



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0081988  
(43) 공개일자 2022년06월16일

- (51) 국제특허분류(Int. Cl.)  
C12N 9/12 (2006.01) C12N 15/10 (2017.01)  
C12N 15/63 (2006.01)
- (52) CPC특허분류  
C12N 9/1276 (2013.01)  
C12N 15/102 (2013.01)
- (21) 출원번호 10-2022-7012030
- (22) 출원일자(국제) 2020년09월11일  
심사청구일자 2022년05월12일
- (85) 번역문제출일자 2022년04월11일
- (86) 국제출원번호 PCT/US2020/050323
- (87) 국제공개번호 WO 2021/050822  
국제공개일자 2021년03월18일
- (30) 우선권주장  
62/899,625 2019년09월12일 미국(US)

- (71) 출원인  
더 제이. 데이비드 글래드스톤 인스티튜트, 어 테  
스터멘터리 트러스트 이스타빌리쉬드 언더 더 윌  
오브 제이. 데이비드 글래드스톤  
미국 캘리포니아 94158 샌프란시스코 오웬스 스트  
리트 1650  
프레지던트 앤드 펠로우즈 오브 하바드 칼리지  
미합중국, 메사추세츠 02138, 캠브리지, 퀸시스트  
리트17
- (72) 발명자  
쉬프먼, 세트  
미국 94158 캘리포니아주 샌프란시스코 오웬스 스트  
리트 1650
- (74) 대리인  
양영준, 이상남

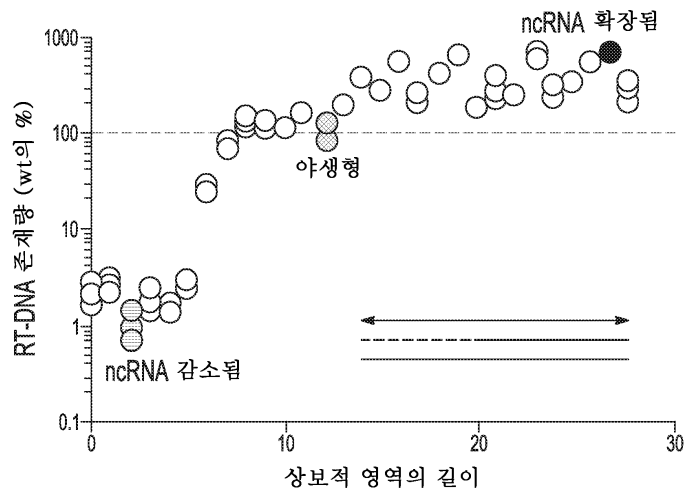
전체 청구항 수 : 총 66 항

(54) 발명의 명칭 DNA 생산을 증진시키는 변형된 박테리아 레트로요소

(57) 요약

멀티카피 단일 가닥 DNA (msDNA)의 생산을 증진시키기 위해 변형된 조작된 레트론이 제공된다. 또한, 이러한 조작된 레트론을 코딩하는 벡터 시스템; 및 조작된 레트론 및 이를 코딩하는 벡터 시스템을 다양한 적용, 예컨대 CRISPR/Cas-매개 게놈 편집, 재조합공학, 세포성 바코딩 및 분자 기록에 사용하는 방법이 또한 개시된다.

대표도 - 도7cb



(52) CPC특허분류

*C12N 15/63* (2013.01)

*C12Y 207/07049* (2013.01)

*C12N 2310/20* (2017.05)

---

## 명세서

### 청구범위

#### 청구항 1

하기를 포함하는 조작된 레트론:

- a) 프리-msr 서열;
- b) 멀티카피 단일 가닥 RNA (msRNA)를 코딩하는 *msr* 유전자;
- c) 멀티카피 단일 가닥 DNA (msDNA)를 코딩하는 *msd* 유전자;
- d) 프리-msr 서열에 대한 서열 상보성을 갖는 자기 상보적 영역을 포함하는 포스트-msd 서열, 여기서 자기 상보적 영역은 조작된 레트론이 msDNA의 생산을 증진시킬 수 있도록 야생형 상보적 영역보다 1 내지 50개 뉴클레오티드 더 긴 길이를 가짐; 및
- e) 역전사효소를 코딩하는 *ret* 유전자.

#### 청구항 2

제1항에 있어서, 관심 이중 서열을 추가로 포함하는 조작된 레트론.

#### 청구항 3

제2항에 있어서, 이중 서열이 *msr* 유전자 또는 *msd* 유전자 내로 삽입되는 것인 조작된 레트론.

#### 청구항 4

제1항에 있어서, *msd* 유전자에 의해 코딩된 단일 가닥 DNA (msDNA)가 *msd* 스템 루프를 포함하며, 여기서 루프가 관심 이중 서열을 포함하는 것인 조작된 레트론.

#### 청구항 5

제2항에 있어서, 이중 서열이 상동성 지정 복구 (HDR) 또는 재조합공학에 의해 표적 유전자좌에서 통합되도록 의도된 편집을 포함하는 뉴클레오티드 서열에 플랭킹하는 3' 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 표적 서열과 혼성화하는 5' 상동성 아암을 포함하는 공여자 폴리뉴클레오티드를 코딩하는 것인 조작된 레트론.

#### 청구항 6

제2항에 있어서, 이중 서열이 CRISPR 프로토스페이서 DNA 서열을 포함하는 것인 조작된 레트론.

#### 청구항 7

제6항에 있어서, CRISPR 프로토스페이서 DNA 서열이 변형된 AAG 프로토스페이서 인접 모티프 (PAM)를 포함하는 것인 조작된 레트론.

#### 청구항 8

제1항에 있어서, 바코드 서열을 추가로 포함하는 조작된 레트론.

#### 청구항 9

제8항에 있어서, 바코드 서열이 msDNA의 헤어핀 루프에 위치하는 것인 조작된 레트론.

#### 청구항 10

제4항에 있어서, *msd* 스템이 적어도 14개 뉴클레오티드 (염기 쌍) 길이인 조작된 레트론.

#### 청구항 11

제4항에 있어서, *msd* 스템이 하나 이상의 미스매치된 뉴클레오티드를 포함할 수 있는 것인 조작된 레트론.

**청구항 12**

제4항에 있어서, *msd* 스템이 5개 또는 6개 미만의 연속 미스매치된 뉴클레오티드를 갖는 것인 조작된 레트론.

**청구항 13**

제4항에 있어서, *msd* 스템이 *msd* 스템 염기에 플랭킹하는 서열에 임의의 미스매치 또는 삽입을 갖지 않는 것인 조작된 레트론.

**청구항 14**

제4항에 있어서, 야생형 *msd* 서열과 비교하여 임의의 *msd* 스템 미스매치 또는 서열 변경이 *msd* 스템의 중간부 이내에 있는 것인 조작된 레트론.

**청구항 15**

제1항에 있어서, *msr* 유전자와 *msd* 유전자가 트랜스 배열 또는 시스 배열로 제공되는 것인 조작된 레트론.

**청구항 16**

제1항에 있어서, *ret* 유전자가 *msr* 유전자 및 *msd* 유전자에 대하여 트랜스 배열로 제공되는 것인 조작된 레트론.

**청구항 17**

제1항에 있어서, *msr* 유전자, *msd* 유전자, 및 *ret* 유전자가 변형된 박테리아 레트론 *msr* 유전자, *msd* 유전자, 및 *ret* 유전자인 조작된 레트론.

**청구항 18**

제1항에 있어서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자가 독립적으로 변형된 미소박테리아 레트론, 변형된 에스케리키아 콜라이(*Escherichia coli*) 레트론, 변형된 살모넬라 엔테리카(*Salmonella enterica*) 레트론, 또는 변형된 비브리오 콜레라에(*Vibrio cholerae*) 레트론인 조작된 레트론.

**청구항 19**

제18항에 있어서, 변형된 에스케리키아 콜라이 레트론이 변형된 EC83 또는 변형된 EC86인 조작된 레트론.

**청구항 20**

제1항의 조작된 레트론을 포함하는 하나 이상의 벡터를 포함하는 벡터 시스템.

**청구항 21**

제20항에 있어서, *msr* 유전자와 *msd* 유전자가 동일한 벡터 또는 상이한 벡터에 의해 제공되는 것인 벡터 시스템.

**청구항 22**

제20항에 있어서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자가 동일한 벡터에 의해 제공되는 것인 벡터 시스템.

**청구항 23**

제22항에 있어서, 동일한 벡터가 *msr* 유전자 및 *msd* 유전자와 작동 가능하게 연결된 프로모터를 포함하는 것인 벡터 시스템.

**청구항 24**

제23항에 있어서, 프로모터가 *ret* 유전자와 추가로 작동 가능하게 연결되는 것인 벡터 시스템.

**청구항 25**

제23항에 있어서, *ret* 유전자와 작동 가능하게 연결된 제2 프로모터를 추가로 포함하는 벡터 시스템.

**청구항 26**

제20항에 있어서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자가 상이한 벡터에 의해 제공되는 것인 벡터 시스템.

**청구항 27**

제20항에 있어서, 하나 이상의 벡터가 바이러스 벡터 또는 비-바이러스 벡터인 벡터 시스템.

**청구항 28**

제27항에 있어서, 비-바이러스 벡터가 플라스미드인 벡터 시스템.

**청구항 29**

제20항에 있어서, 조작된 레트론이 상동성 지정 복구 (HDR) 또는 재조합공학에 의해 표적 유전자좌에서 통합되도록 의도된 편집을 포함하는 뉴클레오티드 서열에 플랭킹하는 3' 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 표적 서열과 혼성화하는 5' 상동성 아암을 포함하는 공여자 폴리뉴클레오티드를 포함하는 것인 벡터 시스템.

**청구항 30**

제29항에 있어서, RNA-가이드된 뉴클레아제를 코딩하는 벡터를 추가로 포함하는 벡터 시스템.

**청구항 31**

제30항에 있어서, RNA-가이드된 뉴클레아제가 Cas 뉴클레아제 또는 조작된 RNA-가이드된 FokI-뉴클레아제인 벡터 시스템.

**청구항 32**

제31항에 있어서, Cas 뉴클레아제가 Cas9 또는 Cpf1인 벡터 시스템.

**청구항 33**

제20항에 있어서, 조작된 레트론이 CRISPR 프로토스페이서 DNA 서열을 포함하는 것인 벡터 시스템.

**청구항 34**

제33항에 있어서, Cas1 및/또는 Cas2 단백질을 코딩하는 벡터를 추가로 포함하는 벡터 시스템.

**청구항 35**

제34항에 있어서, CRISPR 어레이 서열을 포함하는 벡터를 추가로 포함하는 벡터 시스템.

**청구항 36**

제20항에 있어서, 박테리오파지 상동 재조합 단백질을 코딩하는 벡터를 추가로 포함하는 벡터 시스템.

**청구항 37**

제36항에 있어서, 박테리오파지 상동 재조합 단백질을 코딩하는 벡터가 *exo*, *bet* 및 *gam* 유전자를 포함하는 복제 결함 있는  $\lambda$  프로파지인 벡터 시스템.

**청구항 38**

제1항의 조작된 레트론 또는 제20항의 벡터 시스템을 포함하는 단리된 숙주 세포.

**청구항 39**

제38항에 있어서, 원핵, 아키온, 또는 진핵 숙주 세포인 숙주 세포.

**청구항 40**

제39항에 있어서, 진핵 숙주 세포가 포유류 숙주 세포인 숙주 세포.

**청구항 41**

제40항에 있어서, 포유류 숙주 세포가 인간 숙주 세포인 숙주 세포.

**청구항 42**

제38항에 있어서, 인공 세포 또는 유전자 변형된 세포인 숙주 세포.

**청구항 43**

제1항의 조작된 레트론, 제20항의 벡터 시스템 또는 제38항의 숙주 세포를 포함하는 키트.

**청구항 44**

제43항에 있어서, 조작된 레트론을 사용하여 세포를 유전자 변형시키기 위한 지침을 추가로 포함하는 키트.

**청구항 45**

a) 제5항의 조작된 레트론으로 세포를 형질감염시키는 단계;

b) RNA-가이드된 뉴클레아제 및 가이드 RNA를 세포 내로 도입하여 유전자 변형된 세포를 생산하는 단계이며, 여기서 RNA-가이드된 뉴클레아제는 가이드 RNA와 복합체를 형성하고, 상기 가이드 RNA는 복합체가 게놈 표적 유전자좌를 향하도록 하며, RNA-가이드된 뉴클레아제는 게놈 표적 유전자좌에서 게놈 DNA 내에 이중 가닥 파손을 생성하고, 조작된 레트론에 의해 생성된 공여자 폴리뉴클레오티드는 상동성 지정 복구 (HDR)에 의해 그의 5' 상동성 아암과 3' 상동성 아암에 의해 인식된 게놈 표적 유전자좌에 통합되는 것인 단계

를 포함하는, 세포를 유전자 변형시키는 방법.

**청구항 46**

제45항에 있어서, RNA-가이드된 뉴클레아제가 Cas 뉴클레아제 또는 조작된 RNA-가이드된 FokI-뉴클레아제인 방법.

**청구항 47**

제45항에 있어서, Cas 뉴클레아제가 Cas9 또는 Cpf1인 방법.

**청구항 48**

제45항에 있어서, RNA-가이드된 뉴클레아제가 세포의 게놈에 통합된 벡터 또는 재조합 폴리뉴클레오티드에 의해 제공되는 것인 방법.

**청구항 49**

제45항에 있어서, 조작된 레트론이 벡터에 의해 제공되는 것인 방법.

**청구항 50**

제45항에 있어서, 공여자 폴리뉴클레오티드가 유전자 대체, 유전자 녹아웃, 결실, 삽입, 반전 또는 점 돌연변이를 생성하는 데 사용되는 것인 방법.

**청구항 51**

a) 제5항의 조작된 레트론으로 세포를 형질감염시키는 단계; 및

b) 박테리오파지 재조합 단백질을 세포 내로 도입하여 유전자 변형된 세포를 생산하는 단계이며, 여기서 박테리오파지 재조합 단백질은 조작된 레트론에 의해 생성된 공여자 폴리뉴클레오티드가 그의 5' 상동성 아암과 3' 상동성 아암에 의해 인식된 표적 유전자좌에 통합되도록 표적 유전자좌에서 상동 재조합을 매개하는 것인 단계

를 포함하는, 재조합공학에 의해 세포를 유전자 변형시키는 방법.

**청구항 52**

제51항에 있어서, 공여자 폴리뉴클레오티드가 재조합공학에 의해 박테리아 세포 내의 플라스미드, 박테리아 인공 염색체 (BAC) 또는 박테리아 염색체를 변형하는 데 사용되는 것인 방법.

**청구항 53**

제51항에 있어서, 공여자 폴리뉴클레오티드가 유전자 대체, 유전자 녹아웃, 결실, 삽입, 반전 또는 점 돌연변이를 생성할 수 있는 것인 방법.

**청구항 54**

제51항에 있어서, 상기 박테리오파지 재조합 단백질을 세포 내로 도입하는 것이 복제 결함 있는  $\lambda$  프로파지를 박테리아 게놈 내로 삽입하는 것을 포함하는 것인 방법.

**청구항 55**

제54항에 있어서, 박테리오파지가 *exo*, *bet*, 및 *gam* 유전자를 포함하는 것인 방법.

**청구항 56**

제8항의 조작된 레트론으로 세포를 형질감염시키는 것을 포함하는, 세포를 바코딩하는 방법.

**청구항 57**

- a) CRISPR 적응 시스템의 Cas1 단백질 또는 Cas2 단백질을 숙주 세포 내로 도입하는 단계;
- b) 리더 서열과 적어도 하나의 반복 서열을 포함하는 CRISPR 어레이 핵산 서열을 숙주 세포 내로 도입하는 단계이며, 여기서 CRISPR 어레이 핵산 서열은 숙주 세포 내의 게놈 DNA 또는 벡터에 통합되는 것인 단계; 및
- c) 제10항 또는 제11항에 따른 복수 개의 조작된 레트론을 숙주 세포 내로 도입하는 단계이며, 여기서 각각의 레트론은 프로세싱되어 CRISPR 어레이 핵산 서열 내로 삽입될 수 있는 상이한 프로토스페이서 DNA 서열을 포함하는 것인 단계

를 포함하는, 생체내 분자 기록 시스템을 생산하는 방법.

**청구항 58**

제57항에 있어서, Cas1 단백질 또는 Cas2 단백질이 벡터에 의해 제공되는 것인 방법.

**청구항 59**

제57항에 있어서, 조작된 레트론이 벡터에 의해 제공되는 것인 방법.

**청구항 60**

제57항에 있어서, 복수 개의 조작된 레트론이 적어도 3개의 상이한 프로토스페이서 DNA 서열을 포함하는 것인 방법.

**청구항 61**

하기를 포함하는 생체내 분자 기록 시스템을 포함하는 조작된 세포:

- a) CRISPR 적응 시스템의 Cas1 단백질 또는 Cas2 단백질;
- b) 리더 서열과 적어도 하나의 반복 서열을 숙주 세포에 포함하는 CRISPR 어레이 핵산 서열, 여기서 CRISPR 어레이 핵산 서열은 조작된 세포 내의 게놈 DNA 또는 벡터에 통합됨; 및
- c) 제6항에 따른 복수 개의 조작된 레트론, 여기서 각각의 레트론은 프로세싱되어 CRISPR 어레이 핵산 서열 내로 삽입될 수 있는 상이한 프로토스페이서 DNA 서열을 포함함.

**청구항 62**

제61항에 있어서, Cas1 단백질 또는 Cas2 단백질이 벡터에 의해 제공되는 것인 조작된 세포.

**청구항 63**

제61항에 있어서, 조작된 레트론이 벡터에 의해 제공되는 것인 조작된 세포.

**청구항 64**

제61항에 있어서, 복수 개의 조작된 레트론이 적어도 3개의 상이한 프로토스페이서 DNA 서열을 포함하는 것인 조작된 세포.

**청구항 65**

제61항의 조작된 세포 및 생체내 분자 기록을 위한 지침을 포함하는 키트.

**청구항 66**

- a) 제1항의 조작된 레트론 또는 제20항의 벡터 시스템으로 숙주 세포를 형질감염시키는 단계; 및
- b) msDNA가 생산되는 적합한 조건하에 숙주 세포를 배양하는 단계를 포함하는, 재조합 msDNA를 생산하는 방법.

**발명의 설명**

**기술 분야**

[0001]

**우선권**

[0002]

본 출원은 2019년 9월 12일에 출원된 미국 가출원 번호 62/899,625를 우선권 주장하며, 이 가출원의 내용은 그 전체 내용이 구체적으로 본원에 참조로 포함된다.

[0003]

**텍스트 파일로서 제공된 서열 목록이 참조로 포함됨**

[0004]

본원의 서열 목록은 2020년 9월 10일에 작성된 12,288 바이트 크기의 텍스트 파일 "2072305.txt"로서 제공된다. 텍스트 파일의 내용은 그 전체 내용이 본원에 참조로 포함된다.

**배경 기술**

[0005]

레트론은 거의 모든 미소박테리아 (문헌 [Dhundale et al. *Journal of Bacteriology* 164, 914-917 (1985)] 및 드물게 이. 콜라이(*E. coli*) (문헌 [Lampson et al. *Science* 243, 1033-1038 (1989)], 브이. 콜레라에(*V. cholerae*) (문헌 [Inouye et al. *Microbiology and Immunology* 55, 510-513] 및 기타 박테리아에서 발견되는 역전사된 요소이다. 레트론 오페론은 RNA 프라이머 (멀티카피 단일 가닥 RNA, msr), 역전사될 RNA 서열 (멀티카피 단일 가닥 DNA, msd) 및 역전사효소를 그 순서대로 코딩한다. 레트론 전사체는 스스로 폴딩되고 부분적으로 역전사되어 약 80개 염기의 단일 가닥 DNA (ssDNA)를 생성한다. 레트론 유래 DNA는 단일 가닥이지만, 이중 가닥 DNA의 헤어핀을 함유하고 있다. 다수의 레트론 ssDNA는 또한 더 큰 이중 가닥 요소를 형성하기 위해 서로를 보완할 수 있다. 레트론 변이체는 상이한 DNA 길이와 염기 함량을 갖지만, 대체로 이러한 전체적인 형식을 공유한다.

[0006]

레트론에 의해 생성된 ssDNA는 하기 2가지 상황에서 게놈 조작에 사용되었다: 박테리아의 경우, 재조합공학을 위한 λ 레드 베타 재조합효소를 사용하고 (문헌 [Farzadfard et al. *Science* 346, 1256272, (2014)]); 진핵생물의 경우, 효모에서 Cas9 편집을 위한 상동성 지정 복구 (HDR) 템플릿으로서 사용된다 (문헌 [Sharon et al. *Cell* 175, 544-557.e516, (2018)]). 엄청난 가능성에도 불구하고, 이러한 적용은 예상보다 더 낮은 효율성과 상황 제한으로 어려움을 겪었는데, 이는 아마도 레트론의 내인성 형태의 요소에서 비롯되었을 것이다. 이는 (1) ssDNA의 5' 말단과 msr RNA의 2' 히드록실을 연결하는 포스포디에스테르 결합을 가진 분기 구조, (2) 레트론 역전사에 필요할 수 있지만 복구 템플릿의 일부가 아닌 불변 플랭킹 영역, (3) 제한된 총 길이, 및 (4) Pol III 전사에 대한 종결인자로서 기능하는 천연 폴리 T 연장물을 포함한다.

**발명의 내용**

- [0007] 멀티카피 단일 가닥 DNA (msDNA)의 생산을 증진시키기 위해 변형된 조작된 레트론이 제공되어 효율성과 낮은 카피 수와 관련된 기존의 많은 문제를 해결한다. 또한 이러한 조작된 레트론을 코딩하는 벡터 시스템 및 CRISPR/Cas 매개 게놈 편집, 재조합공학, 세포성 바코딩 및 분자 기록과 같은 다양한 적용에서 조작된 레트론 및 벡터 시스템을 사용하는 방법이 본원에 기재되어 있다.
- [0008] 한 측면에서, 조작된 레트론이 제공되며, 이러한 조작된 레트론은 하기를 포함한다: a) 프리-msr 서열; b) 멀티카피 단일 가닥 RNA (msRNA)를 코딩하는 *msr* 유전자; c) 멀티카피 단일 가닥 DNA (msDNA)를 코딩하는 *msd* 유전자; d) 프리-msr 서열에 대한 서열 상보성을 갖는 자기 상보적 영역을 포함하는 포스트-*msd* 서열, 여기서 자기 상보적 영역은 조작된 레트론이 msDNA의 생산을 증진시킬 수 있도록 야생형 상보적 영역보다 적어도 1 내지 50개 뉴클레오티드 더 긴 길이를 가짐; 및 e) 역전사효소를 코딩하는 *ret* 유전자.
- [0009] 자기 상보적 영역은 ncRNA의 3' 말단과 5' 말단 사이에 수소 결합에 의해 형성된다. 특정 실시양태에서, 상보적 영역은 야생형 상보적 영역보다 적어도 1, 적어도 2, 적어도 4, 적어도 6, 적어도 8, 적어도 10, 적어도 12, 적어도 14, 적어도 16, 적어도 18, 적어도 20, 적어도 30, 적어도 40 또는 적어도 50개 뉴클레오티드 더 긴 길이를 갖는다. 예를 들어, 자기 상보적 영역은 야생형 상보적 영역보다 1 내지 50개 뉴클레오티드 범위의 더 긴 길이를 가질 수 있으며, 이는 이러한 범위 내의 임의의 길이, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 또는 50개 뉴클레오티드 더 긴 길이를 포함한다. 특정 실시양태에서, 자기 상보적 영역은 야생형 상보적 영역보다 1 내지 16개 뉴클레오티드 범위의 더 긴 길이를 갖는다.
- [0010] 특정 실시양태에서, *msr* 유전자와 *msd* 유전자는 트랜스 배열 또는 시스 배열로 제공된다. 일부 실시양태에서, *ret* 유전자는 *msr* 유전자 및/또는 *msd* 유전자에 대한 트랜스 배열로 제공된다.
- [0011] 특정 실시양태에서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자는 미소박테리아 레트론 (예를 들어, Mx65, Mx162), 에스케리키아 콜라이(*Escherichia coli*) 레트론 (예를 들어, E67, Ec73, EC83, EC86, EC107), 살모넬라 엔테리카(*Salmonella enterica*) 레트론 (예를 들어, msDNA-St85) 및 비브리오 콜레라에(*Vibrio cholerae*) 레트론 (예를 들어, Vc81, Vc95, Vc137)을 포함하나 이에 제한되지는 않는 박테리아 레트론으로부터 유래된다.
- [0012] 특정 실시양태에서, 조작된 레트론은 관심 이종 서열을 추가로 포함한다. 이종 서열은, 예를 들어 *msr* 유전자 또는 *msd* 유전자 내로 삽입될 수 있다. 예를 들어, 이종 서열은 *msd* 스템 루프의 루프 내로 삽입될 수 있다. 일부 실시양태에서, 이종 서열은 폴리펩티드 또는 펩티드를 코딩한다. 다른 실시양태에서, 이종 서열은 상동성 지정 복구 (HDR) 또는 재조합공학에 의해 게놈 표적 유전자좌에서 통합되도록 의도된 편집을 포함하는 뉴클레오티드 서열에 플랭킹하는 3' 게놈 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 게놈 표적 서열과 혼성화하는 5' 상동성 아암을 포함하는 공여자 폴리뉴클레오티드를 코딩한다. 또 다른 실시양태에서, 이종 서열은 CRISPR 프로토스페이서 DNA 서열을 포함한다. 한 실시양태에서, CRISPR 프로토스페이서 DNA 서열은 변형된 "AAG" 프로토스페이서 인접 모티프 (PAM)를 포함한다.
- [0013] 특정 실시양태에서, 조작된 레트론은 바코드 서열을 추가로 포함한다. 바코드 서열은, 예를 들어, msDNA의 헤어핀 루프에 위치할 수 있다.
- [0014] 또 다른 측면에서, 벡터 시스템이 제공되며, 이러한 벡터 시스템은 본원에 기재된 조작된 레트론을 포함하는 하나 이상의 벡터를 포함한다. 특정 실시양태에서, *msr* 유전자와 *msd* 유전자는 동일한 벡터 또는 상이한 벡터에 의해 제공된다. 일부 실시양태에서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자는 동일한 벡터에 의해 제공되며, 여기서 벡터는 *msr* 유전자 및 *msd* 유전자와 작동 가능하게 연결된 프로모터를 포함한다. 일부 실시양태에서, 프로모터는 *ret* 유전자와 추가로 작동 가능하게 연결된다. 다른 실시양태에서, 벡터는 *ret* 유전자와 작동 가능하게 연결된 제2 프로모터를 추가로 포함한다. 특정 실시양태에서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자는 상이한 벡터에 의해 제공된다.
- [0015] 특정 실시양태에서, 벡터 시스템의 벡터 중 하나 이상은 바이러스 벡터 또는 비-바이러스 벡터 (예를 들어, 플라스미드)이다.
- [0016] 특정 실시양태에서, 벡터 시스템은 공여자 폴리뉴클레오티드 서열에 플랭킹하는 3' 게놈 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 게놈 표적 서열과 혼성화하는 5' 상동성 아암을 포함하는 공여자 폴리뉴클레오티드를 코딩하는 이종 서열을 포함하는 조작된 레트론을 포함한다. 공여자 폴리뉴클레오티드 서열은, 예를 들어, 상동

성 지정 복구 (HDR) 또는 재조합공학에 의해 게놈 표적 유전자좌를 대체하거나 편집할 수 있다.

- [0017] 특정 실시양태에서, 벡터 시스템은 RNA-가이드된 뉴클레아제를 코딩하는 벡터를 추가로 포함한다. 예시적인 RNA-가이드된 뉴클레아제는 Cas 뉴클레아제 (예를 들어, Cas9, Cpf1) 및 조작된 RNA-가이드된 FokI-뉴클레아제를 포함하나 이에 제한되지는 않는다.
- [0018] 특정 실시양태에서, 벡터 시스템은 재조합공학을 위한 박테리오파지 재조합 단백질을 코딩하는 벡터를 추가로 포함한다. 일부 실시양태에서, 벡터는 박테리오파지 재조합 단백질을 코딩하는 복제 결함 있는 프로파지이다.
- [0019] 특정 실시양태에서, 벡터 시스템은 CRISPR 프로토스페이서 DNA 서열을 코딩하는 이중 서열을 포함하는 조작된 레트론을 포함한다. 일부 실시양태에서, 벡터 시스템은 Cas1 또는 Cas2 단백질을 코딩하는 벡터를 추가로 포함한다. 일부 실시양태에서, 벡터 시스템은 CRISPR 어레이 서열을 포함하는 벡터를 추가로 포함한다.
- [0020] 또 다른 측면에서, 단리된 숙주 세포가 제공되며, 이러한 숙주 세포는 본원에 기재된 조작된 레트론 또는 벡터 시스템을 포함한다.
- [0021] 특정 실시양태에서, 숙주 세포는 원핵, 아키온(archeon), 또는 진핵 숙주 세포이다. 예를 들어, 숙주 세포는 박테리아, 원생생물, 진균, 동물 또는 식물 숙주 세포일 수 있다. 일부 실시양태에서, 숙주 세포는 포유류 숙주 세포이다. 숙주 세포는 인간 또는 비인간 포유류 숙주 세포일 수 있다. 다른 실시양태에서, 숙주 세포는 인공 세포 또는 유전자 변형된 세포이다.
- [0022] 또 다른 측면에서, 본원에 기재된 조작된 레트론, 또는 이러한 조작된 레트론을 포함하는 벡터 시스템 또는 숙주 세포를 포함하는 키트가 제공된다. 일부 실시양태에서, 키트는 조작된 레트론을 사용하는 방법에 대한 지침을 추가로 포함한다.
- [0023] 또 다른 측면에서, 세포를 유전자 변형시키는 방법이 제공된다. 일부 경우에, 이러한 방법은 세포를 조작된 레트론으로 형질감염시키는 것을 포함한다. 예를 들어, 상기 방법은 a) 상동성 지정 복구 (HDR)에 의해 게놈 표적 유전자좌에 통합되도록 의도된 편집을 포함하는 뉴클레오티드 서열에 플랭킹하는 3' 게놈 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 게놈 표적 서열과 혼성화하는 5' 상동성 아암을 포함하는 공여자 폴리뉴클레오티드를 코딩하는 이중 서열을 포함하는 조작된 레트론으로 세포를 형질감염시키는 단계; 및 b) RNA-가이드된 뉴클레아제 및 가이드 RNA를 세포 내로 도입하는 단계를 포함할 수 있으며, 여기서 RNA-가이드된 뉴클레아제는 가이드 RNA와 복합체를 형성하고, 상기 가이드 RNA는 복합체가 게놈 표적 유전자좌를 향하도록 하며, RNA-가이드된 뉴클레아제는 게놈 표적 유전자좌에서 게놈 DNA 내에 이중 가닥 파손을 생성하고, 조작된 레트론에 의해 생성된 공여자 폴리뉴클레오티드는 상동성 지정 복구 (HDR)에 의해 그의 5' 상동성 아암과 3' 상동성 아암에 의해 인식되는 게놈 표적 유전자좌에 통합된다. 공여자 폴리뉴클레오티드를 코딩하는 조작된 레트론을 이용한 HDR은, 예를 들어 유전자 대체, 유전자 녹아웃, 결실, 삽입, 반전 또는 점 돌연변이를 생성하는 데 사용될 수 있다. 일부 경우에, 공여자 폴리뉴클레오티드를 코딩하는 조작된 레트론을 이용한 HDR은, 예를 들어, 유전자, 유전자 녹아웃, 결실, 삽입, 반전 또는 점 돌연변이를 복구하는 데 사용될 수 있다. 이러한 방법을 통하여 유전자 변형된 세포를 생성할 수 있다. 일부 실시양태에서, 상기 방법은 유전자 변형된 세포의 표현형을 분류하거나 또는 유전자 변형된 세포의 게놈을 시퀀싱하는 단계를 추가로 포함한다.
- [0024] 또 다른 측면에서, 재조합공학에 의해 세포를 유전자 변형시키는 방법이 제공되며, 이러한 방법은 a) 재조합공학에 의해 게놈 표적 유전자좌에 통합되도록 의도된 편집을 포함하는 뉴클레오티드 서열에 플랭킹하는 3' 게놈 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 게놈 표적 서열과 혼성화하는 5' 상동성 아암을 포함하는 공여자 폴리뉴클레오티드를 코딩하는 이중 서열을 포함하는 조작된 레트론으로 세포를 형질감염시키는 단계; 및 b) 박테리오파지 재조합 단백질을 세포 내로 도입하는 단계를 포함할 수 있으며, 여기서 박테리오파지 재조합 단백질은 조작된 레트론에 의해 생성된 공여자 폴리뉴클레오티드가 그의 5' 상동성 아암과 3' 상동성 아암에 의해 인식된 표적 유전자좌에 통합되어 유전자 변형된 세포를 생산하도록 표적 유전자좌에서 상동 재조합을 매개한다. 공여자 폴리뉴클레오티드를 코딩하는 조작된 레트론을 이용한 재조합공학은, 예를 들어, 유전자 대체, 유전자 녹아웃, 결실, 삽입, 반전 또는 점 돌연변이를 생성하는 데 사용될 수 있다. 특정 실시양태에서, 공여자 폴리뉴클레오티드는 재조합공학에 의해 박테리아 세포에서 플라스미드, 박테리아 인공 염색체 (BAC) 또는 박테리아 염색체를 변형시키는 데 사용된다. 일부 실시양태에서, 상기 방법은 유전자 변형된 세포의 표현형을 분류하거나 또는 유전자 변형된 세포의 게놈을 시퀀싱하는 단계를 추가로 포함한다.
- [0025] 특정 실시양태에서, 박테리오파지 재조합 단백질은 복제 결함 있는  $\lambda$  프로파지를 박테리아 게놈에 삽입함으로써 박테리아 세포 내로 도입된다. 한 실시양태에서, 박테리오파지는 *exo*, *bet* 및 *gam* 유전자를 포함한다.

