



(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2021년04월22일
(11) 등록번호 10-2243092
(24) 등록일자 2021년04월16일

- (51) 국제특허분류(Int. Cl.)
C12N 15/63 (2006.01) A61K 38/00 (2006.01)
C12N 15/10 (2017.01) C12N 15/85 (2006.01)
C12N 9/22 (2006.01)
- (52) CPC특허분류
C12N 15/63 (2013.01)
A61K 38/00 (2013.01)
- (21) 출원번호 10-2020-7023229(분할)
- (22) 출원일자(국제) 2013년12월05일
심사청구일자 2020년08월12일
- (85) 번역문제출일자 2020년08월12일
- (65) 공개번호 10-2020-0098727
- (43) 공개일자 2020년08월20일
- (62) 원출원 특허 10-2019-7022305
원출원일자(국제) 2013년12월05일
심사청구일자 2019년07월29일
- (86) 국제출원번호 PCT/US2013/073307
- (87) 국제공개번호 WO 2014/089290
국제공개일자 2014년06월12일
- (30) 우선권주장
61/734,256 2012년12월06일 미국(US)
(뒷면에 계속)
- (56) 선행기술조사문헌
US20100076057 A1

- (73) 특허권자
시그마-알드리치 컴퍼니., 엘엘씨
미국 미주리 63103 세인트 루이스 스프러스 스트리트 3050
- (72) 발명자
첸, 푸취양
미국 미주리 63103 세인트 루이스 스프러스 스트리트 3050
데이비스, 그레고리 디.
미국 미주리 63103 세인트 루이스 스프러스 스트리트 3050
- (74) 대리인
강명구, 이경민

전체 청구항 수 : 총 31 항

심사관 : 김정태

(54) 발명의 명칭 CRISPR-기초된 유전체 변형과 조절

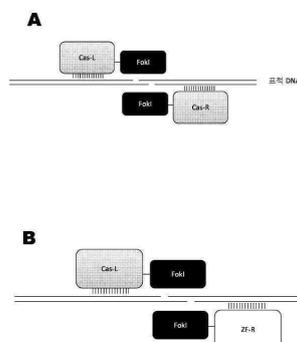
(57) 요약

요약

본 발명은 RNA-유도된 엔도뉴클레아제를 제공하고, 이것은 진핵 세포 또는 배아에서 발현을 위해 가공되고, 그리고 진핵 세포 또는 배아에서 표적화된 유전체 변형을 위해 RNA-유도된 엔도뉴클레아제를 이용하는 방법을 제공한

(뒷면에 계속)

대표도 - 도1



다. 융합 단백질 역시 제공되고, 여기서 각 융합 단백질은 CRISPR/Cas-유사 단백질 또는 이의 단편 및 작동체 도메인을 포함한다. 작동체 도메인은 개열 도메인, 후성 변형 도메인, 전사 활성화 도메인, 또는 전사 억제인자 도메인일 수 있다. 융합 단백질을 이용하여 염색체 서열을 변경하거나 또는 염색체 서열의 발현을 조절하기 위한 방법 역시 제공된다.

(52) CPC특허분류

C12N 15/102 (2013.01)

C12N 15/85 (2013.01)

C12N 9/22 (2013.01)

(30) 우선권주장

61/758,624 2013년01월30일 미국(US)

61/761,046 2013년02월05일 미국(US)

61/794,422 2013년03월15일 미국(US)

명세서

청구범위

청구항 1

다음을 포함하는 벡터:

(a) 진핵 세포에서 최소한 하나의 유도 RNA의 발현을 위해 프로모터 제어 서열에 작동가능하게 연결된 최소한 하나의 유도 RNA를 인코딩하는 DNA 코딩 서열, 이 때 각 유도 RNA는

(i) 표적 부위와 염기쌍을 이룰 수 있는 진핵세포 염색체 서열 내 표적 부위에 상보적인 첫 번째 영역,

(ii) 줄기 및 루프 구조를 형성하는 두 번째 영역, 및

(iii) 실질적으로 단일 가닥인 세 번째 영역

을 포함하고,

이 때, 상기 (i), (ii), 및 (iii)은 5' 에서 3' 방향으로 배열되고, 그리고

(b) 진핵 세포에서의 발현을 위해 프로모터 제어 서열에 작동적으로 연결된 조작된 RNA-유도된 엔도뉴클레아제를 인코딩하는 DNA 코딩 서열, 이 때 RNA-유도된 엔도뉴클레아제는 최소한 하나의 핵 국지화 신호를 포함하는 유형 II CRISPR/Cas9 엔도뉴클레아제이고,

이 때 (a) 및 (b)는 동일 또는 상이한 벡터에 위치하고,

이에 의해, 상기 RNA-유도된 엔도뉴클레아제는 최소한 하나의 유도 RNA에 의해 특정 핵산 서열에 지향되고 상기 RNA-유도된 엔도뉴클레아제는 상기 특정 핵산 서열을 절단함으로써, 상기 핵산 서열은 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환 또는 이의 조합에 의해 변형되는, 벡터.

청구항 2

청구항 1에 있어서, 상기 유도 RNA는 2개의 별개의 분자들을 포함함을 특징으로 하는, 벡터.

청구항 3

청구항 2에 있어서, 상기 유도 RNA의 첫 번째 분자는 유도 RNA의 첫 번째 영역 및 유도 RNA의 두 번째 영역의 줄기의 한 쪽 절반을 포함함을 특징으로 하는, 벡터.

청구항 4

청구항 2에 있어서, 상기 유도 RNA의 두 번째 분자는 유도 RNA의 두 번째 영역의 줄기의 다른 쪽 절반 및 유도 RNA의 세 번째 영역을 포함함을 특징으로 하는, 벡터.

청구항 5

청구항 1-4 중 어느 한 항에 있어서, 상기 유도 RNA의 첫 번째 영역은 약 10 뉴클레오티드 내지 약 25개 보다 많은 뉴클레오티드를 포함함을 특징으로 하는, 벡터.

청구항 6

청구항 1-4 중 어느 한 항에 있어서, 상기 유도 RNA의 두 번째 영역은 약 16 내지 약 60개 뉴클레오티드 길이임을 특징으로 하는, 벡터.

청구항 7

청구항 1-4 중 어느 한 항에 있어서, 상기 유도 RNA의 세 번째 영역은 약 5 내지 약 60개 뉴클레오티드 길이임을 특징으로 하는, 벡터.

청구항 8

청구항 1-4 중 어느 한 항에 있어서, 상기 유형 II CRISPR/Cas9 단백질은 단 하나의 기능적 뉴클레아제 도메인을 포함함을 특징으로 하는, 벡터.

청구항 9

청구항 8에 있어서, 상기 유형 II CRISPR/Cas9 단백질은 비기능적 RuvC-유사 도메인을 포함함을 특징으로 하는, 벡터.

청구항 10

청구항 8에 있어서, 상기 유형 II CRISPR/Cas9 단백질은 비기능적 HNH-유사 도메인을 포함함을 특징으로 하는, 벡터.

청구항 11

청구항 1-4 중 어느 한 항에 있어서, 상기 유형 II CRISPR/Cas9 단백질은 스트렙토코쿠스 종에서 유래함을 특징으로 하는, 벡터.

청구항 12

청구항 11에 있어서, 상기 유형 II CRISPR/Cas9 단백질은 스트렙토코쿠스 피오게네스에서 유래함을 특징으로 하는, 벡터.

청구항 13

청구항 1-4 중 어느 한 항에 있어서, NLS는 RNA-유도된 엔도뉴클레아제의 C-말단에 위치함을 특징으로 하는, 벡터.

청구항 14

청구항 1-4 중 어느 한 항에 있어서, 유형 II CRISPR/Cas9 단백질은 진핵세포 발현에 최적화된 DNA에 의해 인코딩됨을 특징으로 하는, 벡터.

청구항 15

청구항 1-4 중 어느 한 항에 있어서, 벡터들은 추가 발현 제어 서열들을 포함함을 특징으로 하는, 벡터.

청구항 16

청구항 1-4 중 어느 한 항에 있어서, 상기 유도 RNA 및 유형 II CRISPR/Cas9 단백질은 동일한 벡터에서 인코딩됨을 특징으로 하는, 벡터.

청구항 17

청구항 1-4 중 어느 한 항에 있어서, 상기 표적 부위는 프로토스페이서 인접 모티프 (PAM)의 상류에 존재함을 특징으로 하는, 벡터.

청구항 18

청구항 17에 있어서, 상기 PAM은 표적 부위의 바로 하류에 존재함을 특징으로 하는, 벡터.

청구항 19

청구항 17에 있어서, 상기 PAM은 NGG 또는 NGGNG이고, 이 때 N은 임의의 뉴클레오티드로 정의됨을 특징으로 하는, 벡터.

청구항 20

청구항 1-4 중 어느 한 항에 있어서, 상기 벡터들은 플라스미드 벡터들임을 특징으로 하는, 벡터.

청구항 21

청구항 1-4 중 어느 한 항에 있어서, 상기 벡터들은 바이러스 벡터임을 특징으로 하는, 벡터.

청구항 22

청구항 1-4 중 어느 한 항에 있어서, 상기 RNA-유도된 엔도뉴클레아제는 진핵세포 염색체 서열 내 표적 부위에 이중 가닥 절단을 도입하고, 이러한 절단은 DNA 복구 과정에 의해 복구됨으로써 상기 서열을 변형시킴을 특징으로 하는, 벡터.

청구항 23

청구항 1-4 중 어느 한 항에 있어서, 공여자 폴리뉴클레오티드를 추가로 포함함을 특징으로 하는, 벡터.

청구항 24

진핵 세포의 염색체 서열을 변형하는 방법으로서, 청구항 1-4 중 어느 한 항의 벡터, 및 선택적으로, 공여자 서열을 포함하는 최소한 하나의 공여자 폴리뉴클레오티드를 진핵 세포와 함께 배양하는 단계를 포함하고, 이 때 상기 방법은 인간의 생식 계열 유전자 동일성을 변형하기 위한 과정을 포함하지 않고 수술 또는 치료에 의한 인간 또는 동물 신체의 치료를 포함하지 않는, 방법.

청구항 25

청구항 24에 있어서, 상기 염색체 서열은 DNA 복구 과정에 의해 복구되어, 염색체 서열이 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이의 조합에 의해 변형됨을 특징으로 하는, 방법.

청구항 26

청구항 25에 있어서, 상기 DNA 복구 과정은 비상동성 말단 연결 (NHEJ) 복구 과정임을 특징으로 하는, 방법.

청구항 27

청구항 25에 있어서, 상류 및 하류 서열들을 포함하는 공여자 폴리뉴클레오티드를 추가로 포함하고, 이 때 상기 DNA 복구 과정은 상동 재조합 (HDR)이고, 공여자 서열은 염색체 서열내 통합됨을 특징으로 하는, 방법.

청구항 28

청구항 24에 있어서, 상기 염색체 서열은 진핵 세포에 존재하고, 이 때 진핵 세포는 인간 세포, 비-인간 포유류 세포, 또는 비포유류 척추동물 세포임을 특징으로 하는, 방법.

청구항 29

청구항 24에 있어서, 상기 염색체 서열은 진핵 세포에 존재하고, 이 때 진핵 세포는 무척추동물 세포, 곤충 세포, 식물 세포, 효모 세포 또는 단일 세포 진핵 생물체임을 특징으로 하는, 방법.

청구항 30

청구항 29에 있어서, 상기 진핵 세포는 식물 세포임을 특징으로 하는, 방법.

청구항 31

청구항 24에 있어서, 상기 염색체 서열은 비-인간 포유류 배아에 존재함을 특징으로 하는, 방법.

발명의 설명

기술 분야

발명의 분야

본 발명은 표적화된 유전체 변형에 관계한다. 특히, 본 발명은 CRISPR/Cas-유사 단백질을 포함하는 RNA-유도된

[0001]

[0002]

엔도뉴클레아제 또는 융합 단백질 및 표적화된 염색체 서열을 변경하거나 또는 조절하기 위해 상기 단백질을 이용하는 방법에 관계한다.

배경 기술

[0003] 발명의 배경

[0004] 표적화된 유전체 변형은 진핵 세포, 배아, 그리고 동물의 유전자 조작을 위한 강력한 도구이다. 가령, 외인성 서열이 표적화된 유전체학 위치에서 통합될 수 있고 및/또는 특정한 내인성 염색체 서열이 결실되거나, 불활성화되거나, 또는 변형될 수 있다. 현재 방법은 가공된 뉴클레아제 효소, 예를 들면, 예로서, 아연 핑거 뉴클레아제 (ZFNs) 또는 전사 활성화제-유사 작동체 뉴클레아제 (TALENs)의 이용에 의존한다. 이들 키메라 뉴클레아제는 비특이적 DNA 개열 도메인에 연결된 프로그램가능, 서열-특정한 DNA-결합 모듈을 내포한다. 각 새로운 유전체학 표적은 하지만, 신규한 서열-특이적 DNA-결합 모듈을 포함하는 새로운 ZFN 또는 TALEN의 설계를 필요로 한다. 따라서, 이들 맞춤 설계된 뉴클레아제는 준비하는데 값비싸고 시간 소모적인 경향이 있다. 게다가, ZFNs와 TALENs의 특이성은 그들이 부정확한 개열을 매개할 수 있는 정도이다.

[0005] 따라서, 각 새로운 표적화된 유전체학 위치에 대한 새로운 뉴클레아제의 설계를 필요로 하지 않는 표적화된 유전체 변형 기술이 필요하다. 부가적으로, 부정확한 효과가 거의 또는 전혀 없는 증가된 특이성을 갖는 기술이 필요하다.

발명의 내용

[0006] 발명의 요약

[0007] 본 발명의 다양한 양상에는 단리된 RNA-유도된 엔도뉴클레아제의 제공이 포함되고, 여기서 엔도뉴클레아제는 최소한 하나의 핵 국지화 신호, 최소한 하나의 뉴클레아제 도메인, 그리고 엔도뉴클레아제를 개열을 위한 특정한 뉴클레오티드 서열로 표적화하기 위한 유도 RNA와 상호작용하는 최소한 하나의 도메인을 포함한다. 한 구체예에서, 엔도뉴클레아제는 Cas9 단백질로부터 유래될 수 있다. 다른 구체예에서, 엔도뉴클레아제는 최소한 하나의 기능적 뉴클레아제 도메인을 결합하도록 변형될 수 있다. 다른 구체예에서, 엔도뉴클레아제는 세포-투과성 도메인, 마커 도메인, 또는 둘 모두를 더욱 포함할 수 있다. 추가의 구체예에서, 엔도뉴클레아제는 유도 RNA를 포함하는 단백질-RNA 복합체의 부분일 수 있다. 일부 경우에, 유도 RNA는 표적 부위에 상보적인 5' 영역을 포함하는 단일 분자일 수 있다. 본원에서 개시된 RNA-유도된 엔도뉴클레아제 중에서 한 가지를 인코딩하는 단리된 핵산 역시 제공된다. 일부 구체예에서, 핵산은 포유류 세포, 예를 들면, 예로서, 인간 세포에서 번역을 위해 코돈 최적화될 수 있다. 다른 구체예에서, RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산 서열은 프로모터 제어 서열에 작동가능하게 연결될 수 있고, 그리고 임의선택적으로, 백터의 부분일 수 있다. 다른 구체예에서, 프로모터 제어 서열에 작동가능하게 연결될 수 있는, RNA-유도된 엔도뉴클레아제를 인코딩하는 서열을 포함하는 백터는 프로모터 제어 서열에 작동가능하게 연결될 수 있는, 유도 RNA를 인코딩하는 서열을 또한 포함할 수 있다.

[0008] 본 발명의 다른 양상은 진핵 세포 또는 배아에서 염색체 서열을 변경하기 위한 방법을 포괄한다. 상기 방법은 (i) 최소한 하나의 핵 국지화 신호를 포함하는 최소한 하나의 RNA-유도된 엔도뉴클레아제 또는 본원에서 규정된 바와 같은 최소한 하나의 RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산, (ii) 최소한 하나의 유도 RNA 또는 최소한 하나의 유도 RNA를 인코딩하는 DNA, 그리고, 임의선택적으로, (iii) 공여자 서열을 포함하는 최소한 하나의 공여자 폴리뉴클레오티드를 진핵 세포 또는 배아 내로 도입하는 것을 포함한다. 상기 방법은 각 유도 RNA가 염색체 서열 내에 표적화된 부위에 RNA-유도된 엔도뉴클레아제를 향하게 하도록 상기 세포 또는 배아를 배양하는 것을 더욱 포함하고, 여기서 상기 RNA-유도된 엔도뉴클레아제는 표적화된 부위 내에 이중 가닥 절단을 도입하고, 그리고 상기 이중 가닥 절단은 염색체 서열이 변형되도록 DNA 복구 과정에 의해 복구된다. 한 구체예에서, RNA-유도된 엔도뉴클레아제는 Cas9 단백질로부터 유래될 수 있다. 다른 구체예에서, 세포 또는 배아 내로 도입된 RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산은 mRNA일 수 있다. 추가의 구체예에서, 세포 또는 배아 내로 도입된 RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산은 DNA일 수 있다. 추가의 구체예에서, RNA-유도된 엔도뉴클레아제를 인코딩하는 DNA는 유도 RNA를 인코딩하는 서열을 더욱 포함하는 백터의 부분일 수 있다. 일정한 구체예에서, 진핵 세포는 인간 세포, 비-인간 포유류 세포, 줄기 세포, 비포유류 척추동물 세포, 무척추동물 세포, 식물 세포, 또는 단일 세포 진핵 생물체일 수 있다. 일정한 다른 구체예에서, 배아는 비-인간 단세포 동물 배아일 수 있다.

