



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2017-0093246
(43) 공개일자 2017년08월14일

- | | |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)
C12N 15/90 (2006.01) A01K 67/027 (2006.01)
C12N 15/85 (2006.01) C12N 9/22 (2006.01)</p> <p>(52) CPC특허분류
C12N 15/907 (2013.01)
A01K 67/0276 (2013.01)</p> <p>(21) 출원번호 10-2017-7019826</p> <p>(22) 출원일자(국제) 2015년12월18일
심사청구일자 없음</p> <p>(85) 번역문제출일자 2017년07월17일</p> <p>(86) 국제출원번호 PCT/US2015/066681</p> <p>(87) 국제공개번호 WO 2016/100819
국제공개일자 2016년06월23일</p> <p>(30) 우선권주장
62/094,104 2014년12월19일 미국(US)
(뒷면에 계속)</p> | <p>(71) 출원인
리제너론 파마슈티칼스 인코포레이티드
미합중국 뉴욕주 10591 타리타운 올드 소우 밀 리버 로드 777</p> <p>(72) 발명자
보로니나 베라
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 (뒷면에 계속)</p> <p>(74) 대리인
송봉식, 박진우, 정삼영</p> |
|--|---|

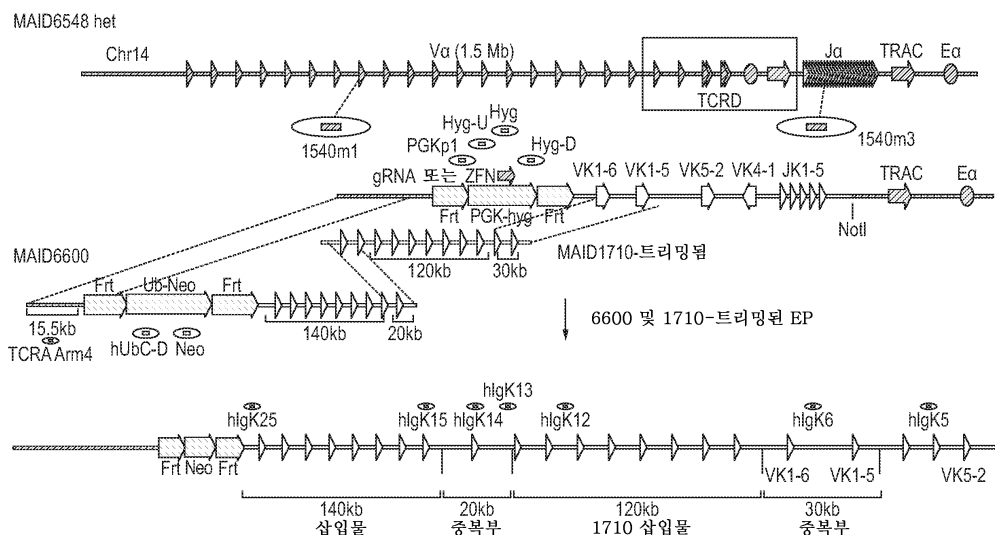
전체 청구항 수 : 총 57 항

(54) 발명의 명칭 **단일 단계 다중 표적화를 통한 표적화된 유전자 변형을 위한 방법 및 조성물**

(57) 요약

세포 내의 표적 게놈 유전자좌에 하나 이상의 표적화된 유전자 변형을 생성시키고, 변형된 게놈 유전자좌를 포함하는 비인간 동물을 생산하는 방법 및 조성물이 제공된다. 방법은, 단일 게놈 표적화 단계로 서로 그리고 표적 게놈 유전자좌와 재조합가능한 2개 이상의 큰 표적화 벡터를 사용한다. 방법은 또한 뉴클레아제 작용제와 조합하여 사용될 수 있다. 또한, 세포의 표적 게놈 유전자좌에서 상동 재조합을 향상시키는 방법 및 조성물이 제공된다. 방법은 하나 이상의 중복 서열을 포함하는 2개 이상의 핵산을 사용한다. 방법은 뉴클레아제 작용제와 조합하여 또는 뉴클레아제 작용제 없이 사용될 수 있다.

대표도 - 도1



(52) CPC특허분류

A01K 67/0278 (2013.01)

C12N 15/8509 (2013.01)

C12N 9/22 (2013.01)

A01K 2217/072 (2013.01)

A01K 2217/075 (2013.01)

A01K 2227/105 (2013.01)

(72) 발명자

잠브로비츠 브라이언

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

(30) 우선권주장

62/167,408 2015년05월28일 미국(US)

62/205,524 2015년08월14일 미국(US)

명세서

청구범위

청구항 1

하기 단계를 포함하는, 세포에서 표적 게놈 유전자좌를 변형시키는 방법:

(a) 상기 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부(break)를 생성시키는 뉴클레아제 작용제(agent)를 상기 세포 내로 도입시키는 단계;

(b) 길이가 적어도 10 kb이며 제1의 5' 상동성 아암(homology arm)과 제1의 3' 상동성 아암에 의해 플랭킹(flanking)된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터(large targeting vector, LTVEC), 및 길이가 적어도 10 kb이며 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC를 상기 세포 내로 도입시키는 단계 -

상기 제1의 LTVEC의 제1의 3' 상동성 아암은 상기 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 제1 중복 서열을 가지고, 상기 제1의 LTVEC의 제1의 5' 상동성 아암과 상기 제2의 LTVEC의 제2의 3' 상동성 아암은 상기 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이며,

상기 표적 게놈 유전자좌는 상기 상응하는 게놈 세그먼트들 사이에서 상기 제1 핵산 삽입물과 상기 제2 핵산 삽입물의 통합에 의해 변형됨 -; 및

(c) 상기 표적 게놈 유전자좌 내로 통합된 상기 제1 핵산 삽입물과 상기 제2 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계.

청구항 2

제1항에 있어서, 상기 제1 핵 삽입물과 상기 제1의 3' 상동성 아암 그리고 상기 제2 핵산 삽입물과 상기 제2의 5' 상동성 아암은 연속 핵산의 중복 단편들이며, 상기 연속 핵산은 상기 표적 게놈 유전자좌 내로의 상기 제1 핵산 삽입물과 상기 제2 핵산 삽입물의 통합에 의해 재형성되는, 방법.

청구항 3

제1항 또는 제2항에 있어서, 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 또는 둘 모두는 상기 세포의 종과 상이한 종으로부터 유래되는, 방법.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 또는 둘 모두는 인간 핵산인, 방법.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 상기 제1 핵산 삽입물과 상기 제2 핵산 삽입물을 합한 크기는 약 50 kb 내지 약 500 kb, 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, 약 275 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 약 350 kb 내지 약 400 kb, 약 400 kb 내지 약 450 kb, 또는 약 450 kb 내지 약 500 kb인, 방법.

청구항 6

제5항에 있어서, 상기 제1 핵산 삽입물과 상기 제2 핵산 삽입물을 합한 크기는 약 100 kb 내지 약 500 kb인, 방법.

청구항 7

제6항에 있어서, 상기 제1 핵산 삽입물과 상기 제2 핵산 삽입물을 합한 크기는 약 300 kb인, 방법.

청구항 8

제1항 내지 제4항 중 어느 한 항에 있어서, 상기 표적화된 세포는, 합한 크기가 약 5 kb 내지 약 500 kb의 범위인 상기 제1 핵산 삽입물과 상기 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함하는, 방법.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 제1의 LTVEC의 제1 중복 서열은 상기 제2의 LTVEC의 제1 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일한, 방법.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서, 상기 제1 중복 서열의 크기는 약 1 kb 내지 약 70 kb인, 방법.

청구항 11

제1항 내지 제10항 중 어느 한 항에 있어서, 상기 제1 중복 서열의 크기는 적어도 10 kb 또는 적어도 20 kb인, 방법.

청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서, 상기 표적 게놈 유전자좌 내로의 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 또는 둘 모두의 통합은 하기 중 하나 이상을 가져오는, 방법:

- (a) 상기 표적 게놈 유전자좌에 외인성 서열의 첨가;
- (b) 상기 표적 게놈 유전자좌에서 내인성 서열의 결실; 및
- (c) 녹인(knock-in), 녹아웃(knockout), 점 돌연변이, 도메인 교체(swap), 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합.

청구항 13

제1항 내지 제12항 중 어느 한 항에 있어서, 상기 제1 LTVEC와 상기 제2 LTVEC의 병용은 단일 LTVEC의 사용과 비교하여 표적화 효율의 증가를 가져오는, 방법.

청구항 14

제13항에 있어서, 상기 표적화 효율의 증가는 적어도 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배인, 방법.

청구항 15

제1항 내지 제14항 중 어느 한 항에 있어서, 상기 제1의 LTVEC 또는 상기 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb인, 방법.

청구항 16

하기 단계를 포함하는, 세포에서 표적 게놈 유전자좌를 변형시키는 방법:

- (a) 상기 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 상기 세포 내로 도입시키는 단계;
- (b) 길이가 적어도 10 kb이며 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 길이가 적어도 10 kb이며 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC, 및 길이가 적어도 10 kb이며 제3의 5' 상동성 아암과 제3의 3' 상동성 아암에 의해 플랭킹된 제3 핵산 삽입물을 포함하는 제3의 LTVEC를 상기 세포 내로 도입

시키는 단계 -

상기 제1의 LTVEC의 제1의 3' 상동성 아암은 상기 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 제1 중복 서열을 가지고, 상기 제2의 LTVEC의 제2의 3' 상동성 아암은 상기 제3의 LTVEC의 제3의 5' 상동성 아암에 상동성인 제2 중복 서열을 가지고, 상기 제1의 LTVEC의 제1의 5' 상동성 아암과 상기 제3의 LTVEC의 제3의 3' 상동성 아암은 상기 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이며,

상기 표적 게놈 유전자좌는 상기 상응하는 게놈 세그먼트들 사이에서 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물의 통합에 의해 변형됨 -; 및

(c) 상기 표적 게놈 유전자좌 내로 통합된 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계.

청구항 17

제16항에 있어서, 상기 제1 핵 삽입물과 상기 제1의 3' 상동성 아암 그리고 상기 제2 핵산 삽입물과 상기 제2의 5' 상동성 아암은 연속 핵산의 중복 단편들이고, 상기 제2 핵 삽입물과 상기 제2의 3' 상동성 아암 그리고 상기 제3 핵산 삽입물과 상기 제3의 5' 상동성 아암은 상기 연속 핵산의 중복 단편들이며, 상기 연속 핵산은 상기 표적 게놈 유전자좌 내로의 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물의 통합에 의해 재형성되는, 방법.

청구항 18

제16항 또는 제17항에 있어서, 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물 중 하나 이상은 상기 세포의 종과 상이한 종으로부터 유래되는, 방법.

청구항 19

제16항 내지 제18항 중 어느 한 항에 있어서, 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물 중 하나 이상은 인간 핵산인, 방법.

청구항 20

제16항 내지 제19항 중 어느 한 항에 있어서, 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물을 합한 크기는 약 50 kb 내지 약 700 kb, 약 50 kb 내지 약 500 kb, 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, 약 275 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 약 350 kb 내지 약 400 kb, 약 400 kb 내지 약 450 kb, 약 450 kb 내지 약 500 kb, 약 500 kb 내지 약 550 kb, 약 550 kb 내지 약 600 kb, 약 600 kb 내지 약 650 kb, 또는 약 650 kb 내지 약 700 kb인, 방법.

청구항 21

제20항에 있어서, 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물을 합한 크기는 약 100 kb 내지 약 700 kb인, 방법.

청구항 22

제21항에 있어서, 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물을 합한 크기는 약 400 kb인, 방법.

청구항 23

제16항 내지 제19항 중 어느 한 항에 있어서, 상기 표적화된 세포는, 합한 크기가 약 5 kb 내지 약 700 kb의 범위인 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함하는, 방법.

청구항 24

제16항 내지 제23항 중 어느 한 항에 있어서, 상기 제1의 LTVEC의 제1 중복 서열은 상기 제2의 LTVEC의 제1 중

복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하고/하거나 상기 제2의 LTVEC의 제2 중복 서열은 상기 제3의 LTVEC의 제2 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일한, 방법.

청구항 25

제16항 내지 제24항 중 어느 한 항에 있어서, 상기 제1 중복 서열의 크기는 약 1 kb 내지 약 70 kb이고/이거나 상기 제2 중복 서열의 크기는 약 1 kb 내지 약 70 kb인, 방법.

청구항 26

제16항 내지 제25항 중 어느 한 항에 있어서, 상기 제1 중복 서열의 크기는 적어도 10 kb 또는 적어도 20 kb이고/이거나 상기 제2 중복 서열의 크기는 적어도 10 kb 또는 적어도 20 kb인, 방법.

청구항 27

제16항 내지 제26항 중 어느 한 항에 있어서, 상기 표적 게놈 유전자좌 내로의 상기 제1 핵산 삽입물, 상기 제2 핵산 삽입물 및 상기 제3 핵산 삽입물 중 하나 이상의 통합은 하기 중 하나 이상을 가져오는, 방법:

- (a) 상기 표적 게놈 유전자좌에 외인성 서열의 첨가;
- (b) 상기 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는
- (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합.

청구항 28

제16항 내지 제27항 중 어느 한 항에 있어서, 상기 제1의 LTVEC, 상기 제2의 LTVEC 또는 상기 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb인, 방법.

청구항 29

제12항 또는 제27항에 있어서, 상기 표적 게놈 유전자좌에서 상기 내인성 서열의 결실은 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 600 kb, 약 600 kb 내지 약 700 kb, 또는 약 700 kb 내지 약 800 kb인, 방법.

청구항 30

제1 핵산과 제2 핵산을 세포 내로 도입시키는 단계 - 상기 제1 핵산과 상기 제2 핵산은 중복 뉴클레오티드 서열을 포함함 - 를 포함하는, 상기 세포의 표적 게놈 유전자좌에서 상동 재조합을 향상시키는 방법.

청구항 31

제30항에 있어서, 상동 재조합은 뉴클레아제 작용제를 사용하지 않으면서 상기 표적 게놈 유전자좌에서 향상되는, 방법.

청구항 32

제30항에 있어서, 상기 표적 게놈 유전자좌에 또는 이 부근에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 상기 세포 내로 도입시키는 단계를 추가로 포함하는, 방법.

청구항 33

제1항 내지 제29항 및 제32항 중 어느 한 항에 있어서, 상기 뉴클레아제 작용제는 징크 핑거 뉴클레아제(zinc finger nuclease, ZFN), 전사 활성화인자 유사 이펙터 뉴클레아제(Transcription Activator-Like Effector

Nuclease, TALEN) 또는 메가뉴클레아제(meganuclease)인, 방법.

청구항 34

제1항 내지 제29항 및 제32항 중 어느 한 항에 있어서, 상기 뉴클레아제 작용제는 CRISPR(Clustered Regularly Interspaced Short Palindromic Repeat, 클러스터링된 규칙적으로 사이 공간을 둔 짧은 회문성 반복체) 관련 (CRISPR-associated, Cas) 단백질 및 가이드 RNA (gRNA)를 포함하는, 방법.

청구항 35

제34항에 있어서, 상기 Cas 단백질은 Cas9인, 방법.

청구항 36

제30항 내지 제35항 중 어느 한 항에 있어서, 상기 방법은 상기 표적 게놈 유전자좌에서 상기 제1 핵산, 상기 제2 핵산 또는 둘 모두의 상동 재조합을 향상시키는, 방법.

청구항 37

제36항에 있어서, 상기 상동 재조합의 향상은 적어도 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배인, 방법.

청구항 38

제30항 내지 제37항 중 어느 한 항에 있어서, 상기 제1 핵산의 중복 서열은 상기 제2 핵산의 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일한, 방법.

청구항 39

제30항 내지 제38항 중 어느 한 항에 있어서, 상기 중복 서열은 약 1 kb 내지 약 70 kb인, 방법.

청구항 40

제30항 내지 제39항 중 어느 한 항에 있어서, 상기 중복 서열은 적어도 20 kb인, 방법.

청구항 41

제30항 내지 제40항 중 어느 한 항에 있어서, 상기 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 표적화 벡터이고, 상기 제2 핵산은 상기 중복 서열을 제외하고는 상기 표적 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는, 방법.

청구항 42

제30항 내지 제40항 중 어느 한 항에 있어서, 상기 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1 표적화 벡터이고, 상기 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터인, 방법.

청구항 43

제41항 또는 제42항에 있어서, 상기 제1 표적화 벡터는 약 20 kb 내지 약 200 kb 범위의 제1의 큰 표적화 벡터 (LTVEC)인, 방법.

청구항 44

제42항에 있어서, 상기 제2 표적화 벡터는 약 20 kb 내지 약 200 kb 범위의 제2의 큰 표적화 벡터 (LTVEC)인, 방법.

청구항 45

제43항 또는 제44항에 있어서, 상기 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb인, 방법.

청구항 46

제44항 또는 제45항에 있어서, 상기 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb인, 방법.

청구항 47

제30항 내지 제46항 중 어느 한 항에 있어서, 상기 중복 서열은 상기 제1 핵산의 3' 말단 및 상기 제2 핵산 서열의 5' 말단에 위치하는, 방법.

청구항 48

제30항 내지 제47항 중 어느 한 항에 있어서, 상기 중복 뉴클레오티드 서열은 상기 표적 게놈 유전자좌로의 재조합 기구(recombination machinery)의 보충을 용이하게 하는, 방법.

청구항 49

제1항 내지 제48항 중 어느 한 항에 있어서, 상기 세포는 인간 세포인, 방법.

청구항 50

제1항 내지 제48항 중 어느 한 항에 있어서, 상기 세포는 비인간(non-human) 세포인, 방법.

청구항 51

제1항 내지 제50항 중 어느 한 항에 있어서, 상기 세포는 만능성(pluripotent) 세포, 조혈 줄기 세포, 신경(neuronal) 줄기 세포 또는 섬유아세포인, 방법.

청구항 52

제51항에 있어서, 상기 만능성 세포는 배아 줄기 (ES) 세포 또는 유도 만능성 줄기 (iPS) 세포인, 방법.

청구항 53

제50항 내지 제52항 중 어느 한 항에 있어서, 상기 세포는 포유류 세포인, 방법.

청구항 54

제53항에 있어서, 상기 포유류 세포는 설치류 세포인, 방법.

청구항 55

제54항에 있어서, 상기 설치류는 마우스 또는 래트인, 방법.

청구항 56

하기 단계를 포함하는, F0 세대 비인간 동물을 생산하는 방법:

(a) 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계 - 상기 비인간 ES 세포는 제1항 내지 제48항 및 제50항 내지 제55항 중 어느 한 항의 방법에 의해 생성되었음 -; 및

(b) 상기 비인간 숙주 배아를 대리모에 임신시키는 단계 -

상기 대리모는 변형을 포함하는 F0 세대 비인간 동물을 생산함 -.

청구항 57

제56항에 있어서, 상기 비인간 동물은 마우스 또는 래트인, 방법.

발명의 설명

기술 분야

- [0001] 관련 출원의 상호 참조
- [0002] 본 출원은 2014년 12월 19일에 출원된 미국 특허 출원 제62/094,104호, 2015년 5월 28일에 출원된 미국 특허 출원 제62/167,408호 및 2015년 8월 14일에 출원된 미국 특허 출원 제62/205,524호와 관련되며, 이들 각각은 모든 목적을 위해 전체적으로 본 명세서에 참고로 포함된다.
- [0003] EFS 웹을 통해 텍스트 파일로 제출된 서열 목록의 참조
- [0004] 472224SEQLIST.txt 파일로 작성된 서열 목록은 16.7 kb이고, 2015년 12월 16일에 생성되었으며, 본 명세서에 참고로 포함된다.

배경 기술

- [0005] 게놈 유전자좌에서 특정 핵산 서열을 첨가, 결실 또는 대체하도록 설계된 표적화 벡터를 사용하는 상동 재조합은 비인간(non-human) 동물에서 요망되는 게놈 변형을 달성하기 위한 일반적인 접근법이다.
- [0006] 상동 재조합을 통한 게놈 변형 기술은 지난 20년에 걸쳐 상당히 진보하였지만, 많은 경우에, 예를 들어 설치류 게놈의 큰 부분이 큰 인간 게놈 단편으로 대체될 때 또는 특정 세포 유형, 예컨대 섬유아세포 또는 다른 체세포를 표적화할 때, 매우 큰 표적화 벡터인 LTVEC를 사용하여 허용가능한 표적화 빈도를 달성하는 데 여전히 어려움이 남아 있다.

발명의 내용

- [0007] 서로 재조합되어 단일의 연속 핵산 세그먼트를 형성할 수 있는 2개 이상의 표적화 벡터를 이용하는 표적화 시스템을 통해 세포 내의 표적 게놈 유전자좌를 변형시키는 방법 및 조성물이 제공된다. 선택적으로, 표적화 벡터는 큰 표적화 벡터(large targeting vector, LTVEC)이다. 선택적으로, LTVEC들은 각각 크기가 적어도 10 kb이다.
- [0008] 본 발명은 세포에서 표적 게놈 유전자좌를 변형시키는 방법을 제공하며, 이 방법은 (a) 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부(break)를 생성시키는 뉴클레아제 작용제(agent)를 세포 내로 도입시키는 단계; (b) 제1의 5' 상동성 아암(homology arm)과 제1의 3' 상동성 아암에 의해 플랭킹(flanking)된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 및 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC를 세포 내로 도입시키는 단계 - 제1의 LTVEC의 제1의 5' 상동성 아암과 제2의 LTVEC의 제2의 3' 상동성 아암은 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이고, 제1의 LTVEC의 제1의 3' 상동성 아암과 제2의 LTVEC의 제2의 5' 아암은 서로 상동성이거나 하나 이상의 추가 LTVEC의 추가 5' 상동성 아암과 3' 상동성 아암에 각각 상동성이고, 이 추가 LTVEC 각각은 추가 5' 상동성 아암과 추가 3' 상동성 아암에 의해 플랭킹된 추가 핵산 삽입물을 포함하며, 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물, 존재하는 경우, 하나 이상의 추가 LTVEC의 하나 이상의 추가 핵산 삽입물, 및 제2 핵산 삽입물의 통합에 의해 변형됨 -; 및 (c) 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물, 존재하는 경우, 하나 이상의 추가 핵산 삽입물, 및 제2 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 포함한다. 선택적으로, 제1의 LTVEC, 제2의 LTVEC, 및 하나 이상의 추가 LTVEC는 각각 크기가 적어도 10 kb이다. 일부 이러한 방법에서, 추가 LTVEC는, 존재할 때 제1의 LTVEC와 제2의 LTVEC 사이에 삽입되는 하나 이상의 다른 LTVEC이다.
- [0009] 또한, 본 발명은 세포에서 표적 게놈 유전자좌를 변형시키는 이중 표적화 방법을 제공하며, 이 방법은 (a) 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계; (b) 크기가 적어도 10 kb이며 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 및 길이가 적어도 10 kb이며 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC를 세포 내로 도입시키는 단계 - 제1의 LTVEC의 제1의 3' 상동성 아암은 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 제1 중복 서열을 가지고, 제1의 LTVEC의 제1의 5' 상동성 아암과 제2의 LTVEC의 제2의 3' 상동성 아암은 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이며, 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물 및 제2 핵산 삽입물의 통합에 의해 변형됨 -; 및 (c) 표적 게놈 유전자좌 내로 통합된 제1 핵산 삽입물 및 제2 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 포함한다.

- [0010] 선택적으로, 제1 핵 삽입물과 제1의 3' 상동성 아암 그리고 제2 핵산 삽입물과 제2의 5' 상동성 아암은 연속 핵산의 중복 단편들이며, 그 연속 핵산은 표적 게놈 유전자좌 내로의 제1 핵산 삽입물과 제2 핵산 삽입물의 통합에 의해 재형성된다.
- [0011] 일부 이러한 방법에서, 세포는 인간 세포이다. 다른 이러한 방법에서, 세포는 비인간 세포이다. 일부 이러한 방법에서, 세포는 만능성(pluripotent) 세포, 조혈 줄기 세포, 신경(neuronal) 줄기 세포 또는 섬유아세포이다. 선택적으로, 만능성 세포는 배아 줄기 (ES) 세포 또는 유도 만능성 줄기 (iPS) 세포이다. 일부 이러한 방법에서, 세포는 포유류 세포이다. 선택적으로, 포유류 세포는 설치류 세포이다. 선택적으로, 설치류 세포는 마우스 세포 또는 래트 세포이다.
- [0012] 상기 방법들 중 일부에서, 뉴클레아제 작용제는 징크 핑거 뉴클레아제(zinc finger nuclease, ZFN), 전사 활성화 인자 유사 이펙터 뉴클레아제(Transcription Activator-Like Effector Nuclease, TALEN) 또는 메가뉴클레아제(meganuclease)이다. 상기 방법들 중 일부에서, 뉴클레아제 작용제는 CRISPR(Clustered Regularly Interspaced Short Palindromic Repeat, 클러스터링된 규칙적으로 사이 공간을 둔 짧은 회문성 반복체) 관련 (CRISPR-associated, Cas) 단백질 및 가이드 RNA (gRNA)를 포함한다. 선택적으로, Cas 단백질은 Cas9이다.
- [0013] 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 세포의 종과 상이한 종으로부터 유래된다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 인간 핵산이다.
- [0014] 일부 방법에서, 제1 핵산 삽입물과 제2 핵산 삽입물을 합한 크기는 약 50 kb 내지 약 500 kb, 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, 약 275 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 약 350 kb 내지 약 400 kb, 약 400 kb 내지 약 450 kb, 또는 약 450 kb 내지 약 500 kb이다. 선택적으로, 제1 핵산 삽입물과 제2 핵산 삽입물을 합한 크기는 약 100 kb 내지 약 500 kb이다. 선택적으로, 제1 핵산 삽입물과 제2 핵산 삽입물을 합한 크기는 약 300 kb이다.
- [0015] 일부 방법에서, 표적화된 세포는, 합한 크기가 약 5 kb 내지 약 500 kb의 범위인 제1 핵산 삽입물과 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다.
- [0016] 일부 방법에서, 제1의 LTVEC의 제1 중복 서열은 제2의 LTVEC의 제1 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하다. 일부 방법에서, 제1 중복 서열의 크기는 약 1 kb 내지 약 70 kb이다. 일부 방법에서, 제1 중복 서열의 크기는 적어도 10 kb 또는 적어도 20 kb이다.
- [0017] 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인(knock-in), 녹아웃(knockout), 점 돌연변이, 도메인 교체(swap), 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합. 선택적으로, 표적 게놈 유전자좌에서 내인성 서열의 결실은 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 600 kb, 약 600 kb 내지 약 700 kb, 또는 약 700 kb 내지 약 800 kb이다.
- [0018] 일부 방법에서, 제1 LTVEC와 제2 LTVEC의 병용은 단일 LTVEC의 사용과 비교하여 표적화 효율의 증가를 가져온다. 선택적으로, 표적화 효율의 증가는 적어도 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배이다.
- [0019] 일부 방법에서, 제1의 LTVEC 또는 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이고, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이다. 일부 방법에서, 제1의 LTVEC 또는 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb이고, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb이다.

- [0020] 또한, 본 발명은 (a) 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계 - 비인간 ES 세포는 상기 방법들 중 어느 것에 의해 생성되었음 -; 및 (b) 비인간 숙주 배아를 대리모에 임신시키는 단계 - 대리모는 변형을 포함하는 F0 세대 비인간 동물을 생산함 - 를 포함하는, F0 세대 비인간 동물을 생산하는 방법을 제공한다. 선택적으로, 비인간 동물은 마우스 또는 래트이다.
- [0021] 또한, 본 발명은 세포에서 표적 게놈 유전자좌를 변형시키는 삼중 표적화 방법을 제공하며, 이 방법은 (a) 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계; (b) 길이가 적어도 10 kb이며 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 길이가 적어도 10 kb이며 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC, 및 길이가 적어도 10 kb이며 제3의 5' 상동성 아암과 제3의 3' 상동성 아암에 의해 플랭킹된 제3 핵산 삽입물을 포함하는 제3의 LTVEC를 세포 내로 도입시키는 단계 - 제1의 LTVEC의 제1의 3' 상동성 아암은 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 제1 중복 서열을 가지고, 제2의 LTVEC의 제2의 3' 상동성 아암은 제3의 LTVEC의 제3의 5' 상동성 아암에 상동성인 제2 중복 서열을 가지고, 제1의 LTVEC의 제1의 5' 상동성 아암과 제3의 LTVEC의 제3의 3' 상동성 아암은 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이며, 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물의 통합에 의해 변형됨 -; 및 (c) 표적 게놈 유전자좌 내로 통합된 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 포함한다.
- [0022] 선택적으로, 제1 핵 삽입물과 제1의 3' 상동성 아암 그리고 제2 핵산 삽입물과 제2의 5' 상동성 아암은 연속 핵산의 중복 단편들이고, 제2 핵산 삽입물과 제2의 3' 상동성 아암 그리고 제3 핵산 삽입물과 제3의 5' 상동성 아암은 그 연속 핵산의 중복 단편들이며, 그 연속 핵산은 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물의 통합에 의해 재형성된다.
- [0023] 일부 이러한 방법에서, 세포는 인간 세포이다. 다른 이러한 방법에서, 세포는 비인간 세포이다. 일부 이러한 방법에서, 세포는 만능성 세포, 조혈 줄기 세포, 신경 줄기 세포 또는 섬유아세포이다. 선택적으로, 만능성 세포는 배아 줄기 (ES) 세포 또는 유도 만능성 줄기 (iPS) 세포이다. 일부 이러한 방법에서, 세포는 포유류 세포이다. 선택적으로, 포유류 세포는 설치류 세포이다. 선택적으로, 설치류 세포는 마우스 세포 또는 래트 세포이다.
- [0024] 일부 이러한 방법에서, 뉴클레아제 작용제는 징크 핑거 뉴클레아제 (ZFN), 전사 활성화인자 유사 이펙터 뉴클레아제 (TALEN) 또는 메가뉴클레아제이다. 일부 이러한 방법에서, 뉴클레아제 작용제는 CRISPR 관련 (Cas) 단백질 및 가이드 RNA (gRNA)를 포함한다. 선택적으로, Cas 단백질은 Cas9이다.
- [0025] 일부 이러한 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물 중 하나 이상은 세포의 종과 상이한 종으로부터 유래된다. 일부 이러한 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물은 인간 핵산이다.
- [0026] 일부 이러한 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 합한 크기는 약 50 kb 내지 약 700 kb, 약 50 kb 내지 약 500 kb, 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, 약 275 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 약 350 kb 내지 약 400 kb, 약 400 kb 내지 약 450 kb, 약 450 kb 내지 약 500 kb, 약 500 kb 내지 약 550 kb, 약 550 kb 내지 약 600 kb, 약 600 kb 내지 약 650 kb, 또는 약 650 kb 내지 약 700 kb이다. 선택적으로, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 합한 크기는 약 100 kb 내지 약 700 kb이다. 선택적으로, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 합한 크기는 약 400 kb이다.
- [0027] 일부 이러한 방법에서, 표적화된 세포는, 합한 크기가 약 5 kb 내지 약 700 kb의 범위인 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다.
- [0028] 일부 이러한 방법에서, 제1의 LTVEC의 제1 중복 서열은 제2의 LTVEC의 제1 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하고/하거나 제2의 LTVEC의 제2 중복 서열은 제3의

LTVEC의 제2 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하다. 일부 이러한 방법에서, 제1 중복 서열의 크기는 약 1 kb 내지 약 70 kb이고/이거나 제2 중복 서열의 크기는 약 1 kb 내지 약 70 kb이다. 일부 이러한 방법에서, 제1 중복 서열의 크기는 적어도 10 kb 또는 적어도 20 kb이고/이거나 제2 중복 서열의 크기는 적어도 10 kb 또는 적어도 20 kb이다.

[0029] 일부 이러한 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물 중 하나 이상의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합. 선택적으로, 표적 게놈 유전자좌에서 내인성 서열의 결실은 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 600 kb, 약 600 kb 내지 약 700 kb, 또는 약 700 kb 내지 약 800 kb이다.

[0030] 일부 방법에서, 제1의 LTVEC, 제2의 LTVEC 또는 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이고, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이고, 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이다. 일부 이러한 방법에서, 제1의 LTVEC, 제2의 LTVEC 또는 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb이고; 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb이고; 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 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 120 kb, 또는 약 120 kb 내지 약 150 kb이다.

[0031] 또한, 본 발명은 (a) 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계 - 비인간 ES 세포는 상기 방법들 중 어느 것에 의해 생성되었음 -; 및 (b) 비인간 숙주 배아를 대리모에 임신시키는 단계 - 대리모는 변형을 포함하는 F0 세대 비인간 동물을 생산함 - 를 포함하는, F0 세대 비인간 동물을 생산하는 방법을 제공한다. 선택적으로, 비인간 동물은 마우스 또는 래트이다.

[0032] 또한, 본 발명은, 제1 핵산과 제2 핵산을 세포 내로 도입시키는 단계 - 제1 핵산과 제2 핵산은 중복 뉴클레오타이드 서열을 포함함 - 를 포함하는, 세포의 표적 게놈 유전자좌에서 상동 재조합을 향상시키는 방법을 제공한다. 일부 이러한 방법에서, 상동 재조합은 단일 핵산만이 세포 내로 도입되는 방법과 비교하여 향상된다.

[0033] 일부 이러한 방법에서, 상동 재조합은 뉴클레아제 작용제를 사용하지 않으면서 표적 게놈 유전자좌에서 향상된다. 일부 이러한 방법은 표적 게놈 유전자좌에 또는 이 부근에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계를 추가로 포함한다. 일부 이러한 방법에서, 뉴클레아제 작용제는 징크 핑거 뉴클레아제 (ZFN), 전사 활성화인자 유사 이펙터 뉴클레아제 (TALEN) 또는 메가뉴클레아제이다. 일부 이러한 방법에서, 뉴클레아제 작용제는 CRISPR 관련 (Cas) 단백질 및 가이드 RNA (gRNA)를 포함한다. 선택적으로, Cas 단백질은 Cas9이다.

[0034] 일부 이러한 방법에서, 방법은 표적 게놈 유전자좌에서 제1 핵산, 제2 핵산 또는 둘 모두의 상동 재조합을 향상시킨다. 일부 이러한 방법은 제1 핵산이 제2 핵산 없이 도입되는 방법과 비교하여 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합을 향상시킨다. 일부 이러한 방법은 제2 핵산이 제1 핵산 없이 도입되는 방법과 비교하여 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합을 향상시킨다. 선택적으로, 상동 재조합의 향상은 적어도 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배이다.

[0035] 일부 이러한 방법에서, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하다. 일부 이러한 방법에서, 중복 서열은 약 1 kb 내지 약 70 kb이다. 선택적으로, 중복 서열은 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 15 kb, 약

15 kb 내지 약 20 kb, 약 20 kb 내지 약 25 kb, 약 25 kb 내지 약 30 kb, 약 30 kb 내지 약 35 kb, 약 35 kb 내지 약 40 kb, 약 40 kb 내지 약 45 kb, 약 45 kb 내지 약 50 kb, 약 50 kb 내지 약 55 kb, 약 55 kb 내지 약 60 kb, 약 60 kb 내지 약 65 kb, 또는 약 65 kb 내지 약 70 kb이다. 일부 이러한 방법에서, 중복 서열은 적어도 5 kb, 적어도 10 kb, 적어도 15 kb, 적어도 20 kb, 적어도 25 kb, 적어도 30kb, 적어도 35 kb, 적어도 40 kb, 적어도 45 kb, 적어도 50 kb, 적어도 55 kb, 적어도 60 kb, 적어도 65 kb 또는 적어도 70 kb이다. 선택적으로, 중복 서열은 적어도 20 kb이다.

[0036] 일부 이러한 방법에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 표적화 벡터이고, 제2 핵산은 중복 서열을 제외하고는 표적 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다. 선택적으로, 제1 표적화 벡터는 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 5 kb, 또는 약 5 kb 내지 약 10 kb이다. 선택적으로, 제1 표적화 벡터는 제1의 큰 표적화 벡터 (LTVEC)이다. 선택적으로, 제1의 LTVEC는 길이가 적어도 10 kb이다. 선택적으로, 제1 표적화 벡터는 약 20 kb 내지 약 200 kb 범위의 제1의 큰 표적화 벡터 (LTVEC)이다. 선택적으로, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다.

[0037] 일부 이러한 방법에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1 표적화 벡터이고, 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터이다. 선택적으로, 제1 표적화 벡터는 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 5 kb, 또는 약 5 kb 내지 약 10 kb이고/이거나 제2 표적화 벡터는 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 5 kb, 또는 약 5 kb 내지 약 10 kb이다. 선택적으로, 제1 표적화 벡터는 제1의 큰 표적화 벡터 (LTVEC)이고/이거나 제2 표적화 벡터는 제2의 큰 표적화 벡터 (LTVEC)이다. 선택적으로, 제1의 LTVEC는 길이가 적어도 10 kb이고/이거나 제2의 LTVEC는 길이가 적어도 10 kb이다. 선택적으로, 제1 표적화 벡터는 약 20 kb 내지 약 200 kb 범위의 제1의 큰 표적화 벡터 (LTVEC)이고/이거나 제2 표적화 벡터는 약 20 kb 내지 약 200 kb 범위의 제2의 큰 표적화 벡터 (LTVEC)이다. 선택적으로, 제1의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이고/이거나 제2의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 선택적으로, 제1의 LTVEC 또는 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 선택적으로, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이고, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다.

[0038] 일부 방법에서, 중복 서열은 제1 핵산의 3' 말단 및 제2 핵산 서열의 5' 말단에 위치한다. 일부 방법에서, 중복 뉴클레오티드 서열은 표적 게놈 유전자좌로의 재조합 기구(recombination machinery)의 보충을 용이하게 한다.

[0039] 일부 이러한 방법에서, 세포는 인간 세포이다. 다른 이러한 방법에서, 세포는 비인간 세포이다. 일부 이러한 방법에서, 세포는 만능성 세포, 조혈 줄기 세포, 신경 줄기 세포 또는 섬유아세포이다. 선택적으로, 만능성 세포는 배아 줄기 (ES) 세포 또는 유도 만능성 줄기 (iPS) 세포이다. 일부 이러한 방법에서, 세포는 포유류 세포이다. 선택적으로, 포유류 세포는 설치류 세포이다. 선택적으로, 설치류 세포는 마우스 세포 또는 래트 세포이다.

[0040] 또한, 본 발명은 (a) 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계 - 비인간 ES 세포는 상기 방법들 중 어느 것에 의해 생성되었음 -; 및 (b) 비인간 숙주 배아를 대리모에 임신시키는 단계 - 대리모는 변형을 포함하는 F0 세대 비인간 동물을 생산함 - 를 포함하는, F0 세대 비인간 동물을 생산하는 방법을 제공한다. 선택적으로, 비인간 동물은 마우스 또는 래트이다.

[0041] 서로 재조합되어 단일의 연속 핵산 세그먼트를 형성할 수 있는 2개 이상의 표적화 벡터를 이용하는 표적화 시스템을 통해 세포 내의 표적 게놈 유전자좌를 변형시키는 방법 및 조성물이 제공된다. 다양한 실시 형태에서, 표적화 벡터는 큰 표적화 벡터 (LTVEC)이다. 선택적으로, LTVEC들은 각각 크기가 적어도 10 kb이다.

[0042] 일 실시 형태에서, 세포에서 표적 게놈 유전자좌를 변형시키는 방법이 제공된다. 이러한 방법은, 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계, 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 및 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는

제2의 LTVEC를 도입시키는 단계 - 제1의 LTVEC의 제1의 5' 상동성 아암과 제2의 LTVEC의 제2의 3' 상동성 아암은 표적 유전자좌 내의 상응하는 세그먼트들에 상동성이고, 제1의 LTVEC의 제1의 3' 상동성 아암과 제2의 LTVEC의 제2의 5' 아암은 서로 상동성이거나 하나 이상의 추가 LTVEC의 추가 5' 상동성 아암과 3' 상동성 아암에 각각 상동성이고, 이 추가 LTVEC 각각은 추가 5' 상동성 아암과 추가 3' 상동성 아암에 의해 플랭킹된 추가 삽입물을 포함하며, 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 삽입물, 존재하는 경우, 하나 이상의 추가 LTVEC의 하나 이상의 추가 삽입물, 및 제2 핵산 삽입물의 통합에 의해 변형됨 - 를 포함한다. 선택적으로, 제1의 LTVEC, 제2의 LTVEC, 및 하나 이상의 추가 LTVEC는 각각 크기가 적어도 10 kb이다. 이 방법은 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물, 존재하는 경우, 하나 이상의 추가 핵산 삽입물, 및 제2 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 추가로 포함한다. 이러한 방법에서, 추가 LTVEC는, 존재할 때 제1의 LTVEC와 제2의 LTVEC 사이에 삽입되는 하나 이상의 다른 LTVEC이다.

[0043] 다른 실시 형태에서, 세포에서 표적 게놈 유전자좌를 변형시키는 이중 표적화 방법이 제공된다. 이러한 방법은 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계, 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 및 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC를 도입시키는 단계를 포함한다. 선택적으로, 제1의 LTVEC는 크기가 적어도 10 kb이고, 제2의 LTVEC는 크기가 적어도 10 kb이다. 이러한 방법에서, 제1의 LTVEC의 제1의 3' 상동성 아암은 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 제1 중복 서열을 가지고, 제1의 LTVEC의 제1의 5' 상동성 아암과 제2의 LTVEC의 제2의 3' 상동성 아암은 표적 유전자좌 내의 상응하는 세그먼트들에 상동성이어서, 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물과 제2 핵산 삽입물의 통합에 의해 변형된다. 이 방법은 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물과 제2 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 추가로 포함한다.

[0044] 일부 이러한 방법에서, 제1 핵 삽입물과 제1의 3' 상동성 아암 그리고 제2 핵산 삽입물과 제2의 5' 상동성 아암은 연속 핵산의 중복 단편들이며, 그 연속 핵산은 표적 게놈 유전자좌 내로의 제1 핵산 삽입물과 제2 핵산 삽입물의 통합에 의해 재형성된다.

[0045] 다른 실시 형태에서, 세포에서 표적 게놈 유전자좌를 변형시키는 삼중 표적화 방법이 제공된다. 이러한 방법은 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계, 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC, 및 제3의 5' 상동성 아암과 제3의 3' 상동성 아암에 의해 플랭킹된 제3 핵산 삽입물을 포함하는 제3의 LTVEC를 도입시키는 단계를 포함한다. 선택적으로, 제1의 LTVEC는 크기가 적어도 10 kb이고, 제2의 LTVEC는 크기가 적어도 10 kb이고, 제3의 LTVEC는 크기가 적어도 10 kb이다. 이러한 방법에서, 제1의 LTVEC의 제1의 3' 상동성 아암은 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 제1 중복 서열을 가지고, 제2의 LTVEC의 제2의 3' 상동성 아암은 제3의 LTVEC의 제3의 5' 상동성 아암에 상동성인 제2 중복 서열을 가지고, 제1의 LTVEC의 제1의 5' 상동성 아암과 제3의 LTVEC의 제3의 3' 상동성 아암은 표적 유전자좌 내의 상응하는 세그먼트들에 상동성이어서, 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물의 통합에 의해 변형된다. 이 방법은 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 추가로 포함한다.

[0046] 일부 이러한 방법에서, 제1 핵 삽입물과 제1의 3' 상동성 아암 그리고 제2 핵산 삽입물과 제2의 5' 상동성 아암은 연속 핵산의 중복 단편들이고, 제2 핵산 삽입물과 제2의 3' 상동성 아암 그리고 제3 핵산 삽입물과 제3의 5' 상동성 아암은 연속 핵산의 중복 단편들이며, 그 연속 핵산은 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물의 통합에 의해 재형성된다.

[0047] 일 실시 형태에서, 세포는 만능성 세포이다. 다른 실시 형태에서, 만능성 세포는 배아 줄기 (ES) 세포이다. 일부 실시 형태에서, 만능성 세포는 조혈 줄기 세포 또는 신경 줄기 세포이다. 다른 실시 형태에서, 세포는 유도 만능성 줄기 (iPS) 세포이다.

[0048] 일 실시 형태에서, 표적 게놈 유전자좌는 세포의 게놈에 존재한다. 다른 실시 형태에서, 표적 게놈 유전자좌는 세포 내의 염색체의 DNA 상에 있다.

[0049] 일 실시 형태에서, 세포는 섬유아세포이다.