- [0026] 또 다른 측면에서, 세포를 바코딩하는 방법이 제공되며, 이러한 방법은 본원에 기재된 바와 같이, 바코드를 포함하는 조작된 레트론으로 세포를 형질감염시키는 것을 포함한다.
- [0027] 또 다른 측면에서, 생체내 분자 기록 시스템을 생산하는 방법이 제공되며, 이러한 방법은 a) CRISPR 적응 시스템의 Cas1 단백질 또는 Cas2 단백질을 숙주 세포 내로 도입하는 단계; b) 리더 서열과 적어도 하나의 반복 서열을 포함하는 CRISPR 어레이 핵산 서열을 숙주 세포 내로 도입하는 단계이며, 여기서 CRISPR 어레이 핵산 서열은 숙주 세포 내의 게놈 DNA 또는 벡터에 통합되는 것인 단계; 및 c) CRISPR 프로토스페이서 DNA 서열을 포함하는 복수 개의 조작된 레트론을 숙주 세포 내로 도입하는 단계이며, 여기서 각각의 레트론은 프로세싱되어 CRISPR 어레이 핵산 서열 내로 삽입될 수 있는 상이한 프로토스페이서 DNA 서열을 포함하는 것인 단계를 포함한다. 특정 실시양태에서, Cas1 단백질 또는 Cas2 단백질은 벡터에 의해 제공된다. 특정 실시양태에서, 조작된 레트론은 벡터에 의해 제공된다. 특정 실시양태에서, 복수 개의 조작된 레트론은 적어도 3개의 상이한 프로토스페이서 DNA 서열을 포함한다.
- [0028] 또 다른 측면에서, 생체내 분자 기록 시스템을 포함하는 조작된 세포가 제공되며, 이러한 조작된 세포는 하기를 포함한다: a) CRISPR 적응 시스템의 Cas1 단백질 또는 Cas2 단백질; b) 리더 서열과 적어도 하나의 반복 서열을 숙주 세포에 포함하는 CRISPR 어레이 핵산 서열, 여기서 CRISPR 어레이 핵산 서열은 조작된 세포 내의 게놈 DNA 또는 벡터에 통합됨; 및 c) 각각 CRISPR 프로토스페이서 DNA 서열을 포함하는 복수 개의 조작된 레트론, 여기서 각각의 레트론은 프로세싱되어 CRISPR 어레이 핵산 서열 내로 삽입될 수 있는 상이한 프로토스페이서 DNA 서열을 포함함. 특정 실시양태에서, Cas1 단백질 또는 Cas2 단백질은 벡터에 의해 제공된다. 특정 실시양태에서, 조작된 레트론은 벡터에 의해 제공된다. 특정 실시양태에서, 복수 개의 조작된 레트론은 적어도 3개의 상이한 프로토스페이서 DNA 서열을 포함한다.
- [0029] 또 다른 측면에서, 본원에 기재된 바와 같이, 생체내 분자 기록 시스템을 포함하는 조작된 세포를 포함하는 키트가 제공된다. 일부 실시양태에서, 키트는 생체내 분자 기록을 위한 지침을 추가로 포함한다.
- [0030] 또 다른 측면에서, 재조합 msDNA를 생산하는 방법이 제공되며, 이러한 방법은 a) 본원에 기재된 조작된 레트론 또는 벡터 시스템으로 숙주 세포를 형질감염시키는 단계; 및 b) msDNA가 생산되는 적합한 조건 하에 숙주 세포를 배양하는 단계를 포함한다.

**도면의 간단한 설명**

- [0031] **도 1a-1d**는 레트론 오페론의 개략도와 레트론의 잠재적 용도를 보여준다. **도 1a**는 msr, msd 및 역전사효소를 코딩하는 레트론 오페론의 개략도를 나타내며, 여기서 역전사효소는 멀티카피 단일 가닥 DNA를 코딩하는 msd 유전자의 일부분의 DNA 카피를 합성할 수 있다. **도 1b**는 재조합공학이 레트론의 잠재적 용도라는 것을 예시하며, 여기서 Beta는 ssDNA를 보호할 수 있고 ssDNA를 상보적 ssDNA 표적, 예를 들어 세포 내 DNA 표적으로 어닐링하는 것을 촉진시킬 수 있다. **도 1c**는 CRISPR/Cas9 유전자 편집이 레트론의 잠재적 용도라는 것을 예시하며, 여기서 레트론은 변이체 또는 돌연변이체 표적 부위를 복구할 수 있는 ssDNA 템플릿을 제공할 수 있다. **도 1d**는 분자 기록이 레트론의 잠재적 용도라는 것을 예시한다 (예를 들어, 그 전체 내용이 본원에 참조로 구체적으로 포함된 WO2018191525 A1에 제공된 바와 같다).
- 도 2a-2b**는 레트론 요소와 그의 어셈블리를 보여준다. **도 2a**는 레트론 요소: (1) msr 및 역전사된 msd의 5' 말단은 2'-5' 연결을 통해 프라이밍 구아노신에 공유 결합되며, 이러한 분기 구조는 게놈 조작에서의 사용을 방해하며, (2) 레트론 역전사에 필요할 수 있는 불변 플랭킹 영역으로, 따라서 이는 쉽게 복구 템플릿의 일부가 될 수 없으며, (3) 현재 제한된 전체 길이를 가진 것으로 생각되는 스템을 보여준다. 게놈 조작의 또 다른 문제는 Pol III 전사에 대한 종결인자로서 기능하는 천연 레트론 폴리 T 연장물이다. **도 2b**는 레트론 오페론의 비-단백질 코딩 (msr-msd) 부분이 중요한 2차 구조를 갖는 전사체를 생산한다는 것과, 역전사효소 (RT)가 이러한 전사체 내의 특정한 개시 부위를 인식하여 전사체를 RT-DNA (msd)로 부분적으로 역전사한다는 것을 예시한다.
- 도 3a-3d**. **도 3a**는 역전사 후 야생형 ec86 (레트론-Eco1 ncRNA라고도 함)의 염기 구조를 나타내며, 여기서 msd DNA는 상위에 있고 (서열식별번호 (SEQ ID NO): 1, GTCAGAAAAACGGGTTTCTGGTTGGCTCGGAGAGCATCAGGCGATGCTCTCCGTTCCAACAAGGAAAACAGACAGTAACCTAGA), msr RNA는 하위 서열이다 (서열식별번호: 2 -AUGCGACCCUUAAGCGAGAGGUUAUCAUUAAGGUCAACCUCUGGAUGUUGUUCGGCAUCCUGCAUUGAAUCUGAGUUACU). **도 3b**는 qPCR 분석에 의해 검출된 바와 같은, ec86의 발현 후에 생산된 ssDNA의 양을 예시한다. **도 3c**는 야생형 및 변이체 msd의 PAGE 분석을 보여준다. **도 3d**는 2가지 변이체 msd, 즉 ec86 야생형으로부터 변경된 레트론-Eco1 v32 ncRNA

(GTCAGAAAAACGGGTTGTCGCCAGTCTGACTGGCGACAAACAGCTTGTAACCTCAGA, 서열식별번호: 3) 및 v32 ncRNA로부터 변경된 레트론-Eco1 v35 ncRNA (GTCAGAAAAACGGGTTGAGAGGTTGCTGCAACCTCTCCATTTTCTGTAACCTCAGA, 서열식별번호: 4)의 염기 구조를 보여준다.

**도 4a-4d**는 확장된 msd ssDNA를 생산하기 위한 발현 시스템을 예시한다. **도 4a**는 더 긴 (변형된) 역전사된 msd ssDNA의 생산을 허용하는 레트론 역전사효소 (RT)로부터 msr/msd를 분할하는 발현 구축물을 보여준다. **도 4b**는 역전사효소 코딩 영역으로부터 (트랜스로) 분리되는 발현 카세트에서의 msr 및 msd의 배열을 보여준다. **도 4c**는 역전사효소 코딩 영역으로부터 (트랜스로) 분리되는 msr/msd 발현 카세트에서의 msd ssDNA의 몇 가지 확장을 예시하며, 이는 msd 영역이 이중 서열을 포함하도록 크게 확장될 수 있다는 것을 보여준다. **도 4d**는 **도 4a-4c**에 나타난 바와 같이 생산된 확장된 msd ssDNA를 포함한 msd ssDNA의 PAGE 분석을 나타낸다.

**도 5**는 변형될 수 있는 레트론 파라미터를 보여준다.

**도 6a-6f**. **도 6a**는 맞춤형된 시퀀싱 프랩 파이프라인을 개략적으로 예시한다. ssDNA는 RNase의 존재 하에 탈분지화 RNA 래리어트 1 (DBR1)로 처리된 후, 템플릿 비의존성 폴리머라제 (TdT)를 사용하여 단일 유형의 여러 개의 폴리뉴클레오티드를 추가하고, 어댑터를 함유하는 역 고정 프라이머를 사용하여 상보적 가닥을 생성하며, 제 2 어댑터를 라이게이션한 다음, 이러한 어댑터 연결된 이중 가닥 DNA를 인벡싱하고 다중화된 시퀀싱을 수행한다 (서열식별번호: 29). **도 6b**는 TdT에 의해 추가된 뉴클레오티드의 수가 제어 가능하다는 것을 보여준다. **도 6c**는 시퀀싱에 의한 검증을 예시하는, 주문형 msd ssDNA ec86 v 32 서열 (GTCAGAAAAACGGGTTGTCGCCAGTCTGACTGGCGACAAACAGCTTGTAACCTCAGA, 서열식별번호: 5)을 나타낸다. **도 6d**는 시퀀싱 결과 (GTCAGAAAAACGGGTTGTCGCCAGTCTGACTGGCGACAAACAGCTTGTAACCTCAG, 서열식별번호: 7)를 예시하는, 예측된 msd ssDNA ec86 v 32 서열 (GTCAGAAAAACGGGTT GTCGCCAGTCTGACTGGCGACAAACAGCTTGTAACCTCAGA, 서열식별번호: 6)을 나타낸다.

**도 6e**는 시퀀싱 결과 (GTCAGAAAAACGGGTTTCTGTTGGCTCGGAGAGCATCAGGCGATGCTCTCTCCGTTCCAACAAGGAAAACAGACAGTAACTCAG, 서열식별번호: 9)를 예시하는, 문헌 야생형 msd ssDNA ec86 서열 (GTCAGAAAAACGGGTTTCTGTTGGCTCGGAGAGCATCAGGCGATGCTCTCTCCGTTCCAACAAGGAAAACAGACAGTAACTCAGA, 서열식별번호: 8)를 보여준다. **도 6f**는 시퀀싱 결과 (AGCCGCGGAACAACTTTTGTATCCGAACCTACTGGATTGCGGCTCAAAAAGTTTGTCCGCAACTGTAATGTAATC, 서열식별번호: 11)를 예시하는, 문헌 야생형 msd ssDNA ec83 서열 (TTGAAGCCGCGGAACAACTTTTGTATCCGAACCTACTGGATTGCGGCTCAAAAAGTTTGTCCGCAACTGTAATGTAATC, 서열식별번호: 10)을 보여준다.

**도 7a-7c**는 msd DNA의 변형을 예시한다. **도 7a**는 레트론 RNA에서의 변화와 msd DNA에서 종료될 바코드의 연결을 개략적으로 보여준다. **도 7b**는 더 긴 포스트-msd 영역이 없는 야생형 레트론과 비교하여 더 긴 포스트-msd 상보적 영역을 갖는 레트론으로부터의 ssDNA 생산 증가를 보여준다. **도 7ca** 및 **7cb**는 레트론 비-코딩 RNA (ncRNA)의 5' 말단과 3' 말단에서 영역의 확장과 감소를 예시한다. **도 7ca**는 사용된 기본 레트론 구축물을 개략적으로 예시하며, 여기서 확장되는 ncRNA 내의 상보적 영역은 흑색 실선으로 표시되고 나머지 ncRNA는 점선으로 표시된다. **도 7cb**는 ncRNA 상보적 영역의 확장이 야생형 서열 (야생형의 존재량은 100%이다)에 비해 RT-DNA의 존재량을 증가시키지만, ncRNA 상보적 영역의 감소는 RT-DNA의 존재량을 감소시킨다는 것을 그래프로 보여준다. 표시된 데이터는 각각의 변이체 (n=3, 표시된 복제물)에 대한 풀링된 실험으로부터의 것이다.

**도 8a-8b**는 ncRNA의 역전사된 스템을 짧게 함으로써 ssDNA의 양을 감소시킬 수 있지만, 스템의 확장이 ssDNA 생산에 부정적인 영향을 미치지 않는다는 것을 그래프로 보여준다. **도 8a**는 변형된 ncRNA 구조의 일부분을 개략적으로 예시하며, 스템 영역은 흑색 실선으로서 표시된 반면, ncRNA의 나머지 부분은 점선으로 표시된다. **도 8b**는 ncRNA 영역을 약 15-30개 뉴클레오티드만큼 확장하는 것이, 비-확장된 야생형 ncRNA 서열에 대해 관찰된 것과 거의 동일한 수준으로 RT-DNA의 존재량을 유지하지만, ncRNA 영역의 길이가 약 14개 미만의 뉴클레오티드로 감소되면, 역전사에 의해 생성된 ssDNA의 양은 비-확장된 야생형 ncRNA 서열과 비교하여 감소한다는 것을 그래프로 보여준다.

**도 9a-9b**는 ncRNA의 역전사된 스템 영역을 파괴하고 고정하는 것의 효과를 보여준다. **도 9a**는 ncRNA의 개략도이며, 여기서 ncRNA의 역전사된 스템 영역은 흑색 실선으로서 표시된다. **도 9b**는 ncRNA 구조 변이체의 역전사된 DNA의 존재량을 야생형 서열과 비교하여 그래프로 보여준다. 데이터는 각각의 변이체에 대한 풀링된 실험으로부터의 것이다. 파괴된 스템, 고정된 스템 및 용인 가능한 선으로 파괴된 스템 ncRNA 구조 변이체에 대한 서열이 본 실시예에 제공된다.

**도 10a-10e**는 ncRNA로부터 역전사된 DNA의 존재량에 대한 ncRNA의 역전사된 영역에서의 삽입과 결실의 효과를 예시한다. **도 10a**는 ncRNA를 개략적으로 나타내며, 여기서 ncRNA의 역전사된 영역은 흑색 점선 및 실선으로서 표시된다. 점선은 msd 스템에 플랭킹하는 영역을 나타낸다. **도 10b**는 일련의 ncRNA 변이체의 역전사에 의해 생산된 RT-DNA 존재량을 그래프로 나타내며, 각각은 야생형 서열과 비교하여, msd 스템 루프를 따라 별개의 위치에서 3개의 염기 결실이 있다. 결실의 위치는 x축을 따라 플롯된다. **도 10c**는 일련의 ncRNA 변이체의 역전사에 의해 생산된 RT-DNA 존재량을 그래프로 나타내며, 각각은 야생형 서열과 비교하여, msd 스템 루프를 따라 별개의 위치에 3개의 염기가 삽입되었다. 삽입의 위치는 x축을 따라 플롯된다. **도 10d**는 일련의 ncRNA 변이체의 역전사에 의해 생산된 RT-DNA 존재량을 그래프로 나타내며, 각각은 야생형 서열과 비교하여, msd 스템 루프를 따라 별개의 위치에 단일 염기 변화가 있다. 삽입의 위치는 x축을 따라 플롯된다. **도 10e**는 **도 10b-10d**에 대해 관찰된 구조적 변화와 결과의 관점에서 msd 루프 위치의 변형 가능성 점수를 그래프로 나타낸다. 변형 가능성 점수는 이러한 변화의 평균 영향을 기반으로 하며, 여기서 데이터는 각각의 변이체에 대한 풀링된 실험으로부터의 것이었다. 풀링된 ncRNA에 대한 스템, 루프 및 플랭킹 영역의 도식이 **도 10b-10c**에 제시된다.

**도 11a-11b**는 CRISPR 기반 게놈 변화를 개선하기 위해 변형된 레트론의 용도를 예시한다. **도 11a**는 게놈 CRISPR 어레이를 변형시키기 위해 CRISPR 인테그라제 Cas1과 Cas2에 의한 레트론 RT-DNA의 통합을 예시하는 개략도이다. **도 11b**는 레트론 유래 스페이서 DNA가 ncRNA의 5' 및 3' 말단에서 자기 상보적 영역을 확장함으로써 증진될 수 있다는 것을 그래프로 보여준다.

### 정의

측정 가능한 값, 예컨대 양, 길이 등을 지칭할 때 본원에 사용된 바와 같은 용어 "약"은 명시된 값으로부터  $\pm 20\%$  또는  $\pm 10\%$ , 보다 바람직하게  $\pm 5\%$ , 보다 더 바람직하게  $\pm 1\%$ , 보다 더 바람직하게  $\pm 0.1\%$ 의 변동을 포괄하는 것을 의미한다.

핵산 분자를 설명하기 위해 본원에 사용된 바와 같은 "재조합"은 게놈, cDNA, 박테리아, 반합성 또는 합성 기원의 폴리뉴클레오티드를 의미하며, 그것의 기원이나 조작에 의해, 그것은 특성상 연관된 폴리뉴클레오티드의 전부 또는 일부분과 연관이 없다.

단백질 또는 폴리펩티드와 관련하여 사용된 바와 같은 용어 "재조합"은 재조합 폴리뉴클레오티드의 발현에 의해 생산된 폴리펩티드를 의미한다. 일반적으로, 관심 유전자는 하기에 추가로 기재되는 바와 같이, 클로닝된 다음 형질전환된 유기체에서 발현된다. 숙주 유기체는 발현 조건하에 단백질을 생산하기 위해 외래 유전자를 발현한다.

본원에 사용된 바와 같은 "세포"는 박테리아, 고세균, 진균, 원생생물, 식물 및 동물을 포함한, 원핵생물, 진핵생물 또는 고세균 유기체로부터 단리된 임의의 유형의 세포를 지칭하며, 조직, 기관 및 생검으로부터의 세포뿐만 아니라 재조합 세포, 시험관내에서 배양된 세포주로부터의 세포, 및 핵산을 포함하는 세포 단편, 세포 성분 또는 소기관을 포함한다. 상기 용어는 또한 인공 세포, 예컨대 나노입자, 리포솜, 폴리머솜, 또는 핵산을 캡슐화한 마이크로캡슐을 포괄한다. 본원에 기재된 방법은, 예를 들어, 단일 세포 또는 세포의 집단을 포함하는 샘플에 대해 수행될 수 있다. 상기 용어는 또한 유전자 변형된 세포를 포함한다.

용어 "형질전환"은 삽입에 사용된 방법에 관계없이 외인성 폴리뉴클레오티드 (예를 들어, 조작된 레트론)를 숙주 세포 내로 삽입하는 것을 지칭한다. 예를 들어, 직접 흡수, 형질도입 또는 f-교배가 포함된다. 외인성 폴리뉴클레오티드는 비-통합 벡터, 예를 들어 플라스미드로서 유지될 수 있거나 또는 대안적으로, 숙주 게놈 내로 통합될 수 있다.

"재조합 숙주 세포", "숙주 세포", "세포", "세포주", "세포 배양물" 및 단세포 실체로서 배양된 미생물 또는 고등 진핵 세포주를 나타내는 기타 이러한 용어는, 재조합 벡터 또는 기타 전달된 DNA에 대한 수용자로서 사용될 수 있거나, 또는 사용되어 온 세포를 지칭하며, 형질감염된 원래 세포의 원래 자손을 포함한다.

"코딩 서열" 또는 선택된 폴리펩티드를 "코딩"하는 서열은 생체내에서 적절한 조절 서열 (또는 "제어 요소")의 제어 하에 놓였을 때 전사되고 (DNA의 경우) 폴리펩티드로 번역되는 (mRNA의 경우) 핵산 분자이다. 코딩 서열의 경계는 5' (아미노) 말단의 출발 코돈과 3' (카르복시) 말단의 번역 정지 코돈에 의해 결정될 수 있다. 코딩 서열은 바이러스, 원핵 또는 진핵 mRNA로부터의 cDNA, 바이러스 또는 원핵 DNA로부터의 게놈 DNA 서열, 및 심지어 합성 DNA 서열을 포함할 수 있으나, 이에 제한되지는 않는다. 전사 종결 서열은 코딩 서열에 대해 3'에 위치할 수 있다.

전형적인 "제어 요소"는 전사 프로모터, 전사 인핸서 요소, 전사 종결 신호, 폴리아데닐화 서열 (번역 정지 코돈에 대해 3'에 위치함), 번역 개시의 최적화를 위한 서열 (코딩 서열에 대해 5'에 위치함) 및 번역 종결 서열을 포함하나 이에 제한되지는 않는다.

"작동 가능하게 연결된"은 언급된 구성 요소가 통상적인 기능을 수행하도록 구성되는 요소의 배열을 지칭한다. 따라서, 적당한 효소가 존재할 때, 코딩 서열과 작동 가능하게 연결된 주어진 프로모터는 코딩 서열의 발현에 영향을 미칠 수 있다. 프로모터는 그의 발현을 지시하는 기능을 하는 한, 코딩 서열과 연속적일 필요는 없다. 따라서, 예를 들어, 중간에 번역되지 않았지만 전사된 서열이 프로모터 서열과 코딩 서열 사이에 존재할 수 있고 프로모터 서열은 여전히 코딩 서열에 "작동 가능하게 연결된" 것으로 간주될 수 있다.

"~에 의해 코딩된"은 폴리펩티드 또는 RNA 서열을 코딩하는 핵산 서열을 지칭한다. 예를 들어, 폴리펩티드 서열 또는 그의 일부분은 핵산 서열에 의해 코딩된 폴리펩티드로부터의 적어도 3 내지 5개의 아미노산, 보다 바람직하게 적어도 8 내지 10개의 아미노산, 보다 더 바람직하게 적어도 15 내지 20개의 아미노산의 아미노산 서열을 함유한다. RNA 서열 또는 그의 일부분은 적어도 3 내지 5개의 뉴클레오티드, 보다 바람직하게 적어도 8 내지 10개의 뉴클레오티드, 보다 더 바람직하게 적어도 15 내지 20개의 뉴클레오티드의 뉴클레오티드 서열을 함유한다.

용어 "단리된", "정제된" 또는 "생물학적으로 순수한"은 그의 천연 상태에서 발견되는 바와 같이 이를 정상적으로 동반하는 성분들이 다양한 정도로 포함되지 않은 물질을 지칭한다. "단리하다"는 원래 공급원 또는 환경으로부터의 분리 정도를 의미한다. "정제하다"는 단리보다 더 높은 분리 정도를 의미한다. "정제된" 또는 "생물학적으로 순수한" 단백질은 다른 물질이 충분히 없기 때문에 어떠한 불순물도 단백질의 생물학적 특성에 물질적으로 영향을 미치지 않거나 또는 다른 불리한 결과를 유발시키지 않는다. 즉, 본 발명의 핵산 또는 펩티드는 재조합 DNA 기술에 의해 생산될 때 세포 물질, 바이러스 물질 또는 배양 배지가 실질적으로 없는 경우 또는 화학적으로 합성될 때 화학적 전구체 또는 기타 화학물질이 없는 경우에 정제된 상태이다. 순도 및 균질성은 전형적으로, 분석 화학 기술, 예를 들어 폴리아크릴아미드 겔 전기영동 또는 고성능 액체 크로마토그래피를 사용하여 결정된다. 용어 "정제된"은 핵산 또는 단백질이 전기영동 겔에서 본질적으로 하나의 밴드를 생성한다는 것을 의미할 수 있다. 변형될 수 있는, 예를 들어 인산화 또는 글리코실화될 수 있는 단백질의 경우, 상이한 변형은 상이한 단리된 단백질을 발생시킬 수 있으며, 이는 별도로 정제될 수 있다.

"실질적으로 정제된"은 일반적으로 특정 물질 (화합물, 폴리뉴클레오티드, 단백질, 폴리펩티드, 펩티드 조성물)이 그것이 존재하는 샘플의 대다수 퍼센트를 차지하도록 이러한 물질을 단리하는 것을 지칭한다. 전형적으로, 샘플 중의 실질적으로 정제된 성분은 샘플의 50%, 바람직하게 80%-85%, 보다 바람직하게 90-95%를 차지한다. 관심 폴리뉴클레오티드 및 폴리펩티드를 정제하는 기술은 관련 기술분야에 널리 공지되어 있으며, 예를 들어 이온 교환 크로마토그래피, 친화성 크로마토그래피, 및 밀도에 따른 침강을 포함한다.

"발현"은 세포에 의한 유전자 산물의 검출 가능한 생산을 지칭한다. 유전자 산물은 문맥에 따라 "유전자 발현"으로서 지칭될 수 있는 전사 산물 (즉, RNA)일 수 있거나, 또는 유전자 산물은 전사 산물의 번역 산물 (즉, 단백질)일 수 있다. "정제된 폴리뉴클레오티드"는 폴리뉴클레오티드와 자연적으로 연합되는 단백질 및/또는 핵산이 본질적으로 없는, 예를 들어 이를 약 50% 미만, 바람직하게 약 70% 미만, 보다 바람직하게 적어도 약 90% 미만 함유하는 관심 폴리뉴클레오티드 또는 그의 단편을 지칭한다. 관심 폴리뉴클레오티드를 정제하기 위한 기술은 관련 기술분야에서 이용 가능하며, 예를 들어, 폴리뉴클레오티드를 함유하는 세포를 카오트로픽제로 붕괴시키는 것, 및 이온 교환 크로마토그래피, 친화성 크로마토그래피 및 밀도에 따른 침강에 의한 폴리뉴클레오티드(들) 및 단백질의 분리를 포함한다.

용어 "형질감염"은 세포에 의한 외래 DNA의 흡수를 지칭하는 데 사용된다. 세포막 내부에 외인성 DNA가 도입되었을 때 세포는 "형질감염"된 것이다. 다수의 형질감염 기술이 일반적으로 관련 기술분야에 공지되어 있다. 예를 들어, 문헌 [Graham et al. (1973) Virology, 52:456, Sambrook et al. (2001) Molecular Cloning, a laboratory manual, 3rd edition, Cold Spring Harbor Laboratories, New York, Davis et al. (1995) Basic Methods in Molecular Biology, 2nd edition, McGraw-Hill, and Chu et al. (1981) Gene 13:197]을 참조한다. 이러한 기술은 하나 이상의 외인성 DNA 모이어티를 적합한 숙주 세포 내로 도입하는 데 사용될 수 있다. 상기 용어는 유전 물질의 안정적이면서 일시적인 흡수를 지칭하며, 펩티드 연결된 DNA 또는 항체 연결된 DNA의 흡수를 포함한다.

백터는 표적 세포에 핵산 서열을 전달할 수 있다 (예를 들어, 바이러스 백터, 비-바이러스 백터, 입자 운반체 및 리포솜). 전형적으로, "백터 구축물", "발현 백터" 및 "유전자 전이 백터"는 관심 핵산의 발현을 지시할 수

있고 핵산 서열을 표적 세포로 전달할 수 있는 임의의 핵산 구축물을 의미한다. 따라서, 상기 용어는 바이러스 벡터뿐만 아니라 클로닝 및 발현 비히클을 포함한다.

"포유류 세포"는 본원에 기재된 바와 같은 조작된 레트론 또는 조작된 레트론을 포함하는 벡터 시스템으로 형질 감염시키기에 적합한 포유류 대상체로부터 유래된 임의의 세포를 지칭한다. 세포는 이종, 자가 또는 동종이게 될 수 있다. 세포는 포유류 대상체로부터 직접 수득된 1차 세포일 수 있다. 세포는 포유류 대상체로부터 수득된 세포의 배양 및 확장으로부터 유래된 세포일 수도 있다. 불멸화된 세포가 또한 이러한 정의 내에 포함된다. 일부 실시양태에서, 세포는 재조합 단백질 및/또는 핵산을 발현하도록 유전자 조작되었다.

용어 "대상체"는 척추동물과 무척추동물을 모두 포함한 동물을 포함하며, 이는 무척추동물, 예컨대 절지동물, 연체동물, 환형동물 및 자포동물; 및 척추동물, 예컨대 개구리, 도롱뇽 및 무족영원류를 포함한 양서류; 도마뱀, 뱀, 거북이, 크로커다일 및 앨리게이터를 포함한 파충류; 어류; 인간 및 비-인간 포유류, 예컨대 침팬지 및 다른 유인원 및 원숭이 종을 포함한 비-인간 영장류를 포함한 포유류; 실험실 동물, 예컨대 마우스, 래트, 토끼, 햄스터, 기니피그, 및 친칠라; 가축, 예컨대 개 및 고양이; 농장 동물, 예컨대 양, 염소, 돼지, 말 및 암소; 및 조류, 예컨대 닭, 칠면조 및 기타 갈금류 조류, 오리, 거위 등을 포함한 가축, 야생 및 사냥용 조류 등을 포함하나 이에 제한되지는 않는다. 일부 경우에, 본원에 개시된 방법은 실험 동물, 수의학 적용, 및 질병을 위한 동물 모델의 개발에 사용되며, 이러한 동물 모델은 마우스, 래트, 및 햄스터를 포함한 설치류; 영장류, 및 트랜스제닉 동물을 포함하나 이에 제한되지는 않는다.

"유전자 전이" 또는 "유전자 전달"은 숙주 세포 내로 관심 DNA 또는 RNA를 안정적으로 삽입하는 방법 또는 시스템을 지칭한다. 이러한 방법은 비-통합 전이된 DNA의 일시적인 발현, 염색체의 복제 및 전이된 레플리콘 (예를 들어, 에피솜)의 발현, 또는 숙주 세포의 게놈 DNA로 전이된 유전 물질의 통합을 초래할 수 있다. 유전자 전달 발현 벡터는 박테리아 플라스미드 벡터, 바이러스 벡터, 비-바이러스 벡터, 알파바이러스, 폭스 바이러스 및 벡시니아 바이러스로부터 유래된 벡터를 포함하나 이에 제한되지는 않는다.

용어 "~로부터 유래된"은 분자의 원래 공급원을 확인하기 위해 본원에 사용되지만, 예를 들어 화학적 합성 또는 재조합 수단에 의해 이루어질 수 있는 분자의 제조 방법을 제한하려는 것은 아니다.

지정된 서열 "로부터 유래된" 폴리뉴클레오티드는 지정된 뉴클레오티드 서열의 영역에 상응하는, 즉 이러한 영역과 동일하거나 이에 상보적인 대략 적어도 약 6개의 뉴클레오티드, 바람직하게 적어도 약 8개의 뉴클레오티드, 보다 바람직하게 적어도 약 10-12개의 뉴클레오티드, 보다 더 바람직하게 적어도 약 15-20개의 뉴클레오티드의 연속된 서열을 포함하는 폴리뉴클레오티드 서열을 지칭한다. 이와 같이 유래된 폴리뉴클레오티드는 반드시 관심 뉴클레오티드 서열로부터 물리적으로 유래될 필요는 없지만, 폴리뉴클레오티드가 유래된 그 영역(들) 내의 염기 서열에 의해 제공된 정보에 기초하여 화학적 합성, 복제, 역전사 및 전사를 포함하나 이에 제한되지는 않는 어떠한 방식으로든 생성될 수 있다. 따라서, 이것은 원래 폴리뉴클레오티드의 센스 또는 안티센스 배향을 나타낼 수 있다.

"바코드"는 바코드와 연관된 핵산 또는 세포를 확인하기 위해 사용되는 하나 이상의 뉴클레오티드 서열을 지칭한다. 바코드는 3-1000개 이상의 뉴클레오티드 길이, 바람직하게 10-250개 뉴클레오티드 길이, 보다 바람직하게 10-30개 뉴클레오티드 길이일 수 있으며, 이는 이러한 범위 내의 임의의 길이, 예컨대 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 또는 1000개 뉴클레오티드 길이를 포함한다. 바코드는, 예를 들어 핵산이 유래된 단일 세포, 세포의 하위 집단, 콜로니 또는 샘플을 확인하기 위해 사용될 수 있다. 바코드는 또한 핵산이 유래된 세포, 콜로니 또는 샘플의 위치 (즉, 위치 바코드), 예컨대 세포 어레이에서의 콜로니의 위치, 다중 웰 플레이트에서의 웰의 위치, 또는 랙에서의 튜브, 플라스크 또는 다른 용기의 위치를 확인하기 위해 사용될 수 있다. 예를 들어, 바코드는 핵산이 유래된 유전자 변형된 세포를 확인하기 위해 사용될 수 있다. 일부 실시양태에서, 바코드는 특정한 유형의 게놈 편집 또는 특정한 유형의 공여자 핵산을 확인하기 위해 사용된다.

용어 "혼성화하다" 및 "혼성화"는 왓슨-크릭 염기 쌍 형성을 통해 복합체를 형성하기에 충분히 상보적인 뉴클레오티드 서열 간의 복합체의 형성을 지칭한다.

용어 "상동 영역"은 또 다른 핵산 영역과의 상동성을 갖는 핵산의 영역을 지칭한다. 따라서, 핵산 분자에 상동 영역이 존재하는지 여부는 동일하거나 상이한 분자 내의 또 다른 핵산 영역을 참조하여 결정된다. 추가로, 핵산은 종종 이중 가닥이기 때문에, 본원에 사용된 바와 같은 용어 "상동 영역"은 핵산 분자들이 서로 혼성화할

수 있는 능력을 지칭한다. 예를 들어, 단일 가닥 핵산 분자는 서로 혼성화할 수 있는 2개의 상동 영역을 가질 수 있다. 따라서, 용어 "상동 영역"은 상보적 서열을 가진 핵산 세그먼트를 포함한다. 상동 영역은 길이가 다양할 수 있지만, 전형적으로 4 내지 500개 뉴클레오티드 (예를 들어, 약 4 내지 약 40개, 약 40 내지 약 80개, 약 80 내지 약 120개, 약 120 내지 약 160개, 약 160 내지 약 200개, 약 200 내지 약 240개, 약 240 내지 약 280개, 약 280 내지 약 320개, 약 320 내지 약 360개, 약 360 내지 약 400개, 약 400 내지 약 440개 등)일 것이다.

본원에 사용된 바와 같은, 용어 "상보적" 또는 "상보성"은 서로 염기 쌍을 형성할 수 있는 폴리뉴클레오티드를 지칭한다. 염기 쌍은 전형적으로, 폴리뉴클레오티드 가닥 간의 역병렬 배향의 뉴클레오티드 단위 간의 수소 결합에 의해 형성된다. 상보적인 폴리뉴클레오티드 가닥은 왓슨-크릭 방식으로 염기 쌍을 형성할 수 있거나 (예를 들어, A와 T, A와 U, C와 G) 또는 이중체의 형성을 허용하는 임의의 다른 방식으로 염기 쌍을 형성할 수 있다. 관련 기술분야의 통상의 기술자가 알고 있는 바와 같이, DNA와 반대로 RNA를 사용할 때는, 티민 (T)이 아닌 우라실 (U)이 아데노신에 상보적인 것으로 간주되는 염기이다. 그러나, 본 발명의 맥락에서 우라실을 표기하는 경우, 달리 명시되지 않는 한, 티민을 대체할 수 있는 능력이 암시된다. "상보성"은 2개의 RNA 가닥 간에, 2개의 DNA 가닥 간에, 또는 RNA 가닥과 DNA 가닥 간에 존재할 수 있다. 일반적으로 2개 이상의 폴리뉴클레오티드는 덜 완벽하거나 100% 미만의 상보성을 가지고 있음에도 불구하고 "상보적"일 수 있고 이중체를 형성할 수 있는 것으로 이해된다. 상보성 영역을 포함하는 각각의 폴리뉴클레오티드 서열의 적어도 연속된 부분이, 이러한 영역 내에 임의의 미스매치 또는 중단 없이 다른 폴리뉴클레오티드와 완벽히 염기 쌍을 형성하는 경우, 이러한 두 서열은 "완벽히 상보적"이거나 또는 "100% 상보적"이다. 각각의 폴리뉴클레오티드 내의 연속되는 상보성 영역이 다른 폴리뉴클레오티드와 완벽하게 혼성화할 수 있는 한은, 2개 이상의 서열은 그 폴리뉴클레오티드 중 하나 또는 둘 다가 부가의 비-상보적 서열을 함유할지라도 "완벽하게 상보적"이거나 또는 "100% 상보적"인 것으로 간주된다. "덜 완벽한" 상보성은 그러한 상보성 영역 내의 연속되는 뉴클레오티드 전부보다 적은 수가 서로 염기 쌍을 형성할 수 있는 상황을 지칭한다. 두 폴리뉴클레오티드 서열 간의 상보성의 백분율을 결정하는 것은 관련 기술분야의 통상의 기술 문제이다.