[0009] 본 발명의 추가 양상은 CRISPR/Cas-유사 단백질 또는 이의 단편 및 작동체 도메인을 포함하는 융합 단백질을 제

공한다. 일반적으로, 융합 단백질은 최소한 하나의 핵 국지화 신호를 포함한다. 융합 단백질의 작동체 도메인은 개별 도메인, 후성 변형 도메인, 전사 활성화 도메인, 또는 전사 억제물질 도메인일 수 있다. 한 구체예에서, 융합 단백질의 CRISPR/Cas-유사 단백질은 Cas9 단백질로부터 유래될 수 있다. 한 가지 반복에서, Cas9 단백질은 최소한 하나의 기능적 뉴클레아제 도메인을 결합하도록 변형될 수 있다. 교체 반복에서, Cas9 단백질은 모든 뉴클레아제 활성을 결합하도록 변형될 수 있다. 한 구체예에서, 작동체 도메인은 개별 도메인, 예를 들면, 예로서, FokI 엔도뉴클레아제 도메인 또는 변형된 FokI 엔도뉴클레아제 도메인일 수 있다. 다른 구체예에서, 한 융합 단백질은 다른 융합 단백질과 이합체를 형성할 수 있다. 이합체는 동종이합체 또는 이형이합체일 수 있다. 다른 구체예에서, 융합 단백질은 아연 핑거 뉴클레아제와 이형이합체를 형성할 수 있고, 여기서 융합 단백질과 아연 핑거 뉴클레아제 둘 모두의 개별 도메인은 FokI 엔도뉴클레아제 도메인 또는 변형된 FokI 엔도뉴클레아제 도메인이다. 또 다른 구체예에서, 융합 단백질은 모든 뉴클레아제 활성을 결합하도록 변형된 Cas9 단백질로부터 유래된 CRISPR/Cas-유사 단백질을 포함하고, 그리고 작동체 도메인은 FokI 엔도뉴클레아제 도메인 또는 변형된 FokI 엔도뉴클레아제 도메인이다. 또 다른 구체예에서, 융합 단백질은 모든 뉴클레아제 활성을 결합하도록 변형된 Cas9 단백질로부터 유래된 CRISPR/Cas-유사 단백질을 포함하고, 그리고 작동체 도메인은 후성 변형 도메인, 전사 활성화 도메인, 또는 전사 억제물질 도메인일 수 있다. 추가의 구체예에서, 본원에서 개시된 융합 단백질 중에서 한 가지는 핵 국지화 신호, 세포-투과성 도메인, 그리고 마커 도메인에서 선택된 최소한 하나의 추가 도메인을 포함할 수 있다. 본원에서 제공된 융합 단백질 중에서 한 가지를 인코딩하는 단리된 핵산 역시 제공된다.

[0010] 본 발명의 또 다른 양상은 세포 또는 배아에서 염색체 서열을 변경하거나 또는 염색체 서열의 발현을 조절하기 위한 방법을 포괄한다. 상기 방법은 다음을 세포 또는 배아 내로 도입하는 것을 포함한다: (a) 최소한 하나의 융합 단백질 또는 최소한 하나의 융합 단백질을 인코딩하는 핵산, 여기서 상기 융합 단백질은 CRISPR/Cas-유사 단백질 또는 이의 단편 및 작동체 도메인을 포함하고, 그리고 (b) 최소한 하나의 유도 RNA 또는 최소한 하나의 유도 RNA를 인코딩하는 DNA, 여기서 유도 RNA는 융합 단백질의 CRISPR/Cas-유사 단백질을 염색체 서열 내에 표적화된 부위로 유도하고, 그리고 융합 단백질의 작동체 도메인은 염색체 서열을 변경하거나 또는 염색체 서열의 발현을 조절한다. 한 구체예에서, 융합 단백질의 CRISPR/Cas-유사 단백질은 Cas9 단백질로부터 유래될 수 있다. 다른 구체예에서, 융합 단백질의 CRISPR/Cas-유사 단백질은 최소한 하나의 기능적 뉴클레아제 도메인을 결합하도록 변형될 수 있다. 또 다른 구체예에서, 융합 단백질의 CRISPR/Cas-유사 단백질은 모든 뉴클레아제 활성을 결합하도록 변형될 수 있다. 융합 단백질이 모든 뉴클레아제 활성을 결합하도록 변형된 Cas9 단백질 및 FokI 개별 도메인 또는 변형된 FokI 개별 도메인을 포함하는 한 구체예에서, 상기 방법은 1개의 융합 단백질 또는 1개의 융합 단백질을 인코딩하는 핵산 및 2개의 유도 RNA 또는 2개의 유도 RNA를 인코딩하는 DNA를 세포 또는 배아 내로 도입하는 것을 포함할 수 있고, 그리고 여기서 1개의 이중 가닥 절단이 염색체 서열 내에 도입된다. 융합 단백질이 모든 뉴클레아제 활성을 결합하도록 변형된 Cas9 단백질 및 FokI 개별 도메인 또는 변형된 FokI 개별 도메인을 포함하는 다른 구체예에서, 상기 방법은 2개의 융합 단백질 또는 2개의 융합 단백질을 인코딩하는 핵산 및 2개의 유도 RNA 또는 2개의 유도 RNA를 인코딩하는 DNA를 세포 또는 배아 내로 도입하는 것을 포함할 수 있고, 그리고 여기서 2개의 이중 가닥 절단이 염색체 서열 내에 도입된다. 융합 단백질이 모든 뉴클레아제 활성을 결합하도록 변형된 Cas9 단백질 및 FokI 개별 도메인 또는 변형된 FokI 개별 도메인을 포함하는 또 다른 구체예에서, 상기 방법은 1개의 융합 단백질 또는 1개의 융합 단백질을 인코딩하는 핵산, 1개의 유도 RNA 또는 1개의 유도 RNA를 인코딩하는 핵산, 그리고 1개의 아연 핑거 뉴클레아제 또는 1개의 아연 핑거 뉴클레아제를 인코딩하는 핵산을 세포 또는 배아 내로 도입하는 것을 포함할 수 있고, 여기서 상기 아연 핑거 뉴클레아제는 FokI 개별 도메인 또는 변형된 FokI 개별 도메인을 포함하고, 그리고 여기서 1개의 이중 가닥 절단이 염색체 서열 내로 도입된다. 융합 단백질이 개별 도메인을 포함하는 일정한 구체예에서, 상기 방법은 최소한 하나의 공여 자 폴리뉴클레오티드를 세포 또는 배아 내로 도입하는 것을 더욱 포함할 수 있다. 융합 단백질이 후성 변형 도메인, 전사 활성화 도메인, 또는 전사 억제물질 도메인에서 선택되는 작동체 도메인을 포함하는 구체예에서, 상기 융합 단백질은 모든 뉴클레아제 활성을 결합하도록 변형된 Cas9 단백질을 포함할 수 있고, 그리고 상기 방법은 1개의 융합 단백질 또는 1개의 융합 단백질을 인코딩하는 핵산, 그리고 1개의 유도 RNA 또는 1개의 유도 RNA를 인코딩하는 핵산을 세포 또는 배아 내로 도입하는 것을 포함할 수 있고, 그리고 여기서 표적화된 염색체 서열의 구조 또는 발현이 변형된다. 일정한 구체예에서, 진핵 세포는 인간 세포, 비-인간 포유류 세포, 줄기 세포, 비포유류 척추동물 세포, 무척추동물 세포, 식물 세포, 또는 단일 세포 진핵 생물체일 수 있다. 일정한 다른 구체예에서, 배아는 비-인간 단세포 동물 배아이다.

[0011] 본 발명의 다른 양상과 반복이 아래에서 상술된다.

도면의 간단한 설명

[0012] 도면의 간단한 설명

도면 1은 단백질 이합체를 이용한 유전체 변형을 도해한다. **(A)**는 DNA 결합을 위한 Cas-유사 단백질 및 FokI 개열 도메인을 각각 포함하는 2개의 융합 단백질로 구성된 이합체에 의해 창출된 이중 가닥 절단을 묘사한다. **(B)**는 Cas-유사 단백질과 FokI 개열 도메인을 포함하는 융합 단백질 및 아연 핑거 (ZF) DNA-결합 도메인과 FokI 개열 도메인을 포함하는 아연 핑거 뉴클레아제로 구성된 이합체에 의해 창출된 이중 가닥 절단을 묘사한다.

도면 2는 유전자 조절 도메인을 포함하는 RNA-유도된 융합 단백질을 이용한 유전자 발현의 조절을 예증한다. **(A)**는 DNA 결합에 이용된 Cas-유사 단백질 및 유전자 발현을 활성화시키거나 또는 억제하는 "A/R" 도메인을 포함하는 융합 단백질을 묘사한다. **(B)**는 DNA 결합을 위한 Cas-유사 단백질 및 근위 DNA 또는 단백질의 공유 변형에 의해 후성 상태에 영향을 주는 후성 변형 도메인 ("Epi-mod")을 포함하는 융합 단백질을 도해한다.

도면 3은 2개 RNA-유도된 엔도뉴클레아제를 이용한 유전체 변형을 도해한다. **(A)**는 틱대기효소로 전환된 2개의 RNA-유도된 엔도뉴클레아제에 의해 창출된 이중 가닥 절단을 묘사한다. **(B)**는 엔도뉴클레아제 활성을 갖는 2개의 RNA-유도된 엔도뉴클레아제에 의해 창출된 2개의 이중 가닥 절단을 묘사한다.

도면 4는 Cas9 핵산, Cas9 가이드용 RNA, 그리고 AAVS1-GFP DNA 공여자로 형질감염된 인간 K562 세포의 형광-활성화된 세포 분류 (FACS)를 제시한다. y축은 적색 통로에서 자동 형광 강도를 나타내고, 그리고 x축은 녹색 형광 강도를 나타낸다. **(A)** 안티 리버스 Cap 유사체, 0.3 nmol의 미리 어닐링된 crRNA-tracrRNA 이중나선, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA와 함께 전사된 10 µg의 Cas9 mRNA로 형질감염된 K562 세포; **(B)** 안티 리버스 Cap 유사체, 0.3 nmol의 키메라 RNA, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA와 함께 전사된 10 µg의 Cas9 mRNA로 형질감염된 K562 세포; **(C)** 전사후 캡핑 반응에 의해 캡핑된 10 µg의 Cas9 mRNA, 0.3 nmol의 키메라 RNA, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA로 형질감염된 K562 세포; **(D)** 10 µg의 Cas9 플라스미드 DNA, 5 µg의 U6-키메라 RNA 플라스미드 DNA, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA로 형질감염된 K562 세포; **(E)** 10 µg의 AAVS1-GFP 플라스미드 DNA로 형질감염된 K562 세포; **(F)** 형질감염 시약 단독으로 형질감염된 K562 세포.

도면 5는 인간 세포 내에 AAVS1 좌위 내로 GFP의 표적화된 통합을 문서화하는 접합부 PCR 분석을 제시한다. 레인 M: 1 kb DNA 분자 마커; 레인 A: 안티 리버스 Cap 유사체, 0.3 nmol의 미리 어닐링된 crRNA-tracrRNA 이중나선, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA와 함께 전사된 10 µg의 Cas9 mRNA로 형질감염된 K562 세포; 레인 B: 안티 리버스 Cap 유사체, 0.3 nmol의 키메라 RNA, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA와 함께 전사된 10 µg의 Cas9 mRNA로 형질감염된 K562 세포; 레인 C: 전사후 캡핑 반응에 의해 캡핑된 10 µg의 Cas9 mRNA, 0.3 nmol의 키메라 RNA, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA로 형질감염된 K562 세포; 레인 D: 10 µg의 Cas9 플라스미드 DNA, 5 µg의 U6-키메라 RNA 플라스미드 DNA, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA로 형질감염된 K562 세포; 레인 E: 10 µg의 AAVS1-GFP 플라스미드 DNA로 형질감염된 K562 세포; 레인 F: 형질감염 시약 단독으로 형질감염된 K562 세포.

발명을 실시하기 위한 구체적인 내용

[0013] 발명의 상세한 설명

[0014] RNA-유도된 엔도뉴클레아제가 본원에서 제공되는데, 이것은 최소한 하나의 핵 국지화 신호, 최소한 하나의 뉴클레아제 도메인, 그리고 엔도뉴클레아제를 개열을 위한 특정한 뉴클레오티드 서열로 표적화하는 유도 RNA와 상호 작용하는 최소한 하나의 도메인을 포함한다. RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산뿐만 아니라 RNA-유도된 엔도뉴클레아제를 이용하여 진핵 세포 또는 배아의 염색체 서열을 변경하는 방법 역시 제공된다. RNA-유도된 엔도뉴클레아제는 특정한 유도 RNA와 상호작용하고, 이들은 각각 상기 엔도뉴클레아제를 특정한 표적화된 부위로 향하게 하고, 상기 부위에서 RNA-유도된 엔도뉴클레아제는 염색체 서열이 변형되도록, DNA 복구 과정에 의해 복구될 수 있는 이중 가닥 절단을 도입한다. 특이성이 유도 RNA에 의해 제공되기 때문에, RNA-기초된 엔도뉴클레아제가 보편적이고 상이한 유전체학 서열을 표적으로 하는 상이한 유도 RNA에서 이용될 수 있다. 본원에서 개시된 방법은 특정한 염색체 서열을 표적으로 하고 변경하고 및/또는 세포 또는 배아의 유전체 내에 표적화된 위치에서 외인성 서열을 도입하는데 이용될 수 있다. 게다가, 표적화는 특정하고 부정확한 효과가 제한된다.

[0015] 본 발명은 융합 단백질을 제공하는데, 여기서 융합 단백질은 CRISPR/Cas-유사 단백질 또는 이의 단편 및 작동체 도메인을 포함한다. 적합한 작동체 도메인은 제한 없이, 개열 도메인, 후성 변형 도메인, 전사 활성화 도메인,

그리고 전사 억제인자 도메인을 포함한다. 각 융합 단백질은 특정한 유도 RNA에 의해 특정한 염색체 서열로 유도되고, 여기서 작동체 도메인은 표적화된 유전체 변형 또는 유전자 조절을 매개한다. 한 양상에서, 융합 단백질은 이합체로서 기능하고, 따라서 표적 부위의 길이를 증가시키고 유전체 내에 이의 독특성의 가능성을 증가시킬 수 있다 (따라서, 부정확한 효과를 감소시킬 수 있다). 가령, 내인성 CRISPR 시스템은 대략 13-20 bp의 DNA 결합 단어 길이에 기초하여 유전체학 위치를 변경한다 (Cong et al., Science, 339:819-823). 이러한 단어 크기에서, 표적 부위의 단지 5-7%만 유전체 내에서 독특하다 (Iseli et al, PLoS One 2(6):e579). 대조적으로, 아연 핑거 뉴클레아제에 대한 DNA 결합 단어 크기는 전형적으로, 30-36 bp 범위에서 변하고, 인간 유전체 내에서 대략 85-87% 독특한 표적 부위를 유발한다. CRISPR-기초된 시스템에 의해 활용되는 더욱 작은 크기산정된 DNA 결합 부위는 원하는 위치, 예를 들면, 질환 SNP, 작은 엑손, 출발 코돈, 그리고 종결 코돈뿐만 아니라 복잡한 유전체 내에 다른 위치 인근에서 표적화된 CRISPR-기초된 뉴클레아제의 설계를 제한하고 복잡하게 만든다. 본 발명은 CRISPR DNA 결합 단어 길이를 확장하기 위한 수단 (즉, 부정확한 활성을 제한하기 위해)을 제공할 뿐만 아니라, 변형된 기능을 갖는 CRISPR 융합 단백질을 더욱 제공한다. 따라서, 개시된 CRISPR 융합 단백질은 증가된 표적 특이성과 독특한 기능성(들)을 갖는다. 표적화된 염색체 서열의 발현을 변경하거나 또는 조절하기 위해 융합 단백질을 이용하는 방법 역시 본원에서 제공된다.

[0017] (I) RNA-유도된 엔도뉴클레아제

[0018] 본 발명의 한 가지 양상은 진핵 세포와 배아, 예를 들면, 예로서, 비-인간 단세포 배아의 핵 내로 엔도뉴클레아제의 진입을 허용하는, 최소한 하나의 핵 국지화 신호를 포함하는 RNA-유도된 엔도뉴클레아제를 제공한다. RNA-유도된 엔도뉴클레아제는 또한, 최소한 하나의 뉴클레아제 도메인 및 유도 RNA와 상호작용하는 최소한 하나의 도메인을 포함한다. RNA-유도된 엔도뉴클레아제는 유도 RNA에 의해 특정한 핵산 서열 (또는 표적 부위)에 지향된다. 유도 RNA는 RNA-유도된 엔도뉴클레아제뿐만 아니라 표적 부위와 상호작용하고, 따라서 일단 표적 부위에 지향되면, RNA-유도된 엔도뉴클레아제는 이중 가닥 절단을 표적 부위 핵산 서열 내로 도입할 수 있다. 유도 RNA가 표적화된 개체에 대한 특이성을 제공하기 때문에, RNA-유도된 엔도뉴클레아제의 엔도뉴클레아제가 보편적이고 상이한 표적 핵산 서열을 개열하는 상이한 유도 RNA에서 이용될 수 있다. 단리된 RNA-유도된 엔도뉴클레아제, RNA-유도된 엔도뉴클레아제를 인코딩하는 단리된 핵산 (즉, RNA 또는 DNA), RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산을 포함하는 벡터, 그리고 RNA-유도된 엔도뉴클레아제 + 유도 RNA를 포함하는 단백질-RNA 복합체가 본원에서 제공된다.

[0019] RNA-유도된 엔도뉴클레아제는 균집된 규칙적으로 산재된 짧은 회귀성 반복부 (CRISPR)/CRISPR-연관된 (Cas) 시스템으로부터 유래될 수 있다. CRISPR/Cas 시스템은 타입 I, 타입 II, 또는 타입 III 시스템일 수 있다. 적합한 CRISPR/Cas 단백질의 무제한적 실례는 Cas3, Cas4, Cas5, Cas5e (또는 CasD), Cas6, Cas6e, Cas6f, Cas7, Cas8a1, Cas8a2, Cas8b, Cas8c, Cas9, Cas10, Cas10d, CasF, CasG, CasH, Csy1, Csy2, Csy3, Cse1 (또는 CasA), Cse2 (또는 CasB), Cse3 (또는 CasE), Cse4 (또는 CasC), Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csz1, Csx15, Csf1, Csf2, Csf3, Csf4, 그리고 Cu1966을 포함한다.