- [0050] 일부 방법에서, 세포는 비인간 세포이다. 다른 방법에서, 세포는 인간으로부터 유래된다. 일부 실시 형태에서, 세포는 포유류 세포이다. 다른 실시 형태에서, 포유류 세포는 설치류로부터 유래된다. 일부의 경우에, 설치류는 마우스, 래트 또는 햄스터이다.
- [0051] 상기 방법들 중 일부에서, 뉴클레아제 작용제는 뉴클레아제를 인코딩하는 핵산 서열을 포함하는 발현 작제물로부터 발현되고, 핵산은 세포에서 활성인 프로모터에 작동가능하게 연결되어 있다. 다른 방법에서, 뉴클레아제 작용제는 뉴클레아제를 인코딩하는 mRNA로부터 발현된다. 일부 이러한 방법에서, 뉴클레아제는 징크 핑거 뉴클레아제 (ZFN)이다. 다른 이러한 방법에서, 뉴클레아제는 전사 활성화인자 유사 이펙터 뉴클레아제 (TALEN)이다. 또 다른 방법에서, 뉴클레아제는 메가뉴클레아제이다.
- [0052] 상기 방법들 중 일부에서, 뉴클레아제 작용제는 CRISPR 관련 (Cas) 단백질 및 가이드 RNA (gRNA)를 포함한다. 일부 이러한 방법에서, Cas 단백질은 Cas9이다.
- [0053] 상기 방법들 중 일부에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 세포의 종과 상이한 종으로부터 유래된다. 일 실시 형태에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및/또는 제3 핵산 삽입물은 상이한 종으로부터 유래된다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물 중 하나 이상은 세포의 종과 상이한 종으로부터 유래된다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 인간 핵산이다. 다른 실시 형태에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및/또는 제3 핵산 삽입물은 인간 핵산이다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물 중 하나 이상은 인간 핵산이다.
- [0054] 일 실시 형태에서, 제1 핵산 삽입물과 제2 핵산 삽입물을 합한 크기는 약 50 kb 내지 약 500 kb, 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, 약 275 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 약 350 kb 내지 약 400 kb, 약 400 kb 내지 약 450 kb, 또는 약 450 kb 내지 약 500 kb이다. 다른 실시 형태에서, 제1 핵산 삽입물과 제2 핵산 삽입물을 합한 크기는 약 100 kb 내지 약 500 kb이다. 또 다른 실시 형태에서, 제1 핵산 삽입물과 제2 핵산 삽입물을 합한 크기는 약 300 kb이다.
- [0055] 일부 실시 형태에서, 표적화된 세포는 약 5 kb 내지 약 500 kb 범위의 제1 핵산 삽입물과 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다.
- [0056] 일 실시 형태에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 합한 크기는 약 50 kb 내지 약 700 kb, 약 50 kb 내지 약 500 kb, 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, 약 275 kb 내지 약 300 kb, 약 300 kb 내지 약 350 kb, 약 350 kb 내지 약 400 kb, 약 400 kb 내지 약 450 kb, 약 450 kb 내지 약 500 kb, 약 500 kb 내지 약 550 kb, 약 550 kb 내지 약 600 kb, 약 600 kb 내지 약 650 kb, 또는 약 650 kb 내지 약 700 kb이다.
- [0057] 일부 실시 형태에서, 표적화된 세포는 약 5 kb 내지 약 700 kb 범위의 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다. 선택적으로, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 합한 크기는 약 100 kb 내지 약 700 kb이다. 일부 실시 형태에서, 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 합한 크기는 약 400 kb이다.
- [0058] 상기 방법들 중 일부에서, 제1의 LTVEC의 제1 중복 서열은 제2의 LTVEC의 제1 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하다. 상기 방법들 중 일부에서, 제2의 LTVEC의 제2 중복 서열은 제3의 LTVEC의 제2 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하다. 상기 방법들 중 어느 것에서, 중복 서열은 약 1 kb 내지 약 70 kb이다. 특정 실시 형태에서, 중복 서열은 적어도 10 kb이다. 다른 특정 실시 형태에서, 중복 서열은 적어도 20 kb이다. 상기 방법들 중 일부에서, 제1 중복 서열 및/또는 제2 중복 서열은 약 1 kb 내지 약 70 kb이다. 일부 방법에서, 제1 중복 서열 및/또는 제2 중복 서열은 적어도 10 kb 또는 적어도 20 kb이다.
- [0059] 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합. 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및

제3 핵산 삽입물 중 하나 이상의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합.

- [0060] 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두의 통합은 표적 게놈 유전자좌에 외인성 서열의 첨가를 가져온다. 일 실시 형태에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및/또는 제3 핵산 삽입물의 통합은 표적 게놈 유전자좌에 외인성 서열의 첨가를 가져온다.
- [0061] 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두의 통합은 표적 게놈 유전자좌에서 내인성 서열의 결실을 가져온다. 다른 실시 형태에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및/또는 제3 핵산 삽입물의 통합은 표적 게놈 유전자좌에서 내인성 서열의 결실을 가져온다. 일부 이러한 방법에서, 표적 게놈 유전자좌에서 내인성 서열의 결실은 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 600 kb, 약 600 kb 내지 약 700 kb, 또는 약 700 kb 내지 약 800 kb이다.
- [0062] 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두의 통합은 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합을 가져온다. 또 다른 실시 형태에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및/또는 제3 핵산 삽입물의 통합은 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합을 가져온다.
- [0063] 상기 방법들 중 일부에서, 제1 LTVEC와 제2 LTVEC의 병용은 단일 LTVEC의 사용과 비교하여 표적화 효율의 증가를 가져온다. 선택적으로, 표적화 효율의 증가는 적어도 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배이다.
- [0064] 일부 방법에서, 제1의 LTVEC 또는 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이고, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이다. 일부 실시 형태에서, 제1의 LTVEC, 제2의 LTVEC 또는 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이고, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이고, 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 10 kb 내지 약 150 kb이다. 다른 실시 형태에서, 제1의 LTVEC, 제2의 LTVEC 또는 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 1 kb 내지 약 5 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 120 kb, 또는 약 120 kb 내지 150 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 1 kb 내지 약 5 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 120 kb, 또는 약 120 kb 내지 150 kb이고; 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 1 kb 내지 약 5 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 120 kb, 또는 약 120 kb 내지 150 kb이고; 제3의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 약 1 kb 내지 약 5 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 120 kb, 또는 약 120 kb 내지 150 kb이다.
- [0065] F0 세대 비인간 동물을 생산하는 방법이 추가로 제공된다. 이러한 방법은, 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계 - 비인간 ES 세포는 상기 방법들 중 어느 것에 의해 생성되었음 - 및 비인간 숙주 배아를 대리모에 임신시켜 대리모가 변형을 포함하는 F0 세대 비인간 동물을 생산하게 하는 단계를 포함한다. 선택적으로, 비인간 동물은 마우스 또는 래트이다.
- [0066] 또한, 본 발명은, 제1 핵산과 제2 핵산을 세포 내로 도입시키는 단계 - 제1 핵산과 제2 핵산은 중복 뉴클레오타이드 서열을 포함함 - 를 포함하는, 세포의 표적 게놈 유전자좌에서 상동 재조합을 향상시키는 방법을 제공한다. 일부 이러한 방법에서, 상동 재조합은 단일 핵산만이 세포 내로 도입되는 방법과 비교하여 향상된다. 일부 이러한 방법에서, 상동 재조합은 뉴클레아제 작용제를 사용하지 않으면서 표적 게놈 유전자좌에서 향상된다. 다

른 이러한 방법은 표적 게놈 유전자좌에 또는 이 부근에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계를 추가로 포함한다.

- [0067] 일 태양에서, 뉴클레아제 작용제를 사용하지 않으면서 세포의 게놈 유전자좌에서 상동 재조합을 향상시키는 방법이 제공되며, 이 방법은 제1 핵산과 제2 핵산을 세포 내로 도입시키는 단계 - 제1 핵산과 제2 핵산은 중복 뉴클레오티드 서열을 포함함 - 를 포함한다.
- [0068] 일 실시 형태에서, 방법은 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합을 향상시킨다. 일부 이러한 방법은 제1 핵산이 제2 핵산 없이 도입되는 방법과 비교하여 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합을 향상시킨다. 일 실시 형태에서, 방법은 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합을 향상시킨다. 일부 이러한 방법은 제2 핵산이 제1 핵산 없이 도입되는 방법과 비교하여 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합을 향상시킨다. 일 실시 형태에서, 방법은 표적 게놈 유전자좌에서 제1 핵산과 제2 핵산의 상동 재조합을 증가시킨다.
- [0069] 일 실시 형태에서, 상동 재조합의 향상은 적어도 0.5배, 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배이다.
- [0070] 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열에 상동성이다. 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하다. 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 그 중복 서열과 100% 동일하다.
- [0071] 일 실시 형태에서, 중복 서열은 약 1 kb 내지 약 70 kb이다. 일부 방법에서, 중복 서열은 적어도 20 kb이다. 일 실시 형태에서, 중복 서열은 약 1 kb 내지 약 5 kb이다. 일 실시 형태에서, 중복 서열은 약 5 kb 내지 약 10 kb이다. 일 실시 형태에서, 중복 서열은 약 10 kb 내지 약 15 kb이다. 일 실시 형태에서, 중복 서열은 약 15 kb 내지 약 20 kb이다. 일 실시 형태에서, 중복 서열은 약 20 kb 내지 약 25 kb이다. 일 실시 형태에서, 중복 서열은 약 25 kb 내지 약 30 kb이다. 일 실시 형태에서, 중복 서열은 약 30 kb 내지 약 35 kb이다. 일 실시 형태에서, 중복 서열은 약 35 kb 내지 약 40 kb이다. 일 실시 형태에서, 중복 서열은 약 40 kb 내지 약 45 kb이다. 일 실시 형태에서, 중복 서열은 약 45 kb 내지 약 50 kb이다. 일 실시 형태에서, 중복 서열은 약 50 kb 내지 약 55 kb이다. 일 실시 형태에서, 중복 서열은 약 55 kb 내지 약 60 kb이다. 일 실시 형태에서, 중복 서열은 약 60 kb 내지 약 65 kb이다. 일 실시 형태에서, 중복 서열은 약 65 kb 내지 약 70 kb이다.
- [0072] 일 실시 형태에서, 중복 서열은 적어도 5 kb이다. 일 실시 형태에서, 중복 서열은 적어도 10 kb이다. 일 실시 형태에서, 중복 서열은 적어도 15 kb이다. 일 실시 형태에서, 중복 서열은 적어도 20 kb이다. 일 실시 형태에서, 중복 서열은 적어도 25 kb이다. 일 실시 형태에서, 중복 서열은 적어도 30 kb이다. 일 실시 형태에서, 중복 서열은 적어도 35 kb이다. 일 실시 형태에서, 중복 서열은 적어도 40 kb이다. 일 실시 형태에서, 중복 서열은 적어도 45 kb이다. 일 실시 형태에서, 중복 서열은 적어도 50 kb이다. 일 실시 형태에서, 중복 서열은 적어도 55 kb이다. 일 실시 형태에서, 중복 서열은 적어도 60 kb이다. 일 실시 형태에서, 중복 서열은 적어도 65 kb이다. 일 실시 형태에서, 중복 서열은 적어도 70 kb이다.
- [0073] 일 실시 형태에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 표적화 벡터이고, 제2 핵산은 중복 서열을 제외하고는 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다.
- [0074] 일 실시 형태에서, 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터이고, 제1 핵산은 중복 서열을 제외하고는 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다.
- [0075] 일 실시 형태에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1 표적화 벡터이고, 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터이다. 일 실시 형태에서, 제1 핵산 삽입물과 제2 핵산 삽입물은 연속 핵산의 중복 단편들이다.
- [0076] 일 실시 형태에서, 표적화 벡터는 약 1 kb 내지 약 2 kb이다. 일 실시 형태에서, 표적화 벡터는 약 2 kb 내지 약 5 kb이다. 일 실시 형태에서, 표적화 벡터는 약 5 kb 내지 약 10 kb이다.
- [0077] 일 실시 형태에서, 표적화 벡터는 큰 표적화 벡터 (LTVEC)이다. 일부 방법에서, 표적화 벡터는 약 20 kb 내지 약 200 kb 범위의 LTVEC이다. 일부 방법에서, 제1 표적화 벡터는 약 20 kb 내지 약 200 kb 범위의 제1의

LTVEC이코/이거나 제2 표적화 벡터는 약 20 kb 내지 약 200 kb 범위의 제2의 LTVEC이다. 일 실시 형태에서, LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 일부 방법에서, 제1 표적화 벡터는 제1의 LTVEC이코/이거나 제2 표적화 벡터는 제2의 LTVEC이다. 일부 방법에서, 제1의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 일부 방법에서, 제2의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이고, 제2의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다.

- [0078] 일 실시 형태에서, LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이고, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다.
- [0079] 일 실시 형태에서, 중복 서열은 제1 핵산의 3' 말단 및 제2 핵산 서열의 5' 말단에 위치한다. 일 실시 형태에서, 중복 서열은 제1 핵산 서열의 5' 말단 및 제2 핵산 서열의 3' 말단에 위치한다.
- [0080] 일 실시 형태에서, 제1 핵산 삽입물 및/또는 제2 핵산 삽입물은 상이한 종으로부터 유래된다. 다른 실시 형태에서, 제1 핵산 삽입물 및/또는 제2 핵산 삽입물은 인간 핵산이다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 세포의 종과 상이한 종으로부터 유래된다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 인간 핵산이다.
- [0081] 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합. 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물 중 하나 이상의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합.
- [0082] 일 실시 형태에서, 게놈 유전자좌 내로의 제1 삽입물 및/또는 제2 삽입물의 통합은 게놈 유전자좌에 외인성 서열의 첨가를 가져온다.
- [0083] 일부 실시 형태에서, 표적화된 세포는 약 5 kb 내지 약 500 kb 범위의 제1 핵산 삽입물과 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다. 일부 방법에서, 표적화된 세포는, 합한 크기가 약 5 kb 내지 약 500 kb의 범위인 제1 핵산 삽입물과 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다.
- [0084] 다른 실시 형태에서, 게놈 유전자좌 내로의 제1 핵산 삽입물 및/또는 제2 핵산 삽입물의 통합은 표적 게놈 유전자좌에서 내인성 서열의 결실을 가져온다. 일 실시 형태에서, 표적 게놈 유전자좌에서 내인성 서열의 결실은 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 600 kb, 약 600 kb 내지 약 700 kb, 또는 약 700 kb 내지 약 800 kb이다.
- [0085] 일부 방법에서, 세포는 인간 세포이다. 다른 방법에서, 세포는 비인간 세포이다. 일부 방법에서, 세포는 만능성 세포, 조혈 줄기 세포, 신경 줄기 세포 또는 섬유아세포이다. 선택적으로, 만능성 세포는 배아 줄기 (ES) 세포 또는 유도 만능성 줄기 (iPS) 세포이다. 일부 방법에서, 세포는 포유류 세포이다. 선택적으로, 포유류 세포는 설치류 세포이다. 선택적으로, 설치류 세포는 마우스 세포 또는 래트 세포이다.
- [0086] 일 실시 형태에서, 세포는 만능성 세포이다. 다른 실시 형태에서, 만능성 세포는 배아 줄기 (ES) 세포이다. 일부 실시 형태에서, 만능성 세포는 조혈 줄기 세포 또는 신경 줄기 세포이다. 다른 실시 형태에서, 세포는 유

도 만능성 줄기 (iPS) 세포이다.

- [0087] 일 실시 형태에서, 표적 게놈 유전자좌는 세포의 게놈에 존재한다. 다른 실시 형태에서, 표적 게놈 유전자좌는 세포 내의 염색체의 DNA 상에 있다.
- [0088] 일 실시 형태에서, 세포는 섬유아세포이다.
- [0089] 일부 방법에서, 세포는 비인간 세포이다. 다른 방법에서, 세포는 인간으로부터 유래된다. 일부 실시 형태에서, 세포는 포유류 세포이다. 다른 실시 형태에서, 포유류 세포는 설치류로부터 유래된다. 일부의 경우에, 설치류는 마우스, 래트 또는 햄스터이다.
- [0090] 상기 방법들 중 일부에서, 중복 뉴클레오티드 서열은 표적 게놈 유전자좌로의 재조합 기구의 보충을 용이하게 한다.
- [0091] 또한, 본 발명은 (a) 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계 - 비인간 ES 세포는 상기 방법들에 의해 생성되었음 -; 및 (b) 비인간 숙주 배아를 대리모에 임신시키는 단계 - 대리모는 변형을 포함하는 F0 세대 비인간 동물을 생산함 - 를 포함하는, F0 세대 비인간 동물을 생산하는 방법을 제공한다. 선택적으로, 비인간 동물은 마우스 또는 래트이다.
- [0092] 다른 태양에서, 뉴클레아제 작용제를 사용하여 세포의 표적 게놈 유전자좌에서 상동 재조합을 향상시키는 방법이 제공되며, 이 방법은 (i) 중복 뉴클레오티드 서열을 포함하는, 제1 핵산과 제2 핵산; 및 (ii) 게놈 유전자좌에 또는 이 부근에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계를 포함한다.
- [0093] 일 실시 형태에서, 방법은 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합을 향상시킨다. 일부 이러한 방법은 제1 핵산이 제2 핵산 없이 도입되는 방법과 비교하여 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합을 향상시킨다. 일 실시 형태에서, 방법은 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합을 향상시킨다. 일부 이러한 방법은 제2 핵산이 제1 핵산 없이 도입되는 방법과 비교하여 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합을 향상시킨다. 일 실시 형태에서, 방법은 표적 게놈 유전자좌에서 제1 핵산과 제2 핵산의 상동 재조합을 증가시킨다. 일 실시 형태에서, 상동 재조합의 향상은 적어도 0.5배, 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배이다.
- [0094] 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열에 상동성이다. 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일하다. 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열과 100% 동일하다.
- [0095] 일 실시 형태에서, 중복 서열은 약 1 kb 내지 약 70 kb이다. 일 실시 형태에서, 중복 서열은 약 1 kb 내지 약 5 kb이다. 일 실시 형태에서, 중복 서열은 약 5 kb 내지 약 10 kb이다. 일 실시 형태에서, 중복 서열은 약 10 kb 내지 약 15 kb이다. 일 실시 형태에서, 중복 서열은 약 15 kb 내지 약 20 kb이다. 일 실시 형태에서, 중복 서열은 약 20 kb 내지 약 25 kb이다. 일 실시 형태에서, 중복 서열은 약 25 kb 내지 약 30 kb이다. 일 실시 형태에서, 중복 서열은 약 30 kb 내지 약 35 kb이다. 일 실시 형태에서, 중복 서열은 약 35 kb 내지 약 40 kb이다. 일 실시 형태에서, 중복 서열은 약 40 kb 내지 약 45 kb이다. 일 실시 형태에서, 중복 서열은 약 45 kb 내지 약 50 kb이다. 일 실시 형태에서, 중복 서열은 약 50 kb 내지 약 55 kb이다. 일 실시 형태에서, 중복 서열은 약 55 kb 내지 약 60 kb이다. 일 실시 형태에서, 중복 서열은 약 60 kb 내지 약 65 kb이다. 일 실시 형태에서, 중복 서열은 약 65 kb 내지 약 70 kb이다.
- [0096] 일 실시 형태에서, 중복 서열은 적어도 5 kb이다. 일 실시 형태에서, 중복 서열은 적어도 10 kb이다. 일 실시 형태에서, 중복 서열은 적어도 15 kb이다. 일 실시 형태에서, 중복 서열은 적어도 20 kb이다. 일 실시 형태에서, 중복 서열은 적어도 25 kb이다. 일 실시 형태에서, 중복 서열은 적어도 30 kb이다. 일 실시 형태에서, 중복 서열은 적어도 35 kb이다. 일 실시 형태에서, 중복 서열은 적어도 40 kb이다. 일 실시 형태에서, 중복 서열은 적어도 45 kb이다. 일 실시 형태에서, 중복 서열은 적어도 50 kb이다. 일 실시 형태에서, 중복 서열은 적어도 55 kb이다. 일 실시 형태에서, 중복 서열은 적어도 60 kb이다. 일 실시 형태에서, 중복 서열은 적어도 65 kb이다. 일 실시 형태에서, 중복 서열은 적어도 70 kb이다.
- [0097] 일 실시 형태에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 표적화 벡터이고, 제2 핵산은 중복 서열을 제외하고는 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다.

- [0098] 일 실시 형태에서, 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터이고, 제1 핵산은 중복 서열을 제외하고는 표적 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다.
- [0099] 일 실시 형태에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1 표적화 벡터이고, 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터이다. 일 실시 형태에서, 제1 핵산 삽입물과 제2 핵산 삽입물은 연속 핵산의 중복 단편들이다.
- [0100] 일 실시 형태에서, 표적화 벡터는 약 1 kb 내지 약 2 kb이다. 일 실시 형태에서, 표적화 벡터는 약 2 kb 내지 약 5 kb이다. 일 실시 형태에서, 표적화 벡터는 약 5 kb 내지 약 10 kb이다.
- [0101] 일 실시 형태에서, 표적화 벡터는 큰 표적화 벡터 (LTVEC)이다. 일부 방법에서, 표적화 벡터는 약 10 kb 내지 약 200 kb 범위의 큰 표적화 벡터이다. 일부 방법에서, 제1 표적화 벡터는 제1의 LTVEC이고/이거나 제2 표적화 벡터는 제2의 LTVEC이다. 일부 방법에서, 제1 표적화 벡터는 약 10 kb 내지 약 200 kb 범위의 제1의 큰 표적화 벡터이고/이거나 제2 표적화 벡터는 약 10 kb 내지 약 200 kb 범위의 제2의 큰 표적화 벡터이다. 일 실시 형태에서, LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 선택적으로, 제1의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 선택적으로, 제2의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이고, 제2의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다.
- [0102] 일 실시 형태에서, LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이고, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다.
- [0103] 일 실시 형태에서, 중복 서열은 제1 핵산의 3' 말단 및 제2 핵산 서열의 5' 말단에 위치한다. 일 실시 형태에서, 중복 서열은 제1 핵산 서열의 5' 말단 및 제2 핵산 서열의 3' 말단에 위치한다.
- [0104] 일 실시 형태에서, 뉴클레아제 작용제는 뉴클레아제를 인코딩하는 핵산 서열을 포함하는 발현 작제물로부터 발현되고, 핵산은 세포에서 활성인 프로모터에 작동가능하게 연결되어 있다. 일 실시 형태에서, 뉴클레아제 작용제는 뉴클레아제를 인코딩하는 mRNA로부터 발현된다. 일 실시 형태에서, 뉴클레아제는 징크 핑거 뉴클레아제 (ZFN)이다. 일 실시 형태에서, 뉴클레아제는 전사 활성화인자 유사 이펙터 뉴클레아제 (TALEN)이다. 일 실시 형태에서, 뉴클레아제는 메가뉴클레아제이다.
- [0105] 일 실시 형태에서, 뉴클레아제 작용제는 CRISPR 관련 (Cas) 단백질 및 가이드 RNA (gRNA)를 포함한다. 일 실시 형태에서, Cas 단백질은 Cas9이다.
- [0106] 일 실시 형태에서, 제1 핵산 삽입물 및/또는 제2 핵산 삽입물은 상이한 종으로부터 유래된다. 다른 실시 형태에서, 제1 핵산 삽입물 및/또는 제2 핵산 삽입물은 인간 핵산이다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 세포의 종과 상이한 종으로부터 유래된다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 인간 핵산이다.
- [0107] 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합. 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물 중 하나 이상의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열

의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합.

- [0108] 일 실시 형태에서, 게놈 유전자좌 내로의 제1 삽입물 및/또는 제2 삽입물의 통합은 게놈 유전자좌에 외인성 서열의 첨가를 가져온다.
- [0109] 일부 실시 형태에서, 표적화된 세포는 약 5 kb 내지 약 500 kb 범위의 제1 핵산 삽입물과 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다. 일부 방법에서, 표적화된 세포는, 합한 크기가 약 5 kb 내지 약 500 kb의 범위인 제1 핵산 삽입물과 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다.
- [0110] 다른 실시 형태에서, 게놈 유전자좌 내로의 제1 핵산 삽입물 및/또는 제2 핵산 삽입물의 통합은 표적 게놈 유전자좌에서 내인성 서열의 결실을 가져온다. 일 실시 형태에서, 표적 게놈 유전자좌에서 내인성 서열의 결실은 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 600 kb, 약 600 kb 내지 약 700 kb, 또는 약 700 kb 내지 약 800 kb이다.
- [0111] 일부 방법에서, 세포는 인간 세포이다. 다른 방법에서, 세포는 비인간 세포이다. 일부 방법에서, 세포는 만능성 세포, 조혈 줄기 세포, 신경 줄기 세포 또는 섬유아세포이다. 선택적으로, 만능성 세포는 배아 줄기 (ES) 세포 또는 유도 만능성 줄기 (iPS) 세포이다. 일부 방법에서, 세포는 포유류 세포이다. 선택적으로, 포유류 세포는 설치류 세포이다. 선택적으로, 설치류 세포는 마우스 세포 또는 래트 세포이다.
- [0112] 일 실시 형태에서, 세포는 만능성 세포이다. 다른 실시 형태에서, 만능성 세포는 배아 줄기 (ES) 세포이다. 일부 실시 형태에서, 만능성 세포는 조혈 줄기 세포 또는 신경 줄기 세포이다. 다른 실시 형태에서, 세포는 유도 만능성 줄기 (iPS) 세포이다.
- [0113] 일 실시 형태에서, 표적 게놈 유전자좌는 세포의 게놈에 존재한다. 다른 실시 형태에서, 표적 게놈 유전자좌는 세포 내의 염색체외 DNA 상에 있다.
- [0114] 일 실시 형태에서, 세포는 섬유아세포이다.
- [0115] 일부 방법에서, 세포는 비인간 세포이다. 다른 방법에서, 세포는 인간으로부터 유래된다. 일부 실시 형태에서, 세포는 포유류 세포이다. 다른 실시 형태에서, 포유류 세포는 설치류로부터 유래된다. 일부의 경우에, 설치류는 마우스, 래트 또는 햄스터이다.
- [0116] 상기 방법들 중 일부에서, 중복 뉴클레오티드 서열은 표적 게놈 유전자좌로의 재조합 기구의 보충을 용이하게 한다.
- [0117] 또한, 본 발명은 (a) 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계 - 비인간 ES 세포는 상기 방법들에 의해 생성되었음 -; 및 (b) 비인간 숙주 배아를 대리모에 임신시키는 단계 - 대리모는 변형을 포함하는 F0 세대 비인간 동물을 생산함 - 를 포함하는, F0 세대 비인간 동물을 생산하는 방법을 제공한다. 선택적으로, 비인간 동물은 마우스 또는 래트이다.
- [0118] 다른 태양에서, 재조합 기구를 표적화 벡터 상에 로딩(loading)함으로써 세포의 표적 게놈 유전자좌에서 상동 재조합을 향상시키는 방법이 제공되며, 이 방법은 제1 핵산과 제2 핵산을 세포 내로 도입시키는 단계 - 제1 핵산과 제2 핵산은 중복 뉴클레오티드 서열을 포함하며, 중복 뉴클레오티드 서열은 표적 게놈 유전자좌로의 재조합 기구의 보충을 용이하게 함 - 를 포함한다.
- [0119] 일 실시 형태에서, 방법은 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합을 향상시킨다. 일부 이러한 방법은 제1 핵산이 제2 핵산 없이 도입되는 방법과 비교하여 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합을 향상시킨다. 일 실시 형태에서, 방법은 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합을 향상시킨다. 일부 이러한 방법은 제2 핵산이 제1 핵산 없이 도입되는 방법과 비교하여 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합을 향상시킨다. 일 실시 형태에서, 상동 재조합의 향상은 적어도 0.5배, 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배이다.
- [0120] 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열에 상동성이다. 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또

는 99.9% 동일하다. 일 실시 형태에서, 제1 핵산의 중복 서열은 제2 핵산의 그 중복 서열과 100% 동일하다.

- [0121] 일 실시 형태에서, 중복 서열은 약 1 kb 내지 약 70 kb이다. 일 실시 형태에서, 중복 서열은 약 1 kb 내지 약 5 kb이다. 일 실시 형태에서, 중복 서열은 약 5 kb 내지 약 10 kb이다. 일 실시 형태에서, 중복 서열은 약 10 kb 내지 약 15 kb이다. 일 실시 형태에서, 중복 서열은 약 15 kb 내지 약 20 kb이다. 일 실시 형태에서, 중복 서열은 약 20 kb 내지 약 25 kb이다. 일 실시 형태에서, 중복 서열은 약 25 kb 내지 약 30 kb이다. 일 실시 형태에서, 중복 서열은 약 30 kb 내지 약 35 kb이다. 일 실시 형태에서, 중복 서열은 약 35 kb 내지 약 40 kb이다. 일 실시 형태에서, 중복 서열은 약 40 kb 내지 약 45 kb이다. 일 실시 형태에서, 중복 서열은 약 45 kb 내지 약 50 kb이다. 일 실시 형태에서, 중복 서열은 약 50 kb 내지 약 55 kb이다. 일 실시 형태에서, 중복 서열은 약 55 kb 내지 약 60 kb이다. 일 실시 형태에서, 중복 서열은 약 60 kb 내지 약 65 kb이다. 일 실시 형태에서, 중복 서열은 약 65 kb 내지 약 70 kb이다.
- [0122] 일 실시 형태에서, 중복 서열은 적어도 5 kb이다. 일 실시 형태에서, 중복 서열은 적어도 10 kb이다. 일 실시 형태에서, 중복 서열은 적어도 15 kb이다. 일 실시 형태에서, 중복 서열은 적어도 20 kb이다. 일 실시 형태에서, 중복 서열은 적어도 25 kb이다. 일 실시 형태에서, 중복 서열은 적어도 30 kb이다. 일 실시 형태에서, 중복 서열은 적어도 35 kb이다. 일 실시 형태에서, 중복 서열은 적어도 40 kb이다. 일 실시 형태에서, 중복 서열은 적어도 45 kb이다. 일 실시 형태에서, 중복 서열은 적어도 50 kb이다. 일 실시 형태에서, 중복 서열은 적어도 55 kb이다. 일 실시 형태에서, 중복 서열은 적어도 60 kb이다. 일 실시 형태에서, 중복 서열은 적어도 65 kb이다. 일 실시 형태에서, 중복 서열은 적어도 70 kb이다.
- [0123] 일 실시 형태에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 표적화 벡터이고, 제2 핵산은 중복 서열을 제외하고는 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다.
- [0124] 일 실시 형태에서, 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터이고, 제1 핵산은 중복 서열을 제외하고는 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다.
- [0125] 일 실시 형태에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1 표적화 벡터이고, 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터이다. 일 실시 형태에서, 제1 핵산 삽입물과 제2 핵산 삽입물은 연속 핵산의 중복 단편들이다.
- [0126] 일 실시 형태에서, 표적화 벡터는 약 1 kb 내지 약 2 kb이다. 일 실시 형태에서, 표적화 벡터는 약 2 kb 내지 약 5 kb이다. 일 실시 형태에서, 표적화 벡터는 약 5 kb 내지 약 10 kb이다.
- [0127] 일 실시 형태에서, 표적화 벡터는 큰 표적화 벡터 (LTVEC)이다. 일부 방법에서, 표적화 벡터는 약 10 kb 내지 약 200 kb 범위의 LTVEC이다. 일부 방법에서, 제1 표적화 벡터는 제1의 LTVEC이고/이거나 제2 표적화 벡터는 제2의 LTVEC이다. 일부 방법에서, 제1 표적화 벡터는 약 10 kb 내지 약 200 kb 범위의 제1의 LTVEC이고/이거나 제2 표적화 벡터는 약 10 kb 내지 약 200 kb 범위의 제2의 LTVEC이다. 일 실시 형태에서, LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 선택적으로, 제1의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 선택적으로, 제2의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이고, 제2의 LTVEC는 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 120 kb, 약 120 kb 내지 약 150 kb, 또는 약 150 kb 내지 약 200 kb이다.
- [0128] 일 실시 형태에서, LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제2의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다. 일부 방법에서, 제1의 LTVEC의 5' 상동성 아암과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이고, 제2의 LTVEC의 5' 상동성 아암

과 3' 상동성 아암의 총합은 10 kb 내지 약 200 kb이다.

- [0129] 일 실시 형태에서, 중복 서열은 제1 핵산의 3' 말단 및 제2 핵산 서열의 5' 말단에 위치한다. 일 실시 형태에서, 중복 서열은 제1 핵산 서열의 5' 말단 및 제2 핵산 서열의 3' 말단에 위치한다.
- [0130] 일 실시 형태에서, 제1 핵산 삽입물 및/또는 제2 핵산 삽입물은 상이한 종으로부터 유래된다. 다른 실시 형태에서, 제1 핵산 삽입물 및/또는 제2 핵산 삽입물은 인간 핵산이다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 세포의 종과 상이한 종으로부터 유래된다. 일부 방법에서, 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두는 인간 핵산이다.
- [0131] 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 또는 둘 모두의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합. 일부 방법에서, 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물 중 하나 이상의 통합은 하기 중 하나 이상을 가져온다: (a) 표적 게놈 유전자좌에 외인성 서열의 첨가; (b) 표적 게놈 유전자좌에서 내인성 서열의 결실; 또는 (c) 녹인, 녹아웃, 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합.
- [0132] 일 실시 형태에서, 게놈 유전자좌 내로의 제1 삽입물 및/또는 제2 삽입물의 통합은 게놈 유전자좌에 외인성 서열의 첨가를 가져온다.
- [0133] 일부 실시 형태에서, 표적화된 세포는 약 5 kb 내지 약 500 kb 범위의 제1 핵산 삽입물과 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다. 일부 방법에서, 표적화된 세포는, 합한 크기가 약 5 kb 내지 약 500 kb의 범위인 제1 핵산 삽입물과 제2 핵산 삽입물을 함께 포함하는 게놈 DNA를 포함한다.
- [0134] 다른 실시 형태에서, 게놈 유전자좌 내로의 제1 핵산 삽입물 및/또는 제2 핵산 삽입물의 통합은 표적 게놈 유전자좌에서 내인성 서열의 결실을 가져온다. 일 실시 형태에서, 표적 게놈 유전자좌에서 내인성 서열의 결실은 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 600 kb, 약 600 kb 내지 약 700 kb, 또는 약 700 kb 내지 약 800 kb이다.
- [0135] 일부 방법에서, 세포는 인간 세포이다. 다른 방법에서, 세포는 비인간 세포이다. 일부 방법에서, 세포는 만능성 세포, 조혈 줄기 세포, 신경 줄기 세포 또는 섬유아세포이다. 선택적으로, 만능성 세포는 배아 줄기 (ES) 세포 또는 유도 만능성 줄기 (iPS) 세포이다. 일부 방법에서, 세포는 포유류 세포이다. 선택적으로, 포유류 세포는 설치류 세포이다. 선택적으로, 설치류 세포는 마우스 세포 또는 래트 세포이다.
- [0136] 일 실시 형태에서, 세포는 만능성 세포이다. 다른 실시 형태에서, 만능성 세포는 배아 줄기 (ES) 세포이다. 일부 실시 형태에서, 만능성 세포는 조혈 줄기 세포 또는 신경 줄기 세포이다. 다른 실시 형태에서, 세포는 유도 만능성 줄기 (iPS) 세포이다.
- [0137] 일 실시 형태에서, 표적 게놈 유전자좌는 세포의 게놈에 존재한다. 다른 실시 형태에서, 표적 게놈 유전자좌는 세포 내의 염색체의 DNA 상에 있다.
- [0138] 일 실시 형태에서, 세포는 섬유아세포이다.
- [0139] 일 실시 형태에서, 세포는 비인간 세포이다. 다른 실시 형태에서, 세포는 인간으로부터 유래된다. 다른 실시 형태에서, 세포는 포유류 세포이다. 다른 실시 형태에서, 포유류 세포는 설치류로부터 유래된다. 다른 실시 형태에서, 설치류는 마우스, 래트 또는 햄스터이다.
- [0140] 또한, 본 발명은 (a) 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계 - 비인간 ES 세포는 상기 방법들에 의해 생성되었음 -; 및 (b) 비인간 숙주 배아를 대리모에 임신시키는 단계 - 대리모는 변형을 포함하는 F0 세대 비인간 동물을 생산함 - 를 포함하는, F0 세대 비인간 동물을 생산하는 방법을 제공한다. 선택적으로, 비인간 동물은 마우스 또는 래트이다.

도면의 간단한 설명

- [0141] 도 1은 하이그로마이신 선택 카세트를 포함하는 마우스 염색체 14 상에 TCR 알파 유전자좌의 이형접합성 변형을 갖는 세포가 표적화되는 게놈 이종 표적화 사건에 대한 개략도를 제공한다. 하이그로마이신 선택 카세트는 정

크 핑거 뉴클레아제 (ZFN) 또는 CRISPR/Cas 복합체에 의해 절단되며, 네오마이신 선택 카세트 및 280 kb가 넘는 인간 면역글로불린 카과 사슬 가변 유전자 세그먼트를 포함하는 2개의 큰 표적화 벡터로 표적화된다. 큰 표적화 벡터는 각각 대략 20 kb의 중복 서열을 포함하며, 이는 큰 표적화 벡터들 사이의 상동 재조합을 가능하게 한다. 표적화 사건은 단일 표적화 단계로 둘 모두의 표적화 벡터로부터 인간 면역글로불린 카과 사슬 가변 유전자 세그먼트를 정확하게 삽입시켰다. 표적화 사건을 확인하는 데 사용된 다양한 프로브의 위치는 원으로 둘러싼 사각형으로 나타나 있다. 마우스 서열은 상향의 사선 해칭(diagonal hatching)으로 표시되어 있고, 인간 서열은 무해칭(no hatching)으로 표시되어 있고, 재조합 부위 및 선택 카세트는 하향의 파선형(dashed) 사선 해칭으로 표시되어 있다. 개략도는 일정한 비율에 따른 것이 아니며, 예를 들어 실제의 가변 유전자 세그먼트 수를 반영하지 않는다.

도 2는 하이그로마이신 선택 카세트를 포함하는 마우스 염색체 14 상에 TCR 알파 유전자좌의 이형접합성 변형을 갖는 세포가 네오마이신 선택 카세트와 120 kb의 인간 면역글로불린 카과 사슬 가변 유전자 세그먼트를 포함하는 하나의 큰 표적화 벡터로 표적화되는 단일 표적화 사건에 대한 개략도를 제공한다. 표적화 사건을 확인하는 데 사용된 다양한 프로브의 위치는 원으로 둘러싼 사각형으로 나타나 있다. 마우스 서열은 상향의 사선 해칭으로 표시되어 있고, 인간 서열은 무해칭으로 표시되어 있고, 재조합 부위 및 선택 카세트는 하향의 파선형 사선 해칭으로 표시되어 있다. 개략도는 일정한 비율에 따른 것이 아니며, 예를 들어 실제의 가변 유전자 세그먼트 수를 반영하지 않는다.

도 3은 CRISPR/Cas9 시스템을 사용한 하이그로마이신 선택 카세트의 표적화 및 파괴에 대한 개략도를 제공하며, 하이그로마이신 유전자의 다양한 서열을 표적화하는 다양한 가이드 RNA (gRNA)에 대한 CRISPR 인식 부위의 하이그로마이신 유전자 내의 위치를 예시한다. 개략도는 일정한 비율에 따른 것이 아니다.

도 4는 하이그로마이신 선택 카세트를 포함하는 마우스 염색체 14 상에 TCR 알파 유전자좌의 이형접합성 변형을 갖는 세포가 표적화되는 게놈 삼중 표적화 사건에 대한 개략도를 제공한다. 하이그로마이신 선택 카세트는 징크 핑거 뉴클레아제 (ZFN) 또는 CRISPR/Cas 복합체에 의해 절단되며, 네오마이신 선택 카세트 및 약 370 kb의 인간 면역글로불린 카과 사슬 가변 유전자 세그먼트를 포함하는 3개의 큰 표적화 벡터로 표적화된다. 큰 표적화 벡터는 각각 약 20 kb 내지 약 60 kb의 중복 서열을 포함하며, 이는 큰 표적화 벡터들 사이의 상동 재조합을 가능하게 한다. 표적화 사건은 단일 표적화 단계로 3개의 모든 표적화 벡터로부터 인간 면역글로불린 카과 사슬 가변 유전자 세그먼트를 정확하게 삽입시켰다. 표적화 사건을 확인하는 데 사용된 다양한 프로브의 위치는 원으로 둘러싼 사각형으로 나타나 있다. 마우스 서열은 상향의 사선 해칭으로 표시되어 있고, 인간 서열은 무해칭으로 표시되어 있고, 재조합 부위 및 선택 카세트는 하향의 파선형 사선 해칭으로 표시되어 있다. 개략도는 일정한 비율에 따른 것이 아니며, 예를 들어 실제의 가변 유전자 세그먼트 수를 반영하지 않는다.

정의

본 명세서에서 상호 교환적으로 사용되는 용어 "단백질", "폴리펩티드" 및 "펩티드"는 코딩된 및 코딩되지 않은 아미노산, 및 화학적으로 또는 생화학적으로 변형되거나 유도체화된 아미노산을 포함하는, 임의의 길이의 아미노산의 폴리머 형태를 포함한다. 상기 용어는 또한 변형된 폴리머, 예를 들어 변형된 펩티드 골격을 갖는 폴리펩티드를 포함한다.

본 명세서에서 상호 교환적으로 사용되는 용어 "핵산" 및 "폴리뉴클레오티드"는 리보뉴클레오티드, 데옥시리보뉴클레오티드, 또는 이들의 유사체 또는 변형된 버전(modified version)을 포함하는, 임의의 길이의 뉴클레오티드의 폴리머 형태를 포함한다. 이들은 단일 가닥, 이중 가닥, 및 다중 가닥 DNA 또는 RNA, 게놈 DNA, cDNA, DNA-RNA 하이브리드, 및 퓨린 염기, 피리미딘 염기, 또는 다른 천연, 화학적으로 변형된, 생화학적으로 변형된, 비천연 또는 유도체화된 뉴클레오티드 염기를 포함하는 폴리머를 포함한다.

표적 게놈 유전자좌는 표적화 벡터에 의한 표적화된 변형에 의해 변형될 게놈의 영역을 의미한다. 이 영역은 표적화 벡터 내의 상동성 아암들에 상응하는 게놈 DNA 세그먼트들의 외측 경계들 내의 영역으로 규정될 수 있다. 표적 게놈 유전자좌는 임의의 또는 모든 유전자 또는 유전자 군, 하나 이상의 인트론, 하나 이상의 엑손, 하나 이상의 조절 서열 등을 포함할 수 있다.

"코돈 최적화"는 대체로, 천연 아미노산 서열을 유지하면서 천연 서열의 적어도 하나의 코돈을 숙주 세포의 유전자에서 더욱 빈번하게 또는 가장 빈번하게 사용되는 코돈으로 대체함으로써 특정 숙주 세포에서의 발현 향상을 위해 핵산 서열을 변형시키는 과정을 포함한다. 예를 들어, Cas 단백질을 인코딩하는 핵산은 자연적으로 발생하는 핵산 서열과 비교하여, 세균 세포, 효모 세포, 인간 세포, 비인간 세포, 포유류 세포, 설치류 세포, 마

우스 세포, 래트 세포, 햄스터 세포, 또는 임의의 다른 숙주 세포를 포함하는 주어진 원핵 또는 진핵 세포에서 사용 빈도가 더 높은 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 방식으로 적합화될 수 있다. 문헌[Nakamura *et al.* (2000) *Nucleic Acids Research* 28:292]이 참조된다. 특정 숙주에서의 발현을 위해 특정 서열의 코돈 최적화를 위한 컴퓨터 알고리즘도 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 포함하여, 두 성분이 정상적으로 기능하게 하고 이들 성분 중의 적어도 하나가 적어도 하나의 다른 성분에 미치는 기능을 매개할 수 있는 가능성을 허용하도록 한다. 예를 들어, 프로모터는, 프로모터가 하나 이상의 전사 조절 인자의 유무에 대응하여 코딩 서열의 전사 수준을 제어하는 경우에, 코딩 서열에 작동가능하게 연결될 수 있다.

용어 "만능성 세포" 또는 "만능성 줄기 세포"는 1개 초과인 분화된 세포 유형으로 발달할 능력을 지니는 미분화 세포를 포함한다. 이러한 만능성 세포는, 예를 들어 포유류 배아 줄기 세포 (ES 세포) 또는 포유류 유도 만능성 줄기 세포 (iPS 세포)일 수 있다.

용어 "배아 줄기 세포" 또는 "ES 세포"는, 시험관내(*in vitro*)에서 미분화된 증식이 가능하고 배아 내로 도입 시에 발달 중인 배아의 어떠한 조직에도 기여할 수 있는 배아 유래 전능성 또는 만능성 세포를 포함한다.

