와, 우성 돌연변이를 포함하는 돌연변이 대립유전자를 포함함). 일부 그러한 경우에, CRISPR RNA 인식 서열은, 야생형 대립유전자 내는 아니고 돌연변이 대립유전자 내에 있을 수 있다. 대안적으로, CRISPR RNA 인식 서열은, 돌연변이 대립유전자 내는 아니고 야생형 대립유전자 내에 있을 수 있다.
- [0115] *III. 표적화 벡터 및 핵산 삽입물*
- [0116] 본 명세서에 개시된 방법 및 조성물은 또한 세포 내의 게놈을 변형시키기 위해 핵산 삽입물 및 상동성 아암들을 포함하는 표적화 벡터를 이용할 수 있다. 그러한 방법에서, 핵산 삽입물은 상동 재조합 이벤트를 통해 상동성 아암들에 의해 결정된 게놈 표적 유전자좌 내로 통합된다. 본 명세서에서 제공된 방법은 상동 재조합 이벤트와 조합하여 뉴클레아제 작용제(nuclease agent)(예를 들어, Cas 단백질)를 이용할 수 있다. 그러한 방법은 상동 재조합과 조합하여 뉴클레아제 절단 부위에서 뉴클레아제 작용제에 의해 생성된 닉(nick) 또는 이중-가닥 브레이크를 사용하여 게놈 표적 유전자좌 내로의 핵산 삽입물의 표적화된 통합을 촉진한다.
- [0117] **A. 1-세포기 배아 이외의 세포에 대한 표적화 벡터 및 핵산 삽입물**
- [0118] **(1) 핵산 삽입물**
- [0119] 하나 이상의 별개의 핵산 삽입물이 본 명세서에 개시된 방법에서 사용될 수 있고, 이들은 별개의 표적화 벡터들을 통해 또는 동일한 표적화 벡터 상에서 세포 내로 도입될 수 있다. 핵산 삽입물은 게놈 표적 유전자좌에서 통합시키고자 하는 DNA의 세그먼트들을 포함한다. 표적 유전자좌에서의 핵산 삽입물의 통합은 표적 유전자좌에 대한 관심 핵산 서열의 부가, 표적 유전자좌에서의 관심 핵산 서열의 결실, 및/또는 표적 유전자좌에서의 관심 핵산 서열의 대체(즉, 결실 및 삽입)를 가져올 수 있다.
- [0120] 핵산 삽입물 또는 대체되는 표적 유전자좌에서의 상응하는 핵산은 코딩 영역, 인트론, 엑손, 비번역 영역, 조절 영역, 프로모터, 인핸서, 또는 이들의 임의의 조합일 수 있다. 더욱이, 핵산 삽입물 또는 대체되는 표적 유전자좌에서의 상응하는 핵산은 임의의 원하는 길이를 가질 수 있으며, 이는, 예를 들어 10 내지 100개 뉴클레오티드 길이, 100 내지 500개 뉴클레오티드 길이, 500개 뉴클레오티드 내지 1 kb 길이, 1 kb 내지 1.5 kb 뉴클레오티드 길이, 1.5 kb 내지 2 kb 뉴클레오티드 길이, 2 kb 내지 2.5 kb 뉴클레오티드 길이, 2.5 kb 내지 3 kb 뉴클레오티드 길이, 3 kb 내지 5 kb 뉴클레오티드 길이, 5 kb 내지 8 kb 뉴클레오티드 길이, 8 kb 내지 10 kb 뉴클레오티드 길이 또는 그 이상을 포함한다. 다른 경우에, 길이는 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 250 kb, 약 250 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 약 350 kb 내지 약 400 kb, 약 400 kb 내지 약 800 kb, 약 800 kb 내지 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb, 내지 약 2.5 Mb, 약 2.5 Mb 내지 약 2.8 Mb, 약 2.8 Mb 내지 약 3 Mb일 수 있다. 또 다른 경우에, 길이는 적어도 100, 200, 300, 400, 500, 600, 700, 800, 또는 900개 뉴클레오티드 또는 적어도 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb, 또는 그 이상일 수 있다. 일부 핵산 삽입물은 심지어 더 작을 수 있다. 일례로서, 약 4개 뉴클레오티드 내지 약 12개 뉴클레오티드 길이의 삽입물이 제한 효소 부위를 생성하기 위해 삽입될 수 있다.
- [0121] 일부 표적화 벡터에서, 핵산 삽입물은 약 5 kb 내지 약 200 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 60 kb 내지 약 70 kb, 약 80 kb 내지 약 90 kb, 약 90 kb 내지 약 100 kb, 약 100 kb 내지 약 110 kb, 약 120 kb 내지 약 130 kb, 약 130 kb 내지 약 140 kb, 약 140 kb 내지 약 150 kb, 약 150 kb 내지 약 160 kb, 약 160 kb 내지 약 170 kb, 약 170 kb 내지 약 180 kb, 약 180 kb 내지 약 190 kb, 약 190 kb 내지 약 200 kb일 수 있다. 대안적으로, 핵산 삽입물은 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 250 kb, 약 250 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 또는 약 350 kb 내지 약 400 kb일 수 있다.
- [0122] 일부 경우에, 표적 유전자좌에서의 핵산의 대체는 약 1 kb 내지 약 200 kb, 약 2 kb 내지 약 20 kb, 또는 약 0.5 kb 내지 약 3 Mb 범위의 핵산 서열의 결실을 가져온다. 일부 경우에, 결실의 정도는 5' 상동성 아암과 3'

상동성 아암의 총 길이보다 더 크다.

- [0123] 일부 경우에, 핵산 서열의 결실의 정도는 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 약 90 kb 내지 약 100 kb, 약 100 kb 내지 약 110 kb, 약 110 kb 내지 약 120 kb, 약 120 kb 내지 약 130 kb, 약 130 kb 내지 약 140 kb, 약 140 kb 내지 약 150 kb, 약 150 kb 내지 약 160 kb, 약 160 kb 내지 약 170 kb, 약 170 kb 내지 약 180 kb, 약 180 kb 내지 약 190 kb, 약 190 kb 내지 약 200 kb, 약 200 kb 내지 약 250 kb, 약 250 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 약 350 kb 내지 약 400 kb, 약 400 kb 내지 약 800 kb, 약 800 kb 내지 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 약 2.5 Mb 내지 약 2.8 Mb, 약 2.8 Mb 내지 약 3 Mb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb의 범위이다.
- [0124] 다른 경우에, 핵산 삽입물 또는 대체되는 표적 유전자좌에서의 상응하는 핵산은 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 150 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb 또는 그 이상일 수 있다.
- [0125] 핵산 삽입물은 게놈 DNA 또는 임의의 다른 유형의 DNA를 포함할 수 있다. 예를 들어, 핵산 삽입물은 원핵생물, 진핵생물, 효모, 조류(예를 들어, 닭), 비인간 포유류, 설치류, 인간, 래트, 마우스, 햄스터, 토끼, 돼지, 소, 사슴, 양, 염소, 고양이, 개, 페렛, 영장류(예를 들어, 마모셋, 붉은털원숭이), 사육된 포유류, 농업용 포유류, 또는 임의의 다른 관심 유기체로부터 유래될 수 있다.
- [0126] 핵산 삽입물 및/또는 표적 유전자좌에서의 핵산은 코딩 서열 또는 비-코딩 서열, 예컨대 조절 요소(예를 들어, 프로모터, 인핸서, 또는 전사 억제인자-결합 요소)를 포함할 수 있다. 예를 들어, 핵산 삽입물은 내인성 유전자의 적어도 하나의 엑손의 녹-인(knock-in) 대립유전자, 또는 전체 내인성 유전자의 녹-인 대립유전자(즉, "유전자-스왑(swap) 녹-인")를 포함할 수 있다.
- [0127] 예를 들어, 핵산 삽입물은 게놈 표적 유전자좌에서의 결실에 대해 표적화되는 서열에 대해 상동성 또는 오르토로그성일 수 있다. 상동성 또는 오르토로그성 핵산 삽입물은 관심 게놈 유전자좌에서의 결실에 대해 표적화되는 서열을 대체할 수 있다. 상동성 서열은 공지된 참조 서열과 동일하거나 실질적으로 유사한 핵산 서열을 포함하여, 그것은 공지된 참조 서열과 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100% 동일하도록 된다. 오르토로그성 서열은 다른 종의 공지된 참조 서열과 기능적으로 동등한 하나의 종으로부터 유래된 핵산 서열을 포함한다. 핵산 삽입물의 삽입이 비인간 핵산 서열의 상동성 또는 오르토로그성 인간 핵산 서열로의 대체를 가져온다면, 이는 유전자좌의 인간화를 가져올 수 있다(즉, 핵산 삽입물이 상응하는 비인간 DNA 서열 대신 그의 내인성 게놈 유전자좌에서 삽입됨).
- [0128] 핵산 삽입물은 또한 조건적 대립유전자(conditional allele)를 포함할 수 있다. 조건적 대립유전자는 미국 특허 출원 공개 제2011/0104799호에 기재된 바와 같은 다기능성 대립유전자일 수 있으며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 예를 들어, 조건적 대립유전자는 하기를 포함할 수 있다: (a) 표적 유전자의 전사에 대해 센스 배향의 구동(actuating) 서열; (b) 센스 또는 안티센스 배향의 약물 선택 카세트(DSC); (c) 안티센스 배향의 관심 뉴클레오티드 서열(nucleotide sequence of interest)(NSI); 및 (d) 역배향의 조건적 역위 모듈(conditional by inversion module)(COIN, 이는 엑손-분할 인트론 및 가역 유전자-트랩-유사 모듈을 이용함). 예를 들어, 미국 특허 출원 공개 제2011/0104799호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 조건적 대립유전자는 제1 재조합효소에 노출 시에 재조합되어, (i) 구동 서열 및 DSC가 결여되어 있고; (ii) 센스 배향의 NSI 및 안티센스 배향의 COIN을 함유하는 조건적 대립유전자를 형성하는 재조합가능 단위를 추가로 포함할 수 있다. 미국 특허 출원 공개 제2011/0104799호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.
- [0129] 일부 핵산 삽입물은 선택 마커를 인코딩하는 폴리뉴클레오티드를 포함한다. 선택 마커는 선택 카세트 내에 함유될 수 있다. 그러한 선택 마커는 네오마이신 포스포트랜스퍼라제(neo^r), 하이그로마이신 B 포스포트랜스퍼라

제(hyg^r), 퓨로마이신-N-아세틸트랜스퍼라제(puro^r), 블라스티시딘 S 데아미나제(bsr^r), 잔틴/구아닌 포스포리보실 트랜스퍼라제(gpt), 또는 단순 포진 바이러스 티미딘 키나제(HSV-k), 또는 이들의 조합을 포함하지만, 이로 제한되지는 않는다. 선택 마커를 인코딩하는 폴리뉴클레오티드는 표적화되는 세포에서 활성인 프로모터에 작동 가능하게 연결될 수 있다. 프로모터의 예는 본 명세서의 어딘가 다른 곳에 기재되어 있다.

[0130] 일부 표적화 벡터에서, 핵산 삽입물은 리포터 유전자를 포함한다. 리포터 유전자의 예는 루시페라제, β -갈락토시다제, 녹색 형광 단백질(GFP), 향상된 녹색 형광 단백질(eGFP), 시안 형광 단백질(CFP), 황색 형광 단백질(YFP), 향상된 황색 형광 단백질(eYFP), 청색 형광 단백질(BFP), 향상된 청색 형광 단백질(eBFP), DsRed, ZsGreen, MmGFP, mPlum, mCherry, tdTomato, mStrawberry, J-Red, mOrange, mKO, mCitrine, 비너스, YPet, 에메랄드, CyPet, Cerulean, T-사파이어, 알칼리성 포스포타제, 및 이들의 조합을 인코딩하는 유전자이다. 그러한 리포터 유전자는 표적화되는 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 프로모터의 예는 본 명세서의 어딘가 다른 곳에 기재되어 있다.

[0131] 일부 표적화 벡터에서, 핵산 삽입물은 하나 이상의 발현 카세트 또는 결실 카세트를 포함한다. 주어진 카세트는 관심 뉴클레오티드 서열, 선택 마커를 인코딩하는 핵산, 및/또는 리포터 유전자, 그리고 이와 함께, 발현에 영향을 미치는 다양한 조절 성분을 포함할 수 있다. 포함될 수 있는 선택가능 마커 및 리포터 유전자의 예는 본 명세서의 어딘가 다른 곳에 상세히 논의되어 있다.

[0132] 일부 표적화 벡터에서, 핵산 삽입물은 부위-특이적 재조합 표적 서열로 플랭킹된 핵산을 포함한다. 전체 핵산 삽입물이 그러한 부위-특이적 재조합 표적 서열에 의해 플랭킹될 수 있지만, 핵산 삽입물 내의 임의의 관심 영역 또는 개별 폴리뉴클레오티드가 또한 그러한 부위에 의해 플랭킹될 수 있다. 핵산 삽입물 또는 핵산 삽입물 내의 임의의 관심 폴리뉴클레오티드를 플랭킹할 수 있는 부위-특이적 재조합 표적 서열은, 예를 들어 loxP, lox511, lox2272, lox66, lox71, loxM2, lox5171, FRT, FRT11, FRT71, attP, att, FRT, rox, 및 이들의 조합을 포함할 수 있다. 일례에서, 부위-특이적 재조합 부위는 핵산 삽입물 내에 함유된 선택 마커 및/또는 리포터 유전자를 인코딩하는 폴리뉴클레오티드를 플랭킹한다. 표적화된 유전자좌에서의 핵산 삽입물의 통합 후에, 부위-특이적 재조합 부위들 사이의 서열은 제거될 수 있다.

[0133] (2) 표적화 벡터

[0134] 핵산 삽입물을 게놈 표적 유전자좌 내로 도입하여 핵산 삽입물 및 핵산 삽입물을 플랭킹하는 상동성 아암들을 포함시키기 위해 표적화 벡터가 사용될 수 있다. 표적화 벡터는 선형 형태 또는 환형 형태일 수 있고, 단일-가닥 또는 이중-가닥일 수 있다. 표적화 벡터는 데옥시리보핵산(DNA) 또는 리보핵산(RNA)일 수 있다. 참조 용이성을 위하여, 상동성 아암들은 본 명세서에서 5' 및 3'(즉, 상류 및 하류) 상동성 아암들을 지칭한다. 이 용어는 표적화 벡터 내에서의 핵산 삽입물에 대한 상동성 아암들의 상대적인 위치에 관한 것이다. 5' 및 3' 상동성 아암은, 본 명세서에서 각각 "5' 표적 서열" 및 "3' 표적 서열"로 지칭되는, 표적화된 유전자좌 내의 영역들에 상응한다. 일부 표적화 벡터는 핵산 삽입물 없이 5' 및 3' 상동성 아암을 포함한다. 그러한 표적화 벡터는 핵산 삽입물을 삽입하지 않고서 5' 표적 서열과 3' 표적 서열 사이의 서열을 결실시키도록 기능할 수 있다.

[0135] 상동성 아암과 표적 서열은, 이들 두 영역이 상동 재조합 반응에 대한 기질로서 작용하기에 충분한 수준의 서열 동일성을 서로에 대해 공유할 때, 서로에게 "상응하거나" 또는 "상응하고 있다". 용어 "상동성"은 상응하는 서열과 동일하거나 그와 서열 동일성을 공유하는 DNA 서열을 포함한다. 주어진 표적 서열과, 표적화 벡터 상에서 확인되는 상응하는 상동성 아암 사이의 서열 동일성은 상동 재조합이 발생하도록 허용하는 임의의 정도의 서열 동일성일 수 있다. 예를 들어, 표적화 벡터(또는 이의 단편)의 상동성 아암과 표적 서열(또는 이의 단편)에 의해 공유되는 서열 동일성의 양은 적어도 50%, 55%, 60%, 65%, 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성일 수 있어서, 이들 서열은 상동 재조합을 거치게 된다. 더욱이, 상동성 아암과 상응하는 표적 서열 사이의 상응하는 상동성 영역은 절단된 인식 부위에서 상동 재조합을 촉진하기에 충분한 임의의 길이를 가질 수 있다. 예를 들어, 주어진 상동성 아암 및/또는 상응하는 표적 서열은 (예컨대, 본 명세서의 어딘가 다른 곳에 기재된 LTVEC 벡터에서 실험된 바와 같이) 적어도 약 5 내지 10 kb, 5 내지 15 kb, 5 내지 20 kb, 5 내지 25 kb, 5 내지 30 kb, 5 내지 35 kb, 5 내지 40 kb, 5 내지 45 kb, 5 내지 50 kb, 5 내지 55 kb, 5 내지 60 kb, 5 내지 65 kb, 5 내지 70 kb, 5 내지 75 kb, 5 내지 80 kb, 5 내지 85 kb, 5 내지 90 kb, 5 내지 95 kb, 5 내지 100 kb, 100 내지 200 kb, 또는 200 내지 300 kb 길이 또는 그 이상의 상응하는 상동성 영역을 포함할 수 있어서, 상동성 아암이 세포의 게놈 내의 상응하는 표적 서열과의 상동성 재조합을 거치기에 충분한 상동성을 갖게 된다.

[0136] 상동성 아암들은 세포에 고유한 유전자좌(예를 들어, 표적화된 유전자좌)에 상응할 수 있거나, 대안적으로

이는, 예를 들어, 도입유전자, 발현 카세트, 또는 DNA의 이중성 또는 외인성 영역을 포함하는, 세포의 게놈 내로 통합된 DNA의 이중성 또는 외인성 세그먼트의 영역에 상응할 수 있다. 대안적으로, 표적화 벡터의 상동성 아암들은 효모 인공 염색체(YAC), 박테리아 인공 염색체(BAC), 인간 인공 염색체의 영역, 또는 적당한 숙주 세포에 함유된 임의의 다른 조작된 영역에 상응할 수 있다. 추가로 또한, 표적화 벡터의 상동성 아암들은 BAC 라이브러리, 코스미드 라이브러리, 또는 P1 파지 라이브러리의 영역에 상응할 수 있거나 이로부터 유래될 수 있다. 소정 경우에, 표적화 벡터의 상동성 아암들은 원핵생물, 효모, 조류(예를 들어, 닭), 비인간 포유류, 설치류, 인간, 래트, 마우스, 햄스터, 토끼, 돼지, 소, 사슴, 양, 염소, 고양이, 개, 페렛, 영장류(예를 들어, 마모셋, 붉은털원숭이), 사육된 포유류, 농업용 포유류, 또는 임의의 다른 관심 유기체에 대해 고유하거나, 이중성이거나, 외인성인 유전자좌에 상응한다. 일부 경우에, 상동성 아암들은, 뉴클레아제 작용제(예를 들어, Cas 단백질)에 의해 유도된 Nick 또는 이중-가닥 브레이크의 부재 하에서 단지 부정확하게 또는 단지 유의하게 낮은 효율로 표적화될 수 있거나 종래의 방법을 사용하여 표적화할 수 없는 세포의 유전자좌에 상응한다. 일부 경우에, 상동성 아암들은 합성 DNA로부터 유래된다.

[0137] 일부 표적화 벡터에서, 5' 및 3' 상동성 아암들은 표적화된 게놈에 상응한다. 대안적으로, 상동성 아암들은 관련 게놈으로부터 유래될 수 있다. 예를 들어, 표적화된 게놈은 제1 세포주(strain)의 마우스 게놈이고, 표적화 아암들은 제2 세포주의 마우스 게놈으로부터 유래되며, 여기서 제1 세포주와 제2 세포주는 상이하다. 소정의 경우에, 상동성 아암들은 동일한 동물의 게놈으로부터 유래되거나 동일한 세포주의 게놈으로부터 유래되고, 예를 들어, 표적화된 게놈은 제1 세포주의 마우스 게놈이고, 표적화 아암들은 동일한 마우스로부터 또는 동일한 세포주로부터의 마우스 게놈으로부터 유래된다.

[0138] 표적화 벡터의 상동성 아암은, 예를 들어, 적어도 5 내지 10 kb, 5 내지 15 kb, 5 내지 20 kb, 5 내지 25 kb, 5 내지 30 kb, 5 내지 35 kb, 5 내지 40 kb, 5 내지 45 kb, 5 내지 50 kb, 5 내지 55 kb, 5 내지 60 kb, 5 내지 65 kb, 5 내지 70 kb, 5 내지 75 kb, 5 내지 80 kb, 5 내지 85 kb, 5 내지 90 kb, 5 내지 95 kb, 5 내지 100 kb, 100 내지 200 kb, 또는 200 내지 300 kb 길이 또는 그 이상을 포함하는, 상응하는 표적 서열과의 상동 재조합 이벤트를 촉진하기에 충분한 임의의 길이를 가질 수 있다. 아래에서 더욱 상세하게 설명되는 바와 같이, 대형 표적화 벡터는 더 큰 길이의 표적화 아암을 사용할 수 있다.

[0139] 표적 유전자좌의 변형을 돕도록 뉴클레아제 작용제(예를 들어, CRISPR/Cas 시스템)가 표적화 벡터와 조합하여 사용될 수 있다. 그러한 뉴클레아제 작용제는 표적화 벡터와 표적 유전자좌 사이에 상동 재조합을 촉진할 수 있다. 뉴클레아제 작용제가 표적화 벡터와 조합하여 사용되는 경우, 표적화 벡터는 뉴클레아제 절단 부위에서 Nick 또는 이중-가닥 브레이크 시에 표적 서열들과 상동성 아암들 사이의 상동 재조합 이벤트의 발생을 촉진하도록 뉴클레아제 절단 부위에 충분히 근접하게 위치된 5' 및 3' 표적 서열에 상응하는 5' 및 3' 상동성 아암을 포함할 수 있다. 용어 "뉴클레아제 절단 부위"는 뉴클레아제 작용제에 의해 Nick 또는 이중-가닥 브레이크가 생성되는 DNA 서열(예를 들어, Cas9 절단 부위)을 포함한다. 표적화 벡터의 5' 및 3' 상동성 아암에 상응하는 표적화된 유전자좌 내의 표적 서열이 뉴클레아제 절단 부위에 "충분히 근접하게 위치되는" 경우는, 그 거리가 인식 부위에서의 Nick 또는 이중-가닥 절단 시에 5' 및 3' 표적 서열과 상동성 아암 사이의 상동 재조합 이벤트의 발생을 촉진하도록 하는 그러한 거리이다. 따라서, 구체적인 경우에, 표적화 벡터의 5' 및/또는 3' 상동성 아암에 상응하는 표적 서열은 주어진 인식 부위의 적어도 1개의 뉴클레오타이드 내에 있거나 주어진 인식 부위의 적어도 10개의 뉴클레오타이드 내지 약 14 kb 내에 존재한다. 일부 경우에, 뉴클레아제 절단 부위는 표적 서열들 중 적어도 하나 또는 둘 모두에 바로 인접한다.

[0140] 표적화 벡터의 상동성 아암들에 상응하는 표적 서열들과 뉴클레아제 절단 부위의 공간적 관계는 다양할 수 있다. 예를 들어, 표적 서열들은 뉴클레아제 절단 부위의 5'에 위치될 수 있거나, 표적 부위들은 뉴클레아제 절단 부위의 3'에 위치될 수 있거나, 표적 서열들은 뉴클레아제 절단 부위를 플랭킹할 수 있다.

[0141] 뉴클레아제 작용제와 (예를 들어, 대형 표적화 벡터를 포함하는) 표적화 벡터의 조합 사용은 표적화 벡터의 단독 사용과 비교하여 증가된 표적화 효율을 가져올 수 있다. 예를 들어, 표적화 벡터가 뉴클레아제 작용제와 함께 사용될 때, 표적화 벡터의 표적화 효율은 표적화 벡터가 단독으로 사용될 때와 비교하여 적어도 2배, 적어도 3배, 적어도 4배, 또는 적어도 10배 증가될 수 있다.

[0142] (3) 대형 표적화 벡터

[0143] 일부 표적화 벡터는 "대형 표적화 벡터" 또는 "LTVEC"인데, 이는 세포에서 상동 재조합을 수행하도록 의도된 다른 접근법에 의해 전형적으로 사용되는 핵산 서열보다 더 큰 핵산 서열에 상응하고 그로부터 유래된 상동성 아암들을 포함하는 표적화 벡터를 포함한다. LTVEC는 또한 세포에서 상동 재조합을 수행하도록 의도된 다른 접근

법에 의해 전형적으로 사용되는 핵산 서열보다 더 큰 핵산 서열을 갖는 핵산 삽입물을 포함하는 표적화 벡터를 포함한다. 예를 들어, LTVEC는, 전통적인 플라스미드-기반 표적화 벡터에 의해서는 크기 제한 때문에 수용될 수 없는 큰 유전자좌의 변형을 가능하게 한다. 예를 들어, 표적화된 유전자좌는, 뉴클레아제 작용제(예를 들어, Cas 단백질)에 의해 유도된 Nick 또는 이중-가닥 브레이크의 부재 하에서 단지 부정확하게 또는 단지 유의하게 낮은 효율로 표적화될 수 있거나 종래의 방법을 사용하여 표적화할 수 없는 세포의 유전자좌일 수 있다(즉, 5' 및 3' 상동성 아암은 그러한 유전자좌에 상응할 수 있다).

[0144] LTVEC의 예는 세균 인공 염색체(BAC), 인간 인공 염색체, 또는 효모 인공 염색체(YAC)로부터 유래된 벡터를 포함한다. LTVEC의 비제한적인 예 및 그의 제조 방법은, 예를 들어, 미국 특허 제6,586,251호; 미국 특허 제6,596,541호; 미국 특허 제7,105,348호; 및 국제특허 공개 WO 2002/036789호(PCT/US01/45375)에 기재되어 있으며, 이들 각각은 본 명세서에 참고로 포함된다. LTVEC는 선형 형태 또는 환형 형태일 수 있다.

[0145] LTVEC는 임의의 길이를 가질 수 있으며, 이는, 예를 들어, 약 50 kb 내지 약 300 kb, 약 50 kb 내지 약 75 kb, 약 75 kb 내지 약 100 kb, 약 100 kb 내지 125 kb, 약 125 kb 내지 약 150 kb, 약 150 kb 내지 약 175 kb, 약 175 kb 내지 약 200 kb, 약 200 kb 내지 약 225 kb, 약 225 kb 내지 약 250 kb, 약 250 kb 내지 약 275 kb 또는 약 275 kb 내지 약 300 kb를 포함한다. 대안적으로, LTVEC는 적어도 10 kb, 적어도 15 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 150 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb 또는 그 이상일 수 있다. LTVEC의 크기는 너무 커서 종래의 검정, 예를 들어, 서던 블롯팅(southern blotting) 및 긴 범위(예를 들어, 1 kb 내지 5 kb) PCR에 의한 표적화 이벤트의 스크리닝을 가능하게 할 수 없다.

[0146] 일부 경우에, LTVEC는 약 5 kb 내지 약 200 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 60 kb 내지 약 70 kb, 약 80 kb 내지 약 90 kb, 약 90 kb 내지 약 100 kb, 약 100 kb 내지 약 110 kb, 약 120 kb 내지 약 130 kb, 약 130 kb 내지 약 140 kb, 약 140 kb 내지 약 150 kb, 약 150 kb 내지 약 160 kb, 약 160 kb 내지 약 170 kb, 약 170 kb 내지 약 180 kb, 약 180 kb 내지 약 190 kb, 또는 약 190 kb 내지 약 200 kb 범위의 핵산 삽입물을 포함한다. 다른 경우에, 핵산 삽입물은 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 250 kb, 약 250 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 또는 약 350 kb 내지 약 400 kb 범위일 수 있다.

[0147] 일부 LTVEC에서, 5' 상동성 아암 및 3' 상동성 아암의 총합은 적어도 10 kb이다. 다른 LTVEC에서, 5' 상동성 아암은 약 5 kb 내지 약 100 kb의 범위이고/이거나 3' 상동성 아암은 약 5 kb 내지 약 100 kb의 범위이다. 각각의 상동성 아암은, 예를 들어, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 약 90 kb 내지 약 100 kb, 약 100 kb 내지 약 110 kb, 약 110 kb 내지 약 120 kb, 약 120 kb 내지 약 130 kb, 약 130 kb 내지 약 140 kb, 약 140 kb 내지 약 150 kb, 약 150 kb 내지 약 160 kb, 약 160 kb 내지 약 170 kb, 약 170 kb 내지 약 180 kb, 약 180 kb 내지 약 190 kb, 또는 약 190 kb 내지 약 200 kb일 수 있다. 5' 및 3' 상동성 아암의 총합은, 예를 들어, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 약 90 kb 내지 약 100 kb, 약 100 kb 내지 약 110 kb, 약 110 kb 내지 약 120 kb, 약 120 kb 내지 약 130 kb, 약 130 kb 내지 약 140 kb, 약 140 kb 내지 약 150 kb, 약 150 kb 내지 약 160 kb, 약 160 kb 내지 약 170 kb, 약 170 kb 내지 약 180 kb, 약 180 kb 내지 약 190 kb, 또는 약 190 kb 내지 약 200 kb일 수 있다. 대안적으로, 각각의 상동성 아암은 적어도 5 kb, 적어도 10 kb, 적어도 15 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 또는 적어도 200 kb일 수 있다. 마찬가지로, 5' 및 3' 상동성 아암의 총합은 적어도 5 kb, 적어도 10 kb, 적어도 15 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 또는 적어도 200 kb일 수 있다.

[0148] 일부 경우에, LTVEC 및 핵산 삽입물은 표적 유전자좌에서 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약

20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb의 결실이 가능하도록 설계된다. 대안적으로, 결실은 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 150 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb 또는 그 이상일 수 있다.

[0149] 다른 경우에, LTVEC 및 핵산 삽입물은 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 250 kb, 약 250 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 또는 약 350 kb 내지 약 400 kb 범위의 외인성 핵산 서열의 표적 유전자좌 내로의 삽입이 가능하도록 설계된다. 대안적으로, 삽입은 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 150 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb 또는 그 이상일 수 있다.

[0150] 또 다른 경우에, 핵산 삽입물 및/또는 결실되는 내인성 유전자좌의 영역은 적어도 100, 200, 300, 400, 500, 600, 700, 800, 또는 900개 뉴클레오타이드 또는 적어도 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 11 kb, 12 kb, 13 kb, 14 kb, 15 kb, 16 kb 또는 그 이상이다.

[0151] B. 1-세포기 배아에 대한 표적화 벡터 및 핵산 삽입물

[0152] 1-세포기 배아에 사용하기 위한 표적화 벡터는 5 kb 길이 이하이고 테옥시리보핵산(DNA) 또는 리보핵산(RNA)일 수 있으며, 이것은 단일-가닥 또는 이중-가닥일 수 있고 이것은 환형 형태 또는 선형 형태일 수 있다. 1-세포기 배아에 사용하기 위한 예시적인 표적화 벡터는 약 50개 뉴클레오타이드 내지 약 5 kb 길이이다. 예를 들어, 1-세포기 배아에 사용하기 위한 표적화 벡터는 약 50 내지 약 100개, 약 100 내지 약 200개, 약 200 내지 약 300개, 약 300 내지 약 400개, 약 400 내지 약 500개, 약 500 내지 약 600개, 약 600 내지 약 700개, 약 700 내지 약 800개, 약 800 내지 약 900개, 또는 약 900 내지 약 1,000개 뉴클레오타이드 길이일 수 있다. 대안적으로, 1-세포기 배아에 사용하기 위한 표적화 벡터는 약 1 kb 내지 약 1.5 kb, 약 1.5 kb 내지 약 2 kb, 약 2 kb 내지 약 2.5 kb, 약 2.5 kb 내지 약 3 kb, 약 3 kb 내지 약 3.5 kb, 약 3.5 kb 내지 약 4 kb, 약 4 kb 내지 약 4.5 kb, 또는 약 4.5 kb 내지 약 5 kb 길이일 수 있다. 대안적으로, 1-세포기 배아에 사용하기 위한 표적화 벡터는, 예를 들어, 5 kb, 4.5 kb, 4 kb, 3.5 kb, 3 kb, 2.5 kb, 2 kb, 1.5 kb, 1 kb, 900개 뉴클레오타이드, 800개 뉴클레오타이드, 700개 뉴클레오타이드, 600개 뉴클레오타이드, 500개 뉴클레오타이드, 400개 뉴클레오타이드, 300개 뉴클레오타이드, 200개 뉴클레오타이드, 100개 뉴클레오타이드, 또는 50개 뉴클레오타이드 길이 이하일 수 있다. 단일-가닥 DNA 공여자의 경우에, 예시적인 표적화 벡터는 약 60개 뉴클레오타이드 내지 약 200개 뉴클레오타이드(예를 들어, 약 60개 뉴클레오타이드 내지 약 80개 뉴클레오타이드, 약 80개 뉴클레오타이드 내지 약 100개 뉴클레오타이드, 약 100개 뉴클레오타이드 내지 약 120개 뉴클레오타이드, 약 120개 뉴클레오타이드 내지 약 140개 뉴클레오타이드, 약 140개 뉴클레오타이드 내지 약 160개 뉴클레오타이드, 약 160개 뉴클레오타이드 내지 약 180개 뉴클레오타이드, 또는 약 180개 뉴클레오타이드 내지 약 200개 뉴클레오타이드)일 수 있다.

[0153] 그러한 표적화 벡터는 표적화된 유전자좌 내의 영역들(각각 5' 표적 서열 및 3' 표적 서열)에 상응하는 5' 및 3' 상동성 아암을 포함한다. 선택적으로, 표적화 벡터는 5' 및 3' 상동성 아암에 의해 플랜킹된 핵산 삽입물(예를 들어, 게놈 표적 유전자좌에서 통합시키고자 하는 DNA의 세그먼트)을 포함한다. 표적 유전자좌에서의 핵산 삽입물의 통합은 표적 유전자좌에 대한 관심 핵산 서열의 부가, 표적 유전자좌에서의 관심 핵산 서열의 결실, 또는 표적 유전자좌에서의 관심 핵산 서열의 대체(즉, 결실 및 삽입)를 가져올 수 있다.

[0154] 상동성 아암과 상응하는 표적 서열 사이의 상응하는 상동성 영역은 상동 재조합을 촉진하기에 충분한 임의의 길이를 가질 수 있다. 1-세포기 배아에 사용하기 위한 예시적인 상동성 아암은 약 20개 뉴클레오타이드 내지 약 2.5 kb 길이(예를 들어, 약 30개 뉴클레오타이드 내지 약 100개 뉴클레오타이드 길이)이다. 예를 들어, 주어진 상동성 아암 및/또는 상응하는 표적 서열은 약 20 내지 약 30개, 약 30 내지 약 40개, 약 40 내지 약 50개, 약 50 내지 약 60개, 약 60 내지 약 70개, 약 70 내지 약 80개, 약 80 내지 약 90개, 약 90 내지 약 100개, 약 100 내지 약 150개, 약 150 내지 약 200개, 약 200 내지 약 250개, 약 250 내지 약 300개, 약 300 내지 약 350개, 약 350 내지 약 400개, 약 400 내지 약 450개, 또는 약 450 내지 약 500개 뉴클레오타이드 길이인 상응하는 상동성 영역을 포함할 수 있어서, 상동성 아암들은 세포의 게놈 내의 상응하는 표적 서열들과의 상동 재조합을 거치

기에 충분한 상동성을 갖는다. 대안적으로, 주어진 상동성 아암 및/또는 상응하는 표적 서열은 약 0.5 kb 내지 약 1 kb, 약 1 kb 내지 약 1.5 kb, 약 1.5 kb 내지 약 2 kb, 또는 약 2 kb 내지 약 2.5 kb 길이의 상응하는 상동성 영역을 포함할 수 있다. 단일-가닥 DNA 공여자의 경우에, 예시적인 상동성 아암들은 약 30개 뉴클레오티드 내지 약 60개 뉴클레오티드(예를 들어, 약 30 내지 약 40개 뉴클레오티드, 약 40개 뉴클레오티드 내지 약 50개 뉴클레오티드, 또는 약 50개 뉴클레오티드 내지 약 60개 뉴클레오티드)일 수 있다.

[0155] 전술된 바와 같이, 상동성 아암들은 세포에 고유한 유전자좌(예를 들어, 표적화된 유전자좌)에 상응할 수 있거나, 대안적으로 이는 세포의 게놈 내로 통합된 DNA의 이중성 또는 외인성 세그먼트의 영역에 상응할 수 있다. 전술된 바와 같이, 5' 및 3' 표적 서열은 바람직하게는, Cas 절단 부위에서 단일-가닥 브레이크(닉) 또는 이중-가닥 브레이크 시에 표적 서열들과 상동성 아암들 사이의 상동 재조합 이벤트의 발생을 촉진하도록 하기에 충분히 근접하게 Cas 절단 부위에 위치된다.

[0156] 핵산 삽입물 또는 결실 및/또는 대체되는 표적 유전자좌에서의 상응하는 핵산은 다양한 길이일 수 있다. 예시적인 핵산 삽입물 또는 결실 및/또는 대체되는 표적 유전자좌에서의 상응하는 핵산은 약 10개 뉴클레오티드 내지 약 5 kb 길이이다. 예를 들어, 핵산 삽입물 또는 결실 및/또는 대체되는 표적 유전자좌에서의 상응하는 핵산은 약 1 내지 약 10개, 약 10 내지 약 20개, 약 20 내지 약 30개, 약 30 내지 약 40개, 약 40 내지 약 50개, 약 50 내지 약 60개, 약 60 내지 약 70개, 약 70 내지 약 80개, 약 80 내지 약 90개, 약 90 내지 약 100개, 약 100 내지 약 110개, 약 110 내지 약 120개, 약 120 내지 약 130개, 약 130 내지 약 140개, 약 140 내지 약 150개, 약 150 내지 약 160개, 약 160 내지 약 170개, 약 170 내지 약 180개, 약 180 내지 약 190개, 약 190 내지 약 200개, 약 200 내지 약 300개, 약 300 내지 약 400개, 약 400 내지 약 500개, 약 500 내지 약 600개, 약 600 내지 약 700개, 약 700 내지 약 800개, 약 800 내지 약 900, 또는 약 900 내지 약 1,000개 뉴클레오티드 길이일 수 있다. 일례로서, 약 4개 뉴클레오티드 내지 약 12개 뉴클레오티드 길이의 삽입물이 제한 효소 부위를 생성하기 위해 삽입될 수 있다. 마찬가지로, 핵산 삽입물 또는 결실 및/또는 대체되는 표적 유전자좌에서의 상응하는 핵산은 약 1 kb 내지 약 1.5 kb, 약 1.5 kb 내지 약 2 kb, 약 2 kb 내지 약 2.5 kb, 약 2.5 kb 내지 약 3 kb, 약 3 kb 내지 약 3.5 kb, 약 3.5 kb 내지 약 4 kb, 약 4 kb 내지 약 4.5 kb, 또는 약 4.5 kb 내지 약 5 kb 길이일 수 있다. 게놈 표적 유전자좌로부터 결실되는 핵산은 또한 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 약 90 kb 내지 약 100 kb, 약 100 kb 내지 약 200 kb, 약 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 600 kb, 약 600 kb 내지 약 700 kb, 약 700 kb 내지 약 800 kb, 약 800 kb 내지 약 900 kb, 약 900 kb 내지 약 1 Mb 또는 그 이상일 수 있다. 대안적으로, 게놈 표적 유전자좌로부터 결실되는 핵산은 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 약 2.5 Mb 내지 약 3 Mb, 약 3 Mb 내지 약 4 Mb, 약 4 Mb 내지 약 5 Mb, 약 5 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb일 수 있다.

[0157] 전술된 바와 같이, 핵산 삽입물은 게놈 DNA 또는 임의의 다른 유형의 DNA를 포함할 수 있고, 핵산 삽입물 또는 결실 및/또는 대체되는 표적 유전자좌에서의 상응하는 핵산은 코딩 영역 또는 비-코딩 영역일 수 있고, 핵산 삽입물은 표적 게놈 유전자좌에서의 결실에 대해 표적화되는 서열에 대해 상동성 또는 오르토로그성일 수 있다. 핵산 삽입물은 또한 조건적 대립유전자, 선택 마커를 인코딩하는 폴리뉴클레오티드, 리포터 유전자, 하나 이상의 발현 카세트, 하나 이상의 결실 카세트, 또는 전술된 바와 같은 부위-특이적 재조합 표적 서열로 플랭킹된 핵산을 포함하는 핵산 삽입물을 포함할 수 있다.

C. 프로모터

[0159] 본 명세서에 기재된 다양한 핵산 서열은 프로모터에 작동가능하게 연결될 수 있다. 그러한 프로모터는, 예를 들어, 다능성 래트 세포, 진행 세포, 포유류 세포, 비인간 포유류 세포, 인간 세포, 설치류 세포, 마우스 세포, 또는 햄스터 세포에서 활성일 수 있다. 1-세포기 배아에서 활성인 프로모터가 또한 사용될 수 있다. 프로모터는, 예를 들어 구성적으로 활성인 프로모터, 조건적 프로모터, 유도성 프로모터, 시간적으로 제한된 프로모터(예를 들어, 발생적으로 조절된 프로모터), 또는 공간적으로 제한된 프로모터(예를 들어, 세포-특이적 또는 조직-특이적 프로모터)일 수 있다. 프로모터의 예는, 예를 들어 국제특허 공개 WO 2013/176772호에서 찾을 수 있으며, 전체적으로 본 명세서에 참고로 포함된다.

[0160] 유도성 프로모터의 예는, 예를 들어, 화학적으로 조절된 프로모터 및 물리적으로 조절된 프로모터를 포함한다.

화학적으로 조절된 프로모터는, 예를 들어, 알코올-조절된 프로모터(예를 들어, 알코올 데하이드로게나제(alcA) 유전자 프로모터), 테트라사이클린-조절된 프로모터(예를 들어, 테트라사이클린-반응성 프로모터, 테트라사이클린 작동자 서열(tetO), tet-On 프로모터, 또는 tet-Off 프로모터), 스테로이드 조절된 프로모터(예를 들어, 래트 글루코코르티코이드 수용체, 에스트로겐 수용체의 프로모터, 또는 엑디손 수용체의 프로모터), 또는 금속-조절된 프로모터(예를 들어, 금속단백질 프로모터)를 포함한다. 물리적으로 조절된 프로모터는, 예를 들어, 온도-조절된 프로모터(예를 들어, 열충격 프로모터) 및 광-조절된 프로모터(예를 들어, 광-유도성 프로모터 또는 광-억제성 프로모터)를 포함한다.

[0161] 조직-특이적 프로모터는, 예를 들어, 뉴런-특이적 프로모터, 신경교-특이적 프로모터, 근육 세포-특이적 프로모터, 심장 세포-특이적 프로모터, 신장 세포-특이적 프로모터, 골 세포-특이적 프로모터, 내피 세포-특이적 프로모터, 또는 면역 세포-특이적 프로모터(예를 들어, B 세포 프로모터 또는 T 세포 프로모터)일 수 있다.

[0162] 발생적으로 조절된 프로모터는, 예를 들어, 단지 배아 발생기 동안에만, 또는 단지 성인 세포에서만 활성화된 프로모터를 포함한다.

[0163] 프로모터는 또한 세포 유형에 기반하여 선택될 수 있다. 예를 들어, 다양한 공지된 프로모터는 진핵 세포, 포유류 세포, 비인간 포유류 세포, 다능성 세포, 비인간 다능성 세포, 인간 다능성 세포, 인간 ES 세포, 인간 성인 줄기 세포, 발생적으로 제한된 인간 선조 세포, 인간 iPS 세포, 인간 세포, 설치류 세포, 래트 세포, 마우스 세포, 햄스터 세포, 섬유아세포, 또는 CHO 세포에서 사용된다.

[0164] IV. 게놈 변형 방법 및 유전자 변형된 비인간 동물의 제조 방법

[0165] A. 게놈 변형 방법

[0166] 2개의 가이드 RNA를 사용하여 단일 게놈 표적 유전자와 내의 상이한 영역들을 표적화함을 통해 세포 내의 게놈을 변형시키기 위한 다양한 방법이 제공된다. 2개 이상의 가이드 RNA(예를 들어, 3개의 가이드 RNA 또는 4개의 가이드 RNA)를 사용하여 단일 게놈 표적 유전자와 내의 상이한 영역들을 표적화하는 방법이 또한 제공된다. 이들 방법은 시험관내에서, 생체외에서, 또는 생체내에서 일어날 수 있다. 그러한 방법들은 이중대립유전자 유전자 변형의 생성을 촉진하고, 게놈 붕괴 또는 다른 표적화된 변형, 예컨대 게놈 내의 핵산 서열의 결실과 외인성 핵산 서열로의 대체의 동시 수행을 포함할 수 있다.

[0167] 표적화 벡터와 표적 유전자와 사이의 상동 재조합에 의한 표적화된 유전자 변형은, 특히 설치류 배아 줄기 세포 이외의 세포 유형에서, 매우 비효율적일 수 있다. 표적 유전자좌에서의 뉴클레아제-유도 이중-가닥 DNA 브레이크와 조합한 표적화 벡터의 사용은 단순 변형, 예컨대 작은 결실 또는 삽입에 대한 이형접합성 표적화 효율을 크게 향상시킬 수 있다.

[0168] 표적화 벡터와 1개의 가이드 RNA(gRNA)에 의해 가이드된 CRISPR/Cas9 뉴클레아제의 조합이 또한 매우 큰 그리고 작은 효율의 유전자 변형, 예컨대 마우스 유전자의 결실 및 그의 인간 대응부(counterpart)로의 동시 대체(인간화)에 대한 이형접합성 표적화 효율을 증가시킬 수 있다. 그러한 변형은 매우 큰(예를 들어, 50 kb 초과) 결실 및 삽입을 수반할 수 있다(실시예 1에서의 *Lrp5*, *C5 (Hc)*, *Ror1*, 및 *Trpa1* 표적화 참조).

[0169] 표적 게놈 유전자좌에서 뉴클레아제, 예컨대 Cas9에 의해 생성된 하나 이상의 이중-가닥 브레이크의 상동성-유도 수복이 행해지는 동안, 하나 이상의 브레이크는 먼저 5' 말단의 절제에 의해 3'-단일-가닥 오버행(overhang)을 생성하도록 처리된다. 이어서, Rad51이 단일-가닥 DNA 상에서 중합되어 상동성 서열을 검색하고, 손상되지 않은 상동성 주형 듀플렉스 DNA(예를 들어, 표적화 벡터) 내로 가닥 침입이 일어나고, 주형으로서 손상되지 않은 상동성 DNA(예를 들어, 표적화 벡터)를 사용하여 하나 이상의 이중-가닥 브레이크의 수복을 촉진하도록 중간 D-루프 구조가 형성된다. 이어서, 염색체 서열은 플랭킹 상동성 영역들을 수반하는 이중 교차 이벤트에 의해 표적화 벡터로부터의 핵산 삽입물에 의해 대체된다. 이러한 과정이 적절하게 진행되는지의 여부는 여러 인자들, 예컨대 핵산 삽입물의 크기, 표적화 벡터의 상동성 아암들에 대해 상동성인 영역들의 길이, 및 (예를 들어, 하나 이상의 이중 가닥 브레이크에 관하여) 표적화 벡터의 상동성 아암들에 대해 상동성인 영역들의 위치에 의해 영향을 받는다.