[0020] 한 구체예에서, RNA-유도된 엔도뉴클레아제는 유형 II CRISPR/Cas 시스템으로부터 유래된다. 특정한 구체예에서, RNA-유도된 엔도뉴클레아제는 Cas9 단백질로부터 유래된다. Cas9 단백질은 스트렙토코쿠스 피오게네스 (스트렙토코쿠스 피오게네스), 스트렙토코쿠스 쉐모필루스 (스트렙토코쿠스 *thermophilus*), 스트렙토코쿠스 (스트렙토코쿠스) 종, 노카르디오프시스 라스손빌레이 (*Nocardioopsis dassonvillei*), 스트렙토미세스 프리스티네스피랄리스 (*Streptomyces pristinaespiralis*), 스트렙토미세스 비리도크로모게네스 (*Streptomyces viridochromogenes*), 스트렙토미세스 비리도크로모게네스 (*Streptomyces viridochromogenes*), 스트렙토스포랑기움 로세움 (*Streptosporangium roseum*), 스트렙토스포랑기움 로세움 (*Streptosporangium roseum*), 알리사이클로바실루스 아시도칼다리우스 (*Alicyclobacillus acidocaldarius*), 바실루스 슈도마이코이데스 (*Bacillus pseudomycooides*), 바실루스 셀레니티레듀센스 (*Bacillus selenitireducens*), 엑시구오박테리움 시비리쿰 (*Exiguobacterium sibiricum*), 락토바실루스 델브루키 (*Lactobacillus delbrueckii*), 락토바실루스 살리바리우스 (*Lactobacillus salivarius*), 마이크로스킬라 마리나 (*Microscilla marina*), 버크홀데리아레스 박테리움 (*Burkholderiales bacterium*), 폴라로모나스 나프탈레니보란스 (*Polaromonas naphthalenivorans*), 폴라로모나스 (*Polaromonas*) 종, 크로코스파에라 왓소니 (*Crocospaera watsonii*), 시아노테세 (*Cyanothece*) 종, 마이크로시스티스 아에루기노사 (*Microcystis aeruginosa*), 시네크코쿠스 (*Synechococcus*) 종, 아세토할로비움 아라바티쿰 (*Acetohalobium arabaticum*), 암모니펙스 데젠씨 (*Ammonifex degensii*), 칼디셀룰로시럽토 베시 (*Caldicellulosiruptor bescii*), 칸디다투스 데술포루디스 (*Candidatus Desulforudis*), 클로스트리듐 보틀리눔

(*Clostridium botulinum*), 클로스트리듐 디피실레 (*Clostridium difficile*), 피네골디아 마그나 (*Fingoldia magna*), 나트라나이로비우스 썬모필루스 (*Natranaerobius thermophilus*), 펠로토마쿨룸 썬모프로피오니쿰 (*Pelotomaculum thermopropionicum*), 아시치오바실루스 칼두스 (*Acidithiobacillus caldus*), 아시치오바실루스 페록시단스 (*Acidithiobacillus ferrooxidans*), 알로크로마티움 비노숨 (*Allochroamatium vinosum*), 마리노박터 (*Marinobacter*) 종, 니트로소코쿠스 할로필루스 (*Nitrosococcus halophilus*), 니트로소코쿠스 왓소니 (*Nitrosococcus watsoni*), 슈도알테로모나스 할로플랭크티스 (*Pseudoalteromonas haloplanktis*), 크테도노박터 라세미페르 (*Ktedonobacter racemifer*), 메타노할로비움 에베스티가툼 (*Methanohalobium evestigatum*), 아나베나 바리아빌리스 (*Anabaena variabilis*), 노둘라리아 스푸미게나 (*Nodularia spumigena*), 노스톡 (*Nostoc*) 종, 아르트로스피라 맥시마 (*Arthrospira maxima*), 아르트로스피라 플라텐시스 (*Arthrospira platensis*), 아르트로스피라 (*Arthrospira*) 종, 링비아 (*Lyngbya*) 종, 미크로콜레우스 크토노플라스테스 (*Microcoleus chthonoplastes*), 오실라토리아 (*Oscillatoria*) 종, 페트로토가 모빌리스 (*Petrotoga mobilis*), 썬모시포 아프리카누스 (*Thermosiphon africanus*), 또는 아카리오클로리스 마리나 (*Acaryochloris marina*)로부터 유래될 수 있다.

[0021] 일반적으로, CRISPR/Cas 단백질은 최소한 하나의 RNA 인식 및/또는 RNA 결합 도메인을 포함한다. RNA 인식 및/또는 RNA 결합 도메인은 유도 RNA와 상호작용한다. CRISPR/Cas 단백질은 또한, 뉴클레아제 도메인 (즉, DNA분해효소 또는 RNA분해효소 도메인), DNA 결합 도메인, 헬리카아제 도메인, RNA분해효소 도메인, 단백질-단백질 상호작용 도메인, 이합체화 도메인뿐만 아니라 다른 도메인을 포함할 수 있다.

[0022] CRISPR/Cas-유사 단백질은 야생형 CRISPR/Cas 단백질, 변형된 CRISPR/Cas 단백질, 또는 야생형 또는 변형된 CRISPR/Cas 단백질의 단편일 수 있다. CRISPR/Cas-유사 단백질은 핵산 결합 친화성 및/또는 특이성을 증가시키고, 효소적 활성을 변경하고, 및/또는 단백질의 다른 성질을 변화시키기 위해 변형될 수 있다. 가령, CRISPR/Cas-유사 단백질의 뉴클레아제 (즉, DNA분해효소, RNA분해효소) 도메인은 변형되거나, 결실되거나, 또는 불활성화될 수 있다. 대안으로, CRISPR/Cas-유사 단백질은 융합 단백질의 기능에 필수적이지 않은 도메인을 제거하기 위해 절두될 수 있다. CRISPR/Cas-유사 단백질은 또한, 융합 단백질의 작동체 도메인의 활성을 최적화하기 위해 절두되거나 또는 변형될 수 있다.

[0023] 일부 구체예에서, CRISPR/Cas-유사 단백질은 야생형 Cas9 단백질 또는 이의 단편으로부터 유래될 수 있다. 다른 구체예에서, CRISPR/Cas-유사 단백질은 변형된 Cas9 단백질로부터 유래될 수 있다. 가령, Cas9 단백질의 아미노산 서열은 단백질의 하나 또는 그 이상의 성질 (가령, 뉴클레아제 활성, 친화성, 안정성 등)을 변경하기 위해 변형될 수 있다. 대안으로, RNA-유도된 개열에 관련되지 않는 Cas9 단백질의 도메인이 단백질로부터 제거될 수 있고, 따라서 변형된 Cas9 단백질은 야생형 Cas9 단백질보다 작다.

[0024] 일반적으로, Cas9 단백질은 최소한 2개의 뉴클레아제 (즉, DNA분해효소) 도메인을 포함한다. 가령, Cas9 단백질은 RuvC-유사 뉴클레아제 도메인 및 HNH-유사 뉴클레아제 도메인을 포함할 수 있다. RuvC와 HNH 도메인은 단일 가닥을 절단하여 DNA에서 이중 가닥 절단을 만들기 위해 함께 작동한다 (Jinek et al., Science, 337: 816-821). 일부 구체예에서, Cas9-유래된 단백질은 단지 한 가지 기능적 뉴클레아제 도메인 (RuvC-유사 또는 HNH-유사 뉴클레아제 도메인 중에서 어느 한쪽)만을 내포하도록 변형될 수 있다. 가령, Cas9-유래된 단백질은 뉴클레아제 도메인 중에서 한 가지가 결실되거나 또는 돌연변이되고, 따라서 이것이 더 이상 기능하지 않도록 (즉, 뉴클레아제 활성이 부재하도록) 변형될 수 있다. 뉴클레아제 도메인 중에서 한 가지가 비활성인 일부 구체예에서, Cas9-유래된 단백질은 이중 가닥 핵산 내로 틸을 도입할 수 있지만 (이런 단백질은 "틈내기효소"로 명명된다), 이중 가닥 DNA를 개열하지 않는다. 가령, RuvC-유사 도메인에서 아스파르트산염에서 알라닌 (D10A) 전환은 Cas9-유래된 단백질을 틸내기효소로 전환한다. 유사하게, HNH 도메인에서 히스티딘에서 알라닌 (H840A 또는 H839A) 전환은 Cas9-유래된 단백질을 틸내기효소로 전환한다. 각 뉴클레아제 도메인은 널리 공지된 방법, 예를 들면, 부위-지향된 돌연변이유발, PCR-매개된 돌연변이유발, 그리고 전체 유전자 합성뿐만 아니라 당분야에서 공지된 다른 방법을 이용하여 변형될 수 있다.

[0025] 본원에서 개시된 RNA-유도된 엔도뉴클레아제는 최소한 하나의 핵 국지화 신호를 포함한다. 일반적으로, NLS은 염기성 아미노산의 스트레치를 포함한다. 핵 국지화 신호는 당분야에서 공지된다 (가령, Lange et al., J. Biol. Chem., 2007, 282:5101-5105를 참조한다). 가령, 한 구체예에서, NLS는 단립형 서열, 예를 들면, PKKKRKV (서열 번호:1) 또는 PKKKRRV (서열 번호:2)일 수 있다. 다른 구체예에서, NLS는 이분 서열일 수 있다. 또 다른 구체예에서, NLS는 KRPAATKKAGQAKKKK (서열 번호:3)일 수 있다. NLS는 RNA-유도된 엔도뉴클레아제의 N 말단, C 말단, 또는 내부 위치에서 위치될 수 있다.

[0026] 일부 구체예에서, RNA-유도된 엔도뉴클레아제는 최소한 하나의 세포-투과성 도메인을 더욱 포함할 수 있다. 한 구체예에서, 세포-투과성 도메인은 제1형 사람 면역결핍바이러스 TAT 단백질로부터 유래된 세포-투과성 펩티드 서열일 수 있다. 실례로서, TAT 세포-투과성 서열은 GRKKRRQRRRPQPQKKRKKV (서열 번호:4)일 수 있다. 다른 구체예에서, 세포-투과성 도메인은 인간 B형 간염 바이러스로부터 유래된 세포-투과성 펩티드 서열인 TLM (PLSSIFSRIGDPPKRRKKV; 서열 번호:5)일 수 있다. 또 다른 구체예에서, 세포-투과성 도메인은 MPG (GALFLGWLGAAGSTMGAPKKRKKV; 서열 번호:6 또는 GALFLGFLGAAGSTMGAWSPKRRKKV; 서열 번호:7)일 수 있다. 추가의 구체예에서, 세포-투과성 도메인은 Pep-1 (KETWETWTEWSQPQKKRKKV; 서열 번호:8), VP22, 단순 헤르페스 바이러스로부터 세포 투과성 펩티드, 또는 폴리아르기닌 펩티드 서열일 수 있다. 세포-투과성 도메인은 상기 단백질의 N 말단, C 말단, 또는 내부 위치에서 위치될 수 있다.

[0027] 또 다른 구체예에서, RNA-유도된 엔도뉴클레아제는 또한, 최소한 하나의 마커 도메인을 포함할 수 있다. 마커 도메인의 무제한적 실례는 형광 단백질, 정제 태그, 그리고 에피토프 태그를 포함한다. 일부 구체예에서, 마커 도메인은 형광 단백질일 수 있다. 적합한 형광 단백질의 무제한적인 실례는 녹색 형광 단백질 (가령, GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), 황색 형광 단백질 (가령, YFP, EYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), 청색 형광 단백질 (가령, EBFP, EBFP2, Azurite, mKalamal, GFPuv, Sapphire, T-sapphire.), 시안색 형광 단백질 (가령, ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), 적색 형광 단백질 (mKate, mKate2, mPlum, DsRed 단위체, mCherry, mRFP1, DsRed-Express, DsRed2, DsRed-단위체, HcRed-탠덤, HcRed1, AsRed2, eqFP611, mRaspberry, mStrawberry, Jred), 그리고 오렌지색 형광 단백질 (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato) 또는 임의의 다른 적합한 형광 단백질을 포함한다. 다른 구체예에서, 마커 도메인은 정제 태그 및/또는 에피토프 태그일 수 있다. 예시적인 태그에는 글루타티온-S-전달효소 (GST), 키틴 결합 단백질 (CBP), 말토오스 결합 단백질, 티오레독신 (TRX), 폴리(NANP), 탠덤 친화성 정제 (TAP) 태그, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6xHis, 비오틴 카르복실 운반 단백질 (BCCP), 그리고 칼모듈린이 포함되지만 이들에 한정되지 않는다.

[0028] 일정한 구체예에서, RNA-유도된 엔도뉴클레아제는 유도 RNA를 포함하는 단백질-RNA 복합체의 부분일 수 있다. 유도 RNA는 RNA-유도된 엔도뉴클레아제와 상호작용하여 엔도뉴클레아제를 특정한 표적 부위로 향하게 하고, 여기서 유도 RNA의 5' 단부는 특정한 프로토스페이서 서열과 염기쌍을 이룬다.

[0030] **(II) 융합 단백질**

[0031] 본 발명의 다른 양상은 CRISPR/Cas-유사 단백질 또는 이의 단편 및 작동체 도메인을 포함하는 융합 단백질을 제공한다. CRISPR/Cas-유사 단백질은 유도 RNA에 의해 표적 부위에 지향되고, 상기 부위에서 작동체 도메인은 표적화된 핵산 서열을 변경하거나 또는 산출할 수 있다. 작동체 도메인은 개열 도메인, 후성 변형 도메인, 전사 활성화 도메인, 또는 전사 억제인자 도메인일 수 있다. 융합 단백질은 핵 국지화 신호, 세포-투과성 도메인, 또는 마커 도메인에서 선택된 최소한 하나의 추가 도메인을 더욱 포함할 수 있다.

[0033] **(a) CRISPR/Cas-유사 단백질**

[0034] 융합 단백질은 CRISPR/Cas-유사 단백질 또는 이의 단편을 포함한다. CRISPR/Cas-유사 단백질은 상기 섹션 (I)에 서 상술된다. CRISPR/Cas-유사 단백질은 융합 단백질의 N 말단, C 말단, 또는 내부 위치에서 위치될 수 있다.

[0035] 일부 구체예에서, 융합 단백질의 CRISPR/Cas-유사 단백질은 Cas9 단백질로부터 유래될 수 있다. Cas9-유래된 단백질은 야생형, 변형된, 또는 이의 단편일 수 있다. 일부 구체예에서, Cas9-유래된 단백질은 단지 한 가지 기능적 뉴클레아제 도메인 (RuvC-유사 또는 HNH-유사 뉴클레아제 도메인 중에서 어느 한쪽)을 내포하도록 변형될 수 있다. 가령, Cas9-유래된 단백질은 뉴클레아제 도메인 중에서 한 가지가 결실되거나 또는 돌연변이되고, 따라서 이것이 더 이상 기능하지 않도록 (즉, 뉴클레아제 활성이 부재하도록) 변형될 수 있다. 뉴클레아제 도메인 중에서 한 가지가 비활성인 일부 구체예에서, Cas9-유래된 단백질은 이중 가닥 핵산 내로 틈을 도입할 수 있지만 (이런 단백질은 "틈내기효소"로 명명된다), 이중 가닥 DNA를 개열하지 않는다. 가령, RuvC-유사 도메인에서 아스파르트산염에서 알라닌 (D10A) 전환은 Cas9-유래된 단백질을 틈내기효소로 전환한다. 유사하게, HNH 도메인에서 히스티딘에서 알라닌 (H840A 또는 H839A) 전환은 Cas9-유래된 단백질을 틈내기효소로 전환한다. 다른 구체예에서, RuvC-유사 뉴클레아제 도메인과 HNH-유사 뉴클레아제 도메인 둘 모두 Cas9-유래된 단백질이 이중 가닥 핵산을 틈내기하거나 또는 개열할 수 없도록 변형되거나 또는 제거될 수 있다. 또 다른 구체예에서, Cas9-유래된 단백질의 모든 뉴클레아제 도메인은 Cas9-유래된 단백질이 모든 뉴클레아제 활성을 결여하도록 변형되거나 또는

제거될 수 있다.

[0036] 상기-설명된 구체예 중에서 한 가지에서, 임의의 또는 모든 뉴클레아제 도메인은 널리 공지된 방법, 예를 들면, 부위-지향된 돌연변이유발, PCR-매개된 돌연변이유발, 그리고 전체 유전자 합성뿐만 아니라 당분야에서 공지된 다른 방법을 이용하여 하나 또는 그 이상의 결실 돌연변이, 삽입 돌연변이, 및/또는 치환 돌연변이에 의해 비활성화될 수 있다. 예시적인 구체예에서, 융합 단백질의 CRISPR/Cas-유사 단백질은 모든 뉴클레아제 도메인이 비활성화되거나 또는 결실된 Cas9 단백질로부터 유래된다.

[0037] **(b) 작동체 도메인**

[0038] 융합 단백질은 또한, 작동체 도메인을 포함한다. 작동체 도메인은 개열 도메인, 후성 변형 도메인, 전사 활성화 도메인, 또는 전사 억제인자 도메인일 수 있다. 작동체 도메인은 융합 단백질의 N 말단, C 말단, 또는 내부 위치에서 위치될 수 있다.

[0040] (i) 개열 도메인

[0041] 일부 구체예에서, 작동체 도메인은 개열 도메인이다. 본원에서 이용된 바와 같이, "개열 도메인"은 DNA를 개열하는 도메인을 지칭한다. 개열 도메인은 임의의 엔도뉴클레아제 또는 엑소뉴클레아제로부터 획득될 수 있다. 개열 도메인이 유래될 수 있는 엔도뉴클레아제의 무제한적 실례에는 제한 엔도뉴클레아제 및 귀소 엔도뉴클레아제가 포함되지만 이들에 한정되지 않는다. 가령, New England Biolabs Catalog 또는 Belfort et al. (1997) *Nucleic Acids Res.* 25:3379-3388을 참조한다. DNA를 개열하는 추가 효소가 알려져 있다 (가령, S1 뉴클레아제; 녹두 뉴클레아제; 췌장 DNA분해효소 I; 미구균 뉴클레아제; 효모 HO 엔도뉴클레아제). Linn et al. (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993을 또한 참조한다. 이들 효소 중에서 하나 또는 그 이상 (또는 이의 기능적 단편)이 개열 도메인의 공급원으로서 이용될 수 있다.

[0042] 일부 구체예에서, 개열 도메인은 유형 II-S 엔도뉴클레아제로부터 유래될 수 있다. 유형 II-S 엔도뉴클레아제는 전형적으로, 인식 부위로부터 여러 염기쌍 떨어진 부위에서 DNA를 개열하고, 그리고 따라서, 분리가능한 인식과 개열 도메인을 갖는다. 이들 효소는 일반적으로, 엇갈린 위치에서 DNA의 각 가닥을 개열하기 위해, 일시적으로 결합하여 이합체를 형성하는 단위체이다. 적합한 유형 II-S 엔도뉴클레아제의 무제한적 실례는 BfiI, BpmI, BsaI, BsgI, BsmBI, BsmI, BspMI, FokI, MboII, 그리고 SapI을 포함한다. 예시적인 구체예에서, 융합 단백질의 개열 도메인은 FokI 개열 도메인 또는 이의 유도체이다.