본원에 사용된 바와 같은 용어 "Cas9"는 임의의 종으로부터의 유형 II 클러스터된 규칙적으로 공간을 둔 짧은 팔린드롬성 반복부 (CRISPR) 시스템 Cas9 엔도뉴클레아제를 포괄하며, 또한 Cas9 엔도뉴클레아제 활성을 보유하는 (즉, 이중 가닥 파손을 생성하기 위해 DNA의 부위 지정 절단을 촉매하는) 그의 생물학적으로 활성인 단편, 변이체, 유사체 및 유도체를 포함한다. Cas9 엔도뉴클레아제는 그의 결합된 가이드 RNA (gRNA)와 상보적인 서열을 포함하는 부위에서 DNA와 결합하고 이를 절단한다. Cas9 표적화를 위해, gRNA는 이중체를 형성하기에 충분한 염기 쌍을 형성할 수 있는 (즉, gRNA는 표적 서열과 혼성화한다), 표적 서열 (예를 들어, 메이저 또는 마이너 대립 유전자)과 "상보적인" 서열을 포함할 수 있다. 부가적으로, gRNA는 PAM 서열과 상보적인 서열을 포함할 수 있으며, 여기서 gRNA는 또한 표적 DNA 내의 PAM 서열과 혼성화한다.

용어 "공여자 폴리뉴클레오티드"는 HDR 또는 재조합공학에 의해 표적 유전자좌에서 게놈에 통합되는 것으로 의도된 편집의 서열을 제공하는 폴리뉴클레오티드를 지칭한다.

"표적 부위" 또는 "표적 서열"은 가이드 RNA (gRNA) 또는 공여자 폴리뉴클레오티드의 상동성 아암에 의해 인식되는 (즉, 혼성화를 위해 충분히 상보적인) 핵산 서열이다. 표적 부위는 대립 유전자에 특이적일 수 있다 (예를 들어, 메이저 또는 마이너 대립 유전자). 예를 들어, 표적 부위는, 예컨대 하나 이상의 뉴클레오티드의 삽입, 하나 이상의 뉴클레오티드의 대체, 하나 이상의 뉴클레오티드의 결실, 또는 그의 조합에 의해 변형시키고자 하는 게놈 부위일 수 있다.

"상동성 아암"은 공여자 폴리뉴클레오티드가 세포에서 편집될 게놈 서열을 표적화하는 데 책임이 있는 공여자 폴리뉴클레오티드의 일부분을 의미한다. 공여자 폴리뉴클레오티드는 전형적으로 게놈 DNA에 대한 의도된 편집을 포함하는 뉴클레오티드 서열에 플랭킹하는 3' 게놈 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 게놈 표적 서열과 혼성화하는 5' 상동성 아암을 포함한다. 상동성 아암은 본원에서 5' 및 3' (즉, 상류 및 하류) 상동성 아암으로서 지칭되며, 이는 공여자 폴리뉴클레오티드 내의 의도된 편집을 포함하는 뉴클레오티드 서열에 대한 상동성 아암의 상대적 위치에 관한 것이다. 5' 및 3' 상동성 아암은 본원에서 각각 "5' 표적 서열" 및 "3' 표적 서열"로서 지칭되는, 변형될 게놈 DNA에서의 표적 유전자좌 내의 영역과 혼성화한다. 예를 들어, 의도된 편집을 포함하는 뉴클레오티드 서열은 5' 및 3' 상동성 아암에 의해 인식된 (즉, 혼성화를 위해 충분히 상보적인) 게놈 표적 유전자좌에서 HDR 또는 재조합공학에 의해 게놈 DNA로 통합될 수 있다.

일반적으로, "CRISPR 적응 시스템"은 Cas 유전자를 코딩하는 서열, 및 리더 서열과 적어도 하나의 반복 서열을

포함한 CRISPR 어레이 핵산 서열을 포함한, CRISPR 관련 ("Cas") 유전자의 발현에 관여하거나 또는 그의 활성을 지시하는 전사체 및 다른 요소를 집합적으로 지칭한다. 일부 실시양태에서, CRISPR 적응 시스템의 하나 이상의 요소는 유형 I, 유형 II 또는 유형 III CRISPR 시스템으로부터 유래된다. Cas1 및 Cas2는 3가지 유형의 CRISPR-Cas 시스템 모두에서 발견되며, 스페이서 획득에 관여한다. 이. 콜라이의 I-E 시스템에서, Cas1과 Cas2는 복합체를 형성하며, 여기서 Cas2 이량체가 2개의 Cas1 이량체를 브리징한다. 이러한 복합체에서 Cas2는 침입 DNA의 이중 가닥 단편과 결합하는 비-효소적 스캐폴딩 역할을 수행하는 반면, Cas1은 DNA의 단일 가닥 플랭크와 결합하고 CRISPR 어레이로의 통합을 촉매한다.

일부 실시양태에서, CRISPR 시스템의 하나 이상의 요소는 내인성 CRISPR 시스템을 포함하는 특정한 유기체, 예컨대 스트렙토코쿠스 피오게네스(*Streptococcus pyogenes*)로부터 유래된다. 일반적으로, CRISPR 시스템은 표적 서열의 부위에서 CRISPR 복합체의 형성을 촉진시키는 요소로서 특징지어진다 (내인성 CRISPR 시스템의 맥락에서 프로토스페이서로서 지칭되기도 함).

일부 실시양태에서, 박터는 CRISPR 효소, 예컨대 Cas 단백질을 코딩하는 효소-코딩 서열과 작동 가능하게 연결된 조절 요소를 포함한다. Cas 단백질의 비-제한적인 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (Csn1 및 Csx12로서 공지되기도 함), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, 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, 그의 상동체, 또는 그의 변형된 버전을 포함한다.

특정 실시양태에서, 본 개시내용은 프로토스페이서 인접 모티프 (PAM)라고 불리는 짧은 (3 - 5 bp) DNA 서열에 인접한 프로토스페이서를 제공한다. PAM은 획득 동안 유형 I 및 유형 II 시스템에 중요하다. 유형 I 및 유형 II 시스템에서, PAM 서열에 인접한 위치에서 프로토스페이서를 절제하고, 스페이서의 다른 쪽 말단은 롤러 메커니즘을 사용하여 커팅하여, CRISPR 어레이에서 스페이서 크기의 규칙성을 유지한다. PAM 서열의 보존은 CRISPR-Cas 시스템마다 상이하하며, Cas1 및 리더 서열과 진화적으로 연결될 수 있다.

일부 실시양태에서, 본 개시내용은 리더 서열에 인접하여 우선적으로 발생하지만 배타적이지 않은 방향성 방식으로 세포 내의 조작된 레트론 시스템을 사용함으로써 세포 내에서 생산되는 규정된 합성 DNA를 CRISPR 어레이로 통합하는 것을 제공한다. 이. 콜라이로부터의 유형 I-E 시스템에서, 리더 서열에 인접한 제1 직접 반복부가 카피되며, 새로 획득한 스페이서는 제1 직접 반복부와 제2 직접 반복부 사이에 삽입되는 것으로 입증되었다.

한 실시양태에서, 프로토스페이서는 규정된 합성 DNA이다. 일부 실시양태에서, 규정된 합성 DNA는 적어도 3, 5, 10, 20, 30, 40 또는 50개 뉴클레오티드, 또는 3 내지 50개, 또는 10 내지 100개, 또는 20 내지 90개, 또는 30 내지 80개, 또는 40 내지 70개, 또는 50 내지 60개 뉴클레오티드 길이이다. 한 실시양태에서, 올리고 뉴클레오티드 서열 또는 규정된 합성 DNA는 변형된 "AAG" 프로토스페이서 인접 모티프 (PAM)를 포함한다.

일부 실시양태에서, 조절 요소는 CRISPR 시스템의 하나 이상의 요소의 발현을 구동하기 위해 CRISPR 시스템의 하나 이상의 요소와 작동 가능하게 연결되어 있다. 일반적으로, SPIDR (스페이서 공간을 둔 직접 반복부)로서 공지되기도 한 CRISPR (클러스터된 규칙적으로 공간을 둔 짧은 팔린드롬성 반복부)은 특정한 박테리아 종에 통상적으로 특이적인 DNA 유전자좌의 패밀리를 구성한다. CRISPR 유전자좌는 이. 콜라이 (문헌 [Ishino et al., J. Bacteriol., 169:5429-5433 (1987); and Nakata et al., J. Bacteriol., 171:3553-3556 (1989)]) 및 관련 유전자에서 인식된 공간을 둔 짧은 서열 반복부 (SSR)의 별개의 클래스를 포함한다. 할로페락스 메디테라네이(*Haloferax mediterranei*), 스트렙토코쿠스 피오게네스, 아나바에나(*Anabaena*), 및 미코박테리움 투베르쿨로시스(*Mycobacterium tuberculosis*)에서도 유사한 공간을 둔 SSR이 확인되었다 (문헌 [Groenen et al., Mol. Microbiol., 10:1057-1065 (1993); Hoe et al., Emerg. Infect. Dis., 5:254-263 (1999); Masepohl et al., Biochim. Biophys. Acta 1307:26-30 (1996); and Mojica et al., Mol. Microbiol., 17:85-93 (1995)] 참조). CRISPR 유전자좌는 전형적으로, 짧은 규칙적으로 이격된 반복부 (SRSR)라고 불리는 반복부의 구조에 의해 다른 SSR과 상이하다 (Janssen et al., OMICS J. Integ. Biol., 6:23-33 (2002); and Mojica et al., Mol. Microbiol., 36:244-246 (2000)). 일반적으로, 상기 반복부는 실질적으로 일정한 길이의 독특한 개체 서열에 의해 규칙적으로 이격된 클러스터에서 발생하는 짧은 요소이다 (상기 문헌 [Mojica et al., (2000)] 참조). 반복 서열이 균주 간에 고도로 보존되긴 하지만, 공간을 둔 반복부의 수와 스페이서 영역의 서열은 전형적으로 균주마다 상이하다 (van Embden et al., J. Bacteriol., 182:2393-2401 (2000)). CRISPR 유전자좌는 아에로피룸(*Aeropyrum*), 피로바쿨룸(*Pyrobaculum*), 숄폴로부스(*Sulfolobus*), 아르케오글로부스(*Archaeoglobus*), 할로카르쿨라(*Halocarcula*), 메타노박테리움(*Methanobacterium*), 메타노코쿠스(*Methanococcus*), 메타노사르키나

(*Methanosarcina*), 메타노피루스(*Methanopyrus*), 피로코쿠스(*Pyrococcus*), 피크로필루스(*Picrophilus*), 썬니오플라스니아(*Thermioplasmia*), 코리네박테리움(*Corynebacterium*), 미코박테리움(*Mycobacterium*), 스트렙토미세스(*Streptomyces*), 아퀴팩스(*Aquifex*), 포르피로모나스(*Porphyromonas*), 클로로비움(*Chlorobium*), 써무스(*Thermus*), 바실루스(*Bacillus*), 리스테리아(*Listeria*), 스태필로코쿠스(*Staphylococcus*), 클로스트리디움(*Clostridium*), 썬모아나어로박터(*Thermoanaerobacter*), 미코플라스마(*Mycoplasma*), 푸소박테리움(*Fusobacterium*), 아자르쿠스(*Azarcus*), 크로모박테리움(*Chromobacterium*), 네이세리아(*Neisseria*), 니트로소모나스(*Nitrosomonas*), 데술포비브리오(*Desulfovibrio*), 지오박터(*Geobacter*), 미로코쿠스(*Myrococcus*), 캄필로박터(*Campylobacter*), 울리넬라(*Wolinella*), 아시네토박터(*Acinetobacter*), 에르위니아(*Erwinia*), 에스케리키아(*Escherichia*), 레지오넬라(*Legionella*), 메틸로코쿠스(*Methylococcus*), 파스퇴렐라(*Pasteurella*), 포토박테리움(*Photobacterium*), 살모넬라(*Salmonella*), 크산토모나스(*Xanthomonas*), 예르시니아(*Yersinia*), 트레포네마(*Treponema*) 및 썬모토가(*Thermotoga*)를 포함하나 이에 제한되지는 않는 40개 초과 원핵생물 (예를 들어, 문헌 [Jansen et al., Mol. Microbiol., 43:1565-1575 (2002); and Mojica et al., (2005)] 참조)에서 확인되었다.

일부 실시양태에서, CRISPR 효소를 코딩하는 효소 코딩 서열은 특정한 세포, 예컨대 진핵 세포에서의 발현을 위해 코돈 최적화된다. 진핵 세포는 특정한 유기체, 예컨대 인간, 마우스, 래트, 토끼, 개 또는 비-인간 영장류를 포함하나 이에 제한되지는 않는 포유류의 것이거나 그로부터 유래될 수 있다. 일반적으로, 코돈 최적화는 천연 아미노산 서열을 유지하면서 숙주 세포의 유전자에서 더 자주 또는 가장 자주 사용되는 코돈으로 천연 서열의 적어도 하나의 코돈 (예를 들어, 약 1개 또는 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개 초과 또는 그보다 많은 코돈)을 대체함으로써 관심 숙주 세포에서의 발현을 증진시키기 위해 핵산 서열을 변형시키는 프로세스를 지칭한다. 다양한 종이 특정한 아미노산의 특정 코돈에 대해 특정한 편향을 나타낸다. 코돈 편향 (유기체 간 코돈 사용량의 차이)은 종종 메신저 RNA (mRNA)의 번역 효율과 상관이 있는데, 이는 특히 번역되는 코돈의 특성과 특정한 전달 RNA (tRNA) 분자의 가용성에 의존적인 것으로 여겨진다. 세포에서 선택된 tRNA의 우위는 일반적으로 펩티드 합성에 가장 자주 사용되는 코돈의 반영이다. 따라서, 유전자는 코돈 최적화를 기반으로 하여 주어진 유기체에서 최적의 유전자 발현에 맞게 조정될 수 있다. 코돈 사용량 표는, 예를 들어 "코돈 사용량 데이터베이스"에서 쉽게 입수할 수 있으며, 이러한 표는 수많은 방식으로 적용될 수 있다. 문헌 [Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000)]을 참조한다. 특정한 숙주 세포에서의 발현을 위해 특정한 서열을 코돈 최적화하는 컴퓨터 알고리즘, 예컨대 진 포지(Gene Forge) [압타젠 (Aptagen; 미국 펜실베이니아주 제이코부스)]가 또한 이용 가능하다. 일부 실시양태에서, CRISPR 효소를 코딩하는 서열 내의 하나 이상의 코돈 (예를 들어, 1, 2, 3, 4, 5, 10, 15, 20, 25, 50개, 또는 그 초과, 또는 모든 코돈)은 특정한 아미노산에 대해 가장 자주 사용되는 코돈에 상응한다.

핵산, 예컨대 조작된 레트론 구축물 또는 조작된 레트론 구축물을 포함하는 벡터를 세포에게 "투여하는 것"은 핵산이 세포막을 가로질러 운반될 수 있는 임의의 수단, 즉 형질도입, 형질감염, 전기천공, 전위, 융합, 식균 작용, 슈팅 또는 탄도 방법 등을 포함한다.

본 개시내용을 추가로 기재하기에 앞서, 개시된 주제가 상술한 특정한 실시양태에 제한되는 것은 아니며, 물론 다양할 수 있다는 점을 이해해야 한다. 본원에 사용되는 전문 용어는 특정한 실시양태만을 설명하기 위한 것으로서, 본 개시내용의 범위가 첨부된 청구범위에 의해서만 제한될 것이기 때문에, 이를 제한하기 위한 것은 아니라는 점도 이해해야 한다.

값의 범위가 제공되는 경우, 문맥상 명확하게 지시되지 않는 한, 하한 단위의 10분의 1까지의 각각의 중간 값, 그러한 범위의 상한과 하한 사이, 및 그러한 언급된 범위 내의 임의의 다른 명시된 또는 중간 값이 본원에 개시된 주제 내에 포괄되는 것으로 이해된다. 이러한 더 작은 범위의 상한 및 하한은 독립적으로 더 작은 범위에 포함될 수 있으며, 명시된 범위에서 구체적으로 배제된 임의의 제한에 따라 본원에 개시된 주제 내에 포괄된다. 명시된 범위가 한계치 중 1가지 또는 2가지 모두를 포함하는 경우, 그러한 1가지 또는 2가지 한계치를 제외한 범위 또한 본원에 개시된 주제에 포함된다.

달리 정의되지 않는 한, 본원에서 사용되는 모든 기술적 및 과학적 용어는 본원에 개시된 주제가 속하는 기술분야에서 통상의 기술자에 의해 통상적으로 이해되는 바와 동일한 의미를 갖는다. 본원에 기재된 방법 및 물질과 유사하거나 동등한 임의의 방법 및 물질이 또한, 본원에 개시된 주제의 실시 또는 시험에 사용될 수 있지만, 바람직한 방법 및 물질이 하기에 설명된다. 본원에 언급된 모든 간행물은 그 간행물이 인용하는 것과 연계해서

방법 및/또는 물질을 개시하고 설명하기 위해 참조로 본원에 포함된다.

본원에 사용된 바와 같고 첨부된 청구범위 내에서, 문맥상 달리 명백히 지시하지 않는 한, 단수 형태는 복수 개의 지시 대상을 포함한다는 점에 유의해야 한다. 따라서, 예를 들어, "세포"에 대한 언급은 이러한 세포 복수 개를 포함하고, "핵산"에 대한 언급은 관련 기술분야의 통상의 기술자에게 공지된 하나 이상의 핵산 및 그의 등가물에 대한 언급 등을 포함한다. 또한 청구범위는 어떠한 임의적 요소를 배제하기 위해 초안을 작성할 수 있다는 점에 유의한다. 이와 같이, 본 진술은 "부정적" 제한 사항의 사용을 포함하는 본원에 기재된 임의의 특징 또는 요소의 나열과 연계해서 "오로지", "단지" 등으로서 유사한 배타적 전문 용어를 사용하기 위한 선행적 근거를 제공하도록 의도되었다.

명확성을 위해 별도의 실시양태의 맥락에서 기재되는, 본원에 개시된 주제의 특징의 특징은 단일 실시양태에서 조합하여 제공될 수도 있음을 인지해야 한다. 반대로, 간결성을 위해 단일 실시양태의 맥락에서 기재되는, 본원에 개시된 주제의 다양한 특징이 또한, 별도로 또는 임의의 적합한 하위 조합으로 제공될 수 있다. 본 개시 내용에 관한 실시양태의 모든 조합은 본원에 개시된 주제에 의해 구체적으로 수용되며, 마치 각각의 모든 조합이 개별적으로 및 명시적으로 개시된 것처럼 본원에 개시된다. 또한, 다양한 실시양태 및 그의 요소의 모든 하위 조합이 또한, 본 개시내용에 구체적으로 수용되며, 이러한 하위 조합 각각의 모두가 본원에 개별적으로 및 명시적으로 개시된 것처럼 본원에 개시된다.

본원에서 논의된 간행물은 본 출원의 출원일 이전에 그의 개시를 위한 목적으로만 제공된다. 본원의 어떠한 것도, 본원에 개시된 주제가 이러한 간행물에 선행할 자격이 없다는 것을 인정하는 것으로 해석되지 않아야 한다. 추가로, 제공된 공개 일자란 실제 공개 일자와 상이할 수 있으며, 이는 독립적으로 확인될 필요가 있을 수 있다.

**발명을 실시하기 위한 구체적인 내용**

[0032] 멀티카피 단일 가닥 DNA (msDNA)의 생산을 증진시키기 위해 변형된, 조작된 레트론이 제공된다. 또한, 이러한 조작된 레트론을 코딩하는 벡터 시스템, 및 조작된 레트론과 이를 코딩하는 벡터 시스템을 다양한 적용, 예컨대 CRISPR/Cas 매개 게놈 편집, 재조합공학, 세포성 마코딩 및 분자 기록에 사용하는 방법이 또한 제공된다.

[0033] 조작된 레트론

[0034] 본 개시내용은 세포에서 msDNA의 생산을 증진시키기 위해 변형되는 조작된 레트론을 제공한다. 조작된 레트론은 프리-msr 서열, 멀티카피 단일 가닥 RNA (msRNA)를 코딩하는 *msr* 유전자; 멀티카피 단일 가닥 DNA (msDNA)를 코딩하는 *msd* 유전자; 포스트-msd 서열 및 역전사효소를 코딩하는 *ret* 유전자를 포함한다. 레트론 코딩된 역전사효소 결과에 의한 DNA의 합성은 *msr* 유전자에 의해 코딩된 단일 가닥 RNA와 연결된 *msd* 유전자에 의해 코딩된 단일 가닥 DNA로 구성되는 DNA/RNA 키메라 산물을 제공할 수 있다. 레트론 msr RNA는 스템 루프 구조의 말단에 보존된 구아노신 잔기를 함유한다. msr RNA의 가닥은 이러한 보존된 구아노신 잔기의 2' 위치에서 2'-5' 포스포디에스테르 연결에 의해 msd 단일 가닥 DNA의 5' 말단에 연결된다.

[0035] 조작된 레트론에서, 포스트-msd 서열은, 예를 들어, 자기 상보적 영역 (프리-msr 서열에 대한 서열 상보성을 갖는다) 내에서 변형되며, 여기서 자기 상보적 영역의 길이는 천연 레트론의 상응하는 영역에 비해 길어진다. 이러한 변형은 msDNA의 증진된 생산을 제공하는 조작된 레트론을 초래한다. 특정 실시양태에서, 상보적 영역은 야생형 자기 상보적 영역보다 적어도 1개, 적어도 2개, 적어도 4개, 적어도 6개, 적어도 8개, 적어도 10개, 적어도 12개, 적어도 14개, 적어도 16개, 적어도 18개, 적어도 20개, 적어도 30개, 적어도 40개 또는 적어도 50개 뉴클레오티드 더 긴 길이를 갖는다. 예를 들어, 자기 상보적 영역은 천연 또는 야생형 상보적 영역보다 1 내지 50개 뉴클레오티드 범위의 더 긴 길이를 가질 수 있으며, 이는 이러한 범위 내의 임의의 길이, 예컨대 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 또는 50개 뉴클레오티드 더 긴 길이를 포함한다. 특정 실시양태에서, 자기 상보적 영역은 야생형 상보적 영역보다 1 내지 16개 뉴클레오티드 범위의 더 긴 길이를 갖는다. 조작된 레트론에 의해 생성된 단일 가닥 DNA는 다양한 적용에 사용될 수 있다.

[0036] 보다 풍부한 ssDNA를 생성하기 위해, 예를 들어, 천연 자기 상보적 3' 및 5' 말단 부분이 볼드체로 강조되어 있는 (위치 1-12 및 158-169), 하기에 제시된 ncRNA 서열식별번호: 12 서열은 위치 1 및 169에서 확장될 수 있다.

```

1  TGCGCACCCCT TAGCGAGAGG TTTATCATT AAGTCAACCT
41 CTGGATGTTG TTTCCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTTT CCTTGTGGGA ACGGAGAGCA TCGCCTGATG
161 CTCTCCGAGC CAACCAGGAA ACCCGTTTTT TCTGACGTAA
201 GGGTGC CGCA
    
```

[0037]

[0038]

예를 들어, 하기 조작된 "ncRNA 확장된" (서열식별번호: 13)에 대해 하기에 제시된 바와 같이, 자기 상보적 영역을 확장하는 부가의 뉴클레오티드는 밑줄과 함께 이탤릭체로 표시된다.

```

1  TGATAAGATT CCGTATGCGC ACCCTTAGCG AGAGGTTTAT
41 CATTAAAGGTC AACCTCTGGA TGTTGTTTCG GCATCCTGCA
81 TTGAATCTGA GTTACTGTCT GTTTTCCTTG TTGGAACGGA
121 GAGCATCGCC TGATGCTCTC CGAGCCAACC AGGAAACCCG
161 TTTTTTCTGA CGTAAGGGTG CGCA TACGGA ATCTTATCA
    
```

[0039]

[0040]

일부 경우에, 부가의 뉴클레오티드가 자기 상보적 영역 내의 임의의 위치에 부가될 수 있으며, 예를 들어, 서열 식별번호: 12 서열의 위치 1-12 및 158-169 내의 어느 곳에도 부가될 수 있다.

[0041]

특정 실시양태에서, 조작된 레트론에 사용된 *msr* 유전자, *msd* 유전자 및 *ret* 유전자의 서열은 박테리아 레트론 오페론으로부터 유래될 수 있다. 대표적인 레트론, 예컨대 미소박테리아 레트론 예컨대 미소코쿠스 크산투스 (*Myxococcus xanthus*) 레트론 (예를 들어, Mx65, Mx162) 및 스티그마텔라 아우란티아카 (*Stigmatella aurantiaca*) 레트론 (예를 들어, Sa163); 에스케리키아 콜라이 레트론 (예를 들어, Ec48, E67, Ec73, Ec78, EC83, EC86, EC107, 및 Ec107); 살모넬라 엔테리카; 비브리오 콜레라에 레트론 (예를 들어, Vc81, Vc95, Vc137); 비브리오 파라하에몰리티쿠스 (*Vibrio parahaemolyticus*) (예를 들어, Vc96); 및 난노시스티스 엑세덴스 (*Nannocystis exedens*) 레트론 (예를 들어, Ne144)을 포함하나 이에 제한되지 않는, 그람 음성 박테리아로부터의 것이 이용 가능하다. 레트론 역전사효소 단백질 서열뿐만 아니라 레트론 *msr* 유전자, *msd* 유전자, *ret* 유전자 핵산 서열은 임의의 공급원으로부터 유래될 수 있다. *msr* 유전자, *msd* 유전자, 및 *ret* 유전자 핵산 서열을 포함한 대표적인 레트론 서열 및 역전사효소 단백질 서열은 미국 국립 생명 공학 정보 센터 (NCBI) 데이터베이스에 등재되어 있다. 예를 들어, NCBI 항목: 수탁 번호 EF428983, M55249, EU250030, X60206, X62583, AB299445, AB436696, AB436695, M86352, M30609, M24392, AF427793, AQ3354, 및 AB079134를 참조하고; (본 출원의 출원일까지 입력된 바와 같은) 모든 서열은 그 전체 내용이 본원에 참조로 포함된다. 이러한 레트론 서열 또는 특정 서열을 포함하는 그의 변이체 중 임의의 것은 변이체 뉴클레오티드, 부가된 뉴클레오티드 또는 더 적은 뉴클레오티드를 포함할 수 있다. 예를 들어, 레트론은 본원에 기재된 레트론 서열 중 임의의 것 (수탁 번호로서 규정된 것을 포함함)과 적어도 약 80-100%의 서열 동일성을 가질 수 있으며, 이는 이러한 범위 내의 임의의 동일성 퍼센트, 예컨대 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 또는 99% 서열 동일성을 포함하며, 본원에 기재된 바와 같은 조작된 레트론 또는 조작된 레트론을 포함하는 벡터 시스템을 구축하는 데 사용될 수 있다.

[0042]

일부 실시양태에서, 재조합 레트론 구축물은 *msr* 유전자, *msd* 유전자 및 *ret* 유전자 사이에 비-천연 간격을 갖는 비-천연 구성을 갖는다. *msr* 유전자와 *msd* 유전자는 내인성 시스 배열로 제공되기보다는 트랜스 배열로 분리될 수 있다. 또한, *ret* 유전자는 *msr* 유전자 또는 *msd* 유전자에 대해 트랜스 배열로 제공될 수 있다. 일부 실시양태에서, *ret* 유전자는 역전사효소에 대한 잠재적 정지 신호를 제거하는 트랜스 배열로 제공되며, 이는 조작된 레트론 구축물로부터 더 긴 단일 가닥 DNA를 생성할 수 있게 한다.

[0043]

일부 실시양태에서, 레트론 구축물은 관심 이중 서열을 포함하기 위해 천연 레트론에 대해 변형된다. 이러한 맥락에서, 레트론은 다양한 적용에서 사용할 수 있도록 이중 서열로 조작될 수 있다. 예를 들어, 하기에 추가로 논의되는 바와 같이, 이중 서열은 관심 단백질 또는 조절 RNA를 코딩하는 핵산, 예를 들어 상동성 지정 복구 (HDR) 또는 재조합 매개 유전자 조작 (재조합공학)에 의해 유전자 편집에 사용하기 적합한 공여자 폴리뉴클레오티드, 또는 분자 기록에 사용하기 위한 CRISPR 프로토타입 DNA 서열을 세포에 제공하기 위해 레트론 구축물에 부가될 수 있다. 이러한 이중 서열은, 예를 들어, *msr* 유전자 또는 *msd* 유전자 내로 삽입되어, msDNA 산물의 일부로서 레트론 역전사효소에 의해 전사되도록 한다.

[0044]

일부 경우에, 관심 이중 서열은 *msd* 스템 루프의 루프 내로 삽입될 수 있다.

[0045]

예를 들어, 조작된 레트론은 다중화를 용이하게 하기 위해 고유한 바코드를 포함할 수 있다. 바코드는 바코드와 연합된 핵산 또는 세포를 확인하기 위해 사용되는 하나 이상의 뉴클레오티드 서열을 포함할 수 있다. 이러

한 바코드는, 예를 들어 msd 코딩된 DNA의 루프 영역 내로 삽입될 수 있다. 바코드는 3-1000개 또는 그 초과 뉴클레오티드 길이, 바람직하게 10-250개 뉴클레오티드 길이, 보다 바람직하게 10-30개 뉴클레오티드 길이일 수 있으며, 이는 이들 범위 내의 임의의 길이, 예컨대 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 40, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 또는 1000개 뉴클레오티드 길이를 포함한다. 일부 실시양태에서, 바코드는 또한 레트론이 유래된 세포, 콜로니 또는 샘플의 위치 (즉, 위치 바코드), 예컨대 세포 어레이에서의 콜로니의 위치, 다중 웰 플레이트에서의 웰의 위치, 랙에서의 튜브의 위치, 또는 실험실 내의 샘플의 위치를 확인하기 위해 사용된다. 특히, 바코드는 레트론을 함유하는 유전자 변형된 세포의 위치를 확인하기 위해 사용될 수 있다. 바코드를 사용하면, 특정한 레트론이 유래된 콜로니로 이를 여전히 다시 추적할 수 있으면서도, 상이한 세포로부터의 레트론을, 시퀀싱을 위해 단일 반응 혼합물에 풀링할 수 있다.

[0046] 또한, 어댑터 서열을 레트론 구축물에 추가하여 높은 처리량 증폭 또는 시퀀싱을 용이하게 할 수 있다. 예를 들어, 한 쌍의 어댑터 서열을 레트론 구축물의 5' 및 3' 말단에 추가하여 동일한 프라이머 세트에 의해 여러 레트론 구축물의 증폭 또는 시퀀싱을 동시에 수행할 수 있다. 레트론 구축물의 증폭은, 예를 들어 세포의 형질감염 또는 벡터로의 라이게이션 전에 수행될 수 있다. 폴리머라제 연쇄 반응 (PCR), 등온 증폭, 핵산 서열 기반 증폭 (NASBA), 전사 매개 증폭 (TMA), 가닥 변위 증폭 (SDA), 및 라이가제 연쇄 반응 (LCR)을 포함하나 이에 제한되지 않는, 레트론 구축물을 증폭시키는 임의의 방법이 사용될 수 있다. 한 실시양태에서, 레트론 구축물은 공통의 5' 및 3' 프라이밍 부위를 포함하여 일련의 범용 프라이머와 병렬로 레트론 서열을 증폭시킬 수 있다. 또 다른 실시양태에서, 일련의 선택적 프라이머는 풀링된 혼합물로부터 레트론 서열의 서브세트를 선택적으로 증폭시키기 위해 사용된다.

[0047] 조작된 레트론을 세포로 전달하는 것은 일반적으로 벡터를 사용하거나 사용하지 않고 달성된다. 조작된 레트론 (또는 이를 함유하는 벡터)은 원핵생물, 진핵생물, 또는 아키온 유기체, 예를 들어 박테리아, 고세균, 진균, 원생생물, 식물 (예를 들어, 외떡잎식물 및 쌍떡잎식물); 및 동물 (예를 들어, 척추동물 및 무척추동물)로부터 임의의 세포를 포함한, 임의의 유형의 세포 내로 도입될 수 있다. 조작된 레트론으로 형질감염될 수 있는 동물의 예는 척추동물, 예컨대 어류, 조류, 포유류 (예를 들어, 인간 및 비-인간 영장류, 농장 동물, 애완동물, 및 실험실 동물), 파충류, 및 양서류를 포함하나 이에 제한되지 않는다. 조작된 레트론으로 형질감염될 수 있는 식물의 예는 곡물, 예컨대 밀, 귀리 및 쌀, 콩과 식물 예컨대 대두 및 완두, 옥수수, 풀, 예컨대 알팔파, 및 목화를 포함한 농작물을 포함하나 이에 제한되지 않는다. 조작된 레트론은 관심 단일 세포 또는 세포 집단 내로 도입될 수 있다. 재조합 세포, 유전자 변형된 세포, 시험관내에서 배양된 세포주로부터의 세포, 및 인공 세포 (예를 들어, 나노입자, 리포솜, 폴리머솜, 또는 핵산을 캡슐화하는 마이크로캡슐)뿐만 아니라 조직, 기관 및 생검으로부터의 세포는 모두 조작된 레트론으로 형질감염될 수 있다. 본 발명의 방법은 또한 세포 단편, 세포 성분 또는 세포 소기관 (예를 들어, 동물 및 식물 세포에서의 미토콘드리아, 식물 세포에서의 색소체 (예를 들어, 엽록체) 및 조류)에도 적용 가능하다. 세포는 조작된 레트론 구축물로 형질감염된 후에 배양되거나 확장될 수 있다.