[0170] 핵산 삽입물 또는 표적 게놈 유전자좌에서의 결실된 서열의 크기가 증가함에 따라, 절제 과정은 더 예측 불가능해지고, 중간 D 루프 구조의 안정성은 감소되어 더 예측 불가능해지고, 재조합 과정의 성공은 일반적으로 감소되고 더 예측 불가능해진다. 예를 들어, 표적화된 변형의 크기가 증가함에 따라, 내부 재조합의 위험이 증가하는데, 이는 특히 대체되는 서열과 삽입되는 서열 사이에 유사성이 존재하는 경우에 그러하다. 그러한 내부 재조합이 일어나는 경우, 상동 재조합 교환이 의도된 표적 영역의 내부에서 일어나고, 전(full) 핵산 삽입물이 게놈

표적 유전자와 내로 혼입되지는 않는다. 게다가, 종래의 사고는, 이중-가닥 브레이크와 돌연변이 또는 삽입 부위 사이의 거리가 (예를 들어, 100 bp 또는 200 bp를 초과하여) 증가함에 따라, HR-매개 삽입의 효율이 감소된다는 것이다. 문헌[Beumer *et al.* (2013) *Genes#Genomes#Genetics* 3:657-664]; 문헌[Elliott *et al.* (1998) *Mol. Cell. Biol.* 18:93-101]; 및 문헌[Byrne *et al.* (2015) *Nucleic Acids Research* 43(3):e21]을 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0171] 표적 게놈 유전자좌에서 큰 결실을 생성하고, 동시에 큰 조각의 외래 DNA를 삽입하는 표적화된 유전자 변형을 달성하기 위해서는, 재조합 중간체로서의 이중 오메가 구조의 형성을 필요로 한다. 변형이 더 클수록, 구조의 안정성은 더 낮다. 1-세포기 배아 이외의 세포 유형에서, 총합 10 kb 이상의 총 상동성을 갖는 LTVEC가 사용될 수 있다. 상동성 아암들이 총합 10 kb 이상의 총 상동성을 갖는 LTVEC는 이중 오메가 재조합 중간체의 안정성을 증가시켜, 뉴클레아제-매개된 동시적 큰 결실 및 대형 핵산 삽입물로의 대체를 촉진시키고, 추가로 상동성 영역들에 인접한 이중-가닥 브레이크들이 표적화 효율을 향상시킬 수 있게 할 뿐만 아니라 상동성 영역들로부터 멀리 떨어져 있는 것들도 표적화 효율을 향상시킬 수 있게 한다.

[0172] 매우 큰 인간화를 수반하는 유전자 변형의 경우, 표적화 벡터를 2개의 gRNA에 의해 가이드된 CRISPR/Cas9 뉴클레아제 시스템과 조합함으로써 1개의 gRNA를 사용하여 달성된 것을 초과하여 표적화 효율을 추가로 향상시킬 수 있다(실시예 1에서의 *Lrp5*, *C5 (Hc)*, *Ror1*, 및 *Trpa1* 인간화 참조). 2개의 gRNA의 사용은 이에 관하여 예기치 않은 결과를 생성한다. 1개의 gRNA를 사용한 표적화 - 이는 이중대립유전자 변형을 낮은 빈도로 생성하거나 전혀 생성하지 않음 - 와 비교하여, 2개의 gRNA를 사용한 표적화는 유의하게 증가된 속도로의 동형접합성 표적화된 세포, 동형접합성 결실된 세포, 및 복합 이형접합성 표적화된 세포(반접합성 표적화된 세포를 포함함)의 생성을 가져온다.

[0173] 단일 표적화 실험에서 3개의 대립 유전자 유형 - 동형접합성 표적화된 유형, 동형접합성 결실된 유형, 및 복합 이형접합성 표적화된 유형(특히 반접합성 표적화된 유형) - 을 생성하는 방법은 표적화된 유전자 변형에 대한 새로운 가능성 및 향상된 효율을 제공한다. 단순 유전자 변형, 예컨대 마우스 ES 세포에서의 유전자의 표적화된 결실 및 유전자 발현을 보고하는 단백질(예를 들어, β -갈락토시다제 또는 형광 단백질)을 인코딩하는 서열로의 그의 대체의 경우, 표적화 벡터와 2개의 gRNA에 의해 가이드된 CRISPR/Cas9 시스템의 조합은 이형접합성 표적화된 ES 세포의 생성을 향상시키며, 이어서 이것은 VelociMouse® 방법에 의해 완전히 ES 세포-유도된 F0 세대 마우스를 생성하는 데 사용될 수 있다. 문헌[Poueymirou *et al.* (2007) *Nat. Biotech.* 25:91-99]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 이들 마우스는 리포터 녹-인 대립유전자를 사용하여 조직 특이적 유전자 발현을 연구하는 데 유용하다. 동일한 실험에서 생성된 동형접합성 표적화된 ES 세포 클론은, 동형접합성 표적화된 유전자 결실을 사용하여 VelociMice로 변환될 수 있는데, 이것은 유전자 녹아웃(knockout)의 표현형 결과뿐만 아니라 리포터로부터의 유전자 발현에 대해서도 연구될 수 있다. 표적 유전자의 동형접합성 CRISPR-유도 결실을 갖는 ES 세포로부터의 VelociMice의 생성은 동형접합성 표적화된 마우스에서 관찰되는 녹아웃 표현형의 검증을 가능하게 하고, 순수한 유전자 결실과 리포터 및 약물 선택 카세트의 삽입이 동반되는 결실 사이의 표현형 차이를 보여줄 수 있다. 표적화된 결실-삽입 대립유전자 및 CRISPR-유도 결실 둘 모두를 보유하는 복합 이형접합성(및 특히 반접합성) ES 세포 클론은 동형접합성 표적화된 클론 및 동형접합성 결실된 클론으로부터 유래된 것들과 동일한 연구 기회를 갖는 VelociMice의 생성을 가능하게 한다. 게다가, 이들 마우스는 단일 ES 세포 클론 단독으로부터의 표적화된 결실 돌연변이 마우스 세포주 및 단순 결실 돌연변이 마우스 세포주 둘 모두를 확립하도록 교배될 수 있다.

[0174] 이러한 이점들은 인간화의 경우로 확장될 때 부가 가치를 갖는다. 마우스 유전자의 인간화의 중요한 용도는 인간-특이적 치료제를 시험할 동물 모델을 생성하는 것이다. 효과적인 모델이 되도록 인간화하기 위해서는, 마우스 유전자는 생물학적 기능 또는 약물과의 적절한 상호작용을 손상시킬 수 있는 마우스 유전자 산물과 인간 유전자 산물 사이의 상호작용을 피하도록 제거되거나 불활성화되어야 한다. 동시에, 인간 유전자는 그의 마우스 대응부의 생물학적 기능을 대체할 수 있어야 한다. 이들 요건은 2개의 gRNA에 의해 가이드된 Cas9 뉴클레아제를 마우스 유전자의 결실 및 인간 유전자로의 대체를 동시에 생성하도록 설계된 표적화 벡터와 조합함으로써 시험될 수 있다. 동형접합성 표적화된 인간화를 갖는 ES 세포로부터 유래된 VelociMice는 마우스 유전자의 동형접합성 CRISPR-유도 결실을 갖는 ES 세포로부터 유래된 VelociMice와 비교될 수 있다. 녹아웃 결실이 관찰가능한 돌연변이 표현형을 야기하고 인간화된 마우스가 이러한 표현형을 발현하지 않고 대신 정상인 경우, 인간 유전자는 마우스 유전자의 생물학적 기능을 대체할 수 있다. 동형접합성 인간화된 마우스 또는 인간화된 대립유전자와 CRISPR-유도 결실 대립유전자의 복합 이형접합성(예를 들어, 반접합성) 조합을 보유하는 마우스 어느 것도 인간-특이적 치료제의 작용 및 효능의 기전을 연구하기 위해 동물 모델로서 사용될 수 있다. 복합 이형접합

성(예를 들어, 반접합성) VelociMice는 또한 종래의 교배에 의해 마우스의 인간화 세포주 및 마우스의 결실 녹아웃 세포주 둘 모두를 생성하는 데 사용될 수 있다. 따라서, 2-gRNA CRISPR 시스템을 표적화 벡터와 조합하는 단일 유전자 표적화 실험으로부터, 치료제의 전임상 시험에 유용한 마우스 모델 및 인간 약물 표적의 마우스 상동체의 생물학적 기능을 연구하기 위한 녹아웃 세포주 둘 모두를 생성하는 유전자 변형된 마우스가 생성된다.

[0175] **(1) 이중대립유전자 유전자 변형을 생성하거나, 또는 그의 빈도를 촉진하거나 증가시키는 방법**

[0176] 세포 내의 게놈에 대해 이중대립유전자 변형을 생성하거나 또는 세포 내의 게놈에 대한 이중대립유전자 변형의 빈도를 촉진하거나 증가시키기 위한 방법이 본 명세서에 제공된다. 그러한 방법은, 예를 들어, 게놈의 붕괴를 가져와서 게놈 DNA의 2개의 서열들 사이의 게놈 DNA의 큰 부분을 제거하고, 이들 2개의 서열은 후속으로 재조합될 수 있다. 그러한 방법은 또한 핵산 삽입물의 삽입 또는 게놈 DNA의 큰 부분의 결실 및 핵산 삽입물로의 대체를 가져올 수 있다.

[0177] 세포 내의 게놈을 변형시키기 위한 본 명세서에 제공된 방법은 게놈을 제1 Cas 단백질, 게놈 표적 유전자와 내의 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA, 게놈 표적 유전자와 내의 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA, 및 tracrRNA와 접촉시키는 단계를 포함한다. 선택적으로, 게놈은 게놈 표적 유전자와 내의 CRISPR RNA 인식 서열에 혼성화되는 추가의 CRISPR RNA, 예컨대 게놈 표적 유전자와 내의 제3 CRISPR RNA 인식 서열에 혼성화되는 제3 CRISPR RNA 및/또는 게놈 표적 유전자와 내의 제4 CRISPR RNA 인식 서열에 혼성화되는 제4 CRISPR RNA와 추가로 접촉될 수 있다. 게놈을 제1 Cas 단백질, 게놈 표적 유전자와 내의 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA, 게놈 표적 유전자와 내의 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA, 및 tracrRNA와 접촉시킴으로써 이중대립유전자 변형이 생성될 수 있다. 하기에 추가로 상세히 기재되는 바와 같이, Cas 단백질, CRISPR RNA, 및 tracrRNA는 임의의 형태로 그리고 임의의 수단에 의해 세포 내로 도입될 수 있다. 마찬가지로, Cas 단백질, CRISPR RNA, 및 tracrRNA의 전부 또는 일부가 동시에 또는 임의의 조합으로 순차적으로 도입될 수 있다. 게놈의 접촉은 직접적으로 일어나거나(즉, 성분이 게놈 그 자체와 직접 접촉하거나) 또는 간접적으로 일어날(즉, 성분이 게놈과 직접 접촉하는 다른 성분과 상호작용할) 수 있다.

[0178] 게놈은 게놈 표적 유전자와를 포함하는 한 쌍의 제1 및 제2 상동 염색체를 포함할 수 있다. 제1 Cas 단백질은 제1 및 제2 CRISPR RNA 인식 서열 중 하나 또는 둘 모두 내에서(즉, 제1 CRISPR RNA 인식 서열 내의 제1 절단 부위에서 그리고/또는 제2 CRISPR RNA 인식 서열 내의 제2 절단 부위에서) 이들 염색체들 중 하나 또는 둘 모두를 절단할 수 있다. 제3 및/또는 제4 CRISPR RNA가 또한 사용되는 경우, 제1 Cas 단백질은 제3 및/또는 제4 CRISPR RNA 인식 서열 중 하나 또는 둘 모두 내에서(즉, 제3 CRISPR RNA 인식 서열 내의 제3 절단 부위에서 그리고/또는 제4 CRISPR RNA 인식 서열 내의 제4 절단 부위에서) 이들 염색체들 중 하나 또는 둘 모두를 절단할 수 있다. 이어서, 절단 이벤트는 염색체들 중 하나 또는 둘 모두에서 적어도 하나의 이중-가닥 브레이크를 생성할 수 있다. 절단 이벤트는 또한 염색체들 중 하나 또는 둘 모두에서 적어도 2개의 이중-가닥 브레이크를 생성할 수 있다. Cas 니카제가 사용되는 경우, 절단 이벤트는 염색체들 중 하나 또는 둘 모두에서 적어도 하나의 단일-가닥 브레이크를 생성하거나 염색체들 중 하나 또는 둘 모두에서 적어도 2개의 단일-가닥 브레이크를 생성할 수 있다. 제3 및/또는 제4 CRISPR RNA가 사용되는 경우, 절단 이벤트는 염색체들 중 하나 또는 둘 모두에서 4개의 단일-가닥 또는 이중-가닥 브레이크 중 적어도 3개를 생성할 수 있다. 이어서, 이중-가닥 브레이크에 의해 생성된 말단 서열들은 재조합을 거칠 수 있거나, 이어서 단일-가닥 브레이크에 의해 생성된 말단 서열들은 재조합을 거칠 수 있다. 이어서, 이중대립유전자 변형을 포함하는 변형된 게놈을 갖는 세포가 확인될 수 있다.

[0179] 예를 들어, 제1 Cas 단백질은 제1 및 제2 상동 염색체에서 제1 CRISPR RNA 인식 서열 내의 제1 절단 부위에서 그리고 적어도 제1 상동 염색체에서 제2 CRISPR RNA 인식 서열 내의 제2 절단 부위에서 게놈을 절단하여, 제1 및 제2 상동 염색체에서 말단 서열들을 생성할 수 있다. 이어서, 말단 서열들은 재조합을 거쳐서, 표적화된 변형을 포함하는 이중대립유전자 변형을 갖는 게놈을 형성할 수 있다. 표적화된 변형은 적어도 제1 염색체에서의 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실을 포함할 수 있다.

[0180] 제1 및 제2 CRISPR RNA 인식 서열은 게놈 표적 유전자와 내의 어디에도 있을 수 있다. 제1 및 제2 CRISPR RNA 인식 서열은 임의의 관심 게놈 영역을 플랭킹할 수 있다. 예를 들어, 제1 및 제2 CRISPR RNA 인식 서열은 유전자에 대한 코딩 서열의 전부 또는 일부, 예컨대 *Lrp5* 유전자와, *C5(Hc)* 유전자와, *Ror1* 유전자와, 또는 *Trpa1* 유전자와를 플랭킹할 수 있다. 제1 및 제2 CRISPR RNA 인식 서열은 또한 *Cmah* 유전자에 대한 코딩 서열의 전부 또는 일부를 플랭킹할 수 있다. 대안적으로, 제1 및 제2 CRISPR RNA 인식 서열은 비-코딩 서열, 예컨대 조절 요소(예를 들어, 프로모터), 또는 코딩 및 비-코딩 서열 둘 모두를 플랭킹할 수 있다. 제3 및 제4 CRISPR RNA

인식 서열은, 예를 들어, 제1 및 제2 CRISPR RNA 인식 서열에 의해 플랭킹되는 관심 게놈 영역 내의 어디에도 있을 수 있다.

[0181] 일례로서, 제3 CRISPR RNA 인식 서열은 제1 CRISPR RNA 인식 서열에 인접할 수 있고, 제4 CRISPR RNA 인식 서열은 제2 CRISPR RNA 인식 서열에 인접할 수 있다. 따라서, 제1 및 제3 CRISPR RNA 인식 서열이 제1 쌍의 CRISPR RNA 인식 서열일 수 있고, 제2 및 제4 CRISPR RNA 인식 서열이 제2 쌍의 CRISPR RNA 인식 서열일 수 있다. 예를 들어, 제1 CRISPR RNA 인식 서열과 제3 CRISPR RNA 인식 서열(및/또는 제2 CRISPR RNA 인식 서열과 제4 CRISPR RNA 인식 서열)은 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 약 900 bp 내지 약 1 kb, 약 1 kb 내지 약 1.5 kb, 약 1.5 kb 내지 약 2 kb, 약 2 kb 내지 약 2.5 kb, 약 2.5 kb 내지 약 3 kb, 약 3 kb 내지 약 3.5 kb, 약 3.5 kb 내지 약 4 kb, 약 4 kb 내지 약 4.5 kb, 약 4.5 kb 내지 약 5 kb, 약 5 kb 내지 약 6 kb, 약 6 kb 내지 약 7 kb, 약 7 kb 내지 약 8 kb, 약 8 kb 내지 약 9 kb, 또는 약 9 kb 내지 약 10 kb만큼 분리될 수 있다. 일례로서, 제1 CRISPR RNA 인식 서열과 제3 CRISPR RNA 인식 서열(및/또는 제2 CRISPR RNA 인식 서열과 제4 CRISPR RNA 인식 서열)은 약 100 bp 내지 약 1 kb만큼 분리될 수 있다. 대안적으로, 제1 CRISPR RNA 인식 서열과 제3 CRISPR RNA 인식 서열(및/또는 제2 CRISPR RNA 인식 서열과 제4 CRISPR RNA 인식 서열)은 중첩될 수 있다.

[0182] 제1 쌍의 CRISPR RNA 인식 서열은 게놈 표적 유전자좌의 5' 말단 부근에 위치될 수 있고, 제2 쌍은 게놈 표적 유전자좌의 3' 말단 부근에 위치될 수 있다. 대안적으로, 제1 및 제2 쌍은 둘 모두 게놈 표적 유전자좌의 5' 말단 부근에 위치될 수 있거나, 둘 모두 표적 유전자좌의 3' 말단 부근에 위치될 수 있다. 대안적으로, 쌍들 중 하나 또는 둘 모두는 게놈 표적 유전자좌 내의 내부에 위치될 수 있다. 예를 들어, 제1 CRISPR RNA 인식 서열 또는 제1 쌍의 CRISPR RNA 인식 서열은 게놈 표적 유전자좌의 5' 말단으로부터 25 bp 미만, 50 bp 미만, 100 bp 미만, 150 bp 미만, 200 bp 미만, 250 bp 미만, 300 bp 미만, 350 bp 미만, 400 bp 미만, 450 bp 미만, 500 bp 미만, 600 bp 미만, 700 bp 미만, 800 bp 미만, 900 bp 미만, 1 kb 미만, 2 kb 미만, 3 kb 미만, 4 kb 미만, 5 kb 미만, 또는 10 kb 미만만큼 떨어져 있을 수 있다. 마찬가지로, 제2 CRISPR RNA 인식 서열 또는 제1 쌍의 CRISPR RNA 인식 서열은 게놈 표적 유전자좌의 3' 말단으로부터 25 bp 미만, 50 bp 미만, 100 bp 미만, 150 bp 미만, 200 bp 미만, 250 bp 미만, 300 bp 미만, 350 bp 미만, 400 bp 미만, 450 bp 미만, 500 bp 미만, 600 bp 미만, 700 bp 미만, 800 bp 미만, 900 bp 미만, 1 kb 미만, 2 kb 미만, 3 kb 미만, 4 kb 미만, 5 kb 미만, 또는 10 kb 미만만큼 떨어져 있을 수 있다.

[0183] 대안적으로, 제1 CRISPR RNA 인식 서열 또는 제1 쌍의 CRISPR RNA 인식 서열은, 예를 들어, 게놈 표적 유전자좌의 5' 말단으로부터 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb만큼 떨어져 있을 수 있다. 마찬가지로, 제2 CRISPR RNA 인식 서열 또는 제1 쌍의 CRISPR RNA 인식 서열은, 예를 들어, 게놈 표적 유전자좌의 3' 말단으로부터 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb만큼 떨어져 있을 수 있다.

[0184] 대안적으로, 제1 CRISPR RNA 인식 서열 또는 제1 쌍의 CRISPR RNA 인식 서열은, 예를 들어, 게놈 표적 유전자좌의 5' 말단으로부터 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 약 900 bp 내지 약 1 kb, 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 3 kb, 약 3 kb 내지 약 4 kb, 약 4 kb 내지 약 5 kb, 약 5 kb 내지 약 6 kb, 약 6 kb 내지 약 7 kb, 약 7 kb 내지 약 8 kb, 약 8 kb 내지 약 9 kb, 약 9 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 또는 약 90 kb 내지 약 100 kb만큼 떨어져 있을 수 있다. 마찬가지로, 제2 CRISPR RNA 인식 서열 또는 제1 쌍의 CRISPR RNA 인식 서열은, 예를 들어, 게놈 표적 유전자좌의 3' 말단으로부터 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp

내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 약 900 bp 내지 약 1 kb, 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 3 kb, 약 3 kb 내지 약 4 kb, 약 4 kb 내지 약 5 kb, 약 5 kb 내지 약 6 kb, 약 6 kb 내지 약 7 kb, 약 7 kb 내지 약 8 kb, 약 8 kb 내지 약 9 kb, 약 9 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 30 kb, 약 30 kb 내지 약 40 kb, 약 40 kb 내지 약 50 kb, 약 50 kb 내지 약 60 kb, 약 60 kb 내지 약 70 kb, 약 70 kb 내지 약 80 kb, 약 80 kb 내지 약 90 kb, 또는 약 90 kb 내지 약 100 kb만큼 떨어져 있을 수 있다.

[0185] 제1 절단 부위와 제2 절단 부위 또는 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열은, 예를 들어, 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb만큼 분리될 수 있다. 제1 절단 부위와 제2 절단 부위 또는 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열은 또한, 예를 들어, 약 3 Mb 내지 약 4 Mb, 약 4 Mb 내지 약 5 Mb, 약 5 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb만큼 분리될 수 있다. 제1 절단 부위와 제2 절단 부위 또는 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열은 또한, 예를 들어, 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 또는 약 900 bp 내지 약 1 kb만큼 분리될 수 있다. 마찬가지로, 제1 쌍의 CRISPR RNA 인식 서열은 제2 쌍의 CRISPR RNA 인식 서열로부터, 예를 들어, 약 25 bp 내지 약 50 bp, 약 50 bp 내지 약 100 bp, 약 100 bp 내지 약 150 bp, 약 150 bp 내지 약 200 bp, 약 200 bp 내지 약 250 bp, 약 250 bp 내지 약 300 bp, 약 300 bp 내지 약 350 bp, 약 350 bp 내지 약 400 bp, 약 400 bp 내지 약 450 bp, 약 450 bp 내지 약 500 bp, 약 500 bp 내지 약 600 bp, 약 600 bp 내지 약 700 bp, 약 700 bp 내지 약 800 bp, 약 800 bp 내지 약 900 bp, 약 900 bp 내지 약 1 kb, 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 약 2.5 Mb 내지 약 3 Mb, 약 3 Mb 내지 약 4 Mb, 약 4 Mb 내지 약 5 Mb, 약 5 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb만큼 분리될 수 있다.

[0186] 대안적으로, 제1 절단 부위와 제2 절단 부위 또는 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열은, 예를 들어, 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb 또는 그 이상만큼 분리될 수 있다. 마찬가지로, 제1 쌍의 CRISPR RNA 인식 서열은 제2 쌍의 CRISPR RNA 인식 서열로부터, 예를 들어, 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb 또는 그 이상만큼 분리될 수 있다.

[0187] 대안적으로, 제1 절단 부위와 제2 절단 부위 또는 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열은 25 bp 미만, 50 bp 미만, 100 bp 미만, 150 bp 미만, 200 bp 미만, 250 bp 미만, 300 bp 미만, 350 bp 미만, 400 bp 미만, 450 bp 미만, 500 bp 미만, 600 bp 미만, 700 bp 미만, 800 bp 미만, 900 bp 미만, 1 kb 미만, 2 kb 미만, 3 kb 미만, 4 kb 미만, 5 kb 미만, 또는 10 kb 미만만큼 분리될 수 있다. 마찬가지로, 제1 쌍의 CRISPR RNA 인식 서열은 제2 쌍의 CRISPR RNA 인식 서열로부터, 예를 들어, 25 bp 미만, 50 bp 미만, 100 bp 미만, 150 bp 미만, 200 bp 미만, 250 bp 미만, 300 bp 미만, 350 bp 미만, 400 bp 미만, 450 bp 미만, 500 bp 미만,

600 bp 미만, 700 bp 미만, 800 bp 미만, 900 bp 미만, 1 kb 미만, 2 kb 미만, 3 kb 미만, 4 kb 미만, 5 kb 미만, 또는 10 kb 미만만큼 분리될 수 있다.

[0188] 제1 및/또는 제2 절단 부위에서 게놈의 절단에 의해 생성된 말단 서열들은 평활 말단 또는 엇갈린 말단일 수 있고, 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실은 제1 및 제2 CRISPR RNA 인식 서열을 포함하여 이들 사이의 핵산 서열의 전부 또는 일부를 포함할 수 있다. 마찬가지로, 제3 및/또는 제4 절단 부위에서 게놈의 절단에 의해 생성된 말단 서열들은 평활 말단 또는 엇갈린 말단일 수 있다. 예를 들어, 결실은 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 핵산 서열의 단지 일부분만을 그리고/또는 제1 CRISPR RNA 인식 서열 및/또는 제2 CRISPR RNA 인식 서열의 단지 일부분만을 포함할 수 있다. 대안적으로, 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실은 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 핵산 서열의 전부를 포함할 수 있다. 마찬가지로, 결실은 제1 CRISPR RNA 인식 서열 및/또는 제2 CRISPR RNA 인식 서열, 또는 이들의 일부분을 포함할 수 있다. 일부 경우에, 결실은 제1 및 제2 CRISPR RNA 인식 서열의 외부에 위치한 서열(즉, 제1 및 제2 CRISPR RNA 인식 서열을 포함하지 않고 이들 사이에 존재하지 않는 서열)을 추가로 포함한다.

[0189] 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실은 임의의 길이일 수 있다. 예를 들어, 결실된 핵산은 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb일 수 있다.

[0190] 대안적으로, 결실된 핵산은, 예를 들어, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb 또는 그 이상일 수 있다. 일부 경우에, 결실된 핵산은 적어도 550 kb, 적어도 600 kb, 적어도 650 kb, 적어도 700 kb, 적어도 750 kb, 적어도 800 kb, 적어도 850 kb, 적어도 900 kb, 적어도 950 kb, 적어도 1 Mb, 적어도 1.5 Mb, 적어도 2 Mb, 적어도 2.5 Mb, 적어도 3 Mb, 적어도 4 Mb, 적어도 5 Mb, 적어도 10 Mb, 적어도 20 Mb, 적어도 30 Mb, 적어도 40 Mb, 적어도 50 Mb, 적어도 60 Mb, 적어도 70 Mb, 적어도 80 Mb, 적어도 90 Mb, 또는 적어도 100 Mb(예를 들어, 염색체의 대부분)일 수 있다.

[0191] 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실은 정확한 결실(precise deletion)일 수 있는데, 여기서는 결실된 핵산이 제1 Cas 단백질 절단 부위와 제2 Cas 단백질 절단 부위 사이의 핵산 서열만으로 이루어져서 변형된 게놈 표적 유전자좌에서의 추가 결실 또는 삽입이 없도록 되어 있다. 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실은 또한, 비상동 말단 결합(non-homologous end joining)(NHEJ)에 의한 부정확한 수복과 일치하는, 제1 및 제2 Cas 단백질 절단 부위를 넘어서 연장되는 부정확한 결실일 수 있으며, 그 결과 변형된 게놈 유전자좌에서의 추가 결실 및/또는 삽입이 발생할 수 있다. 예를 들어, 결실은 제1 및 제2 Cas 단백질 절단 부위를 넘어서 약 1 bp, 약 2 bp, 약 3bp, 약 4 bp, 약 5 bp, 약 10 bp, 약 20 bp, 약 30 bp, 약 40 bp, 약 50 bp, 약 100 bp, 약 200 bp, 약 300 bp, 약 400 bp, 약 500 bp, 또는 그 이상만큼 연장될 수 있다. 마찬가지로, 변형된 게놈 유전자좌는 NHEJ에 의한 부정확한 수복과 일치하는 추가 삽입, 예컨대 약 1 bp, 약 2 bp, 약 3bp, 약 4 bp, 약 5 bp, 약 10 bp, 약 20 bp, 약 30 bp, 약 40 bp, 약 50 bp, 약 100 bp, 약 200 bp, 약 300 bp, 약 400 bp, 약 500 bp, 또는 그 이상의 삽입을 포함할 수 있다.

[0192] 접촉은 외인성 공여자 서열의 부재 또는 외인성 공여자 서열의 존재 하에서 일어날 수 있되, 단, 세포가 1-세포기 배아인 경우, 외인성 공여자 서열은 5 kb 길이 이하이다. 외인성 분자 또는 서열은 세포에 정상적으로 존재하지 않는 분자 또는 서열을 포함한다. 정상적인 존재는 세포의 특정 발생기 및 환경 조건에 대한 존재를 포함한다. 외인성 분자 또는 서열은, 예를 들어, 세포 내의 상응하는 내인성 서열의 돌연변이된 버전, 예컨대 내인성 서열의 인간화 버전을 포함할 수 있다. 대조적으로, 내인성 분자 또는 서열은 특정 환경 조건 하에서 특정 발생기에서 특정 세포에 정상적으로 존재하는 분자 또는 서열을 포함한다.

[0193] 외인성 공여자 서열은 표적화 벡터 내에 있을 수 있고, 게놈 내의 5' 및 3' 표적 서열에 상응하는 5' 및 3' 상동성 아암에 의해 플랭킹된 핵산 삽입물을 포함할 수 있되, 단, 세포가 1-세포기 배아인 경우, 표적화 벡터는 5 kb 길이 이하이다. 1-세포기 배아 이외의 세포 유형에서는, 표적화 벡터가 더 길 수 있다. 1-세포기 배아 이

외의 세포 유형에서, 표적화 벡터는, 예를 들어, 본 명세서에 기재된 바와 같은 대형 표적화 벡터(LTVEC)일 수 있고, 적어도 10 kb일 수 있다. 따라서, 일부 방법에서, 게놈은 표적화 벡터와 추가로 접촉되고, 핵산 삽입물은 5' 표적 서열과 3' 표적 서열 사이에 삽입된다.

[0194] 대안적으로, 외인성 공여자 서열은 핵산 삽입물 없이 5' 및 3' 상동성 아암을 포함할 수 있다. 핵산 삽입물이 없는 그러한 표적화 벡터는 게놈 내의 5' 표적 서열과 3' 표적 서열 사이의 정확한 결실을 촉진시킬 수 있다. 그러한 정확한 결실은, 예를 들어, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 적어도 500 kb, 적어도 550 kb, 적어도 600 kb, 적어도 650 kb, 적어도 700 kb, 적어도 750 kb, 적어도 800 kb, 적어도 850 kb, 적어도 900 kb, 적어도 950 kb, 적어도 1 Mb, 적어도 1.5 Mb, 또는 적어도 2 Mb 또는 그 이상일 수 있다.

[0195] 일부 그러한 방법에서, 5' 및 3' 상동성 아암은 제1 CRISPR RNA의 제1 CRISPR RNA 인식 서열 및/또는 제2 CRISPR RNA의 제2 CRISPR RNA 인식 서열을 포함하는 게놈 표적 유전자좌에서의 5' 및 3' 표적 서열에 상응한다. 제1 및 제2 CRISPR RNA 인식 서열 또는 제1 및 제2 절단 부위는 5' 표적 서열에 인접할 수 있거나, 3' 표적 서열에 인접할 수 있거나, 또는 5' 표적 서열 또는 3' 표적 서열 어느 것에도 인접할 수 없다. 대안적으로, 제1 CRISPR RNA 인식 서열 또는 제1 절단 부위는 5' 표적 서열에 인접할 수 있고, 제2 CRISPR RNA 인식 서열 또는 제2 절단 부위는 3' 표적 서열에 인접할 수 있다. 대안적으로, 제1 CRISPR RNA 인식 서열 또는 제1 절단 부위는 5' 표적 서열 또는 3' 표적 서열 어느 것인가에 인접할 수 있고, 제2 CRISPR RNA 인식 서열 또는 제2 절단 부위는 5' 표적 서열 또는 3' 표적 서열 어느 것에도 인접할 수 없다.

[0196] 예를 들어, 제1 및/또는 제2 CRISPR RNA 인식 서열은 5' 표적 서열과 3' 표적 서열 사이에 위치될 수 있거나, 또는 5' 표적 서열 및/또는 3' 표적 서열에 인접하거나 근접하게 위치될 수 있는데, 예컨대 5' 및/또는 3' 표적 서열의 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 250 kb, 300 kb, 350 kb, 400 kb, 450 kb, 또는 500 kb 이내에 있을 수 있다. 마찬가지로, 제1 및/또는 제2 절단 부위는 5' 표적 서열과 3' 표적 서열 사이에 위치될 수 있거나, 또는 5' 표적 서열 및/또는 3' 표적 서열에 인접하거나 근접하게 위치될 수 있는데, 예컨대 5' 및/또는 3' 표적 서열의 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 250 kb, 300 kb, 350 kb, 400 kb, 450 kb, 또는 500 kb 이내에 있을 수 있다. 예를 들어, 제1 CRISPR RNA 인식 서열 또는 제1 절단 부위는 5' 표적 서열 또는 5' 및 3' 표적 서열 둘 모두의 50 bp, 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 250 kb, 300 kb, 350 kb, 400 kb, 450 kb, 또는 500 kb 이내에 있을 수 있다. 마찬가지로, 제2 CRISPR RNA 인식 서열 또는 제2 절단 부위는 3' 표적 서열 또는 5' 및 3' 표적 서열 둘 모두의 50 bp, 100 bp, 200 bp, 300 bp, 400 bp, 500 bp, 600 bp, 700 bp, 800 bp, 900 bp, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, 10 kb, 20 kb, 30 kb, 40 kb, 50 kb, 60 kb, 70 kb, 80 kb, 90 kb, 100 kb, 110 kb, 120 kb, 130 kb, 140 kb, 150 kb, 160 kb, 170 kb, 180 kb, 190 kb, 200 kb, 250 kb, 300 kb, 350 kb, 400 kb, 450 kb, 또는 500 kb 이내에 있을 수 있다.

[0197] 대안적으로, 제1 및/또는 제2 CRISPR RNA 인식 서열은 5' 및/또는 3' 표적 서열로부터 적어도 50 bp, 적어도 100 bp, 적어도 200 bp, 적어도 300 bp, 적어도 400 bp, 적어도 500 bp, 적어도 600 bp, 적어도 700 bp, 적어도 800 bp, 적어도 900 bp, 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 6 kb, 적어도 7 kb, 적어도 8 kb, 적어도 9 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb만큼 떨어져서 위치될 수 있다. 마찬가지로, 제1 및/또는 제2 절단 부위는 5' 및/또는 3' 표적 서열로부터 적어도 50 bp, 적어도 100 bp, 적어도 200 bp, 적어도 300 bp, 적어도 400 bp, 적어도 500 bp, 적어도 600 bp, 적어도 700 bp, 적어도 800 bp, 적어도 900 bp, 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도

- [0199] 대안적으로, 제1 및/또는 제2 CRISPR RNA 인식 서열 또는 제1 및/또는 제2 절단 부위는 5' 및/또는 3' 표적 서열로부터 50 bp 초과, 100 bp 초과, 200 bp 초과, 300 bp 초과, 400 bp 초과, 500 bp 초과, 600 bp 초과, 700 bp 초과, 800 bp 초과, 900 bp 초과, 1 kb 초과, 2 kb 초과, 3 kb 초과, 4 kb 초과, 5 kb 초과, 6 kb 초과, 7 kb 초과, 8 kb 초과, 9 kb 초과, 10 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과, 또는 100 kb 초과만큼 떨어져서 위치될 수 있다. 예를 들어, 제1 CRISPR RNA 인식 서열 또는 제1 절단 부위는 5' 표적 서열로부터 또는 5' 및 3' 표적 서열 둘 모두로부터 50 bp 초과, 100 bp 초과, 200 bp 초과, 300 bp 초과, 400 bp 초과, 500 bp 초과, 600 bp 초과, 700 bp 초과, 800 bp 초과, 900 bp 초과, 1 kb 초과, 2 kb 초과, 3 kb 초과, 4 kb 초과, 5 kb 초과, 6 kb 초과, 7 kb 초과, 8 kb 초과, 9 kb 초과, 10 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과, 또는 100 kb 초과만큼 떨어져서 위치될 수 있다. 마찬가지로, 제2 CRISPR RNA 인식 서열 또는 제2 절단 부위는 3' 표적 서열로부터 또는 5' 및 3' 표적 서열 둘 모두로부터 50 bp 초과, 100 bp 초과, 200 bp 초과, 300 bp 초과, 400 bp 초과, 500 bp 초과, 600 bp 초과, 700 bp 초과, 800 bp 초과, 900 bp 초과, 1 kb 초과, 2 kb 초과, 3 kb 초과, 4 kb 초과, 5 kb 초과, 6 kb 초과, 7 kb 초과, 8 kb 초과, 9 kb 초과, 10 kb 초과, 20 kb 초과, 30 kb 초과, 40 kb 초과, 50 kb 초과, 60 kb 초과, 70 kb 초과, 80 kb 초과, 90 kb 초과, 또는 100 kb 초과만큼 떨어져서 위치될 수 있다.
- [0200] 본 명세서에 기재된 방법은 이중대립유전자 변형의 빈도를 촉진시키고 증가시킨다. 특히, 게놈을 제1 및 제2 CRISPR RNA 둘 모두와 접촉시킴으로써, 게놈을 단독으로서의 제1 CRISPR RNA 또는 제2 CRISPR RNA와 접촉시키는 것과 비교하여 이중대립유전자 변형의 생성 효율이 증가될 수 있다. 이중대립유전자 변형의 생성 효율은 또한 게놈을 제1, 제2, 및 제3 CRISPR RNA, 또는 제1, 제2, 제3, 및 제4 CRISPR RNA와 접촉시킴으로써 증가될 수 있다. 이중대립유전자 변형은 (예를 들어, 이배체 세포 내의) 상응하는 상동 염색체들 상의 동일한 유전자좌에 대해 동일한 변형이 이루어진 이벤트, 또는 상응하는 상동 염색체들 상의 동일한 유전자좌에 대해 상이한 변형이 이루어진 이벤트를 포함한다. 상동 염색체들은, 동일한 유전자좌에서 동일한 유전자를 갖지만 가능하게는 상이한 대립유전자를 가질 수 있는 염색체들(예를 들어, 감수분열 동안 염기쌍이 형성되는 염색체들)을 포함한다. 용어 대립유전자는 유전자 서열의 하나 이상의 대체 형태들 중 임의의 것을 포함한다. 이배체 세포 또는 유기체에서, 주어진 서열의 2개의 대립유전자는 전형적으로 한 쌍의 상동 염색체 상의 상응하는 유전자좌를 점유한다.
- [0201] 이중대립유전자 변형은 표적화된 변형에 대해 동형접합성을 또는 표적화된 변형에 대해 복합 이형접합성(예를 들어, 반접합성)을 가져올 수 있다. 세포 집단을 사용한 단일 표적화 실험은 표적화된 변형(예를 들어, 유전자좌의 인간화)에 대해 동형접합성인 세포, 그러한 표적화된 변형에 대해 복합 이형접합성인 세포(표적화된 변형에 대해 반접합성인 세포를 포함함), 및 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이에서 동형접합성 붕괴된 세포(즉, 큰 핵산 서열이 2개의 CRISPR RNA 인식 서열들 사이에서 결실됨)를 생성할 수 있다. 동형접합성은 표적 유전자좌의 양쪽 대립유전자(즉, 양쪽 상동 염색체 상의 상응하는 대립유전자들)가 표적화된 변형을 갖는 상황을 포함한다. 복합 이형접합성은 표적 유전자좌의 양쪽 대립유전자(즉, 양쪽 상동 염색체 상의 대립유전자들)가 변형되었지만, 이들이 상이한 방식으로 변형된(예를 들어, 한쪽 대립유전자에서의 표적화된 변형 및 다른 한쪽의 대립유전자의 불활성화 또는 파괴) 상황을 포함한다. 내인성 핵산 서열의 파괴는, 예를 들어, Cas 단백질에 의해 생성된 이중-가닥 브레이크가 비상동 말단 결합(NHEJ)-매개 DNA 수복에 의해 수복될 때 발생할 수 있는데, 이는 핵산 서열의 삽입 또는 결실을 포함하는 돌연변이 대립유전자를 발생시키고, 그럼으로써 그러한 게놈 유전자좌의 파괴를 야기한다. 파괴의 예는 조절 요소(예를 들어, 프로모터 또는 인핸서)의 변경, 미스센스 돌연변이, 넌센스 돌연변이, 프레임-시프트 돌연변이, 절단(truncation) 돌연변이, 널(null) 돌연변이, 또는 소수의 뉴클레오타이드의 삽입 또는 결실(예를 들어, 프레임시프트 돌연변이를 야기함)을 포함한다. 파괴는 대립유전자의 불활성화(즉, 기능의 상실) 또는 상실을 가져올 수 있다.
- [0202] 예를 들어, 이중대립유전자 변형은, 세포가 표적화된 변형을 갖는 한쪽 대립유전자 및 발현될 수 없거나 달리 기능적이지 않은 다른 한쪽의 대립유전자를 갖는 경우, 복합 이형접합성을 가져올 수 있다. 복합 이형접합성은 반접합성을 포함한다. 반접합성은 표적 유전자좌의 단지 한쪽 대립유전자(즉, 2개의 상동 염색체들 중 하나의 상동 염색체 상의 대립유전자)만이 존재하는 상황을 포함한다. 예를 들어, 이중대립유전자 변형은, 표적화된 변형이 한쪽 대립유전자에서 일어나고 다른 한쪽 대립유전자에서는 상응하는 상실 또는 결실을 갖는 경우, 표적화된 변형에 대해 반접합성을 가져올 수 있다.
- [0203] 구체적인 예에서, 이중대립유전자 변형은 한 쌍의 제1 및 제2 상동 염색체에서의 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실을 포함할 수 있다. 이들 결실은 동시에 일어날 수 있거나, 또는 결실이

처음에 제1 상동 염색체에서 일어나고, 이어서 공여자 서열로서 제1 상동 염색체를 사용하여 상동 재조합을 통해, 예컨대 유전자 변환에 의해 제2 상동 염색체 내의 하나 이상의 이중-가닥 브레이크를 수복하여 세포에 의해 동형접합성이 달성될 수 있다. 제1 및 제2 상동 염색체 내의 결실된 핵산 서열은 동일하거나, 부분 중첩되거나, 또는 상이할 수 있다. 대안적으로, 이중대립유전자 변형은 제1 상동 염색체에서의 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실 및 제2 상동 염색체에서의 상응하는 대립유전자 또는 유전자좌의 상실을 포함할 수 있다. 대안적으로, 이중대립유전자 변형은 제1 상동 염색체에서의 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실 및 제2 상동 염색체에서의 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 상응하는 대립유전자 또는 유전자좌의 파괴를 포함할 수 있다.

[0204] 공여자 서열이 사용되는 경우, 이중대립유전자 변형은 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실뿐만 아니라, 한 쌍의 제1 및 제2 상동 염색체에서의 5' 표적 서열과 3' 표적 서열 사이의 핵산 삽입물의 삽입도 포함하여, 동형접합성 변형된 게놈을 생성할 수 있다. 결실 및 삽입이 양쪽 염색체에서 동시에 일어날 수 있거나, 또는 결실 및 삽입이 처음에 제1 상동 염색체에서 일어나고, 이어서 공여자 서열로서 제1 상동 염색체를 사용하여 상동 재조합을 통해, 예컨대 유전자 변환에 의해 제2 상동 염색체 내의 이중-가닥 브레이크(들)를 수복하여 세포에 의해 동형접합성이 달성될 수 있다. 예를 들어, 어떠한 특정 이론에 의해서도 구애되고자 함이 없이, 핵산 삽입물의 삽입이 (Cas 단백질에 의한 절단에 의해 또는 그것 없이) 제1 상동 염색체에서 일어날 수 있고, 이어서 제2 상동 염색체 상에서의 Cas 단백질에 의한 절단에 의해 자극된 유전자 변환 이벤트에 의해 제2 상동 염색체가 변형될 수 있다.

[0205] 대안적으로, 이중대립유전자 변형은 복합 이형접합성 변형된 게놈을 생성할 수 있다. 예를 들어, 표적화된 변형은 제1 및 제2 상동 염색체 둘 모두에서의 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실, 및 제2 상동 염색체에서는 아니고 제1 상동 염색체에서의 핵산 삽입물의 삽입을 포함할 수 있다. 대안적으로, 표적화된 변형은 제1 상동 염색체에서의 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실뿐만 아니라 핵산 삽입물의 삽입, 및 제2 상동 염색체에서의 상응하는 대립유전자 또는 유전자좌의 불활성화 또는 붕괴를 포함할 수 있다. 대안적으로, 이중대립유전자 변형은, 표적화된 변형이 제1 상동 염색체에서의 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실뿐만 아니라 핵산 삽입물의 삽입, 및 제2 상동 염색체에서의 상응하는 대립유전자 또는 유전자좌의 상실 또는 결실을 포함할 수 있는 반접합성 변형된 게놈을 생성할 수 있다.

[0206] 동형접합성 및 복합 이형접합성(특히, 반접합성) 표적화된 유전자 변형이 유리한데, (하기에 더 상세히 기재된) 이들 변형을 갖는 유전자 변형된 동물의 제조 과정이 더 효율적이고 덜 시간-소모적일 수 있기 때문이다. 유전자를 제거하여 그의 부재의 효과를 연구하는 것과 같은 많은 상황에서, 표적화된 유전자 변형에 대해 단지 이형접합성인 것(즉, 한쪽 대립유전자에서는 변형을 갖고 다른 한쪽의 대립유전자에 대해서는 무변화임)만으로는 충분하지 않다. 종래의 표적화 전략을 사용하여, 큰 표적화된 게놈 결실에 대해 이형접합성인 F0 세대 동물이 얻어질 수 있지만, 후속으로, 그러한 결실에 대해 동형접합성인 F1 세대 동물을 생성하기 위해 이들 이형접합성 동물의 이중교배(interbreeding)가 요구된다. 이들 추가의 교배 단계는 비용이 많이 들고 시간 소모적이다. 표적화된 유전자 변형에 대해 동형접합성 또는 복합 이형접합성(특히, 반접합성)인 F0 세대 유전자 변형된 동물을 생성하는 능력은 더 적은 교배 단계가 요구되기 때문에 유의한 효율 이득 및 시간 절약을 가져온다.