용어 "유도 만능성 줄기 세포" 또는 "iPS 세포"는 분화된 성체 세포로부터 직접 유래될 수 있는 만능성 줄기 세포를 포함한다. 인간 iPS 세포는 특정 세트의 리프로그래밍 인자(reprogramming factor)를 비만능성 세포 내로 도입시켜 생성될 수 있으며, 이 리프로그래밍 인자는, 예를 들어 Oct3/4, Sox 패밀리의 전사 인자 (예컨대, Sox1, Sox2, Sox3, Sox15), Myc 패밀리의 전사 인자 (예컨대, c-Myc, l-Myc, n-Myc), Krüppel 유사 패밀리의 전사 인자 (예컨대, KLF1, KLF2, KLF4, KLF5), 및/또는 관련 전사 인자, 예를 들어 NANOG, LIN28, 및/또는 Glis1을 포함할 수 있다. 인간 iPS 세포는, 예를 들어 miRNA, 전사 인자의 작용을 모방하는 작은 분자, 또는 계열 지시자(lineage specifier)의 사용에 의해서도 생성될 수 있다. 인간 iPS 세포는 3가지의 척추동물 배엽층, 예를 들어 내배엽, 외배엽, 또는 중배엽의 어떠한 세포로도 분화하는 이의 능력을 특징으로 한다. 인간 iPS 세포는 또한 적합한 시험관내 배양 조건 하에서 무한히 증식하는 이의 능력을 특징으로 한다. 예를 들어, 문헌[Takahashi and Yamanaka (*Cell* (2006) Vol. 126(4), pp. 663-676)]이 참조된다.

면역글로불린 핵산 서열과 관련하여 용어 "생식세포계열(germline)"은 자손으로 전해질 수 있는 있는 핵산 서열을 포함한다.

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 그리고 C와 G이다. RNA에서, 이들은 전형적으로 C와 G 그리고 U와 A이다. 상보성은 완전하거나 상당하거나/충분할 수 있다. 2개의 핵산 사이의 완전한 상보성은 2개의 핵산이, 이중 나선 구조(duplex)의 모든 염기가 왓슨-크릭 염기 결합(Watson-Crick pairing)에 의해 상보적 염기에 결합되는 이중 나선 구조를 형성할 수 있음을 의미한다. "상당한" 또는 "충분한" 상보성은 한 가닥의 서열이 대향하는 가닥의 서열에 전적으로 및/또는 완전히 상보적이지는 않지만, 일련의 혼성화 조건(예를 들어, 염 농도 및 온도)에서 안정한 하이브리드 복합체를 형성하도록 두 가닥 상의 염기 사이에 충분한 결합이 일어남을 의미한다. 이러한 조건은 혼성화된 가닥의 T_m (용해 온도)을 예측하는 서열 및 표준 수학적 계산을 사용하여 예측될 수 있거나, 통상적인 방법을 사용한 T_m의 실험적 결정에 의해 예측될 수 있다. T_m은 2개의 핵산 가닥 사이에 형성된 혼성화 복합체의 집단이 50% 변성되는 (즉, 이중 가닥 핵산 분자의 집단의 절반이 단일 가닥으로 해리되는) 온도를 포함한다. T_m보다 낮은 온도에서는 혼성화 복합체의 형성이 촉진되는 반면에, T_m보다 높은 온도에서는 혼성화 복합체의 가닥들의 용해 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 핵산의 구조적 특징을 고려한다.

"혼성화 조건"은 하나의 핵산 가닥이 상보적 가닥 상호작용 및 수소 결합에 의해 제2 핵산 가닥에 결합하여 혼성화 복합체를 생성시키는 누적 환경을 포함한다. 이러한 조건은 핵산을 함유하는 수용액 또는 유기 용액의 화학 성분 및 이의 농도 (예를 들어, 염, 킬레이트제, 포름아미드), 및 혼합물의 온도를 포함한다. 다른 인자, 예컨대 인큐베이션 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)]이 참조된다.

혼성화는 2개의 핵산이 상보적 서열을 함유하는 것을 필요로 하지만, 염기 간의 미스매치(mismatch)가 가능하다. 2개의 핵산 간의 혼성화에 적절한 조건은 당업계에 잘 알려진 변수인 핵산 길이 및 상보성 정도에 좌우된다. 2개의 뉴클레오티드 서열 간의 상보성 정도가 클수록, 이들 서열을 갖는 핵산의 하이브리드에 대한 용해 온도(Tm)의 값이 커진다. 쇼트 스트레치(short stretch)의 상보성 (예를 들어, 35개 이하, 30개 이하, 25개 이하, 22개 이하, 20개 이하, 또는 18개 이하의 뉴클레오티드에 걸친 상보성)을 갖는 핵산 간의 혼성화의 경우, 미스매치 위치가 중요해진다 (상기 문헌[Sambrook *et al.*, 11.7-11.8] 참조). 전형적으로, 혼성화 가능한 핵산 길이는 적어도 약 10개의 뉴클레오티드이다. 혼성화 가능한 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. 또한, 온도 및 세척 용액 염 농도는, 필요에 따라, 상보성 영역의 길이 및 상보성 정도와 같은 인자에 따라 조정될 수 있다.

폴리뉴클레오티드의 서열은 특이적으로 혼성화 가능해지기 위해 이의 표적 핵산의 서열과 100% 상보적일 필요는 없다. 더욱이, 폴리뉴클레오티드는 개재 또는 인접 세그먼트가 혼성화 사건 (예를 들어, 루프 구조 또는 헤어핀(hairpin) 구조)에 관여하지 않도록 하나 이상의 세그먼트에 대하여 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, 20개 중 18개의 뉴클레오티드가 표적 영역에 상보적이므로, 특이적으로 혼성화되는 gRNA는 90% 상보성을 나타낼 것이다. 이러한 예에서, 나머지 비상보적 뉴클레오티드는 클러스터를 이루거나 그 사이에 상보적 뉴클레오티드가 배치될 수 있으며, 서로 인접하거나 상보적 뉴클레오티드에 연속적일 필요는 없다.

핵산 내의 핵산 서열의 특정 스트레치들 사이의 상보성 비율(%)은 당업계에 공지된 BLAST 프로그램 (기본 국소 정렬 검색 도구(basic local alignment search tool)) 및 PowerBLAST 프로그램 (문헌[Altschul *et al.* (1990) *J. Mol. Biol.* 215:403-410]; 문헌[Zhang and Madden (1997) *Genome Res.* 7:649-656])을 사용하거나 스미스(Smith)와 워터만(Waterman)의 알고리즘 (문헌[*Adv. Appl. Math.*, 1981, 2, 482-489])을 사용하는 디폴트 세팅(default setting)을 사용한 Gap 프로그램 (Wisconsin Sequence Analysis Package, 유닉스(Unix)용 버전 8; 미국 위스콘신주 매디슨 유니버시티 리서치 파크 소재의 Genetics Computer Group)을 사용하여 통상적으로 결정될 수 있다.

본 명세서에 제공된 방법 및 조성물은 다양한 상이한 성분을 사용한다. 본 발명의 상세한 설명을 통해, 일부 성분이 활성 변이체 및 단편을 가질 수 있는 것으로 인식된다. 이러한 성분은, 예를 들어 뉴클레아제 작용제, Cas 단백질, CRISPR RNA, tracrRNA, 및 가이드 RNA를 포함한다. 각각의 이들 성분에 대한 생물학적 활성은 본 명세서의 다른 곳에 기술되어 있다.

2개의 폴리뉴클레오티드 또는 폴리펩티드 서열과 관련하여, "서열 동일성" 또는 "동일성"은 특정 비교 윈도우(comparison window)에 대하여 최대 일치도로 정렬되는 경우에 동일한 2개의 서열의 잔기에 대하여 언급된다. 서열 동일성 비율이 단백질에 관하여 사용되는 경우에, 동일하지 않은 잔기 위치는 종종 보존적 아미노산 치환에 의해 달라지며, 여기서 아미노산 잔기가 유사한 화학적 특성 (예를 들어, 전하 또는 소수성)을 갖는 다른 아미노산 잔기 대신에 사용되므로, 분자의 기능적 특성을 변경시키지 않음이 인식된다. 서열이 보존적 치환으로 달라지는 경우에는, 서열 동일성 비율은 치환의 보존적 성질을 보정하기 위해 상향 조정될 수 있다. 이러한 보존적 치환에 의해 달라지는 서열은 "서열 유사성" 또는 "유사성"을 갖는다고 한다. 이러한 조정을 가하는 수단은 잘 알려져 있다. 전형적으로, 이는 보존적 치환을 완전 미스매치보다는 부분 미스매치로서 스코어를 주어서 (scoring), 서열 동일성 비율을 증가시키는 것을 포함한다. 따라서, 예를 들어 동일한 아미노산은 스코어 1로 주어지고, 비보존적 치환은 스코어 0으로 주어지며, 보존적 치환은 0과 1 사이의 스코어로 주어진다. 보존적 치환의 스코어는, 예를 들어 프로그램 PC/GENE (미국 캘리포니아주 마운틴 뷰 소재의 Intelligenetics)으로 실행되는 바와 같이 계산된다.

"서열 동일성 비율"은 비교 윈도우에 대하여 최적으로 정렬된 2개의 서열을 비교함으로써 결정된 값을 포함하며, 여기서 비교 윈도우 내의 폴리뉴클레오티드 서열의 부분은 2개의 서열의 최적 정렬을 위한 기준 서열 (부가 또는 결실을 포함하지 않음)과 비교하여, 부가 또는 결실 (즉, 갭)을 포함할 수 있다. 동일한 핵산 염기 또는 아미노산 잔기가 두 서열에 나타나는 위치의 개수를 결정하여 매칭된 위치의 개수를 산출하고, 매칭된 위치의 개수를 비교 윈도우 내의 위치의 총 개수로 나누고, 그 결과에 100을 곱하여 서열 동일성 비율을 산출함으로써 그 비율을 계산한다.

달리 명시되지 않는 한, 서열 동일성/유사성 값은 하기 파라미터를 사용하여 GAP 버전 10을 이용하여 구한 값을

포함한다: GAP 중량(Weight) 50 및 길이 중량(Length Weight) 3, 및 nwsgapdna.cmp 스코어링 매트릭스(scoring matrix)를 이용한 뉴클레오티드 서열의 % 동일성 및 % 유사성; GAP 중량 8 및 길이 중량 2, 및 BLOSUM62 스코어링 매트릭스를 이용한 아미노산 서열의 % 동일성 및 % 유사성; 또는 이의 임의의 등가 프로그램. "등가 프로그램"은 임의의 서열 비교 프로그램을 포함하는데, 대상이 되는 임의의 2개의 서열의 경우, GAP 버전 10에 의해 생성된 상응하는 정렬과 비교할 때, 동일한 뉴클레오티드 또는 아미노산 잔기 매치 및 동일한 서열 동일성 비율을 갖는 정렬을 얻는다.

"상동성" 서열은, 공지된 기준 서열과 동일하거나 실질적으로 유사하여, 그 공지된 기준 서열과 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일한 핵산 서열을 포함한다. "오르토로그성(orthologous)" 서열은 다른 종의 공지된 기준 서열과 기능적으로 동등한 하나의 종의 핵산 서열을 포함한다.

용어 "시험관내"는 인공 환경 및 인공 환경 (예를 들어, 시험관) 내에서 발생하는 과정 또는 반응까지 포함한다. 용어 "생체내(*in vivo*)"는 천연 환경 (예를 들어, 세포 또는 유기체 또는 신체) 및 천연 환경 내에서 발생하는 과정 또는 반응까지 포함한다. 용어 "생체외(*ex vivo*)"는 개체의 신체으로부터 분리된 세포 및 이러한 세포 내에서 발생하는 과정 또는 반응까지 포함한다.

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

값의 범위 지정은 해당 범위 내 또는 그 범위를 한정하는 모든 정수, 및 그 범위 내의 정수로 한정되는 모든 부분적인 범위를 포함한다.

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

문맥상 달리 명확히 지시하지 않는 한, 단수형은 복수의 지시 대상을 포함한다. 예를 들어, 용어 "Cas 단백질" 또는 "적어도 하나의 Cas 단백질"은 이들의 혼합물을 포함하는, 복수의 Cas 단백질을 포함할 수 있다.

발명을 실시하기 위한 구체적인 내용

[0142]

I. 다수의 표적화 벡터를 사용한 게놈 유전자좌의 변형

[0143]

세포 내의 표적 게놈 유전자좌를 변형시키는 방법 및 조성물이 제공된다. 이러한 방법은, 서로 재조합되어 단일의 연속 핵산 세그먼트를 형성할 수 있는 다수의 큰 표적화 벡터 (LTVEC)를 사용한다. 이러한 방법은 단일 표적화 단계로 1개, 2개, 3개, 4개, 5개, 6개 또는 그 이상의 LTVEC를 이용할 수 있다. 또한, 세포의 표적 게놈 유전자좌에서 상동 재조합을 향상시키는 방법 및 조성물이 제공된다. 이러한 방법은 하나 이상의 중복 서열을 포함하는 2개 이상의 핵산을 사용한다. 본 명세서에 개시된 임의의 방법은 시험관내, 생체외, 또는 생체내에서 수행될 수 있다.

[0144]

A. 이중 표적화

[0145]

이중 표적화 방법을 통해 세포 내의 표적 게놈 유전자좌를 변형시키는 방법 및 조성물이 제공된다. 그 방법 및 조성물은, 서로 재조합되어 단일의 연속 핵산 세그먼트를 형성할 수 있는 2개의 큰 표적화 벡터 (LTVEC) (즉, 제1의 LTVEC 및 제2의 LTVEC)를 사용한다. 제1의 LTVEC는 제1 핵산 삽입물을 포함하고, 제2의 LTVEC는 제2 핵산 삽입물을 포함한다. 핵산 삽입물은 5' 상동성 아암과 3' 상동성 아암에 의해 플랭킹되어 있다. 제1 핵산 삽입물과 이의 3' 상동성 아암 그리고 제2 핵산 삽입물과 이의 5' 상동성 아암은 동일한 연속 핵산의 중복 단편 들일 수 있다. 제1의 LTVEC의 3' 상동성 아암과 제2의 LTVEC의 5' 상동성 아암은 중복되고 (즉, 서로 상보적임), 제1 삽입물과 제2 삽입물은 중복 상동성 아암을 플랭킹한다. 이러한 방법은 임의의 순서로 발생할 수 있는 하기 3가지의 재조합 사건을 포함한다: (1) 제1의 LTVEC의 3' 상동성 아암과 제2의 LTVEC의 5' 상동성 아암 사이의 재조합; (2) 제1의 LTVEC의 5' 상동성 아암과 표적 유전자좌의 상응하는 세그먼트 사이의 재조합; 및 (3) 제2의 LTVEC의 3' 상동성 아암과 표적 유전자좌의 상응하는 세그먼트 사이의 재조합. 이러한 3 방향 재조합은 표적 유전자좌에서 제1 핵산 삽입물과 제2 핵산 삽입물 사이에 위치한 상동성 아암의 중복 서열로 연속 핵산을 재구성한다.

[0146]

LTVEC 각각은 또한 표적 게놈 유전자좌 내의 또는 이 부근의 DNA 영역에 상동성인 5' 또는 3' 상동성 아암을 포

함하여 단일의 연속 핵산 세그먼트의 재조합 및 통합을 가능하게 한다. 따라서, 3 방향 재조합 사건에 의해, 단일 표적화 단계로 표적 유전자좌에서 큰 핵산 변형 (즉, 결실, 삽입 및/또는 대체)이 이루어질 수 있다.

[0147] 3가지의 재조합 사건은 임의의 순서로 발생할 수 있다. 일 실시 형태에서, 2개의 LTVEC의 중복 서열들 사이의 재조합 사건은 표적 유전자좌와의 상동 재조합 전에 발생한다. 다른 실시 형태에서, 표적 유전자좌와의 재조합은 2개의 LTVEC 사이의 재조합 전에 발생한다. 또 다른 실시 형태에서, 3가지의 재조합 사건은 동시에 발생할 수 있다.

[0148] 일 실시 형태에서, 세포에서 표적 게놈 유전자좌를 변형시키는 방법이 제공된다. 이러한 방법은, 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC) 및 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC를 도입시키는 단계를 포함하고, 여기서 제1의 LTVEC의 제1의 3' 상동성 아암은 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 중복 서열을 가지고, 제1의 LTVEC의 제1의 5' 상동성 아암과 제2의 LTVEC의 제2의 3' 상동성 아암은 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이며; 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물과 제2 핵산 삽입물의 통합에 의해 변형된다. 이 방법은 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물과 제2 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 추가로 포함한다.

[0149] **B.삼중 표적화**

[0150] 또한, 삼중 표적화 방법을 통해 세포 내의 표적 게놈 유전자좌를 변형시키는 방법 및 조성물이 제공된다. 그 방법 및 조성물은, 서로 재조합되어 단일의 연속 핵산 세그먼트를 형성할 수 있는 3개의 큰 표적화 벡터 (LTVEC) (즉, 제1의 LTVEC, 제2의 LTVEC 및 제3의 LTVEC)를 사용한다. 제1의 LTVEC는 제1 핵산 삽입물을 포함하고, 제2의 LTVEC는 제2 핵산 삽입물을 포함하며, 제3의 LTVEC는 제3 핵산 삽입물을 포함한다. 핵산 삽입물은 5' 상동성 아암과 3' 상동성 아암에 의해 플랭킹되어 있다. 제1 핵산 삽입물과 이의 3' 상동성 아암 그리고 제2 핵산 삽입물과 이의 5' 상동성 아암은 동일한 연속 핵산의 중복 단편들일 수 있다. 제2 핵산 삽입물과 이의 3' 상동성 아암 그리고 제3 핵산 삽입물과 이의 5' 상동성 아암은 동일한 연속 핵산의 중복 단편들일 수 있다. 제1의 LTVEC의 3' 상동성 아암과 제2의 LTVEC의 5' 상동성 아암은 중복되고 (즉, 서로 상보적임), 제1 삽입물과 제2 삽입물은 중복 상동성 아암을 플랭킹한다. 제2의 LTVEC의 3' 상동성 아암과 제3의 LTVEC의 5' 상동성 아암은 중복되고 (즉, 서로 상보적임), 제2 삽입물과 제3 삽입물은 중복 상동성 아암을 플랭킹한다.

[0151] 이러한 방법은 임의의 순서로 발생할 수 있는 하기 4가지의 재조합 사건을 포함한다: (1) 제1의 LTVEC의 3' 상동성 아암과 제2의 LTVEC의 5' 상동성 아암 사이의 재조합; (2) 제2의 LTVEC의 3' 상동성 아암과 제3의 LTVEC의 5' 상동성 아암 사이의 재조합; (3) 제1의 LTVEC의 5' 상동성 아암과 표적 유전자좌의 상응하는 세그먼트 사이의 재조합; 및 (4) 제3의 LTVEC의 3' 상동성 아암과 표적 유전자좌의 상응하는 세그먼트 사이의 재조합. 이러한 4 방향 재조합은 표적 유전자좌에서 제1 핵산 삽입물과 제2 핵산 삽입물 사이 그리고 제2 핵산 삽입물과 제3 핵산 삽입물 사이에 위치한 상동성 아암의 중복 서열로 연속 핵산을 재구성한다.

[0152] 제1의 LTVEC와 제3의 LTVEC는 또한 표적 게놈 유전자좌 내의 또는 이 부근의 DNA 영역에 상동성인 5' 또는 3' 상동성 아암을 포함하여 단일의 연속 핵산 세그먼트의 재조합 및 통합을 가능하게 한다. 따라서, 4 방향 재조합 사건에 의해, 단일 표적화 단계로 표적 유전자좌에서 큰 핵산 변형 (즉, 결실, 삽입 및/또는 대체)이 이루어질 수 있다.

[0153] 4가지의 재조합 사건은 임의의 순서로 발생할 수 있다. 일 실시 형태에서, 3개의 LTVEC의 중복 서열들 사이의 재조합 사건은 표적 유전자좌와의 상동 재조합 전에 발생한다. 다른 실시 형태에서, 표적 유전자좌와의 재조합은 3개의 LTVEC 사이의 재조합 전에 발생한다. 또 다른 실시 형태에서, 4가지의 재조합 사건은 동시에 발생할 수 있다.

[0154] 일 실시 형태에서, 세포에서 표적 게놈 유전자좌를 변형시키는 방법이 제공된다. 이러한 방법은, 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC, 및 제3의 5' 상동성 아암과 제3의 3' 상동성 아암에 의해 플랭킹된 제3 핵산 삽입물을 포함하는 제3의 LTVEC를 도입시키는 단계를 포함하며, 여기서 제1의 LTVEC의 제1의 3' 상동성 아암은 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 중복 서열을 가지고, 제2의 LTVEC의 제2의 3' 상동성 아암은 제3의 LTVEC의 제3의 5' 상동성 아암에 상동성인 중복 서열을 가지고, 제1의 LTVEC의 제1의 5' 상동성 아암과 제3의 LTVEC의 제3의 3'

상동성 아암은 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이며; 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물의 통합에 의해 변형된다. 이 방법은 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 추가로 포함한다.

[0155] C. 다수의 LTVEC를 사용한 표적화

[0156] 단일 표적화 단계로 유전자 변형을 생성시키기 위해 본 명세서에 제공된 표적화 방법은 단일 LTVEC 표적화 방법으로 달성되는 것 이상으로 표적화된 유전자 변형에 대한 새로운 가능성 및 향상된 효율을 제공한다. 서로 재조합가능한 2개, 3개 또는 그 이상의 LTVEC로 표적화하는 것은 더 큰 DNA 세그먼트의 변형을 가능하게 한다. 재조합 사건은 임의의 순서로 발생할 수 있다. 예를 들어, LTVEC들의 중복 서열들 사이의 재조합 사건은 표적 유전자좌와의 상동 재조합 전에 발생할 수 있다. 대안적으로, 표적 유전자좌와의 재조합은 LTVEC 간의 재조합 전에 발생할 수 있거나 재조합 사건들은 동시에 발생할 수 있다.

[0157] 본 명세서에 기술된 표적화 방법은 기존의 단일 LTVEC 표적화 방법에 비해 증가된 표적화 효율, 달성가능한 유전자 변형의 크기 증가 및 큰 게놈 변형을 얻는 데 필요한 표적화 단계 수의 감소를 포함하는 몇 가지 이점을 제공하며, 이는 시간을 절약해 주고 변형된 배아 줄기 세포의 만능성을 유지시켜 준다. 이것은 큰 게놈 변형을 위해 특히 중요한데, 이는 그 방법이 단일 단계로 2개, 3개 또는 그 이상의 LTVEC로부터의 핵산 삽입물들의 조합에 의한 게놈 유전자좌의 변형을 가능하게 하기 때문이다. 따라서, 이러한 변형은 표적화된 게놈 유전자좌 내에서 매우 큰 (예를 들어, >50 kb) 결실, 대체 및 삽입을 가능하게 할 수 있다.

[0158] 예를 들어, 순차적인 방식으로 3개의 LTVEC를 사용하여 표적 게놈 유전자좌를 변형시키고 표적화된 변형을 스크리닝 및 확인하는 데 필요한 시간은 대략 9개월인 반면, 3개의 LTVEC를 동시에 사용하여 단지 약 4개월만에 동일한 변형이 이루어지고 확인될 수 있다.

[0159] 또한, 순차적인 변형은 배아 줄기 세포와 같은 만능성 세포가 변형될 때 만능성 및 생식세포계열 전달 가능성의 상실 위험을 높인다. 배양 중 계대 수가 증가하고 전기천공(electroporation) 수가 증가함에 따라, 염색체 및 핵형 비정상이 축적되어 생식세포계열 기능(germline competence)의 상실을 초래할 수 있다. 예를 들어, 문헌 [Buehr *et al.* (2008) *Cell* 135:1287-1298]; 문헌 [Li *et al.* (2008) *Cell* 135(7): 1299-1310]; 및 문헌 [Liu *et al.* (1997) *Dev. Dyn.* 209:85-91]이 참조되며, 이들 각각은 모든 목적을 위해 전체적으로 본 명세서에 참고로 포함된다. 순차적이 아니고 동시에 다수의 LTVEC를 사용하여 표적화하는 것은 계대 수 및 전기천공 수를 감소시킴으로써, 배아 줄기 세포와 같은 만능성 세포에서 이의 생식세포계열 기능을 보유하면서 유전자 조작을 수행하는 능력을 증가시킨다.

[0160] 특정 실시 형태에서, 유전자 변형은 하나 이상의 내인성 핵산의 변형, 하나 이상의 내인성 핵산의 치환, 내인성 핵산의 이중성 핵산으로의 대체, 녹아웃, 또는 녹인을 포함한다. 특정 예에서, 유전자 변형은 적어도 2개의 큰 표적화 벡터 (LTVEC)를 세포 내로 도입함으로써 도입된다. 다른 예에서, 유전자 변형은 적어도 3개의 큰 표적화 벡터 (LTVEC)를 세포 내로 도입함으로써 도입된다. 이러한 예에서, LTVEC는 세포의 표적 게놈 유전자좌 내로 삽입될 DNA를 포함할 수 있다.

[0161] 일부 실시 형태에서, 표적 게놈 유전자좌를 변형시키는 방법은 유전자 변형을 포유류 세포 내로 도입시키는 단계를 포함한다. 마찬가지로, 본 발명은 유전자 변형을 포함하는 포유류 세포를 제공한다.

[0162] 세포에서 표적화된 유전자 변형을 생성시키는 다양한 방법이 사용될 수 있다. 예를 들어, 상기 기술된 바와 같이, 표적화된 유전자 변형은 상동 재조합 사건을 통해 표적화된 유전자 변형을 생성시킬 시스템을 사용한다. 다른 경우에는, 표적화된 게놈 유전자좌에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제를 사용하여 세포가 변형될 수 있다. 이어서, 단일 가닥 또는 이중 가닥 절단부가 비상동성 말단 결합 경로(non-homologous end joining pathway, NHEJ)에 의해 수복된다. 예를 들어, 큰 표적화 벡터의 사용을 포함하는, 이러한 표적화된 유전자 변형을 생성시키는 예시적인 방법은 본 명세서의 다른 곳에 상세히 논의되어 있다. 또한, 문헌 [Wang *et al.* (2013) *Cell* 153:910-918]; 문헌 [Mandalos *et al.* (2012) *PLOS ONE* 7:e45768:1-9]; 및 문헌 [Wang *et al.* (2013) *Nat Biotechnol.* 31:530-532]이 참조되며, 이들 각각은 본 명세서에 참고로 포함된다.

[0163] 표적화 벡터와 표적 유전자좌 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 절단과 조합된 표적화 벡터의 사용은 결실 또는 삽입과 같은 변형에 대한 표적화 효율을 크게 향상시킬 수 있다. 유사

하계는, 표적 유전자좌에서의 뉴클레아제 유도 단일 가닥 DNA 절단과 조합된 표적화 벡터의 사용은 변형에 대한 표적화 효율을 크게 향상시킬 수 있다.

[0164] 일부 실시 형태에서, LTVEC는 표적 게놈 유전자좌 내에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제 작용제와 조합하여 사용될 수 있다. 이러한 방법은 뉴클레아제 작용제를 세포 내로 도입시키는 단계를 추가로 포함한다. 일 실시 형태에서, 뉴클레아제 작용제는 징크 핑거 뉴클레아제 (ZFN)이다. 다른 실시 형태에서, 뉴클레아제 작용제는 CRISPR/CRISPR 관련 (Cas) 시스템이다.

[0165] 일 실시 형태에서, 다수의 LTVEC를 이용하여 세포에서 표적 게놈 유전자좌를 변형시키는 방법이 제공된다. 이러한 방법은 (a) 표적 게놈 유전자좌 내에 단일가닥 또는 이중가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계; (b) 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 및 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC를 도입시키는 단계 - 제1의 LTVEC의 제1의 5' 상동성 아암과 제2의 LTVEC의 제2의 3' 상동성 아암은 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이고, 제1의 LTVEC의 제1의 3' 상동성 아암과 제2의 LTVEC의 제2의 5' 아암은 서로 상동성이거나 하나 이상의 추가 LTVEC의 추가 5' 상동성 아암과 3' 상동성 아암에 각각 상동성이고, 이 추가 LTVEC 각각은 추가 5' 상동성 아암과 추가 3' 상동성 아암에 의해 플랭킹된 추가 핵산 삽입물을 포함하며; 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물, (존재하는 경우) 하나 이상의 추가 LTVEC의 하나 이상의 추가 핵산 삽입물, 및 제2 핵산 삽입물의 통합에 의해 변형됨 -; 및 (c) 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물, (존재하는 경우) 하나 이상의 추가 핵산 삽입물, 및 제2 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 포함한다. 이러한 방법에서, 추가 LTVEC는, 존재할 때 제1의 LTVEC와 제2의 LTVEC 사이에 삽입되는 하나 이상의 다른 LTVEC이다.

[0166] 일 실시 형태에서, 세포에서 표적 게놈 유전자좌를 변형시키는 이중 표적화 방법이 제공되며, 이 방법은 (a) 표적 게놈 유전자좌 내에 단일가닥 또는 이중가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계; (b) 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 및 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC를 도입시키는 단계 - 제1의 LTVEC의 제1의 3' 상동성 아암은 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 중복 서열을 가지고, 제1의 LTVEC의 제1의 5' 상동성 아암과 제2의 LTVEC의 제2의 3' 상동성 아암은 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이며, 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물 및 제2 핵산 삽입물의 통합에 의해 변형됨 -; 및 (c) 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물 및 제2 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 포함한다. 이러한 방법에서, 제1 핵 삽입물과 제1의 3' 상동성 아암 그리고 제2 핵산 삽입물과 제2의 5' 상동성 아암은 연속 핵산의 중복 단편들이며, 그 연속 핵산은 표적 게놈 유전자좌 내로의 제1 핵산 삽입물과 제2 핵산 삽입물의 통합에 의해 재형성된다.

[0167] 일 실시 형태에서, 세포에서 표적 게놈 유전자좌를 변형시키는 삼중 표적화 방법이 제공되며, 이 방법은 (a) 표적 게놈 유전자좌 내에 단일가닥 또는 이중가닥 절단부를 생성시키는 뉴클레아제 작용제를 세포 내로 도입시키는 단계; (b) 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 제1의 큰 표적화 벡터 (LTVEC), 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2의 LTVEC, 및 제3의 5' 상동성 아암과 제3의 3' 상동성 아암에 의해 플랭킹된 제3 핵산 삽입물을 포함하는 제3의 LTVEC를 도입시키는 단계 - 제1의 LTVEC의 제1의 3' 상동성 아암은 제2의 LTVEC의 제2의 5' 상동성 아암에 상동성인 중복 서열을 가지고, 제2의 LTVEC의 제2의 3' 상동성 아암은 제3의 LTVEC의 제3의 5' 상동성 아암에 상동성인 중복 서열을 가지고, 제1의 LTVEC의 제1의 5' 상동성 아암과 제3의 LTVEC의 제3의 3' 상동성 아암은 표적 게놈 유전자좌 내의 상응하는 게놈 세그먼트들에 상동성이며, 표적 게놈 유전자좌는, 상응하는 게놈 세그먼트들 사이에서 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물의 통합에 의해 변형됨 -; 및 (c) 표적 게놈 유전자좌에 통합된 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물을 포함하는 표적화된 세포를 선택하는 단계를 포함한다. 이러한 삼중 표적화 방법에서, 제1 핵 삽입물과 제1의 3' 상동성 아암 그리고 제2 핵산 삽입물과 제2의 5' 상동성 아암은 연속 핵산의 중복 단편들이고, 제2 핵 삽입물과 제2의 3' 상동성 아암 그리고 제3 핵산 삽입물과 제3의 5' 상동성 아암은 연속 핵산의 중복 단편들이며, 그 연속 핵산은 표적 게놈 유전자좌 내로의 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물의 통합에 의해 재형성된다.

[0168] 일부의 경우에, 2개, 3개 또는 그 이상의 LTVEC를 동시에 도입시킬 수 있다. 대안적으로, 2개, 3개 또는 그 이

상의 LTVEC를 순차적으로 도입시킬 수 있거나 상이한 시점에 도입시킬 수 있다.

[0169] 표적화 시스템의 다양한 성분은, 예를 들어 표적화 벡터, 뉴클레아제 작용제, 표적 게놈 유전자좌, 핵산 삽입물, 관심 폴리뉴클레오티드, 및/또는 다른 성분을 포함할 수 있으며, 이들 각각은 본 명세서의 다른 곳에 상세히 기술되어 있다.

[0170] **D. 다수의 중복 핵산을 사용한 표적화**

[0171] 단일 표적화 단계로 유전자 변형을 생성시키기 위해 본 명세서에 제공된 표적화 방법은 단일 핵산으로 달성되는 것 이상으로 표적화된 유전자 변형에 대한 새로운 가능성 및 향상된 효율을 제공한다. 서로 재조합가능한 2개, 3개 또는 그 이상의 핵산을 사용한 표적화는 더 큰 DNA 세그먼트의 변형을 가능하게 하고 단일 핵산 단독에 비해 향상된 표적화 효율을 제공할 수 있는데, 이는 뉴클레아제 작용제의 부재 하에서도 그러하다. 뉴클레아제 작용제를 사용하지 않는 이러한 방법은 뉴클레아제 작용제를 사용하는 방법에 비해 유리할 수 있는데, 이는 뉴클레아제를 사용하는 방법에 필요한 스크리닝이, 절단을 확인하고 오프 타겟 효과(off-targeted effect)를 검사하는 추가 스크리닝 단계를 포함하여 더 복잡하고 시간 소모적이기 때문이다. 충분한 길이의 중복 영역을 갖는 핵산 (예를 들어, LTVEC)은 표적화된 뉴클레아제의 부재 하에서도 표적 게놈 유전자좌에서 상동 재조합을 향상시킬 수 있다. 예로서, 충분한 길이의 중복 영역을 갖는 2개의 핵산의 사용은 단일 핵산의 사용과 비교하여 표적 게놈 유전자좌에서 상동 재조합을 향상시킬 수 있다. 메커니즘의 이해가 실시예에 필요한 것은 아니지만, 핵산 (예를 들어, LTVCE) 상에 재조합 기구 (예를 들어, ExoI, Rad51, BRCA2 등)를 로딩하여 표적 유전자좌로의 재조합 기구의 보충을 용이하게 함으로써 이러한 상황 하에서 상동 재조합이 향상되는 것으로 여겨진다.

[0172] 세포의 표적 게놈 유전자좌를 변형시키거나 표적 게놈 유전자좌에서 상동 재조합을 향상시키는 방법이 본 명세서에 제공되며, 이 방법은 제1 핵산과 제2 핵산을 세포 내로 도입시키는 단계 - 제1 핵산과 제2 핵산은 중복 서열을 포함함 - 를 포함한다. 제1 핵산과 제2 핵산은, 예를 들어 선형 핵산일 수 있다. 또한, 이러한 방법은 서로 재조합가능한 3개 이상의 핵산을 세포 내로 도입시키는 단계를 포함할 수 있다. 예를 들어, 제1 핵산과 제2 핵산은 제1 중복 서열을 가질 수 있고, 제2 핵산과 제3 핵산은 제2 중복 서열을 가질 수 있다. 일부 방법에서, 뉴클레아제의 도움 없이, 표적 게놈 유전자좌가 변형되거나 표적 게놈 유전자좌에서의 상동 재조합이 향상된다. 다른 방법에서, 표적 게놈 유전자좌에 또는 이 부근에 단일 가닥 또는 이중 가닥 절단부를 생성시키는 뉴클레아제, 예를 들어 징크 핑거 뉴클레아제, TALEN, 메가뉴클레아제, 또는 Cas9 및 가이드 RNA의 도움으로, 표적 게놈 유전자좌가 변형되거나 표적 게놈 유전자좌에서의 상동 재조합이 향상된다.

[0173] 본 방법은 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합을 향상시킬 수 있거나, 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합을 향상시킬 수 있거나, 표적 게놈 유전자좌에서 제1 핵산과 제2 핵산의 둘 모두의 상동 재조합을 향상시킬 수 있다. 예로서, 표적 게놈 유전자좌에서 제1 핵산의 상동 재조합은 제1 핵산이 제2 핵산 없이 도입되는 방법과 비교하여 향상될 수 있다. 마찬가지로, 표적 게놈 유전자좌에서 제2 핵산의 상동 재조합은 제2 핵산이 제1 핵산 없이 도입되는 방법과 비교하여 향상될 수 있다. 상동 재조합의 향상은, 예를 들어 적어도 1.5배, 2배, 2.5배, 3배, 4배, 5배, 6배, 7배, 8배, 9배, 10배, 11배, 12배, 13배, 14배, 15배, 16배, 17배, 18배, 19배 또는 20배일 수 있다. 일부 방법에서, 뉴클레아제 부재 하의 향상은 뉴클레아제 존재 하의 향상에 필적할 수 있다. 예를 들어, 뉴클레아제 존재 하의 향상의 배수 변화는 뉴클레아제 부재 하의 향상과 비교할 때 0.5배, 0.6배, 0.7배, 0.8배, 0.9배, 1.0배, 1.1배, 1.2배, 1.3배, 1.4배, 1.5배, 1.6배, 1.7배, 1.8배, 1.9배, 2배, 3배, 4배, 5배, 6배, 7배, 8배, 9배 또는 10배일 수 있다. 일부의 경우에, 뉴클레아제 부재 하의 향상은 뉴클레아제 존재 하의 향상과 동일하거나 이보다 클 수 있다.

[0174] 제1 핵산의 중복 서열은 제2 핵산의 중복 서열에 상동성일 수 있다. 예를 들어, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열과 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 99.5% 또는 99.9% 동일할 수 있다. 대안적으로, 제1 핵산의 중복 서열은 제2 핵산의 중복 서열과 100% 동일할 수 있다.

[0175] 중복 서열은, 예를 들어 약 1 kb 내지 약 70 kb 또는 그 이상일 수 있다. 예를 들어, 중복 서열은 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 15 kb, 약 15 kb 내지 약 20 kb, 약 20 kb 내지 약 25 kb, 약 25 kb 내지 약 30 kb, 약 30 kb 내지 약 35 kb, 약 35 kb 내지 약 40 kb, 약 40 kb 내지 약 45 kb, 약 45 kb 내지 약 50 kb, 약 50 kb 내지 약 55 kb, 약 55 kb 내지 약 60 kb, 약 60 kb 내지 약 65 kb, 약 65 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 120 kb, 약 120 kb 내지 약 140 kb, 약 140 kb 내지 약 160 kb, 약 160 kb 내지 약 180 kb, 약 180 kb 내지 약 200 kb, 약 200 kb 내지 약 220 kb, 약 220 kb 내지 약 240 kb, 약 240 kb 내지 약 260 kb, 약 260 kb 내지 약 280 kb, 또는 약 280 kb 내지 약 300 kb일 수 있다. 예로서, 중복 서열은 약 20 kb 내지 약 60 kb일 수 있

다. 대안적으로, 중복 서열은 적어도 1 kb, 적어도 5 kb, 적어도 10 kb, 적어도 15 kb, 적어도 20 kb, 적어도 25 kb, 적어도 30 kb, 적어도 35 kb, 적어도 40 kb, 적어도 45 kb, 적어도 50 kb, 적어도 55 kb, 적어도 60 kb, 적어도 65 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 120 kb, 적어도 140 kb, 적어도 160 kb, 적어도 180 kb, 적어도 200 kb, 적어도 220 kb, 적어도 240 kb, 적어도 260 kb, 적어도 280 kb 또는 적어도 300 kb일 수 있다.

[0176] 중복 서열은 제1 핵산과 제2 핵산 내의 임의의 위치에 위치할 수 있다. 예를 들어, 중복 서열은 제1 핵산의 3' 말단 및 제2 핵산의 5' 말단에 위치할 수 있다. 대안적으로, 중복 서열은 제1 핵산의 5' 말단 및 제2 핵산의 3' 말단에 위치할 수 있다.

[0177] 일부 방법에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 표적화 벡터이다. 제2 핵산은 중복 서열을 포함하는 임의의 핵산, 예를 들어 플라스미드, 표적화 벡터 또는 큰 표적화 벡터일 수 있다. 일부 방법에서, 제2 핵산은 중복 서열을 제외하고는 표적 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다. 예를 들어, 제2 핵산은 중복 서열로 본질적으로 이루어지거나 이로 이루어질 수 있다.

[0178] 일부 방법에서, 제1 핵산은 제1의 5' 상동성 아암과 제1의 3' 상동성 아암에 의해 플랭킹된 제1 핵산 삽입물을 포함하는 표적화 벡터이고, 제2 핵산은 제2의 5' 상동성 아암과 제2의 3' 상동성 아암에 의해 플랭킹된 제2 핵산 삽입물을 포함하는 제2 표적화 벡터이다.

[0179] 제1 표적화 벡터는 임의의 크기를 가질 수 있다. 마찬가지로, 제2 표적화 벡터는 임의의 크기를 가질 수 있다. 예를 들어, 제1 표적화 벡터 및/또는 제2 표적화 벡터는 약 1 kb 내지 약 2 kb, 약 2 kb 내지 약 5 kb, 또는 약 5 kb 내지 약 10 kb일 수 있다. 제1 표적화 벡터는 또한 큰 표적화 벡터(LTVEC)일 수 있다. 마찬가지로, 제2 표적화 벡터는 LTVEC일 수 있다. LTVEC의 예시적인 크기는 본 명세서의 다른 곳에 개시되어 있다. 예를 들어, 제1의 LTVEC 및/또는 제2의 LTVEC는 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 120 kb, 약 120 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 450 kb, 약 450 kb 내지 약 500 kb, 약 500 kb 내지 약 550 kb, 약 550 kb 내지 약 600 kb, 약 600 kb 내지 약 650 kb, 약 650 kb 내지 약 700 kb, 약 700 kb 내지 약 750 kb 또는 약 750 kb 내지 약 800 kb일 수 있다.

[0180] 일부 방법에서, 제1 핵산은 LTVEC이고, 제2 핵산은 중복 서열을 포함하는 더 작은 핵산, 예를 들어 플라스미드 또는 표적화 벡터이다. 일부 방법에서, 제2 핵산은 중복 서열을 제외하고는 표적 게놈 유전자좌에 상동성인 뉴클레오티드 서열을 포함하지 않는다. 예를 들어, 제2 핵산은 중복 서열로 본질적으로 이루어지거나 이로 이루어질 수 있다.

[0181] 일부 방법에서, 제1 핵산 삽입물과 제2 핵산 삽입물은 연속 핵산의 중복 단편들이다. 일부 방법에서, 제1 핵산 삽입물 및/또는 제2 핵산 삽입물은 세포의 종과 상이한 종으로부터 유래될 수 있다. 예를 들어, 제1 핵산 삽입물 및/또는 제2 핵산 삽입물은 인간 핵산일 수 있다.

[0182] 본 방법은 표적 게놈 유전자좌 내로의 제1 핵산 삽입물 및/또는 제2 핵산 삽입물의 통합을 가져올 수 있다. 통합은 표적 게놈 유전자좌에 서열의 첨가, 표적 게놈 유전자좌에서 서열의 결실, 또는 표적 게놈 유전자좌에서 서열의 대체를 가져올 수 있다. 예를 들어, 통합은 표적 게놈 유전자좌에 외인성 서열의 첨가, 표적 게놈 유전자좌에서 내인성 서열의 결실, 또는 표적 게놈 유전자좌에서 내인성 서열의 외인성 서열로의 대체를 가져올 수 있다. 표적 게놈 유전자좌에 삽입되는 제1 핵산 삽입물, 제2 핵산 삽입물 또는 제1 핵산 삽입물과 제2 핵산 삽입물의 조합은, 예를 들어 약 5 kb 내지 약 500 kb일 수 있다. 다른 예시적인 핵산 삽입물 및 삽입 크기는 본 명세서의 다른 곳에 개시되어 있다. 표적 게놈 유전자좌에서 결실은, 예를 들어 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 600 kb, 약 600 kb 내지 약 700 kb, 또는 약 700 kb 내지 약 800 kb일 수 있다. 다른 예시적인 결실 크기는 본 명세서의 다른 곳에 개시되어 있다.

[0183] 표적화된 세포는 본 명세서에 제공된 임의의 세포 유형일 수 있고, 표적 게놈 유전자좌는 세포 내의 임의의 DNA 일 수 있다. 예를 들어, 표적 게놈 유전자좌는 세포의 게놈에 존재할 수 있거나 세포 내의 염색체의 DNA 상에 존재할 수 있다.

[0184] **II. 핵산 삽입물 및 표적화 벡터**

[0185] **A. 핵산 삽입물**

[0186] 하나 이상의 핵산 삽입물을 본 명세서에 개시된 방법에 사용할 수 있으며, 이것을 개별적인 표적화 벡터를 통해 또는 동일한 표적화 벡터로 세포 내로 도입시킬 수 있다. 핵산 삽입물은 게놈 표적 유전자좌에 통합될 DNA 세그먼트를 포함한다. 표적 유전자좌에서 핵산 삽입물의 통합은 표적 유전자좌에 대한 관심 핵산 서열의 첨가, 표적 유전자좌에서 관심 핵산 서열의 결실, 및/또는 표적 유전자좌에서 관심 핵산 서열의 대체를 가져올 수 있다.