[0048] 숙주 세포 내로 핵산을 도입하는 방법은 관련 기술분야에 널리 공지되어 있다. 통상적으로 사용되는 방법은 전형적으로 2가 양이온 (예를 들어,  $CaCl_2$ )을 사용하는 화학적으로 유도된 변환, 텍스트란 매개 형질감염, 폴리브렌 매개 형질감염, 리포펙타민 및 LT-1 매개 형질감염, 전기천공법, 원형질 융합, 리포솜에서의 핵산의 캡슐화, 및 조작된 레트론을 포함하는 핵산을 세포핵에 직접 미세 주입하는 것을 포함한다. 예를 들어, 문헌 [Sambrook et al. (2001) Molecular Cloning, a laboratory manual, 3rd edition, Cold Spring Harbor Laboratories, New York, Davis et al. (1995) Basic Methods in Molecular Biology, 2nd edition, McGraw-Hill, and Chu et al. (1981) Gene 13:197] (그 전체 내용이 본원에 참조로 포함된다)을 참조한다.

[0049] *조작된 레트론을 포함하는 벡터 시스템*

[0050] 특정 실시양태에서, 레트론 *msr* 유전자, *msd* 유전자 및 *ret* 유전자는 세포 내의 벡터로부터 생체내에서 발현된다. "벡터"는 세포의 내부로 관심 핵산을 전달하기 위해 사용될 수 있는 물질의 조성이다. 레트론 *msr* 유전자, *msd* 유전자, 및 *ret* 유전자는 숙주 대상체에서 msDNA를 생산하기 위해 단일 벡터를 가진 세포 내로 도입되거나 다수의 별도의 벡터에 도입될 수 있다. 벡터는 전형적으로 레트론 서열과 작동 가능하게 연결된 제어 요소를 포함하며, 이는 대상체 중에서 msDNA의 생체내 생산을 가능하게 한다. 예를 들어, 레트론 *msr* 유전자, *msd* 유전자 및 *ret* 유전자는 레트론 역전사효소 및 msDNA 산물의 발현을 허용하기 위해 프로모터에 작동 가능하게 연결될 수 있다. 일부 실시양태에서, 원하는 관심 산물을 코딩하는 이중 서열 (예를 들어, 폴리펩티드 또는

조절 RNA를 코딩하는 폴리뉴클레오티드, 유전자 편집을 위한 공여자 폴리뉴클레오티드, 또는 분자 기록을 위한 프로토스페이서 DNA)이 *msr* 유전자 또는 *msd* 유전자에 삽입될 수 있다. 조작된 레트론 서열을 포함하는 벡터로 형질감염될 수 있는 임의의 진핵, 아키온 또는 원핵 세포가 msDNA를 생산하기 위해 사용될 수 있다. 다른 레트론 코딩된 산물과 함께 msDNA를 생산할 수 있는 구축물의 능력은 실험적으로 결정될 수 있다.

[0051] 일부 실시양태에서, 조작된 레트론은 하나 이상의 벡터를 포함하는 벡터 시스템에 의해 생산된다. 벡터 시스템에서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자는 동일한 벡터 (즉, 이러한 모든 레트론 요소의 시스템 배열)에 의해 제공될 수 있으며, 여기서 벡터는 *msr* 유전자 및 *msd* 유전자와 작동 가능하게 연결된 프로모터를 포함한다. 일부 실시양태에서, 프로모터는 *ret* 유전자와 추가로 작동 가능하게 연결되어 있다. 다른 실시양태에서, 벡터는 *ret* 유전자와 작동 가능하게 연결된 제2 프로모터를 추가로 포함한다. 대안적으로, *ret* 유전자는 *msr* 유전자와 *msd* 유전자를 포함하지 않는 제2 벡터에 의해 (즉, *msr*-*msd*와 *ret*의 트랜스 배열) 제공될 수 있다. 또 다른 실시양태에서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자는 각각 상이한 벡터 (즉, 모든 레트론 요소의 트랜스 배열)에 의해 제공된다. 선형 폴리뉴클레오티드, 이온성 또는 양친매성 화합물과 연합된 폴리뉴클레오티드, 플라스미드, 및 바이러스를 포함하나 이에 제한되지 않는 다수의 벡터가 이용 가능하다. 따라서, 용어 "벡터"는 자율적으로 복제하는 플라스미드 또는 바이러스를 포함한다. 바이러스 벡터의 예는 아데노바이러스 벡터, 아데노 관련 바이러스 벡터, 레트로바이러스 벡터, 렌티바이러스 벡터 등을 포함하나 이에 제한되지는 않는다. 발현 구축물은 살아 있는 세포에서 복제될 수 있거나, 또는 합성적으로 만들어질 수 있다. 본 출원의 목적상, 용어 "발현 구축물", "발현 벡터" 및 "벡터"는 본 발명의 적용을 일반적이고 예시적인 의미로 입증하기 위해 상호 교환적으로 사용되며, 발명을 제한하려는 의도는 아니다.

[0052] 특정 실시양태에서, 조작된 레트론 서열을 포함하는 핵산은 프로모터의 전사 제어 하에 있다. "프로모터"는 세포의 합성 기구, 또는 유전자의 특이적 전사를 시작하는 데 필요한 도입된 합성 기구에 의해 인식된 DNA 서열을 지칭한다. 용어 프로모터는 RNA 폴리머라제 I, II 또는 III의 개시 부위 주위에 클러스터된 전사 제어 모듈의 특정 군을 지칭하기 위해 본원에서 사용될 것이다. 포유류 세포 발현을 위한 전형적인 프로모터는 특히, SV40 초기 프로모터, CMV 프로모터, 예컨대 CMV 즉시 초기 프로모터 (그 전체 내용이 본원에 참조로 포함된 미국 특허 번호 5,168,062 및 5,385,839 참조), 마우스 유방 종양 바이러스 LTR 프로모터, 아데노바이러스 주요 후기 프로모터 (Ad MLP), 및 단순 포진 바이러스 프로모터를 포함한다. 다른 비-바이러스 프로모터, 예컨대 뮤린 메탈로티오네인 유전자로부터 유래된 프로모터가 또한 포유류 발현에 사용될 것이다. 이들 프로모터 및 기타 프로모터는 관련 기술분야에 널리 공지된 기술을 사용하여 상업적으로 이용 가능한 플라스미드로부터 획득될 수 있다. 예를 들어, 상기 문헌 [Sambrook et al.]을 참조한다. 인핸서 요소는 구축물의 발현 수준을 증가시키기 위해 프로모터와 연합하여 사용될 수 있다. 그 예는 문헌 [Dijkema et al., *EMBO J.* (1985) 4:761]에 기재된 바와 같은 SV40 초기 유전자 인핸서, 문헌 [Gorman et al., *Proc. Natl. Acad. Sci. USA* (1982b) 79:6777]에 기재된 바와 같은, 라우스 육종 바이러스의 긴 말단 반복부 (LTR)로부터 유래된 인핸서/프로모터, 및 문헌 [Boshart et al., *Cell* (1985) 41:521]에 기재된 바와 같은, 인간 CMV로부터 유래된 요소, 예컨대 CMV 인트론 A 서열에 포함된 요소를 포함한다.

[0053] 한 실시양태에서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자를 포함한 조작된 레트론을 발현하기 위한 발현 벡터는 *msr* 유전자, *msd* 유전자 및 *ret* 유전자를 코딩하는 폴리뉴클레오티드와 "작동 가능하게 연결된" 프로모터를 포함한다. 본원에 사용된 바와 같은 "작동 가능하게 연결된" 또는 "전사 제어 하에"라는 문구는 프로모터가 RNA 폴리머라제에 의한 전사의 개시 및 *msr* 유전자, *msd* 유전자 및 *ret* 유전자의 발현을 제어하기 위해 폴리뉴클레오티드와 관련하여 올바른 위치와 배향에 있다는 것을 의미한다.

[0054] 전형적으로, 전사 종결인자/폴리아데닐화 신호가 발현 구축물에 또한 존재할 것이다. 이러한 서열의 예는 상기 문헌 [Sambrook et al.]에 기재된 바와 같이 SV40으로부터 유래된 것뿐만 아니라 소의 성장 호르몬 종결인자 서열 (예를 들어, 미국 특허 번호 5,122,458 참조)을 포함하나 이에 제한되지는 않는다. 부가적으로, 5'-UTR 서열은 코딩 서열에 인접하여 배치되어 그의 발현을 증진시킬 수 있다. 이러한 서열은 내부 리보솜 진입 부위 (IRES)를 포함하는 UTR을 포함할 수 있다.

[0055] IRES의 봉입은 벡터로부터의 하나 이상의 오픈 리딩 프레임의 번역을 허용한다. IRES 요소는 진핵 리보솜 번역 개시 복합체를 끌어당기고 번역 개시를 촉진한다. 예를 들어, 문헌 [Kaufman et al., *Nuc. Acids Res.* (1991) 19:4485-4490; Gurtu et al., *Biochem. Biophys. Res. Comm.* (1996) 229:295-298; Rees et al., *BioTechniques* (1996) 20:102-110; Kobayashi et al., *BioTechniques* (1996) 21:399-402; and Mosser et al., *BioTechniques* (1997) 22:150-161]을 참조한다. 다수의 IRES 서열이 공지되어 있으며, 광범위한 바이러스로부터 유래된 서열, 예컨대 피코르나 바이러스의 리더 서열, 예컨대 뇌심근염 바이러스 (EMCV) UTR (문헌 [Jang et

al. *J. Virol.* (1989) 63:1651-1660]), 소아마비 리더 서열, A형 간염 바이러스 리더, C형 간염 바이러스 IRES, 인간 리노바이러스 유형 2 IRES (문헌 [Dobrikova et al., *Proc. Natl. Acad. Sci.* (2003) 100(25):15125-15130]), 구제역 바이러스로부터의 IRES 요소 (문헌 [Ramesh et al., *Nucl. Acid Res.* (1996) 24:2697-2700]), 지아르디아 바이러스 IRES (문헌 [Garlapati et al., *J. Biol. Chem.* (2004) 279(5):3389-3397]) 등을 포함한다. 다양한 비-바이러스 IRES 서열이 또한 본원에서 사용되며, 이는 효모로부터의 IRES 서열뿐만 아니라 인간 안지오텐신 II 유형 1 수용체 IRES (문헌 [Martin et al., *Mol. Cell Endocrinol.* (2003) 212:51-61]), 섬유모세포 성장 인자 IRES (FGF-1 IRES 및 FGF-2 IRES, 문헌 [Martineau et al. (2004) *Mol. Cell Biol.* 24(17):7622-7635]), 혈관 내피 성장 인자 IRES (문헌 [Baranick et al. (2008) *Proc. Natl. Acad. Sci. U.S.A.* 105(12):4733-4738, Stein et al. (1998) *Mol. Cell Biol.* 18(6):3112-3119, Bert et al. (2006) *RNA* 12(6):1074-1083]) 및 인슐린 유사 성장 인자 2 IRES (문헌 [Pedersen et al. (2002) *Biochem. J.* 363(Pt 1):37-44])를 포함하나 이에 제한되지는 않는다. 이들 요소는, 예를 들어 클론테크 (Clontech; 미국 캘리포니아주 마운틴 뷰), 인비보젠 (Invivogen; 미국 캘리포니아주 샌디에이고), 애드진 (Addgene; 미국 메사추세츠주 캠프리지), 및 진코포이아 (GeneCopoeia; 미국 메릴랜드주 록빌)에서 판매되는 플라스미드에서 쉽게 상업적으로 입수 가능하다. 또한 [IRESite: The database of experimentally verified IRES structures (iresite.org)]을 참조한다. IRES 서열은, 예를 들어, 재조합공학을 위한 다수의 박테리오파지 재조합 단백질 또는 발현 카세트로부터의 레트론 역전사효소와 조합하여 HDR을 위한 RNA-가이드된 뉴클레아제 (예를 들어, Cas9)를 발현하기 위해 벡터에 포함될 수 있다.

[0056] 대안적으로, 바이러스 T2A 펩티드를 코딩하는 폴리뉴클레오티드를 사용하여 단일 벡터로부터 다수의 단백질 산물 (예를 들어, Cas9, 박테리오파지 재조합 단백질, 레트론 역전사효소)을 생산할 수 있다. 하나 이상의 2A 링커 펩티드가 멀티시스트론성 구축물 내의 코딩 서열 사이에 삽입될 수 있다. 자기 절단성인 2A 펩티드는 멀티시스트론성 구축물로부터 공동 발현된 단백질이 등물 수준으로 생산될 수 있도록 한다. 다양한 바이러스로부터의 2A 펩티드가 사용될 수 있으며, 이는 구제역 바이러스, 말 비염 A 바이러스, 토세아 아스그나(*Thosea asigna*) 바이러스 및 돼지 테스코바이러스-1로부터 유래된 2A 펩티드를 포함하나 이에 제한되지는 않는다. 예를 들어, 문헌 [Kim et al. (2011) *PLoS One* 6(4):e18556, Trichas et al. (2008) *BMC Biol.* 6:40, Provost et al. (2007) *Genesis* 45(10):625-629, Furler et al. (2001) *Gene Ther.* 8(11):864-873] (이의 전체 내용이 본원에 참조로 포함된다)을 참조한다.

[0057] 특정 실시양태에서, 발현 구축물은 박테리아 숙주를 형질전환시키는 데 적합한 플라스미드를 포함한다. 다수의 박테리아 발현 벡터가 관련 기술분야의 통상의 기술자에게 공지되어 있으며, 적절한 벡터의 선택은 선택의 문제이다. 박테리아 발현 벡터는 pACYC177, pASK75, pBAD, pBADM, pBAT, pCal, pET, pETM, pGAT, pGEX, pHAT, pKK223, pMal, pProEx, pQE, 및 pZA31을 포함하나 이에 제한되지는 않는다. 박테리아 플라스미드는 항생제 선택 마커 (예를 들어, 암피실린, 카나마이신, 에리트로마이신, 카르베니실린, 스트렙토마이신 또는 테트라사이클린 내성), lacZ 유전자 ( $\beta$ -갈락토시다제가 x-gal 기질로부터 청색 색소를 생산한다), 형광 마커 (예를 들어, GFP, mCherry) 또는 형질전환된 박테리아의 선택을 위한 다른 마커를 함유할 수 있다. 예를 들어, 상기 문헌 [Sambrook et al.]을 참조한다.

[0058] 다른 실시양태에서, 발현 구축물은 효모 세포를 형질전환시키는 데 적합한 플라스미드를 포함한다. 효모 발현 플라스미드는 전형적으로 효모 특이적 복제 기점 (ORI) 및 영양 선택 마커 (예를 들어, HIS3, URA3, LYS2, LEU2, TRP1, MET15, ura4+, leu1+, ade6+), 항생제 선택 마커 (예를 들어, 카나마이신 내성), 형광 마커 (예를 들어, mCherry) 또는 형질전환된 효모 세포의 선택을 위한 다른 마커를 함유한다. 효모 플라스미드는 박테리아 숙주 (예를 들어, 이. 콜라이)와 효모 세포 간의 서플링을 가능하게 하는 성분을 추가로 함유할 수 있다. ORI가 결여되어 있으며 상동 재조합에 의해 숙주 염색체에 통합되는 효모 통합 플라스미드 (YIp); 자율 복제 서열 (ARS)을 함유하고 독립적으로 복제할 수 있는 효모 복제 플라스미드 (YRp); ARS의 일부와 중심체 서열 (CEN)의 일부를 함유하는 낮은 카피 벡터인 효모 중심체 플라스미드 (YCp); 및 세포당 50개 이상의 카피가 안정적으로 전파될 수 있는 2 마이크로미터 서클 (자연 효모 플라스미드)로부터의 단편을 포함하는 높은 카피 수 플라스미드인 효모 에피솜 플라스미드 (YEpl)를 포함할 수 많은 상이한 유형의 효모 플라스미드가 이용 가능하다.

[0059] 다른 실시양태에서, 발현 구축물은 바이러스 또는 바이러스 게놈으로부터 유래된 조작된 구축물을 포함한다. 포유류 세포로의 유전자 전이를 위해 수많은 바이러스 기반 시스템이 개발되었다. 이는 아데노바이러스, 레트로바이러스 ( $\gamma$ -레트로바이러스 및 렌티바이러스), 폭스바이러스, 아데노 관련 바이러스, 바칼로바이러스, 및 단순 포진 바이러스를 포함한다 (예를 들어, 문헌 [Warnock et al. (2011) *Methods Mol. Biol.* 737:1-25; Walther et al. (2000) *Drugs* 60(2):249-271; and Lundstrom (2003) *Trends Biotechnol.* 21(3):117-122]

참조; 그 전체 내용이 본원에 참조로 포함된다). 특정 바이러스가 수용체 매개 세포내이입을 통해 세포에 진입하여 숙주 세포 게놈에 통합될 수 있고 바이러스 유전자를 안정적이고 효율적으로 발현할 수 있는 능력은 외래 유전자를 포유류 세포로 전이시키는 데 있어 매력적인 후보로 만들었다.

[0060] 예를 들어, 레트로바이러스는 유전자 전달 시스템을 위한 편리한 플랫폼을 제공한다. 선택된 서열은 벡터에 삽입되고 관련 기술분야에 공지된 기술을 사용하여 레트로바이러스 입자로 패키징될 수 있다. 이어서, 재조합 바이러스는 단리되어 생체내에서 또는 생체외에서 대상체의 세포로 전달될 수 있다. 다수의 레트로바이러스 시스템이 보고되었다 (미국 특허 번호 5,219,740; 문헌 [Miller and Rosman (1989) *BioTechniques* 7:980-990; Miller, A. D. (1990) *Human Gene Therapy* 1:5-14; Scarpa et al. (1991) *Virology* 180:849-852; Burns et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:8033-8037; Boris-Lawrie and Temin (1993) *Cur. Opin. Genet. Develop.* 3:102-109; and Ferry et al. (2011) *Curr. Pharm. Des.* 17(24):2516-2527]). 렌티바이러스는 레트로바이러스의 한 클래스로, 포유류 세포에 폴리뉴클레오티드를 전달하는데 특히 유용한데, 이는 이들이 분열 세포와 비-분열 세포 모두를 감염시킬 수 있기 때문이다 (예를 들어, 문헌 [Lois et al. (2002) *Science* 295:868-872; Durand et al. (2011) *Viruses* 3(2):132-159] 참조; 본원에 참조로 포함된다).

[0061] 수많은 아데노바이러스 벡터가 또한 보고되었다. 숙주 게놈에 통합되는 레트로바이러스와 달리, 아데노바이러스는 염색체 외에서 지속되므로 삽입 돌연변이 유발과 관련된 위험을 최소화한다 (Haj-Ahmad and Graham, *J. Virol.* (1986) 57:267-274; Bett et al., *J. Virol.* (1993) 67:5911-5921; Mittereder et al., *Human Gene Therapy* (1994) 5:717-729; Seth et al., *J. Virol.* (1994) 68:933-940; Barr et al., *Gene Therapy* (1994) 1:51-58; Berkner, K. L. *BioTechniques* (1988) 6:616-629; and Rich et al., *Human Gene Therapy* (1993) 4:461-476). 부가적으로, 다양한 아데노 관련 바이러스 (AAV) 벡터 시스템이 유전자 전달을 위해 개발되었다. AAV 벡터는 관련 기술분야에 널리 공지된 기술을 사용하여 쉽게 구축될 수 있다. 예를 들어, 미국 특허 번호 5,173,414 및 5,139,941; 국제 공개 번호 WO 92/01070 (1992년 1월 23일에 공개됨) 및 WO 93/03769 (1993년 3월 4일에 공개됨); 문헌 [Lebkowski et al., *Molec. Cell. Biol.* (1988) 8:3988-3996; Vincent et al., *Vaccines* 90 (1990) (Cold Spring Harbor Laboratory Press); Carter, B. J. *Current Opinion in Biotechnology* (1992) 3:533-539; Muzyczka, N. *Current Topics in Microbiol. and Immunol.* (1992) 158:97-129; Kotin, R. M. *Human Gene Therapy* (1994) 5:793-801; Shelling and Smith, *Gene Therapy* (1994) 1:165-169; and Zhou et al., *J. Exp. Med.* (1994) 179:1867-1875]을 참조한다.

[0062] 조작된 레트론을 코딩하는 핵산을 전달하는 데 유용한 또 다른 벡터 시스템은 문헌 [Small, Jr., P. A., et al. (미국 특허 번호 5,676,950; 1997년 10월 14일에 허여됨); 본원에 참조로 포함됨]에 기재된 장용으로 투여된 재조합 폭스바이러스 백신이다.

[0063] 관심 핵산 분자를 전달하기 위해 사용되는 부가의 바이러스 벡터는 백시니아 바이러스 및 조류 폭스바이러스를 포함한 폭스 패밀리의 바이러스로부터 유래된 것을 포함한다. 한 예를 들자면, 관심 핵산 분자 (예를 들어, 조작된 레트론)를 발현하는 백시니아 바이러스 재조합체는 하기와 같이 구축될 수 있다. 특정한 핵산 서열을 코딩하는 DNA는 먼저 적절한 벡터 내로 삽입되어, 백시니아 프로모터 및 플랭킹 백시니아 DNA 서열, 예컨대 티미딘 카나제 (TK)를 코딩하는 서열에 인접하도록 한다. 이어서, 이러한 벡터는 백시니아에 동시에 감염된 세포를 형질감염시키는 데 사용된다. 상동 재조합은 백시니아 프로모터 플러스 관심 서열을 코딩하는 유전자를 바이러스 게놈 내로 삽입시키는 역할을 한다. 그 결과로 생성된 TK-재조합체는 5-브로모데옥시우리딘의 존재 하에서 세포를 배양하고 이에 내성이 있는 바이러스 플라크를 골라냄으로써 선택될 수 있다.

[0064] 대안적으로, 아비폭스바이러스, 예컨대 조류폭스 및 카나리폭스 바이러스를 또한 사용하여 관심 핵산 분자를 전달할 수 있다. 아비폭스 속의 구성원은 감수성 조류 종에서만 대량으로 복제될 수 있고 따라서 포유류 세포에서 감염되지 않기 때문에 아비폭스 벡터의 사용이 인간과 다른 포유류 종에 특히 바람직하다. 재조합 아비폭스 바이러스를 생산하는 방법은 관련 기술분야에 공지되어 있으며, 백시니아 바이러스의 생산과 관련하여 상기 기재된 바와 같이 유전적 재조합을 이용한다. 예를 들어, WO 91/12882; WO 89/03429; 및 WO 92/03545를 참조한다.

[0065] 분자 공역 벡터, 예컨대 문헌 [Michael et al., *J. Biol. Chem.* (1993) 268:6866-6869 and Wagner et al., *Proc. Natl. Acad. Sci. USA* (1992) 89:6099-6103]에 기재된 아데노바이러스 키메라 벡터가 또한 유전자 전달에 사용될 수 있다.

[0066] 예컨대 신드비스 바이러스 (SIN), 쉼리키 삼립열 바이러스 (SFV), 및 베네수엘라 말 뇌염 바이러스 (VEE)이지만, 이에 제한되지 않는 알파바이러스 속의 구성원으로부터 유래된 벡터가 또한, 본 발명의 폴리뉴클

레오티드를 전달하기 위한 바이러스 벡터로서 사용될 것이다. 즉각적인 방법의 실행에 유용한 신드비스 바이러스 유래 벡터에 대한 설명은 문헌 [Dubensky et al. (1996) J. Virol. 70:508-519; 및 국제 공개 번호 WO 95/07995, WO 96/17072; 및 Dubensky, Jr., T. W., et al., 미국 특허 번호 5,843,723 (1998년 12월 1일에 허여됨), 및 Dubensky, Jr., T. W., 미국 특허 번호 5,789,245 (1998년 8월 4일에 허여됨)] (모두 본원에 참조로 포함됨)을 참조한다. 특히 선호되는 것은 신드비스 바이러스와 베네수엘라 말 너염 바이러스로부터 유래된 서열로 구성된 키메라 알파바이러스 벡터이다. 예를 들어, 문헌 [Perri et al. (2003) J. Virol. 77: 10394-10403] 및 국제 공개 번호 WO 02/099035, WO 02/080982, WO 01/81609, 및 WO 00/61772] (그 전체 내용이 본원에 참조로 포함됨)을 참조한다.

[0067] 백시니아 기반 감염/형질감염 시스템은 숙주 세포에서 관심 핵산 (예를 들어, 조작된 레트론)의 유도 가능한 일시적인 발현을 제공하기 위해 편리하게 사용될 수 있다. 이러한 시스템에서, 세포는 박테리오파지 T7 RNA 폴리머라제를 코딩하는 백시니아 바이러스 재조합체에 의해 시험관내에서 먼저 감염된다. 이러한 폴리머라제는 T7 프로모터를 보유하는 템플릿만 전사한다는 점에서 정교한 특이성을 보여준다. 감염 후, 세포는 T7 프로모터에 의해 구동되는 관심 핵산으로 형질감염된다. 백시니아 바이러스 재조합체로부터 세포질에서 발현된 폴리머라제는 형질감염된 DNA를 RNA로 전사한다. 이러한 방법은 다량의 RNA의 높은 수준의 일시적인 세포질 생산을 제공한다. 예를 들어, 문헌 [Elroy-Stein and Moss, Proc. Natl. Acad. Sci. USA (1990) 87:6743-6747; Fuerst et al., Proc. Natl. Acad. Sci. USA (1986) 83:8122-8126]을 참조한다.

[0068] 백시니아 또는 아비폭스 바이러스 재조합체에 의한 감염, 또는 다른 바이러스 벡터를 사용한 핵산의 전달에 대한 대안적 접근 방식으로서, 숙주 세포 내로의 도입 후 높은 수준의 발현을 유도하는 증폭 시스템을 사용할 수 있다. 구체적으로, T7 RNA 폴리머라제에 대한 코딩 영역에 선행하는 T7 RNA 폴리머라제 프로모터가 조작될 수 있다. 이러한 템플릿으로부터 유래된 RNA의 번역은 더 많은 템플릿을 전사할 것인 T7 RNA 폴리머라제를 생성할 것이다. 동반하여, 그의 발현이 T7 프로모터의 제어 하에 있는 cDNA가 있을 것이다. 따라서, 증폭 템플릿 RNA의 번역으로부터 생성된 T7 RNA 폴리머라제의 일부는 원하는 유전자의 전사로 이어질 것이다. 증폭을 시작하기 위해서는 일부 T7 RNA 폴리머라제가 필요하기 때문에, T7 RNA 폴리머라제는 전사 반응을 프라이밍하기 위해 그 템플릿(들)과 함께 세포 내로 도입될 수 있다. 폴리머라제는 단백질로서 도입되거나 또는 RNA 폴리머라제를 코딩하는 플라스미드 상에 도입될 수 있다. T7 시스템 및 세포를 형질전환시키기 위한 T7 시스템의 용도에 대한 추가의 논의는, 예를 들어 국제 공개 번호 WO 94/26911; 문헌 [Studier and Moffatt, J. Mol. Biol. (1986) 189:113-130; Deng and Wolff, Gene (1994) 143:245-249; Gao et al., Biochem. Biophys. Res. Commun. (1994) 200:1201-1206; Gao and Huang, Nuc. Acids Res. (1993) 21:2867-2872; Chen et al., Nuc. Acids Res. (1994) 22:2114-2120]; 및 미국 특허 번호 5,135,855를 참조한다.

[0069] 곤충 세포 발현 시스템, 예컨대 바큘로바이러스 시스템이 또한 사용될 수 있으며, 관련 기술분야의 통상의 기술자에게 공지되어 있으며, 예를 들어 문헌 [*Baculovirus and Insect Cell Expression Protocols* (Methods in Molecular Biology, D.W. Murhammer ed., Humana Press, 2nd edition, 2007) and L. King *The Baculovirus Expression System: A laboratory guide* (Springer, 1992)]에 기재되어 있다. 바큘로바이러스/곤충 세포 발현 시스템을 위한 물질 및 방법은 특히 써모 피셔 사이언티픽 (Thermo Fisher Scientific; 미국 매사추세츠주 윌섬) 및 클론테크 (미국 캘리포니아주 마운틴 뷰)로부터 키트 형태로 상업적으로 이용 가능하다.

[0070] 식물 발현 시스템이 또한 식물 세포를 형질전환시키는 데 사용될 수 있다. 일반적으로, 이러한 시스템은 바이러스 기반 벡터를 사용하여 식물 세포를 이중 유전자로 형질감염시킨다. 이러한 시스템에 대한 설명은, 예를 들어 문헌 [Porta et al., *Mol. Biotech.* (1996) 5:209-221; and Hackland et al., *Arch. Virol.* (1994) 139:1-22]을 참조한다.

[0071] 조작된 레트론 구축물의 발현을 수행하기 하기 위해, 발현 구축물은 세포 내로 전달되어야만 한다. 이러한 전달은 세포주를 형질전환시키기 위한 실험실 절차에서와 같이 시험관내에서 수행될 수 있거나, 또는 특정 질환 상태의 치료에서와 같이 생체내 또는 생체외에서 수행될 수 있다. 전달의 한 가지 메커니즘은 바이러스 감염을 통한 것이며, 여기서는 발현 구축물이 감염성 바이러스 입자에 캡슐화된다.

[0072] 배양된 세포 내로의 발현 구축물의 전이를 위한 몇 가지 비-바이러스 방법이 또한 고려된다. 이들 방법은 인산 칼슘 침전, DEAE-텍스트란, 전기천공, 직접 미세주입, DNA-부하된 리포솜, 리포펙타민-DNA 복합체, 세포 초음파 처리, 고속 미세발사체를 사용한 유전자 폭격, 및 수용체 매개 형질감염의 사용을 포함한다 (예를 들어, 문헌 [Graham and Van Der Eb (1973) *Virology* 52:456-467; Chen and Okayama (1987) *Mol. Cell Biol.* 7:2745-2752; Rippe et al. (1990) *Mol. Cell Biol.* 10:689-695; Gopal (1985) *Mol. Cell Biol.* 5:1188-1190; Tur-

Kaspa et al. (1986) *Mol. Cell. Biol.* 6:716-718; Potter et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:7161-7165); Harland and Weintraub (1985) *J. Cell Biol.* 101:1094-1099); Nicolau & Sene (1982) *Biochim. Biophys. Acta* 721:185-190; Fraley et al. (1979) *Proc. Natl. Acad. Sci. USA* 76:3348-3352; Fehheimer et al. (1987) *Proc Natl. Acad. Sci. USA* 84:8463-8467; Yang et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:9568-9572; Wu and Wu (1987) *J. Biol. Chem.* 262:4429-4432; Wu and Wu (1988) *Biochemistry* 27:887-892] 참조; 본원에 참조로 포함됨). 이러한 기술 중 일부는 생체내 또는 생체의 사용을 위해 성공적으로 적용될 수 있다.

[0073] 일단 발현 구축물이 세포 내로 전달되면, 조작된 레트론 서열을 포함하는 핵산은 상이한 부위에 위치하고 발현될 수 있다. 특정 실시양태에서, 조작된 레트론 서열을 포함하는 핵산은 세포의 계놈에 안정적으로 통합될 수 있다. 이러한 통합은 상동 재조합을 통해 동쪽 위치와 배향에 있을 수 있거나 (유전자 대체), 또는 무작위적이고 비-특이적인 위치에 통합될 수 있다 (유전자 증대). 또 다른 실시양태에서, 핵산은 DNA의 별도의 에피솜 세그먼트로서 세포 내에서 안정적으로 유지될 수 있다. 이러한 핵산 세그먼트 또는 "에피솜"은 숙주 세포 주기와 독립적으로 또는 숙주 세포 주기와 동기화하여 유지 및 복제를 허용하기에 충분한 서열을 코딩한다. 발현 구축물이 세포로 전달되는 방법과 세포에서 핵산이 남아 있는 위치는 이용된 발현 구축물의 유형에 따라 달라진다.

[0074] 또 다른 실시양태에서, 발현 구축물은 단순히 조작된 레트론을 포함하는 네이키드 재조합 DNA 또는 플라스미드로 이루어질 수 있다. 상기 구축물의 전달은 세포막에 물리적으로 또는 화학적으로 투과될 수 있는 전술한 방법 중 임의의 것에 의해 수행될 수 있다. 이는 시험관내 전달에 특히 적용 가능하지만, 생체내 사용에도 적용될 수 있다. 문헌 [Dubensky et al. (*Proc. Natl. Acad. Sci. USA* (1984) 81:7529-7533)]에서는 활성 바이러스 복제와 급성 감염을 명확하게 보여주는 성체 및 신생 마우스의 간 및 비장에 인산칼슘 침전물의 형태로 폴리오마바이러스 DNA를 성공적으로 주사하였다. 문헌 [Benvenisty & Neshif (*Proc. Natl. Acad. Sci. USA* (1986) 83:9551-9555)]에서는 또한, 인산칼슘 침전된 플라스미드의 복강 내 직접 주사가 형질감염된 유전자의 발현을 초래한다는 것을 입증해주었다. 조작된 관심 레트론을 코딩하는 DNA가 또한 생체내에서 유사한 방식으로 전달될 수 있고 레트론 산물을 발현할 수 있는 것으로 고려된다.

[0075] 또 다른 실시양태에서, 네이키드 DNA 발현 구축물은 입자 폭격에 의해 세포로 전달될 수 있다. 이러한 방법은 DNA 코딩된 미세발사체를 사멸시키지 않고서도 세포막을 뚫고 세포 안으로 들어갈 수 있도록 빠른 속도로 상기 미세발사체를 가속시킬 수 있는 능력에 좌우된다 (Klein et al. (1987) *Nature* 327:70-73). 작은 입자를 가속시키는 몇 가지 장치가 개발되었다. 이러한 장치 중 하나는 고전압 방전에 의존하여 전류를 발생시키고, 그 결과 원동력을 제공한다 (Yang et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:9568-9572). 미세발사체는 생물학적으로 불활성인 물질, 예컨대 텅스텐 또는 금 비드로 이루어질 수 있다.

[0076] 추가 실시양태에서, 발현 구축물은 리포솜을 사용하여 전달될 수 있다. 리포솜은 인지질 이중층 막과 내부 수성 매질을 특징으로 하는 소포체 구조이다. 다층 리포솜은 수성 매질에 의해 분리된 여러 지질 층을 갖는다. 이는 인지질이 과량의 수용액에 현탁될 때 자발적으로 형성된다. 지질 성분은 닫힌 구조의 형성 전에 자기 재배열을 진행하고 지질 이중층 사이에 용해된 용질과 물을 포획한다 (Ghosh & Bachhawat (1991) *Liver Diseases, Targeted Diagnosis and Therapy Using Specific Receptors and Ligands*, Wu et al. (Eds.), Marcel Dekker, NY, 87-104). 또한 리포펙타민-DNA 복합체의 사용이 고려된다.