[0043] 일정한 구체예에서, 유형 II-S 개열은 2개의 상이한 개열 도메인 (이들은 각각, CRISPR/Cas-유사 단백질 또는 이의 단편에 부착된다)의 이합체화를 조장하도록 변형될 수 있다. 가령, FokI의 개열 도메인은 일정한 아미노산 잔기를 돌연변이시킴으로써 변형될 수 있다. 무제한적 실례에 의하여, FokI 개열 도메인의 위치 446, 447, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, 그리고 538에서 아미노산 잔기가 변형에 대한 표적이다. 가령, 절대 이형이합체를 형성하는 FokI의 변형된 개열 도메인은 첫 번째 변형된 개열 도메인이 아미노산 위치 490과 538에서 돌연변이를 포함하고, 그리고 두 번째 변형된 개열 도메인이 아미노산 위치 486과 499에서 돌연변이를 포함하는 쌍을 포함한다 (Miller et al., 2007, *Nat. Biotechnol.*, 25:778-785; Szczpek et al., 2007, *Nat. Biotechnol.*, 25:786-793). 가령, 한 도메인에서 위치 490에서 Glu (E)가 Lys (K)로 변화될 수 있고 위치 538에서 Ile (I)가 K로 변화될 수 있고 (E490K, I538K), 그리고 다른 개열 도메인에서 위치 486에서 Gln (Q)이 E로 변화될 수 있고 위치 499에서 I가 Leu (L)로 변화될 수 있다 (Q486E, I499L). 다른 구체예에서, 변형된 FokI 개열 도메인은 3가지 아미노산 변화를 포함할 수 있다 (Doyon et al. 2011, *Nat. Methods*, 8:74-81). 가령, 한 변형된 FokI 도메인 (ELD로 명명된다)은 Q486E, I499L, N496D 돌연변이를 포함할 수 있고, 그리고 다른 변형된 FokI 도메인 (KKR로 명명된다)은 E490K, I538K, H537R 돌연변이를 포함할 수 있다.

[0044] 예시적인 구체예에서, 융합 단백질의 작동체 도메인은 FokI 개열 도메인 또는 변형된 FokI 개열 도메인이다.

[0045] 작동체 도메인이 개열 도메인이고 CRISPR/Cas-유사 단백질이 Cas9 단백질로부터 유래되는 구체예에서, Cas9-유래된 단백질은 엔도뉴클레아제 활성이 제거되도록 본원에서 논의된 바와 같이 변형될 수 있다. 가령, Cas9-유래된 단백질은 이들이 뉴클레아제 활성을 더 이상 소유하지 않도록, RuvC와 HNH 도메인을 돌연변이시킴으로써 변형될 수 있다.

[0047] (ii) 후성 변형 도메인

[0048] 다른 구체예에서, 융합 단백질의 작동체 도메인은 후성 변형 도메인일 수 있다. 일반적으로, 후성 변형 도메인은 DNA 서열을 변화시키지 않으면서 히스톤 구조 및/또는 염색체 구조를 변경한다. 히스톤 및/또는 염색질 구조

에서 변화는 유전자 발현에서 변화를 야기할 수 있다. 후성 변형의 실례는 제한 없이, 히스톤 단백질에서 리신 잔기의 아세틸화 또는 메틸화, 그리고 DNA에서 시토신 잔기의 메틸화를 포함한다. 적합한 후성 변형 도메인의 무제한적 실례는 히스톤 아세틸전달효소 도메인, 히스톤 탈아세틸화효소 도메인, 히스톤 메틸전달효소 도메인, 히스톤 탈메틸효소 도메인, DNA 메틸전달효소 도메인, 그리고 DNA 탈메틸효소 도메인을 포함한다.

[0050] 작동체 도메인이 히스톤 아세틸전달효소 (HAT) 도메인인 구체예에서, HAT 도메인은 EP300 (즉, E1A 결합 단백질 p300), CREBBP (즉, CREB-결합 단백질), CDY1, CDY2, CDYL1, CLOCK, ELP3, ESA1, GCN5 (KAT2A), HAT1, KAT2B, KAT5, MYST1, MYST2, MYST3, MYST4, NCOA1, NCOA2, NCOA3, NCOAT, P/CAF, Tip60, TAFII250, 또는 TF3C4로부터 유래될 수 있다. 이와 같은 한 가지 구체예에서, HAT 도메인은 p300이다.

[0051] 작동체 도메인이 후성 변형 도메인이고 CRISPR/Cas-유사 단백질이 Cas9 단백질로부터 유래되는 구체예에서, Cas9-유래된 단백질은 엔도뉴클레아제 활성이 제거되도록, 본원에서 논의된 바와 같이 변형될 수 있다. 가령, Cas9-유래된 단백질은 그들이 뉴클레아제 활성을 더 이상 소유하지 않도록, RuvC와 HNH 도메인을 돌연변이시킴으로써 변형될 수 있다.

[0053] (iii) 전사 활성화 도메인

[0054] 다른 구체예에서, 융합 단백질의 작동체 도메인은 전사 활성화 도메인일 수 있다. 일반적으로, 전사 활성화 도메인은 유전자의 전사를 증가 및/또는 활성화시키기 위해 전사 제어 원소 및/또는 전사 조절 단백질 (즉, 전사 인자, RNA 중합효소 등)과 상호작용한다. 일부 구체예에서, 전사 활성화 도메인은 제한 없이, 단순 헤르페스 바이러스 VP16 활성화 도메인, VP64 (이것은 VP16의 사합체성 유도체이다), NFκB p65 활성화 도메인, p53 활성화 도메인 1과 2, CREB (cAMP 반응 원소 결합 단백질) 활성화 도메인, E2A 활성화 도메인, 그리고 NFAT (활성화된 T-세포의 핵 인자) 활성화 도메인일 수 있다. 다른 구체예에서, 전사 활성화 도메인은 Gal4, Gcn4, MLL, Rtg3, Gln3, Oaf1, Pip2, Pdr1, Pdr3, Pho4, 그리고 Leu3일 수 있다. 전사 활성화 도메인은 야생형이거나, 또는 본래 전사 활성화 도메인의 변형된 이형일 수 있다. 일부 구체예에서, 융합 단백질의 작동체 도메인은 VP16 또는 VP64 전사 활성화 도메인이다.

[0055] 작동체 도메인이 전사 활성화 도메인이고 CRISPR/Cas-유사 단백질이 Cas9 단백질로부터 유래되는 구체예에서, Cas9-유래된 단백질은 엔도뉴클레아제 활성이 제거되도록, 본원에서 논의된 바와 같이 변형될 수 있다. 가령, Cas9-유래된 단백질은 그들이 뉴클레아제 활성을 더 이상 소유하지 않도록, RuvC와 HNH 도메인을 돌연변이시킴으로써 변형될 수 있다.

[0057] (iv) 전사 억제인자 도메인

[0058] 또 다른 구체예에서, 융합 단백질의 작동체 도메인은 전사 억제인자 도메인일 수 있다. 일반적으로, 전사 억제인자 도메인은 유전자의 전사를 감소시키고 및/또는 종결하기 위해 전사 제어 원소 및/또는 전사 조절 단백질 (즉, 전사 인자, RNA 중합효소 등)과 상호작용한다. 적합한 전사 억제인자 도메인의 무제한적 실례는 유도성 cAMP 초기 억제인자 (ICER) 도메인, Kruppel-연관된 상자 A (KRAB-A) 억제인자 도메인, YY1 글리신 풍부한 억제인자 도메인, Sp1-유사 억제인자, E(sp1) 억제인자, IκB 억제인자, 그리고 MeCP2를 포함한다.

[0059] 작동체 도메인이 전사 억제인자 도메인이고 CRISPR/Cas-유사 단백질이 Cas9 단백질로부터 유래되는 구체예에서, Cas9-유래된 단백질은 엔도뉴클레아제 활성이 제거되도록, 본원에서 논의된 바와 같이 변형될 수 있다. 가령, cas9는 그들이 뉴클레아제 활성을 더 이상 소유하지 않도록, RuvC와 HNH 도메인을 돌연변이시킴으로써 변형될 수 있다.

[0061] **(c) 추가 도메인**

[0062] 일부 구체예에서, 융합 단백질은 최소한 하나의 추가 도메인을 더욱 포함한다. 적합한 추가 도메인의 무제한적 실례는 핵 국지화 신호, 세포-투과성 또는 전위 도메인, 그리고 마커 도메인을 포함한다. 적합한 핵 국지화 신호, 세포-투과성 도메인, 그리고 마커 도메인의 무제한적 실례는 상기 섹션 (I)에서 제공된다.

[0064] **(d) 융합 단백질 이합체**

[0065] 융합 단백질의 작동체 도메인이 개별 도메인인 구체예에서, 최소한 하나의 융합 단백질을 포함하는 이합체가 형성될 수 있다. 이합체는 동종이합체 또는 이형이합체일 수 있다. 일부 구체예에서, 이형이합체는 2개의 상이한 융합 단백질을 포함한다. 다른 구체예에서, 이형이합체는 1개의 융합 단백질 및 추가 단백질을 포함한다.

[0066] 일부 구체예에서, 이합체는 2개의 융합 단백질 단위체가 일차성 아미노산 서열에 대하여 동일한

동종이합체이다. 이합체가 동종이합체인 한 구체예에서, Cas9-유래된 단백질은 그들의 엔도뉴클레아제 활성이 제거되도록, 다시 말하면, 그들이 기능적 뉴클레아제 도메인을 갖지 않도록 변형된다. Cas9-유래된 단백질이 그들의 엔도뉴클레아제 활성이 제거되도록 변형되는 일정한 구체예에서, 각 융합 단백질 단위체는 동일한 Cas9 유사 단백질 및 동일한 개열 도메인을 포함한다. 개열 도메인은 임의의 개열 도메인, 예를 들면, 본원에서 제공된 예시적인 개열 도메인 중에서 한 가지일 수 있다. 특정한 구체예에서, 개열 도메인은 FokI 개열 도메인 또는 변형된 FokI 개열 도메인이다. 이런 구체예에서, 특정한 유도 RNA는 융합 단백질 단위체를 상이하지만 가깝게 인접한 부위로 향하게 할 것이고, 따라서 이합체 형성 시에, 이들 두 단위체의 뉴클레아제 도메인은 표적 DNA 내에 이중 가닥 절단을 창출할 것이다.

[0067] 다른 구체예에서, 이합체는 2개의 상이한 융합 단백질의 이형이합체이다. 가령, 각 융합 단백질의 CRISPR/Cas-유사 단백질은 상이한 CRISPR/Cas 단백질로부터 또는 상이한 세균 종의 이중상동성 CRISPR/Cas 단백질로부터 유래될 수 있다. 가령, 각 융합 단백질은 Cas9-유사 단백질을 포함할 수 있고, 상기 Cas9-유사 단백질은 상이한 세균 종으로부터 유래된다. 이들 구체예에서, 각 융합 단백질은 상이한 표적 부위 (즉, 프로토스페이서 및/또는 PAM 서열에 의해 특정됨)를 인식할 것이다. 가령, 유도 RNA는 이형이합체를 상이하지만 가깝게 인접한 부위에 위치시킬 수 있고, 따라서 그들의 뉴클레아제 도메인은 표적 DNA 내에 효과적인 이중 가닥 절단을 유발한다. 이형이합체는 또한, 틈내는 위치가 상이하도록, 틈내기 활성을 갖는 변형된 Cas9 단백질을 가질 수 있다.

[0068] 대안으로, 이형이합체의 두 융합 단백질은 상이한 작동체 도메인을 가질 수 있다. 작동체 도메인이 개열 도메인인 구체예에서, 각 융합 단백질은 상이한 변형된 개열 도메인을 내포할 수 있다. 가령, 각 융합 단백질은 상기 섹션 (II)(b)(i)에서 상술된 바와 같이, 상이한 변형된 FokI 개열 도메인을 내포할 수 있다. 이들 구체예에서, Cas-9 단백질은 그들의 엔도뉴클레아제 활성이 제거되도록 변형될 수 있다.

[0069] 당업자에 의해 인지되는 바와 같이, 이형이합체를 형성하는 이들 두 융합 단백질은 CRISPR/Cas-유사 단백질 도메인 및 작동체 도메인 둘 모두에서 상이할 수 있다.

[0070] 상기-설명된 구체예 중에서 한 가지에서, 동종이합체 또는 이형이합체는 상기 상술된 바와 같이, 핵 국지화 신호 (NLSs), 세포-투과성, 전위 도메인 및 마커 도메인에서 선택되는 최소한 하나의 추가 도메인을 포함할 수 있다.

[0071] 상기-설명된 구체예 중에서 한 가지에서, Cas9-유래된 단백질 중에서 한쪽 또는 양쪽이 엔도뉴클레아제 활성이 제거되거나 또는 변형되도록 변형될 수 있다.

[0072] 다른 대안적 구체예에서, 이형이합체는 1개의 융합 단백질 및 추가 단백질을 포함한다. 가령, 추가 단백질은 뉴클레아제일 수 있다. 한 구체예에서, 뉴클레아제는 아연 핑거 뉴클레아제이다. 아연 핑거 뉴클레아제는 아연 핑거 DNA 결합 도메인 및 개열 도메인을 포함한다. 아연 핑거는 3개의 뉴클레오티드를 인식하고 이들에 결합한다. 아연 핑거 DNA 결합 도메인은 약 3개 아연 핑거 내지 약 7개 아연 핑거를 포함할 수 있다. 아연 핑거 DNA 결합 도메인은 자연적으로 발생하는 단백질로부터 유래되거나 또는 가공될 수 있다. 가령, Beerli et al. (2002) Nat. Biotechnol. 20:135-141; Pabo et al. (2001) Ann. Rev. Biochem. 70:313-340; Isalan et al. (2001) Nat. Biotechnol. 19:656-660; Segal et al. (2001) Curr. Opin. Biotechnol. 12:632-637; Choo et al. (2000) Curr. Opin. Struct. Biol. 10:411-416; Zhang et al. (2000) J. Biol. Chem. 275(43):33850-33860; Doyon et al. (2008) Nat. Biotechnol. 26:702-708; 그리고 Santiago et al. (2008) Proc. Natl. Acad. Sci. USA 105:5809-5814를 참조한다. 아연 핑거 뉴클레아제의 개열 도메인은 상기 섹션 (II)(b)(i)에서 상술된 임의의 개열 도메인일 수 있다. 예시적인 구체예에서, 아연 핑거 뉴클레아제의 개열 도메인은 FokI 개열 도메인 또는 변형된 FokI 개열 도메인이다. 이런 아연 핑거 뉴클레아제는 FokI 개열 도메인 또는 변형된 FokI 개열 도메인을 포함하는 융합 단백질과 이합체화할 것이다.

[0073] 일부 구체예에서, 아연 핑거 뉴클레아제는 핵 국지화 신호, 세포-투과성 또는 전위 도메인에서 선택되는 최소한 하나의 추가 도메인을 포함할 수 있고, 이들은 상기 상술된다.

[0074] 일정한 구체예에서, 상기 상술된 융합 단백질 중에서 한 가지 또는 최소한 하나의 융합 단백질을 포함하는 이합체는 최소한 하나의 유도 RNA를 포함하는 단백질-RNA 복합체의 부분일 수 있다. 유도 RNA는 융합 단백질의 CRISPR-Cas0 유사 단백질과 상호작용하여 상기 융합 단백질을 특정한 표적 부위로 향하게 하고, 여기서 유도 RNA의 5' 단부는 특정한 프로토스페이서 서열과 염기쌍을 이룬다.

[0076] **(III) RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 핵산**

[0077] 본 발명의 다른 양상은 각각, 상기 섹션 (I)과 (II)에서 설명된 RNA-유도된 엔도뉴클레아제 또는 융합 단백질

중에서 한 가지를 인코딩하는 핵산을 제공한다. 핵산은 RNA 또는 DNA일 수 있다. 한 구체예에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 핵산은 mRNA이다. mRNA는 5' 캡핑되고 및/또는 3' 폴리아데닐화될 수 있다. 다른 구체예에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 핵산은 DNA이다. DNA는 벡터 내에 존재할 수 있다 (아래 참조).

[0078] RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 핵산은 관심되는 진핵 세포 또는 동물에서 단백질로의 효율적인 번역을 위해 코돈 최적화될 수 있다. 가령, 인간, 생쥐, 쥐, 햄스터, 소, 돼지, 고양이, 개, 어류, 양서류, 식물, 효모, 곤충, 기타 등등에서 발현을 위해 코돈 최적화될 수 있다 (참조: www.kazusa.or.jp/codon/에서 Codon Usage Database). 코돈 최적화를 위한 프로그램은 프리웨어로서 가용하다 (가령, genomes.urv.es/OPTIMIZER에서 OPTIMIZER; www.genscript.com/codon_opt.html에서 GenScript로부터 OptimumGene™). 상업적인 코돈 최적화 프로그램 역시 가용하다.

[0079] 일부 구체예에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 DNA는 최소한 하나의 프로모터 제어 서열에 작동가능하게 연결될 수 있다. 일부 반복에서, DNA 코딩 서열은 관심되는 진핵 세포 또는 동물에서 발현을 위해 프로모터 제어 서열에 작동가능하게 연결될 수 있다. 프로모터 제어 서열은 구조성, 조절된, 또는 조직 특이적일 수 있다. 적합한 구조성 프로모터 제어 서열에는 시토메갈로바이러스 극초기 프로모터 (CMV), 유인원 바이러스(SV40) 프로모터, 아데노바이러스 주요 후기 프로모터, 라우스 육종 바이러스 (RSV) 프로모터, 생쥐 유방 종양 바이러스 (MMTV) 프로모터, 포스포글리세린산 키나아제 (PGK) 프로모터, 연장 인자 (ED1)-알파 프로모터, 유비퀴틴 프로모터, 액틴 프로모터, 튜불린 프로모터, 면역글로불린 프로모터, 이들의 단편, 또는 전술한 것들의 임의의 조합이 포함되지만 이들에 한정되지 않는다. 적합한 조절된 프로모터 제어 서열의 실례는 제한 없이, 열 쇼크, 금속, 스테로이드, 항생제, 또는 알코올에 의해 조절된 것들을 포함한다. 조직 특이적 프로모터의 무제한적 실례는 B29 프로모터, CD14 프로모터, CD43 프로모터, CD45 프로모터, CD68 프로모터, 테스민 프로모터, 엘라스타아제-1 프로모터, 엔도글린 프로모터, 섬유결합소 프로모터, Flt-1 프로모터, GFAP 프로모터, GPIIb 프로모터, ICAM-2 프로모터, INF-β 프로모터, Mb 프로모터, NphsI 프로모터, OG-2 프로모터, SP-B 프로모터, SYN1 프로모터, 그리고 WASP 프로모터를 포함한다. 프로모터 서열은 야생형이거나 또는 더욱 효율적인 또는 유효한 발현을 위해 변형될 수 있다. 한 예시적인 구체예에서, 인코딩 DNA는 포유류 세포에서 구조성 발현을 위해 CMV 프로모터에 작동가능하게 연결될 수 있다.