[0187] 본 방법은 종래의 단일 표적화 기법 (즉, 단일 LTVEC)을 사용하여 달성될 수 있는 것보다 크기가 더 큰 핵산 삽입물을 사용한 게놈 유전자좌의 변형을 제공한다. 이러한 방법에서, 핵산 삽입물은 2개, 3개 또는 그 이상의 LTVEC에 포함된다. LTVEC들은, 이들이 서로 재조합되어 2개, 3개 또는 그 이상의 LTVEC로부터의 조합된 핵산 삽입물들을 포함하는 단일의 큰 DNA 세그먼트를 형성할 수 있도록 설계된다.

[0188] 이러한 방법에서, 핵산 삽입물은 5' 상동성 아암과 3' 상동성 아암에 의해 플랭킹되어 있다. 제1 핵산 삽입물을 플랭킹하는 3' 상동성 아암과 제2 핵산 삽입물을 플랭킹하는 5' 상동성 아암은 동일한 연속 핵산의 중복 단편들이며, 그 연속 핵산은 이어서 그 상동성 아암들의 중복 단편들 사이의 재조합에 의해 재형성된다. 이러한 방법에서, 2개의 LTVEC 사이의 재조합은 제1 핵산 삽입물과 제2 핵산 삽입물 사이에 위치한 상동성 아암들의 중복 서열을 갖는 연속 핵산 삽입물을 생성시킨다. 삼중 표적화 방법은 제2의 LTVEC와 제3의 LTVEC 사이의 추가 재조합을 포함하며, 여기서 제2 핵산 삽입물을 플랭킹하는 3' 상동성 아암과 제3 핵산 삽입물을 플랭킹하는 5' 상동성 아암은 동일한 연속 핵산의 중복 단편들이며, 그 연속 핵산은 이어서 그 상동성 아암들의 중복 단편들 사이의 재조합에 의해 재형성된다. 이러한 삼중 표적화 방법에서, 3개의 LTVEC 사이의 재조합은 제1 핵산 삽입물, 제2 핵산 삽입물 및 제3 핵산 삽입물 사이에 위치한 상동성 아암의 중복 서열을 갖는 연속 핵산 삽입물을 생성시킨다. 일 실시 형태에서, 상동성 아암의 중복 서열은 핵산 삽입물의 일부를 포함한다.

[0189] 이와 같이, 이들 방법은 단일 표적화 단계로 2개, 3개 또는 그 이상의 LTVEC로부터의 핵산 삽입물을 사용한 게놈 유전자좌의 변형을 가능하게 함으로써, 핵산 삽입물의 총 크기를 효과적으로 증가시키는 동시에 표적화 단계 수를 감소시킨다.

[0190] 핵산 삽입물 또는 대체되는 표적 유전자좌의 상응하는 핵산은 코딩 영역, 인트론, 엑손, 비번역 영역, 조절 영역, 프로모터, 인핸서, 또는 이들의 임의의 조합일 수 있다. 더욱이, 핵산 삽입물 (즉, 2개, 3개 또는 그 이상의 LTVEC로부터의 조합된 핵산 삽입물들) 또는 대체되는 표적 유전자좌의 상응하는 핵산의 크기는, 예를 들어 10개 내지 100개의 뉴클레오티드의 길이, 100개 내지 500개의 뉴클레오티드의 길이, 500개의 뉴클레오티드 내지 1 kb의 길이, 1 kb 내지 1.5 kb의 길이, 1.5 kb 내지 2 kb의 길이, 2 kb 내지 2.5 kb의 길이, 2.5 kb 내지 3 kb의 길이, 3 kb 내지 5 kb의 길이, 5 kb 내지 8 kb의 길이, 8 kb 내지 10 kb의 길이 또는 그 이상을 포함하는 임의의 요망되는 길이를 가질 수 있다. 다른 경우에, 길이는 약 50 kb 내지 약 700 kb, 약 50 kb 내지 약 500 kb, 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, 약 275 kb 내지 약 300 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 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 450 kb, 약 450 kb 내지 약 500 kb, 약 500 kb 내지 약 550 kb, 약 550 kb 내지 약 600 kb, 약 600 kb 내지 약 650 kb, 약 650 kb 내지 약 700 kb, 약 700 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, 또는 약 2.8 Mb 내지 약 3 Mb일 수 있다. 대안적으로, 2개, 3개 또는 그 이상의 LTVEC로부터의 조합된 핵산 삽입물들 또는 대체되는 표적 유전자좌의 상응하는 핵산은 약 3 Mb 내지 약 4 Mb, 약 4 Mb 내지 약 5 Mb, 약 5 Mb 내지 약 6 Mb, 약 6 Mb 내지 약 7 Mb, 약 7 Mb 내지 약 8 Mb, 약 8 Mb 내지 약 9 Mb, 또는 약 9 Mb 내지 약 10 Mb일 수 있다. 또 다른 경우에, 길이는 적어도 100개, 200개, 300개, 400개, 500개, 600개, 700개, 800개, 또는 900개의 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. 예를 들어, 2개, 3개 또는 그 이상의 LTVEC(들)로부터 조합된 핵산 삽입물들 또는 대체되는 표적 유전자좌의 상응하는 핵산은 적어도 20 kb, 적어도 40 kb, 적어도 60 kb, 적어도 80 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, 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, 적어도 2 Mb, 적어도 2.5 Mb, 적어도 3 Mb, 적어도 4 Mb, 적어도 5 Mb, 적어도 6 Mb, 적어도 7 Mb, 적어도 8 Mb, 적어도 9 Mb, 적어도 10 Mb일 수 있다. 일 실시 형태에서, 핵산 삽입물의 크기는 약 5 kb 내지 약 700 kb이다. 일 실시 형태에서, 핵산 삽입물의 크기는 약 5 kb 내지 약 500 kb이다. 다른 실시 형태에서, 핵산 삽입물의 크기는 약 100 kb 내지 약 700 kb이다. 다른 실시 형태에서, 핵산 삽입물의 크기는 약 100 kb 내지 약 500 kb이다. 특정 실시 형태에서, 핵산 삽입물은 약 140 kb이다. 다른 특정 실시 형태에서, 핵산 삽입물은 약 370 kb이다. 다른 특정 실시 형태에서, 핵산 삽입물은 약 300 kb이다. 다른 특정 실시 형태에서, 핵산 삽입물은 약 400 kb이다.

[0191] 일부의 개별 표적화 벡터 (즉, 다른 표적화 벡터와의 재조합 이전)에서, 핵산 삽입물은 10개 내지 100개의 뉴클레오티드의 길이, 100개 내지 500개의 뉴클레오티드의 길이, 500개의 뉴클레오티드 내지 1 kb의 길이, 1 kb 내지 1.5 kb의 길이, 1.5 kb 내지 2 kb의 길이, 2 kb 내지 2.5 kb의 길이, 2.5 kb 내지 3 kb의 길이, 또는 3 kb 내지 5 kb의 길이일 수 있다. 다른 경우에, 길이는 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 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. 대안적으로, 핵산 삽입물은 약 400 kb 내지 약 450 kb, 약 450 kb 내지 약 500 kb, 약 500 kb 내지 약 550 kb, 약 550 kb 내지 약 600 kb, 약 600 kb 내지 약 650 kb, 약 650 kb 내지 약 700 kb, 약 700 kb 내지 약 750 kb, 또는 약 750 kb 내지 약 800 kb일 수 있다.

[0192] 일부의 경우에, 표적 유전자좌에서 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. 일부의 경우에, 결실의 정도는 5' 상동성 아암과 3' 상동성 아암의 총 길이를 초과한다.

[0193] 일부의 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. 대안적으로, 결실은 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.

[0194] 다른 경우에, 핵산 삽입물 또는 대체되는 표적 유전자좌의 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 120 kb, 적어도 150 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 또는 그 이상일 수 있다.

[0195] 핵산 삽입물은 게놈 DNA 또는 임의의 다른 DNA 유형을 포함할 수 있다. 예를 들어, 핵산 삽입물은 원핵생물, 진핵생물, 효모, 조류 (예를 들어, 닭), 비인간 포유류, 설치류, 인간, 래트, 마우스, 햄스터, 토끼, 피그, 소, 사슴, 양, 염소, 고양이, 개, 흰족제비(ferret), 영장류 (예를 들어, 마모셋, 붉은털원숭이), 가축 포유류 또는

농업용 포유류, 또는 임의의 다른 관심 유기체로부터 유래될 수 있다.

- [0196] 핵산 삽입물 및/또는 표적 유전자좌의 핵산은 코딩 서열 또는 비코딩(non-coding) 서열, 예를 들어 조절 요소 (예컨대, 프로모터, 인핸서 또는 전사 억제인자 결합 요소)를 포함할 수 있다. 예를 들어, 핵산 삽입물은 내인성 유전자의 적어도 하나의 엑손의 녹인 대립유전자, 또는 전체 내인성 유전자의 녹인 대립유전자 (즉, "유전자 교체 녹인")를 포함할 수 있다. 예를 들어, 핵산 삽입물은 게놈 표적 유전자좌에서의 결실을 위해 표적화되는 서열에 상동성 또는 오르토로그성일 수 있다. 상동성 또는 오르토로그성 핵산 삽입물은 관심 게놈 유전자좌에서의 결실을 위해 표적화되는 서열을 대체할 수 있다. 이는, 핵산 삽입물의 삽입이 비인간 핵산 서열의 상동성 또는 오르토로그성 인간 핵산 서열로의 대체를 가져오는 (즉, 핵산 삽입물이 내인성 게놈 유전자좌의 상응하는 비인간 DNA 서열 대신에 삽입되는) 경우 유전자좌의 인간화를 가져올 수 있다.
- [0197] 핵산 삽입물은 또한 조건부 대립유전자(conditional allele)를 포함할 수 있다. 조건부 대립유전자는 전체적으로 참고로 포함되는 미국 특허 출원 공개 제2011/0104799호에 기술된 바와 같은 다기능성 대립유전자일 수 있다. 예를 들어, 조건부 대립유전자는 (a) 표적 유전자의 전사에 대하여 센스 배향의 작동 서열(actuating sequence); (b) 센스 또는 안티센스 배향의 약물 선택 카세트 (DSC); (c) 안티센스 배향의 관심 뉴클레오티드 서열 (NSI); 및 (d) 역배향의 반전 모듈에 의한 조건부(conditional by inversion module, COIN; 이는 엑손 분할 인트론 및 반전성 유전자 트랩 유사 모듈을 이용함)를 포함할 수 있다. 예를 들어, 전체적으로 참고로 포함되는 미국 특허 출원 공개 제2011/0104799호가 참조된다. 조건부 대립유전자는 제1 재조합효소에 노출 시 재조합되어 조건부 대립유전자를 형성하는 재조합가능한 유닛을 추가로 포함할 수 있으며, 그 조건부 대립유전자는 (i) 작동 서열과 DSC가 결합되어 있고; (ii) 센스 배향의 NSI 및 안티센스 배향의 COIN을 함유한다. 미국 특허 출원 공개 제2011/0104799호가 참조된다.
- [0198] 일부 핵산 삽입물은 선택 마커를 인코딩하는 폴리뉴클레오티드를 포함한다. 선택 마커는 선택 카세트에 함유될 수 있다. 이러한 선택 마커는 네오마이신 포스포트랜스퍼라제(neo^r), 하이그로마이신 B 포스포트랜스퍼라제(hyg^r), 퓨로마이신-N-아세틸트랜스퍼라제(puro^r), 블라스티사이드인 S 데아미나아제(bsr^r), 잔틴/구아닌 포스포리보실 트랜스퍼라제(gpt), 또는 단순 포진 바이러스 티미딘 키나제(HSV-k), 또는 이들의 조합을 포함하지만 이에 한정되지 않는다. 선택 마커를 인코딩하는 폴리뉴클레오티드는 표적화되는 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 프로모터의 예는 본 명세서의 다른 곳에 기재되어 있다.
- [0199] 일부 표적화 벡터의 경우, 핵산 삽입물은 리포터 유전자를 포함한다. 리포터 유전자의 예는 루시퍼라제, β-갈락토시다제, 녹색 형광 단백질(GFP), 강화(enhanced) 녹색 형광 단백질(eGFP), 사이안 형광 단백질(CFP), 황색 형광 단백질(YFP), 강화 황색 형광 단백질(eYFP), 청색 형광 단백질(BFP), 강화 청색 형광 단백질(eBFP), DsRed, ZsGreen, MmGFP, mPlum, mCherry, tdTomato, mStrawberry, J-Red, mOrange, mKO, mCitrine, 비너스(Venus), YPet, 에메랄드(Emerald), CyPet, 세룰리언(Cerulean), T-사파이어(Sapphire), 알칼리성 포스파타제, 및 이들의 조합을 인코딩하는 유전자이다. 이러한 리포터 유전자는 표적화되는 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 프로모터의 예는 본 명세서의 다른 곳에 기재되어 있다.
- [0200] 일부 표적화 벡터의 경우, 핵산 삽입물은 하나 이상의 발현 카세트 또는 결실 카세트를 포함한다. 주어진 카세트는, 발현에 영향을 미치는 다양한 조절 성분과 함께, 관심 뉴클레오티드 서열, 선택 마커를 인코딩하는 핵산, 및/또는 리포터 유전자를 포함할 수 있다. 포함될 수 있는 선택가능 마커 및 리포터 유전자의 예는 본 명세서의 다른 곳에 상세히 논의되어 있다.
- [0201] 일부 표적화 벡터의 경우, 핵산 삽입물은 부위 특이적 재조합 표적 서열에 의해 플랭킹된 핵산을 포함한다. 전체 핵산 삽입물이 이러한 부위 특이적 재조합 표적 서열에 의해 플랭킹될 수 있지만, 핵산 삽입물 내의 임의의 영역 또는 개별 관심 폴리뉴클레오티드도 이러한 부위에 의해 플랭킹될 수 있다. 핵산 삽입물 또는 핵산 삽입물의 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 삽입물에 함유된, 선택 마커를 인코딩하는 폴리뉴클레오티드 및/또는 리포터 유전자를 플랭킹한다. 표적화된 유전자좌에서 핵산 삽입물의 통합 후에, 부위 특이적 재조합 부위들 사이의 서열이 제거될 수 있다.
- [0202] **B. 관심 폴리뉴클레오티드**
- [0203] 임의의 관심 폴리뉴클레오티드가 다양한 핵산 삽입물에 함유되어 표적 게놈 유전자좌에 통합될 수 있다. 본 명세서에 개시된 방법은 표적화된 게놈 유전자좌에 통합될 적어도 1개, 2개, 3개, 4개, 5개, 6개 또는 그 이상의

관심 폴리뉴클레오티드를 제공한다.

- [0204] 표적 게놈 유전자좌에 통합될 때 핵산 삽입물 내의 관심 폴리뉴클레오티드는 세포 내로 하나 이상의 유전자 변형을 도입시킬 수 있다. 유전자 변형은 내인성 핵산 서열의 결실 및/또는 표적 게놈 유전자좌 내로의 외인성 또는 이중성 또는 오르토로그성 폴리뉴클레오티드의 첨가를 포함할 수 있다. 일 실시 형태에서, 유전자 변형은 표적 게놈 유전자좌에서 내인성 핵산 서열의 관심 외인성 폴리뉴클레오티드로의 대체를 포함한다. 따라서, 본 명세서에 제공되는 방법은 표적 게놈 유전자좌에서의 녹아웃, 결실, 삽입, 치환 ("녹인"), 점 돌연변이, 도메인 교체, 엑손 교체, 인트론 교체, 조절 서열 교체, 유전자 교체 또는 이들의 조합을 포함하는 유전자 변형의 생성을 가능하게 한다. 이러한 변형은 표적 게놈 유전자좌 내로의 제1, 제2, 제3, 제4, 제5, 제6, 제7 또는 임의의 후속 핵산 삽입물의 통합 시에 발생할 수 있다.
- [0205] 핵산 삽입물 내의 및/또는 표적 게놈 유전자좌에 통합된 관심 폴리뉴클레오티드는 이것이 도입되는 세포에 고유하거나 상동성인 서열을 포함할 수 있거나; 관심 폴리뉴클레오티드는 이것이 도입되는 세포에 외인성일 수 있거나; 관심 폴리뉴클레오티드는 이것이 도입되는 세포에 오르토로그성일 수 있거나; 관심 폴리뉴클레오티드는 이것이 도입되는 세포와 상이한 종으로부터 유래될 수 있다. 서열과 관련하여 "상동성"은 세포에 고유한 서열을 포함한다. 서열과 관련하여, "이중성"은 외래종으로부터 기원하는 서열이거나, 동일한 종으로부터 유래되는 경우 의도적인 인간 개입에 의해 조성 및/또는 게놈 유전자좌 면에서 이의 천연 형태로부터 실질적으로 변형된 서열을 포함한다. 서열과 관련하여 "외인성"은 외래종으로부터 기원하는 서열을 포함한다. "오르토로그성"은 다른 종의 공지된 기준 서열과 기능적으로 동등한 하나의 종으로부터의 폴리뉴클레오티드 (즉, 종 변이체)를 포함한다. 관심 폴리뉴클레오티드는 비인간, 설치류, 햄스터, 마우스, 래트, 인간, 원숭이, 조류, 농업용 포유류 또는 비농업용 포유류를 포함하지만 이에 한정되지 않는 임의의 관심 유기체로부터 유래될 수 있다. 관심 폴리뉴클레오티드는 코딩 영역, 비코딩 영역, 조절 영역 또는 게놈 DNA를 추가로 포함할 수 있다. 따라서, 제1, 제2, 제3, 제4, 제5, 제6, 제7 및/또는 임의의 후속 핵산 삽입물은 이러한 서열을 포함할 수 있다.
- [0206] 일 실시 형태에서, 핵산 삽입물 내의 및/또는 표적 게놈 유전자좌에 통합된 관심 폴리뉴클레오티드는 인간 핵산에 상동성이다. 또 다른 실시 형태에서, 표적 유전자좌에 통합된 관심 폴리뉴클레오티드는 게놈 핵산의 단편이다. 일 실시 형태에서, 게놈 핵산은 마우스 게놈 핵산, 인간 게놈 핵산, 비인간 핵산, 설치류 핵산, 래트 핵산, 햄스터 핵산, 원숭이 핵산, 농업용 포유류 핵산 또는 비농업용 포유류 핵산, 또는 이들의 조합이다.
- [0207] 일 실시 형태에서, 관심 폴리뉴클레오티드는 상기 기술된 바와 같이, 약 500개의 뉴클레오티드 내지 약 200 kb의 범위일 수 있다. 관심 폴리뉴클레오티드는 약 500개의 뉴클레오티드 내지 약 5 kb, 약 5 kb 내지 약 200 kb, 약 5 kb 내지 약 700 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 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 600 kb, 또는 약 600 kb 내지 약 700 kb일 수 있다.
- [0208] 핵산 삽입물 내의 및/또는 표적 게놈 유전자좌에 삽입된 관심 폴리뉴클레오티드는 폴리펩티드를 인코딩할 수 있거나, miRNA를 인코딩할 수 있거나, 긴 비코딩 RNA를 인코딩할 수 있거나, 이것은, 예를 들어 조절 서열, 프로모터 서열, 인핸서 서열, 전사 억제인자 결합 서열을 포함하는 임의의 관심 조절 영역 또는 비코딩 영역, 또는 비단백질 코딩(non-protein-coding) 서열의 결실을 포함할 수 있지만, 단백질 코딩 서열의 결실은 포함하지 않는다. 또한, 핵산 삽입물 내의 및/또는 표적 게놈 유전자좌에 삽입된 관심 폴리뉴클레오티드는 신경계, 골격계, 소화계, 순환계, 근육계, 호흡계, 심혈관계, 림프계, 내분비계, 비뇨계, 생식계 또는 이들의 조합에서 발견되는 단백질을 인코딩할 수 있다.
- [0209] 핵산 삽입물 내의 및/또는 표적 게놈 유전자좌에 통합된 관심 폴리뉴클레오티드는 코딩 서열의 유전자 변형을 포함할 수 있다. 이러한 유전자 변형은 코딩 서열의 결실 돌연변이 또는 두 코딩 서열의 융합을 포함하지만 이에 한정되지 않는다.
- [0210] 핵산 삽입물 내의 및/또는 표적 게놈 유전자좌에 통합된 관심 폴리뉴클레오티드는 돌연변이 단백질을 인코딩하는 폴리뉴클레오티드를 포함할 수 있다. 일 실시 형태에서, 돌연변이 단백질은 결합 특징 변경, 국제화 변경, 발현 변경 및/또는 발현 패턴 변경을 특징으로 한다. 일 실시 형태에서, 핵산 삽입물 내의 및/또는 게놈 표적 유전자좌에 통합된 관심 폴리뉴클레오티드는 적어도 하나의 질병 대립유전자를 포함한다. 이러한 경우에, 질병

대립유전자는 우성 대립유전자일 수 있거나, 질병 대립유전자는 열성 대립유전자이다. 더욱이, 질병 대립유전자는 단일 뉴클레오타이드 다형 (SNP) 대립유전자를 포함할 수 있다. 돌연변이 단백질을 인코딩하는 관심 폴리뉴클레오타이드는, 돌연변이 단백질을 인코딩하는, 포유류, 비인간 포유류, 설치류, 마우스, 래트, 인간, 원숭이, 농업용 포유류 또는 가축 포유류를 포함하지만 이에 한정되지 않는 임의의 유기체의 폴리뉴클레오타이드로부터 유래될 수 있다.

[0211] 핵산 삽입물 내의 및/또는 표적 게놈 유전자좌에 통합된 관심 폴리뉴클레오타이드는 또한, 예를 들어 프로모터 서열, 인핸서 서열, 전사 억제인자 결합 서열, 또는 전사 종결자 서열을 포함하는 조절 서열을 포함할 수 있다. 특정 실시 형태에서, 핵산 삽입물 내의 및/또는 표적 게놈 유전자좌에 통합된 관심 폴리뉴클레오타이드는 비단백질 코딩 서열의 결실을 갖는 폴리뉴클레오타이드를 포함하지만, 단백질 코딩 서열의 결실은 포함하지 않는다. 일 실시 형태에서, 비단백질 코딩 서열의 결실은 조절 서열의 결실을 포함한다. 다른 실시 형태에서, 조절 요소의 결실은 프로모터 서열의 결실을 포함한다. 일 실시 형태에서, 조절 요소의 결실은 인핸서 서열의 결실을 포함한다. 이러한 관심 폴리뉴클레오타이드는 돌연변이 단백질을 인코딩하는, 포유류, 비인간 포유류, 설치류, 마우스, 래트, 인간, 원숭이, 농업용 포유류 또는 가축 포유류를 포함하지만 이에 한정되지 않는 임의의 유기체의 폴리뉴클레오타이드로부터 유래될 수 있다.

[0212] 표적화된 유전자 변형은 관심 폴리뉴클레오타이드에 대한 표적화된 변경을 포함할 수 있다. 이러한 표적화된 변형은 하나 이상의 뉴클레오타이드의 첨가, 하나 이상의 뉴클레오타이드의 결실, 하나 이상의 뉴클레오타이드의 치환, 관심 폴리뉴클레오타이드 또는 이의 일부의 녹아웃, 관심 폴리뉴클레오타이드 또는 이의 일부의 녹인, 내인성 핵산 서열의 이중성 핵산 서열로의 대체, 또는 이들의 조합을 포함하지만 이에 한정되지 않는다. 특정 실시 형태에서, 적어도 1개, 2개, 3개, 4개, 5개, 7개, 8개, 9개, 10개, 100개, 500개 또는 그 이상의 뉴클레오타이드 또는 적어도 10 kb 내지 500 kb 또는 그 이상이 변화되어 표적화된 게놈 변형을 형성한다.

[0213] **C. 표적화 벡터**

[0214] 표적화 벡터는 핵산 삽입물을 게놈 표적 유전자좌 내로 도입시키는 데 사용될 수 있고, 핵산 삽입물 및 핵산 삽입물을 플랭킹하는 상동성 아암을 포함할 수 있다. 표적화 벡터는 선형 형태 또는 원형 형태일 수 있으며, 이것은 단일 가닥 또는 이중 가닥일 수 있다. 표적화 벡터는 데옥시리보핵산 (DNA) 또는 리보핵산 (RNA)일 수 있다. 참조의 용이함을 위해, 상동성 아암은 본 명세서에서 5' 및 3' (즉, 상류 및 하류) 상동성 아암으로 지칭된다. 이 용어는 표적화 벡터 내의 핵산 삽입물에 대한 상동성 아암의 상대적인 위치에 관한 것이다. 5' 상동성 아암과 3' 상동성 아암은 표적화된 유전자좌 내의 영역 또는 다른 표적화 벡터 내의 영역에 상응하며, 이는 본 명세서에서 각각 "5' 표적 서열"과 "3' 표적 서열"로 지칭된다. 일부의 경우에, 상동성 아암은 또한 5' 또는 3' 표적 서열로 기능할 수 있다.

[0215] 본 방법은 서로 재조합가능한 2개, 3개 또는 그 이상의 표적화 벡터를 사용한다. 다양한 실시 형태에서, 표적화 벡터는 본 명세서의 다른 곳에 기술된 바와 같이 큰 표적화 벡터 (LTVEC)이다. 이러한 방법에서, 제1 표적화 벡터, 제2 표적화 벡터 및 제3 표적화 벡터는 각각 5' 상동성 아암과 3' 상동성 아암을 포함한다. 제1 표적화 벡터의 3' 상동성 아암은 제2 표적화 벡터의 5' 상동성 아암과 중복되는 서열 (즉, 중복 서열)을 포함하며, 이는 제1의 LTVEC와 제2의 LTVEC 사이의 상동 재조합을 가능하게 한다.

[0216] 이중 표적화 방법의 경우, 제1 표적화 벡터의 5' 상동성 아암과 제2 표적화 벡터의 3' 상동성 아암은 표적 게놈 유전자좌 (즉, 표적 서열) 내의 상응하는 세그먼트들에 상동성이며, 이는 제1 표적화 벡터 및 제2 표적화 벡터의 상응하는 게놈 세그먼트들과의 상동 재조합을 촉진시키고 표적 게놈 유전자좌를 변형시킨다.

[0217] 삼중 표적화 방법의 경우, 제2 표적화 벡터의 3' 상동성 아암은 제3 표적화 벡터의 5' 상동성 아암과 중복되는 서열 (즉, 중복 서열)을 포함하며, 이는 제2의 LTVEC와 제3의 LTVEC 사이의 상동 재조합을 가능하게 한다. 제1 표적화 벡터의 5' 상동성 아암과 제3 표적화 벡터의 3' 상동성 아암은 표적 게놈 유전자좌 (즉, 표적 서열) 내의 상응하는 세그먼트들에 상동성이며, 이는 제1 표적화 벡터 및 제3 표적화 벡터의 상응하는 게놈 세그먼트들과의 상동 재조합을 촉진시키고 표적 게놈 유전자좌를 변형시킨다.

[0218] 상동성 아암과 표적 서열 또는 두 상동성 아암은, 두 영역이 상동 재조합 반응에 대한 기질(substrate)로 작용하기에 서로 충분한 수준의 서열 동일성을 공유할 때, 서로 "상응한다"거나 "상응하고 있다"고 한다. 용어 "상동성"은 상응하는 서열과 동일하거나 이와 서열 동일성을 공유하는 DNA 서열을 포함한다. 주어진 표적 서열과 표적화 벡터에서 발견되는 상응하는 상동성 아암 사이 (즉, 중복 서열) 또는 두 상동성 아암 사이의 서열 동일성은 상동 재조합이 발생할 수 있게 하는 임의의 정도의 서열 동일성일 수 있다. 예를 들어, 표적화 벡터의 상

동성 아암 (또는 이의 단편)과 다른 표적화 벡터의 표적 서열 또는 표적 게놈 유전자좌의 표적 서열 (또는 이의 단편)이 공유하는 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%의 서열 동일성일 수 있어서, 서열들은 상동 재조합을 일으킨다.

[0219] 더욱이, 상동성 아암과 상응하는 표적 서열 사이의 상응하는 상동성 영역은 표적 게놈 유전자좌에서 상동 재조합을 촉진시키기에 충분한 임의의 길이를 가질 수 있다. 예를 들어, 주어진 상동성 아암 및/또는 상응하는 표적 서열은 (예컨대, 본 명세서의 다른 곳에 기술된 LTVEC 벡터에 기재된 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 kb의 길이 또는 그 이상인 상응하는 상동성 영역을 포함할 수 있어서, 상동성 아암이 세포의 표적 게놈 유전자좌 내의 또는 다른 표적화 벡터 내의 상응하는 표적 서열과의 상동 재조합을 일으키기에 충분한 상동성을 갖게 된다.

[0220] 제1 표적화 벡터의 3' 상동성 아암과 제2 표적화 벡터의 5' 상동성 아암의 중복 서열 또는 제2 표적화 벡터의 3' 상동성 아암과 제3 표적화 벡터의 5' 상동성 아암의 중복 서열은 표적화 벡터들 사이의 상동 재조합을 촉진시키기에 충분한 임의의 길이를 가질 수 있다. 예를 들어, 상동성 아암의 주어진 중복 서열은 적어도 약 1 내지 5 kb, 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 kb의 길이 또는 그 이상인 상응하는 중복 영역을 포함할 수 있어서, 상동성 아암의 중복 서열이 다른 표적화 벡터 내의 상응하는 중복 서열과의 상동 재조합을 일으키기에 충분한 상동성을 갖게 된다. 일 실시 형태에서, 중복 서열은 1 내지 5 kb이다. 일 실시 형태에서, 중복 서열은 약 1 kb 내지 약 70 kb이다. 일 실시 형태에서, 중복 서열은 약 10 kb 내지 약 70 kb이다. 다른 실시 형태에서, 중복 서열은 약 10 kb 내지 약 50 kb이다. 일 실시 형태에서, 중복 서열은 적어도 10 kb이다. 다른 실시 형태에서, 중복 서열은 적어도 20 kb이다. 예를 들어, 중복 서열은 약 1 kb 내지 약 5 kb, 약 5 kb 내지 약 10 kb, 약 10 kb 내지 약 15 kb, 약 15 kb 내지 약 20 kb, 약 20 kb 내지 약 25 kb, 약 25 kb 내지 약 30 kb, 약 30 kb 내지 약 35 kb, 약 35 kb 내지 약 40 kb, 약 40 kb 내지 약 45 kb, 약 45 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 120 kb, 약 120 kb 내지 약 140 kb, 약 140 kb 내지 약 160 kb, 약 160 kb 내지 약 180 kb, 약 180 kb 내지 약 200 kb, 약 200 kb 내지 약 220 kb, 약 220 kb 내지 약 240 kb, 약 240 kb 내지 약 260 kb, 약 260 kb 내지 약 280 kb, 또는 약 280 kb 내지 약 300 kb일 수 있다. 예로서, 중복 서열은 약 20 kb 내지 약 60 kb일 수 있다. 대안적으로, 중복 서열은 적어도 1 kb, 적어도 5 kb, 적어도 10 kb, 적어도 15 kb, 적어도 20 kb, 적어도 25 kb, 적어도 30 kb, 적어도 35 kb, 적어도 40 kb, 적어도 45 kb, 적어도 50 kb, 적어도 60 kb, 적어도 70 kb, 적어도 80 kb, 적어도 90 kb, 적어도 100 kb, 적어도 120 kb, 적어도 140 kb, 적어도 160 kb, 적어도 180 kb, 적어도 200 kb, 적어도 220 kb, 적어도 240 kb, 적어도 260 kb, 적어도 280 kb 또는 적어도 300 kb일 수 있다.

[0221] 상동성 아암은 세포에 고유한 유전자좌 (예를 들어, 표적화된 유전자좌)에 상응할 수 있거나, 대안적으로 이것은, 예를 들어 도입유전자, 발현 카세트, 또는 이중성 또는 외인성 DNA 영역을 포함하는, 세포의 게놈 내로 통합된 이중성 또는 외인성 DNA 세그먼트의 영역에 상응할 수 있다. 대안적으로, 상동성 아암은 세포 내의 표적화 벡터 상의 영역에 상응할 수 있다. 표적화 벡터의 상동성 아암은 효모 인공 염색체 (YAC), 세균 인공 염색체 (BAC), 인간 인공 염색체의 영역, 또는 적절한 숙주 세포에 함유된 임의의 다른 유전자 조작된 영역에 상응할 수 있다. 또한, 표적화 벡터의 상동성 아암은 BAC 라이브러리, 코스미드(cosmid) 라이브러리 또는 P1 파지 라이브러리의 영역에 상응하거나 이로부터 유래될 수 있다. 특정 경우에, 표적화 벡터의 상동성 아암은 원핵생물, 효모, 조류 (예를 들어, 닭), 비인간 포유류, 설치류, 인간, 래트, 마우스, 햄스터, 토끼, 피그, 소, 사슴, 양, 염소, 고양이, 개, 흰족제비, 영장류 (예를 들어, 마모셋, 붉은털원숭이), 가축 포유류, 농업용 포유류, 또는 임의의 다른 관심 유기체에 대해 고유하거나, 이중성이거나 외인성인 유전자좌에 상응한다. 일부의 경우에, 상동성 아암은, 뉴클레아제 작용제 (예를 들어, Cas 단백질)에 의해 유도된 Nick 또는 이중 가닥 절단부의 부재 하에, 종래의 방법을 사용하여 표적화될 수 없거나 단지 부정확하게 또는 단지 유의하게 저효율로 표적화될 수 있는 세포의 유전자좌에 상응한다. 일부의 경우에, 상동성 아암은 합성 DNA로부터 유래된다.

[0222] 일부 표적화 벡터의 경우, 5' 상동성 아암 또는 3' 상동성 아암 중 하나는 표적화된 게놈 유전자좌에 상응하는 한편, 5' 상동성 아암 또는 3' 상동성 아암 중 다른 하나는 다른 표적화 벡터 상의 영역에 상응한다.

- [0223] 일부 표적화 벡터의 경우, 5' 상동성 아암과 3' 상동성 아암은 표적화된 게놈에 상응한다. 대안적으로, 상동성 아암은 관련 게놈으로부터 유래될 수 있다. 예를 들어, 표적화된 게놈은 제1 계통의 마우스 게놈이고, 표적화 아암은 제2 계통의 마우스 게놈으로부터 유래되며, 여기서 제1 계통과 제2 계통은 상이하다. 특정 경우에, 상동성 아암은 동일한 동물의 게놈으로부터 유래되거나 동일한 계통의 게놈으로부터 유래되며, 예를 들어 표적화된 게놈은 제1 계통의 마우스 게놈이고, 표적화 아암은 동일한 마우스로부터 또는 동일한 계통으로부터의 마우스 게놈으로부터 유래된다.
- [0224] 표적화 벡터의 상동성 아암은, 예를 들어 적어도 1 내지 5 kb, 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 kb의 길이 또는 그 이상을 포함하는, 상응하는 표적 서열과의 상동 재조합 사건을 촉진시키기에 충분한 임의의 길이를 가질 수 있다. 하기에 더욱 상세히 기술되는 바와 같이, 큰 표적화 벡터는 길이가 더 긴 표적화 아암을 사용할 수 있다.
- [0225] 뉴클레아제 작용제 (예를 들어, CRISPR/Cas 시스템)가 표적화 벡터와 조합하여 사용되어 표적 유전자좌의 변형을 도울 수 있다. 이러한 뉴클레아제 작용제는 표적화 벡터와 표적 유전자좌 사이의 상동 재조합을 촉진시킬 수 있다. 뉴클레아제 작용제가 표적화 벡터와 조합하여 사용될 때, 표적화 벡터는 뉴클레아제 절단 부위에 충분히 근접하여 위치하는 5' 표적 서열과 3' 표적 서열에 상응하는 5' 상동성 아암과 3' 상동성 아암을 포함하여, 뉴클레아제 절단 부위의 Nick 또는 이중 가닥 절단부에서 표적 서열과 상동성 아암 사이의 상동 재조합 사건의 발생을 촉진시킬 수 있다. 용어 "뉴클레아제 절단 부위"는 Nick 또는 이중 가닥 절단부가 뉴클레아제 작용제에 의해 생성되는 DNA 서열 (예를 들어, Cas9 절단 부위)을 포함한다. 표적화 벡터의 5' 상동성 아암과 3' 상동성 아암에 상응하는 표적화된 유전자좌 내의 표적 서열은, 거리가 인식 부위의 Nick 또는 이중 가닥 절단부에서 5' 표적 서열 및 3' 표적 서열과 상동성 아암 사이의 상동 재조합 사건의 발생을 촉진시키는 정도인 경우 뉴클레아제 절단 부위에 "충분히 근접하여 위치한다". 따라서, 특정 경우에, 표적화 벡터의 5' 상동성 아암 및/또는 3' 상동성 아암에 상응하는 표적 서열은 주어진 인식 부위의 적어도 1개의 뉴클레오티드 이내 또는 주어진 인식 부위의 적어도 10개의 뉴클레오티드 내지 약 14 kb 이내에 존재한다. 일부의 경우에, 뉴클레아제 절단 부위는 표적 서열 중 적어도 하나 또는 둘 모두에 바로 인접해 있다.
- [0226] 표적화 벡터의 상동성 아암에 상응하는 표적 서열과 뉴클레아제 절단 부위의 공간적 관계는 다양할 수 있다. 예를 들어, 표적 서열은 뉴클레아제 절단 부위의 5'에 위치할 수 있거나, 표적 서열은 인식 부위의 3'에 위치할 수 있거나, 표적 서열은 뉴클레아제 절단 부위를 플랭킹할 수 있다.
- [0227] (예를 들어, 큰 표적화 벡터를 포함하는) 표적화 벡터와 뉴클레아제 작용제의 병용은 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 벡터의 단독 사용과 비교할 때 적어도 2배, 적어도 3배, 적어도 4배, 적어도 10배 또는 이들 정수로부터 형성된 범위, 예컨대 2배 내지 10배 이내로 증가될 수 있다.
- [0228] **D. 큰 표적화 벡터**
- [0229] 일부 표적화 벡터는 "큰 표적화 벡터" 또는 "LTVEC"이며, 이는 세포에서 상동 재조합을 수행하고자 하는 다른 접근법에 의해 전형적으로 사용되는 것보다 큰 핵산 서열에 상응하며 이로부터 유래되는 상동성 아암을 포함하는 표적화 벡터를 포함한다. LTVEC는, 예를 들어 길이가 적어도 10 kb일 수 있거나, 5' 상동성 아암과 3' 상동성 아암의 총합은, 예를 들어 적어도 10 kb일 수 있다. 또한, LTVEC는 세포에서 상동 재조합을 수행하고자 하는 다른 접근법에 의해 전형적으로 사용되는 것보다 큰 핵산 서열을 갖는 핵산 삽입물을 포함하는 표적화 벡터를 포함한다. 예를 들어, LTVEC는, 크기 제한 때문에 전통적인 플라스미드 기반 표적화 벡터에 의해 수용될 수 없는 큰 유전자좌의 변형을 가능하게 한다. 예를 들어, 표적화된 유전자좌는 (즉, 5' 및 3' 상동성 아암은) 뉴클레아제 작용제 (예를 들어, Cas 단백질)에 의해 유도된 Nick 또는 이중 가닥 절단부의 부재 하에, 종래의 방법을 사용하여 표적화될 수 없거나 단지 부정확하게 또는 단지 유의하게 저효율로 표적화될 수 있는 세포의 유전자좌일 수 있다 (즉, 그 유전자좌에 상응할 수 있다).
- [0230] 본 명세서에 제공된 방법은 본 명세서의 다른 곳에 기술된 바와 같이 3 방향 또는 4 방향 재조합 사건으로 서로 그리고 표적 게놈 유전자좌와 재조합가능한 2개 또는 3개의 LTVEC를 사용한다. 이들 방법은 단일 LTVEC를 사용하여 달성될 수 없는 큰 유전자좌의 변형을 가능하게 한다.
- [0231] LTVEC의 예는 세균 인공 염색체 (BAC), 인간 인공 염색체 또는 효모 인공 염색체 (YAC)로부터 유래된 벡터를 포

함한다. LTVEC의 예 및 이를 제조하는 방법은, 예를 들어 각각 본 명세서에 참고로 포함되는, 미국 특허 제 6,586,251호, 미국 특허 제6,596,541호, 미국 특허 제7,105,348호 및 국제특허 공개 WO 2002/036789호 (PCT/US01/45375호)에 기술되어 있다. LTVEC는 선형 형태 또는 원형 형태일 수 있다.

[0232] LTVEC는, 예를 들어 약 20 kb 내지 약 300 kb, 약 20 kb 내지 약 50 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 포함하는 임의의 길이를 가질 수 있다. 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(long-range) (예를 들어, 1 kb 내지 5 kb) PCR에 의한 표적화 사건의 스크리닝을 가능하게 할 수 없다.

[0233] 일부의 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 kb의 범위일 수 있다. 일부의 경우에, LTVEC는 약 400 kb 내지 약 450 kb, 약 450 kb 내지 약 500 kb, 약 500 kb 내지 약 550 kb, 약 550 kb 내지 약 600 kb, 약 600 kb 내지 약 650 kb, 약 650 kb 내지 약 700 kb, 약 700 kb 내지 약 750 kb, 또는 약 750 kb 내지 약 800 kb 범위의 핵산 삽입물을 포함한다.

[0234] 일부 LTVEC에서, 5' 상동성 아암과 3' 상동성 아암의 총합은 적어도 10 kb이다. 다른 LTVEC에서, 5' 상동성 아암은 약 1 kb 내지 약 100 kb의 범위이고/이거나 3' 상동성 아암은 약 1 kb 내지 약 100 kb의 범위이다. 5' 상동성 아암과 3' 상동성 아암의 총합은, 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 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' 상동성 아암과 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.

[0235] 일부의 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 600 kb, 약 600 kb 내지 약 700 kb, 약 700 kb 내지 약 800 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 내인성 서열의 결실을 가능하게 하도록 설계된다. 대안적으로, 결실은 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. 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.

[0236] 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 삽입을 가능하게 하도록 설계된다. 대안적으로, 삽입은 약 400 kb 내지 약 450 kb, 약 450 kb 내지 약 500 kb, 약 500 kb 내지 약 550 kb, 약 550 kb 내지 약 600 kb, 약 600 kb 내지 약 650 kb, 약 650 kb 내지 약 700 kb, 약 700 kb 내지 약 750 kb, 또는 약 750 kb 내지 약 800 kb일 수 있다. 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.

[0237] 또 다른 경우에, 핵산 삽입물 및/또는 결실되는 내인성 유전자좌의 영역은 적어도 100개, 200개, 300개, 400개, 500개, 600개, 700개, 800개 또는 900개의 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 이상이다.

[0238] **E. 게놈 및 표적 게놈 유전자좌**

[0239] 본 명세서에 개시된 방법에 의해 변형되는 게놈 또는 게놈 표적 유전자좌는 세포 내의 임의의 DNA 세그먼트 또는 영역을 포함할 수 있다. 게놈 또는 게놈 표적 유전자좌는 세포에 고유할 수 있거나 세포의 게놈 내로 통합된 이중성 또는 외인성 DNA 세그먼트일 수 있거나 이들의 조합일 수 있다. 이러한 이중성 또는 외인성 DNA 세그먼트는 도입유전자, 발현 카세트, 선택 마커를 인코딩하는 폴리뉴클레오티드, 또는 이중성 또는 외인성 게놈 DNA 영역을 포함할 수 있다.

[0240] 또한, 게놈 또는 게놈 표적 유전자좌는 세포 내의 염색체의 DNA, 예를 들어 효모 인공 염색체 (YAC), 세균 인공 염색체 (BAC), 인간 인공 염색체, 또는 적절한 숙주 세포에 함유된 임의의 다른 유전자 조작된 게놈 영역을 포함할 수 있다.

[0241] **III. 뉴클레아제 작용제**

[0242] 본 명세서에 제공된 표적 게놈 유전자좌를 변형시키는 방법 및 조성물은 요망되는 인식 부위에서 닉 또는 이중 가닥 절단을 유도하는 뉴클레아제 작용제를 사용할 수 있다.

[0243] 용어 "뉴클레아제 작용제에 대한 인식 부위"는 닉 또는 이중 가닥 절단이 뉴클레아제 작용제에 의해 유도되는 DNA 서열을 포함한다. 뉴클레아제 작용제에 대한 인식 부위는 세포에 내인성(또는 고유한 것)일 수 있거나, 인식 부위는 세포에 외인성일 수 있다. 특정 실시 형태에서, 인식 부위는 세포에 외인성이므로, 세포의 게놈에서 자연적으로 발생하지 않는다. 또 다른 추가의 실시 형태에서, 인식 부위는 세포에, 그리고 표적 유전자좌에 위치시키고자 하는 관심 폴리뉴클레오티드에 외인성이다. 추가의 실시 형태에서, 외인성 또는 내인성 인식 부위는 숙주 세포의 게놈에 단 한번만 존재한다. 특정 실시 형태에서, 게놈 내에 단 한번만 발생하는 내인성 또는 고유한 부위가 확인된다. 이어서, 이러한 부위를 사용하여, 내인성 인식 부위에 닉 또는 이중 가닥 절단부를 생성시킬 뉴클레아제 작용제를 설계할 수 있다.