[0077] 특정 실시양태에서, 리포솜은 혈구응집 바이러스 (HVJ)와 복합체를 형성할 수 있다. 이는 세포막과의 융합을 용이하게 하고 리포솜 캡슐화된 DNA의 세포 진입을 촉진시키는 것으로 나타났다 (Kaneda et al. (1989) *Science* 243:375-378). 다른 실시양태에서, 리포솜은 핵 비-히스톤 염색체 단백질 (HMG-I)과 복합체를 형성하거나 또는 이와 연계해서 이용될 수 있다 (Kato et al. (1991) *J. Biol. Chem.* 266(6):3361-3364). 또한 추가 실시양태에서, 리포솜은 HVJ 및 HMG-I 둘 다와 복합체를 형성하거나 또는 이들과 연계해서 이용될 수 있다. 이러한 발현 구축물이 시험관내 및 생체내에서 핵산의 전이 및 발현에 성공적으로 이용되었다는 점에서, 이들을 본 발명에 적용할 수 있다. DNA 구축물에 박테리아 프로모터가 이용되는 경우, 리포솜 내에 적절한 박테리아 폴리머라제를 포함하는 것이 또한 바람직할 것이다.

[0078] 핵산을 세포에 전달하기 위해 이용될 수 있는 다른 발현 구축물은 수용체 매개 전달 비히클이다. 이들은 거의 모든 진핵 세포에서 수용체 매개 세포내이입에 의한 거대분자의 선택적 흡수를 이용한다. 다양한 수용체의 세포 유형 특이적 분포로 인해, 전달은 매우 특이적일 수 있다 (Wu and Wu (1993) *Adv. Drug Delivery Rev.* 12:159-167).

[0079] 수용체 매개 유전자 표적화 비히클은 일반적으로 2가지 구성 요소: 세포 수용체 특이적 리간드와 DNA 결합체로

이루어진다. 수용체 매개 유전자 전이를 위해 여러 리간드가 사용되어 왔다. 가장 광범위하게 특징 규명된 리간드는 아시알로오로소뮤코이드 (ASOR) 및 트랜스페린이다 (예를 들어, 문헌 [Wu and Wu (1987), 상기 참조; Wagner et al. (1990) Proc. Natl. Acad. Sci. USA 87(9):3410-3414] 참조). ASOR과 동일한 수용체를 인식하는 합성 신당단백질이 유전자 전달 비히클로서 사용되고 있으며 (문헌 [Ferkol et al. (1993) FASEB J. 7:1081-1091; Perales et al. (1994) Proc. Natl. Acad. Sci. USA 91(9):4086-4090]), 표피 성장 인자 (EGF) 가 또한 편평 암종 세포에 유전자를 전달하는 데 사용되어 왔다 (Myers, EPO 0273085).

[0080] 다른 실시양태에서, 전달 비히클은 리간드 및 리포솜을 포함할 수 있다. 예를 들어, 문헌 [Nicolau et al. (Methods Enzymol. (1987) 149:157-176)]에서는 갈락토스 말단 아시알로글리코시드인 락토실-세라미드를 이용하여 리포솜에 통합하고, 간세포에 의한 인슐린 유전자의 흡수가 증가한다는 것을 관찰하였다. 따라서, 특정한 유전자를 코딩하는 핵산은 또한 리포솜의 유무와 관계없이 임의의 수의 수용체-리간드 시스템에 의해 세포로 특이적으로 전달될 수 있다. 또한, 세포 상의 표면 항원에 대한 항체는 유사하게 표적화 모이어티로서 사용될 수 있다.

[0081] 특정한 예에서, 조작된 레트론을 포함하는 재조합 폴리뉴클레오티드는 양이온성 지질과 조합하여 투여될 수 있다. 양이온성 지질의 예는 리포펙틴, DOTMA, DOPE 및 DOTAP를 포함하나, 이에 제한되지는 않는다. 구체적으로 참조로 포함된 공보 WO/0071096에는 유전자 요법에 효과적으로 사용될 수 있는 상이한 제형, 예컨대 DOTAP:콜레스테롤 또는 콜레스테롤 유도체 제형이 기재되어 있다. 다른 개시에는 또한 나노입자를 포함한 상이한 지질 또는 리포솜 제형, 및 투여 방법에 대해 논의되어 있으며; 이들 개시는 미국 특허 공보 20030203865, 20020150626, 20030032615, 및 20040048787을 포함하나 이에 제한되지는 않으며, 핵산의 투여 및 전달에 관한 제형 및 기타 관련 측면을 개시하는 정도로 구체적으로 참조로 포함된다. 입자를 형성하는 데 사용되는 방법 또한 미국 특허 번호 5,844,107, 5,877,302, 6,008,336, 6,077,835, 5,972,901, 6,200,801, 및 5,972,900에 개시되어 있으며, 이들 측면에 대해 참조로 포함된다.

[0082] 특정 실시양태에서, 유전자 전이는 생체의 조건 하에 더 쉽게 수행될 수 있다. 생체의 유전자 요법은 대상체로부터 세포를 단리하고, 핵산을 시험관내에서 세포로 전달한 다음, 변형된 세포를 대상체 내로 다시 돌려보내는 것을 지칭한다. 이것은 대상체로부터의 세포를 포함하는 생물학적 샘플의 수집을 포함할 수 있다. 예를 들어, 혈액은 정맥천자에 의해 취득될 수 있으며, 고체 조직 샘플은 관련 기술분야에 널리 공지된 방법에 따라 외과적 기술에 의해 취득될 수 있다.

[0083] 항상 그렇지는 않지만, 통상적으로 세포를 받는 대상체 (즉, 수용자)는 세포가 수거되거나 취득되는 대상체이기도 하며, 이는 공여된 세포가 자가 세포라는 이점을 제공한다. 그러나, 세포는 또 다른 대상체 (즉, 공여자), 공여자로부터의 세포의 배양물, 또는 확립된 세포주 배양물로부터 취득될 수 있다. 세포는 치료받을 대상체와 동일하거나 상이한 종으로부터 취득될 수 있으나, 바람직하게 동일한 종의 것이며, 보다 바람직하게 대상체와 동일한 면역학적 프로필의 것이다. 이러한 세포는, 예를 들어, 가까운 친척 또는 일치하는 공여자로부터의 세포를 포함하는 생물학적 샘플로부터 취득될 수 있으며, 그 후 핵산 (예를 들어, 조작된 레트론을 포함함)으로 형질감염되고, 예를 들어 질환 또는 병태의 치료를 위하여, 게놈 변형을 필요로 하는 대상체에게 투여된다.

[0084] *키트*

[0085] 본원에 기재된 바와 같은 조작된 레트론 구축물을 포함하는 키트가 또한 제공된다. 일부 실시양태에서, 키트는 조작된 레트론 구축물, 또는 이러한 레트론 구축물을 포함하는 벡터 시스템을 제공한다. 일부 실시양태에서, 키트에 포함된 조작된 레트론 구축물은 관심 단백질 또는 조절 RNA를 코딩하는 핵산, 세포성 바코드, 예를 들어 상동성 지정 복구 (HDR) 또는 재조합 매개 유전자 조작 (재조합공학)에 의해 유전자 편집에 사용하기 적합한 공여자 폴리뉴클레오티드, 또는 분자 기록에 사용하기 위한 CRISPR 프로토스페이스 DNA 서열을 세포에 제공할 수 있는 이중 서열을 포함한다. 상기 키트에는 다른 작용제, 예컨대 형질감염제, 숙주 세포, 세포를 배양하는 데 적합한 배지, 완충제 등이 포함될 수 있다.

[0086] 키트의 맥락에서, 작용제는 임의의 편리한 패키징 (예를 들어, 스틱 팩, 용량 팩 등)으로 액체 또는 고체 형태로 제공될 수 있다. 키트의 작용제는 동일한 또는 별도의 용기에 존재할 수 있다. 작용제는 또한 동일한 용기에 존재할 수 있다. 상기 구성 요소 외에도, 대상 키트는 대상 방법을 실행하기 위한 (특정 실시양태에서) 지침을 추가로 포함할 수 있다. 이러한 지침은 다양한 형태로 대상 키트에 존재할 수 있으며, 그 중 하나 이상이 키트에 존재할 수 있다. 이러한 지침이 존재할 수 있는 한 가지 형태는 정보가 인쇄된 적절한 매질 또는 기질, 예를 들어, 키트의 패키징 내에, 패키지 삽입물 내에 그 정보가 인쇄된 하나 또는 다수의 종이이다. 이러한 지침의 또 다른 형태는 컴퓨터로 관독 가능한 매체, 예를 들어 정보가 기록된 디스켓, 콤팩트 디스크 (CD), 플래

시 드라이브 등이다. 이러한 지침이 존재할 수 있는 또 다른 형태는 원격 부위의 정보에 접근하기 위해 인터넷을 통해 사용될 수 있는 웹사이트 주소이다.

- [0087] 활용
- [0088] 레트론은 다양한 적용에서 사용할 수 있도록 이중 서열로 조작될 수 있다. 예를 들어, 하기에 추가로 논의되는 바와 같이, 이중 서열은 관심 단백질 또는 조절 RNA를 코딩하는 이중 핵산, 세포성 바코드, 예를 들어, 상동성 지정 복구 (HDR) 또는 재조합 매개 유전자 조작 (재조합공학)에 의해 유전자 편집에 사용하기 적합한 공여자 폴리뉴클레오티드, 또는 분자 기록에 사용하기 위한 CRISPR 프로토스페이서 DNA 서열을 세포에 제공하기 위해 레트론 구축물에 부가될 수 있다. 이러한 이중 서열은, 예를 들어, msr 유전자 또는 msd 유전자 내로 삽입되어, 이중 서열이 msDNA 산물의 일부로서 레트론 역전사효소에 의해 전사되도록 한다.
- [0089] 단백질 또는 RNA의 생산
- [0090] 예를 들어, 조작된 레트론에 의해 생성된 단일 가닥 DNA는 세포에서 원하는 관심 산물을 생산하는 데 사용될 수 있다. 일부 실시양태에서, 레트론은 세포에서 생성된 레트론 msDNA로부터 폴리펩티드를 생산할 수 있도록 관심 폴리펩티드를 코딩하는 이중 서열로 조작된다. 관심 폴리펩티드는 효소, 세포의 매트릭스 단백질, 수용체, 수송체, 이온 채널 또는 기타 막 단백질, 호르몬, 신경펩티드, 항체 또는 세포골격 단백질; 또는 그의 단편, 또는 생물학적으로 활성인 관심 도메인을 포함하나 이에 제한되지 않는 임의의 유형의 단백질/펩티드일 수 있다. 일부 실시양태에서, 단백질은 질환의 치료에 사용하기 위한 치료용 단백질 또는 치료용 항체이다.
- [0091] 다른 실시양태에서, 레트론은 세포에서 레트론으로부터 RNA를 생산할 수 있도록 관심 RNA를 코딩하는 이중 서열로 조작된다. 관심 RNA는 RNA 간섭 (RNAi) 핵산 또는 조절 RNA, 예컨대 마이크로RNA (miRNA), 작은 간섭 RNA (siRNA), 짧은 헤어핀 RNA (shRNA), 작은 핵 RNA (snRNA), 긴 비-코딩 RNA (lncRNA), 안티센스 핵산 등을 포함하나 이에 제한되지 않는 임의의 유형의 RNA일 수 있다.
- [0092] 유전자 편집
- [0093] 일부 실시양태에서, 레트론은 CRISPR/Cas 게놈 편집 시스템과 함께 사용하기에 적합한 공여자 폴리뉴클레오티드를 코딩하는 이중 서열로 조작된다. 공여자 폴리뉴클레오티드는 이러한 공여자 폴리뉴클레오티드가 세포에서 편집될 표적 유전자좌를 표적으로 하는 데 책임이 있는 한 쌍의 상동성 아암에 의해 플랭킹된 의도된 게놈 편집을 포함하는 서열을 포함한다. 공여자 폴리뉴클레오티드는 전형적으로 5' 게놈 표적 서열과 혼성화하는 5' 상동성 아암과 3' 게놈 표적 서열과 혼성화하는 3' 상동성 아암을 포함한다. 상동성 아암은 공여자 폴리뉴클레오티드 내에서 의도된 편집을 포함하는 뉴클레오티드 서열에 대한 상동성 아암의 상대적 위치와 관련된 5' 및 3' (즉, 상류 및 하류) 상동성 아암으로서 본원에서 지칭된다. 5' 및 3' 상동성 아암은 각각 "5' 표적 서열" 및 "3' 표적 서열"로서 본원에서 지칭되는, 변형될 게놈 DNA의 표적 유전자좌 내의 영역과 혼성화한다.
- [0094] 상동성 아암은 표적 유전자좌에서 공여자 폴리뉴클레오티드와 게놈 DNA 간의 상동 재조합을 매개하기 위해 표적 서열과의 혼성화에 대해 충분히 상보적이어야만 한다. 예를 들어, 상동성 아암은 상응하는 게놈 표적 서열과의 적어도 약 80-100% 서열 동일성을 갖는 뉴클레오티드 서열을 포함할 수 있으며, 서열 동일성은 상기 범위 내의 임의의 동일성 퍼센트, 예컨대 적어도 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 포함하며, 여기서 의도된 편집을 포함하는 뉴클레오티드 서열은 5' 및 3' 상동성 아암에 의해 인식된 (즉, 혼성화에 대해 충분한 상보성을 갖는) 게놈 표적 유전자좌에서 HDR에 의해 게놈 DNA에 통합될 수 있다.
- [0095] 특정 실시양태에서, 게놈 표적 서열 (즉, "5' 표적 서열" 및 "3' 표적 서열") 내의 상응하는 상동 뉴클레오티드 서열은 절단을 위한 특이적 부위 및/또는 의도된 편집을 도입하기 위한 특이적 부위에 플랭킹한다. 특이적 절단 부위와 상동 뉴클레오티드 서열 간의 거리 (예를 들어, 각각의 상동성 아암)는 수백 개의 뉴클레오티드일 수 있다. 일부 실시양태에서, 상동성 아암과 절단 부위 간의 거리는 200개 이하의 뉴클레오티드 (예를 들어, 0, 10, 20, 30, 50, 75, 100, 125, 150, 175, 및 200개의 뉴클레오티드)이다. 대부분의 경우, 거리가 더 작을수록 유전자 표적화 비율이 더 높아질 수 있다. 바람직한 실시양태에서, 공여자 폴리뉴클레오티드는 특이적 절단 부위 및 변경될 게놈 표적 서열의 일부분을 모두 포괄하는 게놈의 일부분에 도입될 서열 변화를 제외하고 전체 길이 전반에 걸쳐 표적 게놈 서열과 실질적으로 동일하다.
- [0096] 상동성 아암은 임의의 길이, 예를 들어 10개 이상의 뉴클레오티드, 15개 이상의 뉴클레오티드, 20개 이상의 뉴클레오티드, 50개 이상의 뉴클레오티드, 100개 이상의 뉴클레오티드, 250개 이상의 뉴클레오티드, 300개 이상의 뉴클레오티드, 350개 이상의 뉴클레오티드, 400개 이상의 뉴클레오티드, 450개 이상의 뉴클레오티드, 500개 이

상의 뉴클레오티드, 1000개 이상의 뉴클레오티드 (1 kb), 5000개 이상의 뉴클레오티드 (5 kb), 10000개 이상의 뉴클레오티드 (10 kb) 등의 것일 수 있다. 일부 경우에, 5' 상동성 아암과 3' 상동성 아암은 길이가 서로 실질적으로 동일하다. 그러나, 일부 경우에 5' 상동성 아암과 3' 상동성 아암은 길이가 서로 반드시 동일하지 않다. 예를 들어, 하나의 상동성 아암은 다른 상동성 아암보다 30% 이하 더 짧거나, 다른 상동성 아암보다 20% 이하 더 짧거나, 다른 상동성 아암보다 10% 이하 더 짧거나, 다른 상동성 아암보다 5% 이하 더 짧거나, 다른 상동성 아암보다 2% 이하 더 짧거나, 또는 다른 상동성 아암보다 단지 몇 개의 뉴클레오티드 만큼만 더 작을 수 있다. 다른 경우에, 5' 상동성 아암과 3' 상동성 아암은 길이가 서로 실질적으로 상이할 수 있으며, 예를 들어, 하나의 상동성 아암이 다른 상동성 아암보다 40% 이상 더 짧을 수 있거나, 50% 이상 더 짧을 수 있거나, 종종 60% 이상 더 짧을 수 있거나, 70% 이상 더 짧을 수 있거나, 80% 이상 더 짧을 수 있거나, 90% 이상 더 짧을 수 있거나, 또는 95% 이상 더 짧을 수 있다.

[0097] 공여자 폴리뉴클레오티드는 가이드 RNA에 의해 특정한 게놈 서열 (즉, 변형될 게놈 표적 서열)을 표적으로 하는 RNA-가이드된 뉴클레아제와 조합하여 사용된다. 표적 특이적 가이드 RNA는 게놈 표적 서열과 상보적인 뉴클레오티드 서열을 포함하며, 이에 따라 표적 부위에서 혼성화에 의해 뉴클레아제-gRNA 복합체의 결합을 매개한다. 예를 들어, gRNA는 뉴클레아제-gRNA 복합체가 돌연변이의 부위를 표적으로 하도록 마이너 대립 유전자의 서열과 상보적인 서열로 설계될 수 있다. 돌연변이는 삽입, 결실 또는 치환을 포함할 수 있다. 예를 들어, 돌연변이는 단일 뉴클레오티드 변이, 유전자 융합, 전위, 반전, 복제, 프레임 쉬프트, 미스센스, 넌센스 또는 관심 표현형 또는 질환과 관련된 다른 돌연변이를 포함할 수 있다. 표적화된 마이너 대립 유전자는 흔한 유전적 변이체 또는 희귀한 유전적 변이체일 수 있다. 특정 실시양태에서, gRNA는 단일 염기 쌍 구별을 가진 마이너 대립 유전자와 선택적으로 결합하도록 설계되는데, 예를 들어, 뉴클레아제-gRNA 복합체가 단일 뉴클레오티드 다형성 (SNP)과 결합할 수 있도록 설계된다. 특히, gRNA는 유전자로부터 돌연변이를 제거하기 위해 게놈 편집을 목적으로 질환과 관련된 관심 돌연변이를 표적으로 설계될 수 있다. 대안적으로, gRNA는 메이저 또는 야생형 대립 유전자의 서열과 상보적인 서열이 게놈 편집을 목적으로 대립 유전자에 대한 뉴클레아제-gRNA 복합체를 표적으로 하도록 설계하여 세포의 게놈 DNA 내의 유전자 내로 돌연변이, 예컨대 삽입, 결실 또는 치환을 도입할 수 있다. 이러한 유전자 변형된 세포는, 예를 들어 표현형을 변경하거나, 새로운 특성을 부여하거나, 또는 약물 스크리닝을 위한 질환 모델을 생산하는 데 사용될 수 있다.

[0098] 특정 실시양태에서, 게놈 변형에 사용되는 RNA-가이드된 뉴클레아제는 클러스터된 규칙적으로 공간을 둔 짧은 팔린드롬성 반복부 (CRISPR) 시스템 Cas 뉴클레아제이다. CRISPR 시스템 유형 I, 유형 II 또는 유형 III Cas 뉴클레아제를 포함하여, HDR 메커니즘에 의해 공여자 폴리뉴클레오티드의 통합을 허용하기 위해 DNA의 부위 지정 절단을 촉매할 수 있는 임의의 RNA-가이드된 Cas 뉴클레아제가 게놈 편집에 사용될 수 있다. Cas 단백질의 예는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5e (CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9 (Csn1 또는 Csx12), Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (CasA), Cse2 (CasB), Cse3 (CasE), Cse4 (CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, 및 Cu1966, 및 그의 상동체 또는 변형된 버전을 포함한다.

[0099] 특정 실시양태에서, 유형 II CRISPR 시스템 Cas9 엔도뉴클레아제가 사용된다. 임의의 종으로부터의 Cas9 뉴클레아제, 또는 Cas9 엔도뉴클레아제 활성을 보유하는 (즉, 이중 가닥 파손을 생성하기 위해 DNA의 부위 지정 절단을 촉매하는) 그의 생물학적으로 활성인 단편, 변이체, 유사체 또는 유도체가 본원에 기재된 바와 같이 게놈 변형을 수행하는 데 사용될 수 있다. Cas9는 유기체로부터 물리적으로 유래될 필요는 없지만, 합성적으로 또는 제조함으로써 생산될 수 있다. 다수의 박테리아 종으로부터의 Cas9 서열은 관련 기술분야에 널리 공지되어 있으며, 미국 국립 생명 공학 정보 센터 (NCBI) 데이터베이스에 등재되어 있다. 예를 들어, 하기로부터의 Cas9에 대한 NCBI 항목을 참조한다: 스트렙토코쿠스 피오게네스 (WP\_002989955, WP\_038434062, WP\_011528583); 캄필로박터 제주니(*Campylobacter jejuni*) (WP\_022552435, YP\_002344900), 캄필로박터 콜라이(*Campylobacter coli*) (WP\_060786116); 캄필로박터 페투스(*Campylobacter fetus*) (WP\_059434633); 코리네박테리움 울세란스(*Corynebacterium ulcerans*) (NC\_015683, NC\_017317); 코리네박테리움 디프테리아(*Corynebacterium diphtheria*) (NC\_016782, NC\_016786); 엔테로코쿠스 파에칼리스(*Enterococcus faecalis*) (WP\_033919308); 스피로플라스마 시르피디콜라(*Spiroplasma syrophidicola*) (NC\_021284); 프레보텔라 인터메디아(*Prevotella intermedia*) (NC\_017861); 스피로플라스마 타이완센세(*Spiroplasma taiwanense*) (NC\_021846); 스트렙토코쿠스 이니아에(*Streptococcus iniae*) (NC\_021314); 벨리엘라 발티카(*Belliella baltica*) (NC\_018010); 사이크로플렉수스 토르퀴스(*Psychroflexus torquis*) (NC\_018721); 스트렙토코쿠스 써모필루스(*Streptococcus thermophilus*) (YP\_820832), 스트렙토코쿠스 뮤탄스(*Streptococcus mutans*) (WP\_061046374, WP\_024786433); 리스테리아 이노

쿠아(*Listeria innocua*) (NP\_472073); 리스테리아 모노시토크네스(*Listeria monocytogenes*) (WP\_061665472); 레지오넬라 뉴모필라(*Legionella pneumophila*) (WP\_062726656); 스타필로코쿠스 아우레우스(*Staphylococcus aureus*) (WP\_001573634); 프란시셀라 툴라렌시스(*Francisella tularensis*) (WP\_032729892, WP\_014548420), 엔테로코쿠스 파에칼리스 (WP\_033919308); 락토바실루스 람노수스(*Lactobacillus rhamnosus*) (WP\_048482595, WP\_032965177); 및 네이세리아 메닝기티디스(*Neisseria meningitidis*) (WP\_061704949, YP\_002342100); 이들 모두의 서열 (본 출원의 출원일로서 도입됨)은 그 전체 내용이 본원에 참조로 포함된다. 이들 서열 중 임의의 것, 또는 이들 서열과 적어도 약 70-100% 서열 동일성, 예를 들어 이러한 범위 내의 임의의 동일성 퍼센트, 예컨대 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 또는 99% 서열 동일성을 갖는 서열을 포함하는 그의 변이체가, 본원에 기재된 바와 같이 게놈 편집에 사용될 수 있다. Cas9의 유전적 다양성 및 계통발생학적 분석에 대한 서열 비교와 논의를 위해서, 문헌 [Fonfara et al. (2014) *Nucleic Acids Res.* 42(4):2577-90; Kapitonov et al. (2015) *J. Bacteriol.* 198(5):797-807, Shmakov et al. (2015) *Mol. Cell.* 60(3):385-397, and Chylinski et al. (2014) *Nucleic Acids Res.* 42(10):6091-6105]을 또한 참조한다.

[0100] CRISPR-Cas 시스템은 박테리아 및 고세균에서 자연적으로 발생하며, 여기서 외래 DNA에 대항하여 RNA 매개 적응 면역의 역할을 한다. 박테리아 유형 II CRISPR 시스템은 엔도뉴클레아제인 Cas9를 사용하는데, 이는 상보적인 게놈 표적 서열과 특이적으로 혼성화하는 가이드 RNA (gRNA)와 복합체를 형성하며, 여기서 Cas9 엔도뉴클레아제는 이중 가닥 파손을 생산하기 위한 절단을 촉매한다. Cas9의 표적화는 전형적으로 gRNA 결합 부위 또는 그 근처에 있는 DNA에 5' 프로토스페이서-인접 모티프 (PAM)의 존재에 추가로 의존한다.

[0101] 게놈 표적 부위는 전형적으로 gRNA와 상보적인 뉴클레오티드 서열을 포함하며, 프로토스페이서 인접 모티프 (PAM)를 추가로 포함할 수 있다. 특정 실시양태에서, 표적 부위는 3개의 염기 쌍 PAM 외에 20-30개의 염기 쌍을 포함한다. 전형적으로, PAM의 제1 뉴클레오티드는 임의의 뉴클레오티드일 수 있는 반면, 다른 두 뉴클레오티드는 선택되는 특이적 Cas9 단백질에 의존될 것이다. 예시적인 PAM 서열은 관련 기술분야의 통상의 기술자에게 공지되어 있으며, NNG, NGN, NAG 및 NGG를 포함하나 이에 제한되지 않으며, 여기서 N은 임의의 뉴클레오티드를 나타낸다. 특정 실시양태에서, gRNA에 의해 표적화된 대립 유전자는 대립 유전자 내에 PAM을 생성하는 돌연변이를 포함하며, 여기서 PAM은 대립 유전자에 대한 Cas9-gRNA 복합체의 결합을 촉진시킨다.

[0102] 특정 실시양태에서, gRNA는 5-50개 뉴클레오티드, 10-30개 뉴클레오티드, 15-25개 뉴클레오티드, 18-22개 뉴클레오티드, 또는 19-21개 뉴클레오티드 길이어거나, 또는 언급된 범위 사이의 임의의 길이, 예를 들어 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 또는 35개 뉴클레오티드 길이일 수 있다. 가이드 RNA는 단일 RNA 분자에 crRNA와 tracrRNA 서열을 포함하는 단일 가이드 RNA일 수 있거나, 또는 가이드 RNA는 별도의 RNA 분자에 존재하는 crRNA와 tracrRNA 서열을 갖는 2개의 RNA 분자를 포함할 수 있다.

[0103] 또 다른 실시양태에서, 프레보텔라(*Prevotella*) 및 프란시셀라(*Francisella*) 1 (Cpf1)로부터의 CRISPR 뉴클레아제가 사용된다. Cpf1은 Cas9와 유사한 또 다른 클래스 II CRISPR/Cas 시스템 RNA-가이드된 뉴클레아제이며, 유사하게 사용될 수 있다. Cas9와 달리, Cpf1은 tracrRNA를 필요로 하지 않으며 그의 가이드 RNA 내의 crRNA에만 의존하므로, 표적화를 위해 Cas9보다 더 짧은 가이드 RNA를 Cpf1과 함께 사용할 수 있다는 장점을 제공한다. Cpf1은 DNA 또는 RNA를 절단할 수 있다. Cpf1에 의해 인식된 PAM 부위는 Cas9에 의해 인식된 G-풍부 PAM 부위와 대조적으로, 서열 5'-YTN-3' (여기서 "Y"는 퍼리미딘이고 "N"은 임의의 핵염기이다), 또는 5'-TTN-3'을 갖는다. DNA의 Cpf1 절단은 4개 또는 5개의 뉴클레오티드 오버행을 갖는 점착성 말단을 가진 이중 가닥 파손물을 생산한다. Cpf1의 논의를 위해, 예를 들어 문헌 [Ledford et al. (2015) *Nature.* 526 (7571):17-17, Zetsche et al. (2015) *Cell.* 163 (3):759-771, Murovec et al. (2017) *Plant Biotechnol. J.* 15(8):917-926, Zhang et al. (2017) *Front. Plant Sci.* 8:177, Fernandes et al. (2016) *Postepy Biochem.* 62(3):315-326] (본원에 참조로 포함됨)을 참조한다.

[0104] C2c1은 사용될 수 있는 또 다른 클래스 II CRISPR/Cas 시스템 RNA-가이드된 뉴클레아제이다. C2c1은 Cas9와 유사하게, 표적 부위에 대한 가이드를 위해 crRNA와 tracrRNA 모두에 의존한다. C2c1에 대한 설명을 위해, 예를 들어, 문헌 [Shmakov et al. (2015) *Mol Cell.* 60(3):385-397, Zhang et al. (2017) *Front Plant Sci.* 8:177] (본원에 참조로 포함됨)을 참조한다.

[0105] 또 다른 실시양태에서, 조각된 RNA-가이드된 FokI 뉴클레아제가 사용될 수 있다. RNA-가이드된 FokI 뉴클레아제는 불활성 Cas9 (dCas9)와 FokI 엔도뉴클레아제 (FokI-dCas9)의 융합을 포함하며, 여기서 dCas9 부분은 FokI

에 대한 가이드 RNA 의존적 표적화를 부여한다. 조작된 RNA-가이드된 FokI 뉴클레아제의 설명에 대해서는, 예를 들어 문헌 [Havlicek et al. (2017) Mol. Ther. 25(2):342-355, Pan et al. (2016) Sci Rep. 6:35794, Tsai et al. (2014) Nat Biotechnol. 32(6):569-576] (본원에 참조로 포함됨)을 참조한다.

[0106] RNA-가이드된 뉴클레아제는 단백질의 형태로 제공될 수 있으며, 임의로 여기서 뉴클레아제는 gRNA와 복합체를 형성하거나, 또는 RNA-가이드된 뉴클레아제를 코딩하는 핵산, 예컨대 RNA (예를 들어, 메신저 RNA) 또는 DNA (발현 벡터)에 의해 제공된다. 일부 실시양태에서, RNA-가이드된 뉴클레아제와 gRNA는 모두 벡터에 의해 제공된다. 둘 다는 단일 벡터에 의해 발현되거나 또는 상이한 벡터 상에 별도로 발현될 수 있다. RNA-가이드된 뉴클레아제 및 gRNA를 코딩하는 벡터(들)는 조작된 레트론 *msr* 유전자, *msd* 유전자 및 *ret* 유전자 서열을 포함하는 벡터 시스템에 포함될 수 있다.

[0107] 코돈 사용 빈도는 특정한 세포 또는 유기체에서 RNA-가이드된 뉴클레아제 및/또는 레트론 역전사효소의 생산을 개선하기 위해 최적화될 수 있다. 예를 들어, RNA-가이드된 뉴클레아제 또는 역전사효소를 코딩하는 핵산은 자연적으로 발생하는 폴리뉴클레오티드 서열과의 비교 시, 효모 세포, 박테리아 세포, 인간 세포, 비-인간 세포, 포유류 세포, 설치류 세포, 마우스 세포, 래트 세포, 또는 임의의 다른 관심 숙주 세포에서 더 높은 사용 빈도를 갖는 코돈을 대체하도록 변형될 수 있다. RNA-가이드된 뉴클레아제 또는 역전사효소를 코딩하는 핵산이 세포 내로 도입될 때, 단백질은 일시적으로, 조건부로, 또는 구성적으로 세포에서 발현될 수 있다.

[0108] **제조합공학**

[0109] 제조합공학 (제조합 매개 유전자 조작)은, 예를 들어, 유전자 대체, 유전자 녹아웃, 결실, 삽입, 반전 또는 점 돌연변이를 생성하기 위해 세포 내에서 염색체뿐만 아니라 에피솜 레플리콘을 변형시키는 데 사용될 수 있다. 제조합공학은, 예를 들어, 유전자를 클로닝하거나 또는 마커 또는 태그를 삽입하기 위해 플라스미드 또는 박테리아 인공 염색체 (BAC)를 변형시키는 데 사용될 수 있다. 본원에 기재된 조작된 레트론은 제조합을 위한 선형 단일 가닥 또는 이중 가닥 DNA를 제공하기 위해 제조합공학 적용에 사용될 수 있다. 상동 제조합은 박테리오파지 단백질, 예컨대 Rac 프로파지로부터의 RecE/RecT 또는 박테리오파지 람다로부터의 Red  $\alpha \beta \delta$  에 의해 매개된다. 선형 DNA는 제조합을 허용하기 위해 세포에 존재하는 표적 DNA 분자 (예를 들어, 플라스미드, BAC 또는 염색체)에 대한 5' 및 3' 말단에 충분한 상동성을 가져야 한다.

[0110] 제조합공학에 사용되는 선형 이중 가닥 또는 단일 가닥 DNA 분자 (즉, 공여자 폴리뉴클레오티드)는 선형 DNA 분자가 상동 제조합을 위한 표적 부위를 표적으로 하는 2개의 상동성 아암에 의해 플랭킹된 삽입되도록 의도된 편집을 가진 서열을 포함한다. 제조합공학을 위한 상동성 아암은 전형적으로 13-300개 뉴클레오티드 또는 20 내지 200개 뉴클레오티드 길이, 예를 들어 이러한 범위 내의 임의의 길이, 예컨대 13, 14, 15, 16, 17, 18, 19, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 또는 200개 뉴클레오티드 길이의 범위이다. 일부 실시양태에서, 상동성 아암은 적어도 15개, 적어도 20개, 적어도 30개, 적어도 40개 또는 적어도 50개 또는 그 초과 뉴클레오티드 길이이다. 길이가 40-50개 뉴클레오티드 범위인 상동성 아암은 일반적으로 제조합을 위한 충분한 표적화 효율을 갖지만; 길이가 150개 내지 200개 또는 그 초과 염기 범위인 더 긴 상동성 아암은 표적화 효율을 추가로 개선시킬 수 있다. 일부 실시양태에서, 5' 상동성 아암과 3' 상동성 아암은 길이가 상이하다. 예를 들어, 선형 DNA는 5' 말단에 약 50개의 염기를 가지며 3' 말단에 약 20개의 염기를 가질 수 있으며, 표적화될 영역에 대한 상동성을 갖는다.

[0111] 박테리오파지 상동 제조합 단백질은 단백질로서 세포에 제공되거나 또는 상기 제조합 단백질을 코딩하는 하나 이상의 벡터에 의해 세포에 제공될 수 있다. 일부 실시양태에서, 박테리오파지 제조합 단백질을 코딩하는 하나 이상의 벡터는 조작된 레트론 *msr* 유전자, *msd* 유전자 및 *ret* 유전자 서열을 포함하는 벡터 시스템에 포함된다.