[0207] (2) 유전자 변환 또는 이형접합성의 상실

[0208] 일부 방법에서, 변형시키고자 하는 게놈은 제1 대립유전자에 대해 이형접합성인 세포 내에 있고, 유전자는 제1 대립유전자에 대해 동형접합성이 되도록 변형된다. 용어 이형접합성은 게놈이 하나 이상의 상응하는 염색체 유전자좌에서 상이한 대립유전자들(예를 들어, 상동 염색체들 상의 상응하는 유전자좌들에서 상이한 대립유전자들)을 포함하는 상황을 포함한다. 용어 동형접합성은 게놈이 (예를 들어, 상응하는 상동 염색체들 상의) 상응하는 염색체 유전자좌들에서 동일한 대립유전자를 포함하는 상황을 포함한다. 일부 그러한 방법에서는, 공여자 서열로서 제1 대립유전자를 사용하여 상동 재조합, 예컨대 유전자 변환을 통해 상응하는 제2 대립유전자에서 이중-가닥 브레이크를 수복하여 세포에 의해 동형접합성이 달성될 수 있다. 전형적으로, 유전자 변환의 정도는 수백 개의 염기쌍으로 제한된다. 예를 들어, 문헌[Kasperek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 그러나, 단일 유전자좌 내의 상이한 절단 부위들에서 절단을 유도하는 쌍 형성된 가이드 RNA들의 사용은 더 긴 트랙에 걸쳐 유전자 변환 능력을 촉진시키고 향상시킬 수 있다.

[0209] 그러한 방법은 여러 상황에서 유용할 수 있다. 제1 대립유전자는 돌연변이를 포함할 수 있다. 일부 방법에서,

예를 들어, 제1 대립유전자는 원하는 표적화된 유전자 변형을 함유한다. 그러한 표적화된 유전자 변형에 대해 동형접합성을 달성함으로써, 예를 들어, 그 목적이 그러한 변형에 대해 동형접합성인 비인간 동물을 생성하는 것이라면, 유의한 시간 및 비용 절약을 가져올 수 있다. 다른 방법에서, 제1 대립유전자는 유전자의 야생형 대립유전자이며, 이는 그러한 유전자의 제2 질병-유발 대립유전자에 상응한다. 대안적으로, 제2 대립유전자는 임의의 돌연변이를 포함할 수 있다. 이어서, 이러한 방법을 사용하여, 질병-유발 대립유전자를 그의 천연 염색체 유전자좌에 있는 야생형 대립유전자로 대체하는 궁극적인 유전자 요법 목적을 달성할 수 있다.

[0210] 제1 대립유전자에 대해 이형접합성인 게놈을 변형시켜 제1 대립유전자에 대해 동형접합성이 되게 하는 일부 그러한 방법에서, 게놈은 Cas 단백질, tracrRNA, 및 제2 대립유전자 내의 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA, 및 제2 대립유전자 내의 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA와 접촉되며, 여기서 제1 대립유전자는 제1 상동 염색체 상에 있고, 제2 대립유전자는 제2 상동 염색체 상의 상응하는 유전자좌에 있다(즉, 제1 대립유전자 및 제2 대립유전자는 한 쌍의 제1 및 제2 상동 염색체 내의 상응하는 대립유전자들일 수 있음). 선택적으로, 게놈은 제2 대립유전자 내의 CRISPR RNA 인식 서열에 혼성화되는 추가 CRISPR RNA(예를 들어, 제3 CRISPR RNA, 또는 제3 및 제4 CRISPR RNA)와 접촉될 수 있다. Cas 단백질은 제1 및 제2 CRISPR RNA 인식 서열 중 하나 또는 둘 모두를(즉, 제1 CRISPR RNA 인식 서열 내의 제1 절단 부위에서 그리고/또는 제2 CRISPR RNA 인식 서열 내의 제2 절단 부위에서) 절단할 수 있다. 제1 및/또는 제2 절단 부위에서의 게놈의 절단은 게놈 DNA에서 평활 말단을 생성할 수 있거나 엇갈린 말단을 생성할 수 있다. 이어서, 절단 부위들은 제1 대립유전자와 제2 대립유전자 사이의 재조합을 통해 수복되고, 그 결과 제1 대립유전자에 대해 동형접합성인 변형된 게놈이 생성될 수 있다. 이어서, 변형된 게놈을 갖는 세포가 확인될 수 있다.

[0211] 일부 방법에서, 제1 및/또는 제2 CRISPR RNA 인식 서열은, 제1 대립유전자 내에는 아니고 제2 대립유전자 내에 위치된다. 제1 및/또는 제2 대립유전자는 야생형 대립유전자일 수 있거나, 또는 표적화된 변형 또는 야생형 대립유전자로부터의 다른 변이를 포함할 수 있다. 예를 들어, 제1 대립유전자는 원하는 표적화된 변형을 포함할 수 있고, 제2 대립유전자는 야생형 대립유전자일 수 있다. 대안적으로, 제1 대립유전자는 야생형 대립유전자일 수 있고, 제2 대립유전자는 원치 않는 변형, 예컨대 질병-유발 돌연변이를 포함할 수 있다. 일부 그러한 방법에서는, 공여자 서열로서 제1 대립 유전자를 사용하여 재조합을 통해 제2 대립유전자에서의 질병-유발 돌연변이가 교정되도록 하는 표적화된 유전자 수복 또는 표적화된 유전자 교정이 일어난다.

[0212] 제1 절단 부위와 제2 절단 부위 또는 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열은, 예를 들어, 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb만큼 분리될 수 있다.

[0213] 대안적으로, 제1 절단 부위와 제2 절단 부위 또는 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열은, 예를 들어, 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb만큼 분리될 수 있다.

[0214] 일부 방법에서, 제1 대립유전자와 제2 대립유전자 사이의 서열 차이가 약 100 bp 내지 약 200 bp, 약 200 bp 내지 약 400 bp, 약 400 bp 내지 약 600 bp, 약 600 bp 내지 약 800 bp, 약 800 bp 내지 약 1 kb, 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 3 kb, 약 4 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 20 kb, 약 20 kb 내지 약 40 kb, 약 40 kb 내지 약 60 kb, 약 60 kb 내지 약 80 kb, 약 80 kb 내지 약 100 kb, 약 100 kb 내지 약 150 kb, 약 150 kb 내지 약 200 kb, 약 200 kb 내지 약 300 kb, 약 300 kb 내지 약 400 kb, 약 400 kb 내지 약 500 kb, 약 500 kb 내지 약 1 Mb, 약 1 Mb 내지 약 1.5 Mb, 약 1.5 Mb 내지 약 2 Mb, 약 2 Mb 내지 약 2.5 Mb, 또는 약 2.5 Mb 내지 약 3 Mb에 이른다.

[0215] 다른 방법에서, 제1 대립유전자와 제2 대립유전자 사이의 차이가 적어도 100 bp, 적어도 200 bp, 적어도 300 bp, 적어도 400 bp, 적어도 500 bp, 적어도 600 bp, 적어도 700 bp, 적어도 800 bp, 적어도 800 bp, 적어도 1 kb, 적어도 2 kb, 적어도 3 kb, 적어도 4 kb, 적어도 5 kb, 적어도 6 kb, 적어도 7 kb, 적어도 8 kb, 적어도 9 kb, 적어도 10 kb, 20 kb, 적어도 30 kb, 적어도 40 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도

80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 110 kb, 적어도 120 kb, 적어도 130 kb, 적어도 140 kb, 적어도 150 kb, 적어도 160 kb, 적어도 170 kb, 적어도 180 kb, 적어도 190 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb에 이른다.

[0216] 제1 대립유전자에 대해 이형접합성인 세포에서 게놈을 변형시켜 제1 대립유전자에 대해 동형접합성이 되게 하는 다른 그러한 방법에서, 게놈은 Cas 단백질, tracrRNA, 및 제1 CRISPR RNA 인식 서열에 혼성되는 제1 비-대립유전자-특이적 CRISPR RNA와 접촉된다. 제1 대립유전자는 제1 상동 염색체 상에 있고, CRISPR RNA 인식 서열은 제2 상동 염색체 상의 제1 대립유전자에 상응하는 유전자좌에 대해 동원체 측에(즉, 동원체에 더 가까이) 있다. Cas 단백질은 제1 CRISPR RNA 인식 서열을 절단하여 이중-가닥 브레이크를 생성할 수 있다. 이어서, 재조합이 일어나서 세포를 변형시켜 제1 대립유전자에 대해 동형접합성이 되게 할 수 있다.

[0217] 선택적으로, 세포는 하나 이상의 추가 대립유전자에 대해 이형접합성이고, 제1 CRISPR RNA 인식 서열은 제2 상동 염색체 상의 하나 이상의 추가 대립유전자에 상응하는 유전자좌에 대해 동원체 측에 있고, 재조합에 의해 세포를 변형시켜 하나 이상의 추가 대립유전자에 대해 동형접합성이 되게 한다.

[0218] 선택적으로, 본 방법은 게놈을 제2 상동 염색체에서의 제1 대립유전자에 상응하는 유전자좌에 대해 동원체 측에 있는 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 비-대립유전자-특이적 CRISPR RNA와 접촉시키는 단계를 추가로 포함할 수 있으며, Cas 단백질은 제1 및 제2 CRISPR RNA 인식 서열 중 적어도 하나를 절단하여 적어도 하나의 이중-가닥 브레이크를 생성한다. 선택적으로, 본 방법은 게놈을 제2 상동 염색체 상의 제1 대립유전자에 상응하는 유전자좌에 대해 동원체 측에 있는 CRISPR RNA 인식 서열에 혼성화되는 추가 비-대립유전자-특이적 CRISPR RNA(예를 들어, 제3 CRISPR RNA, 또는 제3 및 제4 CRISPR RNA)와 접촉시키는 단계를 추가로 포함할 수 있다. 이어서, 변형된 게놈을 갖는 세포가 확인될 수 있다.

[0219] 일부 방법에서, 제1(또는 제2, 제3, 또는 제4) CRISPR RNA 인식 서열은, 제1 상동 염색체 상은 아니고 제2 상동 염색체 상에 위치된다. 제1(또는 제2, 제3, 또는 제4) CRISPR RNA 인식 부위는 동원체로부터 약 100 bp 내지 약 1 kb, 약 1 kb 내지 약 10 kb, 약 10 kb 내지 약 100 kb, 약 100 kb 내지 약 1 Mb, 약 1 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb만큼 떨어져 있을 수 있다.

[0220] 제1 대립유전자 및/또는 하나 이상의 추가 대립유전자는 돌연변이, 예컨대 표적화된 변형을 포함할 수 있다. 대안적으로, 제1 대립유전자 및/또는 하나 이상의 추가 대립유전자는 야생형 대립유전자일 수 있고, 제2 상동 염색체 상의 상응하는 유전자좌는 돌연변이, 예컨대 질병-유발 돌연변이를 포함할 수 있다. 제1 대립유전자는 제1 CRISPR RNA 인식 부위로부터 약 100 bp 내지 약 1 kb, 약 1 kb 내지 약 10 kb, 약 10 kb 내지 약 100 kb, 약 100 kb 내지 약 1 Mb, 약 1 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb만큼 떨어져 있을 수 있다. 대안적으로, 제1 대립유전자는 제1 CRISPR RNA 인식 부위로부터 적어도 100 bp, 적어도 1 kb, 적어도 10 kb, 적어도 100 kb, 적어도 1 Mb, 적어도 10 Mb, 적어도 20 Mb, 적어도 30 Mb, 적어도 40 Mb, 적어도 50 Mb, 적어도 60 Mb, 적어도 70 Mb, 적어도 80 Mb, 적어도 90 Mb, 또는 적어도 100 Mb 또는 그 이상 떨어져 있을 수 있다.

[0221] Cas 단백질은 Cas9일 수 있다. 그것은 이중-가닥 DNA의 양쪽 가닥에 대해 뉴클레아제 활성을 가질 수 있거나, 또는 그것은 니카제일 수 있다. 일부 방법에서, Cas 단백질과 제1 CRISPR RNA는 자연적으로 함께 발생되지 않는다.

[0222] 재조합은 이중-가닥 브레이크의 말단소체 측에서의(즉, 말단소체를 향하여 일어나는) 이형접합성의 상실(예를 들어, 극성 또는 지향적 유전자 변환 또는 이형접합성의 상실)을 포함할 수 있다. 이형접합성의 상실에 의해 대체되는 제2 상동 염색체의 영역이 약 100 bp 내지 약 1 kb, 약 1 kb 내지 약 10 kb, 약 10 kb 내지 약 100 kb, 약 100 kb 내지 약 1 Mb, 약 1 Mb 내지 약 10 Mb, 약 10 Mb 내지 약 20 Mb, 약 20 Mb 내지 약 30 Mb, 약 30 Mb 내지 약 40 Mb, 약 40 Mb 내지 약 50 Mb, 약 50 Mb 내지 약 60 Mb, 약 60 Mb 내지 약 70 Mb, 약 70 Mb 내지 약 80 Mb, 약 80 Mb 내지 약 90 Mb, 또는 약 90 Mb 내지 약 100 Mb일 수 있다. 대안적으로, 대체되는 제2 상동 염색체의 영역은 적어도 100 bp, 적어도 1 kb, 적어도 10 kb, 적어도 100 kb, 적어도 1 Mb, 적어도 10 Mb, 적어도 20 Mb, 적어도 30 Mb, 적어도 40 Mb, 적어도 50 Mb, 적어도 60 Mb, 적어도 70 Mb, 적어도 80 Mb, 적어도 90 Mb, 또는 적어도 100 Mb 또는 그 이상일 수 있다. 예를 들어, 염색체의 대부분이 대체될 수 있다.

[0223] **B. 유전자 변형된 비인간 동물의 제조 방법**

[0224] 유전자 변형된 비인간 동물은 본 명세서에 개시된 다양한 방법을 사용하여 생성될 수 있다. 일부 경우에, 유전자 변형된 비인간 동물을 생성하는 방법은 (1) 전술된 방법을 사용하여 다능성 세포의 계능을 변형시키는 단계; (2) 유전자 변형된 다능성 세포를 선택하는 단계; (3) 유전자 변형된 다능성 세포를 숙주 배아 내로 도입하는 단계; 및 (4) 유전자 변형된 다능성 세포를 포함하는 숙주 배아를 대리모 내로 이식하는 단계를 포함한다. 유전자 변형된 다능성 세포로부터의 후손이 생성된다. 공여자 세포는 임의의 단계, 예컨대 배반포기(blastocyst stage) 또는 전상실배기(pre-morula stage)(즉, 4-세포기 또는 8-세포기)에 숙주 배아 내로 도입될 수 있다. 생식선(germline)을 통해 유전자 변형을 전달할 수 있는 후손이 생성된다. 다능성 세포는 본 명세서의 어딘가 다른 곳에 논의된 바와 같은 ES 세포(예를 들어, 마우스 ES 세포 또는 래트 ES 세포)일 수 있다. 예를 들어, 미국 특허 제7,294,754호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0225] 대안적으로, 유전자 변형된 비인간 동물을 생성하는 방법은 (1) 전술된 방법을 사용하여 1-세포기 배아의 계능을 변형시키는 단계; (2) 유전자 변형된 배아를 선택하는 단계; 및 (3) 유전자 변형된 배아를 대리모 내로 이식하는 단계를 포함할 수 있다. 생식선을 통해 유전자 변형을 전달할 수 있는 후손이 생성된다.

[0226] 핵 전달(nuclear transfer) 기술이 또한 비인간 포유류 동물을 생성하는 데 사용될 수 있다. 간략하게 말하면, 핵 전달 방법은 (1) 난모세포(oocyte)를 제핵하거나(enucleating) 제핵된 난모세포를 제공하는 단계; (2) 제핵된 난모세포와 조합될 공여자 세포 또는 핵을 단리하거나 제공하는 단계; (3) 세포 또는 핵을 제핵된 난모세포에 삽입하여 재구성된 세포를 형성하는 단계; (4) 재구성된 세포를 동물의 자궁 내로 이식하여 배아를 형성하는 단계; 및 (5) 배아가 발생되게 하는 단계를 포함할 수 있다. 그러한 방법에서, 난모세포는 일반적으로 죽은 동물로부터 회수되긴 하지만, 또한 살아있는 동물의 난관 및/또는 난소로부터 단리될 수 있다. 난모세포는 제핵 전에 당업자에게 공지된 다양한 배치에서 성숙될 수 있다. 난모세포의 제핵은 당업자에게 널리 공지된 다수의 방식으로 수행될 수 있다. 공여자 세포 또는 핵을 제핵된 난모세포 내로 삽입하여 재구성된 세포를 형성하는 것은 융합 전에 투명대(zona pellucida) 아래의 공여자 세포의 미세주입(microinjection)에 의해 행해질 수 있다. 융합은 접촉/융합 평면(전기융합)을 가로지르는 DC 전기적 펄스의 적용에 의해, 폴리에틸렌 글리콜과 같은 융합-촉진 화합물질에 대한 세포의 노출에 의해, 또는 센다이(Sendai) 바이러스와 같은 불활성화된 바이러스에 의해 유도될 수 있다. 재구성된 세포는 핵 공여자 및 수용자 난모세포의 융합 전, 융합 동안, 및/또는 융합 후 전기적 및/또는 비-전기적 수단에 의해 활성화될 수 있다. 활성화 방법은 전기적 펄스, 화학적으로 유도된 쇼크, 정자에 의한 침투, 난모세포에서 2가 양이온의 수준의 증가, 및 난모세포에서 (키나제 억제제에 의한) 세포 단백질의 인산화 감소를 포함한다. 활성화된 재구성된 세포, 또는 배아는 당업자에게 널리 공지된 배치에서 배양되고 이어서 동물의 자궁으로 전달될 수 있다. 예를 들어, 미국 특허 출원 공개 제20080092249호, 국제특허 공개 WO/1999/005266A2호, 미국 특허 출원 공개 제20040177390호, 국제특허 공개 WO/2008/017234A1호, 및 미국 특허 제7,612,250호를 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0227] 유전자 변형된 비인간 동물을 제조하는 일부 방법은 F0 세대 비인간 동물을 생성하는 방법을 포함한다. 그러한 방법은 비인간 ES 세포 내의 계능을 Cas 단백질, 제1 CRISPR RNA 인식 서열에 혼성화되는 제1 CRISPR RNA, 제2 CRISPR RNA 인식 서열에 혼성화되는 제2 CRISPR RNA, 및 tracrRNA와 접촉시키는 단계를 포함할 수 있다. Cas 단백질은 제1 및 제2 CRISPR RNA 인식 서열 내의 계능을 절단하여 말단 서열들을 생성할 수 있다. 말단 서열들은 재조합을 거쳐서 표적화된 변형을 갖는 계능을 형성할 수 있고, 표적화된 변형은 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의 결실을 포함할 수 있다.

[0228] 본 방법은 (1) 표적화된 변형을 포함하는 비인간 ES 세포를 확인하는 단계; (2) 표적화된 변형을 포함하는 비인간 ES 세포를 비인간 숙주 배아 내로 도입하는 단계; 및 (3) 대리모에서 비인간 숙주 배아를 잉태시키는 단계를 추가로 포함할 수 있다. 이어서, 대리모는 표적화된 변형을 포함하는 F0 세대 비인간 동물을 생성할 수 있다. 유전자 변형된 다능성 또는 전능성(totipotent) 세포(예를 들어, 비인간 ES 세포)를 포함하는 숙주 배아는 배반포기까지 인큐베이션된 후 대리모 내로 이식되어 F0 동물을 생성할 수 있다. 유전자 변형된 계능 유전자좌를 함유하는 동물은 본 명세서에 기재된 바와 같은 대립유전자의 변형(MOA) 검정을 통해 확인될 수 있다.

[0229] 본 명세서에 제공된 다양한 방법은 유전자 변형된 비인간 F0 동물의 생성을 가능하게 하는데, 여기서 유전자 변형된 F0 동물의 세포는 표적화된 변형을 포함한다. F0 동물을 생성하기 위해 사용되는 방법에 따라, 관심 뉴클레오티드 서열을 갖고 재조합효소 카세트 및/또는 선택 카세트가 결합되어 있는 F0 동물 내의 세포의 수는 다양할 것으로 인식된다. 공여자 ES 세포를, 예를 들어 VELOCIMOUSE® 방법을 통해 상응하는 유기체로부터의 전상

실배기 배아(예를 들어, 8-세포기 마우스 배아) 내로 도입함으로써, F0 동물의 더 큰 백분율의 세포 집단이, 표적화된 유전자 변형을 포함하는 관심 뉴클레오티드 서열을 갖는 세포를 포함할 수 있게 된다. 구체적인 경우에, 비인간 F0 동물의 세포 기여(cellular contribution)의 적어도 50%, 60%, 65%, 70%, 75%, 85%, 86%, 87%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%는 표적화된 변형을 갖는 세포 집단을 포함한다. 다른 경우에, F0 동물의 생식 세포들 중 적어도 하나 또는 그 이상은 표적화된 변형을 갖는다.

[0230] 일부 경우에, 유전자 변형된 F0 동물의 세포는 표적화된 변형에 대해 이형접합성 또는 복합 이형접합성이다. 예를 들어, 유전자 변형된 F0 동물의 세포는 표적화된 변형에 대해 반접합성일 수 있다. 다른 경우에, 유전자 변형된 F0 동물의 세포는 표적화된 변형에 대해 동형접합성이다.

[0231] 일부 경우에, 본 명세서에 개시된 방법 및 조성물에 의해 생성된 F0 동물은 야생형 동물과 교배되어, 표적화된 변형에 대해 이형접합성인 F1 세대를 생성할 수 있다. 이어서, F1 세대로부터의 동물을 서로 교배하여, 표적화된 변형에 대해 동형접합성인 F2 동물을 생성할 수 있다. F1 자손은, 특정 프라이머 및/또는 프로브를 사용하여 표적화된 유전자 변형이 존재하는지를 결정하여, 유전자형 결정(genotyping)이 행해질 수 있다.

[0232] C. 게놈 및 표적 게놈 유전자좌

[0233] 본 명세서에 개시된 방법에 의해 변형된 게놈 또는 게놈 표적 유전자좌는 세포 내의 DNA의 임의의 세그먼트 또는 영역을 포함할 수 있다. 게놈 또는 게놈 표적 유전자좌는 세포에 대해 천연일 수 있거나, 세포의 게놈 내로 통합된 DNA의 이종성 또는 외인성 세그먼트일 수 있거나, 이들의 조합일 수 있다. 그러한 DNA의 이종성 또는 외인성 세그먼트는 도입유전자, 발현 카세트, 선택 마커를 인코딩하는 폴리뉴클레오티드, 또는 게놈 DNA의 이종성 또는 외인성 영역을 포함할 수 있다.

[0234] 게놈 또는 게놈 표적 유전자좌는 또한 세포 내의 염색체의 DNA, 예컨대 효모 인공 염색체(YAC), 세균 인공 염색체(BAC), 인간 인공 염색체, 또는 적절한 숙주 세포에 함유된 임의의 다른 조각된 게놈 영역을 포함할 수 있다.

[0235] D. Cas9 및 가이드 RNA의 형태

[0236] 일부 방법에서, 게놈을 접촉시키는 단계는 하나 이상의 Cas 단백질, 하나 이상의 CRISPR RNA, 및 하나 이상의 tracrRNA를 세포 내로 도입하는 단계를 포함한다. 도입은 임의의 수단에 의해 달성될 수 있고, 성분들 중 하나 이상(예를 들어, 성분들 중 2개, 또는 성분들 전부)이 동시에 또는 임의의 조합으로 순차적으로 세포 내로 도입될 수 있다.

[0237] CRISPR RNA와 tracrRNA는 가이드 RNA(gRNA)로서 함께 융합되어 세포 내로 도입될 수 있다. 대안적으로, CRISPR RNA 및 tracrRNA는 별개의 RNA 분자일 수 있다. CRISPR RNA는 RNA의 형태로 또는 CRISPR RNA를 인코딩하는 DNA의 형태로 세포 내로 도입될 수 있다. 마찬가지로, tracrRNA는 RNA의 형태로 또는 tracrRNA를 인코딩하는 DNA의 형태로 세포 내로 도입될 수 있고, gRNA는 RNA의 형태로 또는 gRNA를 인코딩하는 DNA의 형태로 세포 내로 도입될 수 있다.

[0238] Cas 단백질은 단백질, Cas 단백질을 인코딩하는 메신저 RNA(mRNA), 또는 Cas 단백질을 인코딩하는 DNA의 형태로 세포 내로 도입될 수 있다. 일부 방법에서, Cas 단백질, CRISPR RNA, 및 tracrRNA는 단백질-RNA 복합체로서 세포 내로 도입될 수 있다. 마찬가지로, Cas 단백질과 gRNA는 단백질-RNA 복합체로서 세포 내로 도입될 수 있다. Cas 단백질은 세포-투과성 Cas 단백질(예를 들어, 세포-투과 도메인을 갖는 Cas 단백질)일 수 있다.

[0239] Cas 단백질, CRISPR RNA, 또는 tracrRNA를 인코딩하는 DNA는 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 그러한 DNA는 하나 이상의 발현 작제물 내에 존재할 수 있다. 일부 방법에서, 그러한 발현 작제물들 중 하나 이상은 단일 핵산 분자의 성분들일 수 있다. 예를 들어, 하나 이상의 Cas 단백질을 인코딩하는 DNA, 하나 이상의 CRISPR RNA를 인코딩하는 DNA, 및 하나 이상의 tracrRNA를 인코딩하는 DNA는 모두 단일 핵산 분자의 성분들일 수 있다. 대안적으로, 이들은 2개, 3개, 4개, 또는 그 이상의 핵산 분자들 중에서 임의의 조합으로 분리될 수 있다.

[0240] 유사하게, Cas 단백질을 인코딩하는 DNA 또는 gRNA를 인코딩하는 DNA는 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 그러한 DNA는 또한 하나 이상의 발현 작제물 내에 존재할 수 있다. 일부 방법에서, 그러한 발현 작제물들 중 하나 이상은 단일 핵산 분자의 성분들일 수 있다. 예를 들어, 하나 이상의 Cas 단백질을 인코딩하는 DNA 및 하나 이상의 gRNA를 인코딩하는 DNA는 모두 단일 핵산 분자의 성분들일 수 있다. 대안적으로, 이들은 2개, 3개, 4개, 또는 그 이상의 핵산 분자들 중에서 임의의 조합으로 분리될 수 있다.

- [0241] 일부 방법에서, Cas 단백질과 CRISPR RNA 및/또는 tracrRNA는 자연적으로 함께 발생되지 않는다. 일부 방법에서, 예를 들어, Cas 단백질과 제1 CRISPR RNA는 자연적으로 함께 발생되지 않고/않거나, Cas 단백질과 제2 CRISPR RNA는 자연적으로 함께 발생되지 않고/않거나, Cas 단백질과 tracrRNA는 자연적으로 함께 발생되지 않는다.
- [0242] 일부 방법에서, Cas 단백질은 Cas9 단백질이다. Cas 단백질은 이중 폴리펩티드, 예컨대 핵 국제화 신호(NLS)에 융합될 수 있다. Cas 단백질은 완전 절단 활성을 갖고 게놈 DNA 내에 이중-가닥 브레이크(예를 들어, 평할 말단을 갖는 이중-가닥 브레이크)를 생성할 수 있거나, 또는 그것은 게놈 DNA의 가닥만을 절단할 수 있는 니카제일 수 있다.
- [0243] 일부 방법에서, 쌍 형성된 니카제가 사용될 수 있다. 예를 들어, 게놈은 DNA의 대향 가닥들을 절단하는 제1 및 제2 니카제와 접촉되어, 게놈은 이중 니킹(double nicking)을 통해 변형될 수 있다. 제1 니카제는 게놈 DNA의 제1 가닥(즉, 상보성 가닥)을 절단할 수 있고, 제2 니카제는 게놈 DNA의 제2 가닥(즉, 비상보성 가닥)을 절단할 수 있다. 대안적으로, 두 니카제 모두가 동일한 가닥을 절단할 수 있다. 제1 및 제2 니카제는, 예를 들어, 제1 니카제의 RuvC 도메인 내의 촉매 잔기를 돌연변이시키고(예를 들어, 본 명세서의 어딘가 다른 곳에 기재된 D10A 돌연변이) 제2 니카제의 HNH 도메인 내의 촉매 잔기를 돌연변이시킴으로써(예를 들어, 본 명세서의 어딘가 다른 곳에 기재된 H840A 돌연변이) 생성될 수 있다. 대안적으로, 제1 니카제가 두 니카제를 생성하는 데 사용될 수 있다.
- [0244] 일부 그러한 방법에서, 이중 니킹은 엇갈린 말단을 갖는 하나 이상의 이중-가닥 브레이크를 생성하는 데 사용될 수 있다. 예를 들어, 이중 니킹은 제1 및 제2 절단 부위에서 엇갈린 말단을 생성하는 데 사용된다. 제1 니카제는 제1 및 제2 CRISPR RNA가 혼성화되는 제1 및 제2 CRISPR RNA 인식 서열 내의 DNA의 제1 가닥을 절단할 수 있고, 제2 니카제는 제3 및 제4 CRISPR RNA가 혼성화되는 제3 및 제4 표적 CRISPR RNA 인식 서열 내의 DNA의 제2 가닥을 절단할 수 있다. 대안적으로, 제1 니카제는 제1, 제2, 제3, 및 제4 CRISPR RNA 인식 서열을 니킹하는 데 사용될 수 있다. 제1 및 제3 표적 CRISPR RNA 인식 서열은, DNA의 제1 및 제2 가닥 상에 제1 및 제2 니카제에 의해 생성된 니이 이중-가닥 브레이크를 생성하도록 하는 제1 절단 부위를 생성하도록 위치될 수 있다(즉, 제1 절단 부위는 제1 및 제3 CRISPR RNA 인식 서열 내의 니를 포함함). 마찬가지로, 제2 및 제4 표적 CRISPR RNA 인식 서열은, DNA의 제1 및 제2 가닥 상에 제1 및 제2 니카제에 의해 생성된 니이 이중-가닥 브레이크를 생성하도록 하는 제2 절단 부위를 생성하도록 위치될 수 있다(즉, 제2 절단 부위는 제2 및 제4 CRISPR RNA 인식 서열 내의 니를 포함함). 일부 경우에, 제1 CRISPR RNA 인식 서열과 제3 CRISPR RNA 인식 서열 및/또는 제2 CRISPR RNA 인식 서열과 제4 CRISPR RNA 인식 서열 내의 니는 오프-셋(off-set) 니일 수 있다. 오프셋 차이는, 예를 들어, 적어도 약 5 bp, 10 bp, 20 bp, 30 bp, 40 bp, 50 bp, 60 bp, 70 bp, 80 bp, 90 bp, 100 bp 또는 그 이상일 수 있다. 문헌[Ran *et al.* (2013) *Cell* 154:1380-1389]; 문헌[Mali *et al.* (2013) *Nat. Biotech.* 31:833-838]; 및 문헌[Shen *et al.* (2014) *Nat. Methods* 11:399-404]을 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.
- [0245] **E. 핵산 및 단백질을 세포 내로 도입하는 방법**
- [0246] 세포 내로의 핵산의 도입을 가능하게 하는 다양한 방법 및 조성물이 본 명세서에 제공된다. 일부 경우에, 핵산의 도입을 위해 사용된 시스템은 특정 게놈 유전자좌에서의 표적화된 통합을 가능하게 한다. 그러한 시스템은 다양한 성분들을 사용하고, 참조 용이성을 위하여, 용어 "표적화된 게놈 통합 시스템"은 일반적으로 통합 이벤트에 필요한 모든 성분(예를 들어, 하나 이상의 뉴클레아제 작용제, 뉴클레아제 절단 부위, 삽입 DNA 폴리뉴클레오티드, 표적화 벡터, 표적 게놈 유전자좌, 및 관심 폴리뉴클레오티드)를 포함한다.
- [0247] 본 명세서에 제공된 방법은 표적화된 게놈 통합 시스템의 하나 이상의 성분들을 포함하는 하나 이상의 폴리뉴클레오티드 또는 폴리펩티드 작제물을 세포 내로 도입하는 단계를 포함할 수 있다. "도입하는"은 서열이 세포의 내부에 대한 접근을 얻게 하는 방식으로 서열(폴리펩티드 또는 폴리뉴클레오티드)을 세포에 제시하는 것을 포함한다. 본 명세서에 제공된 방법은, 핵산 또는 단백질이 적어도 하나의 세포의 내부에 대한 접근을 획득하기만 한다면, 핵산 또는 단백질을 세포 내로 도입하기 위한 특정 방법에 의존하지 않는다. 다양한 세포 유형 내로 핵산 및 단백질을 도입하기 위한 방법은 당업계에 공지되어 있고, 예를 들어, 안정한 형질감염 방법, 일시적 형질감염 방법, 및 바이러스-매개 방법을 포함한다.
- [0248] 일부 경우에, 본 방법 및 조성물에 사용된 세포는 그의 게놈 내로 안정하게 혼입된 DNA 작제물을 갖는다. 예를 들어, 본 명세서에 개시된 방법에 사용된 세포는 그의 게놈 내로 안정하게 혼입된 기존의 Cas-인코딩 유전자를 가질 수 있다(즉, Cas-준비된(ready) 세포). "안정하게 혼입된" 또는 "안정하게 도입된"은, 뉴클레오티드 서열

이 세포의 게놈 내로 통합되어 이의 자손에 의해 유전될 수 있도록 세포 내로의 폴리뉴클레오티드의 도입을 포함한다. 임의의 프로토콜이 DNA 작제물 또는 표적화된 게놈 통합 시스템의 다양한 성분들의 안정한 혼입을 위해 사용될 수 있다.

[0249] 형질감염 프로토콜 및 폴리펩티드 또는 폴리뉴클레오티드 서열의 세포 내로의 도입을 위한 프로토콜은 다양할 수 있다. 비제한적인 형질감염 방법은 리포솜; 나노입자; 인산칼슘(문헌[Graham *et al.* (1973) *Virology* 52 (2): 456-67], 문헌[Bacchetti *et al.* (1977) *Proc Natl Acad Sci USA* 74 (4): 1590-4], 및 문헌[Kriegler, M (1991). *Transfer and Expression: A Laboratory Manual*. New York: W. H. Freeman and Company. pp. 96-97]); 덴드리머(dendrimer); 또는 양이온성 중합체, 예컨대 DEAE-텍스트란 또는 폴리에틸렌 이민을 사용하는 화학적-기반 형질감염 방법을 포함한다. 비화학적 방법은 전기천공, 초음파-천공(Sono-poration), 및 광학적 형질감염을 포함한다. 입자-기반 형질감염은 유전자 총(gun), 또는 자석-보조 형질감염(문헌[Bertram (2006) *Current Pharmaceutical Biotechnology* 7, 277-28])의 사용을 포함한다. 바이러스 방법이 또한 형질감염에 사용될 수 있다.

[0250] 일부 경우에, 세포 내로의 핵산 또는 단백질의 도입은 전기천공에 의해, 세포질내 주입(intracytoplasmic injection)에 의해, 바이러스 감염에 의해, 아테노바이러스에 의해, 렌티바이러스에 의해, 레트로바이러스에 의해, 형질감염에 의해, 지질-매개 형질감염에 의해, 또는 Nucleofection™에 의해 매개된다.

[0251] 세포(예를 들어, 1-세포기 배아) 내로의 핵산 또는 단백질의 도입은 또한 미세주입에 의해 달성될 수 있다. 1-세포기 배아에서, 미세주입은 모계 및/또는 부계 전핵(pronucleus) 내로 또는 세포질 내로 행해질 수 있다. 미세주입이 단지 하나의 전핵 내로 행해진다면, 부계 전핵이 그의 더 큰 크기 때문에 바람직하다. mRNA의 미세주입은 바람직하게는 (예를 들어, mRNA를 번역 기구(translation machinery)에 직접 전달하기 위해) 세포질 내로 행해지며, 한편 Cas 단백질이나, Cas 단백질을 인코딩하거나 RNA를 인코딩하는 핵산 분자의 미세주입은 바람직하게는 핵/전핵 내로 행해진다. 대안적으로, 미세주입은 핵/전핵 및 세포질 둘 모두 내로의 주입에 의해 수행될 수 있다: 바늘이 먼저 핵/전핵 내로 도입될 수 있고 제1 양이 주입될 수 있고, 1-세포기 배아로부터 바늘을 제거하는 동안, 제2 양이 세포질 내로 주입될 수 있다. Cas 단백질이 세포질 내로 주입되는 경우, Cas 단백질은 바람직하게는 핵 국제화 신호를 포함하여 핵/전핵으로의 전달을 보장한다. 미세주입을 수행하기 위한 방법은 잘 알려져 있다. 예를 들어, 문헌[Nagy *et al.* (Nagy A, Gertsenstein M, Vintersten K, Behringer R., 2003, *Manipulating the Mouse Embryo*. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press)]; 문헌[Meyer *et al.* (2010) *Proc Natl Acad Sci USA* 107:15022-15026] 및 문헌[Meyer *et al.* (2012) *Proc Natl Acad Sci USA* 109:9354-9359]을 포함하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0252] 세포 내로의 핵산 또는 단백질의 도입은 일정 기간에 걸쳐 1회 또는 다수회 수행될 수 있다. 예를 들어, 도입은 일정 기간에 걸쳐 적어도 2회, 일정 기간에 걸쳐 적어도 3회, 일정 기간에 걸쳐 적어도 4회, 일정 기간에 걸쳐 적어도 5회, 일정 기간에 걸쳐 적어도 6회, 일정 기간에 걸쳐 적어도 7회, 일정 기간에 걸쳐 적어도 8회, 일정 기간에 걸쳐 적어도 9회, 일정 기간에 걸쳐 적어도 10회, 적어도 11회, 일정 기간에 걸쳐 적어도 12회, 일정 기간에 걸쳐 적어도 13회, 일정 기간에 걸쳐 적어도 14회, 일정 기간에 걸쳐 적어도 15회, 일정 기간에 걸쳐 적어도 16회, 일정 기간에 걸쳐 적어도 17회, 일정 기간에 걸쳐 적어도 18회, 일정 기간에 걸쳐 적어도 19회, 또는 일정 기간에 걸쳐 적어도 20회 수행될 수 있다.

[0253] 뉴클레아제 작용제 및 표적화 벡터(예를 들어, 1-세포기 배아 이외의 세포에 대한 LTVEC) 둘 모두가 세포 내로 도입되는 경우, 이들은 동시에 도입될 수 있다. 대안적으로, 뉴클레아제 작용제는 표적화 벡터와 별개로 도입될 수 있다. 예를 들어, 뉴클레아제 작용제는 표적화 벡터의 도입 전에 도입될 수 있거나, 이것은 표적화 벡터의 도입 후에 도입될 수 있다.

[0254] **F. 재조합의 기전, 및 비상동 말단 결합, 유전자 변환, 또는 상동 재조합의 출현율(prevalence)을 변경시키기 위한 방법**

[0255] 재조합은 2개의 폴리뉴클레오티드들 사이의 유전 정보의 임의의 교환 과정을 포함하고, 임의의 기전에 의해 일어날 수 있다. 이중-가닥 브레이크(DSB)에 따른 재조합은 하기의 2개의 보존된 DNA 수복 경로를 통해 주로 일어난다: 비상동 말단 결합(NHEJ) 및 상동 재조합(HR). 예를 들어, 문헌[Kasperek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. NHEJ는 상동 주형에 대한 필요성 없이 브레이크 말단들의 서로에 대한 직접 연결에 의한 핵산 내의 이중-가닥 브레이크의 수복을 포함한다. NHEJ에 의한 비인접 서열들의 연결은 종종 이중-가닥 브레이크 부위 부근

에서 결실, 삽입, 또는 전좌를 야기할 수 있다. 재조합은 또한 상동성 유도 수복(HDR) 또는 상동 재조합(HR)을 통해 일어날 수 있다. HDR 또는 HR은 뉴클레오타이드 서열 상동성을 필요로 할 수 있는 핵산 수복의 형태를 포함하고, "표적" 분자(즉, 이중-가닥 브레이크를 겪은 것)의 주형 수복을 위해 "공여자" 분자를 사용하고, 공여자로부터 표적으로의 유전 정보의 전달로 이어진다. 어떠한 특정 이론에 의해서도 구애되고자 함이 없이, 그러한 전달은 브레이킹된 표적과 공여자 사이에 형성되는 헤테로듀플렉스 DNA의 불일치 교정, 및/또는 합성-의존적 가닥 어닐링 - 여기서, 공여자는 표적의 일부가 될 유전 정보를 재합성하는 데 사용됨 -, 및/또는 관련 과정을 포함할 수 있다. 일부 경우에, 공여자 폴리뉴클레오타이드, 공여자 폴리뉴클레오타이드의 일부분, 공여자 폴리뉴클레오타이드의 카피, 또는 공여자 폴리뉴클레오타이드의 카피의 일부분이 표적 DNA 내로 통합된다.

[0256] 대립유전자에 대해 이형접합성인 세포의 게놈을 변형시켜 그러한 대립유전자에 대해 동형접합성이 되게 하는 것과 관련하여, 재조합은 이형접합성 세포로부터 동형접합성 세포를 유도하는 임의의 수단을 포함할 수 있다. 그러한 수단은, 예를 들어, 이형접합성의 상실(LOH), 유전자 변환, 또는 임의의 공지된 재조합 기전에 의해 일어나는 교차 이벤트를 포함할 수 있다. 이론에 의해 구애되고자 함이 없이, LOH는, 예를 들어, 유전자 변환이 있거나 없이 유사분열 재조합을 통해 일어나거나, 또는 염색체 상실 및 중복을 통해 일어날 수 있다. 문헌[Lefebvre *et al.* (2001) *Nat. Genet.* 27:257-258]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 이와 관련하여 유전자 변환은 공여자 서열로부터 고도로 상동성인 수용자로서의 유전 물질의 단방향 전달(즉, 하나의 분자로부터 그의 상동체로의 유전 정보의 비상호 교환)을 포함할 수 있다. 유전자 변환은 임의의 공지된 재조합 기전에 의해 대립유전자를 카피하기 위한 임의의 수단을 포함한다. 예를 들어, 유전자 변환은 온전한 서열로부터 이중-가닥 브레이크를 함유하는 상동성 영역으로의 유전 정보의 비상호 전달을 수반할 수 있고, 그것은 동일한 염색분체 상의 또는 상이한 염색체들 상의 자매 염색분체들, 상동 염색체들, 또는 상동성 서열들 사이에서 일어날 수 있다. 문헌[Chen *et al.* (2007) *Nat. Rev. Genet.* 8:762-775]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 특정한 경우에, 유전자 변환은 상동 염색체로부터의 유전 정보의 카피의 결과로서 상동 재조합으로부터 직접 발생된다. 상동성 서열들이 동일하지 않은 경우, 이것은 국제화된 이형접합성의 상실(LOH)로 이어질 수 있다.

[0257] 일례로서, LOH는 유사분열 교차에 의한 상호 염색분체 교환을 통해, 또는 브레이크-유도 복제에 의한 염색분체 카피에 의해 일어날 수 있다. 어느 경우이든, 이형접합성 변형이 일어날 수 있으며, 이러한 변형에서는 게놈 복제 전에 하나의 염색체가 표적화된다. 대안적으로, 게놈 복제 후에 단일 염색분체가 표적화될 수 있으며, 이후에 염색분체간 유전자 변환이 일어날 수 있다.

[0258] 본 명세서에 개시된 방법들 중 임의의 것에서, 세포는 NHEJ 활성을 증가 또는 감소시키도록 변형된 세포일 수 있다. 마찬가지로, 세포는 유전자 변환 또는 HDR 활성을 증가시키도록 변형된 세포일 수 있다. 그러한 변형은 NHEJ, 유전자 변환, 및/또는 HDR을 조절하는 데 관여하는 유전자의 발현 또는 활성의 변형을 포함할 수 있다. 예를 들어, NHEJ의 활성의 감소 및/또는 HDR의 활성의 증가는 2개의 gRNA에 상응하는 CRISPR RNA 인식 서열들 사이의 게놈 영역의 이중대립유전자 붕괴를 촉진시킬 수 있다. 어떠한 특정 이론에 의해서도 구애되고자 함이 없이, 이중대립유전자 게놈 붕괴가 일어날 수 있는 하나의 기전은 제1 대립유전자 내에서의 NHEJ-매개 수복 또는 HDR-매개 수복 및 HDR 기전, 예컨대 유전자 변환을 통한 동일한 제2 대립유전자의 생성에 의한 것이다(실시예 1 참조). 따라서, (예를 들어, NHEJ 활성을 감소시키거나 HDR 활성을 증가시켜) HDR-매개 경로를 촉진시킴으로써 또한 게놈 영역의 이중대립유전자 붕괴를 촉진시킬 수 있다. 유사하게, 어떠한 특정 이론에 의해서도 구애되고자 함이 없이, 단일 유전자좌를 표적화하는 쌍 형성된 가이드 RNA를 사용함으로써 이형접합성 세포를 동형접합성 세포로 변환시키는 것은, NHEJ 활성이 감소되고 HDR 활성(예를 들어, 유전자 변환 활성)이 상응하여 증가된다면, 촉진될 수 있다.

[0259] NHEJ 활성을 증가 또는 감소시키는 데 또는 HDR 활성을 증가 또는 감소시키는 데 억제제가 사용될 수 있다. 그러한 억제제는, 예를 들어, 소분자 또는 억제성 핵산, 예컨대 짧은 간섭 핵산(예를 들어, 짧은 간섭 RNA(siRNA), 이중-가닥 RNA(dsRNA), 마이크로-RNA(miRNA), 및 짧은 헤어핀 RNA(shRNA)) 또는 유전자 전사체에 특이적인 안티센스 올리고뉴클레오타이드일 수 있다. 억제제는 NHEJ 또는 HDR에 관여하는 효소에서 또는, 예를 들어 인산화, 유비퀴틸화, 및 수모화(sumoylation)를 통한 번역후 변형에 의해 그들의 상류 조절에서 유도될 수 있다.

[0260] 포유류 세포에서, NHEJ는 우세한 DSB 수복 기전이고, 세포 주기 전체에 걸쳐 활성이다. 척추동물에서, "표준적" 또는 "고전적" NHEJ 경로(C-NHEJ)는 DSB를 수복하는 데 DNA-PK, Ku70-80, 아르테미스(Artemis), 리가제 IV(Lig4), XRCC4, CLF, 및 Pol μ 를 포함한 여러 코어 인자를 필요로 한다. 예를 들어, 문헌[Kasperek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897]을 참조하며, 이는 모든 목적을 위하여 전체적으

로 본 명세서에 참고로 포함된다. NHEJ 동안, DNA 말단은 매우 풍부한 말단-보호 Ku 단백질에 의해 결합되는데, 이러한 말단-보호 Ku 단백질은 다른 NHEJ 성분들의 로딩을 위한 독킹 스테이션(docking station)으로서 기능한다.