[0244] 인식 부위의 길이는 달라질 수 있으며, 예를 들어 징크 핑거 뉴클레아제 (ZFN) 쌍에 대하여 약 30 내지 36 bp (즉, 각 ZFN에 대하여 약 15 내지 18 bp)이거나, 전자 활성화인자 유사 이펙터 뉴클레아제 (TALEN)에 대하여 약 36 bp이거나, CRISPR/Cas9 가이드 RNA에 대하여 약 20 bp인 인식 부위를 포함한다.

[0245] 닉 또는 이중 가닥 절단을 요망되는 인식 부위에 유도하는 임의의 뉴클레아제 작용제를 본 명세서에 개시된 방법 및 조성물에 사용할 수 있다. 자연적으로 발생하는 또는 천연 뉴클레아제 작용제가 요망되는 인식 부위에서 닉 또는 이중 가닥 절단을 유도하는 한, 그러한 뉴클레아제 작용제를 사용할 수 있다. 대안적으로, 변형되거나 유전자 조작된 뉴클레아제 작용제를 사용할 수 있다. "유전자 조작된 뉴클레아제 작용제"는 요망되는 인식 부위를 특이적으로 인식하여 여기에서 닉 또는 이중 가닥 절단을 유도하도록 그의 천연 형태로부터 유전자 조작된 (변형 또는 유래된) 뉴클레아제를 포함한다. 따라서, 유전자 조작된 뉴클레아제 작용제는 천연의 자연적으로 발생하는 뉴클레아제 작용제로부터 유래될 수 있거나, 이것은 인공적으로 생성 또는 합성될 수 있다. 뉴클레아제 작용제의 변형은 단백질 절단제(cleavage agent)의 하나의 아미노산 또는 핵산 절단제의 하나의 뉴클레오티드만큼 적을 수 있다. 일부 실시 형태에서, 유전자 조작된 뉴클레아제는 인식 부위에서 닉 또는 이중 가닥 절단을 유도하며, 여기서 인식 부위는 천연 (유전자 조작되지 않거나 변형되지 않은) 뉴클레아제 작용제에 의해

인식되었을 서열이 아니었다. 인식 부위 또는 다른 DNA에 Nick 또는 이중 가닥 절단부를 형성하는 것은 본 명세서에서 인식 부위 또는 다른 DNA의 "컷팅" 또는 "절단"으로 지칭될 수 있다.

[0246] 예시된 인식 부위의 활성 변이체 및 단편도 제공된다. 이러한 활성 변이체는 주어진 인식 부위에 대하여, 적어도 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 포함할 수 있으며, 여기서 활성 변이체는 생물학적 활성을 보유하므로, 서열 특이적 방식으로 뉴클레아제 작용제에 의해 인식되고 절단될 수 있다. 뉴클레아제 작용제에 의한 인식 부위의 이중 가닥 절단을 측정하는 검정은 당업계에 공지되어 있다 (예를 들어, TAQMAN® qPCR 검정, 전체적으로 본 명세서에 참고로 포함되는 문헌 [Frendewey D. et al., *Methods in Enzymology*, 2010, 476:295-307]).

[0247] 뉴클레아제 작용제의 인식 부위는 표적 유전자와 내의 또는 이 부근의 임의의 위치에 위치할 수 있다. 인식 부위는 유전자의 코딩 영역 내에 또는 그 유전자의 발현에 영향을 미치는 조절 영역 내에 위치할 수 있다. 뉴클레아제 작용제의 인식 부위는 인트론, 엑손, 프로모터, 인핸서, 조절 영역, 또는 임의의 비단백질 코딩 영역에 위치할 수 있다. 특정 실시 형태에서, 인식 부위는 선택 마커를 인코딩하는 폴리뉴클레오티드 내에 위치한다. 이러한 위치는 선택 마커의 코딩 영역 내에 또는 선택 마커의 발현에 영향을 미치는 조절 영역 내에 위치할 수 있다. 따라서, 뉴클레아제 작용제의 인식 부위는 선택 마커의 인트론, 프로모터, 인핸서, 조절 영역, 또는 선택 마커를 인코딩하는 폴리뉴클레오티드의 임의의 비단백질 코딩 영역에 위치할 수 있다. 특정 실시 형태에서, 인식 부위의 Nick 또는 이중 가닥 절단부는 선택 마커의 활성을 파괴한다. 기능적 선택 마커의 존재 또는 부재의 검정 방법은 알려져 있다.

[0248] 일 실시 형태에서, 뉴클레아제 작용제는 전사 활성화인자 유사 이펙터 뉴클레아제 (TALEN)이다. TAL 이펙터 뉴클레아제는 원핵생물 또는 진핵생물 유기체의 게놈의 특정 표적 서열에 이중 가닥 절단부를 생성시키는 데 사용될 수 있는 서열 특이적 뉴클레아제의 부류이다. TAL 이펙터 뉴클레아제는 천연 또는 유전자 조작된 전사 활성화인자 유사 (TAL) 이펙터 또는 이의 기능적 부분을 엔도뉴클레아제, 예를 들어 FokI의 촉매 도메인에 융합시킴으로써 생성된다. 독특한 모듈식 TAL 이펙터 DNA 결합 도메인은 잠재적으로 임의의 주어진 DNA 인식 특이성을 갖는 단백질의 설계를 가능하게 한다. 따라서, TAL 이펙터 뉴클레아제의 DNA 결합 도메인은 특정 DNA 표적 부위를 인식하도록 유전자 조작될 수 있으므로, 이를 사용하여 요망되는 표적 서열에 이중 가닥 절단부를 생성시킬 수 있다. 국제특허 공개 WO 2010/079430호; 문헌[Morbiter et al. (2010) PNAS 10.1073/pnas.1013133107]; 문헌[Scholze & Boch (2010) *Virulence* 1:428-432]; 문헌[Christian et al. *Genetics* (2010) 186:757-761]; 문헌[Li et al. (2010) *Nuc. Acids Res.* (2010) doi:10.1093/nar/gkq704]; 및 문헌[Miller et al. (2011) *Nature Biotechnology* 29:143-148]이 참조되며, 이들 모두는 본 명세서에 참고로 포함된다.

[0249] 적합한 TAL 뉴클레아제의 예 및 적합한 TAL 뉴클레아제의 제조 방법이, 예를 들어 미국 특허 출원 공개 제 2011/0239315 A1호, 미국 특허 출원 공개 제 2011/0269234 A1호, 미국 특허 출원 공개 제 2011/0145940 A1호, 미국 특허 출원 공개 제 2003/0232410 A1호, 미국 특허 출원 공개 제 2005/0208489 A1호, 미국 특허 출원 공개 제 2005/0026157 A1호, 미국 특허 출원 공개 제 2005/0064474 A1호, 미국 특허 출원 공개 제 2006/0188987 A1호, 및 미국 특허 출원 공개 제 2006/0063231 A1호 (각각 본 명세서에 참고로 포함됨)에 개시되어 있다. 다양한 실시 형태에서, TAL 이펙터 뉴클레아제는, 예를 들어 관심 유전자와 또는 관심 게놈 유전자와의 표적 핵산 서열 또는 이 부근을 절단하도록 유전자 조작되며, 여기서 표적 핵산 서열은 표적화 벡터에 의해 변형될 서열에 또는 이 부근에 존재한다. 본 명세서에 제공된 다양한 방법 및 조성물에 사용하기에 적합한 TAL 뉴클레아제는 본 명세서에 기술된 바와 같은 표적화 벡터에 의해 변형될 표적 핵산 서열에 또는 이 부근에 결합하도록 특별히 설계된 것을 포함한다.

[0250] 일 실시 형태에서, TALEN의 각 단량체는 2개의 초가변 잔기(hypervariable residue)를 통해 단일 염기쌍을 인식하는 33개 내지 35개의 TAL 반복체를 포함한다. 일 실시 형태에서, 뉴클레아제 작용제는 독립적인 뉴클레아제에 작동가능하게 연결된 TAL 반복체 기반 DNA 결합 도메인을 포함하는 키메라 단백질이다. 일 실시 형태에서, 독립적인 뉴클레아제는 FokI 엔도뉴클레아제이다. 일 실시 형태에서, 뉴클레아제 작용제는 제1 TAL 반복체 기반 DNA 결합 도메인 및 제2 TAL 반복체 기반 DNA 결합 도메인을 포함하며, 여기서 제1 및 제2 TAL 반복체 기반 DNA 결합 도메인은 각각, FokI 뉴클레아제에 작동가능하게 연결되고, 제1 및 제2 TAL 반복체 기반 DNA 결합 도메인은 가변 길이 (12 내지 20 bp)의 스페이서 서열에 의해 분리된 표적 DNA 서열의 각 가닥에서 2개의 연속 표적 DNA 서열을 인식하고, FokI 뉴클레아제 서브유닛은 이량체화되어, 표적 서열에 이중 가닥 절단부를 생성시키는 활성 뉴클레아제를 생성시킨다.

[0251] 본 명세서에 개시된 다양한 방법 및 조성물에 사용되는 뉴클레아제 작용제는 징크 핑거 뉴클레아제 (ZFN)를 추가로 포함할 수 있다. 일 실시 형태에서, ZFN의 각 단량체는 3개 이상의 징크 핑거 기반 DNA 결합 도메인을 포함하며, 여기서 각 징크 핑거 기반 DNA 결합 도메인은 3 bp 서브사이트(subsite)에 결합한다. 다른 실시 형태에서, ZFN은 독립적인 뉴클레아제에 작동가능하게 연결된 징크 핑거 기반 DNA 결합 도메인을 포함하는 키메라 단백질이다. 일 실시 형태에서, 독립적인 엔도뉴클레아제는 FokI 엔도뉴클레아제이다. 일 실시 형태에서, 뉴클레아제 작용제는 제1 ZFN 및 제2 ZFN을 포함하고, 여기서 제1 ZFN 및 제2 ZFN 각각은 FokI 뉴클레아제 서브유닛에 작동가능하게 연결되고, 제1 및 제2 ZFN은 표적 DNA 서열의 각 가닥에서 약 5 내지 7 bp 스페이스에 의해 분리된 2개의 연속 표적 DNA 서열을 인식하고, FokI 뉴클레아제 서브유닛은 이량체화되어, 이중 가닥 절단부를 생성시키는 활성 뉴클레아제를 생성시킨다. 예를 들어, 미국 특허 출원 공개 제20060246567호; 미국 특허 출원 공개 제20080182332호; 미국 특허 출원 공개 제20020081614호; 미국 특허 출원 공개 제20030021776호; 국제특허 공개 WO/2002/057308A2호; 미국 특허 출원 공개 제20130123484호; 미국 특허 출원 공개 제20100291048호; 국제특허 공개 WO/2011/017293A2호; 및 문헌[Gaj *et al.* (2013) *Trends in Biotechnology*, 31(7):397-405]이 참조되며, 이들 각각은 본 명세서에 참고로 포함된다.

[0252] 또 다른 실시 형태에서, 뉴클레아제 작용제는 메가뉴클레아제이다. 메가뉴클레아제는 보존된 서열 모티프에 기반하여 4개의 패밀리로 분류되어 왔는데, 이들 패밀리는 LAGLIDADG, GIY-YIG, H-N-H, 및 His-Cys 박스 패밀리아이다. 이들 모티프는 금속 이온들의 배위(coordination)와 포스포다이에스테르 결합의 가수분해에 참여한다. 메가뉴클레아제는 이의 긴 인식 부위, 및 이의 DNA 기질에서의 일부 서열 다형성을 용인한다는 점이 주목할 만하다. 메가뉴클레아제 도메인, 구조 및 기능은 공지되어 있으며, 예를 들어 문헌[Guhan and Muniyappa (2003) *Crit Rev Biochem Mol Biol* 38:199-248]; 문헌[Lucas *et al.*, (2001) *Nucleic Acids Res* 29:960-9]; 문헌 [Jurica and Stoddard, (1999) *Cell Mol Life Sci* 55:1304-26]; 문헌[Stoddard, (2006) *Q Rev Biophys* 38:49-95]; 및 문헌[Moure *et al.*, (2002) *Nat Struct Biol* 9:764]이 참조된다. 일부 예에서, 자연적으로 발생하는 변이체 및/또는 유전자 조작된 유도체 메가뉴클레아제를 사용한다. 동태(kinetics), 보조인자 상호작용, 발현, 최적 조건 및/또는 인식 부위 특이성을 변형시키고 활성을 스크리닝하는 방법이 공지되어 있으며, 예를 들어 문헌[Epinat *et al.*, (2003) *Nucleic Acids Res* 31:2952-62]; 문헌[Chevalier *et al.*, (2002) *Mol Cell* 10:895-905]; 문헌[Gimble *et al.*, (2003) *Mol Biol* 334:993-1008]; 문헌[Seligman *et al.*, (2002) *Nucleic Acids Res* 30:3870-9]; 문헌[Sussman *et al.*, (2004) *J Mol Biol* 342:31-41]; 문헌[Rosen *et al.*, (2006) *Nucleic Acids Res* 34:4791-800]; 문헌[Chames *et al.*, (2005) *Nucleic Acids Res* 33:e178]; 문헌[Smith *et al.*, (2006) *Nucleic Acids Res* 34:e149]; 문헌[Gruen *et al.*, (2002) *Nucleic Acids Res* 30:e29]; 문헌[Chen and Zhao, (2005) *Nucleic Acids Res* 33:e154]; 국제특허 공개 WO2005105989호; 국제특허 공개 WO2003078619호; 국제특허 공개 WO2006097854호; 국제특허 WO2006097853호; 국제특허 WO2006097784호; 및 국제특허 공개 WO2004031346호가 참조된다.

[0253] I-SceI, I-SceII, I-SceIII, I-SceIV, I-SceV, I-SceVI, I-SceVII, I-CeuI, I-CeuAIIP, I-CreI, I-CrepsbIP, I-CrepsbIIP, I-CrepsbIIIP, I-CrepsbIIVP, I-TliI, I-PpoI, PI-PspI, F-SceI, F-SceII, F-SuvI, F-TevI, F-TevII, I-AmaI, I-AniI, I-ChuI, I-Cmoel, I-CpaI, I-CpaII, I-CsmI, I-CvuI, I-CvuAIP, I-DdiI, I-DdiII, I-DirI, I-DmoI, I-HmuI, I-HmuII, I-HsNIP, I-LlaI, I-MsoI, I-NaaI, I-NanI, I-NcIIP, I-NgrIP, I-NitI, I-NjaI, I-Nsp236IP, I-PakI, I-PboIP, I-PcuIP, I-PcuAI, I-PcuVI, I-PgrIP, I-PobIP, I-PorI, I-PorIIP, I-PbpIP, I-SpBetaIP, I-ScaI, I-SexIP, I-SneIP, I-SpomI, I-SpomCP, I-SpomIP, I-SpomIIP, I-SquIP, I-Ssp6803I, I-SthPhiJP, I-SthPhiST3P, I-SthPhiSTe3bP, I-TdeIP, I-TevI, I-TevII, I-TevIII, I-UarAP, I-UarHGPAIP, I-UarHGPA13P, I-VinIP, I-ZbiIP, PI-MtuI, PI-MtuHIP, PI-MtuHIIP, PI-PfuI, PI-PfuII, PI-PkoI, PI-PkoII, PI-Rma43812IP, PI-SpBetaIP, PI-SceI, PI-TfuI, PI-TfuII, PI-ThyI, PI-TliI, PI-TliII, 또는 이들의 임의의 활성 변이체 또는 단편을 포함하지만 이에 한정되지 않는 임의의 메가뉴클레아제를 본 발명에 사용할 수 있다.

[0254] 일 실시 형태에서, 메가뉴클레아제는 12 내지 40개의 염기쌍의 이중 가닥 DNA 서열을 인식한다. 일 실시 형태에서, 메가뉴클레아제는 게놈의 하나의 완전히 매칭된 표적 서열을 인식한다. 일 실시 형태에서, 메가뉴클레아제는 호밍(homing) 뉴클레아제이다. 일 실시 형태에서, 호밍 뉴클레아제는 LAGLIDADG 패밀리의 호밍 뉴클레아제이다. 일 실시 형태에서, LAGLIDADG 패밀리의 호밍 뉴클레아제는 I-SceI, I-CreI 및 I-DmoI로부터 선택된다.

[0255] 뉴클레아제 작용제는 타입 I, 타입 II, 타입 III 및 타입 IV 엔도뉴클레아제를 포함하는 제한 엔도뉴클레아제를 추가로 포함할 수 있다. 타입 I 및 타입 III 제한 엔도뉴클레아제는 특정 인식 부위를 인식하지만, 전형적으로 뉴클레아제 결합 부위의 가변 위치에서 절단하며, 이는 절단 부위 (인식 부위)로부터 수백 염기쌍만큼 떨어져

있을 수 있다. 타입 II 시스템에서, 제한 활성은 임의의 메틸라제 활성과 무관하며, 절단은 전형적으로 결합 부위 내 또는 이 부근의 특정 부위에서 발생한다. 대부분의 타입 II 효소는 회문 서열을 절단하지만, 타입 IIa 효소는 비회문 인식 부위를 인식하여 인식 부위의 외측을 절단하고, 타입 IIb 효소는 인식 부위 외측의 양측 부위에서 서열을 2회 절단하고, 타입 IIc 효소는 비대칭 인식 부위를 인식하여 일측에서 그리고 인식 부위로부터 약 1개 내지 20개의 뉴클레오티드의 한정된 거리에서 절단한다. 타입 IV 제한효소는 메틸화 DNA를 표적화한다. 제한 효소는, 예를 들어 REBASE 데이터베이스 (웹페이지: rebase.neb.com; 문헌[Roberts et al., (2003) Nucleic Acids Res 31:418-20]), 문헌[Roberts et al., (2003) Nucleic Acids Res 31:1805-12], 및 문헌[Belfort et al., (2002) in Mobile DNA II, pp. 761-783], 문헌[Eds. Craigie et al., (ASM Press, Washington, DC)]에 추가로 설명되고 분류되어 있다.

[0256] 다양한 방법 및 조성물에 사용되는 뉴클레아제 작용제는 또한 CRISPR/CRISPR 관련 (Cas) 시스템 또는 이러한 시스템의 성분을 포함할 수 있다. CRISPR/Cas 시스템은 Cas 유전자의 발현 또는 활성 유도에 관여하는 전사물 및 다른 요소를 포함한다. CRISPR/Cas 시스템은 타입 I, 타입 II 또는 타입 III 시스템일 수 있다. 본 명세서에 개시된 방법 및 조성물은 핵산의 부위 특이적 절단을 위해 CRISPR 복합체 (Cas 단백질과 복합체를 형성한 가이드 RNA (gRNA)를 포함함)를 이용함으로써 CRISPR/Cas 시스템을 사용한다.

[0257] 본 명세서에 개시된 방법에 사용되는 일부 CRISPR/Cas 시스템은 자연적으로 발생하지 않는다. "자연적으로 발생하지 않는" 시스템은 인간의 손의 개입을 나타내는 모든 것, 예를 들어 시스템의 하나 이상의 성분이, 이들의 자연적으로 발생하는 상태로부터 변경되거나 돌연변이되거나, 이들이 자연 상태에서 자연적으로 결합되어 있는 적어도 하나의 다른 성분을 적어도 실질적으로 함유하지 않거나, 이들이 자연적으로 결합되어 있지 않은 적어도 하나의 다른 구성 요소와 결합되어 있는 것을 포함한다. 예를 들어, 일부 CRISPR/Cas 시스템은, 동시에 자연적으로 발생하지 않는 gRNA 및 Cas 단백질을 포함하는, 자연적으로 발생하지 않는 CRISPR 복합체를 사용한다.

[0258] Cas 단백질은 대체로 적어도 하나의 RNA 인식 또는 결합 도메인을 포함한다. 이러한 도메인은 가이드 RNA (gRNA, 하기에 더욱 상세히 기술됨)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 지닌다. 절단은 핵산 분자의 공유 결합의 파괴를 포함한다. 절단은 평활(blunt) 말단 또는 스테거된(staggered) 말단을 생성시킬 수 있으며, 단일 가닥 또는 이중 가닥으로 될 수 있다. Cas 단백질은 완전한 절단 활성을 가져서 표적 게놈 유전자좌에 이중 가닥 절단부 (예를 들어, 평활 말단을 갖는 이중 가닥 절단부)를 생성시킬 수 있거나, 이것은 표적 게놈 유전자좌에 단일 가닥 절단부를 생성시키는 니카제(nickase)일 수 있다.

[0259] 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 호모로그(homolog) 또는 변형된 버전을 포함한다.

[0260] 일부의 경우에, Cas 단백질은 타입 II CRISPR/Cas 시스템이다. 예를 들어, Cas 단백질은 Cas9 단백질일 수 있거나 Cas9 단백질로부터 유래될 수 있다. Cas9 단백질은 전형적으로 보존 구조를 갖는 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.*), 노카르디옵시스 다손빌레이(*Nocardia dasonvillei*), 스트렙토미세스 프리스티나이스피랄리스(*Streptomyces pristinaespiralis*), 스트렙토미세스 비리도크로모게네스(*Streptomyces viridochromogenes*), 스트렙토미세스 비리도크로모게네스, 스트렙토스포랑기움 로세움(*Streptosporangium roseum*), 스트렙토스포랑기움 로세움, 알리시클로바실루스 아시도칼다리우스(*Alicyclobacillus acidocaldarius*), 바실루스 슈도미코이데스(*Bacillus pseudomycooides*), 바실루스 셀레니티레두센스(*Bacillus selenitireducens*), 엑시구오박테리움 시비리쿰(*Exiguobacterium sibiricum*), 락토바실루스 델브루엑키(*Lactobacillus delbrueckii*), 락토바실루스 살리바리우스(*Lactobacillus salivarius*), 마이크로스킬라 마리나(*Microscilla marina*), 부르크홀데리아레스 박테리움(*Burkholderiales bacterium*), 폴라로모나스 나프탈레니보란스(*Polaromonas naphthalenivorans*), 폴라로모나스 종(*Polaromonas sp.*), 크로코스파이라 왓슨이(*Crocospaera watsonii*), 시아노테세 종(*Cyanothecce sp.*), 마이크로시스티스 아이루기노사(*Microcystis aeruginosa*), 시네토크쿠스 종(*Synechococcus sp.*), 아세토할로비움 아라바티쿰(*Acetohalobium arabaticum*), 암모니펙스 데겐시이(*Ammonifex degensii*), 칼디세룰로시류프토르 벅스키이(*Caldicelulosiruptor becskii*), 칸디

다투스 데술포루디스(*Candidatus Desulforudis*), 클로스트리디움 보툴리눔(*Clostridium botulinum*), 클로스트리디움 디피실레(*Clostridium difficile*), 피네골디아 마그나(*Fingoldia magna*), 나트라나이로비우스 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*), 크테도노박테르 라세미페르(*Ktedonobacter racemifer*), 메타노할로비움 에베스티가툼(*Methanohalobium evestigatum*), 아나바이나 바리아빌리스(*Anabaena variabilis*), 노둘라리아 스푸미게나(*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(*Petrotoga mobilis*), 테르모시포 아프리카누스(*Thermosiphon africanus*) 또는 아카리오클로리스 마리나(*Acaryochloris marina*)로부터 유래될 수 있다. Cas9 단백질은 또한 스태필로코쿠스 아우레우스(*Staphylococcus aureus*)로부터 유래될 수 있다. Cas9 패밀리 구성원의 추가의 예는 전체적으로 본 명세서에 참고로 포함되는 국제특허 공개 WO 2014/131833호에 기술된 것을 포함한다. 특정 예에서, Cas9 단백질은 S. 피오게네스 유래의 Cas9 단백질이거나 이로부터 유래된다. S. 피오게네스 유래의 Cas9 단백질의 아미노산 서열은, 예를 들어 수탁 번호 Q99ZW2로 SwissProt 데이터베이스에서 찾아볼 수 있다.

[0261] Cas 단백질은 야생형 단백질 (즉, 자연 상태에서 발생하는 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 이의 일부에 대하여, 적어도 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 포함할 수 있으며, 여기서 활성 변이체는 요망되는 절단 부위에서 절단하는 능력을 보유하므로, 닉 유도 또는 이중 가닥 절단 유도 활성을 보유한다. 닉 유도 또는 이중 가닥 절단 유도 활성의 검정은 공지되어 있으며, 대체로 절단 부위를 함유하는 DNA 기질에 대한 Cas 단백질의 전체 활성 및 특이성을 측정한다.

[0262] Cas 단백질은 핵산 결합 친화성, 핵산 결합 특이성 및/또는 효소 활성을 증가 또는 감소시키도록 변형될 수 있다. Cas 단백질은 또한 단백질의 임의의 다른 활성 또는 특성, 예를 들어 안정성을 변화시키도록 변형될 수 있다. 예를 들어, Cas 단백질의 하나 이상의 뉴클레아제 도메인을 변형, 결실 또는 비활성화시킬 수 있거나, Cas 단백질을 트렁케이션(truncation)하여 단백질의 기능에 필수적이지 않은 도메인을 제거하거나 Cas 단백질의 활성을 최적화 (예를 들어, 향상 또는 저하)시킬 수 있다.

[0263] 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. RuvC와 HNH 도메인 각각은 이중 가닥 DNA의 상이한 가닥을 절단하여 DNA에 이중 가닥 절단부를 생성시킬 수 있다. 예를 들어, 전체적으로 본 명세서에 참고로 포함되는 문헌[Jinek *et al.* (2012) *Science* 337:816-821]이 참조된다.

[0264] 뉴클레아제 도메인들 중 하나 또는 둘 모두는 더 이상 기능적이지 않거나 뉴클레아제 활성이 감소되도록 결실 또는 돌연변이될 수 있다. 뉴클레아제 도메인들 중 하나가 결실 또는 돌연변이되는 경우, 생성된 Cas 단백질 (예를 들어, Cas9)은 닉카제로 지칭될 수 있고, 이중 가닥 DNA 내의 표적 서열에 단일 가닥 절단부를 생성시킬 수 있지만, 이중 가닥 절단부를 생성시킬 수 없다 (즉, 이것은 상보적 가닥 또는 비상보적 가닥 중 어느 한쪽만 절단할 수 있음). 뉴클레아제 도메인들 둘 모두가 결실 또는 돌연변이되는 경우, 생성된 Cas 단백질(예를 들어, Cas9)은 이중 가닥 DNA의 두 가닥을 절단하는 능력이 저하될 것이다 (예를 들어, 뉴클레아제 널(nuclease-null) Cas 단백질). Cas9을 닉카제로 변환시키는 돌연변이의 예는 S. 피오게네스 유래의 Cas9의 RuvC 도메인에서의 D10A (Cas9의 위치 10에서 아스파르트레이트가 알라닌으로 치환됨) 돌연변이이다. 마찬가지로, S. 피오게네스 유래의 Cas9의 HNH 도메인에서의 H939A (아미노산 위치 839에서 히스티딘이 알라닌으로 치환됨) 또는 H840A (아미노산 위치 840에서 히스티딘이 알라닌으로 치환됨)는 Cas9을 닉카제로 변환시킬 수 있다. Cas9을 닉카제로 변환시키는 돌연변이의 다른 예는 S. 테르모필루스 유래의 Cas9에 대한 상응하는 돌연변이를 포함한다. 예를 들어, 각각 전체적으로 본 명세서에 참고로 포함되는 문헌[Sapranaukas *et al.* (2011) *Nucleic Acids Research* 39:9275-9282] 및 국제특허 공개 WO 2013/141680호가 참조된다. 이러한 돌연변이는 부위 특이적 돌연변이 유발, PCR 매개 돌연변이 유발 또는 총 유전자 합성(total gene synthesis)과 같은 잘 알려진 방법을 사용하여 생성될 수 있다. 닉카제를 생성시키는 다른 돌연변이의 예는, 예를 들어, 각각 본 명세서에 참고로 포함되는, 국제특허 공개 WO/2013/176772A1호 및 국제특허 공개 WO/2013/142578A1호에서 찾

아볼 수 있다.

- [0265] Cas 단백질은 또한 융합 단백질일 수 있다. 예를 들어, Cas 단백질은 절단 도메인, 후성적(epigenetic) 변형 도메인, 전사 활성화 도메인 또는 전사 억제인자 도메인에 융합될 수 있다. 전체적으로 본 명세서에 참고로 포함되는 국제특허 공개 WO 2014/089290호가 참조된다. Cas 단백질은 또한 안정성의 증가 또는 감소를 제공하는 이중성 폴리펩티드에 융합될 수 있다. 융합된 도메인 또는 이중성 폴리펩티드는 N 말단, C 말단 또는 Cas 단백질의 내부에 위치할 수 있다.
- [0266] Cas 융합 단백질의 일례는 세포내 국재화(subcellular localization)를 제공하는 이중성 폴리펩티드에 융합된 Cas 단백질이다. 이러한 서열은, 예를 들어 핵에 대한 표적화를 위한 SV40 NLS와 같은 핵 국재화 신호(nuclear localization signal, NLS), 미토콘드리아에 대한 표적화를 위한 미토콘드리아 국재화 신호, ER 보유 신호(retention signal) 등을 포함할 수 있다. 예를 들어, 문헌[Lange *et al.* (2007) *J. Biol. Chem.* 282:5101-5105]이 참조된다. Cas 단백질은, 예를 들어 하나 이상의 핵 국재화 신호 (예를 들어, 2개의 핵 국재화 신호)를 포함할 수 있다. 이러한 세포내 국재화 신호는 N 말단, C 말단 또는 Cas 단백질 내의 임의의 위치에 위치할 수 있다. NLS는 염기성 아미노산의 스트레치를 포함할 수 있으며, 모노파르타이트(monopartite) 서열 또는 바이파르타이트(bipartite) 서열일 수 있다.
- [0267] Cas 단백질은 또한 세포 투과성(cell-penetrating) 도메인을 포함할 수 있다. 예를 들어, 세포 투과성 도메인은 HIV-1 TAT 단백질, 인간 B형 간염 바이러스 유래의 TLM 세포 투과성 모티프, MPG, Pep-1, VP22, 단순 포진 바이러스 유래의 세포 투과성 펩티드, 또는 폴리아르기닌 펩티드 서열로부터 유래될 수 있다. 예를 들어, 전체적으로 본 명세서에 참고로 포함되는 국제특허 공개 WO 2014/089290호가 참조된다. 세포 투과성 도메인은 N 말단, C 말단 또는 Cas 단백질 내의 임의의 위치에 위치할 수 있다.
- [0268] Cas 단백질은 또한 추적 또는 정제를 용이하게 하기 위한 이중성 폴리펩티드, 예를 들어 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, 아자미 그린(Azami Green), 모노머릭 아자미 그린(Monomeric Azami Green), CopGFP, AceGFP, ZsGreen1), 황색 형광 단백질 (예를 들어, YFP, eYFP, 시트린(Citrine), 비너스, YPet, PhiYFP, ZsYellow1), 청색 형광 단백질 (예를 들어, eBFP, eBFP2, 아주라이트(Azurite), mKalamal, GFPuv, 사파이어, T-사파이어), 사이안 형광 단백질(예를 들어, eCFP, 세룰리언(Cerulean), CyPet, AmCyan1, 미도리이시-사이안(Midoriishi-Cyan)), 적색 형광 단백질 (mKate, mKate2, mPlum, DsRed 모노머, mCherry, mRFP1, DsRed-익스프레스(DsRed-Express), DsRed2, DsRed-모노머, HcRed-탠덤(HcRed-Tandem), HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), 오렌지색 형광 단백질 (mOrange, mKO, 쿠사비라-오렌지(Kusabira-Orange), 모노머릭 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(tandem affinity purification, TAP) 태그, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, 적혈구응집소 (HA), nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 히스티딘 (His), 바이오틴 카르복실 운반 단백질 (BCCP), 및 칼모듈린을 포함한다.
- [0269] Cas 단백질은 임의의 형태로 제공될 수 있다. 예를 들어, Cas 단백질은 단백질, 예컨대 gRNA와 복합체를 형성한 Cas 단백질의 형태로 제공될 수 있다. 대안적으로, Cas 단백질은 Cas 단백질을 인코딩하는 핵산, 예를 들어 RNA (예컨대, 메신저 RNA (mRNA)) 또는 DNA의 형태로 제공될 수 있다. 선택적으로, Cas 단백질을 인코딩하는 핵산은 특정 세포 또는 유기체에서 단백질로의 효율적인 번역을 위해 코돈 최적화될 수 있다. Cas 단백질을 인코딩하는 핵산을 세포 내로 도입시킬 때, Cas 단백질은 세포에서 일시적으로, 조건부로 또는 항시적으로 발현될 수 있다.
- [0270] Cas 단백질을 인코딩하는 핵산은 세포의 게놈에 안정하게 통합될 수 있으며, 세포에서 활성인 프로모터에 작동 가능하게 연결될 수 있다. 대안적으로, Cas 단백질을 인코딩하는 핵산은 발현 작제물에서 프로모터에 작동 가능하게 연결될 수 있다. 발현 작제물은 관심 유전자 또는 다른 핵산 서열 (예를 들어, Cas 유전자)의 발현을 유도할 수 있고 이러한 관심 핵산 서열을 표적 세포에 전달할 수 있는 임의의 핵산 작제물을 포함한다. 예를 들어, Cas 단백질을 인코딩하는 핵산은 gRNA를 인코딩하는 DNA를 포함하는 벡터에 존재할 수 있다. 대안적으로, 이것은 gRNA를 인코딩하는 DNA를 포함하는 벡터와는 별개인 벡터 또는 플라스미드에 존재할 수 있다. 발현 작제물에 사용될 수 있는 프로모터는, 예를 들어 만능성 래트, 진핵, 포유류, 비인간 포유류, 인간, 설치류, 마우스 또는 햄스터 세포에서 활성인 프로모터를 포함한다. 다른 프로모터의 예는 본 명세서의 다른 곳에 기재되어 있다.

- [0271] "가이드 RNA" 또는 "gRNA"는 Cas 단백질에 결합하여 Cas 단백질을 표적 DNA 내의 특정 위치로 표적화하는 RNA 분자를 포함한다. 가이드 RNA는 2개의 세그먼트, 즉, "DNA 표적화 세그먼트"와 "단백질 결합 세그먼트"를 포함할 수 있다. "세그먼트"는 RNA의 뉴클레오티드의 연속 스트레치와 같은, 분자의 세그먼트, 섹션 또는 영역을 포함한다. 일부 gRNA는 2개의 별개의 RNA 분자, 즉, "활성화인자(activator)-RNA"와 "표적화인자(targeter)-RNA"를 포함한다. 다른 gRNA는 "단일 분자 gRNA", "단일 가이드 RNA" 또는 "sgRNA"로도 명명될 수 있는 단일 RNA 분자 (단일 RNA 폴리뉴클레오티드)이다. 예를 들어, 각각 본 명세서에 참고로 포함되는, 국제특허 공개 WO/2013/176772A1호, 국제특허 공개 WO/2014/065596A1호, 국제특허 공개 WO/2014/089290A1호, 국제특허 공개 WO/2014/093622A2호, 국제특허 공개 WO/2014/099750A2호, 국제특허 공개 WO/2013142578A1호 및 국제특허 공개 WO 2014/131833A1호가 참조된다. 용어 "가이드 RNA" 및 "gRNA"는 포괄적인 것으로, 이중 분자 gRNA와 단일 분자 gRNA 둘 모두를 포함한다.
- [0272] 예시적인 2분자 gRNA는 crRNA 유사 ("CRISPR RNA" 또는 "표적화인자-RNA" 또는 "crRNA" 또는 "crRNA 반복체") 분자 및 상응하는 tracrRNA 유사 ("트랜스 작용(trans-acting) CRISPR RNA" 또는 "활성화인자-RNA" 또는 "tracrRNA" 또는 "스캐폴드") 분자를 포함한다. crRNA는 gRNA의 DNA 표적화 세그먼트 (단일 가닥), 및 gRNA의 단백질 결합 세그먼트의 dsRNA 이중 나선 구조의 절반을 형성하는 뉴클레오티드의 스트레치 둘 모두를 포함한다.
- [0273] 상응하는 tracrRNA (활성화인자-RNA)는 gRNA의 단백질 결합 세그먼트의 dsRNA 이중 나선 구조의 다른 절반을 형성하는 뉴클레오티드의 스트레치를 포함한다. crRNA의 뉴클레오티드의 스트레치는 tracrRNA의 뉴클레오티드의 스트레치에 상보적이며, 이것과 혼성화되어 gRNA의 단백질 결합 도메인의 dsRNA 이중 나선 구조를 형성한다. 이와 같이, 각 crRNA는 상응하는 tracrRNA를 갖는다고 할 수 있다.
- [0274] crRNA와 이에 상응하는 tracrRNA는 혼성화되어 gRNA를 형성한다. 추가로 crRNA는 표적 서열에 혼성화되는 단일 가닥 DNA 표적화 세그먼트를 제공한다. 세포 내에서의 변형에 사용되는 경우, 주어진 crRNA 또는 tracrRNA 분자의 정확한 서열은 이들 RNA 분자가 사용될 종에 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]이 참조되며, 이들 각각은 본 명세서에 참고로 포함된다.
- [0275] 주어진 gRNA의 DNA 표적화 세그먼트 (crRNA)는 표적 DNA의 서열에 상보적인 뉴클레오티드 서열을 포함한다. gRNA의 DNA 표적화 세그먼트는 혼성화 (즉, 염기쌍 형성)를 통해 서열 특이적 방식으로 표적 DNA와 상호작용한다. 이와 같이, 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 요망되는 서열에 혼성화되도록 변형될 수 있다. 자연적으로 발생하는 crRNA는 Cas9 시스템 및 유기체에 따라 다르지만, 21개 내지 46개의 뉴클레오티드의 길이의 2개의 직접 반복체 (DR)에 의해 플랭킹된 21개 내지 72개의 뉴클레오티드의 길이의 표적화 세그먼트를 종종 함유한다 (예를 들어, 국제특허 공개 WO2014/131833호 참조). S. 피오게네스의 경우, DR은 36개의 뉴클레오티드의 길이이고, 표적화 세그먼트는 30개의 뉴클레오티드의 길이이다. 3'에 위치한 DR은 상응하는 tracrRNA에 상보적이고 이와 혼성화되며, 이는 다시, Cas9 단백질에 결합한다.
- [0276] DNA 표적화 세그먼트의 길이는 약 12개의 뉴클레오티드 내지 약 100개의 뉴클레오티드일 수 있다. 예를 들어, DNA 표적화 세그먼트의 길이는 약 12개의 뉴클레오티드(nt) 내지 약 80개의 nt, 약 12개의 nt 내지 약 50개의 nt, 약 12개의 nt 내지 약 40개의 nt, 약 12개의 nt 내지 약 30개의 nt, 약 12개의 nt 내지 약 25개의 nt, 약 12개의 nt 내지 약 20개의 nt, 또는 약 12개의 nt 내지 약 19개의 nt일 수 있다. 대안적으로, DNA 표적화 세그먼트의 길이는 약 19개의 nt 내지 약 20개의 nt, 약 19개의 nt 내지 약 25개의 nt, 약 19개의 nt 내지 약 30개의 nt, 약 19개의 nt 내지 약 35개의 nt, 약 19개의 nt 내지 약 40개의 nt, 약 19개의 nt 내지 약 45개의 nt, 약 19개의 nt 내지 약 50개의 nt, 약 19개의 nt 내지 약 60개의 nt, 약 19개의 nt 내지 약 70개의 nt, 약 19개의 nt 내지 약 80개의 nt, 약 19개의 nt 내지 약 90개의 nt, 약 19개의 nt 내지 약 100개의 nt, 약 20개의 nt 내지 약 25개의 nt, 약 20개의 nt 내지 약 30개의 nt, 약 20개의 nt 내지 약 35개의 nt, 약 20개의 nt 내지 약 40개의 nt, 약 20개의 nt 내지 약 45개의 nt, 약 20개의 nt 내지 약 50개의 nt, 약 20개의 nt 내지 약 60개의 nt, 약 20개의 nt 내지 약 70개의 nt, 약 20개의 nt 내지 약 80개의 nt, 약 20개의 nt 내지 약 90개의 nt, 또는 약 20개의 nt 내지 약 100개의 nt일 수 있다.
- [0277] 표적 DNA의 뉴클레오티드 서열 (표적 서열)에 상보적인 DNA 표적화 세그먼트의 뉴클레오티드 서열은 적어도 약 12개의 nt의 길이를 가질 수 있다. 예를 들어, DNA 표적화 서열 (즉, 표적 DNA 내의 표적 서열에 상보적인 DNA

표적화 세그먼트 내의 서열)은 적어도 약 12개의 nt, 적어도 약 15개의 nt, 적어도 약 18개의 nt, 적어도 약 19개의 nt, 적어도 약 20개의 nt, 적어도 약 25개의 nt, 적어도 약 30개의 nt, 적어도 약 35개의 nt, 또는 적어도 약 40개의 nt의 길이를 가질 수 있다. 대안적으로, DNA 표적화 서열은 약 12개의 뉴클레오티드(nt) 내지 약 80개의 nt, 약 12개의 nt 내지 약 50개의 nt, 약 12개의 nt 내지 약 45개의 nt, 약 12개의 nt 내지 약 40개의 nt, 약 12개의 nt 내지 약 35개의 nt, 약 12개의 nt 내지 약 30개의 nt, 약 12개의 nt 내지 약 25개의 nt, 약 12개의 nt 내지 약 20개의 nt, 약 12개의 nt 내지 약 19개의 nt, 약 19개의 nt 내지 약 20개의 nt, 약 19개의 nt 내지 약 25개의 nt, 약 19개의 nt 내지 약 30개의 nt, 약 19개의 nt 내지 약 35개의 nt, 약 19개의 nt 내지 약 40개의 nt, 약 19개의 nt 내지 약 45개의 nt, 약 19개의 nt 내지 약 50개의 nt, 약 19개의 nt 내지 약 60개의 nt, 약 20개의 nt 내지 약 25개의 nt, 약 20개의 nt 내지 약 30개의 nt, 약 20개의 nt 내지 약 35개의 nt, 약 20개의 nt 내지 약 40개의 nt, 약 20개의 nt 내지 약 45개의 nt, 약 20개의 nt 내지 약 50개의 nt, 또는 약 20개의 nt 내지 약 60개의 nt의 길이를 가질 수 있다. 일부의 경우에, DNA 표적화 서열은 약 20개의 nt의 길이를 가질 수 있다.

[0278] tracrRNA는 임의의 형태 (예를 들어, 전장(full-length) tracrRNA 또는 활성 부분 tracrRNA)로 존재할 수 있고 다양한 길이를 가질 수 있다. 이것은 1차 전사물 또는 프로세싱(processing)된 형태를 포함할 수 있다. 예를 들어, (단일 가이드 RNA의 일부로서의, 또는 2분자 gRNA의 일부로서의 별개의 분자로서의) tracrRNA는 야생형 tracrRNA 서열의 전부 또는 일부 (예를 들어, 야생형 tracrRNA 서열의 약 20개, 26개, 32개, 45개, 48개, 54개, 63개, 67개, 85개 또는 그 이상, 또는 그 초과 개수의 뉴클레오티드)를 포함하거나 이로 이루어질 수 있다. S. 피오케네스 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]; 국제 특허 공개 WO 2014/093661호가 참조되며, 이들 각각은 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 sgRNA에 포함된다는 것을 나타낸다. 전체적으로 본 명세서에 참고로 포함되는 미국 특허 제8,697,359호가 참조된다.

[0279] DNA 표적화 서열과 표적 DNA 내의 표적 서열 사이의 상보성 비율은 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. 일부의 경우에, DNA 표적화 서열과 표적 DNA 내의 표적 서열 사이의 상보성 비율은 약 20개의 연속 뉴클레오티드에 대하여 적어도 60%이다. 일례에서, DNA 표적화 서열과 표적 DNA 내의 표적 서열 사이의 상보성 비율은 표적 DNA의 상보적 가닥 내의 표적 서열의 5' 말단에 있는 14개의 연속 뉴클레오티드에 대하여 100%이고, 나머지 부분에 대해서는 0%로 낮다. 이러한 경우에, DNA 표적화 서열은 길이가 14개의 뉴클레오티드인 것으로 간주될 수 있다. 다른 예에서, DNA 표적화 서열과 표적 DNA 내의 표적 서열 사이의 상보성 비율은 표적 DNA의 상보적 가닥 내의 표적 서열의 5' 말단에 있는 7개의 연속 뉴클레오티드에 대하여 100%이고, 나머지 부분에 대해서는 0%로 낮다. 이러한 경우에, DNA 표적화 서열은 길이가 7개의 뉴클레오티드인 것으로 간주될 수 있다.