[0112] 부가적으로, 제조합 단백질 *exo*, *bet*, *gam*을 포함하는 결함 있는  $\lambda$  프로파지를 함유하는 DY380; DY380에서 발견된 제조합 유전자 이외에도, 엄격하게 제어된 아라비노스 유도성 *flpe* 유전자 (*flpe*는 2개의 동일한 *frt* 부위 간의 제조합을 매개한다)를 함유하는, DY380으로부터 유래된 EL250; DY380에서 발견된 제조합 유전자 이외에도, 엄격하게 제어된 아라비노스 유도성 *cre* 유전자 (*cre*는 2개의 동일한 *loxP* 부위 간의 제조합을 매개한다)를 함유하는, DY380으로부터 유래된 EL350; *galK* 양성/음성 선택을 사용하는 BAC 제조합공학을 위해 설계되는, DY380으로부터 유래된 SW102; *galK* 양성/음성 선택을 위해 사용될 수도 있지만, EL250과 마찬가지로 *ara* 유도성 *Flpe* 유전자를 함유하는, EL250으로부터 유래된 SW105; 및 *galK* 양성/음성 선택을 위해 사용될 수 있지만, EL350과 마찬가지로 *ara* 유도성 *Cre* 유전자를 함유하는, EL350으로부터 유래된 SW106을 포함하나 이에 제한되지는 않는, 프로파지 제조합 시스템을 함유하는 다수의 박테리아 균주가 제조합공학에 이용 가능하다. 제조합공학은 그러

한 균주의 박테리아 세포를, 재조합공학에 적합한 선형 DNA를 코딩하는 이중 서열을 포함하는 조작된 레트론으로 형질감염시킴으로써 수행될 수 있다. 재조합공학 시스템과 프로토콜의 논의에 대해서는, 예를 들어 문헌 [Sharan et al. (2009) Nat Protoc. 4(2): 206-223, Zhang et al. (1998) Nature Genetics 20:123-128, Muyrers et al. (1999) Nucleic Acids Res. 27: 1555-1557, Yu et al. (2000) Proc. Natl. Acad. Sci U.S.A. 97 (11):5978-5983] (본원에 참조로 포함됨)을 참조한다.

[0113] 분자 기록

[0114] 일부 실시양태에서, 조작된 레트론 구축물에서의 이중 서열은 분자 기록을 허용하는 합성 CRISPR 프로토스페이서 DNA 서열을 포함한다. 내인성 CRISPR Cas1-Cas2 시스템은 정상적으로, 게놈 기반 어레이 내에서 바이러스 핵산의 침입에 대한 서열 특이적 내성을 부여하는 짧은 서열 (즉, 프로토스페이서)을 저장함으로써 바이러스 감염으로부터 유래한 외래 DNA 서열을 추적하기 위해 박테리아와 고세균에 의해 활용된다. 이러한 어레이는 스페이서 서열을 보존할 뿐만 아니라 서열이 획득되는 순서를 기록하여, 획득 사건의 시간적 기록을 생성한다.

[0115] 이러한 시스템은 조작된 레트론을 사용하여 세포 내로 도입되는 "합성 프로토스페이서"의 형태로 게놈 CRISPR 어레이에 임의의 DNA 서열을 기록하도록 적용될 수 있다. 프로토스페이서 서열을 운반하는 조작된 레트론은 CRISPR 시스템 Cas1-Cas2 복합체를 활용함으로써 특이적 게놈 유전자좌에서 합성 CRISPR 프로토스페이서 서열의 통합을 위해 사용될 수 있다. 분자 기록은 안정적인 유전적 기억 추적 코드를 생산함으로써 특정의 생물학적 사건을 추적하는 데 사용될 수 있다. 예를 들어, 문헌 [Shipman et al. (2016) Science 353(6298):aaf1175] 및 국제 특허 출원 공개 공보 WO/2018/191525 (그 전체 내용이 본원에 참조로 포함됨)를 참조한다.

[0116] 일부 실시양태에서, CRISPR-Cas 시스템은 특이적 및 임의의 DNA 서열을 박테리아 게놈에 기록하기 위해 사용된다. DNA 서열은 세포 내의 조작된 레트론에 의해 생산될 수 있다. 예를 들어, 조작된 레트론은 세포 내의 CRISPR 어레이 내로 삽입되는 프로토스페이서를 세포 내에서 생산하는 데 사용될 수 있다. 세포는, 세포 내에서 하나 이상의 합성 프로토스페이서를 생산할 수 있는 하나 이상의 조작된 레트론 (또는 이를 코딩하는 벡터 시스템)을 포함하도록 변형될 수 있으며, 여기서 합성 프로토스페이서가 CRISPR 어레이에 부가된다. 규정된 서열의 기록은 여러 날에 걸쳐 기록되며 여러 양식으로 생성될 수 있다.

[0117] 일부 실시양태에서, 조작된 레트론은 msd 프로토스페이서 핵산 영역 또는 msr 프로토스페이서 핵산 영역을 포함한다. msr 프로토스페이서 핵산 영역의 경우, 프로토스페이서 서열은 먼저 msr RNA에 통합되고, 이는 프로토스페이서 DNA로 역전사된다. 상보적 서열을 가진 2개의 상보적 프로토스페이서 DNA 서열이 혼성화될 때, 또는 이중 가닥 구조 (예컨대 헤어핀)가 단일 가닥 프로토스페이서 DNA에서 형성될 때 (예를 들어, 단일 msDNA는 이중 가닥 DNA 프로토스페이서를 제공하기 위해 적절한 헤어핀 구조를 형성할 수 있다), 이중 가닥 프로토스페이서 DNA가 생산된다.

[0118] 일부 실시양태에서, 제1 조작된 레트론으로부터 생체내에서 생산된 단일 가닥 DNA는 동일한 레트론 또는 제2 조작된 레트론으로부터 생체내에서 생산된 상보적 단일 가닥 DNA와 혼성화될 수 있거나 또는 헤어핀 구조를 형성할 수 있으며, 이어서 CRISPR 어레이 내로 스페이서 서열로서 삽입될 프로토스페이서 서열로서 사용될 수 있다. 조작된 레트론(들)은 CRISPR 어레이에 통합하기 위해 세포 내에 충분한 수준의 프로토스페이서 서열을 제공해야 한다. 세포 내에서 생성된 프로토스페이서를 사용하면, 생체내 분자 기록 시스템이 사용자에게 공지된 정보만 캡처하는 것에서부터 사용자가 이전에 알지 못했던 생물학적 또는 환경 정보를 캡처하는 데까지 확장된다. 예를 들어, 조작된 레트론 구축물 내의 msDNA 프로토스페이서 서열은 생물학적 현상 또는 환경 독소를 위한 센서 경로의 하류에 있는 프로모터에 의해 구동될 수 있다. CRISPR 어레이 내의 프로토스페이서 서열을 캡처하고 저장하는 것으로 사건이 기록된다. 여러 msDNA 프로토스페이서가 상이한 프로모터에 의해 구동되는 경우, 프로모터의 활성이 (프로모터의 상류일 수 있는 임의의 것과 함께) 기록될 뿐만 아니라 프로모터 활성의 상대적 순서 (CRISPR 어레이 내의 스페이서 서열의 상대적인 위치에 기반함)도 기록된다. 기록이 이루어진 후 임의의 시점에서, CRISPR 어레이는 주어진 생물학적 또는 환경적 사건이 발생했는지 여부를 결정하고 CRISPR 어레이 내의 msDNA 유래 스페이서의 존재 및 상대적 위치에 의해 제공된 다수의 사건의 순서를 결정하기 위해 시퀀싱될 수 있다.

[0119] 일부 실시양태에서, 합성 프로토스페이서는 그의 5' 말단에 AAG PAM 서열을 추가로 포함한다. 5' AAG PAM을 포함한 프로토스페이서는 PAM 서열을 포함하지 않은 프로토스페이서보다 더 높은 효율로 CRISPR 어레이에 의해 획득된다.

[0120] 일부 실시양태에서, Cas1 및 Cas2는 조작된 레트론에 의해 생산된 합성 프로토스페이서 서열이 세포 내의

CRISPR 어레이에 의해 획득되기 충분한 수준으로 Cas1 및 Cas2를 발현하는 벡터에 의해 제공된다. 이러한 벡터 시스템은 내인성 Cas 단백질이 결여된 세포에서 분자 기록을 가능하게 하는 데 사용될 수 있다.

- [0121] 본 개시의 예시적인 비-제한적 측면
- [0122] 전술한 본 주제의 실시양태를 포함한 측면은 단독으로, 또는 하나 이상의 다른 측면 또는 실시양태와 조합하여 유익할 수 있다. 전술한 설명을 제한하지 않으면서, 1 내지 60으로 넘버링된 개시내용의 특정 비-제한적 측면이 하기에 제공된다. 본 개시내용을 판독할 때 관련 기술분야의 통상의 기술자에게 명백한 바와 같이, 각각의 개별적으로 넘버링된 측면이 사용될 수 있거나, 또는 이것이 상기 또는 하기에 개별적으로 넘버링된 측면 중 임의의 것과 조합될 수 있다. 이는 그러한 모든 측면의 조합을 지원하기 위해 제공된 것이며, 하기에 명시적으로 제공되는 측면의 조합으로 제한되지는 않는다.
- [0123] 1. a) 프리-*msr* 서열;
- [0124] b) 멀티카피 단일 가닥 RNA (msRNA)를 코딩하는 *msr* 유전자;
- [0125] c) 멀티카피 단일 가닥 DNA (msDNA)를 코딩하는 *msd* 유전자;
- [0126] d) 프리-*msr* 서열에 대한 서열 상보성을 갖는 자기 상보적 영역을 포함하는 포스트-*msd* 서열, 여기서 자기 상보적 영역은 조작된 레트론이 msDNA의 생산을 증진시킬 수 있도록 야생형 자기 상보적 영역보다 적어도 1 내지 50 개 뉴클레오티드 더 긴 길이를 가짐; 및
- [0127] e) 역전사효소를 코딩하는 *ret* 유전자
- [0128] 를 포함하는 조작된 레트론.
- [0129] 2. 자기 상보적 영역이 야생형 자기 상보적 영역보다 적어도 5개, 적어도 10개, 적어도 15개, 적어도 20개, 적어도 25개, 또는 적어도 30개 뉴클레오티드 더 긴 길이를 갖는 것인 조작된 레트론.
- [0130] 3. 측면 1 또는 2에 있어서, *msr* 유전자와 *msd* 유전자가 트랜스 배열 또는 시스 배열로 제공되는 것인 조작된 레트론.
- [0131] 4. 측면 3에 있어서, *ret* 유전자가 *msr* 유전자 및 *msd* 유전자에 대하여 트랜스 배열로 제공되는 것인 조작된 레트론.
- [0132] 5. 측면 1 내지 4 중 어느 한 측면에 있어서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자가 미소박테리아 레트론, 에스케리키아 콜라이 레트론, 또는 비브리오 콜레라에 레트론으로부터 유래되는 것인 조작된 레트론.
- [0133] 6. 측면 5에 있어서, 에스케리키아 콜라이 레트론이 EC83 또는 EC86인 조작된 레트론.
- [0134] 7. 측면 1 내지 6 중 어느 한 측면에 있어서, 관심 이종 서열을 추가로 포함하는 조작된 레트론.
- [0135] 8. 측면 7에 있어서, 이종 서열이 *msr* 유전자 또는 *msd* 유전자 내로 삽입되는 것인 조작된 레트론.
- [0136] 9. 측면 1 내지 7 또는 8 중 어느 한 측면에 있어서, 이종 핵산 세그먼트가 상보적 영역의 염기로부터 9 내지 20개의 염기를 삽입하는 것인 조작된 레트론.
- [0137] 10. 측면 7, 8 또는 9에 있어서, 이종 서열이 *msd*, 포스트-*msd* 또는 msDNA의 헤어핀 루프 내에 있는 것인 조작된 레트론.
- [0138] 11. 측면 1 내지 8 또는 9 중 어느 한 측면에 있어서, 이종 핵산 세그먼트가 *msd* 또는 포스트-*msd*의 위치 20 내지 60에 삽입되는 것인 조작된 레트론.
- [0139] 12. 측면 7 내지 10 또는 11 중 어느 한 측면에 있어서, 이종 서열이 상동성 지정 복구 (HDR) 또는 재조합공학에 의해 표적 유전자좌에서 통합되도록 의도된 편집을 포함하는 뉴클레오티드 서열에 플랭킹하는 3' 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 표적 서열과 혼성화하는 5' 상동성 아암을 포함하는 공여자 폴리뉴클레오티드를 코딩하는 것인 조작된 레트론.
- [0140] 13. 측면 7 내지 11 또는 12 중 어느 한 측면에 있어서, 이종 서열이 CRISPR 프로토스페이스 DNA 서열을 포함하는 것인 조작된 레트론.
- [0141] 14. 측면 13에 있어서, CRISPR 프로토스페이스 DNA 서열이 변형된 AAG 프로토스페이스 인접 모티프 (PAM)를 포함하는 것인 조작된 레트론.

- [0142] 15. 측면 1 내지 14 중 어느 한 측면에 있어서, 바코드 서열을 추가로 포함하는 조작된 레트론.
- [0143] 16. 측면 15에 있어서, 바코드 서열이 msDNA의 헤어핀 루프에 위치하는 것인 조작된 레트론.
- [0144] 17. 측면 1 내지 16 중 어느 한 측면의 조작된 레트론을 포함하는 하나 이상의 벡터를 포함하는 벡터 시스템.
- [0145] 18. 측면 17에 있어서, *msr* 유전자와 *msd* 유전자가 동일한 벡터 또는 상이한 벡터에 의해 제공되는 것인 벡터 시스템.
- [0146] 19. 측면 17 또는 18에 있어서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자가 동일한 벡터에 의해 제공되는 것인 벡터 시스템.
- [0147] 20. 측면 17, 18 또는 19에 있어서, 벡터가 *msr* 유전자 및 *msd* 유전자와 작동 가능하게 연결된 프로모터를 포함하는 것인 벡터 시스템.
- [0148] 21. 측면 20에 있어서, 프로모터가 *ret* 유전자와 추가로 작동 가능하게 연결되는 것인 벡터 시스템.
- [0149] 22. 측면 20 또는 21에 있어서, *ret* 유전자와 작동 가능하게 연결된 제2 프로모터를 추가로 포함하는 벡터 시스템.
- [0150] 23. 측면 17 내지 21 또는 22 중 어느 한 측면에 있어서, *msr* 유전자, *msd* 유전자 및 *ret* 유전자가 상이한 벡터에 의해 제공되는 것인 벡터 시스템.
- [0151] 24. 측면 17 내지 22 또는 23 중 어느 한 측면에 있어서, 하나 이상의 벡터가 바이러스 벡터 또는 비-바이러스 벡터인 벡터 시스템.
- [0152] 25. 측면 24에 있어서, 비-바이러스 벡터가 플라스미드인 벡터 시스템.
- [0153] 26. 측면 17 내지 25 중 어느 한 측면에 있어서, 조작된 레트론이 상동성 지정 복구 (HDR) 또는 재조합공학에 의해 표적 유전자좌에서 통합되도록 의도된 편집을 포함하는 뉴클레오티드 서열에 플랭킹하는 3' 표적 서열과 혼성화하는 3' 상동성 아암 및 5' 표적 서열과 혼성화하는 5' 상동성 아암을 포함하는 공여자 폴리뉴클레오티드를 포함하는 것인 벡터 시스템.
- [0154] 27. 측면 26에 있어서, RNA-가이드된 뉴클레아제를 코딩하는 벡터를 추가로 포함하는 벡터 시스템.
- [0155] 28. 측면 27에 있어서, RNA-가이드된 뉴클레아제가 Cas 뉴클레아제 또는 조작된 RNA-가이드된 FokI-뉴클레아제인 벡터 시스템.
- [0156] 29. 측면 28에 있어서, Cas 뉴클레아제가 Cas9 또는 Cpf1인 방법.
- [0157] 30. 측면 17 내지 29 중 어느 한 측면에 있어서, 조작된 레트론이 CRISPR 프로토스페이서 DNA 서열을 포함하는 것인 벡터 시스템.
- [0158] 31. 측면 30에 있어서, Cas1 또는 Cas2 단백질을 코딩하는 벡터를 추가로 포함하는 벡터 시스템.
- [0159] 32. 측면 30 또는 31에 있어서, CRISPR 어레이 서열을 포함하는 벡터를 추가로 포함하는 벡터 시스템.
- [0160] 33. 측면 29 내지 31 중 어느 한 측면에 있어서, 박테리오파지 상동 재조합 단백질을 코딩하는 벡터를 추가로 포함하는 벡터 시스템.
- [0161] 34. 측면 33에 있어서, 박테리오파지 상동 재조합 단백질을 코딩하는 벡터가 *exo*, *bet* 및 *gam* 유전자를 포함하는 복제 결함 있는  $\lambda$  프로파지인 벡터 시스템.
- [0162] 35. 측면 1 내지 16 중 어느 한 측면의 조작된 레트론 또는 측면 17 내지 34 중 어느 한 측면의 벡터 시스템을 포함하는 단리된 숙주 세포.
- [0163] 36. 측면 35에 있어서, 원핵, 아키온, 또는 진핵 숙주 세포인 숙주 세포.
- [0164] 37. 측면 36에 있어서, 진핵 숙주 세포가 포유류 숙주 세포인 숙주 세포.
- [0165] 38. 측면 37에 있어서, 포유류 숙주 세포가 인간 숙주 세포인 숙주 세포.
- [0166] 39. 측면 35에 있어서, 인공 세포 또는 유전자 변형된 세포인 숙주 세포.
- [0167] 40. 측면 1 내지 16 중 어느 한 측면의 조작된 레트론, 측면 17 내지 34 중 어느 한 측면의 벡터 시스템 또는

측면 35 내지 39 중 어느 한 측면의 숙주 세포를 포함하는 키트.

- [0168] 41. 측면 40에 있어서, 조작된 레트론을 사용하여 세포를 유전자 변형시키기 위한 지침을 추가로 포함하는 키트.
- [0169] 42. a) 측면 1 내지 15 또는 16 (예를 들어, 측면 12)의 조작된 레트론으로 세포를 형질감염시키는 단계;
- [0170] b) RNA-가이드된 뉴클레아제 및 가이드 RNA를 세포 내로 도입하거나 이를 발현하여 유전자 변형된 세포를 생산하는 단계이며, 여기서 RNA-가이드된 뉴클레아제는 가이드 RNA와 복합체를 형성하고, 상기 가이드 RNA는 복합체가 게놈 표적 유전자좌를 향하도록 하며, RNA-가이드된 뉴클레아제는 게놈 표적 유전자좌에서 게놈 DNA 내에 이중 가닥 파손을 생성하고, 조작된 레트론에 의해 생성된 공여자 폴리뉴클레오티드는 상동성 지정 복구 (HDR)에 의해 그의 5' 상동성 아암과 3' 상동성 아암에 의해 인식된 게놈 표적 유전자좌에 통합되는 것인 단계를 포함하는, 세포를 유전자 변형시키는 방법.
- [0171] 43. 측면 42에 있어서, RNA-가이드된 뉴클레아제가 Cas 뉴클레아제 또는 조작된 RNA-가이드된 FokI-뉴클레아제인 방법.
- [0172] 44. 측면 43에 있어서, Cas 뉴클레아제가 Cas9 또는 Cpf1인 방법.
- [0173] 45. 측면 42 내지 44 중 어느 한 측면에 있어서, RNA-가이드된 뉴클레아제가 세포의 게놈에 통합된 벡터 또는 재조합 폴리뉴클레오티드에 의해 제공되는 것인 방법.
- [0174] 46. 측면 42 내지 45 중 어느 한 측면에 있어서, 조작된 레트론이 벡터에 의해 제공되는 것인 방법.
- [0175] 47. 측면 42 내지 45 또는 46 중 어느 한 측면에 있어서, 공여자 폴리뉴클레오티드가 유전자 대체, 유전자 녹아웃, 결실, 삽입, 반전 또는 점 돌연변이를 생성하는 데 사용되는 것인 방법.
- [0176] 48. a) 측면 1 내지 16 (예를 들어, 측면 12)의 조작된 레트론으로 세포를 형질감염시키는 단계; 및
- [0177] b) 박테리오파지 재조합 단백질을 세포 내로 도입하여 유전자 변형된 세포를 생산하는 단계이며, 여기서 박테리오파지 재조합 단백질은 조작된 레트론에 의해 생성된 공여자 폴리뉴클레오티드가 그의 5' 상동성 아암과 3' 상동성 아암에 의해 인식된 표적 유전자좌에 통합되도록 표적 유전자좌에서 상동 재조합을 매개하는 것인 단계를 포함하는, 재조합공학에 의해 세포를 유전자 변형시키는 방법.
- [0178] 49. 측면 48에 있어서, 공여자 폴리뉴클레오티드가 재조합공학에 의해 박테리아 세포 내의 플라스미드, 박테리아 인공 염색체 (BAC) 또는 박테리아 염색체를 변형시키는 데 사용되는 것인 방법.
- [0179] 50. 측면 48 또는 49에 있어서, 공여자 폴리뉴클레오티드가 유전자 대체, 유전자 녹아웃, 결실, 삽입, 반전 또는 점 돌연변이를 생성하는 데 사용되는 것인 방법.
- [0180] 51. 측면 48 내지 50 중 어느 한 측면에 있어서, 박테리오파지 재조합 단백질을 세포 내로 도입하는 것이 복제 결함 있는  $\lambda$  프로파지를 박테리아 게놈 내로 삽입하는 것을 포함하는 것인 방법.
- [0181] 52. 측면 51에 있어서, 박테리오파지가 *exo*, *bet*, 및 *gam* 유전자를 포함하는 것인 방법.
- [0182] 53. 측면 1 내지 15 또는 16 (예를 들어, 측면 15 또는 16)의 조작된 레트론으로 세포를 형질감염시키는 것을 포함하는, 세포를 마코딩하는 방법.
- [0183] 54. a) CRISPR 적응 시스템의 Cas1 단백질 또는 Cas2 단백질을 숙주 세포 내로 도입하는 단계;
- [0184] b) 리더 서열과 적어도 하나의 반복 서열을 포함하는 CRISPR 어레이 핵산 서열을 숙주 세포 내로 도입하는 단계이며, 여기서 CRISPR 어레이 핵산 서열은 숙주 세포 내의 게놈 DNA 또는 벡터에 통합되는 것인 단계; 및
- [0185] c) 측면 1 내지 16 (예를 들어, 측면 13 또는 14)에 따른 복수 개의 조작된 레트론을 숙주 세포 내로 도입하는 단계이며, 여기서 각각의 레트론은 프로세싱되어 CRISPR 어레이 핵산 서열 내로 삽입될 수 있는 상이한 프로토스페이스 DNA 서열을 포함하는 것인 단계를 포함하는, 생체내 분자 기록 시스템을 생산하는 방법.
- [0186] 55. 측면 54에 있어서, Cas1 단백질 또는 Cas2 단백질이 벡터에 의해 제공되는 것인 방법.
- [0187] 56. 측면 54 또는 55에 있어서, 조작된 레트론이 벡터에 의해 제공되는 것인 방법.

- [0191] 57. 측면 54 내지 56 중 어느 한 측면에 있어서, 복수 개의 조작된 레트론이 적어도 3개의 상이한 프로토스페이서 DNA 서열을 포함하는 것인 방법.
- [0192] 58. a) CRISPR 적응 시스템의 Cas1 단백질 또는 Cas2 단백질;
- [0193] b) 리더 서열과 적어도 하나의 반복 서열을 숙주 세포에 포함하는 CRISPR 어레이 핵산 서열, 여기서 CRISPR 어레이 핵산 서열은 조작된 세포 내의 게놈 DNA 또는 벡터에 통합됨; 및
- [0194] c) 측면 1 내지 16 (또는 측면 13 또는 14)에 따른 복수 개의 조작된 레트론, 여기서 각각의 레트론은 프로세싱되어 CRISPR 어레이 핵산 서열 내로 삽입될 수 있는 상이한 프로토스페이서 DNA 서열을 포함함
- [0195] 을 포함하는, 생체내 분자 기록 시스템을 포함하는 조작된 세포.
- [0196] 59. 측면 58에 있어서, Cas1 단백질 또는 Cas2 단백질이 벡터에 의해 제공되는 것인 조작된 세포.
- [0197] 60. 측면 58 또는 59에 있어서, 조작된 레트론이 벡터에 의해 제공되는 것인 조작된 세포.
- [0198] 61. 측면 58 내지 60 중 어느 한 측면에 있어서, 복수 개의 조작된 레트론이 적어도 3개의 상이한 프로토스페이서 DNA 서열을 포함하는 것인 조작된 세포.
- [0199] 62. 측면 58 내지 61 중 어느 한 측면의 조작된 세포 및 생체내 분자 기록을 위한 지침을 포함하는 키트.
- [0200] 63. a) 측면 1 내지 16 중 어느 한 측면의 조작된 레트론 또는 측면 17 내지 34 중 어느 한 측면의 벡터 시스템으로 숙주 세포를 형질감염시키는 단계; 및
- [0201] b) msDNA가 생산되는 적합한 조건하에 숙주 세포를 배양하는 단계
- [0202] 를 포함하는, 재조합 msDNA를 생산하는 방법.
- [0203] **실시예**
- [0204] 하기 실시예는 관련 기술분야의 통상의 기술자들에게 개시된 주제의 제조 및 사용 방법에 대한 개시내용과 설명을 제공하기 위해 제시되었으며, 본 발명자들은 자신 발명으로 간주하는 것의 범위를 제한하기 위한 것이 아니며, 하기 실험이 전부이거나 또는 수행된 유일한 실험이라는 것을 나타내려는 의도도 아니다. 사용된 수치 (예를 들어, 양, 온도 등)에 대한 정확성을 보장하기 위해 노력했지만, 일부 실험 오류와 편차를 고려해야 한다. 특별한 명시되지 않는 한, 부는 중량부이고, 분자량은 중량 평균 분자량이며, 온도는 °C이며, 압력은 대기압 또는 그 근처이다. 표준 약어, 예를 들어, bp, 염기 쌍(들); kb, 킬로염기(들); pl, 피코리터; s 또는 sec, 초(들); min, 분(들); h 또는 hr, 시간(들); aa, 아미노산(들); kb, 킬로염기(들); bp, 염기 쌍(들); nt, 뉴클레오티드(들); i.m., 근육내(로); i.p., 복강내(로); s.c., 피하(로) 등이 사용될 수 있다.
- [0205] **실시예 1: 재료 및 방법**
- [0206] 본 실시예는 본 발명을 개발하는 데 사용되는 재료 및 방법 중 일부를 예시한다.
- [0207] **박테리아 균주, 플라스미드 및 배양 조건**
- [0208] 실험은 통합된 아라비노스 유도성 T7 폴리머라제, 내인성 CRISPR 어레이, 내인성 레트론을 함유하지만, 내인성 Cas1+2; 또는 통합된 아라비노스 유도성 T7 폴리머라제를 함유하는 MG1655의 변이체인 bMS.346은 함유하지 않고 내인성 레트론은 함유하지 않는 BL21-AI 이. 콜라이 [썬모 피셔(Thermo Fisher)]에서 수행되었다.
- [0209] 도 3b, 3c, 4d, 7b, 7cb, 8b, 9b, 및 10b-d에 도시된 레트론 기반 실험에서, 역전사효소는 에리트로마이신 유도성 프로모터를 갖는 플라스미드 (mphR-ec86RT)로부터 발현되었다 (문헌 [Rogers et al., *Nucleic Acids Res.* 2015 Sep 3;43(15):7648-60. doi: 10.1093/nar/gkv616. Epub 2015 Jul 7] (그 전체 내용이 본원에 참조로 포함됨) 참조). msd 및 msr 요소는 유도성 T7 프로모터로부터 함께 (DUET-T7-msr/msd) 또는 별도로 (DUET-T7-msr-T7-msd) 발현되었다.
- [0210] 도 11a-b에 대해 기재된 레트론 생성 프로토스페이서 실험의 경우, 둘 다 유도성 (T7/lac) 프로모터에 의해 발현된, Cas1+2 및 변형된 ec86 msr/msd를 코딩하는 플라스미드 (DUET-msr/msd-Cas1+2)를 각각의 실험에 앞서 세포 내로 형질전환시켰다. 플라스미드를 함유하는 세포는 최대 3주 동안 4°C의 플레이트 상에 콜로니로 유지되었다. 세포는 34°C 하의 LB 배지에서 성장하였으며, 지정된 지속 기간 동안 IPTG, L-아라비노스 및/또는 에리트로마이신을 사용하여 유도되었다.

[0211] **msd의 전기영동 분석**

[0212] 변형된 레트론으로부터 생산된 msd를 시각화하기 위해, 박테리아는 msr 함유, msd 함유 및 역전사효소 함유 전사체를 발현하는 데 필요한 모든 유도인자와 함께 LB에서 4-16시간 동안 배양되었다. 5-25 ml 용적의 배양물을 4°C에서 펠릿화한 다음, 플라스미드 플러스 미디 키트 [퀴아젠(Qiagen)] 또는 미니 키트를 사용하여 준비하였다. 이어서, RNA는 RNaseA와 RNaseT1의 조합을 사용하여 소화시키고, 그 결과로 생성된 msd는 ssDNA/RNA 클린 앤 농축기 키트 [자이모 리서치(Zymo Research)]를 사용하여 정제되었다. msd는 노벡스(Novex) TBE-우레아 겔 (썬모 피셔)에서 실행되고 SYBR 골드 (썬모 피셔)로 사후 염색함으로써 시각화되었다.

[0213] **변이체 라이브러리 구축**

[0214] 레트론 ncRNA 변이체 라이브러리는 애질런트(Agilent) 또는 트위스트(Twist)에 의해 합성 실행당 다수의 라이브러리와 함께 올리고 풀로서 합성되었다. 단일 라이브러리는 이러한 올리고 풀로부터 증폭되고, NEB5a 세포를 클로닝 균주로서 사용하여 골든 게이트 접근 방식을 사용하여 발현 벡터 내로 클로닝되었다. 이러한 클로닝된 라이브러리는 클로닝 균주로부터 정제되고 발현 균주 (BL21-AI 또는 bMS.346)로 전이되었다. 모든 라이브러리는 일루미나(Illumina) 시퀀싱에 의해 정량화되었다.

[0215] **시퀀싱 및 분석**

[0216] 라이브러리 실험에서 RT-DNA 존재량을 정량화하기 위해, 역전사된 DNA는 레트론 변이체의 라이브러리를 함유하는 세포에서의 발현 후 상기 기재된 바와 같이 정제되었다. 5-25 ml 용적의 배양물을 4°C에서 펠릿화한 다음, 플라스미드 플러스 미디 키트 (퀴아젠) 또는 미니 키트를 사용하여 준비하였다. 이어서, RNA는 RNaseA와 RNaseT1의 조합을 사용하여 소화시키고, 그 결과로 생성된 msd는 ssDNA/RNA 클린 앤 농축기 키트 (자이모 리서치)를 사용하여 정제되었다. 변형이 역전사된 요소를 벗어난 변이체 라이브러리에서 (예를 들어, 도 7cb), 역전사된 DNA의 루프에서의 바코드 부착된 영역이 발현 후 증폭되어 일루미나 시퀀싱을 위해 준비되었다. 변형이 역전사된 요소 내부에 있었던 변이체 라이브러리에서 (예를 들어, 도 8b), 정제된 역전사된 DNA의 풀은 말단 테옥시뉴클레오티드 트랜스퍼라제 (TdT)를 사용하여 단일 뉴클레오티드로 3' 말단에서 확장되었으며, 부가의 뉴클레오티드의 길이는 TdT 인큐베이션 시간에 의해 제어되었다 (도 6b). 다음으로, 어댑터 서열, 확장을 위해 사용되는 뉴클레오티드와 상보적인 뉴클레오티드의 연장물, 및 고정 뉴클레오티드 (확장에 사용되는 뉴클레오티드와 상보적이지 않은 모든 염기)로 구성된 역방향 프라이머를 사용하여, 5' 말단에 A 오버행을 남기는 클레노브 단편 (3'→5' exo-)를 사용하여 제2 가닥을 생성하였다. 이러한 오버행 A는 대립하는 5' 말단 상에 아미노 변형된 이중 가닥 어댑터를 부착하기 위해 TA 라이게이션에 사용되었다 (도 6a). 양 말단에 어댑터가 추가된 이러한 뉴클레오티드 풀은 일루미나 시퀀싱을 위해 인덱싱되고 준비되었다. 모든 변이체 라이브러리에서, 세포에 존재하는 플라스미드의 풀은 가변 영역을 증폭시키고 그 영역을 일루미나 시퀀싱을 실시함으로써 정량화하였다. 그 다음, 역전사된 DNA 내의 변이체의 비율과 변이체 플라스미드의 비율을, 공동 발현된 야생형 레트론으로 표준화하여 비교함으로써, 상이한 역전사된 DNA의 상대적인 존재량을 계산하였다.

[0217] 스페이스 획득을 분석하기 위해, 박테리아를 95°C까지 5분 동안 가열하여 용해시킨 다음, 리더-반복 접합부에 플랭킹하고 일루미나 호환 어댑터를 부가적으로 함유하는 프라이머를 사용하여 게놈 어레이의 PCR을 실시하였다. 스페이스 서열은 플랭킹 반복 서열의 존재에 기초하여 생체정보적으로 추출되었으며, 이를 기존의 스페이스 서열에 대항하여 비교하여 확장된 어레이의 백분율과 새로 획득한 스페이스의 위치 및 서열을 결정하였다. 새로운 스페이스는 게놈과 플라스미드 서열에 대항하여 BLAST로 분석되었고 (NCBI), 프로토스페이스의 기원을 결정하기 위해 의도된 프로토스페이스 서열에 대항하여 부가적으로 비교되었다. 이러한 분석은 피톤(Python)에서 맞춤 작성 스크립트를 사용하여 수행되었다.

[0218] 실시예 2: 세포에서 DNA 생산을 증진시키기 위한 박테리아 레트로요소 (레트론)

[0219] 살아 있는 세포의 게놈을 다시 쓰는 것은 새로운 DNA를 필요로 한다. 현재, 작업자들이 이러한 DNA를 외생적으로 합성하여 이를 세포에 전달하며, 여기서 이는 게놈 조작을 위한 템플릿 또는 세포 또는 세포성 사건을 표시하는 바코드의 역할을 한다. 그러나, 특히 복잡한 조직 내에서, 상동성 지정 복구 (HDR) 및 통합 프로세스에서 비효율성을 극복하기에 충분한 존재량의 외인성 DNA를 전달하는 것은 여전히 매우 어려운 일이다. 더욱이, 특정한 DNA 서열을 가진 세포의 하위 집단을 표적으로 할 수 있도록 세포 유형 또는 세포 상태에 따른 전달을 게이트할 방법이 없다.

[0220] 만약 본 발명자들이 RNA와 단백질에 하는 것처럼, 본 발명자들이 선택한 세포 안에서 충분히 요구에 따라 설계된 DNA 서열을 생산할 수 있다면, HDR에서의 비효율성을 극복하고 상이한 세포에 상이한 템플릿을 전달할 수 있

는 능력을 열 수 있을 것이다. 본 발명자들은 이러한 국소적으로 생산된 DNA 템플릿을 사용하여 게놈 편집에서 게놈 작성으로 전환할 수 있다. 맞춤형 DNA는 박테리아 게놈을 기록하기 위한  $\lambda$  레드 베타, 인간 게놈에 치료적 변형을 작성하기 위한 Cas9, 및 살아 있는 세포 내에서 분자 사건의 타이밍을 기록하는 분자 장치를 생성하는 CRISPR 인테그라제 Cas1+2를 포함한, DNA를 변형시키는 단백질의 어레이에 연료를 공급할 것이다.