[0261] 따라서, 본 명세서에 개시된 방법들 중 일부에서, 세포는 C-NHEJ에 관여하는 인자들의 발현 또는 활성을 감소 또는 제거하거나 증가시키도록 변형되었다. 예를 들어, 일부 방법에서, 세포는 DNA-PK, Ku70-80, 아르테미스, 리가제 IV(Lig4), XRCC4, CLF, 및/또는 Pol μ 발현 또는 활성을 감소 또는 제거하도록 변형되었다. 구체적인 방법에서, 세포는 DNA-PK 발현 또는 활성을 감소 또는 제거하거나 DNA-PK 발현 또는 활성(예를 들어, DNA-PKcs의 발현 또는 활성; P97313로 지정된 예시적인 UniProt 서열)을 증가시키도록 변형되었다. DNA-PKcs 억제제의 예는, 예를 들어, NU7026, 및 NU7441을 포함한다. 예를 들어, 미국 특허 제6,974,867호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 구체적인 방법에서, 세포는 리가제 IV 발현 또는 활성을 감소 또는 제거하거나 리가제 IV 발현 또는 활성을 증가시키도록 변형되었다. 리가제 IV 억제제의 예는 SCR7이다.

[0262] ATM(예를 들어, KU55933), CHK1/CHK2(예를 들어, KLD1162 또는 CHIR-124) 및 ATR(예를 들어, VE 821)와 같은 세포 주기 체크포인트 단백질을 표적화하는 억제제는 또한 특정 DNA 수복 억제제의 효과를 상승적으로 향상시키는 데 또는 세포 주기 정지 및/또는 아포토시스(apoptosis)와 같은 의도치 않은 부작용을 방지하는 데 사용될 수 있다(문헌[Ciccica et al. (2010) Mol Cell 40:179]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함됨).

[0263] C-NHEJ의 파괴는 "대체" NHEJ(A-NHEJ) 경로에 의해 매개되는 비정상 결합의 수준을 증가시킬 수 있고, 또한 HR 수복을 증가시킬 수 있다. A-NHEJ 경로는 미세상동성(microhomology)-매개 결합 쪽으로의 편향을 나타내고, C-NHEJ보다 더 느린 속도를 따른다. MRN 복합체(MRE11, RAD50, NBS1), CtIP, XRCC1, PARP, Lig1, 및 Lig3을 포함한 여러 인자들이 참여하는 것으로 제안되어 있다. 문헌[Kasperek & Humphrey (2011) *Seminars in Cell & Dev. Biol.* 22:886-897] 및 문헌[Claybon et al. (2010) *Nucleic Acids Res.* 38(21):7538-7545]을 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0264] 따라서, 본 명세서에 개시된 방법들 중 일부에서, 세포는 A-NHEJ에 관여하는 인자들의 발현 또는 활성을 감소 또는 제거하거나 증가시키도록 변형되었다. 예를 들어, 일부 방법에서, 세포는 MRE11, RAD50, NBS1, CtIP, XRCC1, PARP(예를 들어, PARP1), Lig1, 및/또는 Lig3 발현 또는 활성을 감소 또는 제거하도록 변형되었다. 다른 방법에서, 세포는 MRE11, RAD50, NBS1, CtIP, XRCC1, PARP(예를 들어, PARP1), Lig1, 및/또는 Lig3 발현 또는 활성을 증가시키도록 변형되었다. 구체적인 방법에서, 세포는 PARP1 발현 또는 활성을 감소 또는 제거하거나 PARP1 발현 또는 활성(P11103으로 지정된 예시적인 UniProt 서열)을 증가시키도록 변형되었다. PARP 억제제(예를 들어, NU1025, 이니파리브(Iniparib), 올라파리브(Olaparib))의 예는 니코틴아미드; 아이소퀴놀린 및 다이하이드로아이소퀴놀린; 벤즈이마다졸 및 인돌; 프탈라진-1(2H)-온 및 퀴나졸린; 아이소인돌린 및 그의 유사체 및 유도체; 페난트리딘 및 페난트리딘; 벤조피론 및 그의 유사체 및 유도체; 불포화 하이드록시산(hydroxamic acid) 유도체 및 그의 유사체 및 유도체; 피리다진 - 융합 피리다진 및 그의 유사체 및 유도체를 포함함 -; 및/또는 다른 화합물, 예컨대 카페인, 테오필린, 및 티미딘, 및 이들의 유사체 및 유도체를 포함한다. 예를 들어, 미국 특허 제8,071,579호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0265] C-NHEJ는 또한 HR과 경쟁적 관계를 나타내서 C-NHEJ의 파괴는 또한 증가된 HR 수복으로 이어지도록 할 수 있다. NHEJ와 HR 사이의 그러한 경쟁은 잘 활용될 수 있는데, 이는 NHEJ의 파괴가 감소된 무작위 통합 및 가능하게는 상동 재조합에 의한 증가된 표적 통합을 통한 향상된 유전자 표적화로 이어질 수 있기 때문이다.

[0266] 단일-가닥 어닐링, 유전자 변환, 교차, 및 브레이크-유도 복제를 포함한 여러 형태의 상동 재조합 수복이 있다. 단일-가닥 어닐링은, 절제된 DSB의 양측에 있는 상동성 단일-가닥 서열들이 어닐링되어, 염색체 재구성을 가져오는 미소한 형태의 HR 수복이다. 단일-가닥 어닐링은, 서열 상동성의 2개의 영역을 분리하는 거리에 따라, 다양한 크기의 결실을 생성한다. 유전자 변환은 하나의 분자로부터 그의 상동체로의 유전 정보의 비상호 교환을 포함하며, 상동 염색체로부터의 유전 정보의 카피의 결과로서 HR로부터 직접 발생된다. 상동성 서열들이 동일하지 않은 경우, 이것은 국제화된 LOH로 이어질 수 있다. 통상적으로, 유전자 변환의 정도는 수백 개의 염기쌍으로 제한된다. 그러나, 긴 트랙 유전자 변환이 일부 유전적 백그라운드(genetic background)에서 보고되어 있으며, 이에 RAD51C 결핍이 포함된다. 문헌[Nagaraju et al. (2006) *Mol. Cell. Biol.* 26:8075-8086]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 교차가, 예를 들어, 상동 염색체들

사이에서 일어날 수 있고, 이는, G1에서 일어난다면 상호 전좌로 이어지거나 또는 G2에서 일어난다면 브레이크 부위로부터 원위 말단소체까지 연장되는 LOH 및 비상호 전좌로 이어질 잠재력을 가질 수 있다. 브레이크-유도 복제는 HR의 변이체인데, HR에서는 가닥 침입 후에, DNA 복제가 염색체의 말단까지 내내 계속된다. 따라서, HR 이 LOH를 촉진시킬 수 있는 많은 기전이 있다.

[0267] 따라서, 본 명세서에 개시된 방법들 중 일부에서, 세포는 HR에 관여하는 인자들의 발현 또는 활성을 감소 또는 제거하거나 증가시키도록 변형되었다. 예를 들어, 일부 방법에서, 세포는 RAD51, RAD52, RAD54, RAD55, RAD51C, BRCA1, 및/또는 BRCA2 발현 또는 활성을 증가시키도록 변형되었다. 다른 방법에서, 세포는 RAD51, RAD52, RAD54, RAD55, RAD51C, BRCA1, 및/또는 BRCA2 발현 또는 활성을 감소 또는 제거하도록 변형되었다.

[0268] 일부 방법에서, NHEJ 및/또는 HR을 조절하는 데 관여하는 또 다른 단백질의 발현 또는 활성이 변경될 수 있다. 예를 들어, 일부 방법에서, 세포는 *Chk2* 발현 또는 활성을 감소 또는 제거하도록, *Clspn* 발현 또는 활성을 감소 또는 제거하도록, *Setd2* 발현 또는 활성을 감소 또는 제거하도록, *Kat2a* 발현 또는 활성을 증가시키도록, 그리고/또는 *Rad51* 발현 또는 활성을 증가시키도록 변형되었다. 다른 방법에서, 세포는 *Chk2* 발현 또는 활성을 증가시키도록, *Clspn* 발현 또는 활성을 증가시키도록, *Setd2* 발현 또는 활성을 증가시키도록, *Kat2a* 발현 또는 활성을 감소 또는 제거하도록, 그리고/또는 *Rad51* 발현 또는 활성을 감소 또는 제거하도록 변형되었다.

[0269] Chk2(Chek2 및 Rad53으로도 공지됨; S. 프롬베(*S. pombe*) 상동체는 Cds1임)는 체크포인트-매개 세포 주기 정지, DNA 수복의 활성화, 및 DNA 이중-가닥 브레이크의 존재에 따른 아포토시스에 필요한 세린/트레오닌 단백질 키나제이다. 문헌[Blaikley *et al.* (2014) *Nucleic Acids Research* 42:5644-5656]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. Clspn(Claspin으로도 공지됨; S. 프롬베 상동체는 Mrc1임)은 DNA 손상에 따른 체크포인트 매개 세포 주기 정지에 필요한 단백질이다. S. 프롬베에서의 *Chk2* 또는 *Clspn*의 상동체의 결실은 야생형에 비하여 유의하게 상승된 수준의 브레이크-유도 유전자 변환을 나타내는 과다-재조합(hyper-recombinant) 표현형을 가져오는 것으로 보고되어 있다. 구체적으로, 유전자 변환의 수준은 유의하게 증가되는 것으로 보고되었으며, 한편 비상동 말단 결합(NHEJ), 자매 염색분체 변환(SCC), 및 이형접합성의 상실(LOH)의 수준은 감소되는 것으로 보고되었다. 문헌[Blaikley *et al.* (2014) *Nucleic Acids Research* 42:5644-5656]을 참조한다.

[0270] Kat2a(Gcn5 및 Gcn512로도 공지됨)는, 전사 활성화를 촉진시키고 이중-가닥 브레이크 수복과 관련된 것으로 보고되어 있는 편재적(ubiquitous) 히스톤 아세틸트랜스퍼라제이다. Kat2a-의존성 히스톤 H3 라이신 36(H3K36) 아세틸화는 염색질 접근성을 증가시키고, 절제를 증가시키고, 비상동 말단 결합을 억제하면서 상동 재조합을 촉진시킨다. 문헌[Pai *et al.* (2014) *Nat. Commun.* 5:4091]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. Setd2(Kiaa1732, Kmt3a, 및 Set2로도 공지됨)는 히스톤 H3의 라이신 36(H3K36me3)을 기질로서 탈메틸화된 라이신 36(H3K36me2)을 사용하여 특이적으로 트라이메틸화하는 히스톤 메틸트랜스퍼라제이다. Setd2-의존성 H3K36 메틸화는 염색질 접근성을 감소시키고, 절제를 감소시키고, NHEJ를 촉진시킨다. 문헌[Pai *et al.* (2014) *Nat. Commun.* 5:4091]을 참조한다.

[0271] Rad 51(Reca, Rad51A, 및 DNA 수복 단백질 Rad51 상동체 1로도 공지됨)은, Rad52 및 다른 단백질과 함께, 상동 재조합 동안 가닥 교환을 달성하여, 헤테로듀플렉스 DNA를 형성하고, 헤테로듀플렉스 DNA가 불일치 수복에 분해되어 유전자 변환 트랙을 생성하도록 기능하는 단백질이다. 포유류 세포에서, Rad51 및 Rad52 과발현은 상동 재조합 및 유전자 변환의 빈도를 증가시키는 것으로 보고되어 있다. 문헌[Yanez & Porter (1999) *Gene Ther.* 6:1282-1290] 및 문헌[Lambert & Lopez (2000) *EMBO J.* 19:3090-3099]을 참조하며, 이들은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0272] NHEJ, 유전자 변환, 및/또는 상동성-유도 수복을 조절하는 데 관여하는 유전자의 발현 또는 활성의 변형은 공간적으로 또는 시간적으로 특이적일 수 있으며, 또한 유도성 또는 일시적 및 가역적일 수 있다. 예를 들어, 다양한 형태의 카세트가 특정 세포 또는 조직 유형에서, 특정 발생기에서 또는 유도 시에 결실을 허용하도록 작제될 수 있다. 그러한 카세트는 카세트가 재조합효소 인식 부위에 의해 양측에 플랭킹되는 재조합효소 시스템을 사용할 수 있고, 원하는 세포 유형에서 발현되거나, 원하는 발생기에서 발현되거나, 또는 유도 시에 발현되거나 활성화되는 재조합효소를 사용하여 제거될 수 있다. 그러한 카세트는 널, 조건적, 또는 조합 조건적/널 대립유전자가 생성될 수 있도록 배치된 상이한 재조합효소 인식 부위 쌍들의 어레이를 포함하도록 추가로 작제될 수 있는데, 이는 미국 특허 출원 공개 제2011/0104799호에 기재된 바와 같으며, 이 출원은 모든 목적을 위하여 전체적으로 참고로 포함된다. 재조합효소 유전자의 조절은 다양한 방식으로, 예컨대 재조합효소 유전자를 세포-특이적, 조직-특이적, 또는 발생적으로 조절된 프로모터(또는 다른 조절 요소)에 작동가능하게 연결시킴으로써,

또는 재조합효소 유전자를 특정 세포 유형, 조직 유형, 또는 발생기에서만 활성 상태인 miRNA를 위한 인식 부위를 포함하는 3'-UTR에 작동가능하게 연결시킴으로써 제어될 수 있다. 재조합효소는 또한, 예를 들어, 재조합효소를 이펙터 또는 대사산물(예를 들어, CreER^{T2}, 이의 활성은 타목시펜에 의해 양성 제어됨(positively controlled))의 제어 하에 두는 융합 단백질을 사용함으로써, 또는 재조합효소 유전자를 유도성 프로모터(예를 들어, 활성이 독시사이클린 및 TetR 또는 TetR 변이체에 의해 제어되는 것)의 제어 하에 둠으로써 조절될 수 있다. 다양한 형태의 카세트의 예 및 재조합효소 유전자의 조절 수단이, 예를 들어, 미국 특허 제8,518,392호; 미국 특허 제8,354,389호; 및 미국 특허 제8,697,851호에 제공되어 있으며, 이들 각각은 전체적으로 참고로 포함된다.

G. 세포 및 동물

본 명세서에 제공된 다양한 조성물 및 방법은 세포, 예컨대 동물로부터의 세포를 사용한다. 그러한 세포는 비인간 동물로부터 유래될 수 있다. 그러한 세포는 진핵 세포일 수 있으며, 이에는, 예를 들어 진균 세포(예를 들어, 효모), 식물 세포, 동물 세포, 포유류 세포, 및 인간 세포가 포함된다. 포유류 세포는, 예를 들어, 비인간 포유류 세포, 인간 세포, 설치류 세포, 래트 세포, 마우스 세포, 햄스터 세포, 섬유아세포, 또는 CHO 세포일 수 있다. 진핵 세포는 전능성 세포, 다능성 세포, 예컨대 비인간 다능성 세포(예를 들어, 마우스 배아 줄기(ES) 세포 또는 래트 ES 세포) 또는 인간 다능성 세포, 또는 비다능성 세포일 수 있다. 전능성 세포는 임의의 세포 유형을 생성할 수 있는 미분화된 세포를 포함하고, 다능성 세포는 하나 초과와 분화된 세포 유형으로 발달될 능력을 갖는 미분화된 세포를 포함한다. 그러한 다능성 및/또는 전능성 세포는, 예를 들어, 배아 줄기(ES) 세포 또는 ES-유사 세포, 예컨대 유도 다능성 줄기(iPS) 세포일 수 있다. 배아 줄기 세포는 배아로의 도입 시에 발생 중인 배아의 임의의 조직에 기여할 수 있는 배아-유래 전능성 또는 다능성 세포를 포함한다. ES 세포는 배반포의 내부 세포 집단(inner cell mass)으로부터 유래될 수 있고, 3개의 척추동물 배엽층(내배엽, 외배엽, 및 중배엽) 중 임의의 것의 세포로 분화될 수 있다.

진핵 세포는 또한 1차 체세포가 아닌 세포일 수 있다. 체세포는 배우자(gamete), 생식 세포, 배우자 모세포(gametocyte), 또는 미분화 줄기 세포가 아닌 임의의 세포를 포함할 수 있다.

진핵 세포는 또한 1차 세포를 포함한다. 1차 세포는 유기체, 기관, 또는 조직으로부터 직접 단리된 세포 또는 세포의 배양물을 포함한다. 1차 세포는 형질전환 또는 불멸화 상태 어느 것도 아닌 세포를 포함한다. 이것은, 이전에 조직 배양물 중에 통과되지 않았거나 이전에 조직 배양물 중에 통과되었지만 조직 배양물 중에 무한으로 통과될 수 없는 유기체, 기관, 또는 조직으로부터 얻어진 임의의 세포를 포함한다. 그러한 세포는 통상적인 기술에 의해 단리될 수 있고, 예를 들어, 체세포, 조혈 세포, 내피 세포, 상피 세포, 섬유아세포, 중간엽 세포, 각질세포, 멜라닌세포, 단핵구, 단핵 세포, 지방세포, 지방전구세포, 뉴런, 신경교 세포, 간세포, 골격 근아세포, 및 평활근 세포를 포함한다. 예를 들어, 1차 세포는 결합 조직, 근육 조직, 신경계 조직, 또는 상피 조직으로부터 유래될 수 있다.

진핵 세포는 또한 불멸화 세포를 포함한다. 불멸화 세포는, 통상적으로 무한으로 증식하지 않지만, 돌연변이 또는 변경으로 인해 정상 세포 노화를 빠져나가서 대신에 진행 중인 분열을 유지할 수 있는 다세포 유기체로부터의 세포를 포함한다. 그러한 돌연변이 또는 변경은 자연적으로 발생할 수 있거나 의도적으로 유도될 수 있다. 불멸화 세포의 예는 중국 햄스터 난소(Chinese hamster ovary)(CHO) 세포, 인간 배아 신장 세포(예를 들어, HEK 293 세포), 및 마우스 배아 섬유아세포 세포(예를 들어, 3T3 세포)를 포함한다. 다수의 유형의 불멸화 세포가 당업계에서 잘 알려져 있다.

불멸화 또는 1차 세포는, 전형적으로 재조합 유전자 또는 단백질을 배양하거나 발현시키는 데 사용되는 세포를 포함한다.

진핵 세포는 또한 1-세포기 배아(즉, 수정 난모세포 또는 접합자)를 포함할 수 있다. 그러한 1-세포기 배아는 임의의 유전적 백그라운드(예를 들어, BALB/c, C57BL/6, 129, 또는 이들의 조합)로부터 유래될 수 있고, 새로운 것이거나 냉동된 것일 수 있고, 자연 교배 또는 시험관내 수정으로부터 유래될 수 있다.

세포, 다능성 및/또는 전능성 세포, ES 세포, 공유자 세포, 및/또는 숙주 배아와 관련하여, 용어 "동물"은 포유류, 어류, 및 조류를 포함한다. 포유류는, 예를 들어, 인간, 비인간 영장류, 원숭이, 유인원, 고양이, 개, 말, 황소, 사슴, 들소, 양, 설치류(예를 들어, 마우스, 래트, 햄스터, 기니피그), 가축(예를 들어, 소의 종, 예컨대 암소, 거세 황소 등; 양의 종, 예컨대 양, 염소 등; 및 돼지의 종, 예컨대 피그 및 수퇘지)을 포함한다. 조류는, 예를 들어, 닭, 칠면조, 타조, 거위, 오리 등을 포함한다. 가축 및 농업용 동물이 또한 포함된다. 용어

"비인간 동물"은 인간을 배제한다.

[0281] 마우스 다능성 및/또는 전능성 세포는 129 계통, C57BL/6 계통, 129와 C57BL/6의 혼합체, BALB/c 계통, 또는 스위스 웹스터(Swiss Webster) 계통으로부터 유래될 수 있다. 129 계통의 예는 129P1, 129P2, 129P3, 129X1, 129S1(예를 들어, 129S1/SV, 129S1/SvIm), 129S2, 129S4, 129S5, 129S9/SvEvH, 129S6(129/SvEvTac), 129S7, 129S8, 129T1, 및 129T2를 포함한다. 예를 들어, 문헌[Festing *et al.* (1999) *Mammalian Genome* 10:836]]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. C57BL 계통의 예는 C57BL/A, C57BL/An, C57BL/GrFa, C57BL/Kal_wN, C57BL/6, C57BL/6J, C57BL/6ByJ, C57BL/6NJ, C57BL/10, C57BL/10ScSn, C57BL/10Cr, 및 C57BL/Ola를 포함한다. 마우스 다능성 및/또는 전능성 세포는 또한 상기 언급된 129 계통과 상기 언급된 C57BL/6 계통의 혼합체(예를 들어, 50% 129와 50% C57BL/6)로부터 유래될 수 있다. 마찬가지로, 마우스 다능성 및/또는 전능성 세포는 상기 언급된 129 계통들의 혼합체 또는 상기 언급된 BL/6 계통들의 혼합체로부터 유래될 수 있다(예를 들어, 129S6(129/SvEvTac) 계통). 마우스 ES 세포의 구체적인 예는 VGF1 마우스 ES 세포이다. 예를 들어, 문헌[Auerbach *et al.* (2000) *Biotechniques* 29, 1024-1028, 1030, 1032]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0282] 래트 다능성 및/또는 전능성 세포는 임의의 래트 계통으로부터 유래될 수 있으며, 이에, 예를 들어, ACI 래트 계통, 다크 아구티(Dark Agouti)(DA) 래트 계통, 위스타(Wistar) 래트 계통, LEA 래트 계통, 스프라그 돌리(Sprague Dawley)(SD) 래트 계통, 또는 피셔(Fischer) 래트 계통, 예컨대 피셔 F344 또는 피셔 F6이 포함된다. 래트 다능성 및/또는 전능성 세포는 또한 상기에 열거된 2개 이상의 계통의 혼합체로부터 유래되는 계통으로부터 얻어질 수 있다. 예를 들어, 래트 다능성 및/또는 전능성 세포는 DA 계통 또는 ACI 계통으로부터 유래될 수 있다. ACI 래트 계통은, 흰색의 배 및 발과, *RTI*^{av1} 하플로타입(haplotype)을 가지며, 블랙 아구티를 갖는 것을 특징으로 한다. 그러한 계통은 할란 연구소(Harlan Laboratories)를 비롯한 다양한 공급처로부터 입수가능하다. ACI 래트로부터의 래트 ES 세포주의 일례는 ACI.G1 래트 ES 세포이다. 다크 아구티(DA) 래트 계통은 아구티 코트 및 *RTI*^{av1} 하플로타입을 갖는 것을 특징으로 한다. 그러한 래트는 찰스 리버 연구소(Charles River Laboratories) 및 할란 연구소를 비롯한 다양한 공급처로부터 입수가능하다. DA 래트로부터의 래트 ES 세포주의 예는 DA.2B 래트 ES 세포주 및 DA.2C 래트 ES 세포주이다. 일부 경우에, 래트 다능성 및/또는 전능성 세포는 근교계(inbred) 래트 계통으로부터 유래된다. 예를 들어, 2014년 2월 20일에 출원된 미국 특허 출원 공개 제2014/0235933 A1호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0283] 인간 다능성 세포의 예는 인간 ES 세포, 인간 성인 줄기 세포, 발생적으로 제한된 인간 선조 세포, 및 인간 유도 다능성 줄기(iPS) 세포, 예컨대 초회감작(primed) 인간 iPS 세포 및 나이브(*naïve*) 인간 iPS 세포가 포함된다. 유도 다능성 줄기 세포는 분화된 성체 세포로부터 직접 유래될 수 있는 다능성 줄기 세포를 포함한다. 인간 iPS 세포는 특정 세트의 재프로그래밍 인자들을 세포 내로 도입함으로써 생성될 수 있는데, 이는, 예를 들어, Oct3/4, Sox 패밀리를 전사 인자(예를 들어, Sox1, Sox2, Sox3, Sox15), Myc 패밀리를 전사 인자(예를 들어, c-Myc, l-Myc, n-Myc), 크뢰펠-유사(Krüppel-like) 패밀리를 전사 인자(예를 들어, KLF1, KLF2, KLF4, KLF5), 및/또는 관련 전사 인자, 예컨대 NANOG, LIN28, 및/또는 Glis1을 포함할 수 있다. 인간 iPS 세포는 또한, 예를 들어, miRNA, 전사 인자의 작용을 모방하는 소분자, 또는 계통 지정자(lineage specifier)의 사용에 의해 생성될 수 있다. 인간 iPS 세포는 3개의 척추동물 배엽층, 예를 들어, 내배엽, 외배엽, 또는 중배엽 중 임의의 세포로 분화되는 그의 능력을 특징으로 한다. 인간 iPS 세포는 또한 적합한 시험관내 배양 조건 하에서 무한으로 증식되는 그의 능력을 특징으로 한다. 예를 들어, 문헌[Takahashi and Yamanaka (2006) *Cell* 126:663-676]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 초회감작 인간 ES 세포 및 초회감작 인간 iPS 세포는, 이식 후 상배엽 세포의 특성과 유사한 특성을 발현하고 계통 예정화(specification) 및 분화가 결정된 세포를 포함한다. 나이브 인간 ES 세포 및 나이브 인간 iPS 세포는, 이식 전 배아의 내부 세포 집단의 ES 세포의 특성과 유사한 특성을 발현하고 계통 예정화가 결정되지 않은 세포를 포함한다. 예를 들어, 문헌[Nichols and Smith (2009) *Cell Stem Cell* 4:487-492]을 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다.

[0284] 숙주 배아 내로 이식된 세포는 "공여자 세포"로 지칭될 수 있다. 유전자 변형된 다능성 및/또는 전능성 세포는 숙주 배아와 동일한 계통으로부터 또는 상이한 계통으로부터 유래될 수 있다. 마찬가지로, 대리모는 유전자 변형된 다능성 및/또는 전능성 세포 및/또는 숙주 배아와 동일한 계통으로부터 유래될 수 있거나, 대리모는 유전

자 변형된 다능성 및/또는 전능성 세포 및/또는 숙주 배아와 상이한 계통으로부터 유래될 수 있다.

[0285] 다양한 숙주 배아가 본 명세서에 개시된 방법 및 조성물에 사용될 수 있다. 예를 들어, 표적화된 유전자 변형을 갖는 다능성 및/또는 전능성 세포는 상응하는 유기체로부터의 전상실배기 배아(예를 들어, 8-세포기 배아) 내로 도입될 수 있다. 예를 들어, 미국 특허 제7,576,259호, 미국 특허 제7,659,442호, 미국 특허 제7,294,754호, 및 미국 특허 출원 공개 제2008/0078000 A1호를 참조하며, 이들 각각은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함된다. 다른 경우에, 공여자 ES 세포는 2-세포기, 4-세포기, 8-세포기, 16-세포기, 32-세포기, 또는 64-세포기의 숙주 배아 내로 이식될 수 있다. 숙주 배아는 또한 배반포일 수 있거나, 전배반포 배아, 전상실배기 배아, 상실배기 배아, 비밀집화(uncompacted) 상실배기 배아, 또는 밀집화(compacted) 상실배기 배아일 수 있다. 마우스 배아를 사용하는 경우, 숙주 배아기는 문헌[Theiler (1989) "The House Mouse: Atlas of Mouse Development," Springer-Verlag, New York]에 기재된 테일러 단계(Theiler Stage)에 준거하여, 테일러 단계(TS1), TS2, TS3, TS4, TS5 및 TS6일 수 있다. 예를 들어, 테일러 단계는 TS1, TS2, TS3, 및 TS4로부터 선택될 수 있다. 일부 경우에, 숙주 배아는 투명대를 포함하고, 공여자 세포는 ES 세포이며, 이는 투명대의 구멍을 통해 숙주 배아로 도입된다. 다른 경우에, 숙주 배아는 무대(zona-less) 배아이다. 또 다른 경우에, 상실배기 숙주 배아는 응집된다.

[0286] H. 변형된 계놈을 갖는 세포를 확인하는 방법

[0287] 상기 방법들 중 일부는 변형된 계놈을 갖는 세포를 확인하는 단계를 추가로 포함한다. 표적화된 변형, 예컨대 결실 또는 삽입을 갖는 세포를 확인하기 위해 다양한 방법이 사용될 수 있다. 그러한 방법은 (예를 들어, 제1 CRISPR RNA 인식 서열과 제2 CRISPR RNA 인식 서열 사이의) 표적 유전자좌에서 표적화된 변형을 갖는 하나의 세포를 확인하는 단계를 포함할 수 있다. 변형된 계놈 유전자좌를 갖는 그러한 세포를 확인하기 위해 스크리닝이 행해질 수 있다.

[0288] 스크리닝 단계는 모 염색체의 대립유전자의 변형(MOA)을 평가하기 위한 정량적 검정을 포함할 수 있다. 예를 들어, 정량적 검정은 정량적 PCR, 예컨대 실시간 PCR(qPCR)을 통해 수행될 수 있다. 실시간 PCR은 표적 유전자좌를 인식하는 제1 프라이머 세트 및 비표적화된 참조 유전자좌를 인식하는 제2 프라이머 세트를 이용할 수 있다. 프라이머 세트는 증폭된 서열을 인식하는 형광 프로브를 포함할 수 있다.

[0289] 스크리닝 단계는 또한 보유 검정을 포함할 수 있는데, 이 검정은 표적 계놈 유전자좌 내로의 핵산 삽입물의 올바른 표적화된 삽입과 표적 계놈 유전자좌의 외부에 있는 계놈 위치 내로의 핵산 삽입물의 무작위 유전자도입적 삽입 사이를 구별하는 데 사용되는 검정이다. 표적화된 변형에 대해 스크리닝하기 위한 종래의 검정, 예컨대 긴 범위 PCR 또는 서던 블롯팅은 삽입된 표적화 벡터를 표적화된 유전자좌에 연결시킨다. 그러나, 큰 상동성 아암 크기 때문에, LTVEC는 그러한 종래의 검정에 의한 스크리닝을 허용하지 않는다. LTVEC 표적화를 스크리닝하기 위하여, 대립유전자의 상실(LOA) 및 대립유전자의 획득(GOA)을 포함한 대립유전자의 변형(MOA) 검정이 사용될 수 있다(예를 들어, 미국 특허 출원 공개 제2014/0178879호 및 문헌[Frendewey *et al.* (2010) *Methods Enzymol.* 476:295-307]을 참조하며, 이들은 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함됨). 대립유전자의 상실(LOA) 검정은 종래의 스크리닝 로직을 역전시키고 돌연변이가 유도된 천연 유전자좌의 카피의 수를 정량한다. 올바르게 표적화된 세포 클론에서, LOA 검정은 (X 또는 Y 염색체 상에 있지 않은 유전자에 대해) 2개의 천연 대립유전자들 중 하나를 검출하는데, 이때 다른 하나의 대립유전자는 표적화된 변형에 의해 파괴된다. 동일한 원리가 대립유전자의 획득(GOA) 검정에 대해 역으로 적용되어, 삽입된 표적화 벡터의 카피수를 정량할 수 있다. 예를 들어, GOA 및 LOA 검정의 조합된 사용은, 천연 표적 유전자의 하나의 카피를 상실하고 약물 저항성 유전자 또는 다른 삽입된 마커의 하나의 카피를 획득한 것으로서 올바르게 표적화된 이형접합성 클론을 보여줄 것이다.

[0290] 일례로서, 정량적 폴리머라제 사슬 반응(qPCR)이 대립유전자 정량화의 방법으로서 사용될 수 있지만, 표적 유전자의 0개, 1개, 및 2개의 카피 사이 또는 핵산 삽입물의 0개, 1개, 및 2개의 카피 사이의 차이를 신뢰성있게 구별할 수 있는 임의의 방법이 MOA 검정을 개발하는 데 사용될 수 있다. 예를 들어, 특히 참조 유전자와 비교함으로써, 계놈 DNA 샘플에서의 DNA 주형의 카피의 수를 정량하는 데 TaqMan®이 사용될 수 있다(예를 들어, 미국 특허 제6,596,541호를 참조하며, 이는 모든 목적을 위하여 전체적으로 참고로 포함됨). 참조 유전자는 표적 유전자(들) 또는 유전자좌(들)과 동일한 계놈 DNA에서 정량된다. 따라서, 2회의 TaqMan® 증폭(각각은 그의 각각의 프로브를 가짐)이 수행된다. 하나의 TaqMan® 프로브는 참조 유전자의 "Ct"(역치 주기)를 결정하고, 다른 하나의 프로브는 성공적인 표적화에 의해 대체되는 표적화된 유전자(들) 또는 유전자좌(들)의 영역의 Ct를 결정한다(즉, LOA 검정). Ct는 TaqMan® 프로브들 각각에 대한 출발 DNA의 양을 반영하는 양이며, 즉 덜 풍부한 서

열은 역치 주기에 도달하는 데 더 많은 주기의 PCR을 필요로 한다. TaqMan® 반응에 대한 주형 서열의 카피수를 절반으로 감소시킴으로써 약 1 Ct 단위가 증가될 것이다. 표적 유전자(들) 또는 유전자좌(들) 중 한쪽 대립 유전자가 상동 재조합에 의해 대체된 세포에서의 TaqMan® 반응은 비표적화된 세포로부터의 DNA와 비교될 때 참조 유전자에 대한 Ct의 증가 없이 표적 TaqMan® 반응에 대해 1 Ct의 증가를 가져올 것이다. GOA 검정의 경우, 성공적인 표적화에 의해 표적화된 유전자(들) 또는 유전자좌(들)를 대체하고 있는 핵산 삽입물의 Ct를 결정하기 위해 다른 TaqMan® 프로브가 사용될 수 있다.

[0291] 쌍 형성된 gRNA는 표적 게놈 유전자좌에서의 큰 Cas-매개 결실을 생성할 수 있기 때문에, 그것은 LTVEC에 의한 (즉, 1-세포기 배아 이외의 세포에서의) 올바른 표적화를 검증하기 위한 유용한 강화 표준 LOA 및 GOA 검정일 수 있다. 예를 들어, LOA 및 GOA 검정 단독으로는 올바르게 표적화된 세포 클론을 표적 게놈 유전자좌의 큰 Cas-유도 결실이 게놈 내의 어딘가 다른 곳에서의 LTVEC의 무작위 통합과 일치하는 클론과 구별할 수 없는데, 이는 특히 GOA 검정이 LTVEC 삽입물 내의 선택 카세트에 대한 프로브를 사용하는 경우에 그러하다. 표적화된 세포에서의 선택압(selection pressure)은 선택 카세트에 기초하기 때문에, 게놈 내의 어딘가 다른 곳에서의 LTVEC의 무작위 유전자도입적 통합은 일반적으로 선택 카세트 및 LTVEC의 인접 영역들을 포함하겠지만 LTVEC의 더 원위인 영역은 제외시킬 것이다. 예를 들어, LTVEC의 일부분이 게놈 내로 무작위적으로 통합되고, LTVEC가 3' 상동성 아암에 인접한 선택 카세트와 함께 약 5 kb 이상 길이의 핵산 삽입물을 포함하는 경우, 일반적으로, 5' 상동성 아암은 아니고 3' 상동성 아암은 선택 카세트와 유전자도입적으로 통합될 것이다. 대안적으로, 선택 카세트가 5' 상동성 아암에 인접한 경우, 일반적으로 3' 상동성 아암은 아니고 5' 상동성 아암이 선택 카세트와 유전자도입적으로 통합될 것이다. 일례로서, LOA 및 GOA 검정이 LTVEC의 표적화된 통합을 평가하는 데 사용되고, GOA 검정이 선택 카세트에 대한 프로브를 이용하는 경우, LTVEC의 무작위 유전자도입적 통합과 조합된 표적 게놈 유전자좌에서의 이형접합성 결실은 표적 게놈 유전자좌에서의 LTVEC의 이형접합성 표적화된 통합과 동일한 판독을 제공할 것이다. LTVEC에 의한 올바른 표적화를 검증하기 위하여, 보유 검정이 단독으로 또는 LOA 및/또는 GOA 검정과 조합하여 사용될 수 있다.

[0292] 보유 검정은 (LTVEC의 5' 상동성 아암에 상응하는) 5' 표적 서열 및/또는 (LTVEC의 3' 상동성 아암에 상응하는) 3' 표적 서열에서의 DNA 주형의 카피수를 결정한다. 특히, 선택 카세트에 인접한 상동성 아암에 상응하는 표적 서열에서의 DNA 주형의 카피수를 결정하는 것이 유용하다. 이배체 세포에서, 2개 초과 카피수는 일반적으로 표적 게놈 유전자좌에 존재한다보다는 오히려 표적 게놈 유전자좌의 외부에 무작위로 존재하는 LTVEC의 유전자도입적 통합을 나타내며, 이는 바람직하지 않다. 올바르게 표적화된 클론은 2개의 카피수를 보유할 것이다. 게다가, 그러한 보유 검정에서 2개 미만의 카피수는 일반적으로 결실에 대해 표적화된 영역을 넘어서 연장되는 큰 Cas-매개 결실을 나타내며, 이 또한 바람직하지 않다.

[0293] 이배체 세포에서의 표적 게놈 유전자좌에서의 핵산 삽입물의 표적화된 삽입을 확인하기 위한 예시적인 보유 검정에서는, 먼저, 제1 표적 서열에 혼성화되는 제1 상동성 아암 및 제2 표적 서열에 혼성화되는 제2 상동성 아암에 의해 플랭킹된 핵산 삽입물을 포함하는 대형 표적화 벡터(LTVEC)와 접촉되었으며, 핵산 삽입물은 제1 상동성 아암에 인접한 선택 카세트를 포함하는 세포로부터 DNA를 얻는다. 선택적으로, 선택 카세트는 약물 저항성 유전자를 포함할 수 있다. 이어서, DNA를 제1 표적 서열 내에서 결합하는 프로브, 핵산 삽입물 내에서 결합하는 프로브, 및 기지의 카피수를 갖는 참조 유전자 내에서 결합하는 프로브에 노출시키며, 각각의 프로브는 결합 시에 검출가능한 신호를 발생시킨다. 이어서, 프로브들 각각의 결합으로부터 신호를 검출한다. 참조 유전자 프로브로부터의 신호를 제1 표적 서열 프로브로부터의 신호와 비교하여 제1 표적 서열에 대한 카피수를 결정하고, 참조 유전자 프로브로부터의 신호를 핵산 삽입물 프로브로부터의 신호와 비교하여 핵산 삽입물에 대한 카피수를 결정한다. 1개 또는 2개의 핵산 삽입물 카피수 및 2개의 제1 표적 서열 카피수는 일반적으로 표적 게놈 유전자좌에서의 핵산 삽입물의 표적화된 삽입을 나타내고, 1개 이상의 핵산 삽입물 카피수 및 3개 이상의 제1 표적 서열 카피수는 일반적으로 표적 게놈 유전자좌 이외의 게놈 유전자좌에서의 핵산 삽입물의 무작위 삽입을 나타낸다.

[0294] 제1 표적 서열 프로브의 결합으로부터의 신호는 제1 표적 서열에 대한 역치 주기(Ct) 값을 결정하는 데 사용될 수 있고, 참조 유전자 프로브의 결합으로부터의 신호는 참조 유전자에 대한 역치 주기(Ct) 값을 결정하는 데 사용될 수 있고, 제1 표적 서열의 카피수는 제1 표적 서열 Ct 값과 참조 유전자 Ct 값을 비교함으로써 결정될 수 있다. 마찬가지로, 핵산 삽입물 프로브의 결합으로부터의 신호는 핵산 삽입물에 대한 역치 주기(Ct) 값을 결정하는 데 사용될 수 있고, 핵산 삽입물의 카피수는 제1 표적 서열 Ct 값과 참조 유전자 Ct 값을 비교함으로써 결정될 수 있다.

[0295] LTVEC 내의 핵산 삽입물은, 예를 들어, 적어도 5 kb, 적어도 10 kb, 적어도 20 kb, 적어도 30 kb, 적어도 40

kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 150 kb, 적어도 200 kb, 적어도 250 kb, 적어도 300 kb, 적어도 350 kb, 적어도 400 kb, 적어도 450 kb, 또는 적어도 500 kb일 수 있다. 제1 표적 서열과 선택 카세트에서 프로브들이 결합하는 서열들 사이의 거리가, 예를 들어, 100개 뉴클레오티드 이하, 200개 뉴클레오티드 이하, 300개 뉴클레오티드 이하, 400개 뉴클레오티드 이하, 500개 뉴클레오티드 이하, 600개 뉴클레오티드 이하, 700개 뉴클레오티드 이하, 800개 뉴클레오티드 이하, 900개 뉴클레오티드 이하, 1 kb 이하, 1.5 kb 이하, 2 kb 이하, 2.5 kb 이하, 3 kb 이하, 3.5 kb 이하, 4 kb 이하, 4.5 kb 이하, 또는 5 kb 이하일 수 있다.

[0296] 그러한 방법은 제2 표적 서열의 카피수를 결정하기 위한 추가의 보유 검정을 추가로 포함할 수 있다. 예를 들어, 그러한 방법은 세포의 DNA를 제2 표적 서열과 결합하는 프로브에 노출시키는 단계, 제2 표적 서열 프로브의 결합으로부터 신호를 검출하는 단계, 및 참조 유전자 프로브로부터의 신호를 제2 표적 서열 프로브로부터의 신호와 비교하여 제2 표적 서열에 대한 카피수를 결정하는 단계를 추가로 포함할 수 있다.

[0297] 마찬가지로, 그러한 방법은 추가의 GOA 검정을 추가로 포함하여, 핵산 삽입물 내의 하나 이상의 추가 서열의 카피수를 결정할 수 있다. 예를 들어, 그러한 방법은 세포의 DNA를 핵산 삽입물과 결합하는 하나 이상의 추가 프로브에 노출시키는 단계, 하나 이상의 추가 프로브의 결합으로부터 신호를 검출하는 단계, 및 참조 유전자 프로브로부터의 신호를 하나 이상의 추가 핵산 삽입물 프로브로부터의 신호와 비교하여, 핵산 삽입물 내의 하나 이상의 추가 서열에 대한 카피수를 결정하는 단계를 추가로 포함할 수 있다.

[0298] 마찬가지로, LTVEC가 표적 게놈 유전자좌로부터 내인성 서열을 결실시키도록 설계되는 경우, 또는 쌍 형성된 gRNA가 사용되는 경우(예를 들어, 단일 게놈 표적 유전자좌 내의 상이한 부위에서 쌍 형성된 이중-가닥 브레이크를 생성하고 개재된 내인성 서열을 결실시키기 위해), 그러한 방법은 표적 게놈 유전자좌에서의 내인성 서열의 카피수를 결정하기 위해 LOA 검정을 추가로 포함할 수 있다. 예를 들어, 그러한 방법은 세포의 DNA를 표적 게놈 유전자좌에서의 내인성 서열과 결합하는 프로브에 노출시키는 단계, 내인성 서열 프로브의 결합으로부터 신호를 검출하는 단계, 및 참조 유전자 프로브로부터의 신호를 내인성 서열 프로브로부터의 신호와 비교하여 내인성 서열에 대한 카피수를 결정하는 단계를 추가로 포함할 수 있다.

[0299] 적합한 정량적 검정의 다른 예는 형광-매개 동소 혼성화(fluorescence-mediated in situ hybridization)(FISH), 비교 게놈 혼성화, 등온 DNA 증폭, 고정화된 프로브(들)에 대한 정량적 혼성화, Invader Probes®, MMP assays®, TaqMan® 분자 비콘(Molecular Beacon), 또는 Eclipse™ 프로브 기술을 포함한다(예를 들어, 미국 특허 출원 공개 제2005/0144655호를 참조하며, 이는 모든 목적을 위하여 전체적으로 본 명세서에 참고로 포함됨).

[0300] LTVEC를 사용하지 않고서 생성된 표적화된 유전자 변형의 경우, 표적화된 변형에 대해 스크리닝하기 위한 종래의 검정, 예컨대 긴 범위 PCR, 서던 블롯팅, 또는 생거 서열분석이 사용될 수 있다. 그러한 검정은 전형적으로 삽입된 표적화 벡터와 표적화된 게놈 유전자좌 사이의 결합에 대한 증거를 얻기 위해 사용된다. 예를 들어, 긴 범위 PCR 검정의 경우, 하나의 프라이머는 삽입된 DNA 내의 서열을 인식할 수 있으며, 다른 하나는 표적화 벡터의 상동성 아암의 말단 너머의 표적 유전자좌 서열을 인식한다.

[0301] 상기 또는 하기 인용된 모든 특허 출원, 웹사이트, 다른 간행물, 수탁 번호 등은 각각의 개별 항목이 구체적으로 및 개별적으로 그렇게 참고로 포함되도록 지시되어 있는 것처럼 그와 동일한 정도로 모든 목적을 위해 전체적으로 참고로 포함된다. 다양한 버전(version)의 서열이 다양한 시점에서 수탁 번호와 관련되는 경우, 본 출원의 유효 출원일에서 수탁 번호와 관련된 버전이 의도된다. 유효 출원일은 실제 출원일 또는 적용가능한 경우 수탁 번호를 언급하는 우선 출원의 출원일 중 더 빠른 것을 의미한다. 마찬가지로, 다양한 버전의 간행물, 웹사이트 등이 다양한 시점에서 공개된 경우, 달리 지시되지 않는 한 본 출원의 유효 출원일에서 가장 최근에 공개된 버전이 의도된다. 본 발명의 임의의 특징부, 단계, 요소, 실시 형태, 또는 태양은 달리 구체적으로 지시되지 않는 한 임의의 다른 것과 조합하여 사용될 수 있다. 본 발명은 명확성 및 이해의 목적으로 설명 및 예시에 의해 약간 상세히 기재되어 있기는 하지만, 첨부된 청구범위의 범주 내에서 소정 변화 및 변경이 실시될 수 있음이 명백할 것이다.