[0280] gRNA의 단백질 결합 세그먼트는 서로 상보적인 2개의 뉴클레오티드 스트레치를 포함할 수 있다. 단백질 결합 세그먼트의 상보적 뉴클레오티드는 혼성화되어 이중 가닥 RNA 이중 나선 구조(dsRNA)를 형성한다. 대상 gRNA의 단백질 결합 세그먼트는 Cas 단백질과 상호작용하고, gRNA는 결합된 Cas 단백질을 DNA 표적화 세그먼트를 통해 표적 DNA 내의 특정 뉴클레오티드 서열로 유도한다.

[0281] 가이드 RNA는 추가의 바람직한 특징 (예를 들어, 변형 또는 조절된 안정성; 세포내 표적화; 형광 표지를 이용한 추적; 단백질 또는 단백질 복합체에 대한 결합 부위 등)을 제공하는 변형 또는 서열을 포함할 수 있다. 이러한 변형의 예는, 예를 들어 5' 캡 (예를 들어, 7-메틸구아닐레이트 캡 (m7G)); 3' 폴리아데닐화된 테일 (즉, 3' 폴리(A) 테일); 리보스위치(riboswitch) 서열 (예를 들어, 단백질 및/또는 단백질 복합체에 의한 조절된 안정성 및/또는 조절된 접근성을 가능하게 함); 안정성 제어 서열; dsRNA 이중 나선 구조 (즉, 헤어핀)를 형성하는 서열; RNA를 세포내 위치 (예를 들어, 핵, 미토콘드리아, 엽록체 등)에 표적화하는 변형 또는 서열; 추적 (예를 들어, 형광 분자에 대한 직접 접합, 형광 검출을 용이하게 하는 모이어티(moiety)에 대한 접합, 형광 검출을 가능하게 하는 서열 등)을 제공하는 변형 또는 서열; 단백질 (예를 들어, 전사 활성화인자, 전사 억제인자, DNA 메틸트랜스퍼라제, DNA 데메틸라제, 히스톤 아세틸트랜스퍼라제, 히스톤 데아세틸라제 등을 포함하는, DNA에 작용하는 단백질)에 대한 결합 부위를 제공하는 변형 또는 서열; 및 이들의 조합을 포함한다.

[0282] 가이드 RNA는 임의의 형태로 제공될 수 있다. 예를 들어, gRNA는 2개의 분자 (별개의 crRNA 및 tracrRNA) 또는 1개의 분자 (sgRNA)로서의 RNA 형태로 그리고 선택적으로 Cas 단백질과의 복합체 형태로 제공될 수 있다. gRNA

는 또한 gRNA를 인코딩하는 DNA의 형태로 제공될 수 있다. gRNA를 인코딩하는 DNA는 단일 RNA 분자 (sgRNA) 또는 별개의 RNA 분자 (예를 들어, 별개의 crRNA 및 tracrRNA)를 인코딩할 수 있다. 후자의 경우, gRNA를 인코딩하는 DNA는, 각각 crRNA 및 tracrRNA를 인코딩하는 별개의 DNA 분자들로서 제공될 수 있다. 대안적으로, gRNA를 인코딩하는 DNA는 하나의 DNA 분자로 제공될 수 있다.

[0283] gRNA를 인코딩하는 DNA가 세포 내로 도입될 때, gRNA는 세포 내에서 일시적으로, 조건부로 또는 항시적으로 발현될 수 있다. gRNA를 인코딩하는 DNA는 세포의 게놈에 안정하게 통합될 수 있으며, 세포에서 활성인 프로모터에 작동가능하게 연결될 수 있다. 대안적으로, gRNA를 인코딩하는 DNA는 발현 작제물에서 프로모터에 작동가능하게 연결될 수 있다. 예를 들어, gRNA를 인코딩하는 DNA는 Cas 단백질을 인코딩하는 핵산을 포함하는 벡터에 존재할 수 있다. 대안적으로, 이것은 Cas 단백질을 인코딩하는 핵산을 포함하는 벡터와는 별개인 벡터 또는 플라스미드에 존재할 수 있다. 이러한 발현 작제물에 사용될 수 있는 프로모터는, 예를 들어 만능성 래트, 진핵, 포유류, 비인간 포유류, 인간, 설치류, 마우스 또는 햄스터 세포에서 활성인 프로모터를 포함한다. 이러한 프로모터는, 예를 들어 조건부 프로모터, 유도성 프로모터, 항시적 프로모터 또는 조직 특이적 프로모터일 수 있다. 일부의 경우에, 프로모터는 RNA 폴리머라제 III 프로모터, 예를 들어 인간 U6 프로모터이다.

[0284] 대안적으로, gRNA는 다양한 다른 방법으로 제조될 수 있다. 예를 들어, gRNA는, 예컨대 T7 RNA 폴리머라제를 사용하여 시험관내 전사에 의해 제조될 수 있다 (예를 들어, 각각 모든 목적을 위해 본 명세서에 전체적으로 참고로 포함되는, 국제특허 공개 WO 2014/089290호 및 국제특허 공개 WO 2014/065596호 참조). 가이드 RNA는 화학 합성에 의해 제조되는 합성적으로 생성된 분자일 수도 있다.

[0285] CRISPR/Cas 시스템에 대한 표적 서열은, 결합을 위한 충분한 조건이 존재한다면, gRNA의 DNA 표적화 세그먼트가 결합할 표적 DNA에 존재하는 핵산 서열을 포함한다. 예를 들어, 표적 서열은 가이드 RNA가 상보성을 갖도록 설계된 서열을 포함하며, 여기서 표적 서열과 DNA 표적화 서열 사이의 혼성화는 CRISPR 복합체의 형성을 촉진시킨다. 혼성화를 일으켜서 CRISPR 복합체의 형성을 촉진시키기에 충분한 상보성이 존재한다면, 완전 상보성이 반드시 필요한 것은 아니다. 표적 서열은 또한 하기에 더욱 상세히 설명되는, Cas 단백질에 대한 절단 부위를 포함한다. 표적 서열은, 예를 들어 세포의 핵 또는 세포질에, 또는 미토콘드리아 또는 엽록체와 같은 세포 소기관 내에 위치할 수 있는 임의의 폴리뉴클레오티드를 포함할 수 있다.

[0286] 표적 DNA 내의 표적 서열은 Cas 단백질 또는 gRNA에 의해 표적화될 수 있다 (즉, 이에 의해 결합되거나, 이와 혼성화되거나, 이에 상보적일 수 있다). 적합한 DNA/RNA 결합 조건은 통상 세포 내에 존재하는 생리학적 조건을 포함한다. 다른 적합한 DNA/RNA 결합 조건 (예를 들어, 무세포 시스템에서의 조건)은 당업계에 공지되어 있다 (예를 들어, 문헌[Molecular Cloning: A Laboratory Manual, 3rd Ed. (Sambrook *et al.*, Harbor Laboratory Press 2001)] 참조). Cas 단백질 또는 gRNA와 상보적이며 이와 혼성화되는 표적 DNA의 가닥은 "상보적 가닥"으로 명명될 수 있고, "상보적 가닥"과 상보적인 (따라서 Cas 단백질 또는 gRNA에 상보적이지 않은) 표적 DNA의 가닥은 "비상보적 가닥" 또는 "주형 가닥"으로 명명될 수 있다.

[0287] Cas 단백질은 gRNA의 DNA 표적화 세그먼트가 결합할 표적 DNA에 존재하는 핵산 서열의 내부 또는 외부의 부위에서 핵산을 절단할 수 있다. "절단 부위"는 Cas 단백질이 단일 가닥 절단부 또는 이중 가닥 절단부를 생성시키는 핵산 위치를 포함한다. 예를 들어, CRISPR 복합체 (표적 서열에 혼성화되어 Cas 단백질과 복합체를 형성한 gRNA를 포함함)의 형성은 gRNA의 DNA 표적화 세그먼트가 결합될 표적 DNA에 존재하는 핵산 서열에서 또는 이 부근에서 (예를 들어, 그 핵산 서열로부터 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 20개, 50개 또는 그 이상의 염기쌍 이내에서) 한 가닥 또는 두 가닥의 절단을 일으킬 수 있다. 절단 부위가 gRNA의 DNA 표적화 세그먼트가 결합할 표적 DNA에 존재하는 핵산 서열의 외부에 존재하는 경우, 절단 부위는 여전히 "표적 서열" 내에 있는 것으로 간주된다. 절단 부위는 핵산의 한 가닥에만 또는 두 가닥에 있을 수 있다. 절단 부위는 핵산의 두 가닥의 동일한 위치에 있을 수 있거나 (평활 말단을 생성함), 각 가닥의 상이한 부위에 있을 수 있다 (스태거된 말단 (즉, 돌출부(overhang))를 생성함). 스태거된 말단은, 예를 들어 상이한 가닥의 상이한 절단 부위에 단일 가닥 절단부를 생성시키는 2개의 Cas 단백질을 사용하여 이중 가닥 절단부를 생성시킴으로써 생성될 수 있다. 예를 들어, 제1 니카제는 이중 가닥 DNA (dsDNA)의 제1 가닥에 단일 가닥 절단부를 생성시킬 수 있는 한편, 제2 니카제는 돌출 서열(overhanging sequence)이 생성되도록 dsDNA의 제2 가닥에 단일 가닥 절단부를 생성시킬 수 있다. 일부의 경우에, 제1 가닥의 니카제의 표적 서열은 제2 가닥의 니카제의 표적 서열로부터 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 75, 100, 250, 500 또는 1,000개의 염기쌍 만큼 떨어져 있다.

[0288] Cas9에 의한 표적 DNA의 부위 특이적 절단은 (i) gRNA와 표적 DNA 사이의 염기쌍 형성 상보성과 (ii) 프로토스

페이지 인접 모티프(proto-spacer adjacent motif, PAM)로 명명되는 표적 DNA의 짧은 모티프 둘 모두에 의해 결정되는 위치에서 발생할 수 있다. PAM은 표적 서열을 플랭킹할 수 있다. 선택적으로, 표적 서열은 PAM에 의해 3' 말단에서 플랭킹될 수 있다. 예를 들어, Cas9의 절단 부위는 PAM 서열의 상류 또는 하류에 있는 약 1개 내지 약 10개, 또는 약 2개 내지 약 5개의 염기쌍 (예컨대, 3개의 염기쌍)일 수 있다. 일부의 경우에 (예를 들어, S. 피오게네스 유래의 Cas9 또는 밀접하게 관련된 Cas9가 사용될 때), 비상보적 가닥의 PAM 서열은 5'-N₁GG-3'일 수 있으며, 여기서 N₁은 임의의 DNA 뉴클레오티드이고 표적 DNA의 비상보적 가닥의 표적 서열의 바로 3'에 위치한다. 이와 같이, 상보적 가닥의 PAM 서열은 5'-CCN₂-3'일 것이며, 여기서 N₂는 임의의 DNA 뉴클레오티드이고 표적 DNA의 상보적 가닥의 표적 서열의 바로 5'에 위치한다. 일부 이러한 경우에, N₁ 및 N₂는 상보적일 수 있고, N₁-N₂ 염기쌍은 임의의 염기쌍일 수 있다 (예를 들어, N₁=C와 N₂=G; N₁=G와 N₂=C; N₁=A와 N₂=T; 또는 N₁=T와 N₂=A).

[0289] 표적 서열의 예는 gRNA의 DNA 표적화 세그먼트에 상보적인 DNA 서열, 또는 PAM 서열에 덧붙여진 이러한 DNA 서열을 포함한다. 표적 서열의 일례는 GNNNNNNNNNNNNNNNNNNNGG (GN₁₋₂₀ GG; 서열 번호 1)의 뉴클레오티드 서열을 포함한다. 5' 말단의 구아닌은 세포에서 RNA 폴리머라제에 의한 전사를 용이하게 할 수 있다. 표적 서열의 다른 예는 5' 말단에 2개의 구아닌 뉴클레오티드를 포함하여 시험관내에서 T7 폴리머라제에 의한 효율적인 전사를 용이하게 할 수 있다. 예를 들어, 전체적으로 모든 목적을 위하여 본 명세서에 참고로 포함되는 국제특허 공개 WO 2014/065596호가 참조된다. 다른 표적 서열은 5' G 및 3' GG를 포함하는 서열 번호 1의 길이가 4개 내지 22개인 뉴클레오티드를 가질 수 있다. 또 다른 표적 서열은 서열 번호 1의 길이가 14개 내지 20개인 뉴클레오티드를 가질 수 있다.

[0290] 표적 서열은 세포에 내인성이거나 외인성인 임의의 핵산 서열일 수 있다. 표적 서열은 유전자 생성물 (예를 들어, 단백질)을 코딩하는 서열 또는 비코딩 서열 (예를 들어, 조절 서열 또는 정크(junk) DNA)일 수 있거나, 둘 모두를 포함할 수 있다.

[0291] 뉴클레아제 작용제의 활성 변이체 및 단편 (즉, 유전자 조작된 뉴클레아제 작용제)도 제공된다. 이러한 활성 변이체는 천연 뉴클레아제 작용제에 대하여, 적어도 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 그 이상의 서열 동일성을 포함할 수 있으며, 여기서 활성 변이체는 요망되는 인식 부위에서 절단하는 능력을 보유하므로, 닉 유도 또는 이중 가닥 절단 유도 활성을 보유한다. 예를 들어, 본 명세서에 기술된 임의의 뉴클레아제 작용제는 천연 엔도뉴클레아제 서열로부터 변형될 수 있고, 천연 뉴클레아제 작용제에 의해 인식되지 않았던 인식 부위를 인식하여 여기에서 닉 또는 이중 가닥 절단을 유도하도록 설계될 수 있다. 따라서, 일부 실시 형태에서, 유전자 조작된 뉴클레아제는 상응하는 천연 뉴클레아제 작용제의 인식 부위와 상이한 인식 부위에 닉 또는 이중 가닥 절단을 유도하는 특이성을 갖는다. 닉 유도 또는 이중 가닥 절단 유도 활성의 검정은 공지되어 있으며, 대체로 인식 부위를 포함하는 DNA 기질에 대한 엔도뉴클레아제의 전체 활성 및 특이성을 측정한다.

[0292] 뉴클레아제 작용제를 당업계에 공지된 임의의 수단에 의해 만능성 세포 내로 도입시킬 수 있다. 뉴클레아제 작용제를 인코딩하는 폴리펩티드를 세포 내로 직접 도입시킬 수 있다. 대안적으로, 뉴클레아제 작용제를 인코딩하는 폴리뉴클레오티드를 세포 내로 도입시킬 수 있다. 뉴클레아제 작용제를 인코딩하는 폴리뉴클레오티드가 세포 내로 도입될 때, 뉴클레아제 작용제는 세포 내에서 일시적으로, 조건부로 또는 항시적으로 발현될 수 있다. 따라서, 뉴클레아제 작용제를 인코딩하는 폴리뉴클레오티드는 발현 카세트에 함유될 수 있고, 조건부 프로모터, 유도성 프로모터, 항시적 프로모터 또는 조직 특이적 프로모터에 작동가능하게 연결될 수 있다. 대안적으로, 뉴클레아제 작용제는 뉴클레아제 작용제를 인코딩하는 mRNA로 세포 내로 도입된다.

[0293] 특정 실시 형태에서, 뉴클레아제 작용제를 인코딩하는 폴리뉴클레오티드는 세포의 게놈에 안정하게 통합되며, 세포에서 활성인 프로모터에 작동가능하게 연결되어 있다. 다른 실시 형태에서, 뉴클레아제 작용제를 인코딩하는 폴리뉴클레오티드는 핵산 삽입물을 포함하는 동일한 표적화 벡터에 존재하지만, 다른 경우에는 뉴클레아제 작용제를 인코딩하는 폴리뉴클레오티드는 핵산 삽입물을 포함하는 표적화 벡터와는 별개의 벡터 또는 플라스미드에 존재한다.

[0294] 뉴클레아제 작용제가 뉴클레아제 작용제를 인코딩하는 폴리뉴클레오티드의 도입을 통해 세포에 제공될 때, 뉴클레아제 작용제를 인코딩하는 이러한 폴리뉴클레오티드는 뉴클레아제 작용제를 인코딩하는 자연적으로 발생하는 폴리뉴클레오티드 서열과 비교하여, 관심 세포에서 사용 빈도가 더 높은 코돈으로 치환되도록 변형될 수 있다. 예를 들어, 뉴클레아제 작용제를 인코딩하는 폴리뉴클레오티드는 자연적으로 발생하는 폴리뉴클레오티드 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 포함하는 원핵 또는 진핵 세포에서 사용 빈도가 더 높은 코돈으로 치환되도록 변형될 수 있다.

[0295] 상기 제시된 다양한 방법은 순차적으로 반복되어, 염색체 상의 주어진 표적화된 게놈 유전자좌 내로의 임의의 수의 핵산 삽입물의 표적화된 통합을 가능하게 할 수 있다. 따라서, 다양한 방법은 염색체 상의 표적 게놈 유전자좌 내로 적어도 1개, 2개, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 10개, 11개, 12개, 13개, 14개, 15개, 16개, 17개, 18개, 19개, 20개 또는 그 이상의 핵산 삽입물의 삽입을 제공한다. 특정한 실시 형태에서, 이러한 순차적 타일링(tiling) 방법은 동물 세포 또는 포유류 세포 (즉, 인간, 비인간, 설치류, 마우스, 원숭이, 래트, 햄스터, 가축 포유류 또는 농업용 동물) 유래의 큰 게놈 영역의 염색체 상의 표적화된 게놈 유전자좌로의 재구성을 가능하게 한다. 이러한 경우에, 코딩 및 비코딩 영역 둘 모두를 포함하는 게놈 영역의 전달 및 재구성은 천연 게놈 영역 내에서 발견되는 코딩 영역, 비코딩 영역 및 카피 수 변이를 적어도 부분적으로 보유함으로써 주어진 영역의 복잡성이 보존될 수 있게 한다. 따라서, 다양한 방법은, 예를 들어 세포 내에 "이중성" 또는 "외인성" 게놈 영역을 생성시키는 방법을 제공한다.

[0296] **IV. 선택 마커**

[0297] 본 명세서에서 제공된 다양한 방법 및 조성물은 선택 마커와 조합하여 뉴클레아제 작용제 및 이의 상응하는 인식 부위를 사용할 수 있다. 본 명세서에서 논의된 바와 같이, 선택 마커를 인코딩하는 폴리뉴클레오티드에서의 인식 부위의 위치는 표적 유전자좌에서의 통합 사건을 확인하는 효율적인 방법을 가능하게 한다. 더욱이, 다양한 방법이 본 명세서에 제공되며, 여기서 뉴클레아제 인식 부위를 갖는 대안적인 선택 마커를 사용하여 효율 및 효능을 개선하는데, 이를 통해 다수의 관심 폴리뉴클레오티드가 주어진 표적화된 유전자좌 내에 통합된다.

[0298] 다양한 선택 마커가 본 명세서에 개시된 방법 및 조성물에 사용될 수 있다. 이러한 선택 마커는, 예를 들어 항생제, 예컨대 G418, 하이그로마이신, 블라스티시딘, 네오마이신 또는 퓨로마이신에 대한 내성을 부여할 수 있다. 이러한 선택 마커는 네오마이신 포스포트랜스퍼라제 (neo^r), 하이그로마이신 B 포스포트랜스퍼라제 (hyg^r), 퓨로마이신-N-아세틸트랜스퍼라제 (puro^r) 및 블라스티시딘 S 데아미나제 (bsr^r)를 포함한다. 또 다른 실시 형태에서, 선택 마커는 유도성 프로모터에 작동가능하게 연결되고 선택 마커의 발현은 세포에 독성이다. 이러한 선택 마커의 비제한적인 예는 잔틴/구아닌 포스포리보실 트랜스퍼라제 (gpt), 하이포잔틴-구아닌 포스포리보실트랜스퍼라제 (HGPRT) 또는 단순 포진 바이러스 티미딘 키나제 (HSV-TK)를 포함한다.

[0299] 일 실시 형태에서, 뉴클레아제 인식 부위는 선택 마커를 인코딩하는 유전자 내에 위치한다. 특정 실시 형태에서, 뉴클레아제 인식 부위는 하이그로마이신 유전자 내에 위치한다.

[0300] 선택 마커를 인코딩하는 폴리뉴클레오티드는 세포에서 활성인 프로모터에 작동가능하게 연결되어 있다. 이러한 발현 카세트 및 이의 다양한 조절 성분은 본 명세서의 다른 곳에 더욱 상세히 논의되어 있다.

[0301] **V. 프로모터**

[0302] 본 명세서에 기술된 다양한 핵산 서열은 프로모터에 작동가능하게 연결될 수 있다. 이러한 프로모터는, 예를 들어 만능성인 진핵, 포유류, 비인간 포유류, 인간, 설치류, 마우스 또는 햄스터 세포에서 활성일 수 있다. 프로모터는, 예를 들어 항시적으로 활성인 프로모터, 조건부 프로모터, 유도성 프로모터, 일시적으로 제한된 프로모터 (예컨대, 발달적으로 조절되는 프로모터), 또는 공간적으로 제한된 프로모터 (예를 들어, 세포 특이적 또는 조직 특이적 프로모터)일 수 있다. 프로모터의 예는, 예를 들어 전체적으로 본 명세서에 참고로 포함되는 국제특허 공개 WO 2013/176772호에서 찾아볼 수 있다.

[0303] 유도성 프로모터의 예는, 예를 들어 화학적으로 조절되는 프로모터 및 물리적으로 조절되는 프로모터를 포함한다. 화학적으로 조절되는 프로모터는, 예를 들어 알코올 조절 프로모터 (예컨대, 알코올 데하이드로나제 (alcA) 유전자 프로모터), 테트라사이클린 조절 프로모터 (예컨대, 테트라사이클린 반응성 프로모터, 테트라사이클린 오퍼레이터 서열 (tetO), tet-On 프로모터 또는 tet-Off 프로모터), 스테로이드 조절 프로모터 (예컨대, 래트 글루코코르티코이드 수용체, 에스트로겐 수용체의 프로모터 또는 엑디손 수용체의 프로모터) 또는 금속 조절 프로모터 (예컨대, 금속단백질 프로모터)를 포함한다. 물리적으로 조절되는 프로모터는, 예를 들어 온도 조절 프로모터 (예를 들어, 열충격 프로모터) 및 광 조절 프로모터 (예컨대, 광 유도성 프로모터 또는 광 억제성 프로모터)를 포함한다.

[0304] 조직 특이적 프로모터는, 예를 들어 뉴런 특이적 프로모터, 아교세포 특이적 프로모터, 근육 세포 특이적 프로

모터, 심장 세포 특이적 프로모터, 신장 세포 특이적 프로모터, 골 세포 특이적 프로모터, 내피 세포 특이적 프로모터 또는 면역 세포 특이적 프로모터 (예컨대, B 세포 프로모터 또는 T 세포 프로모터)일 수 있다.

[0305] 발달적으로 조절되는 프로모터는, 예를 들어 배아 발달 단계 동안에만 또는 성체 세포에서만 활성인 프로모터를 포함한다.

[0306] 프로모터는 또한 세포 유형에 기반하여 선택될 수 있다. 예를 들어, 다양한 공지된 프로모터가 진핵 세포, 포유류 세포, 비인간 세포, 비인간 포유류 세포, 만능성 세포, 비인간 만능성 세포, 인간 만능성 세포, 인간 ES 세포, 인간 성체 줄기 세포, 발달적으로 제한된 인간 전구 세포, 인간 iPS 세포, 인간 세포, 설치류 세포, 래트 세포, 마우스 세포, 햄스터 세포, 섬유아세포 또는 CHO 세포에서 사용된다.

[0307] **VI. 발현 카세트**

[0308] 본 명세서에 제공된 표적화 시스템의 다양한 성분 (즉, 뉴클레아제 작용제, 인식 부위, 핵산 삽입물, 관심 폴리뉴클레오티드, 표적화 벡터 (즉, LTVEC), 선택 마커 및 다른 성분)을 포함하는 폴리뉴클레오티드 또는 핵산 분자가 본 명세서에 제공된다.

[0309] 표적화 시스템의 다양한 성분을 포함하는 재조합 폴리뉴클레오티드가 추가로 제공된다. 용어 "재조합 폴리뉴클레오티드" 및 "재조합 DNA 작제물"은 본 명세서에서 상호 교환적으로 사용된다. 재조합 작제물은 핵산 서열의 인공 또는 이중성 조합, 예를 들어 자연 상태에서 함께 발견되지 않는 조절 서열 및 코딩 서열을 포함한다. 다른 실시 형태에서, 재조합 작제물은 상이한 공급원으로부터 유래된 조절 서열 및 코딩 서열, 또는 동일한 공급원으로부터 유래되지만 자연 상태에서 발견되는 것과 상이한 방식으로 배열된 조절 서열 및 코딩 서열을 포함할 수 있다. 이러한 작제물은 그 자체로 사용될 수 있거나 벡터와 함께 사용될 수 있다. 벡터가 사용되는 경우, 벡터의 선택은 당업자에게 잘 알려진 바와 같이 숙주 세포를 형질전환시키는 데 사용되는 방법에 의존한다. 예를 들어, 플라스미드 벡터가 사용될 수 있다. 숙주 세포를 성공적으로 형질전환시키고, 선택하고, 증식시키는 데 필요하고 임의의 단리된 핵산 단편을 포함하는 유전자 요소들이 본 명세서에 제공된다. 스크리닝은 그 중에서도 DNA의 서던 분석, mRNA 발현의 노던(Northern) 분석, 단백질 발현의 면역블롯팅 분석, 또는 표현형 분석에 의해 달성될 수 있다.

[0310] 특정 실시 형태에서, 본 명세서에 기술된 표적화 시스템의 하나 이상의 성분이 원핵 세포, 진핵 세포, 세균, 효모 세포, 포유류 세포 또는 다른 유기체 또는 관심 세포 유형에서의 발현을 위해 발현 카세트에 제공될 수 있다. 카세트는 본 명세서에 제공된 폴리뉴클레오티드에 작동가능하게 연결된 5' 및 3' 조절 서열을 포함할 수 있다. 2개의 단백질 코딩 영역의 결합을 지칭하는 데 사용될 때, '작동가능하게 연결된'은 코딩 영역이 동일한 리딩 프레임(reading frame)에 존재함을 의미한다. 다른 경우에, 단백질을 인코딩하는 핵산 서열이 적당한 전사 조절을 보유하도록 조절 서열 (예를 들어, 프로모터, 인핸서, 사일렌서(silencer) 서열 등)에 작동가능하게 연결될 수 있다.

[0311] 카세트는 유기체 내로 공동 도입되도록 적어도 하나의 추가의 관심 폴리뉴클레오티드를 추가로 함유할 수 있다. 대안적으로, 추가의 관심 폴리뉴클레오티드가 다수의 발현 카세트 상에 제공될 수 있다. 이러한 발현 카세트에는 조절 영역의 전사 조절 하에 있도록 재조합 폴리뉴클레오티드의 삽입을 위한 복수의 제한 부위 및/또는 재조합 부위가 제공된다. 발현 카세트는 추가적으로 선택 마커 유전자를 함유할 수 있다.

[0312] 발현 카세트는 5'-3' 전사 방향으로, 전사 및 번역 개시 영역 (즉, 프로모터), 본 명세서에 제공된 재조합 폴리뉴클레오티드, 및 관심 숙주 세포 또는 포유류 세포에서 기능적인 전사 및 번역 종결 영역 (즉, 종결 영역)을 포함할 수 있다. 본 명세서에 제공된 조절 영역 (즉, 프로모터, 전사 조절 영역, 및 번역 종결 영역) 및/또는 폴리뉴클레오티드는 숙주 세포에 대해 또는 서로에 대해 고유하고/유사할 수 있다. 대안적으로, 본 명세서에 제공된 조절 영역 및/또는 폴리뉴클레오티드는 숙주 세포에 대해 또는 서로에 대해 이중성일 수 있다. 예를 들어, 이중성 폴리뉴클레오티드에 작동가능하게 연결된 프로모터는, 폴리뉴클레오티드가 유래된 종과 상이한 종으로부터 유래되거나, 또는 만약 이것이 동일하고/유사한 종으로부터 유래되는 경우, 하나 또는 둘 모두는 이들의 본래의 형태 및/또는 유전자좌로부터 실질적으로 변형되거나, 프로모터는 작동가능하게 연결된 폴리뉴클레오티드에 대한 고유의 프로모터가 아니다. 대안적으로, 본 명세서에 제공된 조절 영역 및/또는 재조합 폴리뉴클레오티드는 전체적으로 합성일 수 있다.

[0313] 종결 영역은 전사 개시 영역에 대해 고유할 수 있거나, 작동가능하게 연결된 재조합 폴리뉴클레오티드에 대해 고유할 수 있거나, 숙주 세포에 대해 고유할 수 있거나, 프로모터, 재조합 폴리뉴클레오티드, 숙주 세포, 또는 이들의 임의의 조합에 대해 다른 공급원 (즉, 외래 또는 이중성)으로부터 유래될 수 있다.

- [0314] 발현 카세트를 제조함에 있어서, 다양한 DNA 단편을 조작하여 적당한 배향으로 DNA 서열을 제공할 수 있다. 이러한 목적을 위해, 어댑터(adapter) 또는 링커(linker)를 사용하여 DNA 단편들을 결합할 수 있거나 다른 조작을 포함시켜 편리한 제한 부위, 피상적 DNA의 제거, 제한 부위의 제거 등을 제공할 수 있다. 이러한 목적을 위해, 시험관내 돌연변이 유발, 프라이머 수복, 제한, 어닐링, 재치환, 예를 들어 전이(transition) 및 역위(transversion)가 포함될 수 있다.
- [0315] 다수의 프로모터가 본 명세서에 제공된 발현 카세트에 사용될 수 있다. 프로모터는 요망되는 결과에 기반하여 선택될 수 있다. 관심 폴리뉴클레오티드의 발현의 시기, 위치 및/또는 수준을 조절하기 위해 발현 카세트에서 다양한 프로모터를 사용하여 다양한 적용이 향상될 수 있음이 인식된다. 이러한 발현 작제물은 또한 요망되는 경우, 프로모터 조절 영역 (예를 들어, 유도성, 항시적, 환경적으로 또는 발달적으로 조절되는, 또는 세포 또는 조직 특이적/선택적 발현을 부여하는 것), 전사 개시 부위, 리보솜 결합 부위, RNA 프로세싱 신호, 전사 종결 부위 및/또는 폴리아데닐화 신호를 함유할 수 있다.
- [0316] 본 명세서에 제공된 폴리뉴클레오티드를 함유하는 발현 카세트는 또한 형질전환된 세포의 선택을 위한 선택 마커 유전자를 포함할 수 있다. 선택 마커 유전자는 형질전환된 세포 또는 조직의 선택을 위해 이용된다.
- [0317] 적절한 경우, 본 방법 및 조성물에 사용된 서열 (즉, 관심 폴리뉴클레오티드, 뉴클레아제 작용제 등)은 세포에서의 증가된 발현을 위해 최적화될 수 있다. 즉, 유전자는 개선된 발현을 위해, 예를 들어 포유류 신호 코돈, 인간 신호 코돈, 설치류 신호 코돈, 마우스 신호 코돈, 래트 신호 코돈 등을 포함하는, 주어진 관심 세포에서 선호되는 코돈을 사용하여 합성될 수 있다.
- [0318] 일 실시 형태에서, 뉴클레아제 작용제는 뉴클레아제를 인코딩하는 핵산 서열을 포함하는 발현 작제물로부터 발현되고, 핵산은 세포에서 활성인 프로모터에 작동가능하게 연결되어 있다.
- [0319] **VII. 유전자 변형된 비인간 동물을 제조하는 방법**
- [0320] 본 명세서에 개시된 다양한 방법을 사용하여 유전자 변형된 비인간 동물을 생산할 수 있다. 일부의 경우에, 유전자 변형된 비인간 동물을 생산하는 방법은 (1) 본 명세서에 기술된 방법을 사용하여 만능성 세포의 계능을 변형시키는 단계; (2) 유전자 변형된 만능성 세포를 선택하는 단계; (3) 유전자 변형된 만능성 세포를 숙주 배아 내로 도입시키는 단계; 및 (4) 유전자 변형된 만능성 세포를 포함하는 숙주 배아를 대리모에 이식하는 단계를 포함한다. 유전자 변형된 만능성 세포로부터의 자손이 생성된다. 공여체 세포를 임의의 시기, 예를 들어 배반포기(blastocyst stage) 또는 상실기 전시기(pre-morula stage) (즉, 4 세포기 또는 8 세포기)에서 숙주 배아 내로 도입시킬 수 있다. 생식세포계열을 통해 유전자 변형을 전달할 수 있는 자손이 생성된다. 만능성 세포는 본 명세서의 다른 곳에 논의된 바와 같이 ES 세포 (예를 들어, 마우스 ES 세포 또는 래트 ES 세포)일 수 있다. 예를 들어, 본 명세서에 참고로 포함되는 미국 특허 제7,294,754호가 참조된다.
- [0321] 또한, 핵 전달 기법을 사용하여 비인간 포유류 동물을 생산할 수 있다. 간략하게 말하면, 핵 전달 방법은 (1) 난모세포를 제핵하거나 제핵된 난모세포를 제공하는 단계; (2) 제핵된 난모세포와 조합될 공여체 세포 또는 핵을 단리 또는 제공하는 단계; (3) 세포 또는 핵을 제핵된 난모세포 내로 삽입하여, 재구성된 세포를 형성하는 단계; (4) 재구성된 세포를 동물의 자궁에 이식하여 배아를 형성하는 단계; 및 (5) 배아를 발달시키는 단계를 포함할 수 있다. 이러한 방법에서, 난모세포는 대체로 사망한 동물로부터 회수되지만, 살아있는 동물의 난관 및/또는 난소로부터도 단리될 수 있다. 난모세포를 제핵 전에 당업자에게 공지된 다양한 배지에서 성숙시킬 수 있다. 난모세포의 제핵은 당업자에게 잘 알려진 다수의 방식으로 수행할 수 있다. 공여체 세포 또는 핵을 제핵된 난모세포 내로 삽입하여 재구성된 세포를 형성하는 것은 융합 전에 투명대(zona pellucida) 아래로의 공여체 세포의 미세주입에 의해 이루어질 수 있다. 융합은 접촉면/융합면(전기융합)을 가로질러 DC 전기 펄스를 인가함으로써, 폴리에틸렌 글리콜과 같은 융합 촉진 화학 물질에 세포를 노출시킴으로써, 또는 센다이 바이러스(Sendai virus)와 같은 비활성화 바이러스에 의해 유도될 수 있다. 재구성된 세포는 핵 공여체 및 수령체 난모세포의 융합 전에, 융합 중에 및/또는 융합 후에 전기적 및/또는 비전기적 수단에 의해 활성화될 수 있다. 활성화 방법은 전기 펄스, 화학적으로 유도된 충격, 정자에 의한 침투, 난모세포의 2가 양이온 수준의 증가, 및 난모세포의 세포 단백질의 인산화 감소 (키나제 저해제에 의한 바와 같음)를 포함한다. 활성화된 재구성 세포 또는 배아를 당업자에게 잘 알려진 배지에서 배양하고, 이어서 동물의 자궁에 전달할 수 있다. 예를 들어, 각각 본 명세서에 참고로 포함되는, 미국 특허 출원 공개 제20080092249호, 국제특허 공개 WO/1999/005266A2호, 미국 특허 출원 공개 제20040177390호, 국제특허 공개 WO/2008/017234A1호 및 미국 특허 제7,612,250호가 참조된다.

[0322] 본 방법은 (1) 표적화된 변형을 포함하는 비인간 ES 세포를 확인하는 단계; (2) 표적화된 변형을 포함하는 비인간 ES 세포를 비인간 숙주 배아 내로 도입시키는 단계; 및 (3) 비인간 숙주 배아를 대리모에 임신시키는 단계를 포함하는 F0 세대 비인간 동물을 생산하는 방법을 추가로 포함할 수 있다. 이어서 대리모는 표적화된 변형을 포함하는 F0 세대 비인간 동물을 생산할 수 있다. 유전자 변형된 만능성 또는 전능성 세포 (예를 들어, 비인간 ES 세포)를 포함하는 숙주 배아를 배반포기까지 인큐베이션시키고 이어서 대리모에 이식하여 F0 동물을 생산할 수 있다. 유전자 변형된 계놈 유전자좌를 갖는 동물은 본 명세서에 기술된 바와 같이 대립유전자 변형 (MOA) 검정을 통해 확인될 수 있다.

[0323] 본 명세서에 제공된 다양한 방법은 유전자 변형된 비인간 F0 동물의 생산을 가능하게 하며, 여기서 유전자 변형된 F0 동물의 세포는 표적화된 변형을 포함한다. F0 동물을 생산하는 데 사용된 방법에 따라, 표적화된 유전자 변형을 갖는 F0 동물 내의 세포 수가 다양할 것임이 인식된다. 예를 들어, VELOCIMOUSE® 방법을 통한 상용하는 유기체의 상실기 전시기의 배아 (예컨대, 8 세포기 마우스 배아) 내로의 공여체 ES 세포의 도입은 더 높은 비율의 F0 동물 세포 집단이 표적화된 유전자 변형을 갖는 세포를 포함하게 할 수 있다. 특정 경우에, 비인간 F0 동물의 세포 기여 중 적어도 50%, 60%, 65%, 70%, 75%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100%가 표적화된 변형을 갖는 세포 집단을 포함한다. 다른 경우에, F0 동물의 생식 세포 중 적어도 하나 이상이 표적화된 변형을 갖는다.

[0324] 일부의 경우, 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.

[0325] 일부의 경우에, 본 명세서에 개시된 방법 및 조성물에 의해 생산된 F0 동물을 야생형 동물과 교배시켜 표적화된 변형에 대해 이형접합성인 F1 세대를 생산할 수 있다. 이어서 F1 세대의 동물을 서로 교배시켜 표적화된 변형에 대해 동형접합성인 F2 동물을 생산할 수 있다. 특정 프라이머 및/또는 프로브를 사용하여 F1 자손을 유전형 결정(genotyping)하여, 표적화된 유전자 변형이 존재하는지를 결정할 수 있다.

[0326] **VIII. 핵산 및 단백질을 세포 내로 도입시키는 방법**

[0327] 세포 내로의 핵산의 도입을 가능하게 하는 다양한 방법 및 조성물이 본 명세서에 제공된다. 일부의 경우에, 핵산을 도입시키는 데 사용되는 시스템은 특정 계놈 유전자좌에서 표적화된 통합을 가능하게 한다. 이러한 시스템은 다양한 성분을 사용하며, 참조의 용이함을 위해, 용어 "표적화된 계놈 통합 시스템"은 총칭적으로 통합 사건에 필요한 모든 성분 (예를 들어, 하나 이상의 뉴클레아제 작용제, 뉴클레아제 절단 부위, 핵산 삽입물, 표적화 벡터, 표적 계놈 유전자좌 및 관심 폴리뉴클레오티드)을 포함한다.

[0328] 본 명세서에 제공된 방법은 표적화된 계놈 통합 시스템의 하나 이상의 성분을 포함하는 하나 이상의 폴리뉴클레오티드 또는 폴리펩티드 작제물을 세포 내로 도입시키는 단계를 포함할 수 있다. "도입"은 서열이 세포 내부에 접근하도록 하는 방식으로 서열 (폴리펩티드 또는 폴리뉴클레오티드)을 세포에 제시하는 것을 포함한다. 본 명세서에 제공된 방법은, 핵산 또는 단백질이 적어도 하나의 세포의 내부에 접근하는 것만을 제외하면, 핵산 또는 단백질을 세포 내로 도입시키는 특정 방법에 의존하지 않는다. 핵산 및 단백질을 다양한 세포 유형 내로 도입시키는 방법은 당업계에 공지되어 있으며, 예를 들어 안정한 형질감염 방법, 일시적 형질감염 방법 및 바이러스 매개 방법을 포함한다.

[0329] 일부의 경우에, 본 방법 및 조성물에 사용되는 세포는 이미 계놈에 안정하게 혼입된 DNA 작제물을 갖는다. "안정하게 혼입된" 또는 "안정하게 도입된"은, 뉴클레오티드 서열이 세포의 계놈 내로 통합되어 그의 자손에 의해 유전될 수 있도록 하는, 세포 내로의 폴리뉴클레오티드의 도입을 포함한다. DNA 작제물 또는 표적화된 계놈 통합 시스템의 다양한 성분의 안정적인 혼입을 위해 임의의 프로토콜을 사용할 수 있다.

[0330] 형질감염 프로토콜뿐만 아니라 폴리펩티드 또는 폴리뉴클레오티드 서열을 세포 내로 도입시키는 프로토콜도 다양할 수 있다. 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 폴리머, 예를 들어 DEAE-텍스트란 또는 폴리에틸렌이민을 사용하는 화학물질 기반 형질감염 방법을 포함한다. 비화학적 방법은 전기천공, 초음파-천공(Sono-poration), 및 광학적 형질감염을 포함한다. 입자 기반 형질감염은 유전자 총(gene gun), 또는 자석 보조 형질감염(magnet-assisted transfection) (문헌[Bertram (2006) *Current Pharmaceutical Biotechnology* 7, 277-28])의 사용을 포함한다.

바이러스 방법도 형질감염에 사용될 수 있다.

[0331] 일부의 경우에, 세포 내로의 핵산 또는 단백질의 도입은 전기천공, 세포질내 주입, 바이러스 감염, 아데노바이러스, 렌티바이러스, 레트로바이러스, 형질감염, 지질 매개 형질감염 또는 Nucleofection™에 의해 매개된다.

[0332] 세포 내로의 핵산 또는 단백질의 도입을 일정 시간에 걸쳐 1회 또는 수회 수행할 수 있다. 예를 들어, 도입을 일정 시간에 걸쳐 적어도 2회, 일정 시간에 걸쳐 적어도 3회, 일정 시간에 걸쳐 적어도 4회, 일정 시간에 걸쳐 적어도 5회, 일정 시간에 걸쳐 적어도 6회, 일정 시간에 걸쳐 적어도 7회, 일정 시간에 걸쳐 적어도 8회, 일정 시간에 걸쳐 적어도 9회, 일정 시간에 걸쳐 적어도 10회, 일정 시간에 걸쳐 적어도 11회, 일정 시간에 걸쳐 적어도 12회, 일정 시간에 걸쳐 적어도 13회, 일정 시간에 걸쳐 적어도 14회, 일정 시간에 걸쳐 적어도 15회, 일정 시간에 걸쳐 적어도 16회, 일정 시간에 걸쳐 적어도 17회, 일정 시간에 걸쳐 적어도 18회, 일정 시간에 걸쳐 적어도 19회 또는 일정 시간에 걸쳐 적어도 20회 수행할 수 있다.

[0333] 뉴클레아제 작용제와 표적화 벡터 (예를 들어, LTVEC) 둘 모두를 세포 내로 도입시킬 때, 이들을 동시에 도입시킬 수 있다. 대안적으로, 뉴클레아제 작용제를 표적화 벡터와 별개로 도입시킬 수 있다. 예를 들어, 뉴클레아제 작용제를 표적화 벡터의 도입 전에 도입시킬 수 있거나, 이것을 표적화 벡터의 도입 후에 도입시킬 수 있다. 2개 이상의 LTVEC를 세포 내로 도입시킬 때, 이들을 동시에 도입시킬 수 있거나, 대안적으로 이들을 별개로 도입시킬 수 있다.

[0334] **IX. 세포 및 동물**

[0335] 본 명세서에 제공된 다양한 조성물 및 방법은 세포, 예를 들어 동물 유래의 세포를 사용한다. 이러한 세포는 비인간 세포일 수 있으며, 비인간 동물로부터 유래될 수 있다. 이러한 세포는, 예를 들어 진균 세포 (예컨대, 효모), 식물 세포, 동물 세포, 포유류 세포 및 인간 세포를 포함하는 진핵 세포일 수 있다. 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, 만능성 세포, 예를 들어 비인간 만능성 세포 (예를 들어, 마우스 배아 줄기 (ES) 세포 또는 래트 ES 세포) 또는 인간 만능성 세포, 또는 비만능성 세포일 수 있다. 전능성 세포는 모든 세포 유형을 생성시킬 수 있는 미분화 세포를 포함하고, 만능성 세포는 하나 초과와 분화된 세포 유형으로 발달하는 능력을 지니는 미분화 세포를 포함한다. 이러한 만능성 및/또는 전능성 세포는, 예를 들어 배아 줄기 (ES) 세포 또는 ES 유사 세포, 예를 들어 유도 만능성 줄기 (iPS) 세포일 수 있다. 배아 줄기 세포는 배아 내로의 도입 시에 발생 중인 배아의 임의의 조직에 기여할 수 있는 배아 유래 전능성 또는 만능성 세포를 포함한다. ES 세포는 배반포의 내부 세포 덩어리로부터 유래될 수 있으며, 3가지의 척추동물 배엽층 (내배엽, 외배엽 및 중배엽)의 어떠한 세포로도 분화할 수 있다. 이러한 세포는 또한 조혈 줄기 세포 또는 신경 줄기 세포일 수 있다.