[0221] 본원에 기재된 바와 같이, 역전사효소는 상이한 서열을 가진 상이한 DNA를 포함한 풍부한 DNA를 생산하기 위한 해결책이다. 역전사효소는 풍부한 DNA를 생성할 수 있을 뿐만 아니라, 그의 활성은 본 발명자들이 현재 RNA와 단백질 발현을 제어하는 것과 동일한 방식으로 시간과 공간을 넘어 제어될 수 있다. 따라서, 광범위하게 전달된 역전사효소는 표적화된 세포 서브세트에서 풍부한 템플릿 DNA를 생성할 수 있다.

[0222] 역전사효소의 특히 매력적인 클래스는 박테리아로부터 나오며, 레드론이라고 불린다 (Inouye & Inouye, *Annual review of microbiology* 45, 163-186 (1991)) (예를 들어, 도 1a 참조). 레드론은 작고 모듈형이며, 진핵 세포와 적교하며, 다른 단백질에 접근할 수 있는 DNA를 생산하는 것으로 나타났으며, 원핵 세포 (문헌 [Farzadfard & Lu *Science* 346, 1256272 (2014)]) 및 진핵 세포 (문헌 [Sharon et al. *Cell* 175, 544-557 (2018)]) 모두에서 게놈 편집을 위한 템플릿 DNA로서 작용하였다 (도 1b, 1c). 그러나, 본 발명자들이 레드론의 생물학에 관하여 아직 모르는 것이 많고, 이러한 지식 격차는 본 발명자들이 레드론을 사용하여 세포 내부에 완전히 설계된 DNA 서열을 생성하지 못하게 한다.

[0223] 여기서, 본 발명자들은 세포 내에서 CRISPR 호환의 임의의 DNA 서열을 대량으로 생산하기 위해 레드론을 추가로 특징 규명하고 조작함으로써 레드론의 한계를 직접 해결한다. 대부분의 조작은 높은 처리량을 달성하기 위해 이. 콜라이에서 수행되었지만, 본원에 기재된 변형된 레드론과 시스템은 게놈 작성의 맥락에서 개선을 제공하기 위해 진핵 세포 (인간 포함)에서 사용될 수 있다.

[0224] 예를 들어, 일부 경우에는 레드론-Eco1을 예시적인 레드론으로서 사용하였다. 이러한 전사체는 도 2a에 나타난 바와 같이, 역전사효소에 의해 인식되며, 부분적으로 RT-DNA로 역전사된다. 역전사된 레드론-Eco1 DNA (RT-DNA)의 서열은 서열식별번호: 14로서 하기에 제시된다.

```

1 GTCAGAAAAA ACGGGTTTCC TGGTTGGCTC GGAGAGCATC
41 AGGCGATGCT CTCCGTCCA ACAAGAAAA CAGACAGTAA
81 CTCAGA
    
```

[0225]

[0226] 실시예 3: 조작된 레드론 변이체 스크리닝

[0227] 본 발명자들은 이. 콜라이에서 레드론-Eco1 (ec86 또는 레드론-Eco1 ncRNA라고도 함)을 발현하였다. 이러한 야생형 레드론-Eco1 ncRNA의 서열은 서열식별번호: 15로서 하기에 제시된다.

```

1 TGCGCACCCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTTT CCTTGTGGA ACGGAGAGCA TCGCCTGATG
121 CTCTCCGAGC CAACCAGGAA ACCCGTTTTT TCTGACGTAA
161 GGGTGCACA
    
```

[0228]

[0229] 정량적 PCR (qPCR)은 이. 콜라이에서 발현된 레드론-Eco1의 발현이 세포당 약 800-1,000개 카피의 ssDNA를 산출하였다는 것을 보여주었다 (도 3b). 도 3c에 예시된 바와 같이, 이와 같이 생산된 ssDNA는 변성 겔 상에서 시각화, 정량화 및 정제될 수 있다.

[0230] 본 발명자들은 또한 다양한 레드론 요소를 코딩하는 구축물을 만들었다. 예를 들어, 역전사효소는 msr/msd (프라이머-템플릿)로부터 분리되어, msr과 msd가 역전사효소에 대해 (전형적인 시스 배열이 아닌) 트랜스로 공급될 수 있었다 (도 4a, 4b). 이러한 트랜스 배열은 역전사효소에 대한 잠재적 정지 신호를 제거한다. 레드론-Eco1 msr 전용 영역의 서열은 서열식별번호: 16으로서 하기에 제시된다.

```

1 TGCGCACCCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 AAGCTGTTTT TCGCCAG
    
```

[0231]

[0232] 레트론-Eco1 msd 전용 영역의 서열은 서열식별번호: 17로서 하기에 제시된다.

```
1 AATCTGAGTT ACAAGCTGTT TGTCGCCAGT CAGACTGGCG
41 ACAACCCGTT TTTTCTGACG TAAGGGTGCG CA
```

[0233]

[0234] 낮은 처리량 실험을 사용하여, 역전사효소에 의해 허용되는 msd 요소에 대한 변화가 이루어졌다. 예를 들어, 레트론-Eco1 v32 ncRNA 및 레트론-Eco1 v35 ncRNA의 2가지 변이체가 생산되었다. 레트론-Eco1 v32 ncRNA의 서열은 서열식별번호: 18로서 하기에 제시된다.

```
1 TGCGCACCCCT TAGCGAGAGG TTTATCATT AAGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 AAGCTGTTTG TCGCCAGTCA GACTGGCGAC AACCCGTTTT
121TTCTGACGTA AGGGTGCGCA
```

[0235]

[0236] 레트론-Eco1 v35 ncRNA의 서열은 서열식별번호: 19로서 하기에 제시된다.

```
1 TGCGCACCCCT TAGCGAGAGG TTTATCATT AAGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 AAGCTGTTTG TCGCCAGTCA GACTGGCGAC AACCCGTTTT
121 TTCTGACGTA AGGGTGCGCA
```

[0237]

[0238] 레트론-Eco1 v32 ncRNA 및 레트론-Eco1 v35 ncRNA의 주요 섹션은 도 3d에 나타나 있다.

[0239] 다수의 변형이 레트론-Eco1 ncRNA에 행해졌다. 그러나, 모든 변형 시도가 성공적인 것은 아니었다. 일부 경우에는, 변형된 버전의 대다수가 세포에서 ssDNA를 전혀 생산하지 않았다.

[0240] 레트론으로부터 msd 생산의 결정 요인을 더 잘 이해하기 위해, 본 발명자들은 라이브러리 기반 접근 방식을 취하였다. 수만 개의 레트론 변이체를 합성하여 레트론의 각각의 구조적 파라미터를 체계적으로 시험하였다 (도 5). 이러한 변이체를 클로닝하기 위해 골든 게이트 기반 클로닝 전략 (문헌 [Engler et al., PLOS One (Nov. 5, 2008)])이 사용되었고, 이후 변형된 레트론의 대규모 풀이 역전사효소와 함께 다중화된 실험에서 발현되었다. 이들 세포에 의해 생산된 모든 msd를 정제하고, 이를 시퀀싱하며, 이들의 존재량을 발현 균주에서의 그들의 기원 레트론/플라스미드와 비교함으로써, 레트론의 특정한 파라미터의 영향은 ssDNA 생산과 관련이 있는 것으로 정량화되었다. ec86 레트론뿐만 아니라 내부 분지 구조를 가진 ec83 레트론을 포함한 다른 레트론이 사용되었다.

[0241] 레트론 역전사효소는 전형적으로 2' 위치에 있는 msr RNA를 포스포디에스테르 결합을 통해 msd ssDNA의 5' 말단에 연결하는 분지형 RNA-DNA 하이브리드를 생성하기 위해 비-표준 방식으로 프라이밍한다 (Inouye & Inouye, *Annual review of microbiology* 45, 163-186 (1991)). 이. 콜라이는 이러한 결합을 절단하는 효소가 없다. 따라서, 이. 콜라이 내에 있을 때 ec86 레트론은 분지된 상태로 남아 있다. 그러나, ec83은 레트론 고유의 공지되지 않은 메커니즘을 통해 프로세싱되어, 2'-5' 연결을 제거하고 ssDNA를 자유롭게 하는 것으로 또한 보고되었다 (Lim, *Molecular microbiology* 6, 3531-3542 (1992)). 이러한 분리는 게놈 조작에서의 다양한 적용에 혜택을 줄 수 있다.

[0242] 실시예 4: 조작된 레트론 변이체 시퀀싱

[0243] 설계상, 정제된 ssDNA의 풀은 공지되지 않은 부분 (예를 들어, 상이한 말단)을 함유하기 때문에, 레트론 유래 ssDNA를 실험의 관독으로 시퀀싱하면 상당한 복잡성이 발생한다. 이러한 ssDNA는 전통적인 파이프라인을 사용하여 다중화된 시퀀싱을 준비할 수 없다. 이러한 문제를 해결하기 위해, 본 발명자들은 ssDNA를 정제하고, RNAase로 ssDNA를 처리하고, 레트론 유래 ssDNA를 탈분지화하는 것을 포함하는, 맞춤형 시퀀싱 파이프라인을 개발하였다. 정제 및 탈분지된 레트론 ssDNA는 템플릿 비의존성 폴리머라제 (TdT)를 사용하여 단일 유형의 폴리뉴클레오티드의 스트링으로 테일링되었다. 이어서, ssDNA의 상보적 가닥은 어댑터를 함유한 역 고정 프라이머를 사용하여 생성되었다 (도 6a, 6b). 이러한 이중 가닥 DNA에 제2 어댑터를 라이게이션한 다음, 이러한 어댑터 연결된 이중 가닥 DNA를 인덱싱하며 다중화된 시퀀싱을 수행한다.

[0244] 이러한 파이프라인은 합성된 올리고뉴클레오티드, 야생형 및 변형된 ec86 및 야생형 ec83을 사용하여 검증되었다. 상기 방법은 합성된 올리고뉴클레오티드의 정확한 서열을 신뢰성있게 결정한다.

[0245] 흥미롭게도, 이러한 다중화된 단일 분자 방법을 사용하여, ec86 레트론 유래 ssDNA는 전형적으로 통상의 대량

방법론 [예를 들어, 맥삼 길버트(Maxam-Gilbert) 시퀀싱]을 사용하는 문헌에서 보고된 것보다 한 염기 일찍 종료된다. ec83의 절단 및 예측된 내인성 엑소뉴클레아제 프로세싱이 또한 확인되었다 (도 6c-6f). 레트론 유래 ssDNA가 시퀀싱되기 때문에, 레트론의 msd (템플릿) 부분의 변형을 직접 판독할 수 있다.

[0246] 본 발명자들은 또한 역전사되지 않은 ncRNA의 파라미터를 이해하는 것을 목표로 한다. 이러한 파라미터를 판독하기 위해, 비-역전사된 영역에서의 변이체는 msd의 루프 영역에 삽입된 바코드에 연결되었다 (도 7a). 이러한 접근 방식은 변이체가 직접 시퀀싱 되지 않았음에도 불구하고, 예를 들어 ssDNA 생산에 대한 서열 변이의 효과를 조명하였다.

[0247] 실시예 5: 세포 내 DNA의 생산을 증가시키는 변형

[0248] 본 실시예는 msr과 msd 전사체의 분리가, 더 긴 RT-DNA의 생산을 가능하게 한다는 것과, 레트론 자기 상보적 비-코딩 RNA 영역의 길이를 변형시키면, 레트론에 의해 생성된 역전사된 DNA의 존재량이 증가될 수 있다는 것을 예시한다.

[0249] 실시예 3에 기재된 바와 같이, 레트론-EcoI msr 및 msd 요소를 코딩하는 트랜스 구축물의 발현은 역전사효소에 대한 잠재적 정지 신호를 제거하였고, 더 긴 ssDNA를 생성할 수 있게 하였다 (도 4c, 4d).

[0250] 확장된 트랜스 레트론-EcoI msd 서열의 한 예는 레트론-EcoI msd +50으로서 지칭되며 서열식별번호: 20으로서 하기에 제시된다.

```

1 TGGACAATAT TGAATGGAGT CTGATCAACC TTCACACCGA
41 TCTAGAATCG GAATCTGAGT TACAAGCTGT TTGTCGCCAG
81 TCAGACTGGC GACAACCCGT TTTTCTGAC GTAAGGGTGC
121 GCA
    
```

[0251]

[0252] 도 7에 도시된 바와 같이, 레트론-EcoI ncRNA의 5' 및 3' 말단에서의 자기 상보적 영역의 확장은 레트론에 의해 세포에서 생산되는 RT-DNA의 존재량을 크게 증가시킨다. 확장된 레트론-EcoI ncRNA의 한 예는 서열식별번호: 21로서 하기에 제시된다.

```

1 TGATAAGATT CCGTATGCGC ACCCTTAGCG AGAGGTTTAT
41 CATTAAGGTC AACCTCTGGA TGTTGTTTCG GCATCCTGCA
81 TTGAATCTGA GTTACTGTCT GTTTTCTTGG TTGGAACGGA
121 GAGCATCGCC TGATGCTCTC CGAGCCAACC AGGAAACCCG
161 TTTTTTCTGA CGTAAGGGTG CGCATACGGA ATCTTATCA
    
```

[0253]

[0254] 다소 상이한 서열을 가진 확장된 레트론-EcoI v35 ncRNA의 한 예가 하기에 제시된다 (서열식별번호: 22).

```

1 TGATAAGATT CCGTATGCGC ACCCTTAGCG AGAGGTTTAT
41 CATTAAGGTC AACCTCTGGA TGTTGTTTCG GCATCCTGCA
81 TTGAATCTGA GTTACAAGCT GTTTGTCGCC AGTCAGACTG
121 GCGACAACCC GTTTTTCTG ACGTAAGGGT GCGCATACGG
161 AATCTTATCA
    
```

[0255]

[0256] 도 7b-7c에 예시된 바와 같이, 레트론-EcoI ncRNA의 5' 및 3' 말단에서의 자기 상보적 영역의 확장은 레트론에 의해 세포에서 생산되는 RT-DNA의 존재량을 크게 증가시킨다. 예를 들어, msd 서열 자기 상보적 영역의 길이를 증가시킴으로써 ssDNA의 상대적인 양에 있어서의 10배 증가가 초래될 수 있다.

[0257] 반대로, ncRNA 자기 상보적 염기에서의 감소는 RT-DNA의 생산을 크게 감소시켰다. 예를 들어, 자기 상보적 서열이 더 짧은 레트론-EcoI ncRNA에 대한 하나의 서열은 서열식별번호: 23으로서 하기에 제시된다.

```

1 TAGCGAGAGG TTTATCATTA AGGTCAACCT CTGGATGTTG
41 TTTTCGGCATC CTGCATTGAA TCTGAGTTAC TGTCTGTTTT
81 CCTTGTTGGA ACGGAGAGCA TCGCCTGATG CTCTCCGAGC
121 CAACCAGGAA ACCCGTTTTT TCTGACGTA
    
```

[0258]

[0259] 도 7c에 도시된 바와 같이, 이러한 상보적 염기의 수에 있어서의 감소는 RT-DNA의 생산을 크게 감소시켰다.

[0260] 따라서, 프리-msr/포스트-msd 자기 상보적 영역의 확장은 ssDNA의 풀을 증가시킬 수 있다. 역전사 ssDNA의 더 큰 풀은 유전적 변형에 이용될 수 있으며, 박테리아 (재조합공학), 효모 (CRISPEY), 및 포유류 세포에서 계층

편집의 효율성을 증가시킬 수 있다. 살아 있는 세포에서 풍부한 DNA를 생산할 목적으로, 확장된 자기 상보적 영역을 가진 이러한 변이체가 선호된다.

[0261] 실시예 6: Msd 스템 영역은 전부가 아닌 일부의 변형을 허용한다

[0262] 레트론-Eco1 ncRNA 영역의 msd 스템 영역을 변형시켜 스템 2차 구조 (이중 가닥 결합)를 붕괴시켰다. 본 실시예는 레트론에 의해 생산된 역전사된 ssDNA의 존재량에 부정적인 영향을 미치지 않으면서 msd 스템을 변형시킬 수 있다는 것을 예시한다.

[0263] msd 스템을 따라 변형된 위치는 도 8a-8b 및 9a-9b에 예시되어 있다.

[0264] msd 스템 구조의 길이를 변형함으로써 더 짧은 RT-DNA 서열을 생성할 수 있다. 예를 들어, 여전히 야생형 수준의 ssDNA (레트론-Eco1 스템 짧은 OK, 도 8b 참조)를 제공하는 짧은 스템 길이를 가진 레트론-Eco1 ncDNA에 대한 하나의 서열이 서열식별번호: 24로서 하기에 제시된다.

```

1 TGCGCACCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTTT CCTTGTTGGA AGCCTAGCCA ACCAGGAAAC
121 CCGTTTTTTC TGACGTAAGG GTGCGCA
    
```

[0265]

[0266] 그러나, msd 스템 길이가 14개 미만 염기로 감소하기 때문에, 생성되는 RT-DNA의 존재량은 부정적인 영향을 받는다 (도 8b). 너무 짧은 스템을 가진 서열 레트론-Eco1 ncDNA의 하나의 예가 서열식별번호: 25로서 하기에 제시된다 (도 8b, 너무 짧은 스템).

```

1 TGCGCACCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTTT CCTAACCCGT TTTTCTGAC GTAAGGGTGC
121 GCA
    
```

[0267]

[0268] 이와는 대조적으로, 레트론의 스템 영역이 손상된 후 복구될 때, 생성되는 ssDNA의 양은 야생형과 동일하거나 그보다 다소 더 높다. "손상된"은, 예를 들어, 비-상보적 뉴클레오티드를 도입함으로써, 스템의 염기 쌍 형성을 약화시키는 것을 의미한다. 도 9b에 의해 예시된 바와 같이, 5개의 염기를 연속적으로 변화시킴으로써 ncRNA 스템의 염기가 손상되면, RT-DNA의 존재량이 급격히 감소하게 된다. 5개의 미스매치된 염기 (손상된 스템, 도 9b)를 가진 이러한 레트론-Eco1 ncRNA에 대한 서열이 서열식별번호: 26으로서 하기에 제시된다.

```

1 TGCGCACCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTTT CGAACATGGA ACGGAGAGCA TCGCCTGATG
121 CTCTCCGAG CAACCAGGAA ACCCGTTTTT TCTGACGTAA
161 GGGTGCGCA
    
```

[0269]

[0270] 그러나, 이러한 염기가 스템 2차 구조를 보존하기 위한 스템의 반대편의 상보적인 변화에 의해 보상된다면, RT-DNA 존재량은 보존된다 (예를 들어, 고정된 스템, 도 9b 참조). 도 9b에 도시된 '고정된 스템' 레트론-Eco1 ncRNA에 대한 서열이 서열식별번호: 27로서 하기에 제시된다.

```

1 TGCGCACCCT TAGCGAGAGG TTTATCATTA AGGTCAACCT
41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC
81 TGTCTGTTTT CGAACATGGA ACGGAGAGCA TCGCCTGATG
121 CTCTCCGAG CATGGTCGAA ACCCGTTTTT TCTGACGTAA
141 GGGTGCGCA
    
```

[0271]

[0272] 스템의 기저부로부터 9 내지 20개의 염기인 영역에서의 스템에 대한 변형은, 이것으로 인해 스템 구조가 손상되는 경우일지라도 허용된다 (도 9b, 허용 가능한 손상된 스템 참조). 이러한 허용 가능한 손상된 스템은 스템의 중간 부분에 변형 (미스매치)을 갖는다. 도 9b의 허용 가능한 손상된 스템을 갖는 레트론-Eco1 ncRNA에 대한 서열의 하나의 예가 서열식별번호: 28로서 하기에 제시된다.

1 TGC GCACCCCT TAGCGAGAGG TTTATCATT AAGTCAACCT  
 41 CTGGATGTTG TTTCGGCATC CTGCATTGAA TCTGAGTTAC  
 81 TGTCTGTTTT CCTTGTGGGA ACGCTCTCCA TCGCCTGATG  
 121 CTCTCCGAGC CAACCAGGAA ACCCGTTTTT TCTGACGTAA  
 161 GGGTGC GCA

[0273]

[0274]

따라서, 살아 있는 세포에서 DNA를 생산하기 위해, ncRNA의 서열은 스템 구조의 염기가 보존되는 한 변형될 수 있다. msd 스템의 중간부, 스템의 기저부로부터의 약 9 내지 20개 염기에 대한 변형은 허용되며, ssDNA의 역전사에 악영향을 미치지 않는다.

[0275]

실시예 7: ncRNA msd 스템 영역 중심이 변형에 대해 내성이 더 높다

[0276]

역전사된 (msd) 레트론-Eco1 ncRNA 영역 내의 다양한 위치에 작은 변형이 이루어졌으며, 이러한 변형의 영향은 상이한 ncRNA 변이체로부터 역전사된 ssDNA의 양에 대해 측정되었다.

[0277]

**도 10**에 예시된 바와 같이, ncRNA의 역전사된 영역 전체에 걸친 작은 변형에 대한 허용치는 가변적이다. **도 10b**는 msd 상보적 (스템) 영역을 따라 다양한 위치로부터 3개의 염기를 결실시킨 효과를 나타낸다. 도시된 바와 같이, msd 스템의 중간으로부터 3개의 염기를 결실시켜도 부작용이 없었으며, 여전히 높은 수준의 ssDNA 생산으로 이어졌다 (**도 10b**). 그러나, 3개 염기의 결실이 상보적 (스템) 영역의 기저부 근처 또는 스템에 플랭킹하는 영역 내에서 이루어졌을 때, 더 낮은 수준의 ssDNA 생산이 관찰되었다 (**도 10b**). msd 영역의 중간 및 플랭킹 부분 내의 3개의 염기 삽입 (**도 10c**)과 단일 염기 변화 (**도 10d**)에 대해서도 유사한 효과가 관찰되었다. msd 스템 영역의 중간부는 ssDNA 생산에 있어서의 유의미한 감소가 관찰되지 않도록 여러 뉴클레오티드 (예를 들어 5개 미만의 뉴클레오티드)의 삽입 및/또는 결실을 허용했지만, msd 스템의 기저부에서의 플랭킹 서열에 대한 그러한 변형은 허용되지 않았다. msd 스템의 기저부와 플랭킹 영역의 변형은 ssDNA의 역전사를 감소시켰다.

[0278]

**도 10e**는 **도 10b-10d**의 데이터를 기반으로 계산된 msd 스템 영역 내의 위치에 대한 변형 가능성 점수를 그래프로 예시한다. 살아 있는 세포에서 더 높은 수준의 DNA를 생산하기 위해, ncRNA의 서열은 높은 변형 가능성 점수를 가진 영역에서 변형되어야 하며, 낮은 변형 가능성 점수를 가진 영역의 변형은 피해야 한다.

[0279]

실시예 8: 조작된 레트론의 적용

[0280]

조작된 레트론이 있는 세포에서 맞춤형으로 DNA를 생성시키는 것은 살아 있는 세포에서의 게놈 편집에서 게놈 작성으로의 전환을 가능하게 한다. 이러한 전환은 본 발명자들이 이전에 존재했던 서열에 의해 제한되지 않으면서 세포를 치료적으로 변형시키게 해줄 것이다. 현재, 새로운 DNA HDR 템플릿은 시간이 지남에 따라 감소하는 대형 단일 용량으로서 전달될 수 밖에 없다. 이러한 대형 단일 용량 전달의 비효율성은 그것들이 계내에서 작성될 수 없다는 것을 의미하며, 대신 시험관내에서 작성되어야만 하며, 선택과 확장이 뒤따라야 한다. 모든 실험과 소수의 치료제가 이러한 전략과 양립하는 것은 아니다.

[0281]

CRISPR 기반 치료제의 잠재력을 완전히 실현하기 위해, 본 발명자들은 게놈이 필요한 시기와 장소를 정확하게 제작하도록 설계된 DNA 서열을 제공한다. 이와 같이 설계된 서열은 **도 11a**에 예시된 바와 같이 제공될 수 있다. ncRNA의 5' 말단과 3' 말단에서 자기 상보적 영역을 확장함으로써 변형의 효율이 증가하며, 레트론 역전사 효소는 그로부터 원하는 ssDNA를 풍부하게 생산하기 위해 동원될 수 있다.

[0282]

맞춤형 DNA를 생산하는 것은 또한, 게놈 조작의 새로운 적용을 가능하게 하는데, 이는 치료제를 직접적으로 목표로 하는 것이 아니라 질환의 생물학을 이해하는 것을 목표로 한다. 이것은 세포의 게놈에 대한 변형이 살아 있는 시스템 내에서 데이터를 작성하기 위해 사용되는 분자 기록의 분야이다. CRISPR 인테그라제 Cas1+2를 사용하여 CRISPR 어레이 내의 짧은 DNA 서열을 캡처하고, 시간 경과에 따라 일지를 기록하여 사건 기록을 저장하는 이러한 접근 방식의 설명에 대해서는, 예를 들어 문헌 [Shipman et al. (2017) Nature 547(7663):345-349, and Shipman et al. (2016) Science 353(6298):aaf1175] (그 전체 내용이 본원에 참조로 포함됨)을 참조한다. 이러한 접근 방식은 세포에서 데이터 수집 장치의 백 엔드로서 사용될 수 있다. 그러나, 그것은 전사와 같은 생물학적 사건에 의해 구동되는, 세포에서 DNA 바코드를 생성하는 프런트 엔드를 필요로 한다. 레트론 유래 DNA는 이러한 유형의 기술을 가능하게 할 수 있고, 본 발명자들이 이전에는 결코 달성되지 않았던 수준의 디테일로 복잡한 생물학적 프로세스를 이해할 수 있게 해준다.

[0283]

궁극적으로, 레트론을 길들이는 것은 게놈 조작을 목표로 하는 본 발명자들의 분자 생명공학의 레퍼토리를 확장시켜 준다. 레트론은 유전자 요법, 세포 제어, 및 세포 조작에 관심이 있는 과학자들에게 광범위하게 영향을

미치면서, 살아 있는 세포에서 임의의 DNA 서열을 맞춤형으로 만들기 위해 모듈식 구성 요소로 설계될 수 있다.

[0284]

참고문헌

- 1 Inouye, M. & Inouye, S. msDNA and bacterial reverse transcriptase. *Annual review of microbiology* **45**, 163-186, doi:10.1146/annurev.mi.45.100191.001115 (1991).
- 2 Farzadfard, F. & Lu, T. K. Synthetic biology. Genomically encoded analog memory with precise in vivo DNA writing in living cell populations. *Science* **346**, 1256272, doi:10.1126/science.1256272 (2014).
- 3 Sharon, E. *et al.* Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell* **175**, 544-557.e516, doi:10.1016/j.cell.2018.08.057 (2018).
- 4 Dhundale, A. R., Furuichi, T., Inouye, S. & Inouye, M. Distribution of multicopy single-stranded DNA among myxobacteria and related species. *Journal of bacteriology* **164**, 914-917 (1985).
- 5 Lampson, B. C. *et al.* Reverse transcriptase in a clinical strain of *Escherichia coli*: production of branched RNA-linked msDNA. *Science* **243**, 1033-1038 (1989).

[0285]

- 6 Inouye, K., Tanimoto, S., Kamimoto, M., Shimamoto, T. & Shimamoto, T. Two novel retron elements are replaced with retron-Vc95 in *Vibrio cholerae*. *Microbiology and immunology* **55**, 510-513, doi:10.1111/j.1348-0421.2011.00342.x (2011).
- 7 Lim, D. Structure and biosynthesis of unbranched multicopy single-stranded DNA by reverse transcriptase in a clinical *Escherichia coli* isolate. *Molecular microbiology* **6**, 3531-3542 (1992).
- 8 Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. CRISPR-Cas encoding of a digital movie into the genomes of a population of living bacteria. *Nature*, doi:10.1038/nature23017 (2017).
- 9 Shipman, S. L., Nivala, J., Macklis, J. D. & Church, G. M. Molecular recordings by directed CRISPR spacer acquisition. *Science*, doi:10.1126/science.aaf1175 (2016).

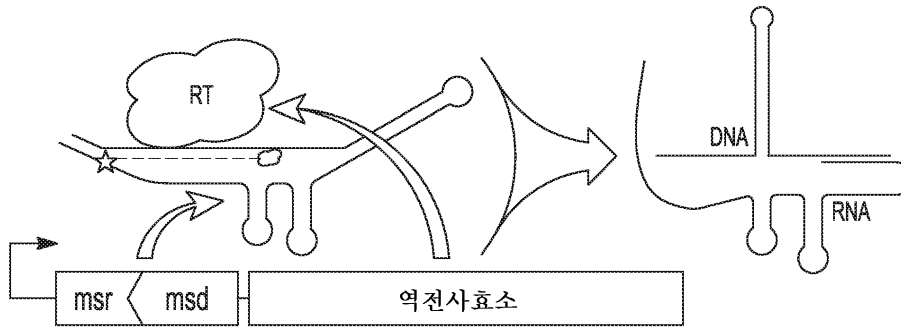
[0286]

[0287]

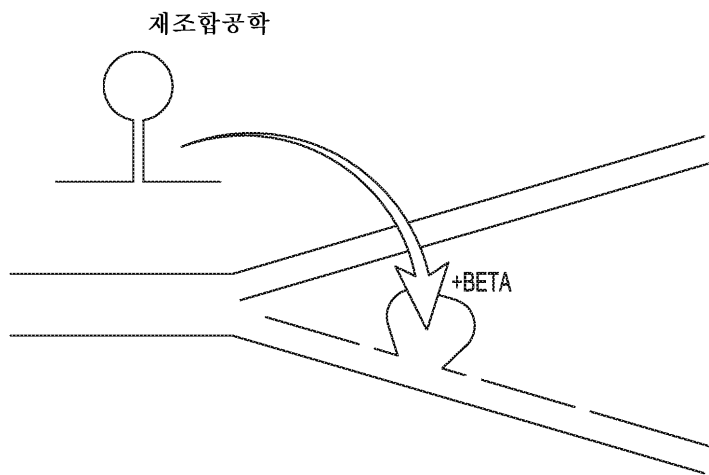
본 개시내용은 그의 구체적 실시양태를 참조하여 설명되었지만, 관련 기술분야의 통상의 기술자는 본 개시내용의 진정한 요지와 범위를 벗어나지 않고서도 다양한 변화가 이루어질 수 있으며 등가물이 대체될 수 있음을 이해해야 한다. 또한, 특정한 상황, 재료, 물질의 조성, 프로세스, 프로세스 단계 또는 단계들을 본 개시내용의 목적, 요지 및 범위에 맞게 조정하기 위해 많은 변형이 이루어질 수 있다. 이러한 모든 변형은 본원에 첨부된 청구범위의 범위 내에 포함되도록 의도된다.