[0080] 일정한 구체예에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 서열은 시험관내 mRNA 합성을 위한 파지 RNA 중합효소에 의해 인식되는 프로모터 서열에 작동가능하게 연결될 수 있다. 이런 구체예에서, 시험관내-전사된 RNA는 아래 섹션 (IV)와 (V)에서 상술된 방법에서 이용을 위해 정제될 수 있다. 가령, 프로모터 서열은 T7, T3, 또는 SP6 프로모터 서열 또는 T7, T3, 또는 SP6 프로모터 서열의 변이일 수 있다. 예시적인 구체예에서, 융합 단백질을 인코딩하는 DNA는 T7 RNA 중합효소를 이용한 시험관내 mRNA 합성을 위해 T7 프로모터에 작동가능하게 연결된다.

[0081] 대안적 구체예에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 서열은 세균 또는 진핵 세포에서 RNA-유도된 엔도뉴클레아제 또는 융합 단백질의 시험관내 발현을 위해 프로모터 서열에 작동가능하게 연결될 수 있다. 이런 구체예에서, 발현된 단백질은 아래 섹션 (IV)와 (V)에서 상술된 방법에서 이용을 위해 정제될 수 있다. 적합한 세균 프로모터는 제한 없이, T7 프로모터, *lac* 오페론 프로모터, *trp* 프로모터, 이들의 변이, 그리고 이들의 조합을 포함한다. 예시적인 세균 프로모터는 *trp*와 *lac* 프로모터의 하이브리드인 *tac*이다. 적합한 진핵 프로모터의 무제한적 실례는 상기에서 열거된다.

[0082] 추가의 양상에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 DNA는 또한, 폴리아데닐화 신호 (가령, SV40 polyA 신호, 소 성장 호르몬 (BGH) polyA 신호 등) 및/또는 최소한 하나의 전사 종결 서열에 연결될 수 있다. 부가적으로, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 서열은 또한, 최소한 하나의 핵 국지화 신호, 최소한 하나의 세포-투과성 도메인, 및/또는 최소한 하나의 마커 도메인을 인코딩하는 서열에 연결될 수 있고, 이들은 상기 섹션 (I)에서 상술된다.

[0083] 다양한 구체예에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 DNA는 벡터 내에 존재할 수 있다. 적합한 벡터는 플라스미드 벡터, 파지미드, 코스미드, 인공/꼬마염색체, 트랜스포손, 그리고 바이러스 벡터 (가령, 렌티바이러스 벡터, 아데노 연관된 바이러스 벡터 등)를 포함한다. 한 구체예에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 DNA는 플라스미드 벡터 내에 존재한다. 적합한 플라스미드 벡터의 무제한적 실례는 pUC, pBR322, pET, pBluescript, 그리고 이들의 변이체를 포함한다. 상기 벡터는 추가 발현 제어 서열 (가령, 인핸서 서열, Kozak 서열, 폴리아데닐화 서열, 전사 종결 서열 등), 선별가능 마커 서열 (가령, 항

생체 내성 유전자), 복제 기점, 기타 등등을 포함할 수 있다. 추가 정보는 "Current Protocols in Molecular Biology" Ausubel et al., John Wiley & Sons, New York, 2003 또는 "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001에서 발견될 수 있다.

[0084] 일부 구체예에서, RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 서열을 포함하는 발현 벡터는 유도 RNA를 인코딩하는 서열을 더욱 포함할 수 있다. 유도 RNA를 인코딩하는 서열은 일반적으로, 관심되는 세포 또는 배아에서 유도 RNA의 발현을 위해 최소한 하나의 전사 제어 서열에 작동가능하게 연결된다. 가령, 유도 RNA를 인코딩하는 DNA는 RNA 중합효소 III (Pol III)에 의해 인식되는 프로모터 서열에 작동가능하게 연결될 수 있다. 적합한 Pol III 프로모터의 실례에는 포유류 U6, U3, H1, 그리고 7SL RNA 프로모터가 포함되지만 이들에 한정되지 않는다.

[0086] **(IV) RNA-유도된 엔도뉴클레아제를 이용하여 염색체 서열을 변경하기 위한 방법**

[0087] 본 발명의 다른 양상은 진핵 세포 또는 배아에서 염색체 서열을 변경하기 위한 방법을 포괄한다. 상기 방법은 (i) 최소한 하나의 핵 국지화 신호를 포함하는 최소한 하나의 RNA-유도된 엔도뉴클레아제 또는 최소한 하나의 핵 국지화 신호를 포함하는 최소한 하나의 RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산, (ii) 최소한 하나의 유도 RNA 또는 최소한 하나의 유도 RNA를 인코딩하는 DNA, 그리고, 임의선택적으로, (iii) 공여자 서열을 포함하는 최소한 하나의 공여자 폴리뉴클레오티드를 진핵 세포 또는 배아 내로 도입하는 것을 포함한다. 상기 방법은 각 유도 RNA가 염색체 서열 내에 표적화된 부위에 RNA-유도된 엔도뉴클레아제를 향하게 하도록 상기 세포 또는 배아를 배양하는 것을 더욱 포함하고, 여기서 상기 RNA-유도된 엔도뉴클레아제는 표적화된 부위 내에 이중 가닥 절단을 도입하고, 그리고 상기 이중 가닥 절단은 염색체 서열이 변형되도록 DNA 복구 과정에 의해 복구된다.

[0088] 일부 구체예에서, 상기 방법은 1개의 RNA-유도된 엔도뉴클레아제 (또는 인코딩 핵산) 및 1개의 유도 RNA (또는 인코딩 DNA)를 세포 또는 배아 내로 도입하는 것을 포함할 수 있고, 여기서 RNA-유도된 엔도뉴클레아제는 표적화된 염색체 서열 내에 1개의 이중 가닥 절단을 도입한다. 임의선택적 공여자 폴리뉴클레오티드가 존재하지 않는 구체예에서, 염색체 서열 내에 이중 가닥 절단은 비상동성 말단 연결 (NHEJ) 복구 과정에 의해 복구될 수 있다. NHEJ가 오류 가능하기 때문에, 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단의 복구 동안 일어날 수 있다. 따라서, 표적화된 염색체 서열은 변형되거나 또는 비활성화될 수 있다. 가령, 단일 뉴클레오티드 변화 (SNP)가 변경된 단백질 산물을 발생시킬 수 있거나, 또는 코딩 서열의 해독틀에서 이동이 단백질 산물이 만들어지지 않도록 상기 서열을 비활성화시키거나 또는 "녹아웃"시킬 수 있다. 임의선택적 공여자 폴리뉴클레오티드가 존재하는 구체예에서, 공여자 폴리뉴클레오티드 내에 공여자 서열은 이중 가닥 절단의 복구 동안 표적화된 부위에서 염색체 서열과 교환되거나 또는 이것 내로 통합될 수 있다. 가령, 공여자 서열이 각각, 염색체 서열의 표적화된 부위의 상류와 하류 서열과 실제적인 서열 동일성을 갖는 상류와 하류 서열과 측면에서 접하는 구체예에서, 공여자 서열은 상동 재조합 과정에 의해 매개된 복구 동안 표적화된 부위에서 염색체 서열과 교환되거나 또는 이것 내로 통합될 수 있다. 대안으로, 공여자 서열이 양립성 오버행과 측면에서 접하는 (또는 양립성 오버행이 RNA-유도된 엔도뉴클레아제에 의해 원지에서 산출되는) 구체예에서, 공여자 서열은 이중 가닥 절단의 복구 동안 비상동성 복구 과정에 의해 개월된 염색체 서열과 직접적으로 결합될 수 있다. 염색체 서열 내로 공여자 서열의 교환 또는 통합은 표적화된 염색체 서열을 변경하거나 또는 외인성 서열을 세포 또는 배아의 염색체 서열 내로 도입한다.

[0089] 다른 구체예에서, 상기 방법은 2개의 RNA-유도된 엔도뉴클레아제 (또는 인코딩 핵산) 및 2개의 유도 RNA (또는 인코딩 DNA)를 세포 또는 배아 내로 도입하는 것을 포함할 수 있고, 여기서 RNA-유도된 엔도뉴클레아제는 염색체 서열 내에 2개의 이중 가닥 절단을 도입한다. 도면 3B를 참조한다. 이들 두 절단은 여러 염기쌍 내에 있거나, 수십 개의 염기쌍 내에 있거나, 또는 수천 여 개의 염기쌍에 의해 분리될 수 있다. 임의선택적 공여자 폴리뉴클레오티드가 존재하지 않는 구체예에서, 결과의 이중 가닥 절단은 두 개월 부위 사이에 서열이 상실되고 및/또는 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단(들)의 복구 동안 일어날 수 있도록 비상동성 복구 과정에 의해 복구될 수 있다. 임의선택적 공여자 폴리뉴클레오티드가 존재하는 구체예에서, 공여자 폴리뉴클레오티드 내에 공여자 서열은 상동성-기초된 복구 과정에 의해 (가령, 공여자 서열이 각각, 염색체 서열 내에 표적화된 부위의 상류와 하류 서열과 실제적인 서열 동일성을 갖는 상류와 하류 서열과 측면에서 접하는 구체예에서) 또는 비상동성 복

구 과정에 의해 (가령, 공여자 서열이 양립성 오버행과 측면에서 접하는 구체예에서), 이중 가닥 절단의 복구 동안 염색체 서열과 교환되거나 또는 이것 내로 통합될 수 있다.

[0090] 또 다른 구체예에서, 상기 방법은 이중 가닥 서열 (또는 인코딩 핵산)의 한쪽 가닥을 개열하기 위해 변형된 1개의 RNA-유도된 엔도뉴클레아제 및 2개의 유도 RNA (또는 인코딩 DNA)를 세포 또는 배아 내로 도입하는 것을 포함할 수 있고, 여기서 각 유도 RNA는 RNA-유도된 엔도뉴클레아제를 특정한 표적 부위로 향하게 하고, 상기 부위에서 변형된 엔도뉴클레아제는 이중 가닥 염색체 서열의 한쪽 가닥을 개열하고 (즉, 틈내기하고), 그리고 여기서 2개의 틈은 마주보는 가닥 내에 있고 이중 가닥 절단을 구성할 만큼 충분히 근접하게 위치한다. **도면 3A**를 참조한다. 임의선택적 공여자 폴리뉴클레오티드가 존재하지 않는 구체예에서, 결과의 이중 가닥 절단은 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단의 복구 동안 일어날 수 있도록 비상동성 복구 과정에 의해 복구될 수 있다. 임의선택적 공여자 폴리뉴클레오티드가 존재하는 구체예에서, 공여자 폴리뉴클레오티드 내에 공여자 서열은 상동성-기초된 복구 과정에 의해 (가령, 공여자 서열이 각각, 염색체 서열 내에 표적화된 부위의 상류와 하류 서열과 실제적인 서열 동일성을 갖는 상류와 하류 서열과 측면에서 접하는 구체예에서) 또는 비상동성 복구 과정에 의해 (가령, 공여자 서열이 양립성 오버행과 측면에서 접하는 구체예에서), 이중 가닥 절단의 복구 동안 염색체 서열과 교환되거나 또는 이것 내로 통합될 수 있다.

[0091]

[0092] **(a) RNA-유도된 엔도뉴클레아제**

[0093] 상기 방법은 최소한 하나의 핵 국지화 신호를 포함하는 최소한 하나의 RNA-유도된 엔도뉴클레아제 또는 최소한 하나의 핵 국지화 신호를 포함하는 최소한 하나의 RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산을 세포 또는 배아 내로 도입하는 것을 포함한다. 이런 RNA-유도된 엔도뉴클레아제 및 RNA-유도된 엔도뉴클레아제를 인코딩하는 핵산은 각각, 상기 섹션 (I)과 (III)에서 설명된다.

[0094] 일부 구체예에서, RNA-유도된 엔도뉴클레아제는 단리된 단백질로서 세포 또는 배아 내로 도입될 수 있다. 이런 구체예에서, RNA-유도된 엔도뉴클레아제는 단백질의 세포 흡수를 용이하게 하는 최소한 하나의 세포-투과성 도메인을 더욱 포함할 수 있다. 다른 구체예에서, RNA-유도된 엔도뉴클레아제는 mRNA 분자로서 세포 또는 배아 내로 도입될 수 있다. 또 다른 구체예에서, RNA-유도된 엔도뉴클레아제는 DNA 분자로서 세포 또는 배아 내로 도입될 수 있다. 일반적으로, 융합 단백질을 인코딩하는 DNA 서열은 관심되는 세포 또는 배아에서 기능하는 프로모터 서열에 작동가능하게 연결된다. DNA 서열은 선형일 수 있거나, 또는 DNA 서열은 벡터의 부분일 수 있다. 또 다른 구체예에서, 융합 단백질은 상기 융합 단백질 및 유도 RNA를 포함하는 RNA-단백질 복합체로서 세포 또는 배아 내로 도입될 수 있다.

[0095] 대안적 구체예에서, RNA-유도된 엔도뉴클레아제를 인코딩하는 DNA는 유도 RNA를 인코딩하는 서열을 더욱 포함할 수 있다. 일반적으로, RNA-유도된 엔도뉴클레아제 및 유도 RNA를 인코딩하는 각 서열은 각각, 세포 또는 배아에서 RNA-유도된 엔도뉴클레아제 및 유도 RNA의 발현을 허용하는 적절한 프로모터 제어 서열에 작동가능하게 연결된다. RNA-유도된 엔도뉴클레아제 및 유도 RNA를 인코딩하는 DNA 서열은 추가 발현 제어, 조절, 및/또는 처리 서열(들)을 더욱 포함할 수 있다. RNA-유도된 엔도뉴클레아제 및 유도 RNA를 인코딩하는 DNA 서열은 선형이거나 또는 벡터의 부분일 수 있다

[0097] **(b) 유도 RNA**

[0098] 상기 방법은 또한, 최소한 하나의 유도 RNA 또는 최소한 하나의 유도 RNA를 인코딩하는 DNA를 세포 또는 배아 내로 도입하는 것을 포함한다. 유도 RNA는 RNA-유도된 엔도뉴클레아제와 상호작용하여 엔도뉴클레아제를 특정한 표적 부위로 향하게 하고, 상기 부위에서 유도 RNA의 5' 단부는 염색체 서열 내에 특정한 프로토스페이스 서열과 염기쌍을 이룬다.

[0099] 각 유도 RNA는 3가지 영역을 포함한다: 염색체 서열 내에 표적 부위에 상보적인 5' 단부에서 첫 번째 영역, 줄기 루프 구조를 형성하는 두 번째 내부 영역, 그리고 본질적으로 단일 가닥으로 남아있는 세 번째 3' 영역. 각 유도 RNA의 첫 번째 영역은 각 유도 RNA가 융합 단백질을 특정한 표적 부위로 유도하도록 상이하다. 각 유도 RNA의 두 번째와 세 번째 영역은 모든 유도 RNA에서 동일할 수 있다.

[0100] 유도 RNA의 첫 번째 영역은 유도 RNA의 첫 번째 영역이 표적 부위와 염기쌍을 이룰 수 있도록, 염색체 서열 내에 표적 부위에서 서열 (즉, 프로토스페이스 서열)에 상보적이다. 다양한 구체예에서, 유도 RNA의 첫 번째 영역은 약 10개 뉴클레오티드 내지 약 25개 보다 많은 뉴클레오티드를 포함할 수 있다. 가령, 유도 RNA의 첫 번째

영역 및 염색체 서열 내에 표적 부위 사이에 염기 대합의 영역은 길이에서 약 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22, 23, 24, 25개, 또는 25개 보다 많은 뉴클레오타이드 일 수 있다. 예시적인 구체예에서, 유도 RNA의 첫 번째 영역은 길이에서 약 19, 20, 또는 21개 뉴클레오타이드이다.

[0101] 유도 RNA는 또한, 이차 구조를 형성하는 두 번째 영역을 포함한다. 일부 구체예에서, 이차 구조는 줄기 (또는 헤어핀) 및 루프를 포함한다. 루프와 줄기의 길이는 변할 수 있다. 가령, 루프는 길이에서 약 3 내지 약 10개 뉴클레오타이드 범위에서 변할 수 있고, 그리고 줄기는 길이에서 약 6 내지 약 20개 염기쌍 범위에서 변할 수 있다. 줄기는 1 내지 약 10개 뉴클레오타이드의 하나 또는 그 이상의 돌출을 포함할 수 있다. 따라서, 두 번째 영역의 전반적인 길이는 길이에서 약 16 내지 약 60개 뉴클레오타이드 범위에서 변할 수 있다. 예시적인 구체예에서, 루프는 길이에서 약 4개 뉴클레오타이드이고, 그리고 줄기는 약 12개 염기쌍을 포함한다.

[0102] 유도 RNA는 또한, 본질적으로 단일 가닥으로 남아있는 3' 단부에서 세 번째 영역을 포함한다. 따라서, 세 번째 영역은 관심되는 세포 내에 임의의 염색체 서열에 대한 상보성을 갖지 않고, 그리고 유도 RNA의 나머지 부분에 상보성을 갖지 않는다. 세 번째 영역의 길이는 변할 수 있다. 일반적으로, 세 번째 영역은 길이에서 약 4개 보다 많은 뉴클레오타이드이다. 가령, 세 번째 영역의 길이는 길이에서 약 5 내지 약 60개 뉴클레오타이드 범위에서 변할 수 있다.