[0302] [표 1]

서열의 설명

서열 번호	유형	설명
1	RNA	gRNA
2	RNA	gRNA
3	RNA	crRNA
4	RNA	crRNA
5	RNA	crRNA
6	RNA	tracrRNA
7	RNA	tracrRNA
8	DNA	CRISPR RNA 인식 서열
9	DNA	CRISPR RNA 인식 서열
10	DNA	CRISPR RNA 인식 서열
11	DNA	<i>C5 (Hc)</i> gRNA A DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 100 bp)
12	DNA	<i>C5 (Hc)</i> gRNA B DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 500 bp)
13	DNA	<i>C5 (Hc)</i> gRNA C DNA-표적화 세그먼트 (표적 유전자와 중점들로부터 38200 및 37500 bp)
14	DNA	<i>C5 (Hc)</i> gRNA D DNA-표적화 세그먼트 (표적 유전자와 중점들로부터 43500 및 32200 bp)
15	DNA	<i>C5 (Hc)</i> gRNA E DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 500 bp)
16	DNA	<i>C5 (Hc)</i> gRNA E2 DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 100 bp)
17	DNA	<i>Lrp5</i> gRNA A DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 50 bp)
18	DNA	<i>Lrp5</i> gRNA B DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 500 bp)
19	DNA	<i>Lrp5</i> gRNA B2 DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 1000 bp)
20	DNA	<i>Lrp5</i> gRNA C DNA-표적화 세그먼트 (표적 유전자와 중점들로부터 29900 및 38430 bp)
21	DNA	<i>Lrp5</i> gRNA D DNA-표적화 세그먼트 (표적 유전자와 중점들로부터 29950 및 38380 bp)
22	DNA	<i>Lrp5</i> gRNA E2 DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 1000 bp)
23	DNA	<i>Lrp5</i> gRNA E DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 500 bp)
24	DNA	<i>Lrp5</i> gRNA F DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 50 bp)
25	DNA	<i>Ror1</i> gRNA A DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 200 bp)
26	DNA	<i>Ror1</i> gRNA B DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 1000 bp)
27	DNA	<i>Ror1</i> gRNA D DNA-표적화 세그먼트 (표적 유전자와 중점들로부터 54300 및 55500 bp)
28	DNA	<i>Ror1</i> gRNA C DNA-표적화 세그먼트 (표적 유전자와 중점들로부터 54500 및 55300 bp)
29	DNA	<i>Ror1</i> gRNA E DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 1000 bp)
30	DNA	<i>Ror1</i> gRNA F DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 200 bp)

[0303]

서열 번호	유형	설명
31	DNA	<i>Trpa1</i> gRNA A DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 100 bp)
32	DNA	<i>Trpa1</i> gRNA A2 DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 500 bp)
33	DNA	<i>Trpa1</i> gRNA B DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 1000 bp)
34	DNA	<i>Trpa1</i> gRNA C DNA-표적화 세그먼트 (표적 유전자와 중점들로부터 25600 및 19740 bp)
35	DNA	<i>Trpa1</i> gRNA D DNA-표적화 세그먼트 (표적 유전자와 중점들로부터 26970 및 18370 bp)
36	DNA	<i>Trpa1</i> gRNA E2 DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 1000 bp)
37	DNA	<i>Trpa1</i> gRNA E DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 500 bp)
38	DNA	<i>Trpa1</i> gRNA F DNA-표적화 세그먼트 (표적 유전자와 중점으로부터 100 bp)
39	DNA	190045 정방향 프라이머
40	DNA	190061 정방향 프라이머
41	DNA	190068 정방향 프라이머
42	DNA	190030 정방향 프라이머
43	DNA	190033 정방향 프라이머 (도 5에서의 SV 48.3에 대한 정방향 프라이머와 동일함)
44	DNA	190013 정방향 프라이머
45	DNA	190045 역방향 프라이머
46	DNA	190061 역방향 프라이머
47	DNA	190068 역방향 프라이머
48	DNA	190030 역방향 프라이머
49	DNA	190033 역방향 프라이머 (도 5에서의 SV 48.3에 대한 역방향 프라이머와 동일함)
50	DNA	190013 역방향 프라이머
51	DNA	C2 프로브 (B6) - 도 5에서의 SNV 0.32
52	DNA	T3 프로브 (B6) - 도 5에서의 SNV 1.2
53	DNA	T6 프로브 (B6) - 도 5에서의 SNV 11.1
54	DNA	T7 프로브 (B6) - 도 5에서의 SNV 13.2
55	DNA	T8 프로브 (B6) - 도 5에서의 SNV 17.5
56	DNA	T9 프로브 (B6) - 도 5에서의 SNV 25.8
57	DNA	T10 프로브 (B6) - 도 5에서의 SNV 33.0
58	DNA	T11 프로브 (B6) - 도 5에서의 SNV 38.3
59	DNA	T13 프로브 (B6) - 도 5에서의 SNV 49.6
60	DNA	T14 프로브 (B6) - 도 5에서의 SNV 57.2
61	DNA	C2 프로브 (129) - 도 5에서의 SNV 0.32
62	DNA	T3 프로브 (129) - 도 5에서의 SNV 1.2
63	DNA	T6 프로브 (129) - 도 5에서의 SNV 11.1
64	DNA	T7 프로브 (129) - 도 5에서의 SNV 13.2
65	DNA	T8 프로브 (129) - 도 5에서의 SNV 17.5
66	DNA	T9 프로브 (129) - 도 5에서의 SNV 25.8
67	DNA	T10 프로브 (129) - 도 5에서의 SNV 33.0
68	DNA	T11 프로브 (129) - 도 5에서의 SNV 38.3
69	DNA	T13 프로브 (129) - 도 5에서의 SNV 49.6
70	DNA	T14 프로브 (129) - 도 5에서의 SNV 57.2
71	DNA	C2 정방향 프라이머 - 도 5에서의 SNV 0.32
72	DNA	T3 정방향 프라이머 - 도 5에서의 SNV 1.2

[0304]

서열 번호	유형	설명
73	DNA	T6 정방향 프라이머 - 도 5에서의 SNV 11.1
74	DNA	T7 정방향 프라이머 - 도 5에서의 SNV 13.2
75	DNA	T8 정방향 프라이머 - 도 5에서의 SNV 17.5
76	DNA	T9 정방향 프라이머 - 도 5에서의 SNV 25.8
77	DNA	T10 정방향 프라이머 - 도 5에서의 SNV 33.0
78	DNA	T11 정방향 프라이머 - 도 5에서의 SNV 38.3
79	DNA	T13 정방향 프라이머 - 도 5에서의 SNV 49.6
80	DNA	T14 정방향 프라이머 - 도 5에서의 SNV 57.2
81	DNA	C2 역방향 프라이머 - 도 5에서의 SNV 0.32
82	DNA	T3 역방향 프라이머 - 도 5에서의 SNV 1.2
83	DNA	T6 역방향 프라이머 - 도 5에서의 SNV 11.1
84	DNA	T7 역방향 프라이머 - 도 5에서의 SNV 13.2
85	DNA	T8 역방향 프라이머 - 도 5에서의 SNV 17.5
86	DNA	T9 역방향 프라이머 - 도 5에서의 SNV 25.8
87	DNA	T10 역방향 프라이머 - 도 5에서의 SNV 33.0
88	DNA	T11 역방향 프라이머 - 도 5에서의 SNV 38.3
89	DNA	T13 역방향 프라이머 - 도 5에서의 SNV 49.6
90	DNA	T14 역방향 프라이머 - 도 5에서의 SNV 57.2
91	DNA	도 5에서의 SV 13.7에 대한 정방향 프라이머
92	DNA	도 5에서의 SV 13.7에 대한 역방향 프라이머
93	DNA	도 5에서의 SV 20.0에 대한 정방향 프라이머
94	DNA	도 5에서의 SV 20.0에 대한 역방향 프라이머
95	DNA	도 5에서의 SV 36.9에 대한 정방향 프라이머
96	DNA	도 5에서의 SV 36.9에 대한 역방향 프라이머
97	DNA	도 5에서의 SV 56.7에 대한 정방향 프라이머
98	DNA	도 5에서의 SV 56.7에 대한 역방향 프라이머
99	DNA	도 1에서의 m-l-r 프라이머
100	DNA	도 1에서의 m-s'-f 프라이머
101	DNA	도 1에서의 m-A 프라이머
102	DNA	도 1에서의 h-l-r 프라이머
103	DNA	도 1에서의 m-5'-r 프라이머
104	DNA	도 1에서의 h-5'-r 프라이머
105	DNA	도 1에서의 m-F 프라이머
106	DNA	도 1에서의 m-E2 프라이머
107	RNA	Cmah gRNA A DNA-표적화 세그먼트
108	RNA	Cmah gRNA B DNA-표적화 세그먼트
109	DNA	Cmah 유전자좌
110	DNA	gRNA A 컷의 상류에 있는 Cmah 유전자좌
111	DNA	gRNA B 컷의 하류에 있는 Cmah 유전자좌
112	RNA	gRNA A 및 B에 의해 절제된 Cmah 유전자좌 서열
113	DNA	도 13에서의 C5 프라이머 m-5'-F
114	DNA	도 13에서의 C5 프라이머 m-5'-R
115	DNA	도 13에서의 C5 프라이머 h-5'-R
116	DNA	도 15에서의 Cmah TaqMan® 정방향 프라이머
117	DNA	도 15에서의 Cmah TaqMan® 프로브
118	DNA	도 15에서의 Cmah TaqMan® 역방향 프라이머
119	DNA	7064retU 정방향 프라이머
120	DNA	7064retU 역방향 프라이머
121	DNA	7064retU TaqMan® 프로브
122	DNA	7064retD 정방향 프라이머

[0305]

서열 번호	유형	설명
123	DNA	7064retD 역방향 프라이머
124	DNA	7064retD TaqMan® 프로브
125	DNA	7140retU 정방향 프라이머
126	DNA	7140retU 역방향 프라이머
127	DNA	7140retU TaqMan® 프로브
128	DNA	7140retD 정방향 프라이머
129	DNA	7140retD 역방향 프라이머
130	DNA	7140retD TaqMan® 프로브
131	DNA	mADAM6-2 LOA 정방향 프라이머
132	DNA	mADAM6-2 LOA 역방향 프라이머
133	DNA	mADAM6-2 LOA 프로브
134	DNA	hIgH31 LOA 정방향 프라이머
135	DNA	hIgH31 LOA 역방향 프라이머
136	DNA	hIgH31 LOA 프로브
137	DNA	hIgH9 LOA 정방향 프라이머
138	DNA	hIgH9 LOA 역방향 프라이머
139	DNA	hIgH9 LOA 프로브
140	DNA	hIgH1 LOA 정방향 프라이머
141	DNA	hIgH1 LOA 역방향 프라이머
142	DNA	hIgH1 LOA 프로브
143	DNA	Neo GOA 정방향 프라이머
144	DNA	Neo GOA 역방향 프라이머
145	DNA	Neo GOA 프로브
146	DNA	5' IgH Arm1 보유 검정 정방향 프라이머
147	DNA	5' IgH Arm1 보유 검정 역방향 프라이머
148	DNA	5' IgH Arm1 보유 검정 프로브
149	DNA	mIgM398 보유 검정 정방향 프라이머
150	DNA	mIgM398 보유 검정 역방향 프라이머
151	DNA	mIgM398 보유 검정 프로브
152	DNA	mIgM1045 보유 검정 정방향 프라이머
153	DNA	mIgM1045 보유 검정 역방향 프라이머
154	DNA	mIgM1045 보유 검정 프로브
155	DNA	3' IgH Arm2 보유 검정 정방향 프라이머
156	DNA	3' IgH Arm2 보유 검정 역방향 프라이머
157	DNA	3' IgH Arm2 보유 검정 프로브
158	DNA	mIgHp2 모체 정방향 프라이머
159	DNA	mIgHp2 모체 역방향 프라이머
160	DNA	mIgHp2 모체 프로브
161	DNA	mIgKd2 모체 정방향 프라이머
162	DNA	mIgKd2 모체 역방향 프라이머
163	DNA	mIgKd2 모체 프로브
164	DNA	hIgK5 모체 정방향 프라이머
165	DNA	hIgK5 모체 역방향 프라이머
166	DNA	hIgK5 모체 프로브
167	DNA	3' gRNA_I DNA-표적화 서열
168	DNA	3' gRNA_II DNA-표적화 서열
169	DNA	5' gRNA_I DNA-표적화 서열
170	DNA	5' gRNA_II DNA-표적화 서열
171	DNA	5' IgH Arm2 보유 검정 정방향 프라이머
172	DNA	5' IgH Arm2 보유 검정 역방향 프라이머
173	DNA	5' IgH Arm2 보유 검정 프로브
174	DNA	3' IgH Arm1 보유 검정 정방향 프라이머
175	DNA	3' IgH Arm1 보유 검정 역방향 프라이머
176	DNA	3' IgH Arm1 보유 검정 프로브

[0306]

[0307]

실시예

[0308]

실시예 1. 1개의 가이드 RNA 또는 2개의 가이드 RNA를 사용한 CRISPR/Cas9-매개 표적화.

[0309]

재료 및 방법

[0310]

ES 세포 배양, 스크리닝, 및 전기천공

[0311]

본 명세서에 기재된 실험은 본 발명자들의 C57BL6NTac/129S6SvEvF1 혼성 XY ES 세포주인 VGF1을 사용하여 수행하였다(문헌[Poueymirou et al. (2007) *Nat. Biotechnol.* 25:91-99]; 문헌[Valenzuela et al. (2003) *Nat. Biotechnol.* 21:652-659]). ES 세포를 이전에 기재된 바와 같이 배양하였다(문헌[Matise et al. (2000) in Joyner, A.L. ed. *Gene Targeting: a practical approach*, pp. 100-132, Oxford University Press, New York]).

[0312]

0.12 ml의 최종 부피로 2 mm 겹 큐벳(gap cuvette)에서 750만개 세포로 전기천공(EP)을 수행하였다. EP에 대한 전기 조건은 BTX ECM 630 전기천공 시스템(미국 매사추세츠주 홀리스톤 소재의 Harvard Apparatus)을 사용하여 700 V, 400 옴 저항, 및 25 마이크로F 캐패시턴스였다. EP당 LTVEC의 양은 0.0015 mg이고, Cas9 발현 플라스미드는 0.005 mg이고, sgRNA 발현 플라스미드는 0.010 mg이었다. 일부 EP는 푸로마이신 저항성을 부여하는 플라스미드 100 ng을 첨가하여 수행하여, LTVEC에 의해 발현되는 네오마이신 저항성을 선택함 없이 클론의 선택

을 가능하게 하였다. EP 후에, 세포를 2개의 15 cm 젤라틴 처리 디시(gelatinized dish) 상에 플레이팅하고, 배지를 매일 변경하였다. 100 ug/ml의 G-418 설페이트 또는 0.0015 mg/ml의 푸로마이신을 함유하는 선택 배지는 EP 후 48시간째에 시작하였으며, EP 후 10일째까지 계속하였다. 콜로니를 PBS 중에 골라 넣고, 0.05% 트립신이 담긴 96웰 디시에 첨가하고, 15분 동안 해리되게 하고, 배지로 중화시키고, 스크리닝을 위해 DNA의 단리를 위해 사용하였다.

[0313] 대립유전자의 변형 방법(문헌[Frendewey *et al.* (2010) *Methods Enzymol.* 476:295-307])을 사용하여 올바르게 표적화된 ES 세포 클론을 확인하고 마우스 대립유전자 유전자형을 결정하였다.

[0314] 가이드 서열의 설계

[0315] 상류 및 하류 둘 모두의, *Lrp5* 또는 다른 표적화된 유전자의 결실된 부분 내부의 50 bp, 100 bp, 500 bp, 또는 1 kb 위치를 둘러싸는 대략 200 bp의 DNA를 CRISPR 설계 툴(crispr.mit.edu) 내로 넣어서 가능한 gRNA 서열을 검색하였다. 이어서, 잠재적인 gRNA 서열들을 여과하여, 이들이 단지 내인성 DNA의 절단만을 가능하게 하고 LTVEC 내의 인간화 삽입물은 그렇게 하지 않음을 보장하였다.

[0316] 단일 가이드 RNA 클로닝

[0317] sgRNA를 시ーム리스(seamless) RNA 발현을 위한 77 bp 스캐폴드에 융합된 BsmBI 부위에서 pMB_sgRNA(U6 프로모터) 내로 듀플렉스 올리고(IDT)로서 클로닝하거나, GeneCopeia로부터 검증된 발현 플라스미드로서 구매하였다(LRP5 가이드 A, B, B2, E2, E, 및 F). 사내 제조된 플라스미드를 PCR 및 생거 서열분석에 의해 확인하였다.

[0318] 유전자형 확인을 위한 DNA 주형

[0319] 표적화 벡터 및 Cas9를 발현하는 플라스미드 및 여러 가이드 RNA(gRNA) 중 하나를 발현하는 플라스미드 또는 상이한 gRNA 조합을 발현하는 2개의 플라스미드로 전기천공된 ES 세포로부터 유래된 클론인, ES 세포로부터 DNA를 정제하였다. 마우스 표적 유전자좌의 표적화된 결실 및 표적화 벡터의 삽입을 갖거나 Cas9/gRNA-유도 결실을 갖는 것으로 대립유전자의 변형(즉, 대립유전자의 상실 또는 대립유전자의 획득) 정량적 PCR 검정에 의해 확인된 클론을 종래의 추적 PCR 검정을 위해 선택하였다.

[0320] 올리고뉴클레오타이드 설계

[0321] 각각의 gRNA 조합에 대해 2개의 PCR 검정을 설계하였다. 첫 번째 PCR은 상이한 gRNA 조합들의 CRISPR RNA 인식 서열들 사이의 붕괴를 검출하기 위한 결실 검정이었다. 5' 검정인 두 번째 PCR 검정은 2개의 PCR 검정을 포함하였다. 첫 번째 것은 인간화된 대립유전자에 대한 5' 인간 검정이고 마우스-인간 접합부를 가로질러 설계되었다. 두 번째 것은 내인성 마우스 대립유전자에 대한 5' 마우스 검정이고 5' 표적화된 결실 접합부를 가로질러 설계되었다.

[0322] PCR 반응 및 TOPO 클로닝

[0323] TaKaRa LA Taq DNA 폴리머라제(Cat. # RR002M)를 사용하여 ES 세포 DNA 주형을 증폭하였다. 각각의 PCR 검정 반응 혼합물은 물 음성 대조군을 사용하여 실시하였다. 검정 혼합물은 하기를 함유하였다: 0.005 mL의 ES 세포 DNA 주형; 1X LA PCR 완충액 II(Mg²⁺ 플러스); 0.01 mM dNTP 혼합물; 0.0075 mM 정방향 올리고(Forward Oligo)(각각); 0.0075 mM 역방향 올리고(Reverse Oligo)(각각); 5000개 단위/mL의 LA Taq 폴리머라제; 및 0.025 mL에 이르기까지의 ddH₂O.

[0324] PCR 열사이클 프로그램은 하기로 이루어졌다: 1분 동안 94℃; 이후에 30초 동안 94℃의 35회 사이클, 30초 동안 60℃ 어닐링 구배, 및 증폭된 kb당 1분 동안 68℃; 이후에 10분 동안 72℃에서 중합.

[0325] PCR 산물을 Invitrogen 1 kb 플러스 DNA 래더(1 kb + DNA ladder)(Cat. # 10787-018) 및/또는 Invitrogen 50 bp DNA 래더(Cat. # 10416-014)를 사용하여 2% 아가로스 겔 상에서 전기영동에 의해 분획화하였다. 나머지 PCR 산물을 서열분석을 위한 Invitrogen의 TOPO TA 클로닝 키트(Cat. # K4575-02)로부터의 사용설명서에 따라 pCR4-TOPO 벡터 내로 클로닝하였다. 클로닝 반응물을 One Shot Top10 세포 내로 화학적으로 형질전환시키고, 0.06 mg/mL의 X-gal 플레이트 및 0.025 mg/mL의 카나마이신 한천 플레이트 상에서 플레이팅하였다.

[0326] 서열분석

[0327] 백색 콜로니를 0.025 mg/mL의 카나마이신이 담긴 LB 내로 접종하고, 37℃에서 진탕하면서 하룻밤 인큐베이션하

였다. 각각의 콜로니는 검정된 산물들의 집단으로부터 하나의 앰플리콘을 나타내었다. QIAGEN 플라스미드 미니프렙 키트(Cat. # 12123)를 사용하여 각각의 세균 배양물로부터 DNA를 추출하였다. 삽입물의 DNA 서열을 0.002 mL의 TOPO 클로닝된 PCR, 1x PCRx 향상제 용액(10x 스톱)(Cat. X11495-017), 0.0075 mM 올리고(M13F 또는 M13R), 및 0.015 mL에 이르기까지의 ddH₂O를 포함하는 서열분석 반응 혼합물에서 결정하였다.

[0328] 서열분석의 분석

[0329] 서열분석 결과를 정해지지 않은 서열 및 pCR4-TOPO 벡터 서열에 대해 트리밍하여, PCR 삽입 서열을 분리하였다. 이어서, 서열분석된 단편을 참조물에 맞추어 정렬하고, 변이를 분석하였다.

[0330] 붕괴된 클론의 서열분석

[0331] 붕괴된 양성 클론으로부터의 PCR 산물을 제조업체의 사용설명서(Invitrogen cat. # K4575-02)에 따라 pCR4-TOPO 벡터 내로 클로닝하고, 이어서 One Shot Top10 세포 내로 화학적으로 형질전환시키고 0.060 mg/mL의 X-gal 플레이트 및 0.025 mg/mL의 카나마이신 한천 플레이트 상에서 플레이트팅하였다. QIAGEN 플라스미드 미니프렙 키트(Cat. # 12123)를 사용하여 세균 배양물로부터 DNA를 추출하였다. 이어서, 삽입물 서열분석 결과를 예측된 붕괴 참조물에 맞추어 정렬하고, indel 변이를 분석하였다. Cas9가, PAM으로부터 gRNA에 의해 인식된 서열 내로 3개의 염기쌍을 절단하도록 예측되었다. 예측된 절단 내의 서열을 참조물로부터 결실시키고, 나머지를 사용하여 결과에 맞추어 정렬하였다.

[0332] 단일 뉴클레오타이드 변이체(SNV)에 대한 TaqMan® 대립유전자 식별 검정

[0333] TaqMan® 대립유전자 식별 반응은 0.008 mL이며, 이는 게놈 DNA, 각각의 다형성에 대한 특이적 프로브/프라이머, 및 TaqMan® 유전자 발현 PCR 마스터 혼합물을 함유하였다. 프로브들은 Life Technologies(Thermo)로부터 그리고 프라이머들은 IDT로부터 주문하였다. 대립유전자 129에 대한 프로브는 VIC 염료로 표지화하였고; 대립유전자 B6에 대한 프로브는 FAM 염료로 표지화하였다. 각각의 TaqMan® 대립유전자 검정을 384웰 플레이트 상에서 4회 반복하여 수행하고, Applied BioSystems ViiA 7 플랫폼 상에서 실시하였다. SNV PCR 사이클링 프로그램은 하기와 같았다: 10분 동안 95°C 이후에, 하기의 40회 사이클을 행하였다: 15초 동안 95°C, 60초 동안 60°C, 및 30초 동안 60°C. 실시 및 결과 평가의 분석을 ViiA 7 소프트웨어 v1.1을 사용하여 행하였다.

[0334] FISH 분석

[0335] 선택된 ES 세포 클론을 형광 동소 혼성화(FISH)를 사용하여 Cell Line Genetics(미국 위스콘신주 매디슨 소재) 또는 Van Andel Institute(미국 미시간주 그랜드 래피즈 소재)에 의해 그들의 표준 절차에 의해 분석하였다. 본 발명자들은 2-컬러 분석을 위한 프로브로서 마우스 및 인간 BAC를 제공하였다.

[0336] 향상된 게놈 붕괴 및/또는 표적 유전자좌의 인간화

[0337] 설치류 유전자의 전부 또는 일부의 정확한 단일-단계 결실 및 선택적으로 그의 인간 상동체의 전부 또는 일부에 의한 동시 대체를 달성하기 위하여, 본 발명자들은 전기천공에 의해 설치류 ES 세포 내로 하기의 핵산 분자를 도입하였다: (1) LTVEC; (2) Cas9 엔도뉴클레아제를 인코딩하는 플라스미드 또는 mRNA; 및 (3) 하나 이상의 CRISPR 단일 가이드 RNA(gRNA)를 인코딩하는 하나 이상의 플라스미드 또는 이들 gRNA 자체. 각각의 실험에서는, LTVEC를 선행화하였다. 일부 실험에서, LTVEC는, 설치류 유전자를 결실시키고 인간 유전자를 삽입하는 상동 재조합 이벤트를 유도하도록 설계된 설치류 DNA의 상동성 아암들에 의해 플랭킹된 유전자 산물(단백질 또는 RNA)을 인코딩하는 인간 유전자의 전부 또는 일부를 포함하였다. 다른 실험에서, LTVEC는 별개의 유전자좌, 예컨대 *Ch25h* 유전자좌를 표적화하도록 설계되었다. 어느 경우이든, LTVEC는 또한 항생제 약물(예를 들어, G418)에 대해 저항성을 부여하는 효소(예를 들어, 네오마이신 포스포트랜스퍼라제)의 발현을 유도하는 약물 선택 카세트를 가졌다.

[0338] LTVEC를 받아들여 그것을 게놈 내로 혼입시킨 ES 세포는 조직 배양 디시 상에서 항생제 약물이 담긴 성장 배지 중에서 성장하여 콜로니를 형성할 수 있었다. 본 발명자들은 LTVEC 분자보다 500 내지 1,000배 더 많은 CRISPR/Cas9-인코딩 및 gRNA-인코딩 핵산 분자를 도입하였기 때문에, LTVEC-함유 약물 저항성 콜로니의 대부분은 또한 적어도 일시적으로 CRISPR/Cas9 성분을 함유하였다. 본 발명자들은 약물 저항성 콜로니들을 고르고, 대립유전자의 변형 방법에 의해 그들을 스크리닝하여(문헌[Valenzuela *et al.* (2003) *Nat. Biotech.* 21:652-660]; 문헌[Frendewey *et al.* (2010) *Methods Enzymol.* 476:295-307]; 이들은 전체적으로 본 명세서에 참고로 포함됨), 올바르게 표적화된 인간화된 대립유전자를 가진 클론을 확인하였다. 게다가, 보유 검정으로서 지칭되

는, LTVEC의 상동성 아암들 내의 서열을 인식하는 실시간 PCR 검정을 사용하여 마우스 게놈 내로의 LTVEC의 올바른 표적화를 검증하였다. 이들 보유 검정의 카피수를 결정함으로써, 올바르게 표적화된 ES 클론 - 이는 2개의 카피수를 보유함 - 을 표적 마우스 유전자좌의 큰 Cas9-유도 결실이 게놈 내의 어딘가 다른 곳에서의 LTVEC의 무작위 통합과 일치하는 클론 - 이 경우에 보유 검정은 3개(또는 그 이상)의 카피수를 가짐 - 과 구별하는 것을 돕도록 추가의 명확화를 제공하였다. 표적 마우스 유전자좌에서 큰 Cas9-매개 결실을 생성하는 쌍 형성된 gRNA의 능력은 앞서 기재된 바와 같은 표준 LOA 및 GOA 검정이 추가의 명확화를 제공하고 올바른 표적화를 검증하도록 보유 검정에 의해 강화될 수 있음을 의미하였다. 따라서, 보유 검정을 설계하고 LOA 및 GOA 검정과 함께 사용하였다.

[0339] 각각의 실험에서는, 2개 중 어느 하나 또는 2개의 gRNA를 사용하였다. 사용된 gRNA들은 개별적으로 표적 유전자좌의 5' 말단(즉, 표적화된 마우스 유전자 결실), 표적 유전자좌의 중간, 또는 표적 유전자좌의 3' 말단 부근에서 Cas9 절단을 유도하였다. 2개의 gRNA를 사용하였을 경우, 1개의 gRNA는 표적 유전자좌의 5' 말단 부근에서 Cas9 절단을 유도하였고, 다른 1개의 gRNA는 표적 유전자좌의 중간에서 또는 표적 유전자좌의 3' 말단 부근에서 Cas9 절단을 유도하였다.

[0340] *Lrp5* 유전자좌

[0341] 한 세트의 실험에서는, 엑토도메인을 인코딩하는 마우스 *Lrp5*(저밀도 지질단백질 수용체-관련 단백질 5) 유전자의 일부분의 68 kb 결실 및 인간 *LRP5* 유전자로부터의 상동성 서열의 91 kb 단편에 의한 동시 대체를 생성하도록 LTVEC를 설계하였다(도 1). LTVEC는 결실이 의도된 마우스 *Lrp5* 유전자의 68 kb 서열을 플랭킹하는 마우스 *Lrp5* 유전자좌의 부분들로부터 유래된 7 kb 및 33 kb의 게놈 DNA를 함유하는 상동성 아암들에 의해 플랭킹된 인간 *LRP5* 유전자의 91 kb 단편을 포함하였다. 별도의 실험에서는, *Lrp5* 인간화(humanizing) LTVEC를, 결실에 대해 표적화된 마우스 *Lrp5* 유전자의 영역 내에 이중-가닥 브레이크를 생성하도록 설계된, Cas9를 인코딩하는 플라스미드 및 8개의 gRNA(A, B, B2, C, D, E2, E, F) 중 하나를 인코딩하는 제2 플라스미드와 조합하였다. gRNA들은 인간 *LRP5* 유전자의 삽입된 부분에서의 임의의 서열의 인식을 피하도록 설계하였다. 다른 실험에서, 본 발명자들은 LTVEC 및 Cas9-인코딩 플라스미드를, 결실에 대해 표적화된 마우스 *Lrp5* 유전자의 영역 내의 상이한 부위들을 표적화하는 2개의 상이한 gRNA를 인코딩하는 플라스미드와 조합하였다.

[0342] 약물-저항성 ES 세포 클론을 결실 내의 서열에 대해 그리고 약물 선택 카세트 및 인간 유전자 삽입물 내의 서열에 대해, 대립유전자의 변형 검정에 의해 표적화된 인간화에 대해 스크리닝하였다(문헌[Valenzuela et al. (2003) *Nat. Biotechnol.* 21:652-659]; 문헌[Frendewey et al. (2010) *Methods Enzymol.* 476:295-307]). 클론이 2개의 내인성 마우스 유전자 서열들 중 하나를 상실하고 인간 삽입물의 1개의 카피를 획득하고, 또한(LTVEC의 상동성 아암에 위치된) 보유 서열의 2개의 카피를 보유하였다면, 그러한 클론은 올바르게 표적화된 것으로 스코어링하였다. 이러한 스크리닝에 대한 2개의 보유 검정은 하기의 프라이머 및 프로브를 사용하는 TaqMan® 검정이었다: 7064retU 정방향 프라이머 CCTCTGAGCTTTCCTTGCAG (서열 번호 119); 7064retU 역방향 프라이머 CCTAGACAACAGACACTGTATCA (서열 번호 120); 7064retU TaqMan® 프로브 TTCTGCCCTTGAAAGGAGAGGC (서열 번호 121); 7064retD 정방향 프라이머 CCTCTGAGGCCACCTGAA (서열 번호 122); 7064retD 역방향 프라이머 CCCTGACAAGTTCTGCCTTCTAC (서열 번호 123); 7064retD TaqMan® 프로브 TGCCCAAGCCTCTGCAGCTTT (서열 번호 124).

[0343] *Lrp5* 유전자의 CRISPR/Cas9-보조 인간화의 결과가 표 2에 요약되어 있다. LTVEC 단독으로 ES 세포 내로 도입하였을 때에는, 1.9%의 스크리닝된 약물 저항성 클론이 올바르게 표적화된 이형접합성 인간화된 대립유전자를 가졌다(표 2에서의 이형접합성 표적화(Het. Targ.) 열을 참조하는데, 이 열은 비표적화된 대립유전자가 전혀 돌연변이되지 않았거나 작은 CRISPR-유도 돌연변이, 예컨대 NHEJ에 의해 야기된 작은 결실을 가진 클론을 포함한다). 대조적으로, LTVEC와 8개의 시험된 gRNA(A, B, B2, C, D, E2, E 및 F; 표 1 참조) 중 7개의 gRNA에 의해 가이드된 Cas9 엔도뉴클레아제의 조합은 2.1 내지 7.8% 범위의 효율로 올바르게 표적화된 단일대립유전자(monoallelic) 이형접합성 돌연변이를 생성하였다. B2 및 D에 의한 Cas9-가이드된 절단의 경우, 단일대립유전자 표적화에 더하여, 이중대립유전자 동형접합성 인간화를 1.0 내지 2.1%의 빈도로 검출하였다. 본 발명자들은, 심지어 작고 간단한 결실 대립유전자에 대해서조차도, LTVEC 그 자체에 의한 이중대립유전자 표적화를 결코 관찰하지 못하였다. 동형접합성 *Lrp5* 인간화 ES 세포는 VELOCIMOUSE® 방법(문헌[Poueymirou et al. (2007) *Nat. Biotech.* 25:91-99], 이는 전체적으로 본 명세서에 참고로 포함됨)에 의해, 표현형 및 약물 효능 연구에 즉시 사용가능한 완전하게 ES 세포-유도된 마우스로 직접 변환될 수 있다.

[0344] 예측된 절단 부위 또는 그 부근에서의 gRNA/Cas9-유도 NHEJ 돌연변이를 검출하도록 고안된 MOA 검정은 시험된

모든 gRNA에 대한 돌연변이 활성을 입증하였다(데이터는 도시되지 않음). 검정된 모든 클론들 중에서 검출된 단일대립유전자 또는 이중대립유전자 gRNA-유도 돌연변이의 비율은 유전자좌 및 위치에 따라 다양하였다. gRNA 돌연변이 활성과 LTVEC 표적화 사이에 강한 상관관계는 없었지만, 최저의 표적화 효율이 최저 돌연변이 빈도를 갖는 gRNA와 종종 관련되었다.

[0345] 결실에 대해 표적화된 *Lrp5* 유전자의 영역의 상이한 말단들을 인식하는 2개의 gRNA를 조합함으로써, 총 인간화 표적화 효율을 증가시켰는데, 이는 시험된 5개의 조합 중 3개에 대해 동형접합성 표적화 이벤트의 빈도를 증가 시킴으로써 주로 이루어졌다. gRNA들의 조합은 gRNA들에 의해 프래그래밍된 Cas9 절단 부위들 사이에 큰 결실을 생성할 잠재력을 갖기 때문에, 본 발명자들은 또한, 한쪽의 *Lrp5* 대립유전자 상에는 표적화된 인간화를 그리고 다른 한쪽의 대립유전자 상에는 큰 CRISPR-유도 결실을 갖는 반접합성 ES 세포 클론을 관찰하였다(gRNA 조합 A + F, 표 2). 게다가, gRNA 조합들 중 2개(A + F 및 A + E2)의 경우, 본 발명자들은 특유의 유전자형을 갖는 ES 세포 클론을 확인하였다: 양쪽 *Lrp5* 대립유전자 상의 큰 CRISPR-매개 결실.

[0346] [표 2]

개별 gRNA 들 및 조합된 gRNA 들을 사용한 *Lrp5* 엑토도메인의 CRISPR/Cas9-보조 인간화에 대한 스크리닝 결과.

gRNA	표적화된 결실의 5'/3' 말단으로부터 gRNA 부위까지의 거리(bp)	대립유전자 유형에 따른 표적화 효율				
		이형접합성 표적화 (Het. Targ.) (% 효율)	반접합성 표적화 (Hemi. Targ.) (% 효율)	동형접합성 표적화 (Homo. Targ.) (% 효율)	총 표적화 (Total Targ.) (% 효율)	동형접합성 결실 (Homo. Del.) (% 효율)
A	50 (5')	7.8			7.8	
B	500 (5')	4.2			4.2	
B2	1000 (5')	6.2		1.0	7.2	
C	29900 (5')/ 38430 (3')	4.1			4.1	
D	29950 (5')/ 38380 (3')	5.2		2.1	7.3	
E2	1000 (3')	2.1			2.1	
E	500 (3')	0.0			0.0	
F	50 (3')	4.2			4.2	
A + F	A: 50 (5') F: 50 (3')	6.6	2.9	2.2	11.7	2.9
B + E	B: 500 (5') E: 500 (3')	2.5			2.5	
B2 + E2	B2: 1000 (5') E2: 1000 (3')	4.2		2.1	6.3	
A + E	A: 50 (5') E: 500 (3')	4.6		6.2	10.8	
A + E2	A: 50 (5') E2: 1000 (3')	2.0		4.0	6.0	4.0
없음	N/A	1.9			1.9	

[0347]

[0348] 표 2에 입증된 바와 같이, 1개의 gRNA보다는 오히려 단일 유전자좌를 표적화하는 2개의 gRNA를 사용하였을 때, 이중대립유전자 표적화를 갖는 클론의 백분율의 상당한 증가가 관찰되었는데(도 2a 참조), 이는 gRNA 조합의 사용이 이중대립유전자 변형을 촉진시킴을 나타낸다. 도 2a는 LTVEC 및 2개의 가이드 RNA(A 및 B)를 사용하는 마우스 유전자의 결실 및 상응하는 인간 버전에 의한 대체의 동시 수행에 대한 일반적인 개략도를 나타낸다. 2개의 gRNA를 사용할 때 훨씬 더 높은 빈도로 관찰되는 특유의 돌연변이 대립유전자 유형은 동형접합성 붕괴된 대립유전자(도 2b; Δ/Δ), 동형접합성 표적화된 대립유전자(도 2c; Hum/Hum), 반접합성 표적화된 대립유전자(도 2d; Hum/ Δ), 및 다른 복합 이형접합성 표적화된 대립유전자(예를 들어, 한쪽 대립유전자는 LTVEC-표적화된 인간화를 갖고 다른 한쪽의 대립유전자는 CRISPR-유도 돌연변이, 예컨대 작은 결실을 가짐)(도 2e)를 포함한다.

[0349] MOA 검정에 기초한 유전자형을 뒷받침하고 확인하기 위해 여러 PCR 검정을 수행하였다. 프라이머들이 도 1에 나타나 있고 표 1에서 찾을 수 있다. *Lrp5* LTVEC는 인간 삽입물과, 인접한 마우스 게놈 서열 사이의 물리적 연결에 대해 검정하는 PCR에 의해 표적화를 입증하기에 충분히 짧은(6.9 kb) 5' 상동성 아암을 가졌다(도 1). 본 발명자들은 이형접합성, 반접합성, 또는 동형접합성으로 스코어링된 클론으로부터의 DNA를 갖지만 모체 ES 세포주로부터의 또는 이중대립유전자 큰 결실을 갖는 것으로 스코어링된 클론으로부터의 DNA는 갖지 않는 예측된 7.5 kb PCR 산물을 관찰하였으며(도 3a), 이에 따라 MOA(즉, LOA 및 GOA) 스크리닝에 의해 제조된 표적화 세포가 확인되었으며, 추론된 이중대립유전자 큰 결실이 뒷받침되었다. 5' -De1-J PCR 검정 - 이는 결실 및 삽입 접합부에서의 서열을 조사함 - (도 3b)은 모체 ES 세포주로부터의 그리고 대부분의 이형접합성 인간화 클론으로부터의 DNA를 갖는 330 bp 산물을 생성하였다(데이터는 도시되지 않음). 이형접합성 클론 AW-C3의 경우, 5' -De1-J 검정은 예측된 것보다 더 작은 산물을 생성하였는데(도 3b), 이는 gRNA A/Cas9 절단이 비표적화된 대립유전자 상에서 작은 결실 돌연변이를 유도하였음을 시사하며, 이때 이 또한 gRNA A 절단에 대한 MOA 검정에 의해

검출되었다(데이터는 도시되지 않음). 예측된 바와 같이, 5' -Del-J 검정은 반접합성, 동형접합성, 및 이중대립유전자 결실 대립유전자를 갖는 클론에 대해 음성이었다. 5' -Ins-J PCR(도 3b) - 이는 인간 DNA 삽입물의 5' 말단과, 인접한 마우스 플랭킹 서열 사이의 접합부에서의 서열을 조사함 - 은 이형접합성, 반접합성, 및 동형접합성 클론에서 478 bp 산물을 생성하였는데, 이들은 적어도 하나의 표적화된 인간화 대립유전자를 갖기 때문이다. 5' -Ins-J PCR 검정은 이중대립유전자 큰 결실을 갖는 클론에 대한 산물은 생성하지 않았다(도 3b). 반접합성 및 이중대립유전자 결실 클론에서의 큰 결실을 확인하기 위하여, 본 발명자들은 이중 gRNA 표적 부위들의 외부에 있는 서열을 인식하는 프라이머들을 사용하여 PCR을 수행하였다. Del(A + F) PCR - 이는 A gRNA 부위와 F gRNA 부위 사이의 결실에 대해 검정하였음 - (도 1)은 클론 AW-A8 및 BO-F10으로부터의 DNA를 갖는 대략 360 bp의 단일 산물을 생성하였는데(도 3b), 이는 *Lrp5* 대립유전자들 중 적어도 하나가 큰 결실을 가졌음을 확인시켜 준다. 마찬가지로, Del(A + E2) PCR - 이는 A gRNA 부위와 E2 gRNA 부위 사이의 큰 결실에 대해 검정하였음 - 은 클론 BA-A7로부터의 DNA를 갖는 대략 250 bp의 단일 산물을 생성하였다. 접합부, LOA, 및 GOA 검정과 함께, 결실 PCR은 이중대립유전자 큰 결실 유전자형을 뒷받침한다. 도 3a 및 도 3b에 나타난 검정 결과는 형광 동소 혼성화(FISH; 도 4a 내지 도 4c)에 더하여 표 2에 요약된 이중대립유전자 유전자형을 확인하기 위하여 본 발명자들이 수행한 것과 유사한 검정의 대표적인 예이다.

[0350] 형광 동소 혼성화(FISH)를 사용하여 *Lrp5* 유전자의 동형접합성 표적화된 인간화를 확인하였다. 정량적 및 종래의 PCR 검정에 의해, *Lrp5* 인간화 LTVEC(도 1)가 Cas9 및 2개의 gRNA(A + F 또는 A + E2)와 조합된 표적화 실험으로부터 동형접합성 표적화된 것으로 스코어링된 ES 세포 클론을 FISH 및 핵형 분석을 위한 상업용 세포검사 서비스(cytology service)로 보냈다. 마우스 *Lrp5* 유전자를 갖는 세균 인공 염색체(BAC)는 적색 형광 마커로 표지화하여 내인성 *Lrp5* 유전자좌를 확인하기 위한 프로브로 사용하고, 인간 *LRP5* 유전자를 갖는 BAC는 녹색 형광 마커로 표지화하여 인간 삽입물로 표적화된 염색체들을 확인하기 위한 프로브로 사용하였다. 표지화된 BAC 프로브를 표적화된 클론으로부터의 중기 확산체(metaphase spread)에 혼성화하고, 형광 현미경법에 의해 시각화하였다. 확산체 상의 염색체를 DAPI(4',6-다이아미디노-2-페닐인돌)로 염색하여 시각화하고, 각각의 클론에 대한 개별 핵형들을 김자(Giemsa) 염색에 의해 결정하였다. 전형적인 결과가 클론 AW-D9에 대해 도 4a에 나타나 있으며, 이것은 정상 40XY 핵형을 갖는 것으로 확인되었다(도시되지 않음). 도 4a에서의 합성 사진은 적색 마우스 BAC 프로브 신호 및 녹색 인간 BAC 프로브 신호 둘 모두가 *Lrp5* 유전자의 공지된 위치인 마우스 염색체 19의 두 카피 모두 상의 세포학적 밴드 B에 공국재화되었음을 보여준다. 도 4c에서의 합성 사진은 다른 클론(BA-D5)에 대한 동일한 동형접합성 표적화를 나타낸다. 이들 결과는 인간화 LTVEC(도 1)에서의 인간 *LRP5* 유전자의 91 kb 단편이 클론 AW-D9 및 BA-D5에서의 두 염색체 19 상동체 모두 상의 의도된 마우스 *Lrp5* 유전자좌에 올바르게 삽입되었음을 확인시켜 준다. 대조적으로, 도 4b에서의 합성 사진은 적색 마우스 BAC 프로브 신호 및 녹색 인간 BAC 프로브 신호 둘 모두가 마우스 염색체 19의 단일 카피 상의 세포학적 밴드 B(실선 화살표)에 공국재화되었으며, 한편 단지 적색 마우스 BAC 프로브 신호만이 마우스 염색체 19의 다른 하나의 카피 상의 세포학적 밴드 B에 국재화됨을 나타낸다. 이들 결과는 인간화 LTVEC(도 1)에서의 인간 *LRP5* 유전자의 91 kb 단편이 염색체 19의 단지 하나의 카피 상의 의도된 마우스 *Lrp5* 유전자좌에 올바르게 삽입되었음(이형접합성 표적화)을 확인시켜 준다. 이들은 또한 (도시되지 않은 다른 대조군과 함께) 인간 BAC 프로브가 마우스 *Lrp5* 유전자좌에 교차-혼성화되지 않고 단지 인간 *LRP5* 삽입물만을 인식함을 나타낸다.

[0351] 소정 클론에서 명백한 비상동 말단-결합 수복에 의해 양쪽 대립유전자에서 형성된 동일한 CRISPR-유도 indel 둘 연변이의 존재는 F1H4 혼성 세포(이는 50% 129SvS6 계통 및 50% C57BL/6N 계통으로 구성됨)에서의 유전자 변환 이벤트의 발생을 시사하였다. 2개의 gRNA를 사용하는 경우 향상된 이중대립유전자 표적화의 근간이 되는 기전에 대한 이해를 얻기 위하여, LTVEC 및 A + F 또는 A + E2 gRNA 조합에 의한 표적화 후에 동형접합성 인간화 또는 동형접합성 CRISPR-유도 큰 결실을 갖는 7개의 클론을 스크리닝하였다.

[0352] 도 5는 2개의 가이드 RNA에 의해 매개된 유전자 변환 이벤트를 조사하도록 설계된 검정의 예를 나타낸다. 구체적으로, F1H4 혼성 ES 세포(이는 50% 129 SvS6 계통 및 50% C57BL/6N 계통으로 구성됨)에서 이형접합성의 상실(LOH)을 분석함으로써 유전자 변환의 가능성을 조사하였다. 유전자 변환이 129SvS6 (129)과 C57BL/6N (B6) 사이의 공지된 다형성에서의 이형접합성의 상실에 의해 입증될 수 있고, 이에 따라 이들 2개의 대립유전자 유형들 사이를 구별하도록 PCR 검정을 설계하였다. 129 대립유전자와 B6 대립유전자 사이의 차이를 검출하도록 설계된 종래의 PCR에 의해 구조적 변이체(SV) 다형성을 검정하였다. 하기에 사용된 SV 검정들 중 단지 하나만이 도 5에 나타나 있지만, 개념은 각각에 대해 동일하다. B6 마우스 계통과 129 마우스 계통 사이의 구조적 변이(SV)에 기초하여 프라이머를 설계하였으며, 이는 표 1에 나타나 있다. 프라이머 설계 조건은, 약 25 bp SV를 확인하고 약 300 bp PCR 산물을 생성하도록 제한되었으며; 이들 조건은 임의의 변화가 겔 전기영동에 의해 가시적이

되도록 선택하였다.