도면

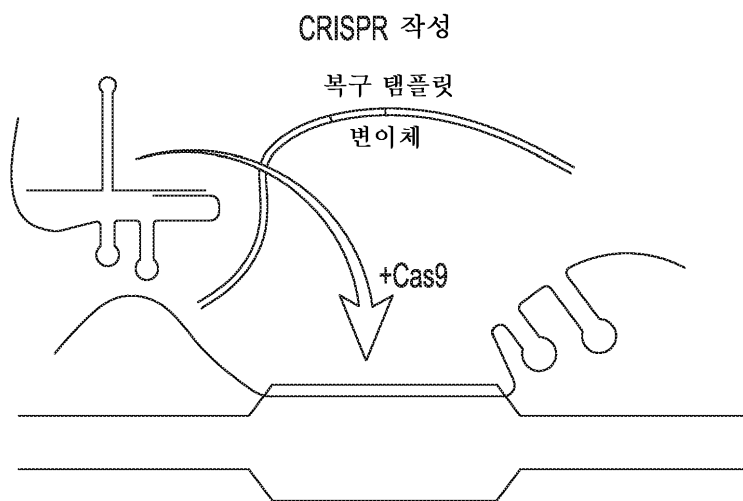
도면1a



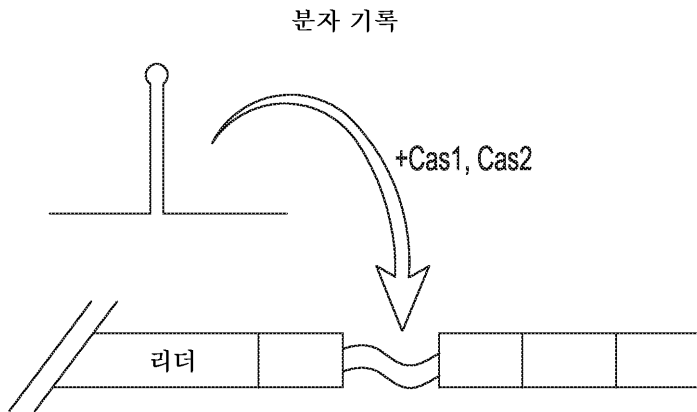
도면1b



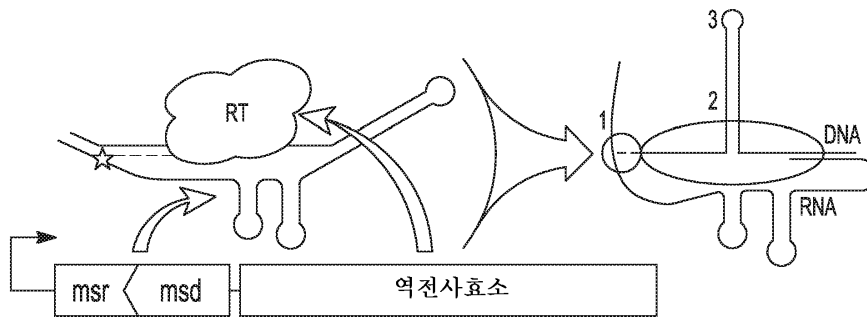
도면1c



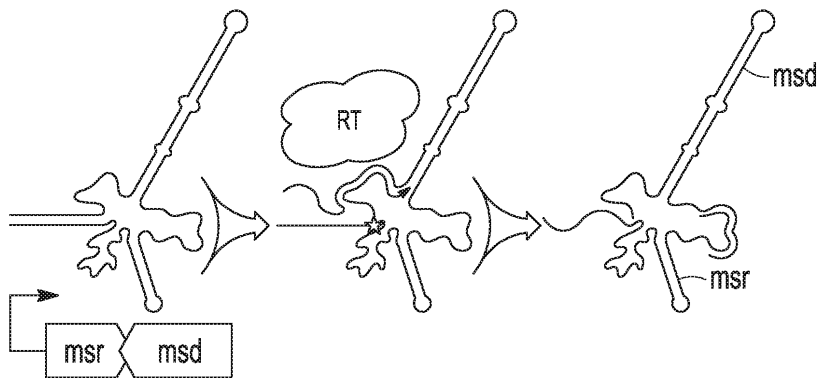
도면1d



도면2a

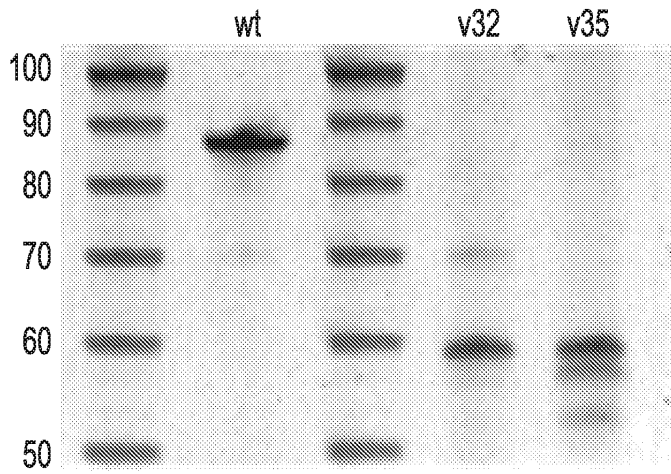


도면2b





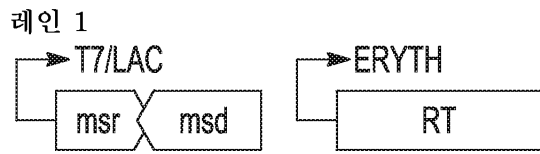
도면3c



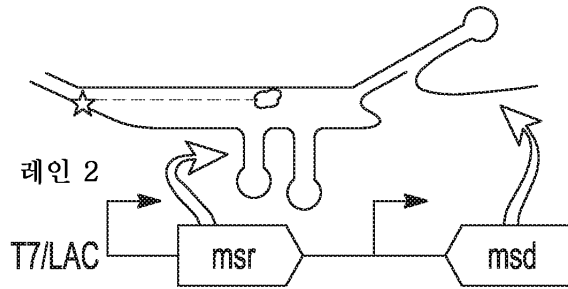
도면3d



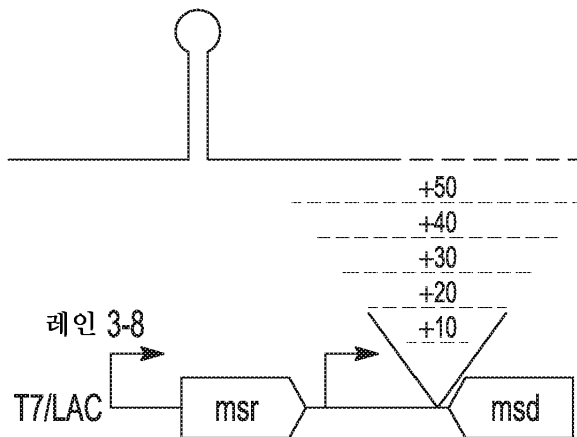
도면4a



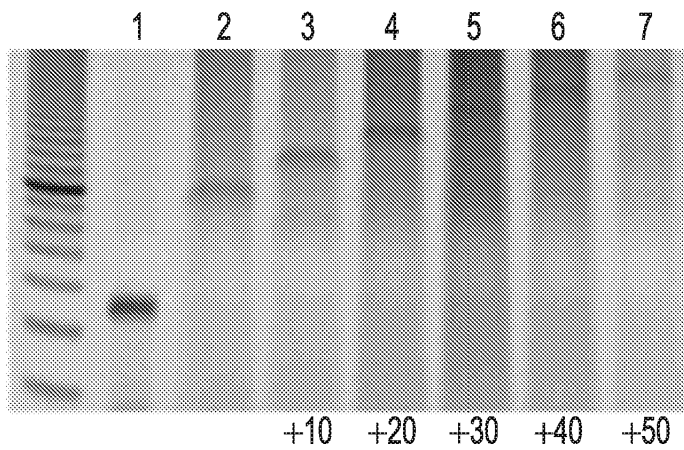
도면4b



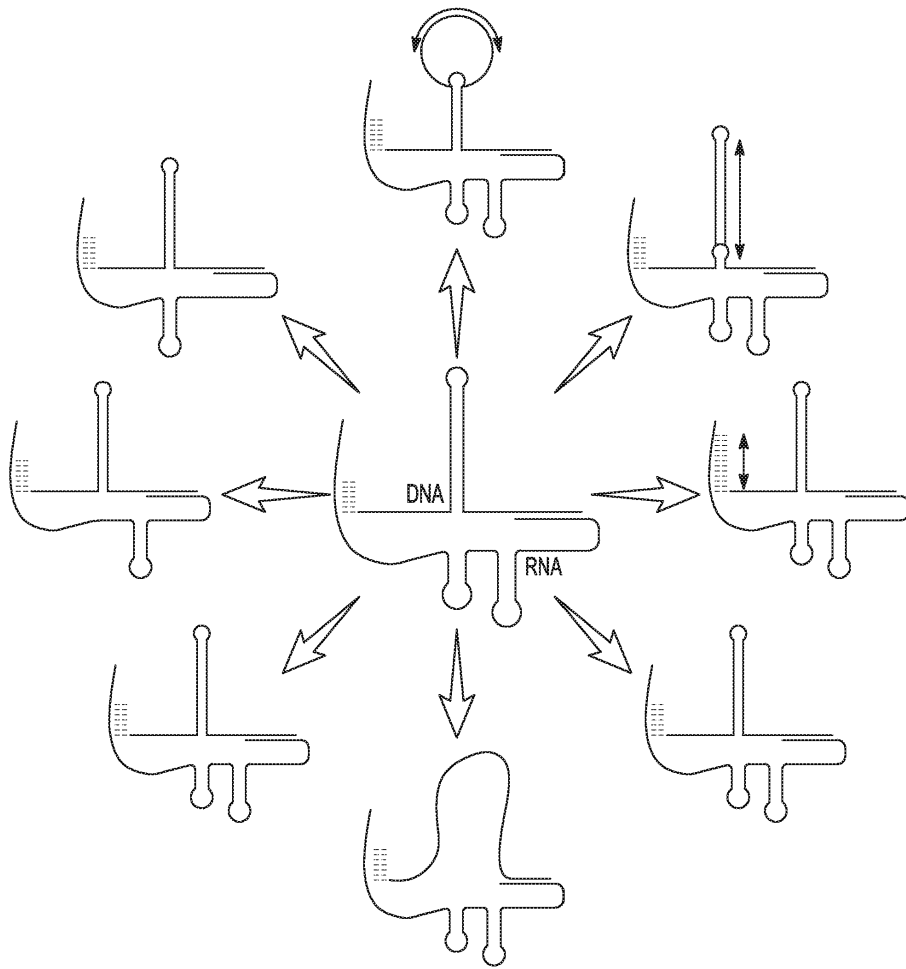
도면4c



도면4d



도면5





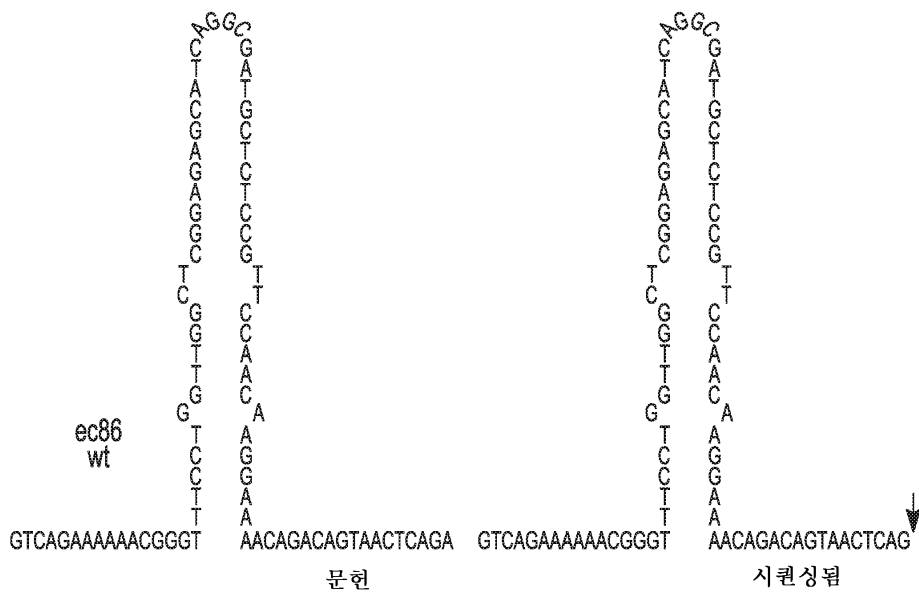
도면6c



도면6d

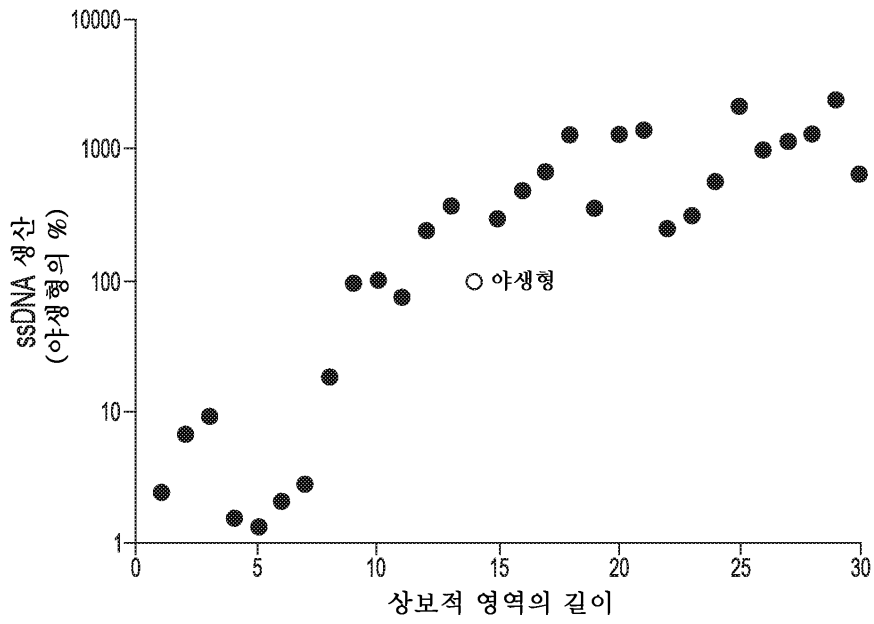


도면6e

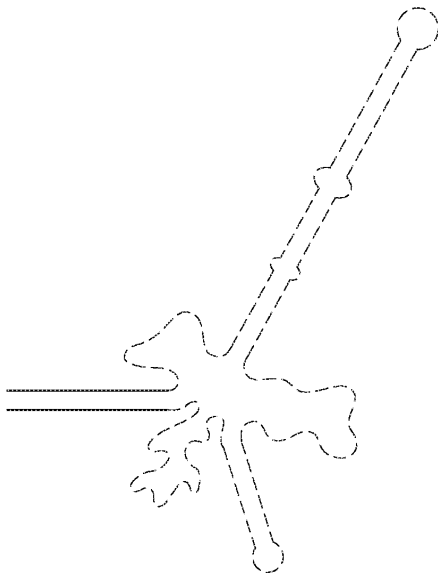




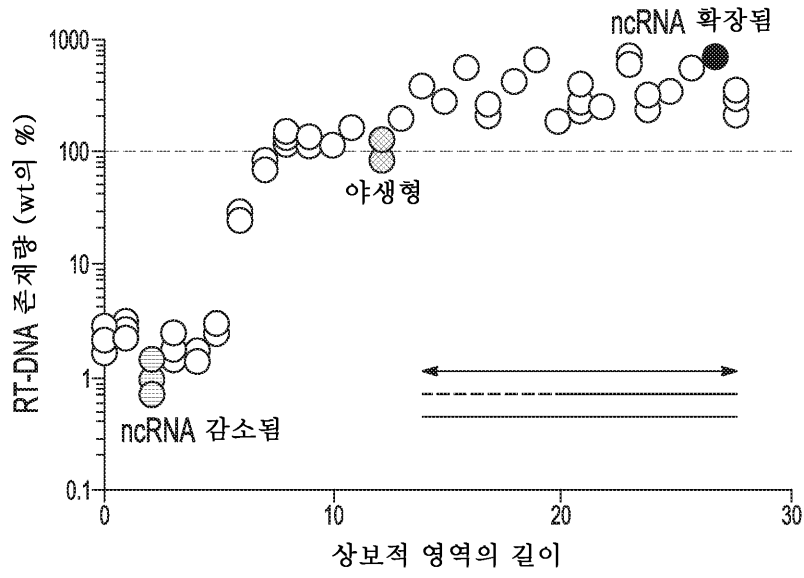
도면7b



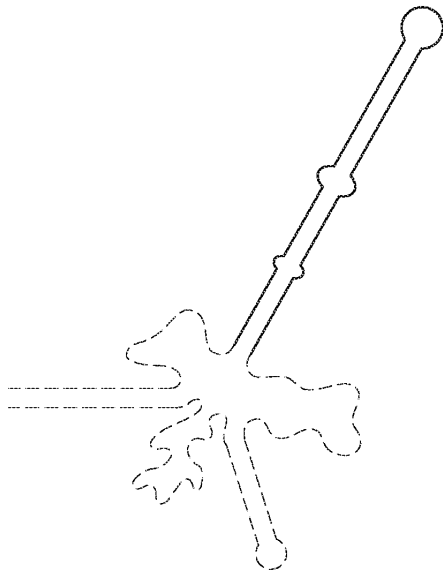
도면7ca



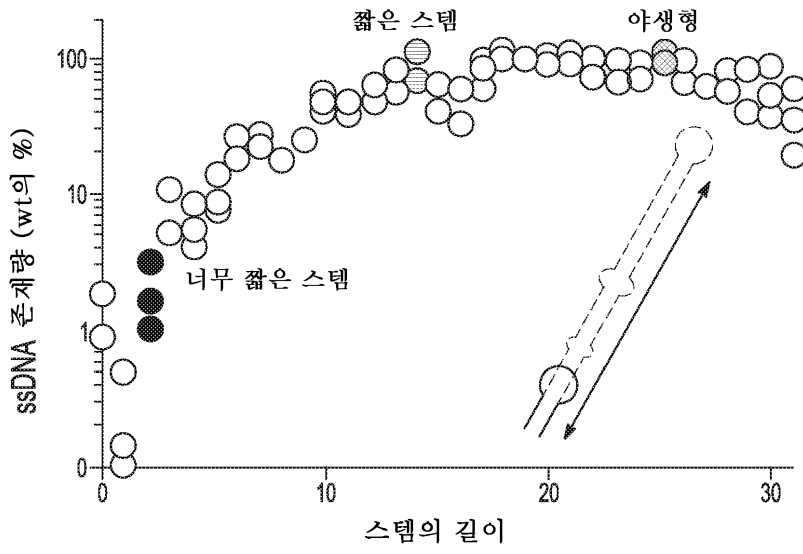
도면7cb



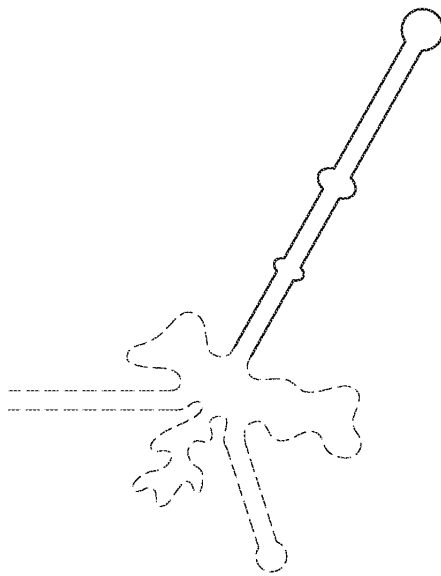
도면8a



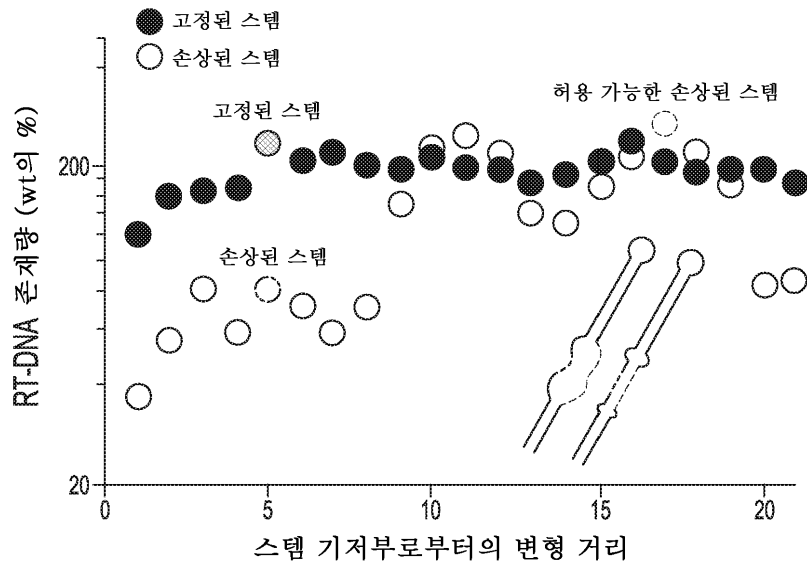
도면8b



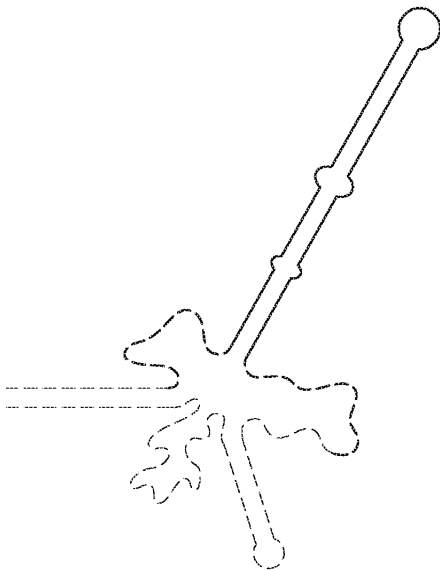
도면9a



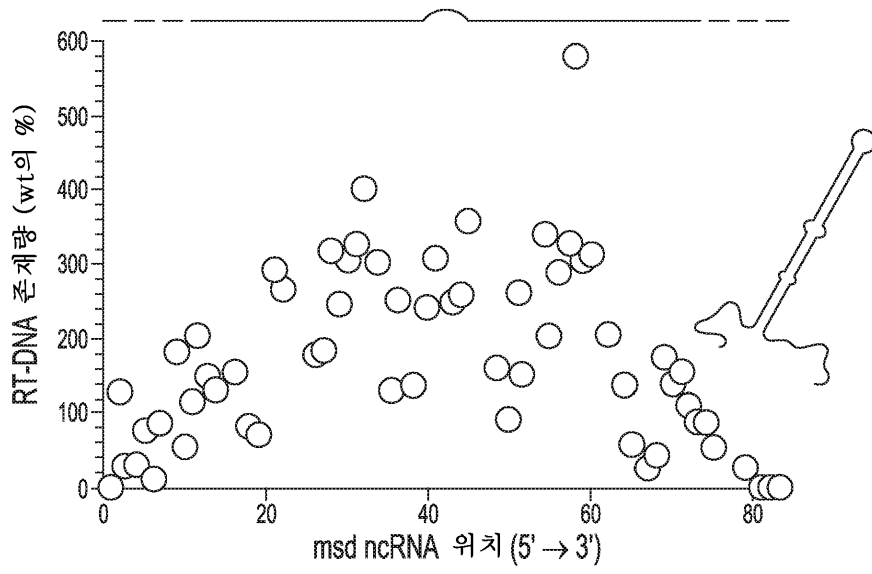
도면9b



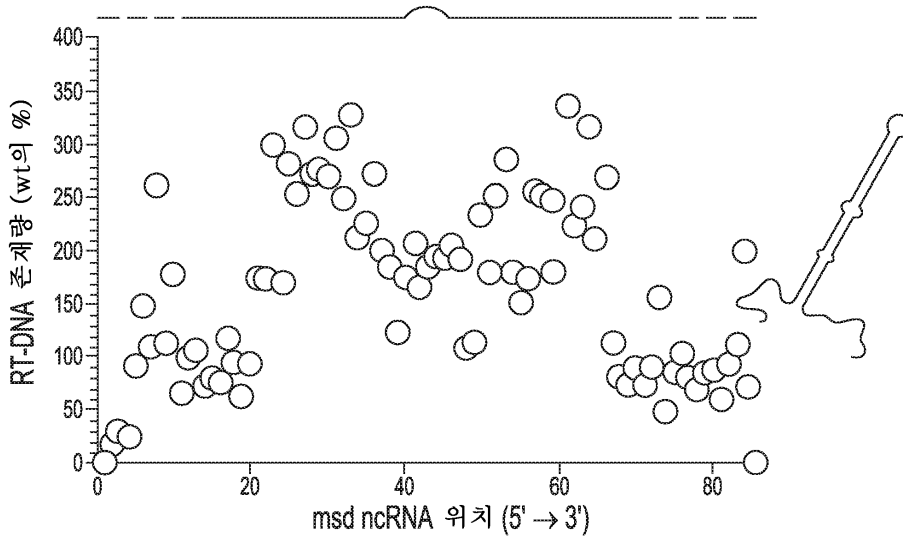
도면10a



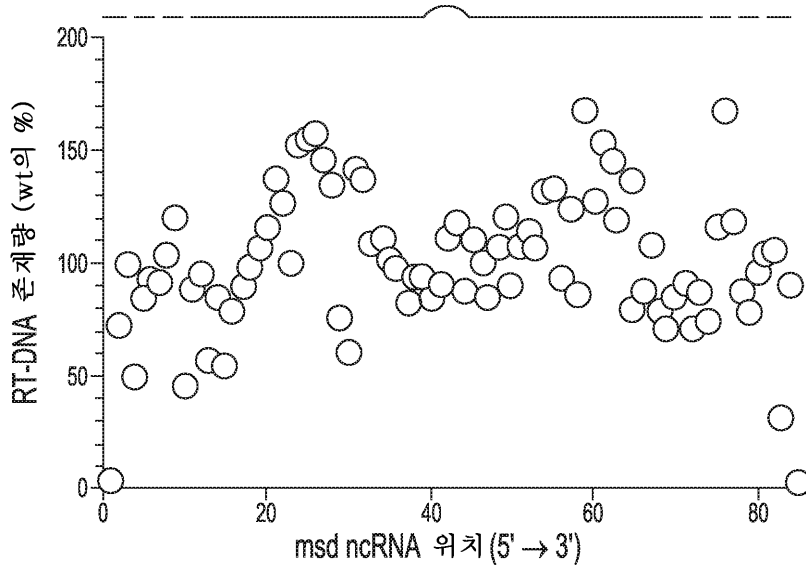
도면10b



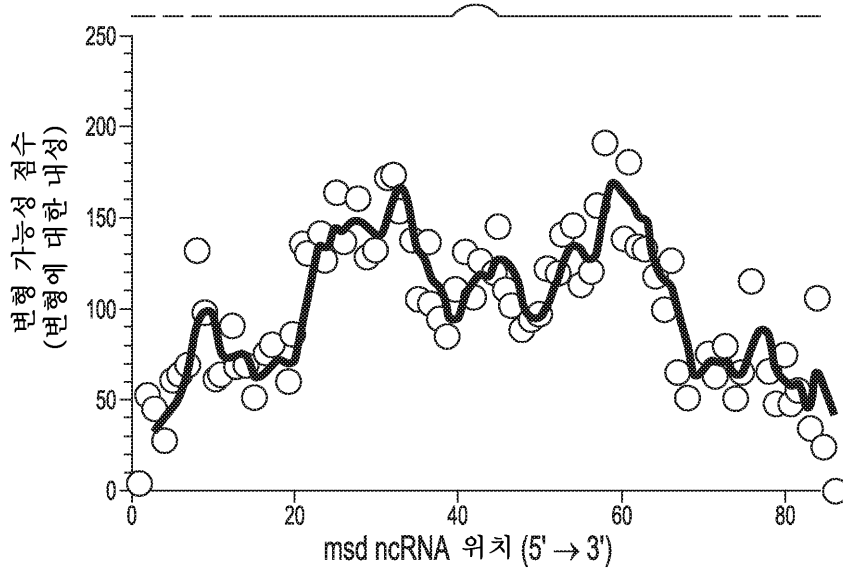
도면10c



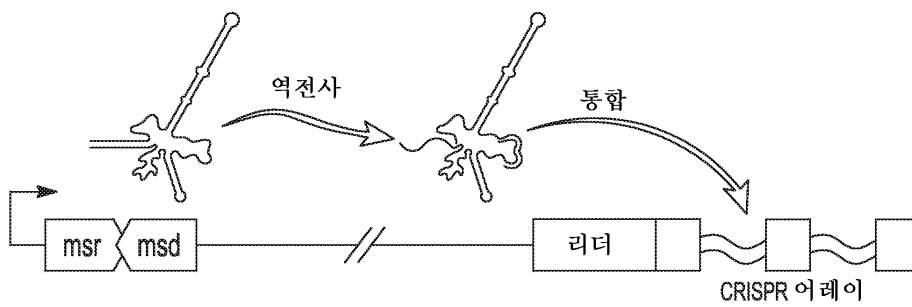
도면10d



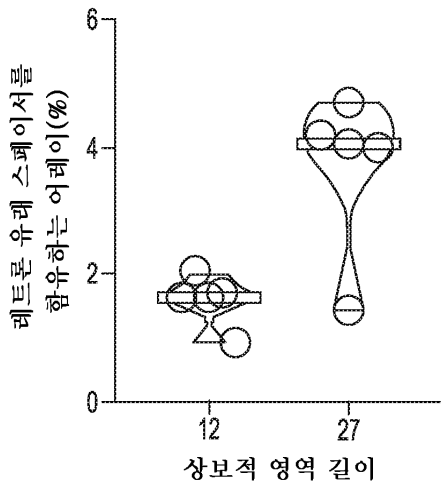
도면10e



도면11a



도면11b



서열목록

SEQUENCE LISTING

<110> Shipman, Seth

The J. David Gladstone Institutes, a testamentary trust

Established under the Will of J. David Gladstone

President and Fellows of Harvard College

<120> MODIFIED BACTERIAL RETROELEMENT WITH ENHANCED DNA PRODUCTION

<130> 3730.112W01

<150> US 62/899,625

<151> 2019-09-12

<160> 29

<170> FastSEQ for Windows Version 4.0

<210> 1

<211> 86

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 1

gtcagaaaa acgggtttcc tggttggctc ggagagcatc aggcgatgct ctccgttcca

60

|  |    |
|--|----|
| acaaggaaaa cagacagtaa ctcaga                                     | 86 |
| <210> 2  |    |
| <211> 69   |    |
| <212> RNA  |    |
| <213> Artificial Sequence  |    |
| <220>  |    |
| <223> A synthetic oligonucleotide                                |    |
| <400> 2  |    |
| augcgcaccc uuagcgagag guuuaucauu aagguaacc ucuggauguu guuucggcau | 60 |
| ccugcauug  | 69 |
| <210> 3  |    |
| <211> 57   |    |
| <212> DNA  |    |
| <213> Artificial Sequence  |    |
| <220>  |    |
| <223> A synthetic oligonucleotide                                |    |
| <400> 3  |    |
| gtcagaaaaa acgggttgtc gccagtctga ctggcgacaa acagcttgta actcaga   | 57 |
| <210> 4  |    |
| <211> 55   |    |
| <212> DNA  |    |
| <213> Artificial Sequence  |    |
| <220>  |    |
| <223> A synthetic oligonucleotide                                |    |
| <400> 4  |    |
| gtcagaaaaa acgggtggag gttgctgcaa cctctccatt ttcttgtaac tcaga     | 55 |
| <210> 5  |    |
| <211> 57   |    |
| <212> DNA  |    |
| <213> Artificial Sequence  |    |
| <220>  |    |
| <223> A synthetic oligonucleotide                                |    |
| <400> 5  |    |
| gtcagaaaaa acgggttgtc gccagtctga ctggcgacaa acagcttgta actcaga   | 57 |

<210> 6  
 <211> 57  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> A synthetic oligonucleotide

<400> 6  
 gtcagaaaaa acgggttgtc gccagtctga ctggcgacaa acagcttgta actcaga 57

<210> 7  
 <211> 56  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> A synthetic oligonucleotide

<400> 7  
 gtcagaaaaa acgggttgtc gccagtctga ctggcgacaa acagcttgta actcag 56

<210> 8  
 <211> 88  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> A synthetic oligonucleotide

<400> 8  
 gtcagaaaaa acgggtttcc tggttggctc ggagagcatc aggcatgct ctctccgttc 60

caacaaggaa aacagacagt aactcaga 88

<210> 9  
 <211> 87  
 <212> DNA  
 <213> Artificial Sequence

<220>  
 <223> A synthetic oligonucleotide

<400> 9  
 gtcagaaaaa acgggtttcc tggttggctc ggagagcatc aggcatgct ctctccgttc 60

caacaaggaa aacagacagt aactcag 87  
 <210> 10  
 <211> 83  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
 <400> 10  
 ttgaagccgc ggaacaaact tttgatccg caacctactg gattgcggt caaaaagttt 60  
  
 gttccgcaac tgtaaatgta atc 83  
 <210> 11  
 <211> 79  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
 <400> 11  
 agcccgaggaa caacttttt gatccgcaac ctactggatt gcggtcaaa aagtttgttc 60  
 cgcaactgta aatgtaatc 79  
 <210> 12  
 <211> 169  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
 <400> 12  
 tgccacacct tagcgagagg tttatcatta agtcaacct ctggatgttg ttcggcatc 60  
  
 ctgcattgaa tctgagttac tgtctgtttt ccttgttgga acggagagca tcgcctgatg 120  
 ctctccgagc caaccaggaa acccgttttt tctgacgtaa gggtagcga 169  
 <210> 13  
 <211> 199  
 <212> DNA  
 <213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 13

|   |     |
|---|-----|
| tgataagatt ccgtatgcgc acccttagcg agaggtttat cattaaggtc aacctctgga | 60  |
| tgttgtttcg gcacctgca ttgaatctga gttactgtct gtttccttg ttggaacgga   | 120 |
| gagcatcgcc tgatgctctc cgagccaacc aggaaaccg tttttctga cgtaagggtg   | 180 |
| cgcatacggga atcttatca   | 199 |

<210> 14

<211> 86

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 14

|   |    |
|---|----|
| gtcagaaaaa acgggtttcc tggttggctc ggagagcatc aggcgatgct ctccgttcca | 60 |
| acaaggaaaa cagacagtaa ctcaga                                      | 86 |

<210> 15

<211> 169

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 15

|  |     |
|--|-----|
| tgcgcacct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcgcatc    | 60  |
| ctgcattgaa tctgagttac tgtctgtttt ccttgttggga acggagagca tcgcctgatg | 120 |
| ctctccgagc caaccaggaa acccgttttt tctgacgtaa gggtagcga              | 169 |

<210> 16

<211> 97

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 16

tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc 60  
 ctgcattgaa tctgagttac aagctgtttg tcgccag 97  
 <210> 17  
 <211> 72  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
 <400> 17  
 aatctgagtt acaagctgtt tgctgccagt cagactggcg acaaccggtt tttctgacg 60  
  
 taagggtgcg ca 72  
 <210> 18  
 <211> 140  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
 <400> 18  
 tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc 60  
 ctgcattgaa tctgagttac aagctgtttg tcgccagtca gactggcgac aaccggtttt 120  
 ttctgacgta aggtgcgca 140  
 <210> 19  
 <211> 140  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
  
 <400> 19  
 tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc 60  
 ctgcattgaa tctgagttac aagctgtttg tcgccagtca gactggcgac aaccggtttt 120  
 ttctgacgta aggtgcgca 140  
 <210> 20  
 <211> 123

<212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
 <400> 20  
 tggacaatat tgaatggagt ctgatcaacc ttcacaccga tctagaatcg gaatctgagt 60  
 tacaagctgt ttgtcgccag tcagactggc gacaaccctg ttttctgac gtaagggtgc 120  
 gca 123

<210> 21  
 <211> 199  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
 <400> 21  
 tgataagatt ccgatgctc acccttagcg agaggtttat cattaaggtc aacctctgga 60  
 tgttgtttcg gcactctgca ttgaatctga gttactgtct gttttccttg ttggaacgga 120  
 gagcatcgcc tgatgctctc cgagccaacc aggaaaccg tttttctga cgtaagggtg 180  
 cgcatacggg atcttatca 199

<210> 22  
 <211> 170  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <223> A synthetic oligonucleotide  
 <400> 22  
 tgataagatt ccgatgctc acccttagcg agaggtttat cattaaggtc aacctctgga 60  
 tgttgtttcg gcactctgca ttgaatctga gttacaagct gtttgcgcc agtcagactg 120  
 gcgacaacc gtttttctg acgtaagggt gcgcatacgg aatcttatca 170

<210> 23  
 <211> 149  
 <212> DNA  
 <213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 23

|   |     |
|---|-----|
| tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc ctgcattgaa | 60  |
| tctgagttac tgtctgtttt ccttgttga acggagagca tcgcctgatg ctctccgagc  | 120 |
| caaccaggaa acccgttttt tctgacgta                                   | 149 |

<210> 24

<211> 147

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 24

|  |     |
|--|-----|
| tgccaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc | 60  |
| ctgcattgaa tctgagttac tgtctgtttt ccttgttga agcctagcca accaggaaac | 120 |
| ccgttttttc tgacgtaagg gtgcgca                                    | 147 |

<210> 25

<211> 123

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 25

|   |     |
|---|-----|
| tgccaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc  | 60  |
| ctgcattgaa tctgagttac tgtctgtttg cctaaccctg tttttctgac gtaagggtgc | 120 |
| gca   | 123 |

<210> 26

<211> 169

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 26

tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc 60  
 ctgcattgaa tctgagttac tgtctgtttt cgaacatgga acggagagca tgcctgatg 120  
 ctctccgagc caaccaggaa acccgttttt tctgacgtaa ggggtgcgca 169

<210> 27

<211

> 169

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 27

tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc 60  
 ctgcattgaa tctgagttac tgtctgtttt cgaacatgga acggagagca tgcctgatg 120  
 ctctccgagc catggtcgaa acccgttttt tctgacgtaa ggggtgcgca 169

<210> 28

<211> 169

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 28

tgcgcaccct tagcgagagg tttatcatta aggtcaacct ctggatgttg tttcggcatc 60  
  
 ctgcattgaa tctgagttac tgtctgtttt ccttgttggga acgctctcca tgcctgatg 120  
 ctctccgagc caaccaggaa acccgttttt tctgacgtaa ggggtgcgca 169

<210> 29

<211> 12

<212> DNA

<213> Artificial Sequence

<220>

<223> A synthetic oligonucleotide

<400> 29

ccccccccc cc 12