[0336] 진핵 세포는 또한 1차 체세포가 아닌 세포일 수 있다. 체세포는 배우자(gamete), 생식 세포, 생식모세포(gametocyte) 또는 미분화 줄기 세포가 아닌 임의의 세포를 포함할 수 있다.

[0337] 진핵 세포는 또한 1차 세포를 포함한다. 1차 세포는 유기체, 기관 또는 조직으로부터 직접 단리된 세포 또는 세포 배양물을 포함한다. 1차 세포는 형질전환되지 않은 또는 비불멸인 세포를 포함한다. 이것은, 조직 배양으로 이전에 계대되지 않았거나 조직 배양으로 이전에 계대되었지만 조직 배양으로 무한히 계대될 수 없는, 유기체, 기관 또는 조직으로부터 얻어진 임의의 세포를 포함한다. 이러한 세포는 종래의 기법에 의해 단리될 수 있으며, 예를 들어 체세포, 조혈 세포, 내피 세포, 상피 세포, 섬유아세포, 중간엽 세포, 각질세포, 멜라닌세포, 단핵구, 단핵 세포, 지방세포, 지방전구세포, 뉴런, 아교 세포, 간세포, 골격 근아세포 및 평활근 세포를 포함한다. 예를 들어, 1차 세포는 결합 조직, 근육 조직, 신경계 조직 또는 상피 조직으로부터 유래될 수 있다.

[0338] 진핵 세포는 또한 불멸화 세포를 포함한다. 불멸화 세포는, 정상적으로는 무한히 증식하지 않을 것이지만, 돌연변이 또는 변경으로 인해 정상적인 세포 노화를 피했고 대신에 계속 분열할 수 있는 다세포 유기체 유래의 세포를 포함한다. 이러한 돌연변이 또는 변경은 자연적으로 발생하거나 의도적으로 유도될 수 있다. 불멸화 세포의 예는 차이나이즈 햄스터 난소 (CHO) 세포, 인간 배아 신장 세포 (예를 들어, HEK 293 세포) 및 마우스 배아 섬유아세포 (예를 들어, 3T3 세포)를 포함한다. 많은 불멸화 세포 유형이 당업계에 널리 공지되어 있다.

[0339] 불멸화 또는 1차 세포는, 전형적으로 배양에 또는 재조합 유전자 또는 단백질을 발현하는 데 사용되는 세포를 포함한다.

[0340] 세포, 만능성 및/또는 전능성 세포, ES 세포, 공여체 세포, 및/또는 숙주 배아와 관련하여, 용어 "동물"은 포유

류, 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, 소 종, 예를 들어 암소, 거세 황소 등; 양 종, 예를 들어 양, 염소 등; 및 돼지 종, 예를 들어 피그 및 수태지)을 포함한다. 조류는, 예를 들어 닭, 칠면조, 타조, 거위, 오리 등을 포함한다. 가축 및 농업용 동물도 포함된다. 용어 "비인간 동물"은 인간을 제외한다.

[0341] 마우스 만능성 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]이 참조된다. 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 만능성 및/또는 전능성 세포는 또한 전술한 129 계통과 전술한 C57BL/6 계통의 혼합체 (예를 들어, 50% 129와 50% C57BL/6)로부터 유래될 수 있다. 마찬가지로, 마우스 만능성 및/또는 전능성 세포는 전술한 129 계통의 혼합체 또는 전술한 BL/6 계통의 혼합체 (예를 들어, 129S6 (129/SvEvTac) 계통)으로부터 유래될 수 있다. 마우스 ES 세포의 특정 예는 VGF1 마우스 ES 세포이다. 예를 들어, 전체적으로 본 명세서에 참고로 포함되는 문헌[Auerbach *et al.* (2000) *Biotechniques* 29, 1024-1028, 1030, 1032]이 참조된다.

[0342] 래트 만능성 및/또는 전능성 세포는, 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 F6을 포함하는 임의의 래트 계통으로부터 유래될 수 있다. 래트 만능성 및/또는 전능성 세포는 또한 상기에 열거된 2개 이상의 계통의 혼합체로부터 유래되는 계통으로부터 얻어질 수 있다. 예를 들어, 래트 만능성 및/또는 전능성 세포는 DA 계통 또는 ACI 계통으로부터 유래될 수 있다. ACI 래트 계통은 블랙 아구티(black agouti)를 가지며 배와 발이 흰색이고 *RTI^{av1}* 하플로타입(haplotype)을 갖는 것을 특징으로 한다. 이러한 계통은 할란 연구소(Harlan Laboratories)를 포함하는 다양한 공급처로부터 입수가 가능하다. ACI 래트 유래의 래트 ES 세포주의 예는 ACI.G1 래트 ES 세포이다. 다크 아구티 (DA) 래트 계통은 아구티 털(agouti coat) 및 *RTI^{av1}* 하플로타입을 갖는 것을 특징으로 한다. 이러한 래트는 찰스 리버 연구소(Charles River Laboratories) 및 할란 연구소를 포함하는 다양한 공급처로부터 입수가 가능하다. DA 래트 유래의 래트 ES 세포주의 예는 DA.2B 래트 ES 세포주 및 DA.2C 래트 ES 세포주이다. 일부의 경우에, 래트 만능성 및/또는 전능성 세포는 근친교배 래트 계통으로부터 유래된다. 예를 들어, 둘 모두 전체적으로 본 명세서에 참고로 포함되는, 2014년 2월 20일에 출원된 미국 특허 출원 공개 제2014/0235933 A1호 및 2014년 4월 16일에 출원된 미국 특허 출원 공개 제2014/0310828 A1호가 참조된다.

[0343] 인간 만능성 세포의 예는 인간 ES 세포, 인간 성체 줄기 세포, 발달적으로 제한된 인간 전구 세포, 및 인간 유도 만능성 줄기 (iPS) 세포, 예를 들어 감작(primed) 인간 iPS 세포 및 미감작(*naïve*) 인간 iPS 세포를 포함한다. 예를 들어, 전체적으로 본 명세서에 참고로 포함되는, 2014년 10월 15일에 출원된 미국 특허 출원 제 14/515,503호가 참조된다. 유도 만능성 줄기 세포는 분화된 성체 세포로부터 직접 유래될 수 있는 만능성 줄기 세포를 포함한다. 인간 iPS 세포는 특정 세트의 리프로그래밍 인자(reprogramming factor)를 세포 내로 도입시켜 생성될 수 있으며, 이 리프로그래밍 인자는, 예를 들어 Oct3/4, Sox 패밀리를 전사 인자 (예컨대, Sox1, Sox2, Sox3, Sox15), Myc 패밀리를 전사 인자 (예컨대, c-Myc, l-Myc, n-Myc), 크뢰펠 유사 패밀리를 (Krüppel-like family, KLF) 전사 인자 (예컨대, KLF1, KLF2, KLF4, KLF5), 및/또는 관련 전사 인자, 예를 들어 NANOG, LIN28 및/또는 Glis1을 포함할 수 있다. 인간 iPS 세포는, 예를 들어 miRNA, 전사 인자의 작용을 모방하는 작은 분자, 또는 계열 지시자의 사용에 의해서도 생성될 수 있다. 인간 iPS 세포는 3가지의 척추동물 배엽층, 예를 들어 내배엽, 외배엽, 또는 중배엽의 어떠한 세포로도 분화하는 이의 능력을 특징으로 한다. 인간 iPS 세포는 또한 적합한 시험관내 배양 조건 하에서 무한히 증식하는 이의 능력을 특징으로 한다. 예를 들어, 문헌[Takahashi and Yamanaka (2006) *Cell* 126:663-676]이 참조된다. 감작 인간 ES 세포 및 감작 인간 iPS 세포는 이식후 배반엽상피 세포(epiblast cell)의 특징과 유사한 특징을 나타내며 계열 특수화 및 분화에 대해 수임된 세포를 포함한다. 미감작 인간 ES 세포 및 미감작 인간 iPS 세포는 이식전 배아의 내부 세포 덩어리의 ES 세포의 특징과 유사한 특징을 나타내며 계열 특수화에 대해 수임되지 않은 세포를 포함한다. 예를 들어, 문헌[Nichols and Smith (2009) *Cell Stem Cell* 4:487-492]이 참조된다.

[0344] 숙주 배아에 이식된 세포는 "공여체 세포"로 지칭될 수 있다. 유전자 변형된 만능성 및/또는 전능성 세포는 숙

주 배아와 동일한 계통 또는 상이한 계통으로부터 유래될 수 있다. 마찬가지로, 대리모는 유전자 변형된 만능성 및/또는 전능성 세포 및/또는 숙주 배아와 동일한 계통으로부터 유래될 수 있거나, 대리모는 유전자 변형된 만능성 및/또는 전능성 세포 및/또는 숙주 배아와 상이한 계통으로부터 유래될 수 있다.

[0345] 다양한 숙주 배아를 본 명세서에 개시된 방법 및 조성물에 사용할 수 있다. 예를 들어, 표적화된 유전자 변형을 갖는 만능성 및/또는 전능성 세포를 상응하는 유기체 유래의 상실기 전시기의 배아 (예컨대, 8 세포기의 배아) 내로 도입시킬 수 있다. 예를 들어, 미국 특허 제7,576,259호, 미국 특허 제7,659,442호, 미국 특허 제7,294,754호 및 미국 특허 출원 공개 제20080078000 A1호가 참조되며, 이들 모두는 전체적으로 본 명세서에 참고로 포함된다. 다른 경우에, 공여체 ES 세포는 상실기 전시기, 예를 들어 2 세포기, 4 세포기, 8 세포기, 16 세포기, 32 세포기 또는 64 세포기의 숙주 배아에 이식될 수 있다. 또한, 숙주 배아는 배반포일 수 있거나 배반포전의 배아, 상실기 전시기의 배아, 상실기의 배아, 비치밀화(uncompacted) 상실기의 배아 또는 치밀화(compact) 상실기의 배아일 수 있다. 마우스 배아를 사용할 때, 숙주 배아의 시기는 문헌[Theiler (1989) "The House Mouse: Atlas of Mouse Development", Springer-Verlag, New York]에 기재된 테일러 시기(Theiler stage)를 참조하여 테일러 시기 1(TS1), TS2, TS3, TS4, TS5 및 TS6일 수 있다. 예를 들어, 테일러 시기는 TS1, TS2, TS3 및 TS4로부터 선택될 수 있다. 일부의 경우에, 숙주 배아는 투명대를 포함하고, 공여체 세포는 투명대의 구멍을 통해 숙주 배아 내로 도입되는 ES 세포이다. 다른 경우에, 숙주 배아는 투명대가 더 적은 (zona-less) 배아이다. 또 다른 경우에, 상실기의 숙주 배아는 응집된다.

[0346] **X. 변형된 표적 게놈 유전자좌를 갖는 세포를 확인하는 방법**

[0347] 상기 방법들 중 일부는 변형된 표적 게놈 유전자좌 (예를 들어, 변형된 게놈)를 갖는 세포를 확인하는 단계를 추가로 포함한다. 다양한 방법을 사용하여 결실 또는 삽입과 같은 표적화된 변형을 갖는 세포를 확인할 수 있다. 이러한 방법은 표적 유전자좌에 표적화된 변형을 갖는 하나의 세포를 확인하는 단계를 포함할 수 있다. 스크리닝을 수행하여 변형된 게놈 유전자좌를 갖는 이러한 세포를 확인할 수 있다.

[0348] 스크리닝 단계는 모염색체의 대립유전자 변형 (MOA)을 평가하는 정량적인 검정을 포함할 수 있다. 예를 들어, 실시간 PCR과 같은 정량적인 PCR (qPCR)을 통해 정량적인 검정을 수행할 수 있다. 실시간 PCR은 표적 유전자좌를 인식하는 제1 프라이머 세트 및 비표적화된 기준 유전자좌를 인식하는 제2 프라이머 세트를 이용할 수 있다. 프라이머 세트는 증폭된 서열을 인식하는 형광 프로브를 포함할 수 있다.

[0349] 다른 경우에, 표적화된 유전자 변형을 갖는 세포는, 예를 들어 서던 블롯 분석, DNA 서열분석, PCR 분석 또는 표현형 분석을 포함하는 방법을 사용하여 선택된다. 이어서, 이러한 세포를 본 명세서에 기술된 다양한 방법 및 조성물에 사용한다.

[0350] 적합한 정량적인 검정의 다른 예는 형광 매개 원위치 혼성화(fluorescence-mediated in situ hybridization, FISH), 비교 게놈 혼성화, 등은 DNA 증폭, 고정화된 프로브(들)에 대한 정량적인 혼성화, Invader Probes®, MMP assays®, TAQMAN® 분자 비콘(Molecular Beacon) 또는 Eclipse™ 프로브 기술을 포함한다 (예를 들어, 전체적으로 본 명세서에 참고로 포함되는 미국 특허 출원 공개 제2005/0144655호 참조).

[0351] 스크리닝 단계는 대체로 또한 아암 특이적(arm-specific) 검정을 포함하는데, 이 검정은 표적 게놈 유전자좌 내로의 핵산 삽입물의 정확한 표적화된 삽입을 표적 게놈 유전자좌 외부의 게놈 위치 내로의 핵산 삽입물의 무작위적 유전자도입식(transgenic) 삽입과 구별하는 데 사용되는 검정이며, 또한 이를 사용하여 단일 작제물로의 2개 이상의 중복 LTVEC의 정확한 조립을 검출한다. 표적화된 변형을 스크리닝하는 종래의 검정, 예를 들어 긴 범위 PCR 또는 서던 블롯팅은 삽입된 표적화 벡터를 표적화된 유전자좌에 연결시킨다. 그러나, 큰 상동성 아암의 크기로 인해, LTVEC는 이러한 종래의 검정에 의한 스크리닝을 허용하지 않는다. LTVEC 표적화를 스크리닝하기 위해, 대립유전자 상실(loss-of-allele, LOA) 검정 및 대립유전자 획득(gain-of-allele, GOA) 검정을 포함하는 대립유전자 변형 (MOA) 검정을 사용할 수 있다 (예를 들어, 모든 목적을 위해 전체적으로 본 명세서에 참고로 포함되는, 미국 특허 출원 공개 제2014/0178879호 및 문헌[Frendewey *et al.* (2010) *Methods Enzymol.* 476:295-307] 참조). 대립유전자 상실 (LOA) 검정은 종래의 스크리닝 논리를 반전시킨 것이며, 돌연변이가 유도된 고유의 유전자좌의 카피 수를 정량화한다. 정확하게 표적화된 세포 클론의 경우, LOA 검정은 (X 또는 Y 염색체 상에 존재하지 않는 유전자에 대하여) 두 고유의 대립유전자 중 하나를 검출하는데, 다른 대립유전자는 표적화된 변형에 의해 파괴되어 있다. 동일한 원리를 역으로 대립유전자 획득 (GOA) 검정으로 적용하여, 삽입된 표적화 벡터의 카피 수를 정량화할 수 있다. 예를 들어, GOA 검정과 LOA 검정의 병용은, 한 카피의 고유의 표적 유전자를 상실하고 한 카피의 약물 내성 유전자 또는 다른 삽입된 마커를 획득한 것으로서 정확하게 표적화된 이형접합성 클론을 밝혀낼 것이다.

[0352] 예로서, 정량적인 폴리머라제 연쇄 반응 (qPCR)을 대립유전자 정량화 방법으로 사용할 수 있지만, 0개, 1개 및 2개 카피의 표적 유전자 간의 차이 또는 0개, 1개 및 2개 카피의 핵산 삽입물 간의 차이를 확실히 구별할 수 있는 임의의 방법을 사용하여 MOA 검정을 개발할 수 있다. 예를 들어, TAQMAN®을 사용하여 게놈 DNA 샘플에서 DNA 주형의 카피 수를, 특히 기준 유전자와 비교하여 정량화할 수 있다 (예를 들어, 모든 목적을 위해 전체적으로 본 명세서에 참고로 포함되는 미국 특허 제6,596,541호 참조). 기준 유전자를 표적 유전자(들) 또는 유전자좌(들)와 동일한 게놈 DNA에서 정량화한다. 따라서, 2회의 TAQMAN® 증폭 (이의 각각의 프로브마다)을 수행한다. 하나의 TAQMAN® 프로브는 기준 유전자의 "Ct" (역치 사이클(Threshold Cycle))를 결정하는 한편, 다른 프로브는 성공적인 표적화에 의해 대체되는 표적화된 유전자(들) 또는 유전자좌(들) 영역의 Ct를 결정한다 (즉, LOA 검정). Ct는 각각의 TAQMAN® 프로브에 대한 출발 DNA량을 반영하는 양이며, 즉, 덜 풍부한 서열은 역치 사이클에 도달하기 위해 더 많은 PCR 사이클을 필요로 한다. TAQMAN® 반응을 위해 주형 서열의 카피 수를 절반으로 감소시키는 것은 약 1 Ct 단위의 증가를 가져올 것이다. 표적 유전자(들) 또는 유전자좌(들) 중 하나의 대립유전자가 상동 재조합에 의해 대체된 세포에서의 TAQMAN® 반응은 비표적화된 세포의 DNA와 비교할 때 기준 유전자에 대한 Ct의 증가 없이 표적 TAQMAN® 반응에 대하여 1 Ct의 증가를 가져올 것이다. GOA 검정의 경우, 다른 TAQMAN® 프로브를 사용하여 성공적인 표적화에 의해 표적화된 유전자(들) 또는 유전자좌(들)를 대체하는 핵산 삽입물의 Ct를 결정할 수 있다.

[0353] LTVEC에 의한 정확한 표적화를 검증하기 위해 표준 LOA 및 GOA 검정을 증대시키는 것이 유용할 수 있다. 예를 들어, LOA 검정과 GOA 검정 단독으로는 정확하게 표적화된 세포 클론을, 표적 게놈 유전자좌의 Cas 유도 결실이 게놈의 다른 곳에서의 LTVEC의 무작위 통합과 일치하는 클론과 구별할 수 없다. 표적화된 세포에서의 선택 압력은 선택 카세트에 기반하기 때문에, 게놈의 다른 곳에서의 LTVEC의 무작위적 유전자도입식 통합은 대체로 LTVEC의 선택 카세트 및 인접 영역을 포함하지만 LTVEC의 더 먼 영역을 제외할 수 있다. 예를 들어, LTVEC의 일부가 게놈 내로 무작위 통합되고, LTVEC가 3' 상동성 아암에 인접한 선택 카세트와 함께 길이가 대략 5 kb 이상인 핵산 삽입물을 포함하는 경우, 일부의 경우에, 5' 상동성 아암이 아닌 3' 상동성 아암이 선택 카세트와 함께 유전자도입식으로 통합될 것이다. 대안적으로, 선택 카세트가 5' 상동성 아암에 인접한 경우, 일부의 경우에, 3' 상동성 아암이 아닌 5' 상동성 아암이 선택 카세트와 함께 유전자도입식으로 통합될 것이다. 예로서, LOA 검정과 GOA 검정을 사용하여 LTVEC의 표적화된 통합을 평가하고, GOA 검정이 LTVEC의 선택 카세트 또는 임의의 다른 독특한 (아암이 아닌) 영역에 대한 프로브를 이용하는 경우, LTVEC의 무작위적 유전자도입식 통합과 조합된 표적 게놈 유전자좌에서의 이형접합성 결실은 표적 게놈 유전자좌에서의 LTVEC의 이형접합성 표적화된 통합과 동일한 판독값을 제공할 것이다. LTVEC에 의한 정확한 표적화를 검증하기 위해, 아암 특이적 검정을 LOA 검정 및/또는 GOA 검정과 함께 사용할 수 있다.

[0354] 아암 특이적 검정은 LTVEC 상동성 아암에서 DNA 주형의 카피 수를 결정한다. 이러한 상동성 아암은, 다른 LTVEC와 중복되지 않고 세포의 표적 서열과 상응하는 LTVEC의 상동성 아암 (예를 들어, 마우스 세포의 게놈 표적 서열과 중복되는 상동성 아암 (mArm))을 포함할 수 있다. 이러한 상동성 아암은 또한 2개의 중복 LTVEC에 존재하는 중복 상동성 아암 (예를 들어, 제1의 LTVEC의 3' 상동성 아암과 제2의 LTVEC의 5' 상동성 아암의 중복 인간 서열 (hArm))을 포함할 수 있다. 다수의 중복 LTVEC가 세포 내로 도입되는 실험의 경우, 스크리닝은 대체로 LOA 검정, 모든 독특한 삽입된 서열에 대한 GOA 검정, 및 모든 상동성 영역 (즉, LTVEC와 세포의 표적 서열 사이의 그리고 2개의 상이한 중복 LTVEC 사이의 상동성 영역)에 대한 아암 특이적 검정을 포함한다. 예로서, 야생형 마우스 표적 유전자좌를 인간화하기 위해 마우스 세포 내로 도입되는 3개의 중복 LTVEC의 경우, 이형접합성 표적화된 삽입에 대한 예상 카피 수는 하기와 같을 것이다: 2 카피의 5' mArm (5' 마우스 표적 서열과 중복되는 상동성 아암), 1 카피의 hArm1 (LTVEC 1과 LTVEC 2 사이의 중복 서열), 1 카피의 hArm2 (LTVEC 2와 LTVEC 3 사이의 중복 서열) 및 2 카피의 3' mArm (3' 마우스 표적 서열과 중복되는 상동성 아암). 상기 예에서, 2 초과 카피의 mArm 카피 수는 대체로 표적 게놈 유전자좌에서라기보다는 표적 게놈 유전자좌의 외부에서의 무작위적 유전자도입식 LTVEC 통합을 나타낼 것이며, 이는 바람직하지 않다. 정확히 표적화된 클론은 2의 mArm 카피 수를 보유할 것이다. 또한, 이러한 아암 특이적 검정에서 2 미만의 mArm 카피 수는 대체로 결실에 대해 표적화된 영역을 벗어나 연장되는 큰 Cas 매개 결실을 나타낼 것이며, 이 또한 바람직하지 않다. 마찬가지로, 이형접합성 표적화된 변형의 경우, hArm1과 hArm2에 대한 1의 카피 수는 대체로 모든 3개의 LTVEC가 단일 작제물로 조립되었음을 나타낼 것이다.

[0355] 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. 서열의 다양한 버전이 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 간행물, 웹사이트 등의 다양한 버전이 다양한 시점에서 공개된 경우, 달리 표시되지 않는 한 본 출원의 유효 출원일에서 가장 최근에 공개된 버전이 의도된다. 본 발명의 임의의 특징부, 단계, 요소, 실시 형태, 또는 태양은 달리 구체적으로 표시되지 않는 한 임의의 다른 것과 조합하여 사용될 수 있다. 본 발명은 명료함과 이해를 목적으로 예시 및 예로서 어느 정도 상세히 설명되었지만, 특정 변화 및 변형이 첨부된 청구범위의 범주 내에서 실시될 수 있음이 명백할 것이다.

[0356] 실시예

[0357] 하기 실시예는 당업자에게 본 발명을 제조하고 사용하는 방법의 완전한 개시 및 설명을 제공하기 위해 제시되며, 본 발명자들이 그들의 발명으로 간주하는 것의 범주를 제한하고자 하는 것이 아니고, 하기 실험이 수행되는 모든 또는 유일한 실험인 것을 나타내는 것으로 의도되지 않는다. 사용된 숫자 (예를 들어, 양, 온도 등)와 관련하여 정확성을 확보하기 위한 노력이 이루어졌지만, 일부 실험 오차 및 편차가 고려되어야 한다. 달리 표시되지 않는 한, 부는 중량부이고, 분자량은 중량 평균 분자량이며, 온도는 섭씨 단위이고, 압력은 대기압 또는 대기압 부근이다.

[0358] 실시예 1: 징크 핑거 뉴클레아제와 조합하여 2개의 LTVEC를 사용한 TCR 알파 유전자좌의 표적화

[0359] 2개의 큰 표적화 벡터 (LTVEC)가 단일 표적화 단계로 게놈 유전자좌를 변형시키도록 이중 표적화 시스템을 설계하였다. 도 1에 도시된 바와 같이, 하이그로마이신 선택 카세트를 포함하는 마우스 염색체 14 상에 TCR 알파 유전자좌의 이형접합성 변형을 갖는 세포를 이중 표적화 시스템에 의해 표적화하여 추가의 Igκ 가변 유전자 세그먼트를 포함하는 ES 세포를 생성시켰다.

[0360] 도 1에 요약된 이러한 이중 표적화 접근법은, ES 세포의 표적 유전자좌에 또는 이 부근에 이중 가닥 절단부를 생성시키는 엔도뉴클레아제 (예를 들어, 징크 핑거 뉴클레아제)를 인코딩하는 뉴클레오티드 서열과 함께 2개의 상이한 큰 표적화 벡터 (LTVEC)의 이중 표적화 또는 공동 전기천공을 포함한다.

[0361] 이러한 접근법에서, 제1의 큰 표적화 벡터 (MAID 1710으로 표지됨)는 인간 Vκ1-5 및 Vκ1-6 유전자 세그먼트의 서열을 포함하는 30 kb의 3' 상동성 아암, 인간 Vκ3-7 내지 Vκ3-15 유전자 세그먼트를 포함하는 120 kb의 서열, 및 인간 Vκ1-16 유전자 세그먼트를 포함하는 20 kb의 5' 영역 ("중복 영역")을 포함하였다. 제2의 큰 표적화 벡터 (MAID 6600으로 표지됨)는 20 kb의 3' 중복 영역 (제1 벡터에서와 동일한 인간 Vκ1-16 유전자 세그먼트를 포함하는 영역), 인간 Vκ1-17 내지 Vκ2-30 유전자 세그먼트를 포함하는 140 kb의 서열, FRT-Ub-Neo-FRT 선택 카세트 및 15.5 kb의 3' 마우스 TCR A 상동성 아암을 포함하였다.

[0362] 하이그로마이신 내성 유전자 내의 표적 서열을 인식하고 절단하여 표적 TCR A 유전자좌에서 2개의 LTVEC의 상동 재조합을 촉진시키는, 징크 핑거 뉴클레아제 (ZFN)를 설계하였다. 도 1에서 생성된 ES 세포 (MAID 6548, 모든 인간 Jκ 세그먼트 및 4개의 기능적 인간 Vκ 유전자 세그먼트에 대해 이형접합성임)를 상기 기술된 2개의 큰 표적화 벡터 (MAID6600 및 MAID1700-트리밍(trimming)됨) 및 ZFN (1/2)의 각 절단을 발현하는 2개의 플라스미드 - 이는 하이그로마이신 내성 유전자의 인식 서열에 결합하여 표적 부위 (TGCGATCGTGGCGCCGAtcttagCCAGACGAGCGGGTTCGG (서열 번호 2); 절단 부위는 소문자로 나타나 있음) (표 1 참조)에서 이중 가닥 절단을 촉매함 - 와 함께 전기천공하였다. 하이그로마이신을 표적화하도록 하기 2개의 추가 ZFN을 설계하였다: 뉴클레오티드 서열 CGCTGCGCCGATCTtagccaGACGAGCGGGTTCGG (서열 번호 3)에서 하이그로마이신 유전자를 표적화하는 ZFN(3/4); 및 뉴클레오티드 서열 AGCGTGTCGACCTGATGcagctcTCGGAGGGCGAAGAA (서열 번호 4)에서 하이그로마이신 유전자를 표적화하는 ZFN(5/6) (표 1 참조).

[0363] [표 1]

하이그로마이신 징크 핑거 뉴클레아제 결합 및 절단 부위 (상보적 가닥은 표시되지 않음)

징크 핑거 뉴클레아제	징크 핑거 절단	결합 서열 (5'-3')	절단 부위 (5'-3')
Hyg-ZFN(1,2)	Hyg-ZF1	TGCGATCGTGGCGCCGA (서열 번호 5)	TCTTAG
	Hyg-ZF2	CGAACCCGCTCGTCTGG (서열 번호 6)	(서열 번호 11)
Hyg-ZFN(3,4)	Hyg-ZF3	CGTGGCGCCGATCT (서열 번호 7)	TAGCCA
	Hyg-ZF4	CGAACCCGCTCGTC (서열 번호 8)	(서열 번호 12)
Hyg-ZFN(5,6)	Hyg-ZF5	AGCGTGTCGACCTGATG (서열 번호 9)	CAGCTC
	Hyg-ZF6	TTCTTCGCCCTCCGA (서열 번호 10)	(서열 번호 13)

[0364] [0365] 2개의 큰 표적화 벡터를 상동 재조합에 의해 DNA 서열 내로 삽입하여, Hyg 선택 카세트를 함유하고 이를 둘러싸

는 영역을 대체하였다. 생성된 ES 세포는 내인성 TCR A 유전자좌에 인간 J_κ1 내지 J_κ5 및 V_κ4-1 내지 V_κ2-30 유전자 세그먼트를 포함하는 인간 면역글로불린 가변 도메인을 함유하였다. 도 1에 나타내고 하기 표 2에 열거된 프로브와 프라이머를 사용하는 상기 기술된 TAQMAN® 검정 (본 명세서에 참고로 포함되는 문헌[Lie and Petropoulos, 1998. *Curr. Opin. Biotechnology* 9:43-48])을 사용하여 2개의 큰 표적화 벡터의 성공적인 혼입을 확인하였다 (GOA = 대립유전자 획득; LOA = 대립유전자 상실; 카피 수 = 서열의 카피 수를 확인하여 유전자 도입식 통합 대 표적화된 통합을 추적함; hArm1 = 제1의 큰 표적화 벡터 (MAID 1710)의 30 kb의 3' 상동성 아암; hArm2 = 제1의 큰 표적화 벡터 (MAID 1710)와 제2의 큰 표적화 벡터 (MAID 6600)의 20 kb의 중복부, mArm = 제2 표적화 벡터 (MAID 6600)의 15.5 kb의 5' 상동성 아암, WT 마우스 대조군 - 마우스 TCR A 유전자좌에 존재하는 서열). 아암 특이적 검정으로 지칭되는 LTVEC의 상동성 아암의 서열을 인식하는 실시간 PCR 검정을 사용하여 마우스 게놈 내로의 LTVEC의 정확한 표적화를 검증하였다. 이러한 아암 특이적 검정의 카피 수를 결정하는 것은, 예를 들어 2의 mArm 카피 수를 보유하는 정확하게 표적화된 ES 클론을, 표적 마우스 유전자좌의 Cas9 유도 결실이 게놈의 다른 곳에서의 LTVEC의 무작위적 통합과 일치하는 클론 - 이 경우 mArm 카피 수는 3 (또는 그 이상)일 것임 - 과 구별하는 것을 돕도록 추가의 설명을 제공하였다.

[0366] [표 2]

TAQMAN 프라이머 및 프로브

유전자	검정	정방향 프라이머	프로브	역방향 프라이머
HYG	LOA	TGCGGCCGATCTTA GCC (서열 번호 14)	ACGAGCGGGTTCGG CCCATTC (서열 번호 15)	TTGACCGATTCCTT GCGG (서열 번호 16)
HYG-U	LOA	CGACGTCTGTCGAG AAGTTTCTG (서열 번호 17)	AGTTCGACAGCGTG TCCGACCTGA (서열 번호 18)	CACGCCCTCTACA TCGAA (서열 번호 19)
Hyg-D	LOA	TGTCGGGGCGTACAC AAATCG (서열 번호 20)	CCGTCTGGACCGAT GGCTGTGT (서열 번호 21)	GGCGTCGGTTTCC ACTATC (서열 번호 22)
PGKp1 (Hyg 프로모터)	LOA	CAAATGGAAGTAGC ACGTCTCACT (서열 번호 23)	CTCGTGCAGATGGA CAGCACCGC (서열 번호 24)	CCGCTGCCCAAAAG G (서열 번호 25)
hIlgK6	hArm1 의 카피수	GTC AAGCACTGCTG GCACAC (서열 번호 26)	AACCCTTGTGCTAT TGAATTGCTATGCT GTCAG (서열 번호 27)	TGTGTAGACCCTC CGCCAC (서열 번호 28)
hIlgK12 (MAID 1710 삽입물)	GOA	TTGCCTTTCTCACA CCTGCAG (서열 번호 29)	CAGCCCATCCTGTC ACTTCGCTGGA (서열 번호 30)	TGGCCCAACAGTAC AGCTCAG (서열 번호 31)
hIlgK13	hArm2 의 카피수	TCAGTCAATCACCT TTCCACG (서열 번호 32)	TCCCAGGTAGCCT CATGAACCAATGTT (서열 번호 33)	CACATTACTGAGTC CCCACAGGG (서열 번호 34)
hIlgK14	hArm2 의 카피수	CATTGTCAAAGAAG CACTGGAAATG (서열 번호 35)	ACCATTGCAGTTTA CCCACGGTTAGGAT TTTT (서열 번호 36)	TCTTGCAATGGGAT CATCAGATG (서열 번호 37)
Neo	GOA	GGTGGAGAGGCTAT TCGGC (서열 번호 38)	TGGGCACAACAGAC AATCGGCTG (서열 번호 39)	GAACACGGCGGCAT CAG (서열 번호 40)
hIlgK15	GOA	CAGGTGCAAAGGTG ACCACAG (서열 번호 41)	I02102TGGGTCTGTC CCATCCATGCA (서열 번호 42)	GGCAGCCTGAGTGT CAGAGC (서열 번호 43)
hIlgK25	GOA	GTTACGGCCCCACA GACTCTC (서열 번호 44)	TCCTCTTGAGCA ACCATGAAGTTCCC T (서열 번호 45)	CCTGAAGCCATGAG GGCAG (서열 번호 46)
hUbc-D (Neo 프로모터)	GOA	AGGGTAGGCTCTCC TGAATCG (서열 번호 47)	ACAGGCGCCGGACC TCTGGT (서열 번호 48)	CCAAAGAACTGAC GCCTCAC (서열 번호 49)
TCRA Arm4	mArm 의 카피수	GCGCCACATGAATT TGACCAG (서열 번호 50)	TGTACCCAATCTTC CAAAGAAAGAGCT G (서열 번호 51)	GGCATCCTGTCTC CCTTC (서열 번호 52)
모 1540m1	WT 마우스 대조군	CAGTAAGGGAAGA GACTACAACAGCAT (서열 번호 53)	TGCACACTGCTCAC CACTGCAAGCTAT (서열 번호 54)	TGCTGGTGCCCCA TCT (서열 번호 55)
모 1540m3	WT 마우스 대조군	GA ACTCAGCTATGA TAGTGTGGAATGTA (서열 번호 56)	CAGCCAGCAGCTG TGGGTCTC (서열 번호 57)	GCTCAGGGAGAACA CAGAACTTAGA (서열 번호 58)
hIlgK5	MAID 6548 서열	CCCCGTCTCCTCC TTTTTC (서열 번호 59)	TCATGTCCATTAAC CCATTTACCTTTTGC CCA (서열 번호 60)	TGCAAGTGCTGCCA GCAAG (서열 번호 61)

[0367]

[0368]

ES 세포에서 생성된 표적화된 유전자좌는 하기의 접합부 서열을 가졌으며, 여기서 마우스 서열은 괄호 안에 표

기되고, 인간 서열은 보통 글꼴로 표기되며, 다중 클로닝 부위는 볼드체로 표기되고, Frt 서열은 이탤릭체로 표기되어 있다 (표 3).

[0369] [표 3]

이중 ES 세포 표적화로부터 생성된 유전자좌의 접합부 서열

접합부	서열 번호	서열
마우스 Tcra/5' Frt	62	(GTCTTTTTGTCTTCACAGTTGAGCTTCA TCAAAGTCACATGGGTAAACTCTATGGAG TAGTCAGAACACACTCTCA) GAAGGGACTC CTGATTTCAAAGGGTACCGAAGTTCCTATT <i>CCGAAGTTCCTATTCTCTAGAAAGTATAGGA</i> <i>ACTTC</i>
3' Frt/인간 IgK	63	<i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG</i> <i>AAAGTATAGGA</i> <i>ACTTCCTAGGGTTTCACCG</i> GTGGCGCGCTAACAGAGAGGAAAGTCAAA TTATAAAGAATATGAGATTCAGAATTCTGA TTAAGTGTGG
인간 IgK/마우스 Tcra	64	GATAAATTATTTTGTTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAG GCGGCC GCCACCGCGGTGGAGCTC (AGGTTTCCG GTA

[0370]

[0371] 단리된 ES 세포 콜로니의 대립유전자 변형 (MOA) 스크리닝 결과, 2.81%의 표적화 효율에 대하여, 스크리닝된 960개의 콜로니 중 27개의 정확하게 표적화된 클론이 확인되었다.

[0372] 추가의 면역글로불린 가변 영역 유전자 세그먼트를 포함하는 TCR A 유전자좌를 생성시키기 위한 대안적인 전략은 연속적인 큰 표적화 벡터들을 사용한 연속 표적화를 포함한다 (예를 들어, 도 2 참조). 이와 같이, 모든 인간 J_κ 유전자 세그먼트 및 4개의 기능적 인간 V_κ 유전자 세그먼트에 대해 이형접합성인 ES 세포 (MAID 6548)를, 5'에서 3'의 방향으로, 15.5 kb의 5' 마우스 상동성 아암, Frt-Ub-Neo-Frt 선택 카세트, V_κ3-7 내지 V_κ3-15 유전자 세그먼트를 포함하는 120 kb의 단편, 및 V_κ1-5 및 V_κ1-6 유전자 세그먼트를 포함하는 30 kb의 3' 인간 상동성 아암 (MAID 6548 서열에도 존재함)을 포함하는 큰 표적화 벡터와 함께 전기천공하였다. 상기 표 2에 열거되고 도 2에 나타낸 프로브 및 프라이머를 사용하여 상기 기술된 TAQMAN® 검정으로 성공적인 혼입을 확인하였다: Hyg, hIgK5, hIgK6, hIgK12, Neo, 모 1540 ml, 모 1540ml. 특히, TCRA Arm4 및 hIgK6 프로브를 아암 특이적 프로브로 사용하여 LTVEC의 정확한 계능 표적화를 입증하였다. 프라이머와 프로브의 추가 세트인 hIgK10을 또한 사용하여 성공적인 혼입을 확인할 수 있다: 정방향 프라이머 - CGATTATGACTGGTTAGGTAGAAAGGTG (서열 번호 65); 프로브 - GCCACTGGTTCTCCAATGTTTCAATCCAT (서열 번호 66); 역방향 프라이머 - GGGAGTACTGGAGATCCCTAAGC (서열 번호 67).

[0373] ES 세포에서 생성된 표적화된 유전자좌는 하기의 접합부 서열을 가졌으며, 여기서 마우스 서열은 괄호 안에 표기되고, 인간 서열은 보통 글꼴로 표기되며, 다중 클로닝 부위는 볼드체로 표기되고, Frt 서열은 이탤릭체로 표기되어 있다 (표 4).

[0374] [표 4]

단일 ES 세포 표적화로부터 생성된 유전자좌의 접합부 서열

접합부	서열 번호	서열
마우스 Tcra/5' Frt	68	(TTGAGCTTCATCAAAGTCACATGGGTAA CTCTATGGAGTAGTCAGAACACACTCTTCA) GAAGGGACTCCTGATTTCAAAGGGTAC <i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAGAA</i> <i>AGTATAGGA</i> <i>ACTTC</i>
3' Frt/인간 IgK	69	<i>GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG</i> <i>AAAGTATAGGA</i> <i>ACTTCCTAGGGTTTCACCG</i> GTGGCGCGCCAGGACCCAGGCTCTGACACT CAGGCTGCCAATACAATTGCCATGAAGACA GATGTTGATG
인간 IgK/마우스 Tcra	64	GATAAATTATTTTGTTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAG GCGGCC GCCACCGCGGTGGAGCTC (AGGTTTCCG GTA

[0375]

[0376] 단리된 콜로니의 MOA 스크리닝 결과, 1.1%의 표적화 효율에 대하여, 스크리닝된 440개의 콜로니 중 5개의 정확하게 표적화된 클론이 확인되었다 (LTVEC 단독). LTVEC + ZFN 또는 LTVEC + CRISPR-Cas9로 표적화된 단리된 콜로니의 스크리닝 결과는 표 9에 나타나 있다.

[0377] 도 2에 도시된 단일 표적화의 완료 시에, 추가 V_K를 포함하는 큰 표적화 벡터들을 사용하여 ES 세포를 연속적으로 표적화하여 기능적 인간 면역글로불린 V_K 유전자 세그먼트의 전체 레퍼토리에 부가시켰다.

[0378] 또 다른 대안적인 전략에서는, 선택 (예를 들어, 하이그로마이신) 카세트(들)의 징크 핑거 뉴클레아제 또는 CRISPR 매개 파괴를 포함하는 이중 (2개의 큰 표적화 벡터) 또는 단일 (1개의 큰 표적화 벡터) 표적화 계획을 사용하여 연속적인 추가의 인간 Ig V_K 유전자 세그먼트의 이중 또는 단일 표적화를 달성할 수 있다.

[0379] 상기 기술된 표적화된 ES 세포를 공여체 ES 세포로 사용하여 VELOCIMOUSE® 방법에 의해 상실기 전시기의 배아, 예를 들어 8 세포기의 마우스 배아 내로 도입시켰다 (예를 들어, 미국 특허 제7,576,259호, 미국 특허 제 7,659,442호, 미국 특허 제7,294,754호 및 미국 특허 출원 공개 제2008-0078000 A1호 참조). 공여체 ES 세포를 포함하는 마우스 배아를 배반포기까지 인큐베이션하고, 이어서 대리모에 이식하여, 공여체 ES 세포로부터 완전히 유래된 F0 마우스를 생산한다. 독립적으로 키메라 인간 IgK V-마우스 Tcra C 유전자를 지니는 공여체 ES 세포로부터 완전히 유래된 F0 마우스를, 독특한 유전자 서열의 존재를 검출하는 대립유전자 변형 검정을 사용하여 유전자형 결정에 의해 확인한다.

[0380] **실시예 2: CRISPR/Cas 시스템과 조합하여 2개의 LTVEC를 사용한 하이그로마이신 유전자의 표적화**

[0381] 징크 핑거 뉴클레아제를 이용하는 실시예 1에 기술된 이중 표적화 방법을 또한 CRISPR/Cas9 시스템으로 수행하였다.

[0382] 하이그로마이신 내성 유전자 내의 다양한 표적 서열 (CRISPR 인식 서열)을 인식하도록 다양한 가이드 RNA (gRNA)를 설계하였다. 하이그로마이신 유전자 내의 CRISPR 인식 서열은 하기와 같았다: gRNA#1: ACGAGCGGGTTCGGCCATTCGG (서열 번호 70); gRNA#6: CTTAGCCAGACGACGGGTTCGG (서열 번호 71); gRNA#10: GCCGATCTTAGCCAGACGACGG (서열 번호 72); 및 gRNA#16: CGACCTGATGCAGCTCTCGGAGG (서열 번호 73). 하이그로마이신 유전자 내의 인식 서열의 위치는 도 3에 도시되어 있으며, 이는 표적화 벡터 MAID 1545에서 하이그로마이신의 CRISPR/Cas 매개 파괴를 도시한다. gRNA#1, gRNA#6, gRNA#10 및 gRNA#16을 스크리닝하였고, 이들은 하이그로마이신 유전자를 특이적으로 표적화하는 것으로 확인되었다 (도 3 참조). 다양한 하이그로마이신 특이적 gRNA를 사용한 1차 스크리닝의 결과가 표 5에 제공되어 있다.

[0383] [표 5]

하이그로마이신 특이적 gRNA 를 사용한 1 차 스크리닝 결과.

gRNA/ZFN	1 차 후보의 총수 (2 개의 플레이트/전기천공)	양성으로 재확인된 후보 수/ 재확인된 총수
gRNA1	5	2/2
gRNA6	6	1/1
gRNA10	19	5/5
gRNA16	91	8/8
ZFN 1/2	10	4/4

[0384] ES 세포, 예를 들어 도 1에서 생성된 ES 세포 (MAID 6548, 모든 인간 J_K 세그먼트 및 4개의 기능적 인간 V_K 유전자 세그먼트에 대해 이형접합성임)를, 단일 벡터 또는 Cas9 및 gRNA (예를 들어, gRNA#1, gRNA#6, gRNA#10 또는 gRNA#16)를 인코딩하는 다수의 벡터 - 이들 벡터는 하이그로마이신 내성 유전자 내의 표적 부위를 인식하여 절단함 - 와 더불어 2개의 큰 표적화 벡터 (실시예 1에 기술되어 있음)와 함께 전기천공하였다.