[0103] 유도 RNA의 두 번째와 세 번째 영역 (또한, 보편적인 또는 골격 영역으로 불린다)의 합동된 길이는 길이에서 약 30 내지 약 120개 뉴클레오타이드 범위에서 변할 수 있다. 한 양상에서, 유도 RNA의 두 번째와 세 번째 영역의 합동된 길이는 길이에서 약 70 내지 약 100개 뉴클레오타이드 범위에서 변한다.

[0104] 일부 구체예에서, 유도 RNA는 3가지 영역 모두를 포함하는 단일 분자를 포함한다. 다른 구체예에서, 유도 RNA는 2개의 별개의 분자를 포함할 수 있다. 첫 번째 RNA 분자는 유도 RNA의 첫 번째 영역 및 유도 RNA의 두 번째 영역의 "줄기"의 한쪽 절반을 포함할 수 있다. 두 번째 RNA 분자는 유도 RNA의 두 번째 영역의 "줄기"의 다른 절반 및 유도 RNA의 세 번째 영역을 포함할 수 있다. 따라서, 이러한 구체예에서, 첫 번째와 두 번째 RNA 분자는 각각, 서로에 상보적인 뉴클레오타이드의 서열을 내포한다. 가령, 한 구체예에서, 첫 번째와 두 번째 RNA 분자는 각각, 다른 서열에 염기쌍을 이루고 기능적 유도 RNA를 형성하는 서열 (약 6 내지 약 20개 뉴클레오타이드의)을 포함한다.

[0105] 일부 구체예에서, 유도 RNA는 RNA 분자로서 세포 또는 배아 내로 도입될 수 있다. RNA 분자는 시험관내에서 전사될 수 있다. 대안으로, RNA 분자는 화학적으로 합성될 수 있다.

[0106] 다른 구체예에서, 유도 RNA는 DNA 분자로서 세포 또는 배아 내로 도입될 수 있다. 이런 사례에서, 유도 RNA를 인코딩하는 DNA는 관심되는 세포 또는 배아에서 유도 RNA의 발현을 위한 프로모터 제어 서열에 작동가능하게 연결될 수 있다. 가령, RNA 코딩 서열은 RNA 중합효소 III (Po1 III)에 의해 인식되는 프로모터 서열에 작동가능하게 연결될 수 있다. 적합한 Po1 III 프로모터의 실례에는 포유류 U6 또는 H1 프로모터가 포함되지만 이들에 한정되지 않는다. 예시적인 구체예에서, RNA 코딩 서열은 생쥐 또는 인간 U6 프로모터에 연결된다. 다른 예시적인 구체예에서, RNA 코딩 서열은 생쥐 또는 인간 H1 프로모터에 연결된다.

[0107] 유도 RNA를 인코딩하는 DNA 분자는 선형 또는 환상일 수 있다. 일부 구체예에서, 유도 RNA를 인코딩하는 DNA 서열은 벡터의 부분일 수 있다. 적합한 벡터는 플라스미드 벡터, 파지미드, 코스미드, 인공/꼬마염색체, 트랜스포손, 그리고 바이러스 벡터를 포함한다. 예시적인 구체예에서, RNA-유도된 엔도뉴클레아제를 인코딩하는 DNA는 플라스미드 벡터 내에 존재한다. 적합한 플라스미드 벡터의 무제한적 실례는 pUC, pBR322, pET, pBluescript, 그리고 이들의 변이체를 포함한다. 벡터는 추가 발현 제어 서열 (가령, 인핸서 서열, Kozak 서열, 폴리아데닐화 서열, 전사 종결 서열 등), 선별가능 마커 서열 (가령, 항생제 내성 유전자), 복제 기점, 기타 등등을 포함할 수 있다.

[0108] RNA-유도된 엔도뉴클레아제와 유도 RNA 둘 모두 DNA 분자로서 세포 내로 도입되는 구체예에서, 각각은 별개의 분자의 부분 (가령, 융합 단백질 코딩 서열을 내포하는 하나의 벡터 및 유도 RNA 코딩 서열을 내포하는 두 번째 벡터)일 수 있거나 또는 둘 모두 동일한 분자의 부분 (가령, 융합 단백질과 유도 RNA 둘 모두에 대한 코딩 (및 조절) 서열을 내포하는 하나의 벡터)일 수 있다.

[0110] **표적 부위**

[0111] 유도 RNA와 함께 RNA-유도된 엔도뉴클레아제는 염색체 서열 내에서 표적 부위에 지향되고, 여기서 RNA-유도된 엔도뉴클레아제는 염색체 서열 내에 이중 가닥 절단을 도입한다. 표적 부위는 상기 서열 바로 뒤에 (하류) 공통 서열이 뒤따른다는 점을 제외하고, 서열 제한을 갖지 않는다. 이러한 공통 서열은 또한, 프로토스케이스 인접한

모티프 (PAM)로서 알려져 있다. PAM의 실례에는 NGG, NGGNG, 그리고 NNAGAAW (여기서 N은 임의의 뉴클레오티드로서 규정되고, 그리고 W는 A 또는 T로서 규정된다)가 포함되지만 이들에 한정되지 않는다. 상기 섹션 (IV)(b)에서 상술된 바와 같이, 유도 RNA의 첫 번째 영역 (5' 단부에서)은 표적 서열의 프로토스페이서에 상보적이다. 전형적으로, 유도 RNA의 첫 번째 영역은 길이에서 약 19 내지 21개 뉴클레오티드이다. 따라서, 일정한 양상에서, 염색체 서열 내에 표적 부위의 서열은 5'-N₁₉₋₂₁-NGG-3'이다. PAM은 이탤릭체로 표시된다.

[0112] 표적 부위는 유전자의 코딩 영역, 유전자의 인트론, 유전자의 제어 영역, 유전자 사이에 비코딩 영역 등에 있을 수 있다. 유전자는 단백질 코딩 유전자 또는 RNA 코딩 유전자일 수 있다. 유전자는 관심되는 임의의 유전자일 수 있다.

[0114] **임의선택적 공여자 폴리뉴클레오티드**

[0115] 일부 구체예에서, 상기 방법은 최소한 하나의 공여자 폴리뉴클레오티드를 배아 내로 도입하는 것을 더욱 포함한다. 공여자 폴리뉴클레오티드는 최소한 하나의 공여자 서열을 포함한다. 일부 양상에서, 공여자 폴리뉴클레오티드의 공여자 서열은 내인성 또는 선천적 염색체 서열에 상응한다. 가령, 공여자 서열은 표적화된 부위에서 또는 이와 가까운 염색체 서열의 부분과 본질적으로 동일할 수 있지만, 최소한 하나의 뉴클레오티드 변화를 포함한다. 따라서, 공여자 서열은 선천적 서열과의 통합 또는 교환 시에, 표적화된 염색체 위치에서 서열이 최소한 하나의 뉴클레오티드 변화를 포함하도록, 표적화된 부위에서 야생형 서열의 변형된 이형을 포함할 수 있다. 가령, 변화는 하나 또는 그 이상의 뉴클레오티드의 삽입, 하나 또는 그 이상의 뉴클레오티드의 결실, 하나 또는 그 이상의 뉴클레오티드의 치환, 또는 이들의 조합일 수 있다. 변형된 서열의 통합의 결과로서, 세포 또는 배아/동물은 표적화된 염색체 서열로부터 변형된 유전자 산물을 생산할 수 있다.

[0116] 다른 양상에서, 공여자 폴리뉴클레오티드의 공여자 서열은 외인성 서열에 상응한다. 본원에서 이용된 바와 같이, "외인성" 서열은 세포 또는 배아에 선천적이지 않은 서열, 또는 세포 또는 배아의 유전체에서 선천적 위치가 상이한 위치에 있는 서열을 지칭한다. 가령, 외인성 서열은 유전체 내로 통합 시에, 세포 또는 배아/동물이 통합된 서열에 의해 코딩된 단백질을 발현할 수 있도록, 외인성 프로모터 제어 서열에 작동가능하게 연결될 수 있는 단백질 코딩 서열을 포함할 수 있다. 대안으로, 외인성 서열은 이의 발현이 내인성 프로모터 제어 서열에 의해 조절되도록 염색체 서열 내로 통합될 수 있다. 다른 반복에서, 외인성 서열은 전사 제어 서열, 다른 발현 제어 서열, RNA 코딩 서열, 기타 등등일 수 있다. 염색체 서열 내로 외인성 서열의 통합은 "녹인"으로 명명된다.

[0117] 당업자에 의해 인지될 수 있는 바와 같이, 공여자 서열의 길이는 변할 수 있고 변할 것이다. 가령, 공여자 서열은 길이에서 여러 뉴클레오티드에서부터 수백 개의 뉴클레오티드 내지 수십만 개의 뉴클레오티드까지 변할 수 있다.

[0118] 상류와 하류 서열을 포함하는 공여자 폴리뉴클레오티드. 일부 구체예에서, 공여자 폴리뉴클레오티드 내에 공여자 서열은 각각, 염색체 서열 내에 표적화된 부위의 상류와 하류에 위치한 서열에 실제적인 서열 동일성을 갖는 상류 서열 및 하류 서열과 측면에서 접한다. 이들 서열 유사성 때문에, 공여자 폴리뉴클레오티드의 상류와 하류 서열은 공여자 서열이 염색체 서열 내로 통합 (또는 이것과 교환)될 수 있도록, 공여자 폴리뉴클레오티드와 표적화된 염색체 서열 사이에 상동성 재조합을 허용한다.

[0119] 상류 서열은 본원에서 이용된 바와 같이, 표적화된 부위의 상류에 염색체 서열과 실제적인 서열 동일성을 공유하는 핵산 서열을 지칭한다. 유사하게, 하류 서열은 표적화된 부위의 하류에 염색체 서열과 실제적인 서열 동일성을 공유하는 핵산 서열을 지칭한다. 본원에서 이용된 바와 같이, 관용구 "실제적인 서열 동일성"은 최소한 약 75% 서열 동일성을 갖는 서열을 지칭한다. 따라서, 공여자 폴리뉴클레오티드에서 상류와 하류 서열은 표적화된 부위의 상류 또는 하류에 서열과 약 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 가질 수 있다. 예시적인 구체예에서, 공여자 폴리뉴클레오티드에서 상류와 하류 서열은 표적화된 부위의 상류 또는 하류에 염색체 서열과 약 95% 또는 100% 서열 동일성을 가질 수 있다. 한 구체예에서, 상류 서열은 표적화된 부위의 즉시 상류에 위치한 (즉, 표적화된 부위에 인접한) 염색체 서열과 실제적인 서열 동일성을 공유한다. 다른 구체예에서, 상류 서열은 표적화된 부위로부터 상류에 약 100개 뉴클레오티드 내에 위치되는 염색체 서열과 실제적인 서열 동일성을 공유한다. 따라서, 예로서, 상류 서열은 표적화된 부위로부터 상류에 약 1 내지 약 20, 약 21 내지 약 40, 약 41 내지 약 60, 약 61 내지 약 80, 또는 약 81 내지 약 100개 뉴클레오티드 내에 위치되는 염색체 서열과 실제적인 서열 동일성을 공유할 수 있다. 한 구체예에서, 하류 서열은 표적화된 부위의 즉시 하류에 위치한 (즉, 표적화된 부위에 인접한) 염색체 서열과 실제적인 서열 동일성을 공유한다. 다른 구체예에서, 하류 서열은 표적화된

부위로부터 하류에 약 100개 뉴클레오티드 내에 위치되는 염색체 서열과 실제적인 서열 동일성을 공유한다. 따라서, 예로서, 하류 서열은 표적화된 부위로부터 하류에 약 1 내지 약 20, 약 21 내지 약 40, 약 41 내지 약 60, 약 61 내지 약 80, 또는 약 81 내지 약 100개 뉴클레오티드 내에 위치되는 염색체 서열과 실제적인 서열 동일성을 공유할 수 있다.

[0120] 각 상류 또는 하류 서열은 길이에서 약 20개 뉴클레오티드 내지 약 5000개 뉴클레오티드 범위에서 변할 수 있다. 일부 구체예에서, 상류와 하류 서열은 약 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, 2500, 2600, 2800, 3000, 3200, 3400, 3600, 3800, 4000, 4200, 4400, 4600, 4800, 또는 5000개 뉴클레오티드를 포함할 수 있다. 예시적인 구체예에서, 상류와 하류 서열은 길이에서 약 50 내지 약 1500개 뉴클레오티드 범위에서 변할 수 있다.

[0121] 표적화된 염색체 서열에 서열 유사성을 갖는 상류와 하류 서열을 포함하는 공여자 폴리뉴클레오티드는 선형 또는 환상일 수 있다. 공여자 폴리뉴클레오티드가 환상인 구체예에서, 이것은 벡터의 부분일 수 있다. 가령, 벡터는 플라스미드 벡터일 수 있다.

[0122] 표적화된 개열 부위(들)을 포함하는 공여자 폴리뉴클레오티드. 다른 구체예에서, 공여자 폴리뉴클레오티드는 RNA-유도된 엔도뉴클레아제에 의해 인식되는 최소한 하나의 표적화된 개열 부위를 부가적으로 포함할 수 있다. 공여자 폴리뉴클레오티드에 부가된 표적화된 개열 부위는 공여자 서열의 상류 또는 하류 또는 상류와 하류 둘 모두에 배치될 수 있다. 가령, 공여자 서열은 RNA-유도된 엔도뉴클레아제에 의한 개열 시에, 공여자 서열이 RNA-유도된 엔도뉴클레아제에 의한 개열 시에 산출된 염색체 서열 내에 것들과 양립하는 오버행과 측면에서 접하도록, 표적화된 개열 부위와 측면에서 접할 수 있다. 따라서, 공여자 서열은 비상동성 복구 과정에 의한 이중 가닥 절단의 복구 동안 개열된 염색체 서열과 결합될 수 있다. 일반적으로, 표적화된 개열 부위(들)를 포함하는 공여자 폴리뉴클레오티드는 환상일 것이다 (가령, 플라스미드 벡터의 부분일 수 있다).

[0123] 임의선택적 오버행을 갖는 짧은 공여자 서열을 포함하는 공여자 폴리뉴클레오티드. 다른 대안적 구체예에서, 공여자 폴리뉴클레오티드는 RNA-유도된 엔도뉴클레아제에 의해 산출된 오버행과 양립하는 임의선택적 짧은 오버행을 갖는 짧은 공여자 서열을 포함하는 선형 분자일 수 있다. 이런 구체예에서, 공여자 서열은 이중 가닥 절단의 복구 동안 개열된 염색체 서열과 직접적으로 결합될 수 있다. 일부 경우에, 공여자 서열은 약 1,000개보다 적거나, 약 500개보다 적거나, 약 250개보다 적거나, 또는 약 100개 보다 적은 뉴클레오티드일 수 있다. 일정한 경우에, 공여자 폴리뉴클레오티드는 평할 말단을 갖는 짧은 공여자 서열을 포함하는 선형 분자일 수 있다. 다른 반복에서, 공여자 폴리뉴클레오티드는 5' 및/또는 3' 오버행을 갖는 짧은 공여자 서열을 포함하는 선형 분자일 수 있다. 오버행은 1, 2, 3, 4, 또는 5개 뉴클레오티드를 포함할 수 있다.

[0124] 전형적으로, 공여자 폴리뉴클레오티드는 DNA일 것이다. DNA는 단일 가닥 또는 이중 가닥 및/또는 선형 또는 환상일 수 있다. 공여자 폴리뉴클레오티드는 DNA 플라스미드, 세균 인공 염색체 (BAC), 효모 인공 염색체 (YAC), 바이러스 벡터, DNA의 선형 조각, PCR 단편, 나신 핵산, 또는 전달 운반체, 예를 들면, 리포솜 또는 폴록사머로 복합화된 핵산일 수 있다. 일정한 구체예에서, 공여자 서열을 포함하는 공여자 폴리뉴클레오티드는 플라스미드 벡터의 부분일 수 있다. 이들 환경 중에서 한 가지에서, 공여자 서열을 포함하는 공여자 폴리뉴클레오티드는 최소한 하나의 추가 서열을 더욱 포함할 수 있다.

[0126] **(e) 세포 또는 배아 내로 도입**

[0127] RNA-표적화된 엔도뉴클레아제(들) (또는 인코딩 핵산), 유도 RNA(들) (또는 인코딩 DNA), 그리고 임의선택적 공여자 폴리뉴클레오티드(들)는 다양한 수단에 의해 세포 또는 배아 내로 도입될 수 있다. 일부 구체예에서, 세포 또는 배아는 형질감염된다. 적합한 형질감염 방법은 인산칼슘-매개된 형질감염, 뉴클레오펙션 (또는 전기천공), 양이온성 중합체 형질감염 (가령, DEAE-덱스트란 또는 폴리에틸렌이민), 바이러스 형질도입, 비로솜 형질감염, 비리온 형질감염, 리포솜 형질감염, 양이온성 리포솜 형질감염, 면역리포솜 형질감염, 비리포솜 지질 형질감염, 덴드리머 형질감염, 열 쇼크 형질감염, 마그네토펙션, 리포펙션, 유전자 총 전달, 임팔렉션, 소노포레이션, 광학적 형질감염, 그리고 핵산의 소유 작용제-증강된 흡수를 포함한다. 형질감염 방법은 당분야에서 널리 공지된다 (가령, "Current Protocols in Molecular Biology" Ausubel et al., John Wiley & Sons, New York, 2003 또는 "Molecular Cloning: A Laboratory Manual" Sambrook & Russell, Cold Spring Harbor Press, Cold Spring Harbor, NY, 3rd edition, 2001을 참조한다). 다른 구체예에서, 이들 분자는 현미주입에 의해 세포 또는 배아 내로 도입된다. 전형적으로, 배아는 관심되는 종의 수태된 단세포 단계 배아이다. 가령, 이들 분자는 단세포

포 배아의 전핵 내로 주입될 수 있다.