[0353] 클론에 대해 PCR을 실시하기 전에, 검정을 검증하고 B6, 129 계통으로부터 그리고 F1H4 ES 세포주로부터의 야생형 ES-세포 DNA에 대해 최적화하였다. B6 또는 129 대립유전자에 특이적인 구별가능한 PCR 밴드들을 생성하고, F1H4 DNA를 사용하여 이들 동일한 2개의 구별가능한 밴드를 생성하는 데 일관성이 있는 프라이머 세트들을 클론에 대한 시험을 위해 선택하였다. 염색체 19(*Lrp5* 유전자의 위치)의 경우, 6개의 프라이머 세트 - ID 190045, 190061, 190068, 190030, 190033, 190013 - 를 대립유전자의 변형(MOA) 검정 및 종래의 PCR에 의해 "동형접합성 표적화된" 또는 "동형접합성 붕괴된" 것으로서 유전자형 결정된 *Lrp5* 인간화 클론에 대해 사용하기 위해 선택하였다. SV PCR 검정은 *Lrp5* 유전자좌로부터 염색체의 말단소체 말단 - *Lrp5* 유전자좌로부터 약 13.7 내지 약 56.2 Mb의 범위임 - 으로 염색체 19를 따라 간격을 두고 행해졌다. *Lrp5* 유전자좌로부터 염색체 19에 대한 SV 검정의 대략적인 거리(단위: Mb)는 하기와 같다: 검정 190045의 경우 13.7, 검정 190061의 경우 19.0, 검정 190068의 경우 35.0, 검정 190030의 경우 37.4, 검정 190033의 경우 48.3, 및 검정 190013의 경우 56.2. 단지 검정 190033만이 도 5에 나타나 있고(SV 48.3으로서 나타나 있음), 검정 190045, 190061, 190068, 190030, 190033, 및 190013에 대한 프라이머는 표 1에 나타나 있다.

[0354] 이들 클론으로부터의 DNA에 대해서뿐만 아니라 F1H4 대조군 DNA, 129 대조군 DNA, 및 B6 대조군 DNA에 대해서도 PCR을 실시하였다. PCR 산물을 6% 폴리아크릴아미드 겔 상에서 전기영동에 의해 분획화하였으며, 이어서 이것을 GelRed로 염색하였다. 2개의 밴드를 생성한 클론은 F1H4 대조군에 매칭되었는데, F1H4 대조군은 이전의 최적화로부터 상부 밴드는 129 대립유전자에 특이적이고 하부 밴드는 B6 대립유전자에 특이적임을 보여주었다. 단지 하나의 밴드를 생성한 클론은 단지 B6 밴드만을 또는 단지 129 밴드만을 나타내었다. 클론 AW-A7, AW-F10, BA-D5, BA-F2, BC-H9, 및 BR-B4는 모든 6개의 검정에 대해 단지 B6 밴드만을 보여주었으며, 한편 클론 BO-A8은 모든 6개의 검정에 대해 단지 129 밴드만을 보여주었다. 앞서 언급된 바와 같이, 이들 클론은 MOA 및/또는 PCR에 의해 동형접합성 표적화된 또는 동형접합성 붕괴된 것으로서 유전자형 결정되었고, 다양한 gRNA 조합(A + F, A + E2, B2, 및 D)을 포함하였다. 단지 단일 대립유전자 밴드만의 존재는 유전자 변환 이벤트가 일어나고 있음을 시사하였다 - 만약 변환이 없었다면, 두 밴드 모두 여전히 F1H4 대조군에서와 같이 존재할 것이다.

[0355] 게다가, 129 대립유전자와 B6 대립유전자 사이의 단일 뉴클레오타이드 변이체(SNV)를 TaqMan® 대립유전자 식별 검정에 의해 검정하였다. 도 5에서의 염색체 19 맵 상에서의 SNV 검정의 대략적인 위치들이 화살촉 모양으로 나타나 있으며, 이와 함께 *Lrp5* 유전자좌로부터의 그들의 거리(단위: Mb)가 그 아래에 주어져 있다. *Lrp5* 유전자좌로부터의 거리(단위: Mb)는 하기와 같다: *Lrp5*에 대해 동원체 측으로 0.32(C2), *Lrp5*에 대해 말단소체 측으로 1.2(T3), *Lrp5*에 대해 말단소체 측으로 11.1(T6), *Lrp5*에 대해 말단소체 측으로 13.2(T7), *Lrp5*에 대해 말단소체 측으로 17.5(T8), *Lrp5*에 대해 말단소체 측으로 25.8(T9), *Lrp5*에 대해 말단소체 측으로 33.0(T10), *Lrp5*에 대해 말단소체 측으로 38.3(T11), *Lrp5*에 대해 말단소체 측으로 49.6(T13), 그리고 *Lrp5*에 대해 말단소체 측으로 57.2(T14). 129-특이적 및 B6-특이적 프로브 및 프라이머 쌍이 표 1에 나타나 있다.

[0356] 표 3은 SV 및 SNV 대립유전자 둘 모두에 대하여 LOH에 의한 *Lrp5* 표적 유전자좌로부터 말단소체 측 방향으로 염색체 19의 긴 아암에 걸쳐 명백한 유전자 변환 이벤트를 나타낸 ES 세포 클론의 7개의 예를 나타낸다. *Lrp5* 인간화 LTVEC(도 1)를 지시된 바와 같이 1개 또는 2개의 gRNA와 조합한 독립적인 표적화 실험으로부터 ES 세포 클론을 유도하였다. gRNA 인식 부위의 위치들이 도 5에서의 *Lrp5* 유전자의 도면 상에 나타나 있다(좌측방향을 가리키는 두꺼운 화살표). 유전자형 결정 검정은 7개의 클론 중 6개가 *Lrp5* 유전자의 동형접합성 표적화된 인간화를 가졌으며, 한편 1개는 동형접합성 붕괴(gRNA 부위들 사이의 큰 결실)를 가졌음을 나타내었다. 7개의 클론 중 6개에서는, 129 대립유전자는 상실되었으며, 단지 B6 대립유전자만이 남았다. 다른 하나의 클론에서는, B6 대립유전자는 상실되었으며, 단지 129 대립유전자만이 남았다. 모든 클론은 *Lrp5* 유전자좌의 동원체 측에서 검정된 대립유전자에 대해서는 이형접합성을 유지하였다(즉, 모든 클론은 C2 SNV 검정에 대해 이형접합성 B6/129였다). 7개의 클론에서 관찰된 LOH는, 동형접합성 유전자 변형된 대립유전자가, LTVEC가 1개, 또는 더 빈번하게는, 2개의 gRNA와 조합될 때 얻어진 하나의 기전이, 한쪽 대립유전자에 대한 제1 표적화된 유전자 변형 이후의 하나의 염색체로부터 그의 상동체로의 표적화된 유전자 변형을 카피하는 상동성 유도된 재조합 유전자 변환 이벤트를 나타낸다.

[0357] [표 3]

이형접합성의 상실 검정 결과.

클론	gRNA	<i>Lrp5</i> 대립유전자 유형	이형접합성의 상실 검정 (SV 및 SNV)
AW-A7	A + F	동형접합성 표적화됨	단지 B6 대립유전자만이 검출됨
AW-F10	A + F	동형접합성 붕괴	단지 B6 대립유전자만이 검출됨
BO-A8	A + F	동형접합성 표적화됨	단지 129 대립유전자만이 검출됨
BA-D5	A + E2	동형접합성 표적화됨	단지 B6 대립유전자만이 검출됨
BA-F2	A + E2	동형접합성 표적화됨	단지 B6 대립유전자만이 검출됨
BC-H9	B2	동형접합성 표적화됨	단지 B6 대립유전자만이 검출됨
BR-B4	D	동형접합성 표적화됨	단지 B6 대립유전자만이 검출됨

[0358]

[0359] *C5(Hc)* 유전자좌

[0360]

다른 세트의 실험에서는, 상보적 성분 5(*C5* 또는 *Hc*(용혈성 상보체))에 대한 마우스 유전자의 76 kb 결실 및 상동성 인간 *C5* 유전자의 97 kb 단편에 의한 동시 대체를 생성하도록 LTVEC를 설계하였다(도 6). 표적 유전자좌는 *C5(Hc)* 유전자의 엑손 2로부터 정지 코돈까지 포함하였다. LTVEC는 결실이 의도된 마우스 *C5(Hc)* 유전자의 76 kb 서열을 플랭킹하는 마우스 *C5(Hc)* 유전자좌의 부분들로부터 유래된 35 kb 및 31 kb의 게놈 DNA를 함유하는 상동성 아암들에 의해 플랭킹된 인간 *C5* 유전자의 97 kb 단편을 포함하였다. 별도의 실험에서는, *C5(Hc)* 인간화 LTVEC를, 결실에 대해 표적화된 마우스 *C5(Hc)* 유전자의 영역 내에 이중-가닥 브레이크를 생성하도록 설계된, Cas9를 인코딩하는 플라스미드 및 6개의 gRNA(A, B, C, D, E, 및 E2; 표 1 참조) 중 하나를 인코딩하는 제 2 플라스미드와 조합하였다. gRNA들은 인간 *C5* 유전자의 삽입된 부분에서의 임의의 서열의 인식을 피하도록 설계하였다. 다른 실험에서, 본 발명자들은 LTVEC 및 Cas9-인코딩 플라스미드를, 결실에 대해 표적화된 마우스 *C5(Hc)* 유전자의 영역 내의 상이한 부위들을 표적화하는 2개의 상이한 gRNA를 인코딩하는 플라스미드와 조합하였다. 일부 실험에서는, *Ch25h* 유전자좌를 표적화하는 대조군 LTVEC를 *C5(Hc)* 인간화 LTVEC 대신 사용하였다. *Ch25h*의 전체 코딩 서열(약 1 kb)을 결실시키고 푸로마이신 및 네오마이신 선택 카세트를 *Ch25h* 유전자좌 내로 삽입하도록 설계된 대조군 LTVEC를 *C5(Hc)* 유전자좌에서의 상동 재조합에 대해 표적화되지 않은 약물-저항성 클론을 선택하기 위한 수단으로서 사용하였다.

[0361]

C5(Hc) 유전자의 CRISPR/Cas9-보조 인간화의 결과가 표 4에 나타나 있고, 이는 *Lrp5* 유전자의 CRISPR/Cas9-보조 인간화에 대해 얻어진 결과와 유사하다. LTVEC 단독 사용에 의한 표적화 효율은 *C5(Hc)* 인간화에 대한 것이 *Lrp5*에 대한 것보다 더 높았지만(6.1%), Cas9 및 gRNA들의 첨가는 시험된 6개의 gRNA 중 4개에 대해 표적화 효율을 향상시켰다. *Lrp5*와 마찬가지로, *C5(Hc)* 인간화를 위해 gRNA들을 조합함으로써(즉, 2개의 gRNA를 사용함으로써) 총 표적화 효율을 추가로 증가시켰는데, 이는 반접합성 및 동형접합성 표적화 이벤트의 빈도를 증가시킴으로써 주로 이루어졌다. 본 발명자들은 또한 (1.8% 내지 3.6%의 빈도로 관찰되는) 양쪽 대립유전자 상에 큰 CRISPR-유도 결실을 갖는 ES 세포 클론을 발견하였다. 게다가, *Ch25h* 유전자좌를 표적화하는 LTVEC를 2개의 *C5(Hc)* gRNA와 조합하여 사용하였을 때, 2개의 gRNA CRISPR RNA 인식 서열들 사이에서 붕괴된 동형접합성 대립유전자를 갖는 클론이 1.2% 내지 6%의 빈도로 관찰되었는데, 이는 붕괴 이벤트가 표적 유전자좌에서의 상동 재조합 이벤트와는 관계없이 일어남을 나타낸다. *Lrp5*와 마찬가지로, 보유 검정을 사용하여 올바르게 표적화된 클론을 확인하였다. 이러한 스크리닝에 대한 2개의 보유 검정은 하기의 프라이머 및 프로브를 사용하는 TaqMan® 검정이었다: 7140retU 정방향 프라이머 CCCAGCATCTGACGACACC (서열 번호 125); 7140retU 역방향 프라이머 GACCACTGTGGGCATCTGTAG (서열 번호 126); 7140retU TaqMan® 프로브 CCGAGTCTGCTTACTGTTAGCATCA (서열 번호 127); 7140retD 정방향 프라이머 CCCGACACCTTCTGAGCATG (서열 번호 128); 7140retD 역방향 프라이머 TGCAGGCTGAGTCAGGATTG (서열 번호 129); 7140retD TaqMan® 프로브 TAGTCACGTTTTGTGACACCCAGA (서열 번호 130).

[0362] [표 4]

개별 gRNA 들 및 조합된 gRNA 들을 사용한 *C5(Hc)* 유전자의 CRISPR/Cas9-보조 인간화에 대한 스크리닝 결과.

gRNA	표적화된 결실의 5'/3' 말단으로부터 gRNA 부위까지의 거리(bp)	LTVEC	대립유전자 유형에 따른 표적화 효율				
			이형접합성 표적화 (% 효율)	반접합성 표적화 (% 효율)	동형접합성 표적화 (% 효율)	총 표적화 (% 효율)	동형접합성 결실 (% 효율)
A	100 (5')	C5	16.6			16.6	
B	500 (5')	C5	14.5			14.5	
C	38200 (5')/37500 (3')	C5	11.4			11.4	
D	43500 (5')/32200 (3')	C5	7.3			7.3	
E	500 (3')	C5	4.2			4.2	
E2	100 (3')	C5	6.2			6.2	
A + C	A: 100 (5') C: 37500 (3')	C5	19.6	7.1	0.6	27.3	0.6
A + C	A: 100 (5') C: 37500 (3')	Ch25h	N/A	N/A	N/A	N/A	6.0
A + E2	A: 100 (5') E2: 100 (3')	C5	19.0	3.6	1.2	23.8	3.0
A + E2	A: 100 (5') E2: 100 (3')	Ch25h	N/A	N/A	N/A	N/A	1.2
없음	N/A	C5	6.1			6.1	

[0363]

[0364]

형광 동소 혼성화(FISH)를 사용하여 *C5(Hc)* 유전자의 동형접합성 표적화된 인간화를 확인하였다. 정량적 및 종래의 PCR 검정에 의해, *C5(Hc)* 인간화 LTVEC(도 6)가 Cas9 및 2개의 gRNA와 조합된 표적화 실험으로부터 동형접합성 표적화된 것으로 스코어링된 ES 세포 클론을 FISH 및 핵형 분석을 위한 상업용 세포검사 서비스로 보냈다. 마우스 *C5(Hc)* 유전자를 갖는 세균 인공 염색체(BAC)는 적색 형광 마커로 표지화하여 내인성 유전자좌를 확인하기 위한 프로브로 사용하고, 인간 *C5* 유전자를 갖는 BAC는 녹색 형광 마커로 표지화하여 인간 삽입물로 표적화된 염색분체를 확인하기 위한 프로브로 사용하였다. 표지화된 BAC 프로브를 표적화된 클론으로부터의 중기 확산체에 혼성화하고, 형광 현미경법에 의해 시각화하였다. 확산체 상의 염색체를 DAPI(4',6-다이아미디노-2-페닐인돌)로 염색하여 시각화하고, 각각의 클론에 대한 개별 핵형들을 김자 염색에 의해 결정하였다. 전형적인 결과가 클론 O-E에 대하여 도 7b에 나타나 있다. 도 7b에서의 합성 사진은 적색 마우스 BAC 프로브 신호 및 녹색 인간 BAC 프로브 신호 둘 모두가 *C5(Hc)* 유전자의 공지된 위치인 마우스 염색체 2의 두 카피 모두 상의 *C5(Hc)* 유전자좌에 공국재화되었음을 보여준다. 이들 결과는 인간화 LTVEC(도 6)에서의 인간 *C5* 유전자의 97 kb 단편이 클론 O-E3에서의 두 염색체 2 상동체 모두 상의 의도된 마우스 *C5(Hc)* 유전자좌에 올바르게 삽입되었음을 확인시켜 준다. 대조적으로, 도 7a에서의 합성 사진은 적색 마우스 BAC 프로브 신호 및 녹색 인간 BAC 프로브 신호 둘 모두가 마우스 염색체 2의 단일 카피(실선 화살표) 상에 공국재화되었으며, 한편 단지 적색 마우스 BAC 프로브 신호만이 마우스 염색체 2의 다른 하나의 카피 상의 *C5(Hc)* 유전자좌에 국재화됨을 나타낸다. 이들 결과는 인간화 LTVEC(도 6)에서의 인간 *C5* 유전자의 97 kb 단편이 클론 Q-E9에서의 염색체 2의 단지 하나의 카피 상의 의도된 마우스 *C5(Hc)* 유전자좌에 올바르게 삽입되었음(이형접합성 표적화)을 확인시켜 준다.

[0365]

Ror1 유전자좌

[0366]

다른 세트의 실험에서는, 마우스 *Ror1*(티로신-단백질 키나제 막관통 수용체 ROR1) 유전자의 110 kb 결실 및 상동성 인간 *ROR1* 유전자의 134 kb 단편에 의한 동시 대체를 생성하도록 LTVEC를 설계하였다(도 8). LTVEC는 결실이 의도된 마우스 *Ror1* 유전자의 110 kb 서열을 플랭킹하는 마우스 *Ror1* 유전자좌의 부분들로부터 유래된 41.8 kb 및 96.4 kb의 게놈 DNA를 함유하는 상동성 아암들에 의해 플랭킹된 인간 *ROR1* 유전자의 134 kb 단편을 포함하였다. 별도의 실험에서는, *Ror1* 인간화 LTVEC를, 결실에 대해 표적화된 마우스 *Ror1* 유전자의 영역 내에 이중-가닥 브레이크를 생성하도록 설계된, Cas9를 인코딩하는 플라스미드 및 6개의 gRNA(A, B, C, D, E, 및 F; 표 1 참조) 중 하나를 인코딩하는 제2 플라스미드와 조합하였다. gRNA들은 인간 *ROR1* 유전자의 삽입된 부분에서의 임의의 서열의 인식을 피하도록 설계하였다. 다른 실험에서, 본 발명자들은 LTVEC 및 Cas9-인코딩 플라스미드를, 결실에 대해 표적화된 *Ror1* 유전자 내의 상이한 부위들을 표적화하는 2개의 상이한 gRNA를 인코딩하는 플라스미드와 조합하였다.

[0367]

Ror1 유전자의 CRISPR/Cas9-보조 인간화의 결과가 표 5에 나타나 있고, 이는 *Lrp5* 및 *C5(Hc)* 유전자의 CRISPR/Cas9-보조 인간화에 대해 얻어진 결과와 유사하다. LTVEC의 단독 사용에 의한 표적화 효율은 0.3%였고, Cas9 및 gRNA들의 첨가는 시험된 6개의 gRNA 중 2개에 대해 표적화 효율을 약간 증가시켰다. A gRNA와 F gRNA를 조합함으로써 총 *Ror1* 표적화 효율을 6.3%로 증가시켰는데, 이는 이형접합성 및 반접합성 표적화 이벤트 둘 모두의 빈도를 증가시킴으로써 이루어졌다. 본 발명자들은 또한 (1.6%의 빈도로 관찰되는) 양쪽 대립유전자 상에 큰 CRISPR-유도 결실을 갖는 ES 세포 클론을 발견하였다.

[0368] [표 5]

개별 gRNA 들 및 조합된 gRNA 들을 사용한 *Ror1* 유전자의 CRISPR/Cas9-보조 인간화에 대한 스크리닝 결과.

gRNA	표적화된 결실의 5'/3' 말단으로부터 gRNA 부위까지의 거리(bp)	대립유전자 유형에 따른 표적화 효율				
		이형접합성 표적화 (% 효율)	반접합성 표적화 (% 효율)	동형접합성 표적화 (% 효율)	총 표적화 (% 효율)	동형접합성 결실 (% 효율)
A	200 (5')	0.7			0.7	
B	1000 (5')	0.0			0.0	
D	54300 (5')/ 55500 (3')	0.7			0.7	
C	54500 (5')/ 55300 (3')	0.0			0.0	
E	1000 (3')	0.0			0.0	
F	200 (3')	0.3			0.3	
A + F	A: 200 (5') F: 200 (3')	4.2	2.1		6.3	1.6
없음	N/A	0.3			0.3	

[0369]

[0370] *Trpa1* 유전자좌

[0371] 다른 세트의 실험에서는, 마우스 *Trpa1*(과도 수용기 전위 양이온 채널(transient receptor potential cation channel)), 하위패밀리 A, 구성원 1) 유전자의 45.3 kb 결실 및 상동성 인간 *TRPA1* 유전자의 54.5 kb 단편에 의한 동시 대체를 생성하도록 LTVEC를 설계하였다(도 9). LTVEC는 결실이 의도된 마우스 *Trpa1* 유전자의 45.3 kb 서열을 플랭킹하는 마우스 *Trpa1* 유전자좌의 부분들로부터 유래된 41.0 kb 및 58.0 kb의 게놈 DNA를 함유하는 상동성 아암들에 의해 플랭킹된 인간 *TRPA1* 유전자의 54.5 kb 단편을 포함하였다. 별도의 실험에서는, *Trpa1* 인간화 LTVEC를, 결실에 대해 표적화된 마우스 *Trpa1* 유전자의 영역 내에 이중-가닥 브레이크를 생성하도록 설계된, Cas9를 인코딩하는 플라스미드 및 8개의 gRNA(A, A2, B, C, D, E2, E, 및 F; 표 1 참조) 중 하나를 인코딩하는 제2 플라스미드와 조합하였다. gRNA들은 인간 *TRPA1* 유전자의 삽입된 부분에서의 임의의 서열의 인식을 피하도록 설계하였다. 다른 실험에서, 본 발명자들은 LTVEC 및 Cas9-인코딩 플라스미드를, 결실에 대해 표적화된 *Trpa1* 유전자 내의 상이한 부위들을 표적화하는 2개의 상이한 gRNA를 인코딩하는 플라스미드와 조합하였다.

[0372] *Trpa1* 유전자의 CRISPR/Cas9-보조 인간화의 결과가 표 6에 나타나 있고, 이는 *Lrp5* 및 *C5(Hc)* 유전자의 CRISPR/Cas9-보조 인간화에 대해 얻어진 결과와 유사하다. LTVEC의 단독 사용에 의한 표적화 효율은 0.3%였고, Cas9 및 gRNA들의 첨가는 시험된 8개의 gRNA 중 6개에 대해 표적화 효율을 증가시켰다. B gRNA와 F gRNA를 조합함으로써 총 *Trpa1* 표적화 효율을 3.4%로 증가시켰는데, 이는 이형접합성, 반접합성, 및 동형접합성 표적화 이벤트의 빈도를 증가시킴으로써 이루어졌다. 본 발명자들은 또한 (0.3%의 빈도로 관찰되는) 양쪽 대립유전자 상에 큰 CRISPR-유도 결실을 갖는 ES 세포 클론을 발견하였다.

[0373] [표 6]

개별 gRNA 들 및 조합된 gRNA 들을 사용한 *Trpa1* 유전자의 CRISPR/Cas9-보조 인간화에 대한 스크리닝 결과.

gRNA	표적화된 결실의 5'/3' 말단으로부터 gRNA 부위까지의 거리(bp)	대립유전자 유형에 따른 표적화 효율				
		이형접합성 표적화 (% 효율)	반접합성 표적화 (% 효율)	동형접합성 표적화 (% 효율)	총 표적화 (% 효율)	동형접합성 결실 (% 효율)
A	100 (5')	1.0			1.0	
A2	500 (5')	2.1			2.1	
B	1000 (5')	1.4			1.4	
C	25600 (5')/ 19740 (3')	1.0			1.0	
D	26970 (5')/ 18370 (3')	2.1			2.1	
E2	1000 (3')	0.0			0.0	
E	500 (3')	0.0			0.0	
F	100 (3')	0.7			0.7	
B + F	B: 1000 (5') F: 100 (3')	2.8	0.3	0.3	3.4	0.3
없음	N/A	0.3			0.3	

[0374]

[0375] 이들 실시예가 예시하는 바와 같이, 간격이 넓은 부위들에서의 이중 가이드 RNA들의 사용은 단일 gRNA와 비교하여 이형접합성 인간화의 향상을 개선했었다. 게다가, 이중 가이드 RNA들의 사용은 단일 gRNA와 비교하여 이중 대립유전자 이벤트를 촉진시켰다. 1개의 gRNA에 의해 표적화하는 것과 대조적으로, 2개의 gRNA에 의해 표적화하는 것은, 양쪽 대립유전자가 표적화된 인간화를 갖는 동형접합성 표적화된 세포(Hum/Hum), 어느 대립유전자도 인간화 LTVEC에 의해 표적화되지 않지만 양쪽 다 큰 결실을 갖는 동형접합성 결실된 세포(Δ/Δ), 및 한쪽 대립유전자는 표적화된 인간화를 갖고 다른 한쪽은 큰 이중 gRNA/Cas9-유도 결실을 갖는 반접합성 표적화된 세포

(Hum/ Δ)의 생성을 가져온다. 첫째, 본 발명자들은 양쪽 표적 대립유전자에서 정확하고 동일한 매우 큰 인간화를 갖는 올바르게 표적화된 클론(예를 들어, 표적화된 유전자 변형에 대해 동형접합성인 세포)을 발견하였다. 본 발명자들이 1개의 gRNA를 사용하여 *Lrp5* 인간화를 달성하였을 때 동형접합성 표적화된 클론이 또한 관찰되었지만, 이것은 2개의 gRNA를 사용하였을 때보다 훨씬 더 낮은 빈도로 발생되었다(표 2 참조). 마찬가지로, 본 발명자들은 1개의 gRNA를 사용하여 *C5(Hc)* 인간화 또는 *Trpa1* 인간화를 달성하였을 때에는 동형접합성 표적화를 관찰하지 못하였지만, 표적화 벡터와 함께 2개의 gRNA를 사용하였을 때에는 동형접합성 표적화를 관찰하였다(표 4 및 표 6 참조). 유사하게, 본 발명자들은 *Lrp5* 표적화, *C5(Hc)* 표적화, *Ror1* 표적화, 및 *Trpa1* 표적화에 대한 유전자 변형에 대해 반접합성인 올바르게 표적화된 클론을 발견하였다(즉, 이것은 한쪽 대립유전자 상에서는 정확하게 표적화된 인간화를, 그리고 다른 한쪽 대립유전자 상에서는 매우 큰, 때로는 유전자 제거인, 결실을 가졌다). 그러한 변형은 1개의 gRNA를 사용하여 *Lrp5*, *C5(Hc)*, *Ror1*, 또는 *Trpa1* 인간화를 달성하였을 때에는 전혀 일어나지 않았다(표 2, 표 4, 표 5, 및 표 6을 각각 참조).

[0376] 둘째, 본 발명자들은 양쪽 표적화된 대립유전자 상에서 두 gRNA 모두에 의해 가이드된 Cas9 절단 이벤트에 의해 유도된 동일한 매우 큰 결실(45 kb 초과)을 갖는 클론을 발견하였다(즉, 세포는 표적 유전자좌에서 큰, 때로는 유전자 제거인, 결실에 대해 동형접합성이었다). 이들 유형의 돌연변이는 동일한 유전자에 대해 유도된 표적화 벡터를 필요로 하지 않았다. 예를 들어, 표 4에 나타난 바와 같이, 본 발명자들은 Cas9 및 2개의 gRNA를 이들 gRNA에 의해 표적화된 것과 관련되지 않은 상이한 유전자에 대해 유도된 표적화 벡터와 조합함으로써 동형접합성 CRISPR-유도 결실을 갖는 ES 세포를 얻었다. 따라서, 2개의 gRNA에 의해 가이드된 Cas9 뉴클레아제는 표적화 벡터의 첨가 없이 세포 내에 큰 결실을 유도할 수 있다. 그러한 경우에, 약물 저항성 유전자를 발현하는 벡터에 의해 제공되는 일시적 또는 안정한 약물 선택은 DNA를 흡수한 ES 세포에 대한 농축에 의해 회귀한 동형접합성 결실 클론의 단리를 촉진시킬 수 있다.

[0377] 실시예 2. 조합된 gRNA들에 의해 유도된 큰 결실의 분석.

[0378] 조합된 gRNA들에 의해 유도된 큰 결실에 대한 대립유전자 구조

[0379] 2개의 gRNA에 의해 가이드된 Cas9 절단 이벤트에 의해 유도된 큰 결실을 포함하는 클론에 대해 추가 서열 분석을 수행하였다(표 7 참조). 이들 큰 결실은 동일한 유전자좌에서의 LTVEC-유도 상동 재조합 이벤트와 무관한 것으로 나타났는데, 이는 본 발명자들이 gRNA들을 *Lrp5* LTVEC 또는 거의 30 Mb만큼 떨어져 있는 *Ch25h* 유전자를 표적화하는 것과 조합하였을 때 대략적으로 동일한 빈도로 *Lrp5* 유전자좌에서 큰 결실을 얻었다는 점에서 그러하다(데이터는 도시되지 않음). 큰 결실을 특성화하기 위하여, 본 발명자들은 4회의 인간화로부터의 37개의 클론 - 15개는 반접합성 클론이고, 22개는 이중대립유전자 큰 결실을 갖는 클론임 - 에 대해 결실-스패닝 PCR을 수행하고, PCR 산물의 개별 클론들을 서열분석하였다. 서열은 38 kb 내지 109 kb 범위의 큰 결실을 확인시켜 주었다. ES 세포 클론들 중 2개(*Lrp5* 클론 AW-A8 및 BP-D3)는 예측된 Cas9 절단 부위들 사이에서 정확한 결실(68.2 kb)을 완전히 수복하였으며, 한편 하나의 클론(*Hc* 클론 P-B12)은 38.1 kb 결실에 더하여 단일 염기쌍 삽입을 가졌다. ES 세포 클론들 중 27개는 비상동 말단 결합(NHEJ)에 의한 부정확한 수복과 일치하는, Cas9 절단 부위를 넘어서 연장된 결실을 가졌다. 나머지 7개의 ES 세포 클론은 명백한 NHEJ-유도 결실 및 삽입을 조합한 돌연변이(예를 들어, *Lrp5* 클론 BP-F6 및 *Hc* 클론 O-E4)를 가졌으며, 이들 중 4개는 본 발명자들이 그들의 소스 게놈 유전자좌에 매핑할 수 있는 200 bp 초과 삽입을 가졌다(데이터는 도시되지 않음). *Lrp5* 클론 BO-E9에의 210 bp 삽입은 동원체 방향으로 gRNA F 표적 부위의 대략 2,600 bp 외부에 놓여 있는 동일한 서열(염색체 19+, 3589138-3589347)에 대해 반전된 배향 상태에 있다. 이 서열은 *Lrp5* LTVEC의 긴 3' 상동성 아암에 존재하였다. *Lrp5* 클론 BP-F6 및 BP-G7은, 본 발명자들이 *Lrp5* gRNA A 및 F를 Cas9 및 *Lrp5*로부터 말단소체 방향으로 30 Mb만큼 떨어진 *Ch25h* 유전자를 표적화한 LTVEC와 조합한 실험으로부터 유래되었다. 클론 BP-F6은 *Ch25h* LTVEC의 한쪽 말단으로부터 유래된 것으로 보이는 266 bp 삽입을 가졌는데, 이는 그것이 *Ch25h* 부근에 있고 또한 LTVEC의 긴 아암에 존재하는 서열(염색체 19+, 34478136-34478298)과 동일한 163 bp 단편에 연결된 벡터 골격의 일부와 동일한 103 bp 단편으로 구성되었다는 점에서 그러하며; 이러한 단편을 내인성 염색체 서열에 대해 반전된 배향으로 결실에서 삽입하였다. *Hc* 클론 O-E4는 gRNA A 인식 부위로부터 대략 3.1 kb만큼 떨어진 결실된 서열 내에서 발견되는 동일한 서열에 대해 반전된 254 bp 삽입을 가졌다. *Hc* 클론 S-D5에서의 1,304 bp 삽입은 하기의 2개의 단편으로 구성되었다: 예측된 gRNA E2-유도 Cas9 절단 부위로부터 대략 1.4 kb만큼 떨어져 있는 결실된 서열 내에서 발견되는 것과 동일한 서열과 동일한 배향에 있는 1,238 bp 조각 및 gRNA E2 컷 부위의 25 bp 외부에 있는 동일한 서열의 반전된 배향으로의 중복인 두 번째 66 bp 조각.

[0380] [표 7]

조합된 gRNA 들에 의해 유도된 큰 결실에 대한 대립유전자 구조

유전자	ES 세포 클론	유전자형 ¹	gRNA	표적화된 결실 내의 위치 (bp)	결실의 크기 (kb)	결실된 추가의 서열 (bp)	삽입 (bp)	PCR 클론
<i>Lrp5</i>	AW-A8	Δ/Δ	A + F	5'-50/50-3'	68.2	-	-	40
	BO-E9	Δ/Δ				12	210	17
	BP-D3	Δ/Δ				-	-	11
	BP-F6	Δ/Δ				30	266	6
	BP-G7	Δ/Δ				77		9
	BA-A7	Δ/Δ	A + E2	5'-50/1,000-3'	67.3	7		19
	BA-C7	Δ/Δ				84		32
<i>Hc</i>	N-A11	Δ/Δ	A + C	5'-100/38,200-3'	38.1	14		12
	N-D4	Δ/Δ				10		15
	N-D11	Hum/Δ				20		10
	N-E1	Hum/Δ				10		1
	N-E9	Hum/Δ				10		13
	N-E9	Hum/Δ				20		16
	O-C5	Hum/Δ				31		21
	O-D2	Hum/Δ				5		12
	O-E4	Hum/Δ				19	254	18
	O-E5	Hum/Δ				35	2	16
	O-E6	Hum/Δ				6		17
	O-F11	Hum/Δ				12	7	18
	O-F12	Hum/Δ				41		6
	P-B12	Δ/Δ				35		1
	P-C12	Δ/Δ					1	7
	P-D1	Δ/Δ				20		15
	P-G8	Δ/Δ				33		10
	Q-F5	Hum/Δ	A + E2	5'-100/100-3'	75.6	5		2
	Q-F10	Δ/Δ				3	3	15
	R-A5	Δ/Δ				46		13
	R-A7	Δ/Δ				18		14
	R-A9	Hum/Δ				37		15
	R-C8	Hum/Δ				261		8
	R-D12	Hum/Δ				180		11
	R-F11	Hum/Δ				182		10
	S-A11	Δ/Δ				19		11
	S-D5	Δ/Δ				122		11
<i>Ror1</i>	Y-B5	Δ/Δ	A + F	5'-200/200-3'	109	46		1
	Y-C7	Δ/Δ				216	1304	8
	Y-E1	Δ/Δ				18		6
	AD-C7	Δ/Δ				23		7
<i>Trpa1</i>	AD-C7	Δ/Δ	B + F	5'-1,000/100-3'	44.6	12		3
						30		8

¹Hum/+, 2 개의 천연 대립유전자 중 하나의 표적화된 인간화로, 이는 이형접합성 유전자형을 생성함; Hum/Δ, 한쪽 대립유전자는 표적화된 인간화를 갖고 다른 한쪽은 큰 Cas9-gRNA-유도 결실을 갖는 이중대립유전자 변형으로, 이는 반접합성 유전자형을 생성함; Hum/Hum, 양쪽 대립유전자가 표적화된 인간화를 갖는 이중대립유전자 변형으로, 이는 동형접합성 유전자형을 생성함; Δ/Δ, 양쪽 대립유전자가 큰 Cas9-gRNA-유도 결실을 갖는 이중대립유전자 변형.

[0381]

[0382] 동형접합성 대립유전자에서의 유전자 변환에 대한 증거

[0383]

이중대립유전자 큰 결실을 갖는 22개의 ES 세포 클론 중 21개는 단지 단일의 특유 서열을 가졌는데(표 7), 이는 이들이 동형접합성 대립유전자였음을 나타낸다. *Hc* 클론 S-A11의 경우, 본 발명자들은 12개의 PCR 클론 중 11 개에서 동일한 서열을 발견하였다. 상이한 서열을 갖는 단일 클론은 2개의 상이한 결실 대립유전자를 시사할 수 있지만, 본 발명자들은 또한 *Hc* 반접합성 클론들 중 2개, 즉 N-D11 및 O-F12에 대해서도 동일한 결과를 확인 하였다. 다수의 클론에서의 별개의 동형접합성 결실 대립유전자들은, 하나의 염색체 상의 결실이 상동 염색체 상의 Cas9 절단의 상동 재조합 수복에 대한 주형으로서의 역할을 한 유전자 변환 기전에 의해, 그들이 발생되었 을 수 있음을 시사하였다. 본 발명자들은 VGF1 ES 세포주의 129S6SvEvTac (129) 및 C57BL/6NTac (B6) F1 혼성 조성물(문헌[Poueymirou et al. (2007) *Nat. Biotechnol.* 25:91-99]; 문헌[Valenzuela et al. (2003) *Nat. Biotechnol.* 21:652-659])을 이용하여, 염색체 19 상의 *Lrp5* 유전자좌(하기에 사용된 5개의 SV 검정 및 10개의 SNV 검정에 대해서는 도 5를 참조) 및 염색체 2 상의 *Hc* 유전자좌(도시되지 않음) 주위의 계통들 사이의 구조 (SV) 및 단일 뉴클레오티드(SNV) 변이체에 대한 이형접합성의 상실로서의 유전자 변환에 대해 검정하였다(문헌 [Lefebvre et al. (2001) *Nat. Genet.* 27:257-258]). 임의의 이형접합성의 상실이 전체 염색체 상실의 결과가 아니었음을 확인하기 위하여, 본 발명자들은 129 계통과 B6 계통 사이의 동일한 부위에서 염색체 카피수(CCN) 검정을 수행하였다. *Lrp5* 인간화 또는 결실된 대립유전자의 경우, 본 발명자들은 *Lrp5*로부터 말단초체 방향으로 1.2 Mb만큼 떨어진 곳으로부터 염색체 19의 긴 아암의 말단까지 위치된 다수의 SV 및 SNV를 검정하였다(도 5). *Lrp5*의 위치가 동원체에 가깝기 때문에, 본 발명자들은 SV는 발견하지 못하였고 유전자의 동원체 측에서

단지 하나의 SNV만을 발견하였다. *Hc*의 경우, 본 발명자들은 염색체 2 상의 유전자 양측에서 다수의 SV 및 SNV에 대해 검정할 수 있었다(도시되지 않음). *Lrp5* 클론들 중 6개에 대한 결과가 도 10a 내지 도 10e 및 도 11a 내지 도 11c에 나타나 있다.

[0384] 도 10a 내지 도 10e는 5개의 SV 검정에 대한 결과를 나타내는데, 5개의 SV 검정의 위치들은 *Lrp5*로부터 13.7 Mb만큼 떨어진 곳으로부터 긴 아암의 말단소체 말단 부근에서 56.7 Mb만큼 떨어진 곳까지의 범위였다. 5개의 SV 검정은 129, B6, 및 VGF1 대조군에서 129(더 큰) 및 B6(더 작은) 대립유전자에 대해 2개의 상이한 크기의 산물을 생성하였다. 염색체 19 맵 상의 SV 검정의 대략적인 위치들이 도 5에 나타나 있다(검정 SV 13.7, 검정 SV 20.0, 검정 SV 36.9, 검정 SV 48.3, 및 검정 SV 56.7을 참조한다). 검정 수는 *Lrp5*에 대해 말단소체 측으로의 Mb의 수를 나타낸다. 이들 검정에 대한 프라이머가 표 1에 나타나 있으며, 결과는 도 10a 내지 도 10e에 나타나 있다. 클론들 중 2개, BC-H9(*Lrp5*^{Hum/Hum}, gRNA B2) 및 BR-B4(*Lrp5*^{Hum/Hum}, gRNA D)는 B6 SV 대립유전자 전부를 보유한 이형접합성의 상실을 나타내었으며, 한편 제3 클론, BO-A8(*Lrp5*^{Hum/Hum}, gRNA A + F)은 129 대립유전자 전부를 보유하였다. 나머지 3개의 클론, BO-F10(*Lrp5*^{Hum/Hum}, gRNA A + F), BO-G11(*Lrp5*^{Hum/Hum}, gRNA A + F), 및 BP-G7(*Lrp5*^{Δ/Δ}, gRNA A + F)은 이형접합성을 유지하였다.

[0385] 게다가, 129 대립유전자와 B6 대립유전자 사이의 단일 뉴클레오타이드 변이체(SNV)를 TaqMan® 대립유전자 식별 검정에 의해 검정하였다. 도 5에서의 염색체 19 맵 상에서의 SNV 검정의 대략적인 위치들이 화살촉 모양으로 나타나 있으며, 이와 함께 그 바로 아래에는 검정 수가 있고, *Lrp5* 유전자좌로부터의 그들의 거리(단위: Mb)가 그 아래에 주어져 있다. *Lrp5* 유전자좌로부터의 거리(단위: Mb)는 하기와 같다: *Lrp5*에 대해 동원체 측으로 0.32(C2), *Lrp5*에 대해 말단소체 측으로 1.2(T3), *Lrp5*에 대해 말단소체 측으로 11.1(T6), *Lrp5*에 대해 말단소체 측으로 13.2(T7), *Lrp5*에 대해 말단소체 측으로 17.5(T8), *Lrp5*에 대해 말단소체 측으로 25.8(T9), *Lrp5*에 대해 말단소체 측으로 33.0(T10), *Lrp5*에 대해 말단소체 측으로 38.3(T11), *Lrp5*에 대해 말단소체 측으로 49.6(T13), 그리고 *Lrp5*에 대해 말단소체 측으로 57.2(T14). 129-특이적 및 B6-특이적 프로브 및 프라이머 쌍이 표 1에 나타나 있다. SV 검정에 의한 말단소체 측 이형접합성의 상실(LOH)을 보여준 3개의 클론(BC-H9, BO-A8, 및 BR-B4)에 대한 결과는 도 11a 내지 도 11c에 나타나 있다. SNV 검정(도 11a 내지 도 11c; 데이터는 도시되지 않음)은 *Lrp5*의 말단소체 측에서의 염색체 19의 긴 아암에 걸친 유전자 변환 이벤트를 확인시켜 주었지만(SNV 1.2 및 SNV 57.2; 각각 도 11b 및 도 11c 참조), SNV 0.32 검정(도 11a 참조)은 모든 클론은 동원체 측에서 *Lrp5*로부터 320 kb만큼 떨어져 있는 대립유전자에 대해 이형접합성을 유지하였음을 보여주었다. 검정된 24개의 *Lrp5*^{Hum/Hum} 또는 *Lrp5*^{Δ/Δ} 클론 중에서, 본 발명자들은 *Lrp5*의 말단소체 측에서의 염색체 19의 전체 긴 아암에 걸쳐 이형접합성의 상실의 증거를 가진 6개를 찾아내었다. 클론들 중 5개(4개의 *Lrp5*^{Hum/Hum} 및 1개의 *Lrp5*^{Δ/Δ})는 이형접합성 B6으로부터 동형접합성 B6으로 변환되었으며, 한편 6개의 클론(*Lrp5*^{Hum/Hum})은 동형접합성 129로 변환되었다. CCN 검정은 염색체 19의 2개의 카피의 보유를 입증하였다. 21개의 *Hc* 동형접합성 클론에 대한 유사한 이형접합성의 상실 검정은 2개, R-E2(*Hc*^{Hum/Hum}, gRNA A + F) 및 R-E8(*Hc*^{Δ/Δ}, gRNA A + F)가, *Hc* 유전자의 동원체 측의 모든 대립유전자에 대해 이형접합성을 보유하면서, 말단소체 측에서의 모든 SV 및 SNV에 대해 동형접합성 129에 대해 이형접합성의 상실을 나타내었음을 보여주었다. CCN 검정은 염색체 2의 무상실을 나타내었다.

[0386] 본 발명자들의 결과는 CRISPR/Cas9가 100 kb 초과인 큰 단일-단계 인간화에 대한 상동성-유도 수복을 향상시킬 수 있음을 최초로 입증하며, 이는 대규모 게놈 조작에 대한 가능성을 확장시킨다. LTVEC와 gRNA/Cas9의 조합의 가장 현저하고 예기치 않은 이득은 동형접합성 표적화된 인간화를 촉진시키는 그들의 능력이었다. 이중대립유전자 돌연변이 및 동형접합성 표적화 이벤트가 다른 CRISPR/Cas9 실험에서 보고되어 있지만, 대부분의 이들 유전자 변형 및 삽입은 본 발명자들의 인간화 대립유전자보다 수 자리수만큼 더 작았다. CRISPR/Cas9의 사용 전에, 본 발명자들은 LTVEC에 의한 동형접합성 표적화를 결코 확인하지 못하였으며, 또한 본 발명자들은 별개의 유전자들을 표적화하는 다수의 LTVEC들을 조합하였을 때 하나 초과인 유전자의 동시 표적화를 관찰하지 못하였다. 이러한 경험을 고려해 볼 때, gRNA/Cas9-유도 동형접합성 표적화는, 양쪽 대립유전자를 개별적으로 표적화하는 2개의 LTVEC보다는 오히려, 한쪽 대립유전자에 대한 초기 표적화 이벤트가 하나 이상의 Cas9 컷에 의해 촉진된 다른 한쪽 대립유전자의 상동성 변환에 대한 주형으로서의 역할을 할 수 있음을 시사하였다. 이중 gRNA/Cas9-유도 큰 이중대립유전자 결실이 또한 동형접합성이라고 밝혀진 사실(표 7)은 유전자 변환 이전에 대한 추가의 뒷받침을 제공하였다.

[0387] 이형접합성의 상실 검정(도 5)은 표적 유전자의 말단소체 측에서의 염색체의 큰 단편을 커버하는 다수의 대립유전자의 대규모 유전자 변환이 동형접합성 인간화 및 큰 결실의 일부를 담당함을 입증하였다. 이러한 유형의 긴 범위 유도성 유전자 변환은 세포 주기의 G2 기에서의 상동 염색체의 복제된 염색분체들 사이의 유사분열 재조합과 일치한다(문헌[Lefebvre et al. (2001) *Nat. Genet.* 27:257-258])(도 12). 단지 소수의 동형접합성 이벤트만을 설명하였지만, 이 기전은 gRNA/Cas9 절단을 사용하여 염색체의 대부분에 걸쳐 다수의 대립유전자에 대해 이형접합성으로부터 동형접합성으로의 대규모 변환을 촉진시킬 수 있는 수단을 제공할 수 있었다. 그러나, 동형접합성 이벤트들의 대부분은 국소 유전자 변환의 결과였던 것으로 보이는데, 이의 기전은 추가의 조사를 할 만한 가치가 있다.