[0386] 2개의 큰 표적화 벡터를 상동 재조합에 의해 DNA 서열 내로 삽입하여, Hyg 선택 카세트를 함유하고 이를 둘러싸는 영역을 대체하였다. TAQMAN® 검정을 사용하여 2개의 큰 표적화 벡터의 성공적인 혼입을 확인하였다.

[0387] 상기 기술된 표적화된 ES 세포를 공여체 ES 세포로 사용하여 VELOCIMOUSE® 방법에 의해 상실기 전시기의 배아, 예를 들어 8 세포기의 마우스 배아 내로 도입시킬 것이다 (예를 들어, 미국 특허 제7,576,259호, 미국 특허 제 7,659,442호, 미국 특허 제7,294,754호 및 미국 특허 출원 공개 제2008-0078000 A1호 참조). 유전자 변형된 ES 세포를 포함하는 마우스 배아를 배반포기까지 인큐베이션하고, 이어서 대리모에 이식하여, 공여체 ES 세포로부터 완전히 유래된 F0 마우스를 생산한다. 공여체 ES 세포로부터 완전히 유래된 F0 마우스를, 독특한 유전자 서열의 존재를 검출하는 대립유전자 변형 검정을 사용하여 유전자형 결정에 의해 확인할 것이다.

[0388] **실시예 3: 징크 핑거 뉴클레아제와 조합하여 3개의 LTVEC를 사용한 TCR 알파 유전자좌의 표적화**

[0389] 3개의 큰 표적화 벡터 (LTVEC)가 단일 표적화 단계로 게놈 유전자좌를 변형시키도록 삼중 표적화 시스템을 설계하였다. 도 4에 도시된 바와 같이, 하이그로마이신 선택 카세트를 포함하는 마우스 염색체 14 상에 TCR 알파

유전자좌의 이형접합성 변형을 갖는 세포를 삼중 표적화 시스템에 의해 표적화하여 추가의 $Ig\kappa$ 가변 유전자 세그먼트를 포함하는 ES 세포를 생성시켰다.

[0390] 도 4에 요약된 이러한 삼중 표적화 접근법은, ES 세포의 표적 유전자좌에 또는 이 부근에 이중 가닥 절단부를 생성시키는 엔도뉴클레아제 (예를 들어, 징크 핑거 뉴클레아제 또는 Cas9 및 gRNA)를 인코딩하는 뉴클레오티드 서열과 함께 3개의 상이한 큰 표적화 벡터 (LTVEC) (MAID 6647, MAID 6600 및 MAID 1710)의 삼중 표적화 또는 공동 전기천공을 포함한다.

[0391] 이러한 접근법에서, 제1의 큰 표적화 벡터 (MAID 1710으로 표지됨)는 인간 $V_{\kappa}1-5$ 및 $V_{\kappa}1-6$ 유전자 세그먼트의 서열을 포함하는 30 kb의 3' 상동성 아암, 인간 $V_{\kappa}3-7$ 내지 $V_{\kappa}3-15$ 유전자 세그먼트를 포함하는 120 kb의 서열, 및 인간 $V_{\kappa}1-16$ 유전자 세그먼트를 포함하는 20 kb의 5' 영역 ("중복 영역")을 포함한다. 제2 큰 표적화 벡터 (MAID 6600으로 표지됨)는 20 kb의 3' 중복 영역 (제1 벡터에서와 동일한 인간 $V_{\kappa}1-16$ 유전자 세그먼트를 포함하는 영역), 인간 $V_{\kappa}1-17$ 내지 $V_{\kappa}2-24$ 유전자 세그먼트를 포함하는 140 kb의 서열, 및 인간 $V_{\kappa}3-25$ 내지 $V_{\kappa}2-30$ 을 포함하는 60 kb의 5' 영역 ("중복 영역")을 포함하였다. 제3의 큰 표적화 벡터 (MAID 6647로 표지됨)는 60 kb의 3' 중복 영역 (제2 벡터에서와 동일한 $V_{\kappa}3-25$ 내지 $V_{\kappa}2-30$ 을 포함하는 영역), 인간 $V_{\kappa}3-31$ 내지 $V_{\kappa}2-40$ 을 포함하는 90 kb의 서열, FRT-Ub-Neo-FRT 선택 카세트 및 15.5 kb의 5' 마우스 TCR A 상동성 아암을 포함하였다.

[0392] 하이그로마이신 내성 유전자 내의 표적 서열을 인식하고 절단하여 표적 TCR A 유전자좌에서 3개의 LTVEC의 상동 재조합을 촉진시키는, 징크 핑거 뉴클레아제 (ZFN)를 설계하였다. 도 4에서 생성된 ES 세포 (MAID 6548, 모든 인간 J_{κ} 세그먼트 및 4개의 기능적 인간 V_{κ} 유전자 세그먼트에 대해 이형접합성임)를 상기 기술된 3개의 큰 표적화 벡터 (MAID6600-트리밍됨, MAID1700-트리밍됨 및 MAID6647) 및 ZFN (1/2)의 각 절단을 발현하는 2개의 플라스미드 - 이는 하이그로마이신 내성 유전자의 인식 서열에 결합하여 표적 부위 (TGCGATCGTCTGCGGCCGAtcttagCCAGACGAGCGGGTTCGG (서열 번호 2); 절단 부위는 소문자로 나타나 있음) (표 1 참조)에서 이중 가닥 절단을 촉매함 - 와 함께 전기천공하였다.

[0393] 3개의 큰 표적화 벡터를 상동 재조합에 의해 DNA 서열 내로 삽입하여, Hyg 선택 카세트를 함유하고 이를 둘러싸는 영역을 대체하였다. 생성된 ES 세포는 내인성 TCR A 유전자좌에 인간 $J_{\kappa}1$ 내지 $J_{\kappa}5$ 및 $V_{\kappa}4-1$ 내지 $V_{\kappa}2-40$ 유전자 세그먼트를 포함하는 인간 면역글로불린 가변 도메인을 함유하였다. 도 4에 나타내고 상기 표 2 및 하기 표 6에 열거된 프로브와 프라이머를 사용하는 상기 기술된 TAQMAN® 검정 (본 명세서에 참고로 포함되는 문헌[Lie and Petropoulos, 1998. *Curr. Opin. Biotechnology* 9:43-48])을 사용하여 3개의 큰 표적화 벡터의 성공적인 혼입을 확인하였다 (GOA = 대립유전자 획득; LOA = 대립유전자 상실; 카피 수 = 서열의 카피 수를 확인하여 유전자도입식 통합 대 표적화된 통합을 추적함; hArm1 = 제1의 큰 표적화 벡터 (MAID 1710)의 30 kb의 3' 상동성 아암; hArm2 = 제1의 큰 표적화 벡터 (MAID 1710)와 제2의 큰 표적화 벡터 (MAID 6600)의 20 kb의 중복부, hArm3 = 제2 표적화 벡터 (MAID 6600)와 제3 표적화 벡터 (MAID6647)의 60 kb의 중복부, mArm = 제3 표적화 벡터 (MAID 6647)의 15.5 kb의 5' 상동성 아암, WT 마우스 대조군 - 마우스 TCR A 유전자좌에 존재하는 서열). 아암 특이적 검정으로 지칭되는 LTVEC의 상동성 아암의 서열을 인식하는 실시간 PCR 검정을 사용하여 마우스 게놈 내로의 LTVEC의 정확한 표적화를 검증하였다. 이러한 아암 특이적 검정의 카피 수를 결정하는 것은, 마우스 프로브 (mArm)에 대해 2의 카피 수를 그리고 인간 프로브 (hArm1)에 대해 1의 카피 수를 보유하는 정확하게 표적화된 ES 클론을, 표적 마우스 유전자좌의 Cas9 유도 결실이 게놈의 다른 곳에서의 LTVEC의 무작위적 통합과 일치하는 클론 - 이 경우 마우스 프로브 (mArm)에 대한 카피 수는 3 (또는 그 이상)이고 인간 프로브 (hArm1)에 대한 카피 수는 2 (또는 그 이상)일 것임 - 과 구별하는 것을 돕도록 추가의 설명을 제공하였다. 요망되는 유전자좌 내로의 상동 재조합에 의한 3개의 LTVEC의 정확한 조립을 검출하기 위해, 본 발명자들은 아암 특이적 TAQMAN® 검정을 이용하였다. hArm2와 hArm3에 대한 1의 예상 카피 수는 모든 3개의 LTVEC가 단일 작제물로 조립되었음을 나타내었다.

[0394]

[표 6]

TAQMAN 프라이머 및 프로브

유전자	검정	정방향 프라이머	프로브	역방향 프라이머
HYG	LOA	TGCGGCCGATCTTAG CC (서열 번호 14)	ACGAGCGGGTTCGGCC CATTTC (서열 번호 15)	TTGACCGATTTCCTTGC GG (서열 번호 16)
HYG-U	LOA	CGACGTCTGTCGAGA AGTTTCTG (서열 번호 17)	AGTTCGACAGCGTGTC CGACCTGA (서열 번호 18)	CACGCCCTCCTACATC GAA (서열 번호 19)
Hyg-D	LOA	TGTCGGGCGTACACA AATCG (서열 번호 20)	CCGTCTGGACCGATGG CTGTGT (서열 번호 21)	GGGCGTCGGTTTCCAC TATC (서열 번호 22)
hIgK6	hArm1 의 카피수	GTCAAGCACTGCTGG CACAC (서열 번호 26)	AACCCTTGCTATTGA ATTGCTATGCTGTCAG (서열 번호 27)	TGTTGTAGACCCTCCG CCAC (서열 번호 28)
hIgK12 (MAID 1710 삽입물)	GOA	TTGCTTTTCTCACACC TGCAG (서열 번호 29)	CAGCCCATCTGTCAC TTCGCTGGA (서열 번호 30)	TGGCCCAACAGTACA GCTCAG (서열 번호 31)
hIgK13	hArm2 의 카피수	TCAGTCAATCACCTTT CCCAGC (서열 번호 32)	TCCCCAGGTAGCCTCA TGAACCAATGTT (서열 번호 33)	CACATTACTGAGTCCC CACAGGG (서열 번호 34)
hIgK14	hArm2 의 카피수	CATTGTCAAAGAAGC ACTGGAAATG (서열 번호 35)	ACCATTGCAGTTTACCC ACGGTTAGGATTTTT (서열 번호 36)	TCTTGCAATGGGATCA TCAGATG (서열 번호 37)
Neo	GOA	GGTGGAGAGGCTATT CGGC (서열 번호 38)	TGGGCACAACAGACAA TCGGCTG (서열 번호 39)	GAACACGGCGGCATC AG (서열 번호 40)
hIgK15	GOA	CAGGTGCAAAGGTGA CCACAG (서열 번호 41)	102102TGGGTCTGCCC ATCCATGCA (서열 번호 42)	GGCAGCCTGAGTGTC AGAGC (서열 번호 43)
hIgK25	hArm3 의 카피수	GTTCAAGCCCCACAG ACTCTC (서열 번호 44)	TCCTCTCTGGAGCAAC CATGAAGTTCCT (서열 번호 45)	CCTGAAGCCATGAGG GCAG (서열 번호 46)
TCRA Arm4	mArm 의 카피수	GCGCCACATGAATTT GACCAG (서열 번호 50)	TGTACCCAATCTTCCAA AGAAAGAGCTG (서열 번호 51)	GGCATCCTGTCCTCCC TTC (서열 번호 52)
모 1540m1	WT 마우스 대조군	CAGTAAGGGAAGAG ACTACAACAGCAT (서열 번호 53)	TGCACACTGCTACCA CTGCAAGCTAT (서열 번호 54)	TGCTGGTGGCCCCATC T (서열 번호 55)
모 1540m3	WT 마우스 대조군	GAACTCAGCTATGAT AGTGTGCAATGTA (서열 번호 56)	CAGCCCAGCAGCTGTG GGTTCTC (서열 번호 57)	GCTCAGGGAGAACAC AGAAGTTAGA (서열 번호 58)
hIgK5	MAID 6548 서열 (표 2 참조)	CCCCGTCTCTCCTT TTTC (서열 번호 59)	TCATGTCCATTAACCCA TTACCTTTGCCCA (서열 번호 60)	TGCAAGTGCTGCCAG CAAG (서열 번호 61)
hIgK22	hArm3 의 카피수	TGGCTCCAAGAACAG TTTGCC (서열 번호 74)	CCCTGACTTGCTGCTC AACTCACAGCC (서열 번호 75)	GGTCCAGTGGAATCT GCCATG (서열 번호 76)
hIgK21	GOA	CATTGGCTACATAT CAAAGCCG (서열 번호 77)	CCTGAGCCAGGGAACA GCCACTGATA (서열 번호 78)	ACATGGCTGAGGCAG ACACC (서열 번호 79)
hIgK26	GOA	TGGGCCGTTATGCTA GTACCA (서열 번호 80)	TGGCTTTACCCCTTTG AAGGGCCC (서열 번호 81)	CACAGCTGAAGCAGG ATGAGC (서열 번호 82)
hIgK30	GOA	TCTCTGAGCAGCCAT CCCC (서열 번호 83)	TTCTCCTTTGGGTGTA GGGCACCAGC (서열 번호 84)	ACCAGGCATGGCAGA AAGG (서열 번호 85)

[0395]

[0396]

ES 세포에서 생성된 표적화된 유전자좌는 하기의 집합부 서열을 가졌으며, 여기서 마우스 서열은 괄호 안에 표기되고, 인간 서열은 보통 글꼴로 표기되며, 다중 클로닝 부위는 볼드체로 표기되고, Frt 서열은 이탤릭체로 표기되어 있다 (표 7).

[0397] [표 7]

삼중 ES 세포 표적화로부터 생성된 유전자좌의 접합부 서열

접합부	서열 번호	서열
마우스 Tcra/5' Frt	62	(GTCTTTTTTGTCTTCACAGTTGAGCTTCA TCAAAGTCACATGGGTTAAACTCTATGGAG TAGTCAGAACACTCTTCA)GAAGGGACTC CTGATTTCAAAGGGTACCGAAGTTCCTATT CCGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTC
3' Frt/인간 IgK	86	GAAGTTCCTATTCCGAAGTTCCTATTCTCTAG AAAGTATAGGAACTTCCTAGGGTTTACCGGTG GCGCGCCTGAGTAGTGCTTTAGGTGTGTAATCA CCAAAGATTTAGTGAAGTCCCTGTGCAAGGAG
인간 IgK/마우스 Tcra	64	GATAAATTATTTGTCAGACAACAATAA AAATCAATAGCACGCCCTAAGAGCGGCC GCCACCGCGGTGGAGCTC(AGGTTTCCG GTACTTAACAACAGAGCACAGATTTAGT GGTGAGGGACTCT)

[0398]

[0399] 단리된 ES 세포 콜로니의 대립유전자 변형 (MOA) 스크리닝은 0.4%의 표적화 효율을 가져왔다 (표 8 참조).

[0400] [표 8]

3 개의 LTVEC 를 사용한 표적화에 대한 대립유전자 변형 (MOA) 스크리닝 결과

LTVEC 수	뉴클레아제	결실	삽입물	효율
3 개의 LTVEC	ZFN	hyg	370 kb 인간	0.4%
3 개의 LTVEC	gRNA#16/Cas9	hyg	370 kb 인간	0.4%
3 개의 LTVEC	없음	hyg	370 kb 인간	0%

[0401]

[0402] 상기 기술된 표적화된 ES 세포를 공여체 ES 세포로 사용하여 VELOCIMOUSE® 방법에 의해 상실기 전시기의 배아, 예를 들어 8 세포기의 마우스 배아 내로 도입시켰다 (예를 들어, 미국 특허 제7,576,259호, 미국 특허 제 7,659,442호, 미국 특허 제7,294,754호 및 미국 특허 출원 공개 제2008-0078000 A1호 참조). 공여체 ES 세포를 포함하는 마우스 배아를 배반포기까지 인큐베이션하고, 이어서 대리모에 이식하여, 공여체 ES 세포로부터 완전히 유래된 F0 마우스를 생산한다. 독립적으로 키메라 인간 IgK V-마우스 Tcra C 유전자를 지니는 공여체 ES 세포로부터 완전히 유래된 F0 마우스를, 독특한 유전자 서열의 존재를 검출하는 대립유전자 변형 검정을 사용하여 유전자형 결정에 의해 확인한다.

[0403] 실시예 4: CRISPR/Cas 시스템과 조합하여 3개의 LTVEC를 사용한 하이그로마이신 유전자의 표적화

[0404] 징크 핑거 뉴클레아제를 이용하는 실시예 3에 기술된 삼중 표적화 방법을 또한 CRISPR/Cas9 시스템으로 수행하였다.

[0405] 하이그로마이신 내성 유전자 내의 다양한 표적 서열 (CRISPR 인식 서열)을 인식하도록 다양한 가이드 RNA (gRNA)를 설계하였다. 하이그로마이신 유전자 내의 CRISPR 인식 서열은 하기와 같다: gRNA#1: ACGAGCGGGTTCGGCCATTCGG (서열 번호 70); gRNA#6: CTTAGCCAGACGAGCGGGTTCGG (서열 번호 71); gRNA#10: GCCGATCTTAGCCAGACGAGCGG (서열 번호 72); 및 gRNA#16: CGACTGTATGCAGCTCTCGGAGG (서열 번호 73). 하이그로마이신 유전자 내의 인식 서열의 위치는 도 3에 도시되어 있다. gRNA#1, gRNA#6, gRNA#10 및 gRNA#16을 스크리닝하였고, 이들은 하이그로마이신 유전자를 특이적으로 표적화하는 것으로 확인되었다 (도 3 및 표 5 참조).

[0406] MAID 6548 ES 세포 (모든 인간 J_K 세그먼트 및 4개의 기능적 인간 V_K 유전자 세그먼트에 대해 이형접합성임)를, Cas9 및 gRNA#16을 인코딩하는 벡터들 - 이들 벡터는 하이그로마이신 내성 유전자 내의 표적 부위를 인식하여 절단함 - 과 더불어 실시예 3에 기술된 바와 같은 3개의 큰 표적화 벡터와 함께 전기천공하였다.

[0407] 3개의 큰 표적화 벡터를 상동 재조합에 의해 DNA 서열 내로 삽입하여, Hyg 선택 카세트들을 함유하고 이를 둘러싸는 영역을 대체하였다. 실시예 3에 기술된 TAQMAN® 검정을 사용하여 3개의 큰 표적화 벡터의 성공적인 혼입을 확인하였다.

[0408] ES 세포에서 생성된 표적화된 유전자좌는 표 7에 나타난 접합부 서열을 가졌으며, 여기서 마우스 서열은 괄호 안에 표기되고, 인간 서열은 보통 글꼴로 표기되며, 다중 클로닝 부위는 볼드체로 표기되고, Frt 서열은 이탤릭체로 표기되어 있다.

[0409] 단리된 ES 세포 콜로니의 대립유전자 변형 (MOA) 스크리닝은 0.4%의 표적화 효율을 가져왔다 (표 8 참조).

[0410] 상기 기술된 표적화된 ES 세포를 공여체 ES 세포로 사용하여 VELOCIMOUSE® 방법에 의해 상실기 전시기의 배아,

예를 들어 8 세포기의 마우스 배아 내로 도입시킬 것이다 (예를 들어, 미국 특허 제7,576,259호, 미국 특허 제 7,659,442호, 미국 특허 제7,294,754호 및 미국 특허 출원 공개 제2008-0078000 A1호 참조). 유전자 변형된 ES 세포를 포함하는 마우스 배아를 배반포기까지 인큐베이션하고, 이어서 대리모에 이식하여, 공여체 ES 세포로부터 완전히 유래된 F0 마우스를 생산한다. 공여체 ES 세포로부터 완전히 유래된 F0 마우스를, 독특한 유전자 서열의 존재를 검출하는 대립유전자 변형 검정을 사용하여 유전자형 결정에 의해 확인할 것이다.

[0411] 실시예 5: 2개의 LTVEC 사이의 중복 서열을 통한 LTVEC 표적화의 향상

[0412] 실시예 1에 기술된 이중 표적화 시스템을 사용하여, 2개의 큰 표적화 벡터 (LTVEC)를 사용하는 단일 표적화 단계로 게놈 유전자좌를 변형시켰다. 도 1에 도시된 바와 같이, 하이그로마이신 선택 카세트를 포함하는 마우스 염색체 14 상에 TCR 알파 유전자좌의 이형접합성 변형을 갖는 세포를 이중 표적화 시스템에 의해 표적화하여 추가의 Igk 가변 유전자 세그먼트를 포함하는 ES 세포를 생성시켰다. 2개의 상이한 LTVEC를 마우스 배아 줄기 (ES) 세포 내로 함께 공동 전기천공하였다. 선택적으로, 엔도뉴클레아제 (징크 핑거 뉴클레아제 (ZFN) 또는 CRISPR-Cas9)를 인코딩하는 핵산을 공동 전기천공하여 표적 유전자좌에 또는 이 부근에 이중 가닥 절단부를 생성시켰다.

[0413] 실시예 1에서와 같이, LTVEC (MAID 1710으로 표지됨)는 인간 Vκ1-5 및 Vκ1-6 유전자 세그먼트의 서열을 포함하는 30 kb의 3' 상동성 아암, 인간 Vκ3-7 내지 Vκ3-15 유전자 세그먼트를 포함하는 120 kb의 서열, 및 인간 Vκ1-16 유전자 세그먼트를 포함하는 20 kb의 5' 영역 ("중복 영역")을 포함하였다. 제2의 LTVEC (MAID 6600으로 표지됨)는 20 kb의 3' 중복 영역 (제1 벡터에서와 동일한 인간 Vκ1-16 유전자 세그먼트를 포함하는 영역), 인간 Vκ1-17 내지 Vκ2-30 유전자 세그먼트를 포함하는 140 kb의 서열, FRT-Ub-Neo-FRT 선택 카세트 및 15.5 kb의 3' 마우스 TCR A 상동성 아암을 포함하였다.

[0414] 성공적인 표적화는 상동 재조합에 의해 2개의 LTVEC를 DNA 서열 내로 삽입시켜, Hyg 선택 카세트를 포함하고 이를 둘러싸는 영역을 대체하였다. 생성된 ES 세포는 내인성 TCR A 유전자좌에 인간 Jκ1 내지 Jκ5 및 Vκ4-1 내지 Vκ2-30 유전자 세그먼트를 포함하는 인간 면역글로불린 가변 도메인을 함유하였다. 도 1 및 표 2에 나타낸 프로브 및 프라이머를 사용하는 상기 기술된 TAQMAN 검정 (본 명세서에 참고로 포함되는 문헌[Lie and Petropoulos, 1998. *Curr. Opin. Biotechnology* 9:43-48])을 사용하여 2개의 큰 표적화 벡터의 성공적인 혼입을 확인하였다.

[0415] 비교로서, 실시예 1에 기술된 단일 LTVEC 시스템을 또한 사용하여 동일한 게놈 유전자좌를 변형시켰는데, 이때 단일 LTVEC를 단독으로 사용하거나 ZFN 또는 CRISPR-Cas9와 조합하여 사용하였다 (도 2 참조). 상기 표 2에 열거되고 도 2에 나타낸 프로브 및 프라이머를 사용하여 상기 기술된 TAQMAN 검정으로 성공적인 혼입을 확인하였다.

[0416] 표 9는 단일 LTVEC를 (단독으로, ZFN과 함께, 또는 Cas9와 함께) 사용하거나, 2개의 LTVEC를 (단독으로, ZFN과 함께, 또는 Cas9와 함께) 동시에 사용하거나, 2개의 LTVEC + 제3의 LTVEC를 (단독으로, ZFN과 함께, 또는 Cas9와 함께) 동시에 사용하는 표적화 실험의 표적화 효율을 비교한 것이다. 표 9에 제시된 표적화 효율은 표 2의 TAQMAN 프라이머 및 프로브를 사용하여 최초 스크리닝, 확인 스크리닝 및 재확인 스크리닝을 통해 정확하게 표적화된 것으로 결정된, 스크리닝된 ESC 클론의 비율(%)이다. 단일 LTVEC를 단독으로 사용한 표적화는 1.1%의 정확하게 표적화된 클론을 생성시켰다. ZFN에 의한 절단은 단일 LTVEC의 표적화 효율을 4.4%로 증가시켰고, CRISPR-Cas9에 의한 절단은 단일 LTVEC의 표적화 효율을 5.5%로 증가시켰다. 놀랍게도, 20 kb의 중복 서열을 갖는 2개의 LTVEC를 사용한 표적화는 뉴클레아제가 사용되지 않았을 때에도 1.4%의 표적화 효율을 가져왔다. 표적화 효율은 ZFN이 사용되었을 때 2.81%로 그리고 Cas9가 사용되었을 때 1.6%로 증가하였다.

[0417] [표 9]

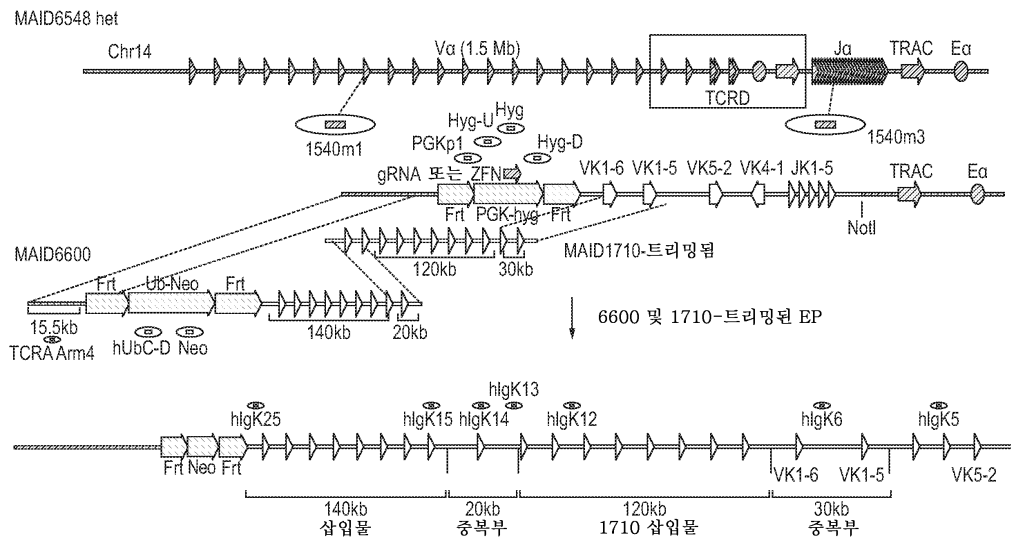
공동 전기천공된 타일링된 LTVEC의 표적화 효율

	MAID 6598: 제 2 삽입			MAID 6600: 제 2 삽입 및 제 3 삽입			MAID 6647: 제 2 삽입, 제 3 삽입 및 제 4 삽입		
	LTVEC 단독	LTVEC + ZFN	LTVEC + Cas9	2 개의 LTVEC 단독	2 개의 LTVEC + ZFN	2 개의 LTVEC + Cas9	3 개의 LTVEC 단독	3 개의 LTVEC + ZFN	3 개의 LTVEC + Cas9
표적화 효율	5 / 440 (1.1%)	17 / 384 (4.4%)	21 / 384 (5.5%)	13 / 960 (1.4%)	27 / 960 (2.8%)	15 / 960 (1.6%)	0 / 960 (0%)	4 / 960 (0.4%)	4 / 960 (0.4%)

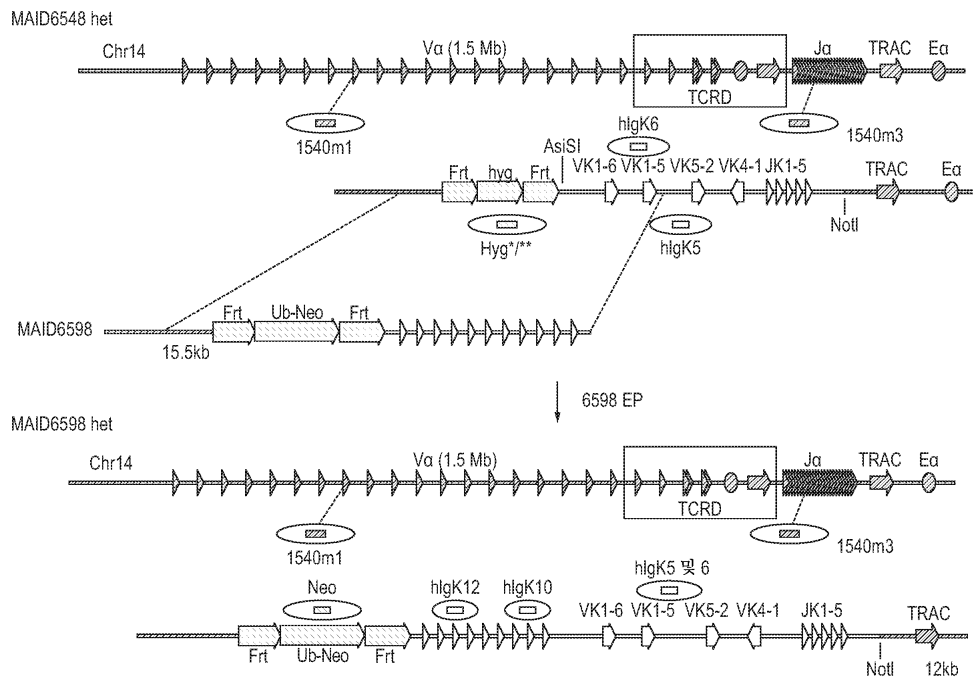
[0418]

도면

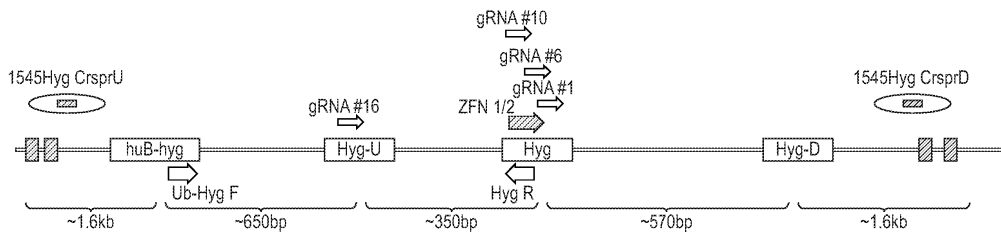
도면1



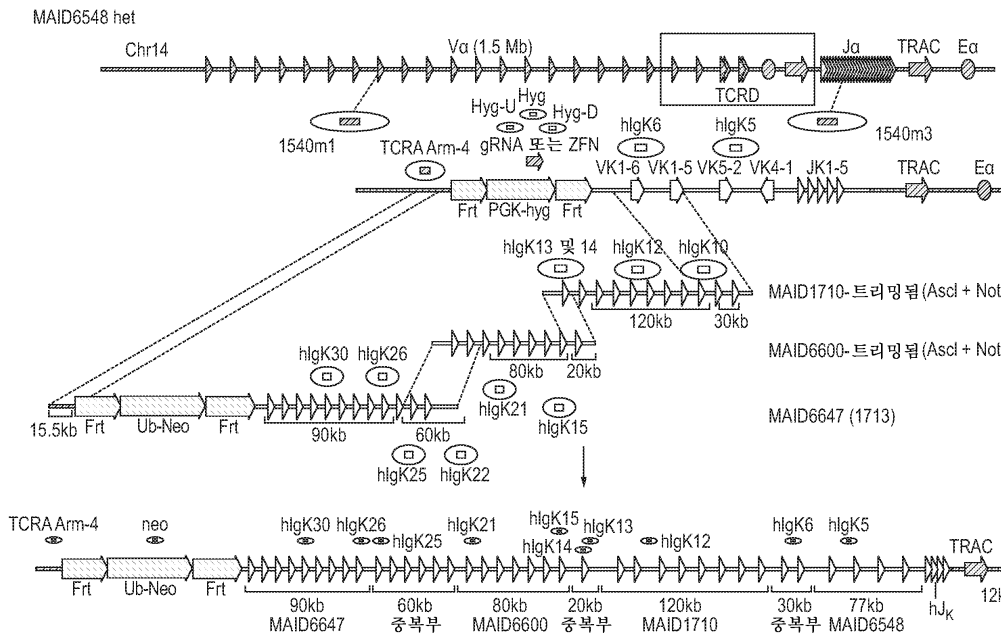
도면2



도면3



도면4



서열 목록

SEQUENCE LISTING

<110> Voronina, Vera

Macdonald, Lynn

Zambrowicz, Brian

Murphy, Andrew J.

<120> Methods and Compositions for Targeted

Genetic Modification Through Single-Step Multiple Targeting

<130> 57766-472224

<150> US 62/094,104

<151> 2014-12-19

<150> US 62/167,408

<151> 2015-05-28

<150> US 62/205,524

<151> 2015-08-14

<160> 86

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 23

<212>

DNA

<213> Artificial Sequence

<220>

<223> a genomic target sequence that is linked to a
guide RNA (gRNA)

<220>

<221> misc_feature

<222> (2)...(21)

<223> n = A,T,C or G

<400> 1

gnnnnnnnnn nnnnnnnnnn ngg 23

<210> 2

<211> 42

<212> DNA

<213> Artificial Sequence

<220>

<223> ZFN1/2 recognition site

<400> 2

tgcgatcgc tgcggccgac ttagccagac gagcgggttc gg 42

<210> 3

<211> 36

<212> DNA

<213> Artificial Sequence

<220>

<223> ZFN3/4 recognition site

<400> 3

cgctgcggcc gatcttagcc agacgagcgg gttcgg 36

<210> 4

<211> 39
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> ZFN5/6 recognition site
 <400> 4
 agcgtgtccg acctgatgca gctctcggag ggccaagaa 39
 <210> 5
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> ZFN1 Binding sequence
 <400> 5
 tgcgatcgct gcggccga 18
 <210> 6
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> ZFN2 Binding sequence
 <400> 6
 ccgaaccgc tegtctgg 18
 <210> 7
 <211> 15
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> ZFN3 Binding sequence
 <400> 7
 cgctgaggcc gatct 15
 <210> 8
 <211> 15
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> ZFN4 Binding sequence
 <400> 8
 ccgaaccgc tcgtc 15
 <210>
 > 9
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> ZFN5 Binding sequence
 <400> 9
 agcgtgtccg acctgatg 18
 <210> 10
 <211> 15
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> ZFN6 Binding sequence
 <400> 10
 ttcttcgcc tccga 15
 <210> 11
 <211> 6
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> ZFN1/2 cleavage site
 <400> 11
 tcttag 6
 <210> 12
 <211> 6
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> ZFN3/4cleavage site
 <400> 12
 tagcca 6
 <210> 13
 <211> 6
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> ZFN5/6 cleavage site
 <400> 13
 cagctc 6
 <210> 14
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> HYG forward primer
 <400> 14
 tgcggccgat cttagcc 17

 <210> 15
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> HYG probe
 <400> 15
 acgagcgggt tcggccatt c 21
 <210> 16
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> HYG reverse primer

<400> 16	
ttgaccgatt ccttgcgg	18
<210> 17	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> HYG-U forward primer	
<400> 17	
cgacgtctgt cgagaagttt ctg	23
<210> 18	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> HYG-U probe	
<400> 18	
agttcgacag cgtgtccgac ctga	24
<210> 19	
<211> 19	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> HYG-U reverse primer	
<400> 19	
cagccctcc tacatcgaa	19
<210> 20	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> HYG-D forward primer	
<400> 20	

tgtcgggcgt acacaaatcg	20
<210> 21	
<211> 22	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> HYG-D probe	
<400> 21	
ccgtctggac cgatggctgt gt	22
<210> 22	
<211> 20	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> HYG-D reverse primer	
<400> 22	
gggcgtcgggt ttccactatc	20
<210> 23	
<211> 24	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> PGKp1 (Hyg Promoter) forward primer	
<400> 23	
caaatggaag tagcacgtct cact	24
<210> 24	
<211> 23	
<212> DNA	
<213> Artificial Sequence	
<220>	
<223> PGKp1 (Hyg Promoter) probe	
<400> 24	
ctcgtgcaga tggacagcac cgc	23

<210> 25
 <211> 15
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> PGKp1 (Hyg Promoter) reverse primer
 <400> 25
 ccgctgcccc aaagg 15
 <210> 26
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgk6 forward primer
 <400> 26
 gtcaagcact gctggcacac 20

 <210> 27
 <211> 33
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgk6 probe
 <400> 27
 aacccttgtg ctattgaatt gctatgctgt cag 33
 <210> 28
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgk6 reverse primer
 <400> 28
 tgtttagac cctccgccac 20
 <210> 29
 <211> 21

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK12 (MAID 1710 insert) forward primer
 <400> 29
 ttgcctttct cacacctgca g 21

<210> 30
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK12 (MAID 1710 insert) probe
 <400> 30
 cagcccatcc tgcacttcg ctgga 25

<210> 31
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK12 (MAID 1710 insert) reverse primer
 <400> 31
 tggccaaca gtacagctca g 21

<210> 32
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK13 forward primer
 <400> 32
 tcagtcaatc acctttccca gc 22

<210> 33
 <211> 28
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> hIgK13 probe
 <400> 33
 tccccaggta gcctcatgaa ccaatgtt 28
 <210> 34
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK13 reverse primer
 <400> 34
 cacattactg agtccccaca ggg 23
 <210> 35
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK14 forward primer
 <400> 35
 cattgtcaaa gaagcactgg aaatg 25

 <210> 36
 <211> 32
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK14 probe
 <400> 36
 accattgcag tttaccacag gtaggattt tt 32
 <210> 37
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> hIgK14 reverse primer
 <400> 37
 tcttgcaatg ggatcatcag atg 23
 <210> 38
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Neo forward primer
 <400> 38
 ggtggagagg ctattcggc 19

 <210> 39
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Neo probe
 <400> 39
 tgggcacaac agacaatcgg ctg 23
 <210> 40
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Neo reverse primer
 <400> 40
 gaacacggcg gcatcag 17
 <210> 41
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK15 forward primer
 <400> 41

caggtgcaaa ggtgaccaca g 21

<210> 42
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK15 probe
 <400> 42

tgggtcctgc ccatccatgc a 21

<210> 43
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK15 reverse primer
 <400> 43

ggcagcctga gtgtcagagc 20

<210> 44
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK25 forward primer
 <400> 44

gttcaggccc cacagactct c 21

<210> 45
 <211> 29
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK25 probe
 <400> 45

tcctctctgg agcaaccatg aagttccct 29

<210> 46
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgk25 reverse primer
 <400> 46
 cctgaagcca tgaggcag 19
 <210> 47
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hUbC-D (Neo Promoter) forward primer
 <400> 47
 agggtaggct ctctgaatc g 21
 <210> 48
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hUbC-D (Neo Promoter) probe
 <400> 48
 acaggcgccg gacctctggt 20
 <210> 49
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hUbC-D (Neo Promoter) reverse primer
 <400> 49
 ccaaagaaac tgagcctca c 21
 <210> 50
 <211> 21

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> TCRA Arm4 forward primer
 <400> 50
 gcgccacatg aattgacca g 21

<210> 51
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> TCRA Arm4 probe
 <400> 51
 tgtacceaat cttcaaaga aagagctg 28

<210> 52
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> TCRA Arm4 reverse primer
 <400> 52
 ggcatcctgt cctcccttc 19

<210> 53
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Parental 1540ml forward primer
 <400> 53
 cagtaaggga agagactaca acagcat 27

<210> 54
 <211> 27
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> Parental 1540m1 probe
 <400> 54
 tgcacactgc tcaccactgc aagctat 27
 <210> 55
 <211> 17
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Parental 1540m1 reverse primer
 <400> 55
 tgctggtggc cccatct 17
 <210> 56
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Parental 1540m3 forward primer
 <400> 56
 gaactcagct atgatagtgt cgaatgta 28

 <210> 57
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Parental 1540m3 probe
 <400> 57
 cagcccagca gctgtgggtt ctc 23
 <210> 58
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> Parental 1540m3 reverse primer
 <400> 58
 gctcagggag aacacagaac ttaga 25
 <210> 59
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK5 forward primer
 <400> 59
 ccccgctctc ctctttttc 20

 <210> 60
 <211> 32
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK5 probe
 <400> 60
 tcatgtccat taaccattt accttttgcc ca 32
 <210> 61
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK5 reverse primer
 <400> 61
 tgcaagtgct gccagcaag 19
 <210> 62
 <211> 156
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> mouse Tcra/5' Frt junction sequence
 <400> 62

gcttttttg ttcttcacag ttgagcttca tcaaagtcac atgggttaaa ctctatggag 60

tagtcagaac acactcttca gaagggactc ctgatttcaa aggttaccga agttcctatt 120

ccgaagttcc tattctctag aaagtatagg aacttc 156

<210> 63

<211> 132

<212> DNA

<213> Artificial Sequence

<220>

<223> 3' Frt/human IgK junction sequence

<400> 63

gaagtctcta ttccgaagtt cctattctct agaaagtata ggaacttcct agggtttcac 60

cggtggcgcg cctaacagag aggaaagtca aattataaag aatagagat tcagaattct 120

gattaactgt gg 132

<210> 64

<211> 124

<212>

DNA

<213> Artificial Sequence

<220>

<223> Human IgK/mouse Tcra junction sequence

<400> 64

gataaattat tttgtcagac aacaataaaa atcaatagca cgcctaaga gcggccgcca 60

ccgcggtgga gctcaggttt ccggtactta acaacagagc acagatttag tggtagagga 120

ctct 124

<210> 65

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> hIgK10 forward primer

<400> 65

cgattatgac tggtaggta gaaaggtg 28

<210> 66

<211>

> 32

<212> DNA

<213> Artificial Sequence

<220>

<223> hIgK10 probe

<400> 66

gccactgggtt tctccaaatg ttttcaatcc at 32

<210> 67

<211> 24

<212> DNA

<213> Artificial Sequence

<220>

<223> hIgK10 reverse primer

<400> 67

gggagtactt ggagatccct aagc 24

<210> 68

<211> 136

<212> DNA

<213> Artificial Sequence

<220>

<223> mouse Tera/5' Frt junction sequence

<400> 68

ttgagcttca tcaaagtcac atgggttaaa ctctatggag tagtcagaac aactcttca 60

gaagggactc ctgatttcaa agggtagcga agttcctatt ccgaagttcc tattctctag 120

aaagtatagg aacttc 136

<210> 69

<211> 132

<212> DNA

<213> Artificial Sequence

<220>

<223> 3' Frt/human IgK junction sequence

<400> 69

gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttctt agggtttcac 60

cggtggcgcg ccaggacceca ggctctgaca ctgaggetgc caatacaatt gccatgaaga 120

cagatgttga tg 132

<210> 70

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> gRNA#1 CRISPR recognition sequence

<400> 70

acgagcgggt tcggccatt cgg 23

<210> 71

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> gRNA#6 CRISPR recognition sequence

<400> 71

cttagccaga cgagcgggtt cgg 23

<210> 72

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> gRNA#10 CRISPR recognition sequence

<400> 72

gccgatctta gccagacgag cgg 23

<210> 73

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> gRNA#16 CRISPR recognition sequence

<400> 73

cgacctgatg cagctctcgg agg 23

<210> 74
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK22 forward primer
 <400> 74
 tggctccaag aacagtttgc c 21
 <210> 75
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK22 probe
 <400> 75
 ccctgacttt gctgctcaac tcacagcc 28
 <210> 76
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK22 reverse primer
 <400> 76
 ggtccagtgg aatctgcat g 21
 <210> 77
 <211> 23
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK21 forward primer
 <400> 77
 catttgcta catatcaaag ccg 23
 <210> 78
 <211> 27

<212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK21 probe
 <400> 78
 cctgagccag ggaacagccc actgata 27

<210> 79
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK21 reverse primer
 <400> 79
 acatggctga ggcagacacc 20

<210> 80
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK26 forward primer
 <400> 80
 tgggccgtta tgctagtacc a 21

<210> 81
 <211> 25
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK26 probe
 <400> 81
 tggctttacc ccttttgaag ggccc 25

<210> 82
 <211> 21
 <212> DNA

<213> Artificial Sequence
 <220>
 <223> hIgK26 reverse primer
 <400> 82
 cacagctgaa gcaggatgag c 21
 <210> 83
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK30 forward primer
 <400> 83
 tctctgagca gccatcccc 19
 <210> 84
 <211> 27
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK30 probe
 <400> 84
 ttctcctttg gtgtagaggg caccagc 27
 <210> 85
 <211> 19
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> hIgK30 reverse primer
 <400> 85
 accagcatg gcagaaagg 19
 <210> 86
 <211> 130
 <212> DNA
 <213> Artificial Sequence
 <220>

<223> 3' Frt/human IgK junction

<400> 86

gaagttccta ttccgaagtt cctattctct agaaagtata ggaacttcct agggtttcac 60

cggtggcgcg ccigagtagt gctttaggtg tgtaatcacc aaagatttag tgaagtcct 120

gtgcaaggag 130