- [0128] RNA-표적화된 엔도뉴클레아제 (들) (또는 인코딩 핵산), 유도 RNA(들) (또는 유도 RNA를 인코딩하는 DNA), 그리고 임의선택적 공여자 폴리뉴클레오티드(들)는 동시에 또는 순차적으로 세포 또는 배아 내로 도입될 수 있다. RNA-표적화된 엔도뉴클레아제 (들) (또는 인코딩 핵산) 내지 유도 RNA(들) (또는 인코딩 DNA)의 비율은 일반적으로, 이들이 RNA-단백질 복합체를 형성할 수 있도록 대략 화학양론적일 것이다. 한 구체예에서, RNA-표적화된 엔도뉴클레아제를 인코딩하는 DNA 및 유도 RNA를 인코딩하는 DNA는 플라스미드 벡터 내에서 함께 전달된다.
- [0130] **(f) 세포 또는 배아 배양**
- [0131] 상기 방법은 유도 RNA(들)가 RNA-유도된 엔도뉴클레아제(들)를 염색체 서열 내에 표적화된 부위(들)로 향하게 하고, 그리고 RNA-유도된 엔도뉴클레아제(들)가 염색체 서열 내에 최소한 하나의 이중 가닥 절단을 도입하도록, 세포 또는 배아를 적절한 조건 하에 유지하는 것을 더욱 포함한다. 이중 가닥 절단은 염색체 서열이 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합에 의해 변형되도록 DNA 복구 과정에 의해 복구될 수 있다.
- [0132] 공여자 폴리뉴클레오티드가 세포 또는 배아 내로 도입되는 구체예에서, 이중 가닥 절단은 비상동성 말단 연결 (NHEJ) 복구 과정에 의해 복구될 수 있다. NHEJ가 오류 가능하기 때문에, 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단의 복구 동안 일어날 수 있다. 따라서, 염색체 서열에서 서열이 코딩 영역의 해독틀이 이동될 수 있고, 그리고 염색체 서열이 비활성화되거나 또는 "적중"되도록 변형될 수 있다. 비활성화된 단백질-코딩 염색체 서열은 야생형 염색체 서열에 의해 코딩된 단백질을 발생시키지 못한다.
- [0133] 상류와 하류 서열을 포함하는 공여자 폴리뉴클레오티드가 세포 또는 배아 내로 도입되는 구체예에서, 이중 가닥 절단은 공여자 서열이 염색체 서열 내로 통합되도록 상동 재조합 (HDR) 과정에 의해 복구될 수 있다. 따라서, 외인성 서열이 세포 또는 배아의 유전체 내로 통합될 수 있거나, 또는 표적화된 염색체 서열이 야생형 염색체 서열에 대한 변형된 서열의 교환에 의해 변형될 수 있다.
- [0134] 표적화된 개별 부위를 포함하는 공여자 폴리뉴클레오티드가 세포 또는 배아 내로 도입되는 구체예에서, RNA-유도된 엔도뉴클레아제는 표적화된 염색체 서열과 공여자 폴리뉴클레오티드 둘 모두를 개별할 수 있다. 선형 공여자 폴리뉴클레오티드는 NHEJ 과정을 통해, 공여자 폴리뉴클레오티드 및 개별된 염색체 서열 사이에 결찰에 의해 이중 가닥 절단의 부위에서 염색체 서열 내로 통합될 수 있다.
- [0135] 짧은 공여자 서열을 포함하는 선형 공여자 폴리뉴클레오티드가 세포 또는 배아 내로 도입되는 구체예에서, 짧은 공여자 서열은 NHEJ 과정을 통해, 이중 가닥 절단의 부위에서 염색체 서열 내로 통합될 수 있다. 통합은 짧은 공여자 서열 및 이중 가닥 절단의 부위에서 염색체 서열 사이에 평활 말단의 결찰을 통해 진행할 수 있다. 대안으로, 통합은 개별된 염색체 서열에서 RNA-표적화 엔도뉴클레아제에 의해 산출된 것들과 양립하는 오버행과 측면에서 접하는 짧은 공여자 서열 사이에 접촉 말단 (즉, 5' 또는 3' 오버행을 갖는)의 결찰을 통해 진행할 수 있다.
- [0136] 일반적으로, 세포는 세포 성장 및/또는 유지에 적절한 조건 하에 유지된다. 적합한 세포 배양 조건은 당분야에서 널리 공지되고 예로서, Santiago et al. (2008) PNAS 105:5809-5814; Moehle et al. (2007) PNAS 104:3055-3060; Urnov et al. (2005) Nature 435:646-651; 그리고 Lombardo et al (2007) Nat. Biotechnology 25:1298-1306에서 설명된다. 당업자는 세포를 배양하기 위한 방법이 당분야에서 공지되고 세포 유형에 따라 변할 수 있고 변할 것이라는 것을 인지한다. 모든 사례에서, 특정 세포 유형에 대한 최선의 기술을 결정하기 위해 일과적인 최적화가 이용될 수 있다.
- [0137] 배아는 시험관내에서 배양될 수 있다 (가령, 세포 배양에서). 전형적으로, 배아는 적절한 온도에서, 그리고 필요하면, RNA 엔도뉴클레아제 및 유도 RNA의 발현을 허용하는데 필요한 O₂/CO₂ 비율을 갖는 적절한 배지에서 배양된다. 배지의 적합한 무제한적 실례는 M2, M16, KSOM, BMOC, 그리고 HTF 배지를 포함한다. 당업자는 배양 조건이 배아의 종류에 따라 변할 수 있고 변할 것이라는 것을 인지할 것이다. 모든 사례에서, 특정 종류의 배아에 대한 최선의 배양 조건을 결정하기 위해 일과적인 최적화가 이용될 수 있다. 일부 경우에, 세포주는 시험관내-배양된 배아 (가령, 배아 줄기 세포주)로부터 유래될 수 있다.
- [0138] 대안으로, 배아는 배아를 암컷 숙주의 자궁 내로 이전함으로써 생체내에서 배양될 수 있다. 대체로 말하면, 암컷 숙주는 배아와 동일한 또는 유사한 종으로부터 유래된다. 바람직하게는, 암컷 숙주는 상상 임신이다. 상상

임신 암컷 숙주를 준비하는 방법은 당분야에서 공지된다. 부가적으로, 배아를 암컷 숙주 내로 이전하는 방법은 알려져 있다. 배아를 생체내에서 배양하는 것은 배아가 발달하도록 허용하고, 그리고 배아로부터 유래된 동물의 생존 출생을 유발할 수 있다. 이런 동물은 신체의 모든 세포에서 변형된 염색체 서열을 포함할 것이다.

[0140] (g) 세포와 배아 유형

[0141] 다양한 진핵 세포와 배아가 상기 방법에서 이용하기에 적합하다. 가령, 세포는 인간 세포, 비-인간 포유류 세포, 비포유류 척추동물 세포, 무척추동물 세포, 곤충 세포, 식물 세포, 효모 세포, 또는 단일 세포 진핵 생물체일 수 있다. 일반적으로, 배아는 비-인간 포유류 배아이다. 특정한 구체예에서, 배아는 단세포 비-인간 포유류 배아이다. 단세포 배아를 비롯한 예시적인 포유류 배아는 제한 없이, 생쥐, 쥐, 햄스터, 설치류, 토끼, 고양이, 개, 양, 돼지, 소, 말, 그리고 영장류 배아를 포함한다. 또 다른 구체예에서, 세포는 줄기 세포일 수 있다. 적합한 줄기 세포는 제한 없이, 배아 줄기 세포, ES-유사 줄기 세포, 태아 줄기 세포, 성체 줄기 세포, 만능성 줄기 세포, 유도된 만능성 줄기 세포, 다능성 줄기 세포, 소기능성 줄기 세포, 단분화능 줄기 세포 등을 포함한다. 예시적인 구체예에서, 세포는 포유류 세포이다.

[0142] 적합한 포유류 세포의 무제한적 실례는 중국 햄스터 난소 (CHO) 세포, 아기 햄스터 신장 (BHK) 세포; 생쥐 골수종 NSO 세포, 생쥐 배아 섬유모세포 3T3 세포 (NIH3T3), 생쥐 B 림프종 A20 세포; 생쥐 흑색종 B16 세포; 생쥐 근모세포 C2C12 세포; 생쥐 골수종 SP2/0 세포; 생쥐 배아 중간엽 C3H-10T1/2 세포; 생쥐 암종 CT26 세포, 생쥐 전립선 DuCuP 세포; 생쥐 유방 EMT6 세포; 생쥐 간암 Hepa1c1c7 세포; 생쥐 골수종 J5582 세포; 생쥐 상피 MTD-1A 세포; 생쥐 심근 MyEnd 세포; 생쥐 신장 RenCa 세포; 생쥐 췌장 RIN-5F 세포; 생쥐 흑색종 X64 세포; 생쥐 림프종 YAC-1 세포; 쥐 교모세포종 9L 세포; 쥐 B 림프종 RBL 세포; 쥐 신경모세포종 B35 세포; 쥐 간암 세포 (HTC); 버팔로 쥐 간 BRL 3A 세포; 개 신장 세포 (MDCK); 개 유방 (CMT) 세포; 쥐 골육종 D17 세포; 쥐 단핵구/대식세포 DH82 세포; 원숭이 신장 SV-40 형질전환된 섬유모세포 (COS7) 세포; 원숭이 신장 CVI-76 세포; 아프리카 녹색 원숭이 신장 (VERO-76) 세포; 인간 배아 신장 세포 (HEK293, HEK293T); 인간 경부 암종 세포 (HELA); 인간 폐 세포 (W138); 인간 간 세포 (Hep G2); 인간 U2-OS 골육종 세포, 인간 A549 세포, 인간 A-431 세포, 그리고 인간 K562 세포를 포함한다. 포유류 세포주의 광범위한 목록은 American Type Culture Collection 카탈로그 (ATCC, Manassas, VA)에서 발견될 수 있다.

[0144] (V) 융합 단백질을 이용하여 염색체 서열을 변경하거나 또는 염색체 서열의 발현을 조절하기 위한 방법

[0145] 본 발명의 다른 양상은 세포 또는 배아에서 염색체 서열을 변경하거나 또는 염색체 서열의 발현을 조절하기 위한 방법을 포괄한다. 상기 방법은 다음을 세포 또는 배아 내로 도입하는 것을 포함한다: (a) 최소한 하나의 융합 단백질 또는 최소한 하나의 융합 단백질을 인코딩하는 핵산, 여기서 상기 융합 단백질은 CRISPR/Cas-유사 단백질 또는 이의 단편 및 작동체 도메인을 포함하고, 그리고 (b) 최소한 하나의 유도 RNA 또는 상기 유도 RNA를 인코딩하는 DNA, 여기서 유도 RNA는 융합 단백질의 CRISPR/Cas-유사 단백질을 염색체 서열 내에 표적화된 부위로 유도하고, 그리고 융합 단백질의 작동체 도메인은 염색체 서열을 변경하거나 또는 염색체 서열의 발현을 조절한다.

[0146] CRISPR/Cas-유사 단백질 또는 이의 단편 및 작동체 도메인을 포함하는 융합 단백질은 상기 섹션 (II)에서 상술된다. 일반적으로, 본원에서 개시된 융합 단백질은 최소한 하나의 핵 국지화 신호를 더욱 포함한다. 융합 단백질을 인코딩하는 핵산은 상기 섹션 (III)에서 설명된다. 일부 구체예에서, 융합 단백질은 단리된 단백질로서 세포 또는 배아 내로 도입될 수 있다 (이것은 세포-투과성 도메인을 더욱 포함할 수 있다). 게다가, 단리된 융합 단백질은 유도 RNA를 포함하는 단백질-RNA 복합체의 부분일 수 있다. 다른 구체예에서, 융합 단백질은 RNA 분자로서 세포 또는 배아 내로 도입될 수 있다 (이것은 캡핑되고 및/또는 폴리아데닐화될 수 있다). 또 다른 구체예에서, 융합 단백질은 DNA 분자로서 세포 또는 배아 내로 도입될 수 있다. 가령, 융합 단백질 및 유도 RNA는 구별된 DNA 분자로서 또는 동일한 DNA 분자의 일부로서 세포 또는 배아 내로 도입될 수 있다. 이런 DNA 분자는 플라스미드 벡터일 수 있다.

[0147] 일부 구체예에서, 상기 방법은 최소한 하나의 아연 핑거 뉴클레아제를 세포 또는 배아 내로 도입하는 것을 더욱 포함한다. 아연 핑거 뉴클레아제는 섹션 (II)(d)에서 전술된다. 또 다른 구체예에서, 상기 방법은 최소한 하나의 공여자 폴리뉴클레오티드를 세포 또는 배아 내로 도입하는 것을 더욱 포함한다. 공여자 폴리뉴클레오티드는 상기 섹션 (IV)(d)에서 상술된다. 분자를 세포 또는 배아 내로 도입하기 위한 수단뿐만 아니라 세포 또는 배아를 배양하기 위한 수단은 각각, 섹션 (IV)(e)와 (IV)(f)에서 전술된다. 적합한 세포와 배아는 섹션 (IV)(g)에서 전술된다.

[0148] 융합 단백질의 작동체 도메인이 개열 도메인 (가령, FokI 개열 도메인 또는 변형된 FokI 개열 도메인)인 일정한 구체예에서, 상기 방법은 1개의 융합 단백질 (또는 1개의 융합 단백질을 인코딩하는 핵산) 및 2개의 유도 RNA (또는 2개의 유도 RNA를 인코딩하는 DNA)를 세포 또는 배아 내로 도입하는 것을 포함할 수 있다. 이들 2개의 유도 RNA는 융합 단백질을 염색체 서열 내에 2개의 상이한 표적 부위로 향하게 하고, 여기서 상기 융합 단백질은 2개의 개열 도메인이 이중 가닥 절단을 염색체 서열 내에 도입할 수 있도록 이합체화한다 (가령, 동종이합체를 형성한다). 도면 1A를 참조한다. 임의선택적 공여자 폴리뉴클레오티드가 존재하지 않는 구체예에서, 염색체 서열 내에 이중 가닥 절단은 비상동성 말단 연결 (NHEJ) 복구 과정에 의해 복구될 수 있다. NHEJ가 오류 가능하기 때문에, 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단의 복구 동안 일어날 수 있다. 따라서, 표적화된 염색체 서열은 변형되거나 또는 비활성화될 수 있다. 가령, 단일 뉴클레오티드 변화 (SNP)가 변경된 단백질 산물을 발생시킬 수 있거나, 또는 코딩 서열의 해독틀에서 이동이 단백질 산물이 만들어지지 않도록 상기 서열을 비활성화시키거나 또는 "녹아웃"시킬 수 있다. 임의선택적 공여자 폴리뉴클레오티드가 존재하는 구체예에서, 공여자 폴리뉴클레오티드 내에 공여자 서열은 이중 가닥 절단의 복구 동안 표적화된 부위에서 염색체 서열과 교환되거나 또는 이것 내로 통합될 수 있다. 가령, 공여자 서열이 각각, 염색체 서열의 표적화된 부위의 상류와 하류 서열과 실제적인 서열 동일성을 갖는 상류와 하류 서열과 측면에서 접하는 구체예에서, 공여자 서열은 상동 재조합 과정에 의해 매개된 복구 동안 표적화된 부위에서 염색체 서열과 교환되거나 또는 이것 내로 통합될 수 있다. 대안으로, 공여자 서열이 양립성 오버행과 측면에서 접하는 (또는 양립성 오버행이 RNA-유도된 엔도뉴클레아제에 의해 원지에서 산출되는) 구체예에서, 공여자 서열은 이중 가닥 절단의 복구 동안 비상동성 복구 과정에 의해 개열된 염색체 서열과 직접적으로 결합될 수 있다. 염색체 서열 내로 공여자 서열의 교환 또는 통합은 표적화된 염색체 서열을 변경하거나 또는 외인성 서열을 세포 또는 배아의 염색체 서열 내로 도입한다.

[0149] 융합 단백질의 작동체 도메인이 개열 도메인 (가령, FokI 개열 도메인 또는 변형된 FokI 개열 도메인)인 다른 구체예에서, 상기 방법은 2개의 상이한 융합 단백질 (또는 2개의 상이한 융합 단백질을 인코딩하는 핵산) 및 2개의 유도 RNA (또는 2개의 유도 RNA를 인코딩하는 DNA)를 세포 또는 배아 내로 도입하는 것을 포함할 수 있다. 이들 융합 단백질은 상기 섹션 (II)에서 상술된 바와 같이 상이할 수 있다. 각 유도 RNA는 융합 단백질을 염색체 서열 내에 특정한 표적 부위로 향하게 하고, 여기서 이들 융합 단백질은 2개의 개열 도메인이 이중 가닥 절단을 염색체 서열 내로 도입할 수 있도록 이합체화한다 (가령, 이형이합체를 형성한다). 임의선택적 공여자 폴리뉴클레오티드가 존재하지 않는 구체예에서, 결과의 이중 가닥 절단은 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단의 복구 동안 일어날 수 있도록, 비상동성 복구 과정에 의해 복구될 수 있다. 임의선택적 공여자 폴리뉴클레오티드가 존재하는 구체예에서, 공여자 폴리뉴클레오티드 내에 공여자 서열은 상동성-기초된 복구 과정에 의해 (가령, 공여자 서열이 각각, 염색체 서열 내에 표적화된 부위의 상류와 하류 서열과 실제적인 서열 동일성을 갖는 상류와 하류 서열과 측면에서 접하는 구체예에서) 또는 비상동성 복구 과정에 의해 (가령, 공여자 서열이 양립성 오버행과 측면에서 접하는 구체예에서), 이중 가닥 절단의 복구 동안 염색체 서열과 교환되거나 또는 이것 내로 통합될 수 있다.