[0388] 긴 범위 유도성 유전자 변환에 대한 추가의 증거는, *Lrp5* gRNA A 및 F를 인코딩하는 플라스미드, Cas9를 인코딩하는 플라스미드, 및 *Lrp5*로부터 말단소체 측 방향으로 30 Mb만큼 떨어져 있는 *Ch25h* 유전자를 표적화한 LTVEC를 사용하여 F1H4 혼성 ES 세포(이는 50% 129SvS6 계통 및 50% C57BL/6N 계통으로 구성됨)를 전기영동한 후에 얻어진 3개의 클론의 분석에 의해 제공되었다. 3개의 클론은 초기에는 2개의 gRNA들 사이의 예측된 결실 내부에서의 TaqMan® 검정을 사용한 1차 스크리닝 후에 야생형으로 스코어링되었지만(5' 말단에서 500 bp만큼 그리고 3' 말단에서 2 kb만큼 떨어져 있음), 129 대립유전자와 B6 대립유전자 사이의 단일 뉴클레오티드 변이체(SNV)들을 검정하는 후속의 TaqMan® 대립유전자 식별 검정은 뜻밖에도 이형접합성의 상실을 보여주었다. 사용된 SNV 검정은 1개의 동원체 측 검정(SNV 0.32) 및 2개의 말단소체 측 검정(SNV 1.2 및 SNV 57.2)이었다(도 5 참조). 표 8에 나타난 바와 같이, 동원체 측 SNV 검정(0.32 Mb)은 모든 3개의 클론에서 이형접합성의 보유를 확인시켜 주었다. 그러나, 말단소체 측 SNV 검정 둘 모두는 BP-E7 및 BP-H4가 129 대립유전자에 대해 동형접합성임을 보여주었고, 말단소체 측 SNV 검정 둘 모두는 BP-E6이 B6 대립유전자에 대해 동형접합성임을 보여주었다. 모든 3개의 클론은 염색체 19의 2개의 카피의 보유를 보여주었고, 모든 3개의 클론은 LTVEC 표적화(즉, *Ch25h* 유전자좌를 표적화함)에 대해 유전자도입을 나타내었다. 이들 결과는 표적화된 CRISPR/Cas9 절단을 사용한 강제(forced) 동형접합성에 대한 가능성을 열어준다.

[0389] [표 8]

SNV 대립유전자 식별 검정에 대한 스크리닝 결과.

클론	SNV 0.32	SNV 1.2	SNV 57.2
BP-E7	129 / B6	129 / 129	129 / 129
BP-H4	129 / B6	129 / 129	129 / 129
BP-E6	129 / B6	B6 / B6	B6 / B6

[0390]

[0391] 여러 가능한 기전이 마우스 F1H4 혼성 ES 세포(이는 50% 129SvS6 계통 및 50% C57BL/6N 계통으로 구성됨)에서의 CRISPR/Cas9-보조 LTVEC 인간화 실험에서 관찰된 결과를 설명할 수 있다(도 16a 내지 도 16f 참조). 그러한 기전은 유사분열 교차에 의한 상호 염색분체 교환을 통해(도 16a 내지 도 16c), 또는 브레이크-유도 복제에 의한 염색분체 카피에 의해(도 16d 및 도 16e) 일어날 수 있다. 둘 중 어떠한 경우에도, 이형접합성 변형이 일어날 수 있는데, 여기서는 129 염색체 또는 B6 염색체가 게놈 복제 전에 LTVEC에 의해 표적화된다(도 16a 및 도 16d 참조). 대안적으로, 단일 129 염색분체 또는 단일 B6 염색분체가 게놈 복제 후에 LTVEC에 의해 표적화된 후, 염색분체간 유전자 변환이 행해질 수 있다(도 16b 및 도 16e 참조). 대안적으로, 표적 게놈 유전자좌에서의 LTVEC 표적화의 결여가 있을 수 있지만, Cas9 절단이 129 또는 B6 염색체 상에서 일어날 수 있다(도 16c 및 도 16f 참조). 이러한 후자의 가능성은 BP-E7, BP-H4, 및 BP-E6 클론에 대해 관찰된 결과를 설명할 수 있다. 잠재적인 결과가 도 16a 내지 도 16f에 나타나 있다. 도 16f의 경우, Cas9가 129 염색분체를 절단하는 경우, B6 대립유전자를 보유하는 이형접합성의 상실(LOH)을 관찰하는 것이 또한 가능하다. 전술된 실험에서는, 양쪽 대립유전자가 표적화되거나(Hum/Hum) 또는 양쪽 대립유전자가 야생형 대립유전자가 되는(+/+) 이형접합성의 상실 이벤트가 관찰되었다.

[0392] 실시예 3. 표적화 효율에 대한 LTVEC 상동성 아암 크기의 영향

[0393] 표적화 효율에 대한 상동성 아암 크기의 영향을 결정하기 위하여, 상보적 성분 5(*C5* 또는 *Hc*(용혈성 상보체))에 대한 마우스 유전자의 76 kb 결실 및 상동성 인간 *C5* 유전자의 97 kb 단편에 의한 동시 대체를 생성하도록 설계된 2개의 LTVEC를 비교하였다(도 13). 표적 유전자좌는 *C5(Hc)* 유전자의 엑손 2로부터 정지 코돈까지 포함하였다. 첫 번째 LTVEC는 결실이 의도된 마우스 *C5(Hc)* 유전자의 76 kb 서열을 플랭킹하는 마우스 *C5(Hc)* 유전자좌의 부분들로부터 유래된 35 kb 및 31 kb의 게놈 DNA를 함유하는 상동성 아암들에 의해 플랭킹된 인간 *C5* 유전자의 97 kb 단편을 포함하였다(도 13에서의 표적화 벡터 표지화된 LTVEC를 참조한다). 두 번째 LTVEC는 결실이 의도된 마우스 *C5(Hc)* 유전자의 76 kb 서열을 플랭킹하는 마우스 *C5(Hc)* 유전자좌의 부분들로부터 유래된 5 kb

각각의 게놈 DNA를 함유하는 상동성 아암들에 의해 플랭킹된 인간 *C5* 유전자의 97 kb 단편을 포함하였다(도 13에서의 표적화 벡터 표지화된 sTVEC를 참조한다).

[0394] 별도의 실험에서는, *C5(Hc)* 인간화 LTVEC들을, 결실에 대해 표적화된 마우스 *C5(Hc)* 유전자의 영역 내에 이중-가닥 브레이크를 생성하도록 설계된, Cas9를 인코딩하는 플라스미드 및 6개의 gRNA(A, B, C, D, E, 및 E2; 표 1 참조) 중 1개 또는 2개를 인코딩하는 제2 플라스미드와 조합하였다. gRNA들은 인간 *C5* 유전자의 삽입된 부분에서의 임의의 서열의 인식을 피하도록 설계하였다.

[0395] *C5(Hc)* 유전자의 CRISPR/Cas9-보조 인간화의 결과가 표 9에 나타나 있다. 제1 LTVEC 단독(35 kb 및 31 kb의 상동성 아암)으로의 표적화 효율은 제2 LTVEC 단독(5 kb 및 5 kb의 상동성 아암)으로의 표적화 효율보다 더 높았다. 그러나, gRNA A 및 E2와 조합될 때 각각의 LTVEC의 총 표적화 효율은 거의 동일하였는데(표 9 참조), 이는 5 kb의 상동성 아암의 크기(즉, 총합 10 kb)가 LTVEC 표적화와 조합하여 CRISPR/Cas9를 사용하여 *C5(Hc)* 유전자좌를 표적화할 때 관찰되는 표적화 효율의 증가를 촉진시키는 데 충분함을 나타낸다.

[0396] [표 9]

상이한 상동성 아암 크기를 갖는 LTVEC를 사용한 *C5(Hc)* 유전자의 CRISPR/Cas9-보조 인간화에 대한 스크리닝 결과.

		LOA 스크리닝에 의한 표적화 효율(%)			
gRNA	LTVEC	이형접합성	반접합성	동형접합성	총계
A + E2	1	12.8	6.4	3.8	23.0
없음	1	8.3	0	0	8.3
A + E2	2	7.2	13.1	4.3	24.6
없음	2	5.4	0	0	5.4

[0397]

[0398] 실시예 4. 표적화 효율에 대한 CRISPR RNA 인식 서열들 사이의 더 짧은 거리의 영향

[0399] 표적화 효율에 대한 CRISPR RNA 인식 서열들 및 절단 부위들 사이의 더 짧은 거리의 영향을 결정하기 위하여, 시티딘 모노포스페이트-N-아세틸뉴라민산 하이드록실라제(*Cmah*)에 대한 마우스 유전자의 18.2 kb 결실 및 *lacZ* 리포터 및 하이그로마이신 저항성 선택 카세트를 포함하는 삽입물에 의한 동시 대체를 생성하도록 LTVEC를 설계하였다. LTVEC는 간격이 좁은 서열을 표적화하는 2개의 gRNA를 사용하였다(도 14). 표적 유전자좌는 *Cmah* 유전자의 처음 5개의 엑손을 포함하였다. LTVEC는 결실이 의도된 마우스 *Cmah* 유전자의 18.2 kb 서열을 플랭킹하는 마우스 *Cmah* 유전자좌의 부분들로부터 유래된 120 kb 및 57 kb의 게놈 DNA를 함유하는 상동성 아암들에 의해 플랭킹된 8.8 kb *lacZ-hyg^r* 삽입물을 포함하였다. LTVEC를 결실에 대해 표적화된 마우스 *Cmah* 유전자의 영역의 5' 말단 부근에서 이중-가닥 브레이크를 생성하도록 설계된 Cas9 및 2개의 gRNA(A 및 B)를 인코딩하는 플라스미드들과 조합하였다. 2개의 gRNA는 결실이 의도된 서열의 5' 말단에서의 ATG 부근의 간격이 좁은 서열들을 표적화하였으며, 표적화된 절단 부위들은 27 bp만큼 떨어져 있었다(도 15 참조). 2개의 gRNA에 의해 가이드된 Cas9에 의한 절단은 27 bp의 평활-말단의 절제된 서열을 생성한다. LTVEC 단독을 대조군으로서 사용하였다.

[0400] 약물-저항성 ES 세포 클론을 결실 내의 서열에 대해 그리고 약물 선택 카세트 및 인간 유전자 삽입물 내의 서열에 대해, 대립유전자의 변형 검정에 의해 표적화된 인간화에 대해 스크리닝하였다(문헌[Valenzuela et al. (2003) *Nat. Biotechnol.* 21:652-659]; 문헌[Frendewey et al. (2010) *Methods Enzymol.* 476:295-307]). 클론이 2개의 내인성 마우스 유전자 서열들 중 하나를 상실하였고 *lacZ-hyg^r* 삽입물의 하나의 카피를 획득하였다면, 그러한 클론은 올바르게 표적화된 것으로 스코어링하였다. 게다가, 보유 검정으로서 지칭되는, LTVEC의 상동성 아암들 내의 서열을 인식하는 실시간 PCR 검정을 사용하여 마우스 게놈 내로의 LTVEC의 올바른 표적화를 검증하였다. 이들 보유 검정의 카피수를 결정함으로써, 올바르게 표적화된 ES 클론 - 이는 2개의 카피수를 보유함 - 을 표적 마우스 유전자좌의 큰 Cas9-유도 결실이 게놈 내의 어딘가 다른 곳에서의 LTVEC의 무작위 통합과 일치하는 클론 - 이 경우에 보유 검정은 3개(또는 그 이상)의 카피수를 가짐 - 과 구별하는 것을 돕도록 추가의 명확화를 제공하였다. 표적 마우스 유전자좌에서 큰 Cas9-매개 결실을 생성하는 쌍 형성된 gRNA의 능력은 앞서 기재된 바와 같은 표준 LOA 및 GOA 검정이 보유 검정에 의해 추가의 명확화를 제공하고 올바른 표적화를 검증하도록 강화될 수 있었음을 의미하였다. 따라서, 보유 검정을 설계하고 LOA 및 GOA 검정과 함께 사용하였다.

[0401] *Cmah* 표적화 실험의 결과가 표 10에 요약되어 있다. LTVEC 단독 사용에 의한 대조군 표적화 실험에서, 스크리닝된 클론들의 5.4%(3/56)는 이형접합성(Het) 결실-대체 돌연변이를 가졌고; 클론의 95%는 *Cmah* 유전자좌에서 야생형(WT)을 유지하였다. CRISPR 표적화 실험에서, 본 발명자들은 수개의 WT 클론에 더하여 5개의 상이한 돌연변이 대립유전자 유형을 관찰하였다. 본 발명자들은 3가지 유형의 LTVEC-표적화된 대립유전자를 관찰하였다: (1) Het; (2) Hom(동형접합성 결실-대체); 및 (3) Hemi(한쪽 대립유전자 상에서는 결실-대체, 그리고 다른 한쪽

대립유전자 상에서는 gRNA/Cas9-유도 돌연변이). 이들 3가지 유형은 스크리닝된 모든 클론들의 43.5%(106/244)를 구성한다. LTVEC 단독 사용과 비교하여, 본 발명자들은 적어도 한쪽의 대립유전자가 표적화된 *Cmah* 유전자 표적화의 8배 향상을 관찰하였다. 본 발명자들은 또한 단지 gRNA/Cas9-indel 돌연변이만을 갖는 2가지 유형의 대립유전자를 관찰하였다: (1) Het, 여기서는 2개의 WT 대립유전자 중 한쪽 상에서 indel을 검출하였음; 및 (2) 이중대립유전자 indel 돌연변이, 이는 동형접합성(Hom) 또는 반접합성(Hemi)일 수 있었음. 스크리닝된 클론들의 단지 3.7%만이 WT를 유지하였으며, 이때 *Cmah* 유전자좌에서는 검출 가능한 돌연변이가 없었다. 전체적으로, 클론의 94% 초과가 gRNA A와 gRNA B의 조합을 사용하였을 때 Cas9-유도 돌연변이를 가졌다.

[표 10]

Cmah 표적화에 대한 스크리닝 결과.

대조군 표적화 (LTVEC 단독)				CRISPR 표적화: LTVEC + Cas9 + gRNA A + B			
돌연변이	대립유전자 유형	유전자형	스크리닝된 클론의 %	돌연변이	대립유전자 유형	유전자형	스크리닝된 클론의 %
결실-대체	Het	<i>lacZ</i> +	5.4	결실-대체	Het	<i>lacZ</i> +	2.5
결실-대체	Hom	<i>lacZ</i> / <i>lacZ</i>	0	결실-대체	Hom	<i>lacZ</i> / <i>lacZ</i>	15
결실-대체/Indel	Hemi	<i>lacZ</i> /indel	0	결실-대체/Indel	Hemi	<i>lacZ</i> /indel	26
Indel	Het	indel/+	0	Indel	Het	indel/+	11
Indel	Hom 또는 Hemi	indel/indel	0	Indel	Hom 또는 Hemi	indel/indel	42
없음	WT	+/+	95	없음	WT	+/+	3.7

실시예 5. 1-세포기 배아에서 쌍 형성된 gRNA를 사용한 큰 붕괴

1-세포기 배아에서 큰 표적화된 결실을 달성하기 위하여, 엑토도메인을 인코딩하는 마우스 *Lrp5*(저밀도 지질단백질 수용체-관련 단백질 5) 유전자의 일부분의 68 kb 결실 및 선택적으로 2개의 60-뉴클레오타이드 상동성 아암에 의해 플랭킹된 4-뉴클레오타이드 삽입물을 갖는 단일-가닥 DNA 공여자 서열(124개 뉴클레오타이드 길이)의 사용을 통한 4-뉴클레오타이드 삽입에 의한 동시 대체를 생성하도록 실험을 설계하였다. 4-뉴클레오타이드 삽입물은 표적 유전자좌 내로 삽입 시에 제한 효소 부위를 생성하였다. 별도의 실험에서는, 단백질 형태의 Cas9 단백질을 세포질 주입(cytoplasmic injection)(CI)에 의해 전달하거나, mRNA 형태의 Cas9를 전핵 주입(pronuclear injection)(PNI) 또는 전기천공(EP)에 의해 전달하였다. Cas9를 결실에 대해 표적화된 마우스 *Lrp5* 유전자의 영역 내에 이중-가닥 브레이크를 생성하도록 설계된 2개의 gRNA(A + F)와, 그리고 선택적으로 상동 재조합 공여자와 조합하였다. gRNA들을 RNA 형태로 주입하였다. 이어서, 생성된 단일대립유전자 및 이중대립유전자 돌연변이의 빈도를 평가하였다.

결과는 표 11에 요약되어 있으며, 이는 2개의 가이드 RNA에 대한 표적 부위들 사이의 NHEJ-매개 결실 또는 ssDNA 공여자에 의한 상동성-유도 수복에 의해 보조된 결실을 포함한다. 쌍 형성된 가이드 RNA들 및 Cas9를 세포질 주입을 통해 ssDNA 공여자와 함께 도입하였을 때 이중대립유전자 돌연변이가 관찰되었다. 각각의 관찰된 이중대립유전자 돌연변이에서, 하나의 염색체는 NHEJ-매개 결실을 통해 변형되었고, 하나의 염색체는 HDR-보조 결실을 통해 변형되었다. 이들 결과는 세포질 mRNA 피에조(piezo) 주입이 공여자에 의한 동형접합성 재조합에 대한 잠재성을 가지면서 일관된 상동성 유도 수복을 가져옴을 나타낸다.

[표 11]

Cas9 전달 방법의 비교.

표적	HR 공여자	전달	Cas9	농도 (mg) Cas9/sgRNA/공여자	단일대립유전자 돌연변이		이중대립유전자 돌연변이	
					NHEJ 붕괴	HDR 붕괴	NHEJ 붕괴	HDR 붕괴
Lrp5 - 68 kb 붕괴 (A+F sgRNA)	-	PNI	단백질	50/60/0	7%	N/A	0%	N/A
		CI	mRNA	100/100/0	15%	N/A	N/A	N/A
		EP	mRNA	250/150/0	3%	N/A	0%	N/A
Lrp5 - 68 kb 붕괴 (A+F sgRNA)	+	PNI	N/A	N/A	N/A	N/A	N/A	N/A
		CI	mRNA	100/50/100	16.6%	19.4%	5.5%	5.5%
		EP	mRNA	250/200/250	1.4%	2.1%	0%	0%

실시예 6. 표적화된 삽입과 유전자도입적 삽입 사이 및 표적화된 결실과 표적화된 영역을 넘어서 연장된 결실 사이를 구별하기 위한 보유 검정

표준 대립유전자의 변형(MOA) 스크리닝 전략(예를 들어, 도 17a 참조)은 각각의 샘플에 대한 생물학적 반복 실험의 4개의 Ct 값들의 평균을 모든 샘플의 Ct 중위(median)와 비교함으로써 TaqMan® 카피수를 결정한다. 대립유전자의 상실의 경우, 결실에 대해 표적화되어 있는 표적 게놈 유전자좌의 영역의 상류(mTU) 및 하류(mTD) 영역에 대해 TaqMan® 프로브를 사용한다. 대립유전자의 획득의 경우, 네오마이신 저항성 카세트에 대해 TaqMan

⑧ 프로브를 사용한다. 그러나, 그러한 프로브들은 핵산 삽입물의 임의의 영역에 대해 설계될 수 있다. 이배체의 이형접합성 표적화된 클론의 경우, mTU, mTD, 및 Neo 프로브 각각에 대한 TaqMan® 카피수는 1이어야 한다. 이배체의 동형접합성 표적화된 클론의 경우, mTU 및 mTD 각각에 대한 TaqMan® 카피수는 0이어야 하고, Neo에 대한 TaqMan® 카피수는 2이어야 한다. 마찬가지로, 이배체의 비표적화된 클론의 경우, mTU 및 mTD 각각에 대한 TaqMan® 카피수는 2이어야 하고, Neo에 대한 TaqMan® 카피수는 0이어야 한다. 이배체의 이형접합성 붕괴된 클론의 경우, mTU 및 mTD에 대한 TaqMan® 카피수는 1이어야 하고, Neo에 대한 카피수는 0이어야 한다. 이배체의 동형접합성 붕괴된 클론의 경우, mTU, mTD, 및 Neo 각각에 대한 TaqMan® 카피수는 0이어야 한다.

[0411] 그러나, 쌍 형성된 gRNA는 표적 게놈 유전자좌에서의 큰 Cas-매개 결실을 생성할 수 있기 때문에, 그것은 LTVEC에 의한 올바른 표적화를 검증하기 위한 유용한 강화 표준 LOA 및 GOA 검정일 수 있다. 예를 들어, LOA 및 GOA 검정 단독으로는 올바르게 표적화된 세포 클론을 표적 게놈 유전자좌의 큰 Cas-유도 결실이 게놈 내의 어딘가 다른 곳에서의 LTVEC의 무작위 통합과 일치하는 클론과 구별할 수 없다. 표적화된 세포에서의 선택압은 선택 카세트에 기초하기 때문에, 게놈 내의 어딘가 다른 곳에서의 LTVEC의 무작위 유전자도입적 통합은 일반적으로 선택 카세트 및 LTVEC의 인접 영역들을 포함하겠지만 LTVEC의 더 원위인 영역은 제외시킬 것이다. 예를 들어, LOA 및 GOA 검정이 LTVEC의 표적화된 통합을 평가하는 데 사용되고, GOA 검정이 선택 카세트에 대한 프로브를 이용하는 경우, LTVEC의 무작위 유전자도입적 통합과 조합된 표적 게놈 유전자좌에서의 이형접합성 결실은 표적 게놈 유전자좌에서의 LTVEC의 이형접합성 표적화된 통합과 동일한 판독을 제공할 것이다. LTVEC에 의한 올바른 표적화를 검증하기 위하여, 보유 검정이 단독으로 또는 LOA 및/또는 GOA 검정과 조합하여 사용될 수 있다.

[0412] TaqMan® 보유 검정을 행할 때, 5' 상동성 아암에 대한 5' 표적 서열에 상응하는 상류 프로브(retU 프로브) 및 3' 상동성 아암에 대한 3' 표적 서열에 상응하는 하류 프로브(retD 프로브)가 사용된다(도 17b를 참조하는데, 이는 네오마이신 선택을 사용하여 CRISPR/Cas9-보조 인간화에 대해 스크리닝하기 위해 GOA 및 LOA 검정과 조합하여 TaqMan® 보유 검정을 사용하는 것을 나타냄). 도 17b는 또한 핵산 삽입물 내의 상이한 프로브들이 GOA 검정에 어떻게 사용될 수 있는지를 나타낸다(상류 hTU 프로브 및 하류 hTD 프로브 참조). 상이한 유형의 표적화된 변형 및 유전자도입적 삽입에 대한 GOA, LOA, 및 보유 검정 결과가 표 12에 나타나 있다.

[0413] [표 12]

상이한 변형 유형에 대한 GOA, LOA, 및 보유 검정의 예측된 카피수 판독치.

변형 유형	retU	mTGU	mTM	mTGD	retD	Neo
Homo 표적화됨	2	0	0	0	2	2
Het 표적화됨	2	1	1	1	2	1
Het 표적화됨 + Het 붕괴	2	0	0	0	2	1
Het 붕괴	2	1	1	1	2	0
Het 붕괴 및 유전자도입적 삽입	2	1	1	1	3*	1*
Homo 붕괴	2	0	0	0	2	0
Homo 붕괴 및 유전자도입적 삽입	2	0	0	0	3*	1*

[0414]

[0415] TaqMan 보유 검정은 또한 쌍 형성된 gRNA들을 사용하여 CRISPR/Cas9-보조 결실에 대해 스크리닝하기 위해 LOA 검정과 조합하여 사용될 수 있다(도 17c 참조). 그러한 검정에서, retU 및 retD에 대한 카피수는 모든 경우에 2개를 유지해야 한다. 2개 미만의 카피수는 결실에 대해 표적화되는 영역을 넘어서 연장되는 큰 Cas9-매개 결실을 나타낸다. 상이한 유형의 붕괴-관련 변형에 대한 LOA 및 보유 검정 결과가 표 13에 나타나 있다.

[0416] [표 13]

상이한 붕괴 변형에 대한 LOA 및 보유 검정의 예측된 카피수 판독치.

변형 유형	retU	mTGU	mTM	mTGD	retD
Homo 붕괴	2	0	0	0	2
Het 붕괴	2	1	1	1	2
Het 붕괴 및 추가의 하류 결실	2	1	1	1	1*

[0417]

[0418] 실시예 7. 4개의 가이드 RNA를 사용한 CRISPR/Cas9-매개 표적화.

[0419] 변형된 마우스 면역글로불린 중쇄 유전자좌의 대략 900 kb 영역의 정확한 단일-단계 결실 및 loxP 부위들에 의해 플랜킹된 Pgk-Neo 삽입물(네오마이신 포스포트랜스퍼라제 유전자에 작동가능하게 연결된 포스포글리세레이트 키나제 I 프로모터)에 의한 동시 대체를 달성하기 위하여, 본 발명자들은 전기천공에 의해 마우스 ES 세포 내로 하기의 핵산 분자를 도입하였다: (1) LTVEC; (2) Cas9 엔도뉴클레아제를 인코딩하는 플라스미드; 및 (3) 4개의 CRISPR 단일 가이드 RNA(gRNA)를 인코딩하는 하나 이상의 플라스미드. 각각의 실험에서는, LTVEC를 선형화하였

다. 변형에 대해 표적화된 유전자좌는, 가변 영역 유전자 세그먼트(V_H , D_H , J_H)가 인간 대응부로 대체된 마우스 면역글로불린 중쇄 유전자좌의 대략 900 kb 영역이었다(도 18 참조). LTVEC는, 표적 유전자좌의 대략 900 kb 영역을 결실시키고, 네오마이신 포스포트랜스퍼라제의 발현을 유도하여 G418에 대한 저항성을 부여하는 약물 선택 카세트를 삽입하는 상동 재조합 이벤트를 유도하도록 설계된 19 kb 5' 상동성 아암 및 13 kb 3' 상동성 아암에 의해 플랭킹된 대략 2 kb의 길이를 갖는 Pgk-Neo 삽입물을 포함하였다.

[0420] 사용된 4개의 gRNA 중에서, 2개는 표적 유전자좌의 5' 말단 부근에서 Cas9 절단을 유도하고(도 18에서의 5' gRNA_I 및 5' gRNA_II), 2개는 표적 유전자좌의 3' 말단 부근에서 Cas9 절단을 유도하였다(도 18에서의 3' gRNA_I 및 3' gRNA_II). 5' gRNA_I 표적 서열과 5' gRNA_II 표적 서열은 서로 대략 150 bp만큼 떨어져 있고, 3' gRNA_I 표적 서열과 3' gRNA_II 표적 서열은 중첩되었으며, 이때 3' gRNA_II 표적 부위는 3' gRNA_I 표적 부위에 대해 1 bp만큼 이동되었다.

[0421] LTVEC를 받아들여 그것을 게놈 내로 혼입시킨 ES 세포는 조직 배양 디시 상에서 항생제 약물이 담긴 성장 배지 중에서 성장하여 콜로니를 형성할 수 있었다. 본 발명자들은 약물 저항성 콜로니들을 고르고, 대립유전자의 변형 방법에 의해 그들을 스크리닝하여(문헌[Valenzuela *et al.* (2003) *Nat. Biotech.* 21:652-660]; 문헌[Frendewey *et al.* (2010) *Methods Enzymol.* 476:295-307]; 이들은 전체적으로 본 명세서에 참고로 포함됨), 올바르게 표적화된 인간화된 대립유전자를 가진 클론을 확인하였다(하기 표 14 참조). 게다가, 보유 검정으로 서 지칭되는, LTVEC의 상동성 아암들 내의 서열을 인식하는 실시간 PCR 검정을 사용하여 마우스 게놈 내로의 LTVEC의 올바른 표적화를 검증하였다(하기 표 14 참조).

[0422] [표 14]

LTVEC 및 4개의 gRNA에 의한 표적화를 확인하기 위해 사용된 프로브.

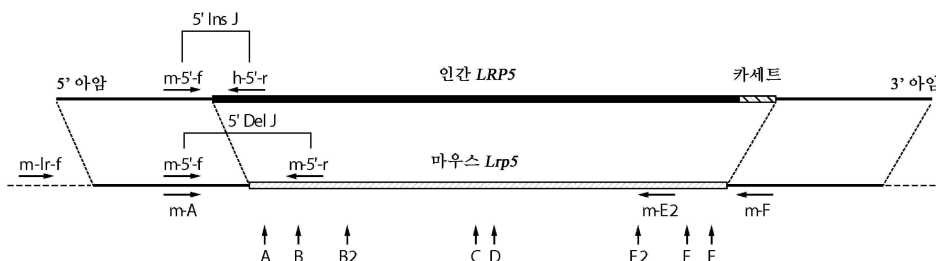
프로브	검정	서열 번호	하기까지의 거리	Kb
5' IgH 아암 1	보유	148	5' gRNA	1.5
5' IgH 아암 2	보유	173	5' gRNA	0.315
hIgH31	LOA	136	5' gRNA	154
hIgH1	LOA	142	5' gRNA	747
hIgH1	LOA	142	3' gRNA	116
hIgH9	LOA	139	3' gRNA	3.2
3' IgH 아암 1	보유	176	3' gRNA	0.484
3' IgH 아암 2	보유	157	3' gRNA	1.2
mIgM-398	보유	151	3' gRNA	6.4
mIgM-1045	보유	154	3' gRNA	7

[0423]

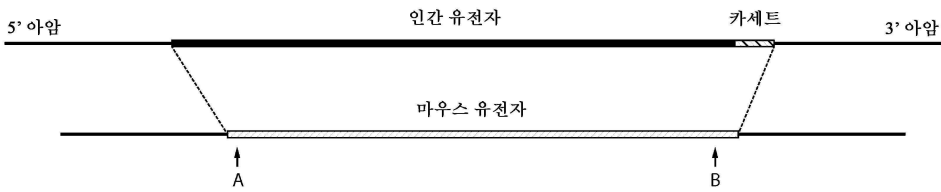
[0424] 생성된 표적화된 ES 세포에서는, 양쪽 대립유전자에서 대략 900 kb 영역이 결실되고 Pgk-Neo 삽입물로 대체되었다(도 18 참조). 이러한 큰 결실 및 대체는 뜻밖에도 높은 효율로 달성되었다(이중대립유전자 결실의 경우 대략 1.2% 효율).

도면

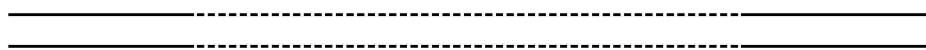
도면1



도면2a



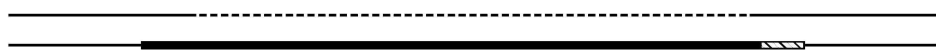
도면2b



도면2c



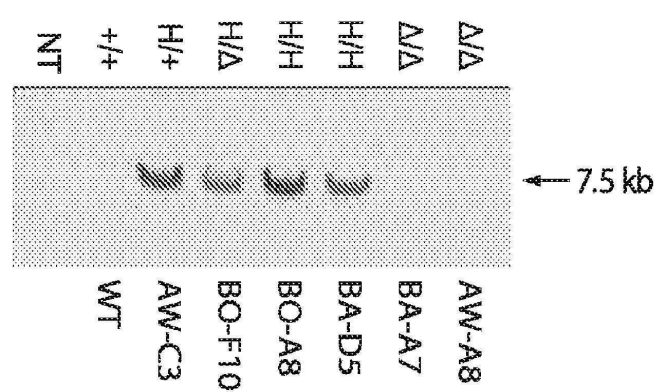
도면2d



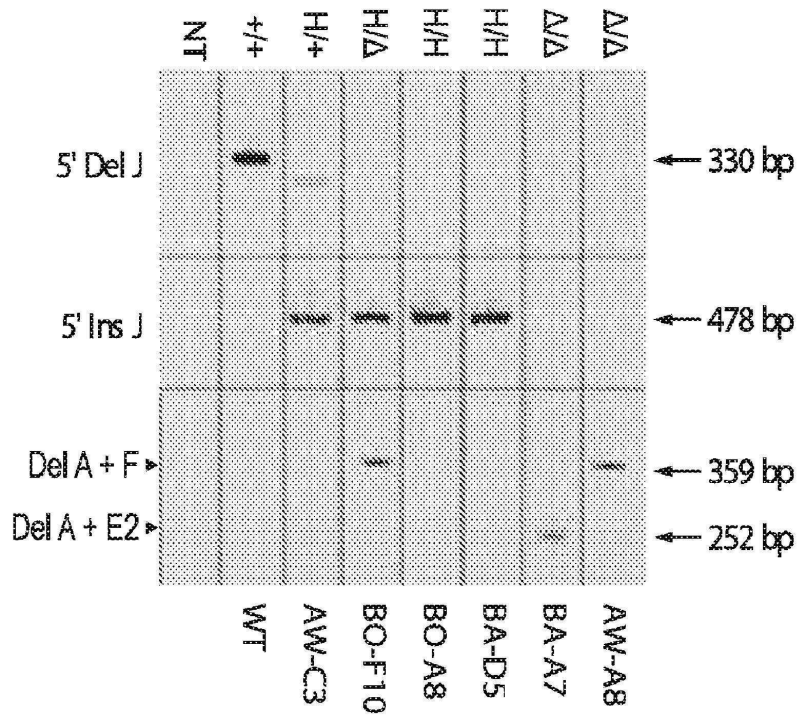
도면2e



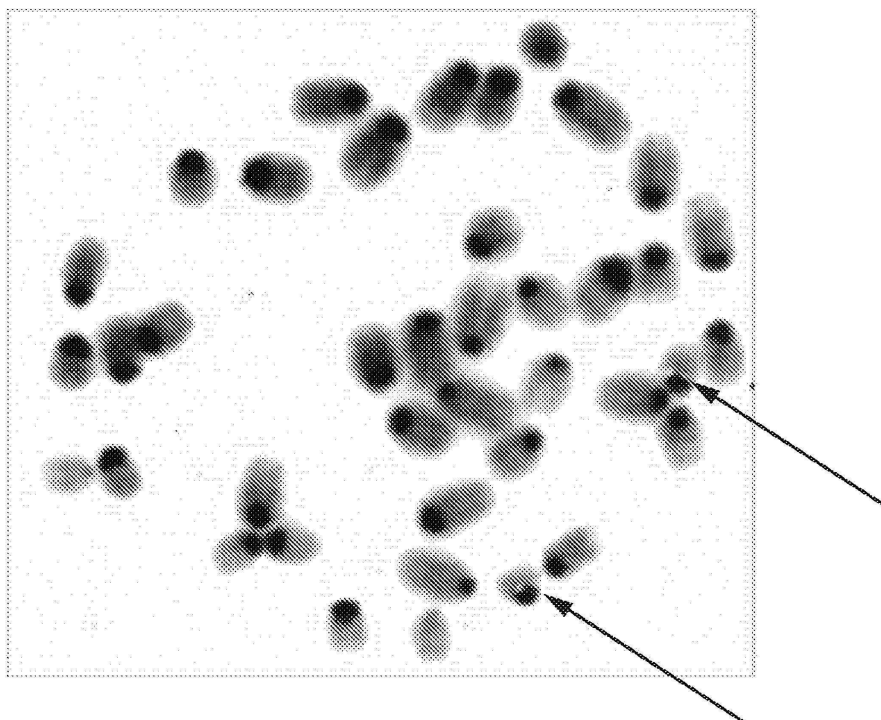
도면3a



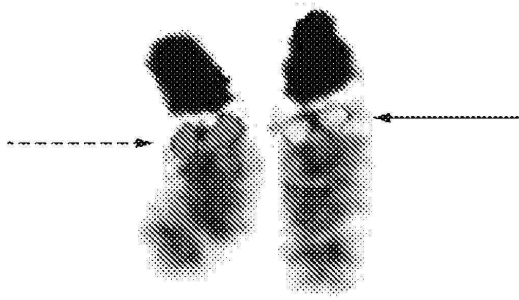
도면3b



도면4a



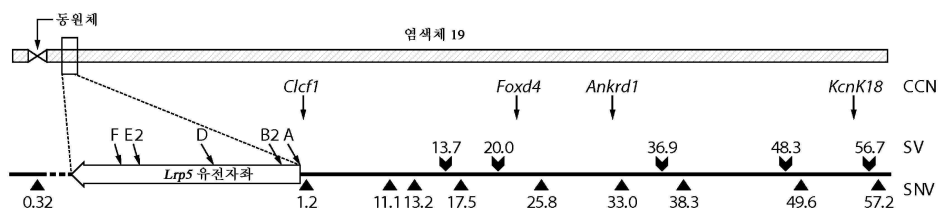
도면4b



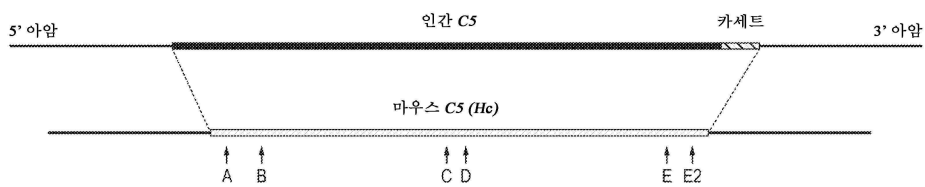
도면4c



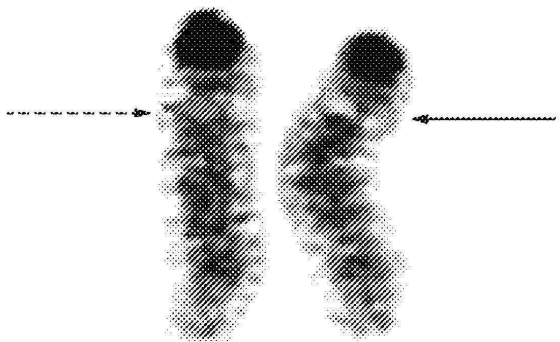
도면5



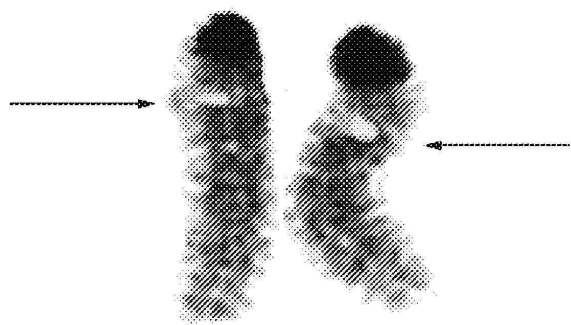
도면6



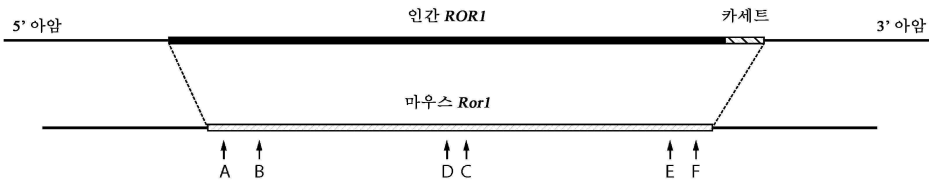
도면7a



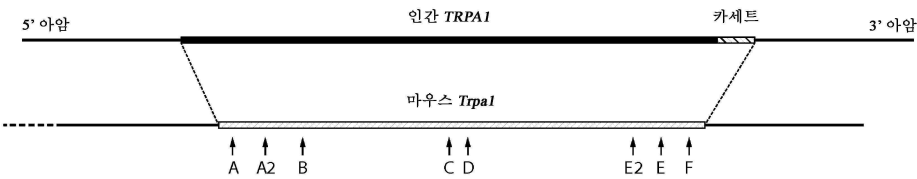
도면7b



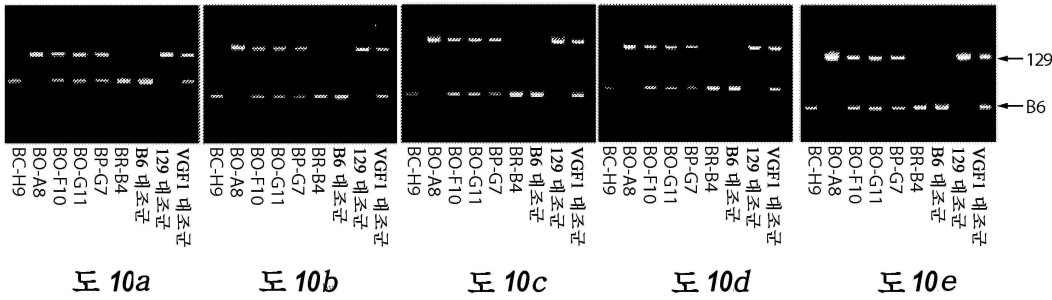
도면8



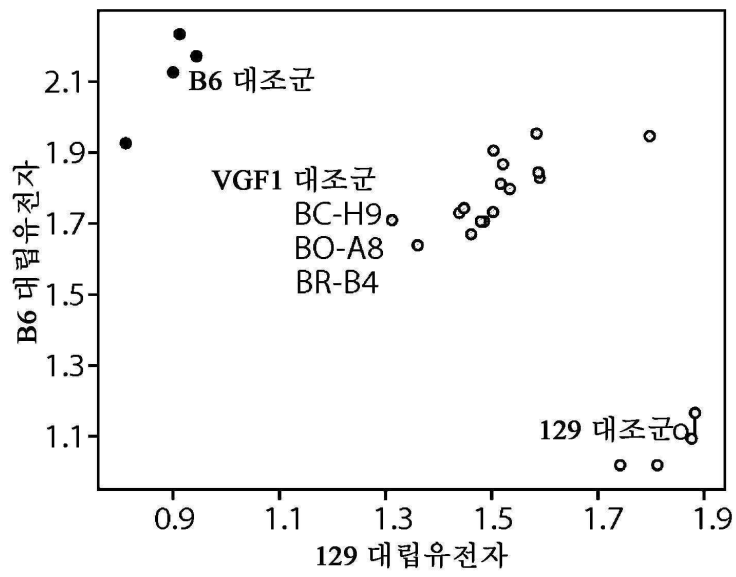
도면9



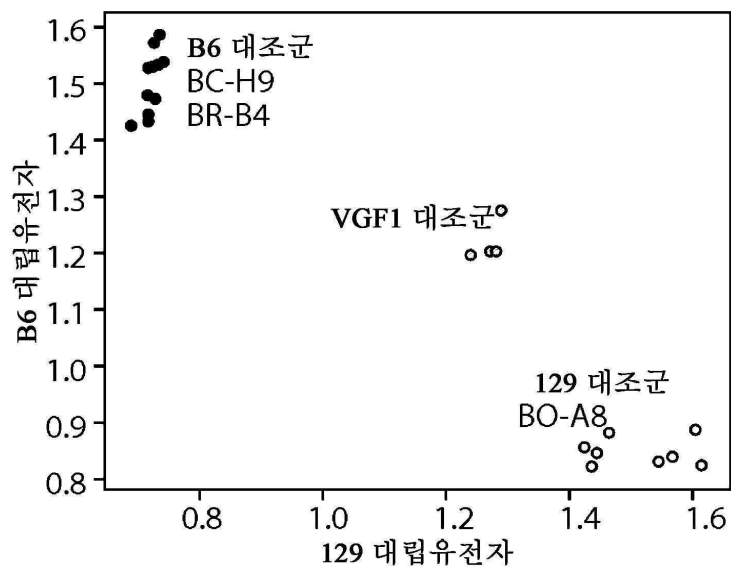
도면10



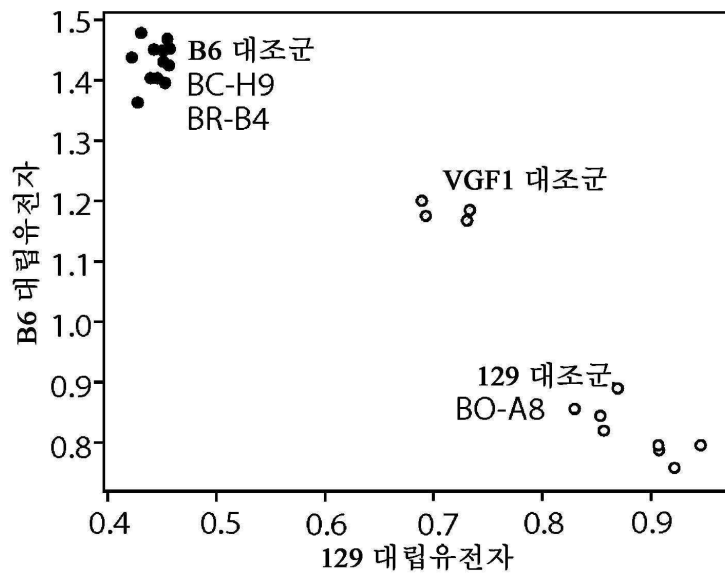
도면11a



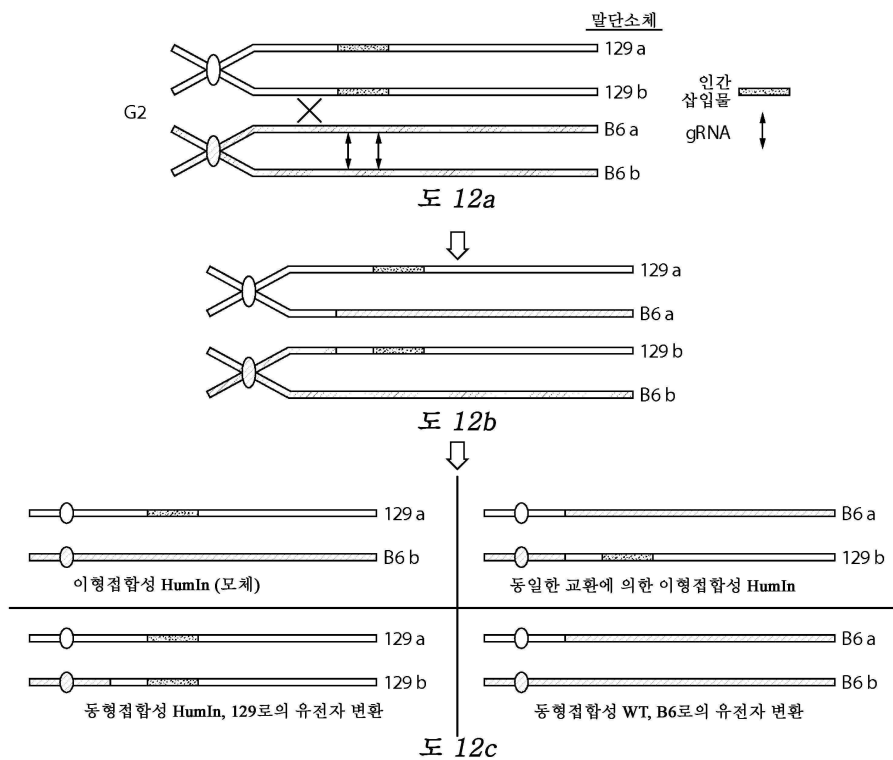
도면11b



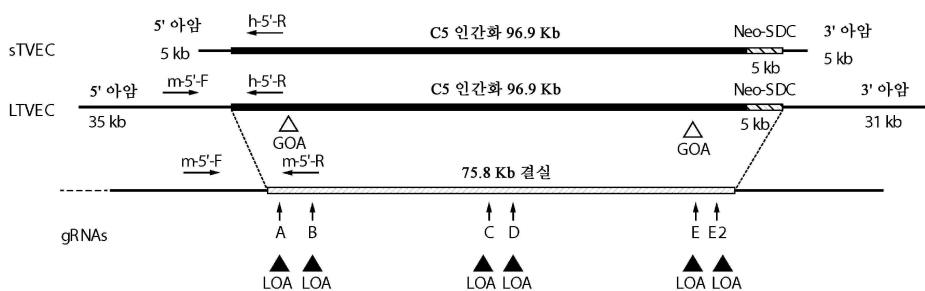
도면11c



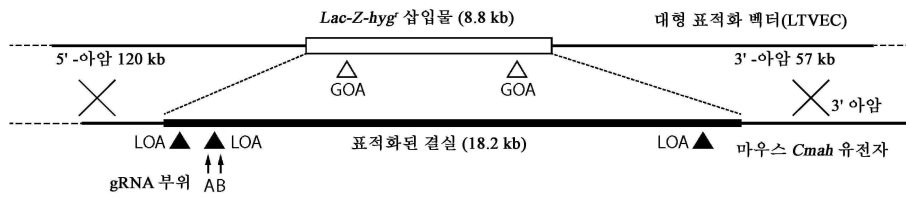
도면12



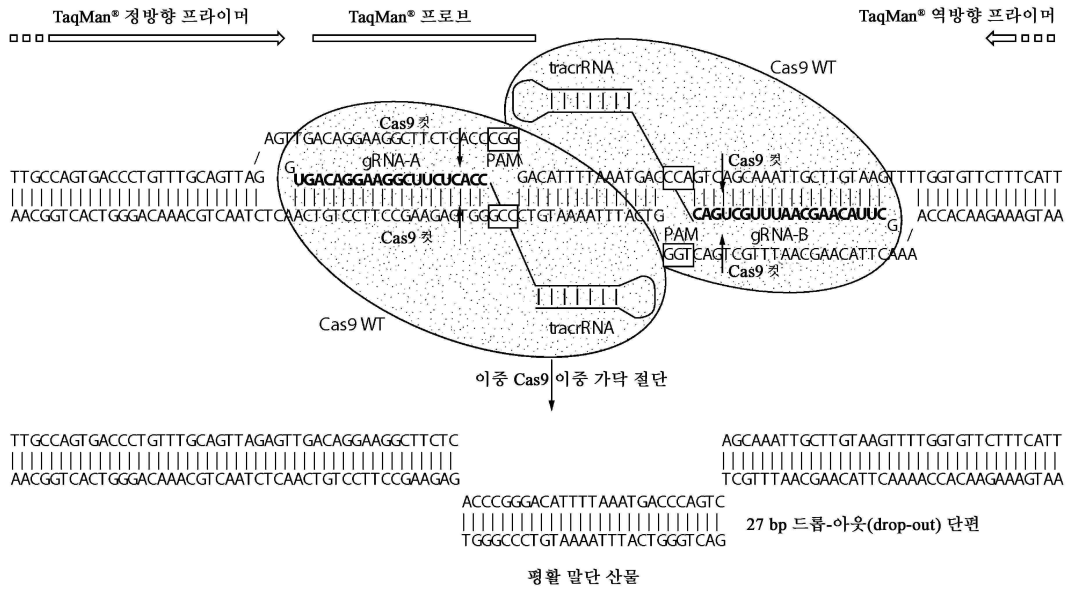
도면13



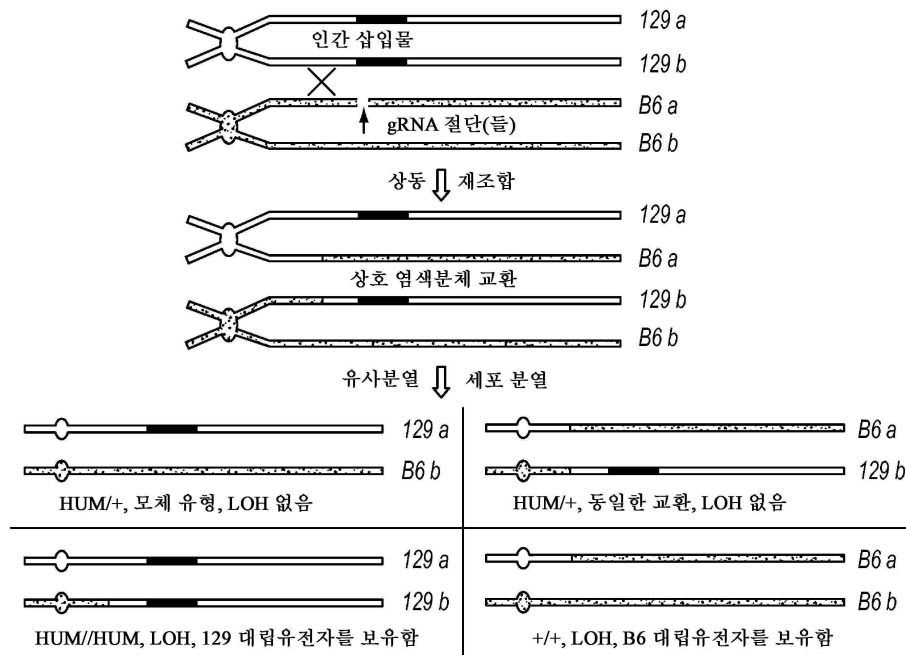
도면14



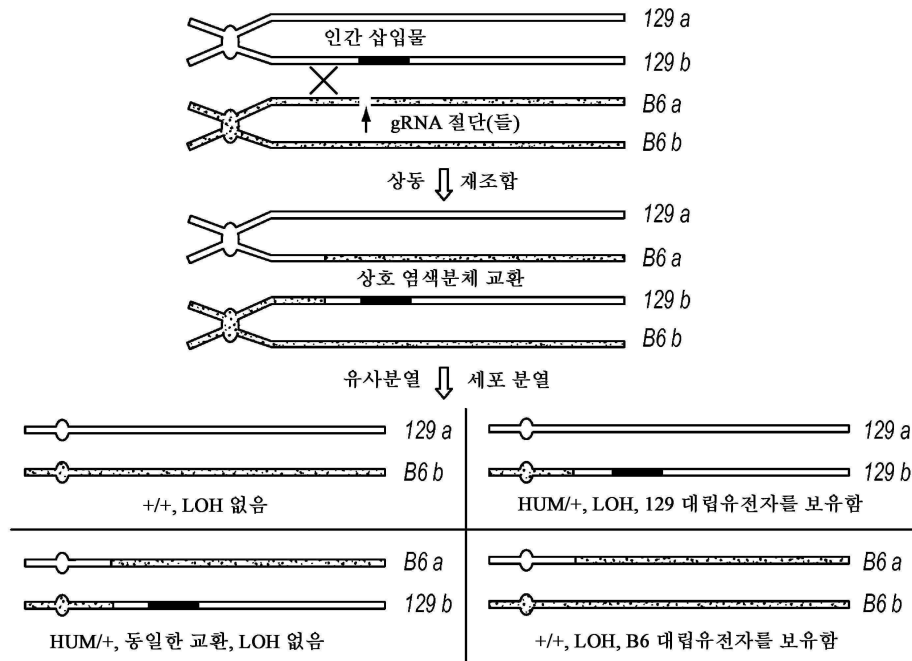
도면15



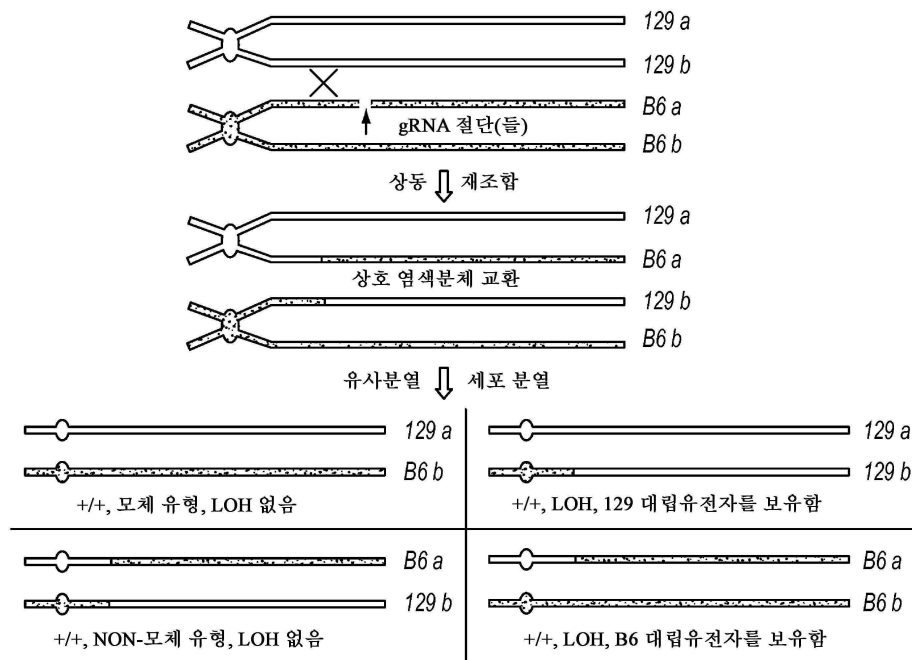
도면16a



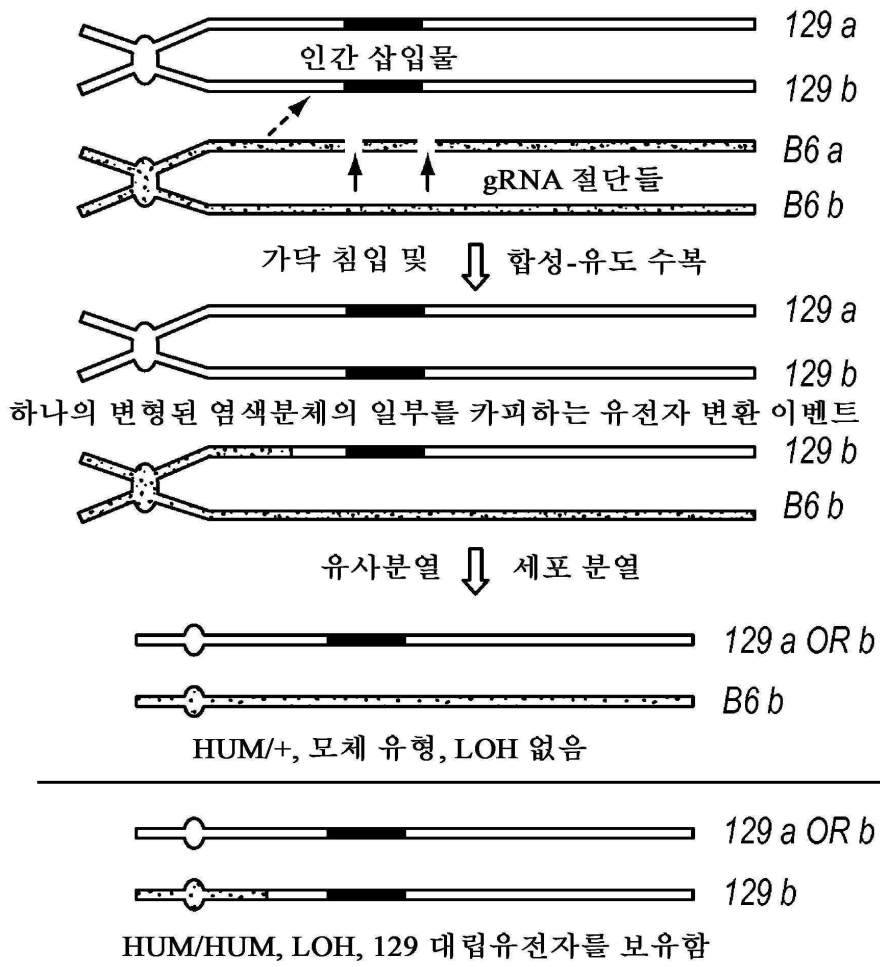
도면16b



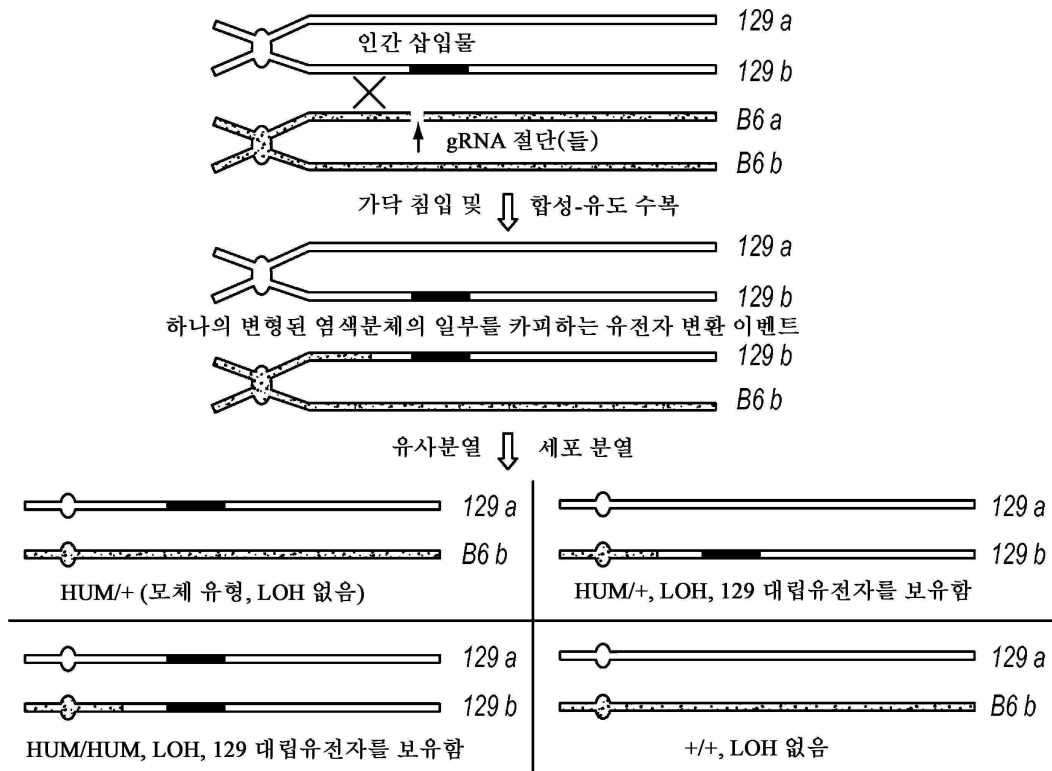
도면16c



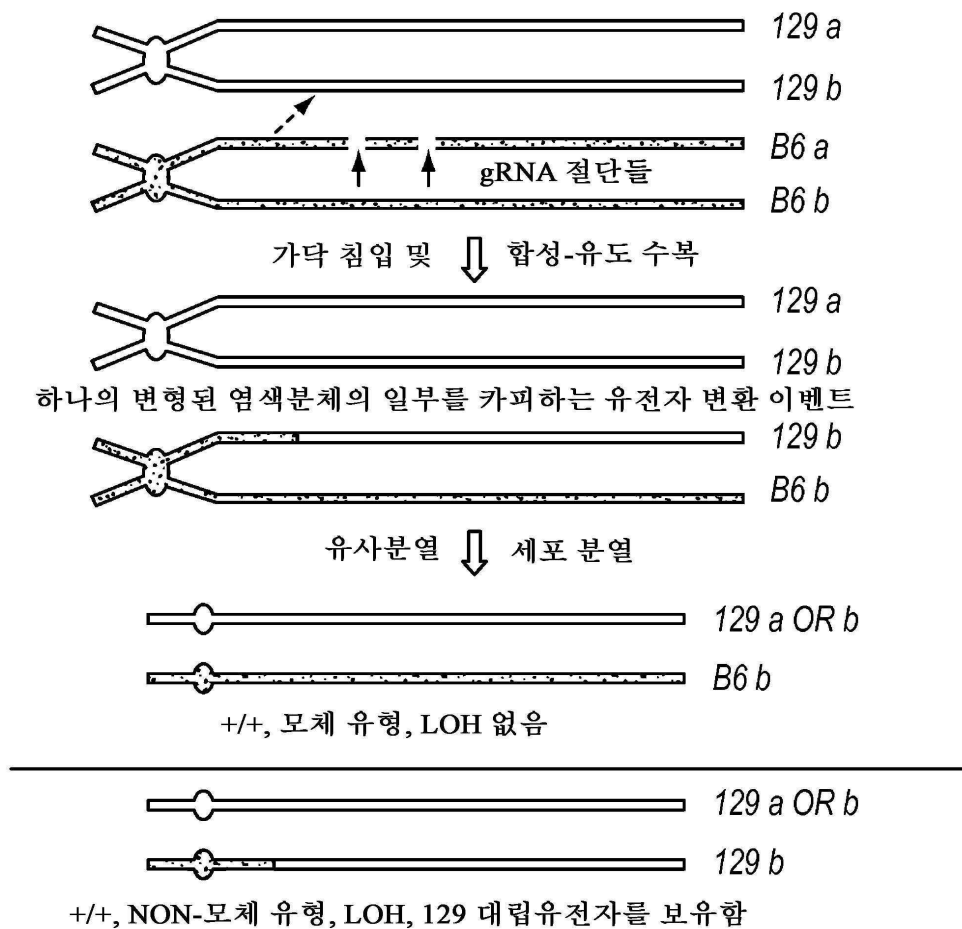
도면16d



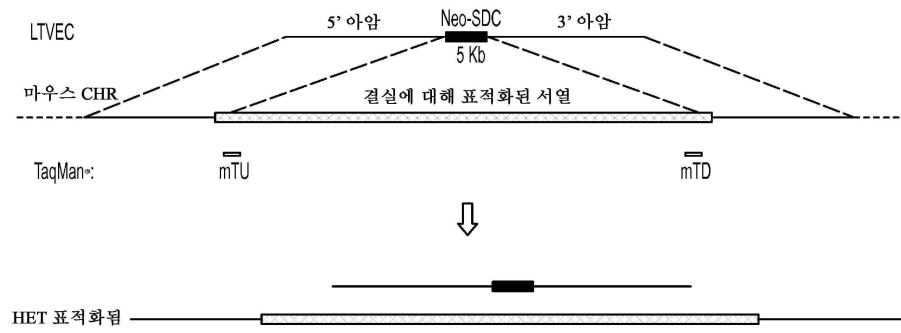
도면16e



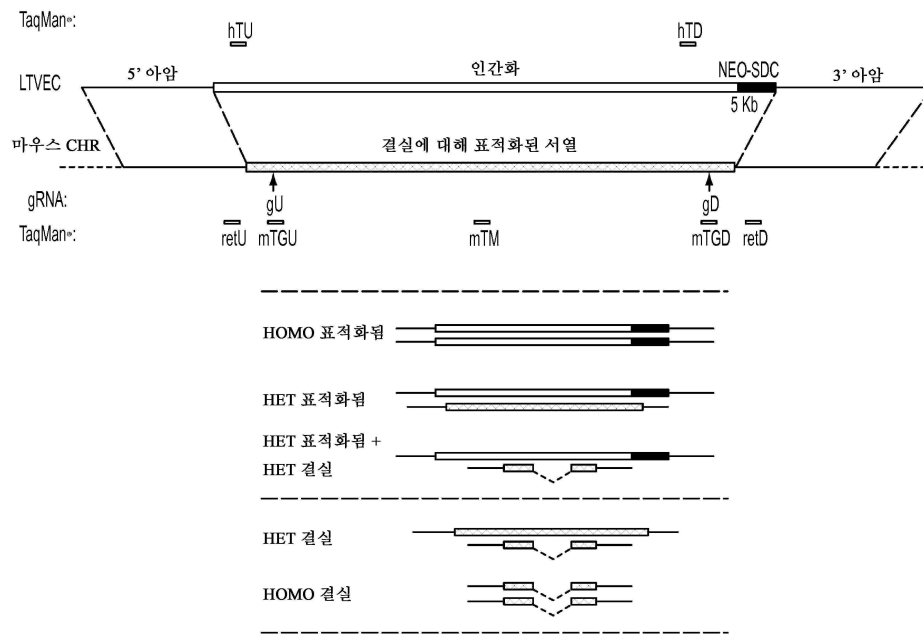
도면16f



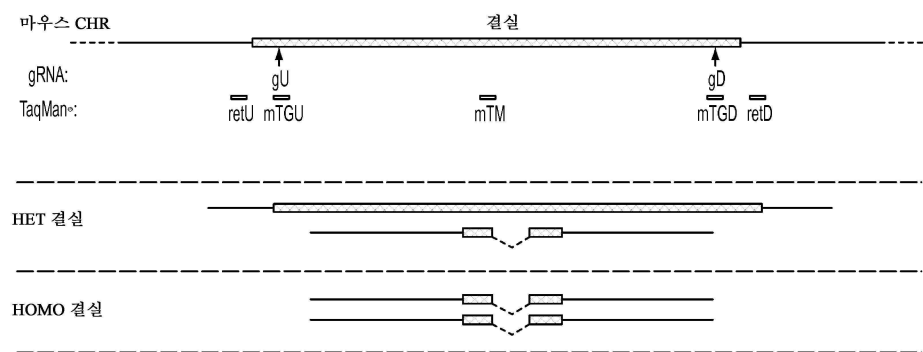
도면17a



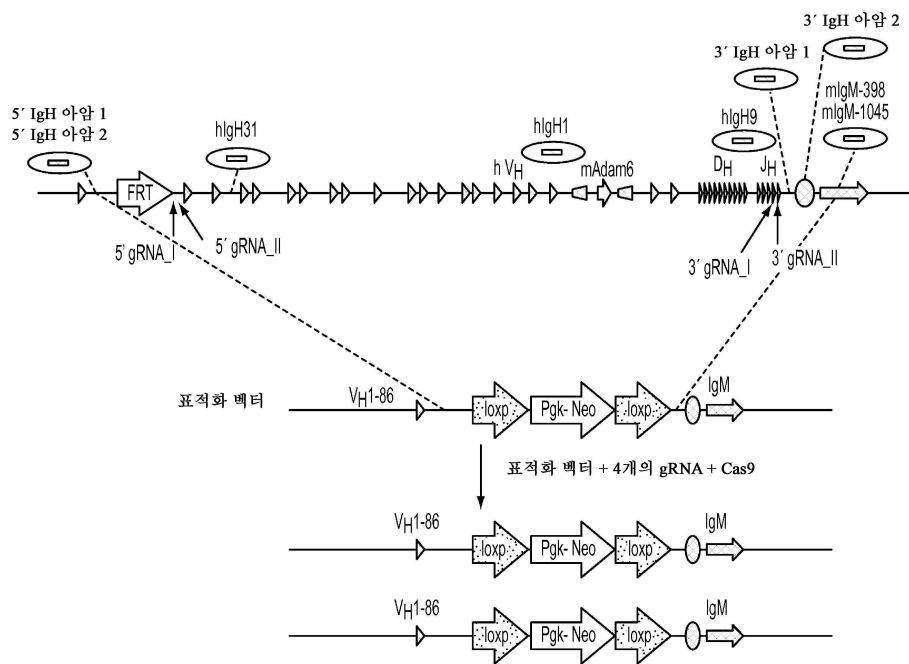
도면17b



도면17c



도면18



서열 목록

SEQUENCE LISTING

<110> Frendewey, David

Lai, Ka-Man Venus

Auerbach, Wojtek

Lee, Jeffrey D.

Mujica, Alexander O.

Droguett, Gustavo

Trzaska, Sean

Hunt, Charleen

Gagliardi, Anthony

Valenzuela, David M.

Voronina, Vera

Macdonald, Lynn

Murphy, Andrew J.

Yancopoulos, George D.

<120> METHODS AND COMPOSITIONS FOR TARGETED

GENETIC MODIFICATION USING PAIRED GUIDE RNAS

<130> 57766-472225

<150>
 > US 62/083,005
 <151> 2014-11-21
 <150> US 62/182,314
 <151> 2015-06-19
 <150> US 62/211,421
 <151> 2015-08-28
 <160> 176
 <170> FastSEQ for Windows Version 4.0
 <210> 1
 <211> 80
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> gRNA
 <400> 1
 guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cguuaucaac uugaaaaagu 60
 ggcaccgagu cggugcuuuu 80
 <210> 2
 <211> 42
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> gRNA
 <400> 2
 guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc cg 42
 <210> 3
 <211> 30
 <212> RNA
 <213> Artificial Sequence
 <220>
 <223> crRNA
 <400> 3
 guuuuagagc uagaaauagc aaguuaaaau 30

<210> 4	
<211> 33	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> crRNA	
<400> 4	
guuuuagagc uagaaauagc aaguuaaaau aag	33
<210> 5	
<211> 26	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> crRNA	
<400> 5	
gaguccgagc agaagaagaa guuuua	26
<210> 6	
<211> 12	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> tracrRNA	
<400> 6	
aaggcuaguc cg	12
<210> 7	
<211> 50	
<212> RNA	
<213> Artificial Sequence	
<220>	
<223> tracrRNA	
<400> 7	
aaggcuaguc cguaaucaac uugaaaaagu ggcaccgagu cggugcuuuu	50
<210> 8	
<211> 23	

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> CRISPR RNA recognition sequence
 <220>
 <221> misc_feature
 <222> (2)...(21)
 <223> n = a, t, c, or g

<400> 8
 gnnnnnnnnnn nnnnnnnnnnn ngg 23

<210> 9
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> CRISPR RNA recognition sequence
 <220>
 <221> misc_feature
 <222> (1)...(21)
 <223> n = a, t, c, or g
 <400> 9

nnnnnnnnnnn nnnnnnnnnnn ngg 23

<210> 10
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> CRISPR RNA recognition sequence
 <220>
 <221> misc_feature
 <222> (3)...(23)

<223> n = a, t, c, or g
 <400> 10

ggnnnnnnnnn nnnnnnnnnn nnngg	25
<210> 11	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> C5 (Hc) gRNA A DNA-targeting segment	
<400> 11	
atcacaaacc agttaaccgg	20
<210> 12	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> C5 (Hc) gRNA B DNA-targeting segment	
<400> 12	
tttcagacga gccgaccgg	20
<210> 13	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> C5 (Hc) gRNA C DNA-targeting segment	
<400> 13	
tgtgtgtcat agcgatgtcg	20
<210> 14	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> C5 (Hc) gRNA D DNA-targeting segment	
<400> 14	
aacaggtacc ctatcctcac	20
<210> 15	

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> C5 (Hc) gRNA E DNA-targeting segment
 <400> 15
 ggcccgacc tagtctctct 20

<210> 16
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> C5 (Hc) gRNA E2 DNA-targeting segment
 <400> 16
 tcgtggttgc atgcgcactg 20

<210> 17
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Lrp5 gRNA A DNA-targeting segment
 <400> 17
 gggaaccac agcatactcc 20

<210> 18
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Lrp5 gRNA B DNA-targeting segment
 <400> 18
 gaatcatgca cggtacccc 20

<210> 19
 <211> 20

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> Lrp5 gRNA B2 DNA-targeting segment
 <400> 19
 tgctcctatg gggaggcgcg 20
 <210> 20
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Lrp5 gRNA C DNA-targeting segment
 <400> 20
 actgagatca atgaccccgga 20
 <210> 21
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Lrp5 gRNA D DNA-targeting segment
 <400> 21
 gggtcgcccc gaacctctac 20
 <210> 22
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Lrp5 gRNA E2 DNA-targeting segment
 <400> 22
 cttggataac attgataccc 20
 <210> 23
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Lrp5 gRNA E DNA-targeting segment
 <400> 23
 ggggcagagc ccttatatca 20

 <210> 24
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Lrp5 gRNA F DNA-targeting segment
 <400> 24
 tcgctcacat taatccctag 20

 <210> 25
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Ror1 gRNA A DNA-targeting segment
 <400> 25
 tgtgggcctt tgctgatcac 20

 <210> 26
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Ror1 gRNA B DNA-targeting segment
 <400> 26

 aatctatgat cctatgcct 20

 <210> 27
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> Ror1 gRNA D DNA-targeting segment

<400> 27

tgccaatagc agtgacttga

20

<210> 28

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Ror1 gRNA C DNA-targeting segment

<400> 28

gggaagaatg ggctattgtc

20

<210> 29

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Ror1 gRNA E DNA-targeting segment

<400> 29

ggttgtttgt gctgatgacg

20

<210> 30

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Ror1 gRNA F DNA-targeting segment

<400> 30

ccgtcctagg ccttctacgt

20

<210> 31

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> Trpa1 gRNA A DNA-targeting segment

<400> 31

gtactgggga atcggtggtc	20
<210> 32	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Trpa1 gRNA A2 DNA-targeting segment	
<400> 32	
cacgcactcc aaatttatcc	20
<210> 33	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Trpa1 gRNA B DNA-targeting segment	
<400> 33	
ctaagtgtgt atcagtacat	20
<210> 34	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Trpa1 gRNA C DNA-targeting segment	
<400> 34	
tgccctgcac aataagcgca	20
<210> 35	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Trpa1 gRNA D DNA-targeting segment	
<400> 35	
actcattgaa acgttatggc	20

<210> 36
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Trp1 grNA E2 DNA-targeting segment
 <400> 36
 agtaagggtg gattaaattc 20
 <210> 37
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Trp1 grNA E DNA-targeting segment
 <400> 37
 gccatctaga ttcattgtaac 20
 <210> 38
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Trp1 grNA F DNA-targeting segment
 <400> 38
 gactagaaat gttctgcacc 20
 <210> 39
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 190045 forward primer
 <400> 39
 gagctcatag ccaacagctt g 21
 <210> 40

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 190061 forward primer
 <400> 40
 atgcatcaga tcacgctcag 20
 <210> 41
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 190068 forward primer
 <400> 41
 gtccttgtgg catttccaac 20
 <210> 42
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 190030 forward primer
 <400> 42
 ccagtatggt gtcagttaat agcg 24

 <210> 43
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 190033 forward primer
 <400> 43
 ctgtgcagaa agcagcctc 19
 <210> 44
 <211> 20
 <212> DNA

<213> Artificial Sequence

<220>

<223> 190013 forward primer

<400> 44

cctctccctc taggcacctg 20

<210> 45

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> 190045 reverse primer

<400> 45

tctttaaggg ctccgttgtc 20

<210> 46

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> 190061 reverse primer

<400> 46

aagaccaacc attcaccag 20

<210> 47

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> 190068 reverse primer

<400> 47

ttcccagtc aagtcaaagg 20

<210> 48

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> 190030 reverse primer
 <400> 48
 ctgttatctg caaggcaccc 20

 <210> 49
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 190033 reverse primer
 <400> 49
 acaactggat cctgattcgc 20

 <210> 50
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 190013 reverse primer
 <400> 50
 taagagggca tgggtgagac 20

 <210> 51
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> C2 probe (B6)
 <400> 51
 aattcagaag acctatcgta 20

 <210> 52
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T3 probe (B6)

<400> 52	
tatgtgtata ggtgtttgga t	21
<210> 53	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> T6 probe (B6)	
<400> 53	
tacattgcta aatgaaacc	19
<210> 54	
<211> 16	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> T7 probe (B6)	
<400> 54	
cgcagtcattg cacata	16
<210> 55	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> T8 probe (B6)	
<400> 55	
ttataaagcc cagtatgtac	20
<210> 56	
<211> 14	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> T9 probe (B6)	
<400> 56	
tgctgcataa tcag	14

<210> 57
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T10 probe (B6)
 <400> 57
 tcaggagtga attggata 18

<210> 58
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T11 probe (B6)
 <400> 58
 ctgctactta cctttg 16

<210> 59
 <211> 13
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T13 probe (B6)
 <400> 59
 aggaggaaaa cgc 13

<210> 60
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T14 probe (B6)
 <400> 60
 cctttgttcc tcataag 17

<210> 61

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> C2 probe (129)
 <400> 61
 aattcagaag acctattgta 20
 <210> 62
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T3 probe (129)
 <400> 62
 tatgtgtata ggtgtttgca t 21
 <210> 63
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T6 probe (129)
 <400> 63
 cattgtctaca tgaaac 16
 <210> 64
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T7 probe (129)
 <400> 64
 cgcagtcattg cacgta 16
 <210> 65
 <211> 23
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> T8 probe (129)
 <400> 65
 tgagaattta taaagcccaa tat 23
 <210> 66
 <211> 14
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T9 probe (129)
 <400> 66
 tgctgcatga tcag 14
 <210> 67
 <211> 15
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T10 probe (129)
 <400> 67
 tcaggagtga atcgg 15
 <210> 68
 <211> 16
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T11 probe (129)
 <400> 68
 ctgctagtta cctttg 16
 <210> 69
 <211> 15
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> T13 probe (129)
 <400> 69
 aggaggaaga cgcag 15

<210> 70
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T14 probe (129)
 <400> 70
 ctttgttctt cataagc 17

<210> 71
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> C2 forward primer
 <400> 71
 atgagggatt tccttaatca gacaa 25

<210> 72
 <211> 29
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T3 forward primer
 <400> 72
 tggatatgttt attcttactc aaggttttg 29

<210> 73
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T6 forward primer

<400> 73	
gggcaactga tggaaagaac tc	22
<210> 74	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> T7 forward primer	
<400> 74	
gactgacgca caaacttgct ctt	23
<210> 75	
<211> 26	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> T8 forward primer	
<400> 75	
cccaaagcat ataacaagaa caaatg	26
<210> 76	
<211> 18	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> T9 forward primer	
<400> 76	
gcaggacgca ggcgttta	18
<210> 77	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> T10 forward primer	
<400> 77	
gcctctcat ggcagtctac atc	23

<210> 78
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T11 forward primer
 <400> 78
 cctgccctt gatgagtgtt 20

<210> 79
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T13 forward primer
 <400> 79
 ccctctttga tatgctcgtg tgt 23

<210> 80
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T14 forward primer
 <400> 80
 tccacaggt ccatgtcttt aa 22

<210> 81
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> C2 reverse primer
 <400> 81
 agactacaat gagctacat cataaggt 28

<210> 82

<211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T3 reverse primer
 <400> 82
 caaccatcta aaactccagt tcca 24
 <210> 83
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T6 reverse primer
 <400> 83
 tgtgtaacag gacagttgaa tgtagaga 28
 <210> 84
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T7 reverse primer
 <400> 84
 cttaaaaccc gccctgcat 19
 <210> 85
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T8 reverse primer
 <400> 85
 ctacaggaga tgtggctgtt ctatgt 26
 <210> 86
 <211> 22
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> T9 reverse primer
 <400> 86
 tcagcgtgat tcgcttgtag tc 22
 <210> 87
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T10 reverse primer
 <400> 87
 tgcataagctg ttgtaataat gacaag 26
 <210> 88
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T11 reverse primer
 <400> 88
 tgcagcatct ctgtcaagca a 21
 <210> 89
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> T13 reverse primer
 <400> 89
 gcaacaacat aaccacagc ataa 24
 <210> 90
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> T14 reverse primer
 <400> 90
 gctaagcgtt tggaagaaat tcc 23

<210> 91
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> SV 13.7 forward primer
 <400> 91
 taggctctaa ggatgctggc 20

<210> 92
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> SV 13.7 reverse primer
 <400> 92
 aagcagcttc aaaccctctg 20

<210> 93
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> SV 20.0 forward primer
 <400> 93
 ttacttg gcc ttggaactgc 20

<210> 94
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> SV 20.0 reverse primer

<400> 94	
tgattcgtaa tcgtcactgc c	21
<210> 95	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> SV 36.9 forward primer	
<400> 95	
tcctgtcccg agaaactgtc	20
<210> 96	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> SV 36.9 reverse primer	
<400> 96	
agctggcttt cagagagctg	20
<210> 97	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> SV 56.7 forward primer	
<400> 97	
ttagaaagtg ccaaccaggc	20
<210> 98	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> SV 56.7 reverse primer	
<400> 98	
ctctggctag gaacaatggc	20

<210> 99
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> m-lr-f primer
 <400> 99
 gttaggtgca gggctctactc agctg 25

<210> 100
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> m-5'-f primer
 <400> 100
 ggaggagagg agaagcagcc 20

<210> 101
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> m-A primer
 <400> 101
 ggaggagagg agaagcagcc 20

<210> 102
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> h-lr-r primer
 <400> 102
 gcaaacagcc ttcttccac attcgg 26

<210> 103

<211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> m-5'-r primer
 <400> 103
 ttgctttcag tagttcaggt gtgc 24
 <210> 104
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> h-5'-r primer
 <400> 104
 ggcgttgtca ggaagttgcc 20
 <210> 105
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> m-F primer
 <400> 105
 tgaagttgag aggcacatga gg 22

 <210> 106
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> m-E2 primer
 <400> 106
 tagagtagcc acaggcagca aagc 24
 <210> 107
 <211> 21
 <212> RNA

<213> Artificial Sequence

<220>

<223> Cmah gRNA A DNA-targeting segment

<400> 107

gugacaggaa ggcucucac c 21

<210> 108

<211> 21

<212> RNA

<213> Artificial Sequence

<220>

<223> Cmah gRNA B DNA-targeting segment

<400> 108

gcuuacaagc aauugcuga c 21

<210> 109

<211> 108

<212> DNA

<213> Artificial Sequence

<220>

<223> Cmah locus

<400> 109

ttgccagtga ccctgtttgc agtttagagt gacaggaagg cttctcacc gggacatttt 60

aaatgaccca gtcagcaaat tgcttgtaag ttttggtgtt ctttcatt 108

<210> 110

<211> 46

<212> DNA

<213> Artificial Sequence

<220>

<223> Cmah locus upstream of gRNA A cut

<400> 110

ttgccagtga ccctgtttgc agtttagagt gacaggaagg cttctc 46

<210> 111

<211> 35

<212> DNA

<213> Artificial Sequence

<220>

<223> Cmah locus downstream of gRNA B cut

<400> 111

agcaaattgc ttgtaagttt tgggtgttctt tcatt 35

<210> 112

<211> 27

<212> DNA

<213> Artificial Sequence

<220>

<223> Cmah locus sequence excised by gRNAs A and B

<400> 112

acccgggaca ttttaaatga ccagtc 27

<210> 113

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> C5 primer m-5'-F

<400> 113

accccagcat ctgacgacac c 21

<210> 114

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> C5 primer m-5'-R

<400> 114

agaaagaccg cagtggaacc 20

<210> 115

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> C5 primer h-5'-R
 <400> 115
 tccccacatg cctagtagga g 21
 <210> 116
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cmah TaqMan forward primer
 <400> 116
 gtgaccctgt ttgcagttag ag 22
 <210> 117
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cmah TaqMan probe
 <400> 117
 acaggaaggc ttctcacccg gga 23
 <210> 118
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Cmah TaqMan reverse primer
 <400> 118
 tgaagtcagg aaactgttcc aatg 24
 <210> 119
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 7064retU forward primer
 <400> 119

cctcctgagc tticctttgc ag	22
<210> 120	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 7064retU reverse primer	
<400> 120	
cctagacaac acagacactg tatca	25
<210> 121	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 7064retU TaqMan probe	
<400> 121	
ttctgccctt gaaaaggaga ggc	23
<210> 122	
<211> 18	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 7064retD forward primer	
<400> 122	
cctctgaggc cacctgaa	18
<210> 123	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 7064retD reverse primer	
<400> 123	
ccctgacaag ttctgccttc tac	23

<210> 124
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 7064retD TaqMan probe
 <400> 124
 tgcccaagcc tctgcagctt t 21
 <210> 125
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 7140retU forward primer
 <400> 125
 cccagcatct gacgacacc 19
 <210> 126
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 7140retU reverse primer
 <400> 126
 gaccactgtg ggcacatgta g 21
 <210> 127
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 7140retU TaqMan probe
 <400> 127
 ccgagtctgc tgttactgtt agcatca 27
 <210> 128
 <211> 20

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> 7140retD forward primer
 <400> 128
 cccgacacct tctgagcatg 20

<210> 129
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 7140retD reverse primer
 <400> 129
 tgcaggctga gtcaggatgtt g 21

<210> 130
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 7140retD TaqMan probe
 <400> 130
 tagtcacgtt ttgtgacacc ccaga 25

<210> 131
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> mADAM6-2 LOA forward primer
 <400> 131
 agggctgagg gagaacatat ac 22

<210> 132
 <211> 20
 <212> DNA

<213> Artificial Sequence

<220>

<223> mADAM6-2 LOA reverse primer

<400> 132

aggcctgatg caggagctat 20

<210> 133

<211> 26

<212> DNA

<213> Artificial Sequence

<220>

<223> mADAM6-2 LOA probe

<400> 133

tcctctcagc tggattaaca gcatca 26

<210> 134

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> hIgH31 LOA forward primer

<400> 134

atcacactca tcccatcccc 20

<210> 135

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> hIgH31 LOA reverse primer

<400> 135

cacaggaag caggaactgc 20

<210> 136

<211> 29

<212> DNA

<213> Artificial Sequence

<220>

<223> hIgh31 LOA probe
 <400> 136
 cccttccta agtaccacag agtgggctc 29
 <210> 137
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgh9 LOA forward primer
 <400> 137
 tcctccaacg acaggtccc 19

 <210> 138
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgh9 LOA reverse primer
 <400> 138
 gatgaactga cgggcacagg 20
 <210> 139
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgh9 LOA probe
 <400> 139
 tccctggaac tctgccccga caca 24
 <210> 140
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgh1 LOA forward primer
 <400> 140

cagtcccggtt gatccagcc	19
<210> 141	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> hIgH1 LOA reverse primer	
<400> 141	
ggatatgcag cactgtgcc a	21
<210> 142	
<211> 30	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> hIgH1 LOA probe	
<400> 142	
cccatcaggg atttgtatc tctgtggacg	30
<210> 143	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Neo GOA forward primer	
<400> 143	
ggtggagagg ctattcggc	19
<210> 144	
<211> 17	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Neo GOA reverse primer	
<400> 144	
gaacacggcg gcatcag	17

<210> 145	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> Neo GOA probe	
<400> 145	
tgggcacaac agacaatcgg ctg	23
<210> 146	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 5' IgH Arml retention assay forward primer	
<400> 146	
aggatgctgg gaaacagac	19
<210> 147	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 5' IgH Arml retention assay reverse primer	
<400> 147	
gaccactctc aggactctca	20
<210> 148	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 5' IgH Arml retention assay probe	
<400> 148	
tggaaggtcc caaaggaaac caca	24
<210> 149	
<211> 22	

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgM398 retention assay forward primer

<400> 149

gagctcacac cttgaccttt ca 22

<210> 150

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgM398 retention assay reverse primer

<400> 150

tggtgggacg aacacattta ca 22

<210> 151

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgM398 retention assay probe

<400> 151

ccagctgtcg cagagatgaa cccc 24

<210> 152

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgM1045 retention assay forward primer

<400> 152

tcctccaca gacatcctaa cc 22

<210> 153

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgM1045 retention assay reverse primer

<400> 153

gttagcggac ttgctgagga a 21

<210> 154

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgM1045 retention assay probe

<400> 154

tcaccatccc cccctccttt gc 22

<210> 155

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> 3' IgH Arm2 retention assay forward primer

<400> 155

ggtcattgtgg caaggctatt tg 22

<210> 156

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> 3' IgH Arm2 retention assay reverse primer

<400> 156

agcctggact ttcggtttgg t 21

<210> 157

<211> 30

<212> DNA

<213> Artificial Sequence

<220>

<223> 3' IgH Arm2 retention assay probe

<400> 157

ccactaggta aacttgtagc tgtggtttga 30

<210> 158

<211> 17

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgHp2 parental forward primer

<400> 158

gccatgcaag gccaaagc 17

<210> 159

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgHp2 parental reverse primer

<400> 159

agttcttgag ccttaggtg ctag 24

<210> 160

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgHp2 parental probe

<400> 160

ccaggaaaat gctgccagag cctg 24

<210> 161

<211> 20

<212> DNA

<213> Artificial Sequence

<220>

<223> mIgKd2 parental forward primer

<400> 161
gcaaacaata accactggcc 20
<210> 162
<211> 19
<212> DNA
<213> Artificial Sequence
<220>
<223> mIgKd2 parental reverse primer
<400> 162
ggccacattc catgggttc 19
<210>
> 163
<211> 37
<212> DNA
<213> Artificial Sequence
<220>
<223> mIgKd2 parental probe
<400> 163
ctgttcctct aaaactggac tccacagtaa atggaaa 37
<210> 164
<211> 20
<212> DNA
<213> Artificial Sequence
<220>
<223> hIgK5 parental forward primer
<400> 164
ccccgtctc ctctttttc 20
<210> 165
<211> 19
<212> DNA
<213> Artificial Sequence
<220>
<223> hIgK5 parental reverse primer
<400> 165

tgcaagtgct gccagcaag	19
<210> 166	
<211> 32	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> hIgK5 parental probe	
<400> 166	
tcatgtccat taacccattt accttttgcc ca	32
<210> 167	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 3' gRNA_I DNA-targeting sequence	
<400> 167	
gactactacg gtatggacgt c	21
<210> 168	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 3' gRNA_II DNA-targeting sequence	
<400> 168	
gctactacgg tatggacgtc t	21
<210> 169	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> 5' gRNA_I DNA-targeting sequence	
<400> 169	
gaagctgact agtttacgca	20

<210> 170
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 5' gRNA_II DNA-targeting sequence
 <400> 170
 gtagcattct tacacctagc a 21
 <210> 171
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 5' IgH Arm2 retention assay forward primer
 <400> 171
 ctgagcatac tgctgcctaa cac 23
 <210> 172
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 5' IgH Arm2 retention assay reverse primer
 <400> 172
 gagcagtgca tttcttagtt aagga 25
 <210> 173
 <211> 26
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 5' IgH Arm2 retention assay probe
 <400> 173
 tgaaatggca gttcttctcc agctgg 26
 <210> 174
 <211> 20

<212
 > DNA
 <213> Artificial Sequence
 <220>
 <223> 3' IgH Arm1 retention assay forward primer
 <400> 174
 ggtggagtcc ctggatgatg 20
 <210> 175
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 3' IgH Arm1 retention assay reverse primer
 <400> 175
 atccctccag ccataggatt g 21
 <210> 176
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> 3' IgH Arm1 retention assay probe
 <400> 176
 ctttggaggc tcatttgagg gagatgct 28