[0150] 융합 단백질의 작동체 도메인이 개열 도메인 (가령, FokI 개열 도메인 또는 변형된 FokI 개열 도메인)인 또 다른 구체예에서, 상기 방법은 1개의 융합 단백질 (또는 1개의 융합 단백질을 인코딩하는 핵산), 1개의 유도 RNA (또는 1개의 유도 RNA를 인코딩하는 DNA), 그리고 1개의 아연 핑거 뉴클레아제 (또는 아연 핑거 뉴클레아제를 인코딩하는 핵산)를 세포 또는 배아 내로 도입하는 것을 포함할 수 있고, 여기서 아연 핑거 뉴클레아제는 FokI 개열 도메인 또는 변형된 FokI 개열 도메인을 포함한다. 유도 RNA는 융합 단백질을 특정한 염색체 서열로 향하게 하고, 그리고 아연 핑거 뉴클레아제는 다른 염색체 서열에 지향되고, 여기서 융합 단백질 및 아연 핑거 뉴클레아제는 융합 단백질의 개열 도메인 및 아연 핑거 뉴클레아제의 개열 도메인이 이중 가닥 절단을 염색체 서열 내로 도입할 수 있도록 이합체화한다. 도면 1B를 참조한다. 임의선택적 공여자 폴리뉴클레오티드가 존재하지 않는 구체예에서, 결과의 이중 가닥 절단은 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단의 복구 동안 일어날 수 있도록, 비상동성 복구 과정에 의해 복구될 수 있다. 임의선택적 공여자 폴리뉴클레오티드가 존재하는 구체예에서, 공여자 폴리뉴클레오티드 내에 공여자 서열은 상동성-기초된 복구 과정에 의해 (가령, 공여자 서열이 각각, 염색체 서열 내에 표적화된 부위의 상류와 하류 서열과 실제적인 서열 동일성을 갖는 상류와 하류 서열과 측면에서 접하는 구체예에서) 또는 비상동성 복구 과정에 의해 (가령, 공여자 서열이 양립성 오버행과 측면에서 접하는 구체예에서), 이중 가닥 절단의 복구 동안 염색체 서열과 교환되거나 또는 이것 내로 통합될 수 있다.

- [0151] 융합 단백질의 작동체 도메인이 전사 활성화 도메인 또는 전사 억제인자 도메인인 또 다른 구체예에서, 상기 방법은 1개의 융합 단백질 (또는 1개의 융합 단백질을 인코딩하는 핵산) 및 1개의 유도 RNA (또는 1개의 유도 RNA를 인코딩하는 DNA)를 세포 또는 배아 내로 도입하는 것을 포함할 수 있다. 유도 RNA는 융합 단백질을 특정한 염색체 서열로 향하게 하고, 여기서 전사 활성화 도메인 또는 전사 억제인자 도메인은 각각, 표적화된 염색체 서열의 발현을 활성화시키거나 또는 억제한다. **도면 2A를 참조한다.**
- [0152] 융합 단백질의 작동체 도메인이 후성 변형 도메인인 대안적 구체예에서, 상기 방법은 1개의 융합 단백질 (또는 1개의 융합 단백질을 인코딩하는 핵산) 및 1개의 유도 RNA (또는 1개의 유도 RNA를 인코딩하는 DNA)를 세포 또는 배아 내로 도입하는 것을 포함할 수 있다. 유도 RNA는 융합 단백질을 특정한 염색체 서열로 향하게 하고, 여기서 후성 변형 도메인은 표적화된 염색체 서열의 구조를 변경한다. **도면 2A를 참조한다.** 후성 변형은 아세틸화, 히스톤 단백질의 메틸화 및/또는 뉴클레오티드 메틸화를 포함한다. 일부 경우에, 염색체 서열의 구조적 변형은 염색체 서열의 발현에서 변화를 야기한다.
- [0154] **(VI) 유전적으로 변형된 세포와 동물**
- [0155] 본 발명은 예로서, 본원에서 설명된 방법을 이용하여, RNA-유도된 엔도뉴클레아제-매개된 또는 융합 단백질-매개된 과정을 이용하여 변형된 최소한 하나의 염색체 서열을 포함하는 유전적으로 변형된 세포, 비-인간 배아, 그리고 비-인간 동물을 포괄한다. 본 발명은 관심되는 염색체 서열에 표적화된 RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 최소한 하나의 DNA 또는 RNA 분자 또는 융합 단백질, 최소한 하나의 유도 RNA, 그리고 임의선택적으로 하나 또는 그 이상의 공여자 폴리뉴클레오티드(들)를 포함하는 세포를 제공한다. 본 발명은 또한, 관심되는 염색체 서열에 표적화된 RNA-유도된 엔도뉴클레아제 또는 융합 단백질을 인코딩하는 최소한 하나의 DNA 또는 RNA 분자, 최소한 하나의 유도 RNA, 그리고 임의선택적으로 하나 또는 그 이상의 공여자 폴리뉴클레오티드(들)를 포함하는 비-인간 배아를 제공한다.
- [0156] 본 발명은 최소한 하나의 변형된 염색체 서열을 포함하는 유전적으로 변형된 비-인간 동물, 비-인간 배아, 또는 동물 세포를 제공한다. 변형된 염색체 서열은 (1) 비활성화되고, (2) 변경된 발현을 갖거나 또는 변경된 단백질 산물을 생산하고, 또는 (3) 통합된 서열을 포함하도록 변형될 수 있다. 염색체 서열은 본원에서 설명된 방법을 이용하여, RNA 유도된 엔도뉴클레아제-매개된 또는 융합 단백질-매개된 과정으로 변형된다.
- [0157] 논의된 바와 같이, 본 발명의 한 가지 양상은 최소한 하나의 염색체 서열이 변형된 유전적으로 변형된 동물을 제공한다. 한 구체예에서, 유전적으로 변형된 동물은 최소한 하나의 비활성화된 염색체 서열을 포함한다. 변형된 염색체 서열은 상기 서열이 전사되지 않고 및/또는 기능적 단백질 산물이 생산되지 않도록 비활성화될 수 있다. 따라서, 비활성화된 염색체 서열을 포함하는 유전적으로 변형된 동물은 "녹아웃" 또는 "조건적 녹아웃"으로 명명될 수 있다. 비활성화된 염색체 서열은 결실 돌연변이 (즉, 하나 또는 그 이상의 뉴클레오티드의 결실), 삽입 돌연변이 (즉, 하나 또는 그 이상의 뉴클레오티드의 삽입), 또는 넨센스 돌연변이 (즉, 종결 코돈이 도입되도록 다른 뉴클레오티드에 대한 단일 뉴클레오티드의 치환)를 포함할 수 있다. 돌연변이의 결과로서, 표적화된 염색체 서열이 비활성화되고 기능적 단백질이 생산되지 않는다. 비활성화된 염색체 서열은 외인성으로 도입된 서열을 포함하지 않는다. 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 또는 그 이상의 염색체 서열이 비활성화되는 유전적으로 변형된 동물 역시 본원에서 포함된다.
- [0158] 다른 구체예에서, 변형된 염색체 서열은 변이체 단백질 산물을 코딩하도록 변경될 수 있다. 가령, 변형된 염색체 서열을 포함하는 유전적으로 변형된 동물은 변경된 단백질 산물이 생산되도록, 표적화된 점 돌연변이(들) 또는 다른 변형을 포함할 수 있다. 한 구체예에서, 염색체 서열은 최소한 하나의 뉴클레오티드가 변화되고, 그리고 발현된 단백질이 하나의 변화된 아미노산 잔기를 포함하도록 변형될 수 있다 (미스센스 돌연변이). 다른 구체예에서, 염색체 서열은 하나 이상의 아미노산이 변화되도록, 하나 이상의 미스센스 돌연변이를 포함하도록 변형될 수 있다. 부가적으로, 염색체 서열은 발현된 단백질이 단일 아미노산 결실 또는 삽입을 포함하도록, 3개의 뉴클레오티드 결실 또는 삽입을 갖도록 변형될 수 있다. 변경된 또는 변이체 단백질은 야생형 단백질과 비교하여 변경된 성질 또는 활성, 예를 들면, 변경된 기질 특이성, 변경된 효소 활성, 변경된 운동성 비율 등을 가질 수 있다.
- [0159] 다른 구체예에서, 유전적으로 변형된 동물은 최소한 하나의 염색체로 통합된 서열을 포함할 수 있다. 통합된 서열을 포함하는 유전적으로 변형된 동물은 "녹인" 또는 "조건적 녹인"으로 명명될 수 있다. 염색체로 통합된 서열은 예로서, 이중상동성 단백질, 내인성 단백질, 또는 둘 모두의 조합을 인코딩할 수 있다. 한 구체예에서, 이중상동성 단백질 또는 내인성 단백질을 인코딩하는 서열은 염색체 서열이 비활성화되지만 외인성 서열이 발현되도록, 단백질을 인코딩하는 염색체 서열 내로 통합될 수 있다. 이런 경우에, 이중상동성 단백질 또는 내인성 단

백질을 인코딩하는 서열은 프로모터 제어 서열에 작동가능하게 연결될 수 있다. 대안으로, 이중상동성 단백질 또는 내인성 단백질을 인코딩하는 서열은 염색체 서열의 발현에 영향을 주지 않으면서 염색체 서열 내로 통합될 수 있다. 가령, 단백질을 인코딩하는 서열은 "안전한 항구" 좌위, 예를 들면, Rosa26 좌위, HPRT 좌위, 또는 AAV 좌위 내로 통합될 수 있다. 본 발명은 또한, 단백질(들)을 인코딩하는 서열을 비롯하여, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 또는 그 이상 서열이 유전체 내로 통합되는 유전적으로 변형된 동물을 포괄한다.

[0160] 단백질을 인코딩하는 염색체로 통합된 서열은 관심되는 단백질의 야생형 형태를 인코딩할 수 있거나 또는 단백질의 변경된 이형이 생산되도록 최소한 하나의 변형을 포함하는 단백질을 인코딩할 수 있다. 가령, 질환 또는 장애에 관련된 단백질을 인코딩하는 염색체로 통합된 서열은 생산된 단백질의 변경된 이형이 연관된 장애를 유발하거나 또는 강력하게 하도록 최소한 하나의 변형을 포함할 수 있다. 대안으로, 질환 또는 장애에 관련된 단백질을 인코딩하는 염색체로 통합된 서열은 단백질의 변경된 이형이 연관된 장애의 발달에 대항하여 보호하도록 최소한 하나의 변형을 포함할 수 있다.

[0161] 추가의 구체예에서, 유전적으로 변형된 동물은 기능적 인간 단백질을 인코딩하는 최소한 하나의 염색체로 통합된 서열을 포함하는 "인간화" 동물일 수 있다. 기능적 인간 단백질은 유전적으로 변형된 동물에서 상응하는 오르소로그 없음을 가질 수 있다. 대안으로, 유전적으로 변형된 동물이 유래되는 야생형 동물은 기능적 인간 단백질에 상응하는 오르소로그를 포함할 수 있다. 이러한 경우에, "인간화" 동물 내에 이중상동성 서열은 기능적 단백질이 만들어지지 않도록 비활성화되고, 그리고 "인간화" 동물은 인간 단백질을 인코딩하는 최소한 하나의 염색체로 통합된 서열을 포함한다.

[0162] 또 다른 구체예에서, 유전적으로 변형된 동물은 단백질의 발현 패턴이 변경되도록, 단백질을 인코딩하는 최소한 하나의 변형된 염색체 서열을 포함할 수 있다. 가령, 단백질의 발현을 제어하는 조절 영역, 예를 들면, 프로모터 또는 전사 인자 결합 부위는 단백질이 과다생산되거나, 또는 단백질의 조직 특이적 또는 일시적 발현이 변경되거나, 또는 이들의 조합이 되도록 변경될 수 있다. 대안으로, 단백질의 발현 패턴은 조건적 녹아웃 시스템을 이용하여 변경될 수 있다. 조건적 녹아웃 시스템의 무제한적 실례는 Cre-lox 재조합 시스템을 포함한다. Cre-lox 재조합 시스템은 핵산 분자 내에 특정한 부위 (lox 부위) 사이에 핵산 서열의 재조합을 촉매작용할 수 있는 부위 특이적 DNA 재조합효소인 Cre 재조합효소 효소를 포함한다. 일시적인 조직 특이적 발현을 산출하기 위해 이러한 시스템을 이용하는 방법은 당분야에서 공지된다. 일반적으로, 유전적으로 변형된 동물은 염색체 서열에 접하는 lox 부위에서 산출된다. lox-접하는 염색체 서열을 포함하는 유전적으로 변형된 동물은 이후, Cre 재조합효소를 발현하는 다른 유전적으로 변형된 동물과 교차될 수 있다. lox-접하는 염색체 서열을 포함하는 자손 동물 및 Cre 재조합효소가 이후 생산되고, 그리고 lox-접하는 염색체 서열이 재조합되어, 단백질을 인코딩하는 염색체 서열의 결실 또는 역전을 야기한다. Cre 재조합효소의 발현은 염색체 서열의 일시적으로 및 조건적으로 조절된 재조합을 산출하기 위해 일시적으로 및 조건적으로 조절될 수 있다.

[0163] 이들 구체예 중에서 한 가지에서, 본원에서 개시된 유전적으로 변형된 동물은 변형된 염색체 서열에 대해 이형 접합성일 수 있다. 대안으로, 유전적으로 변형된 동물은 변형된 염색체 서열에 대해 동형접합성일 수 있다.

[0164] 본원에서 개시된 유전적으로 변형된 동물은 하나 이상의 변형된 염색체 서열을 포함하는 동물을 창출하기 위해 또는 하나 또는 그 이상의 변형된 염색체 서열에 대해 동형접합성인 동물을 창출하기 위해 교배될 수 있다. 가령, 동일한 변형된 염색체 서열을 포함하는 2마리 동물이 교배되어 변형된 염색체 서열에 대해 동형접합성 동물을 창출할 수 있다. 대안으로, 상이한 변형된 염색체 서열을 갖는 동물이 교배되어 양쪽 변형된 염색체 서열을 포함하는 동물을 창출할 수 있다.

[0165] 가령, 비활성화된 염색체 서열 유전자 "x"를 포함하는 첫 번째 동물은 인간 유전자 "X" 단백질을 인코딩하는 염색체로 통합된 서열을 포함하는 두 번째 동물과 교차되어, 비활성화된 유전자 "x" 염색체 서열 및 염색체로 통합된 인간 유전자 "X" 서열 둘 모두를 포함하는 "인간화" 유전자 "X" 자손을 발생시킬 수 있다. 또한, 인간화 유전자 "X" 동물은 인간화 유전자 "Y" 동물과 교차되어 인간화 유전자 X/유전자 Y 자손을 창출할 수 있다. 당업자는 많은 조합이 가능하다는 것을 인지할 것이다.

[0166] 다른 구체예에서, 변형된 염색체 서열을 포함하는 동물은 변형된 염색체 서열을 다른 유전자 배경과 합동하기 위해 교배될 수 있다. 무제한적 실례에 의하여, 다른 유전자 배경은 야생형 유전자 배경, 결실 돌연변이를 갖는 유전자 배경, 다른 표적화된 통합을 갖는 유전자 배경, 그리고 비표적화된 통합을 갖는 유전자 배경을 포함할 수 있다.

[0167] 용어 "동물"은 본원에서 이용된 바와 같이, 비-인간 동물을 지칭한다. 동물은 배아, 소아, 또는 성체일 수

있다. 적합한 동물은 척추동물, 예를 들면, 포유동물, 조류, 파충류, 양서류, 조개류, 그리고 어류를 포함한다. 적합한 포유동물의 실례는 제한 없이, 설치류, 반려 동물, 가축, 그리고 영장류를 포함한다. 설치류의 무제한적 실례는 생쥐, 쥐, 햄스터, 게르빌루스쥐, 그리고 기니 피그를 포함한다. 적합한 반려 동물에는 고양이, 개, 토끼, 헤지호그, 그리고 흰담비가 포함되지만 이들에 한정되지 않는다. 가축의 무제한적 실례는 말, 염소, 양, 돼지, 소, 라마, 그리고 알파카를 포함한다. 적합한 영장류에는 카푸친 원숭이, 침팬지, 여우원숭이, 마카크, 마모셋, 타마린, 거미 원숭이, 다람쥐 원숭이, 그리고 버빗 원숭이가 포함되지만 이들에 한정되지 않는다. 조류의 무제한적 실례는 닭, 칠면조, 오리, 그리고 거위를 포함한다. 대안으로, 동물은 무척추동물, 예를 들면, 곤충, 선충, 기타 등등일 수 있다. 곤충의 무제한적 실례는 초파리 (*Drosophila*) 및 모기를 포함한다. 예시적인 동물은 쥐이다. 적합한 쥐 혈통의 무제한적 실례는 Dahl Salt-Sensitive, Fischer 344, Lewis, Long Evans Hooded, Sprague-Dawley, 그리고 Wistar를 포함한다. 한 구체예에서, 동물은 유전적으로 변형된 생쥐가 아니다. 본 발명을 위한 적합한 동물의 각 전술한 반복에서, 동물은 외인성으로 도입된, 무작위로 통합된 트랜스포손 서열을 포함하지 않는다.

[0168] 본 발명의 추가 양상은 최소한 하나의 변형된 염색체 서열을 포함하는 유전적으로 변형된 세포 또는 세포주를 제공한다. 유전적으로 변형된 세포 또는 세포주는 본원에서 개시된 유전적으로 변형된 동물 중에서 한 가지로부터 유래될 수 있다. 대안으로, 염색체 서열은 본원에서 설명된 방법을 이용하여, 상기 본원에서 (동물에서 염색체 서열 변형을 설명하는 단락에서) 설명된 바와 같이 세포에서 변형될 수 있다. 본 발명은 또한, 상기 세포 또는 세포주의 용해물을 포괄한다.

[0169] 일반적으로, 이들 세포는 진핵 세포이다. 적합한 숙주 세포는 곰팡이류 또는 효모, 예를 들면, 피치아 (*Pichia*), 사카로미세스 (*Saccharomyces*), 또는 슈조사카로미세스 (*Schizosaccharomyces*); 곤충 세포, 예를 들면, 스포도프테라 프루기페르다 (*Spodoptera frugiperda*)로부터 SF9 세포 또는 노랑초파리 (*Drosophila melanogaster*)로부터 S2 세포; 그리고 동물 세포, 예를 들면, 생쥐, 쥐, 햄스터, 비-인간 영장류, 또는 인간 세포를 포함한다. 예시적인 세포는 포유류이다. 포유류 세포는 일차 세포일 수 있다. 일반적으로, 이중 가닥 절단에 민감한 임의의 일차 세포가 이용될 수 있다. 이들 세포는 다양한 세포 유형, 예를 들면, 섬유모세포, 근모세포, T 또는 B 세포, 대식세포, 상피 세포, 기타 등등일 수 있다.

[0170] 포유류 세포주가 이용될 때, 상기 세포주는 아직 설명되지 않은 임의의 확립 세포주 또는 일차 세포주일 수 있다. 세포주는 유착성 또는 비유착성이거나, 또는 세포주는 당업자에게 공지된 표준 기술을 이용하여, 유착성, 비유착성 또는 기관형적 성장을 고무하는 조건 하에 성장될 수 있다. 적합한 포유류 세포와 세포주의 무제한적 실례는 본원에서 섹션 (IV)(g)에서 제공된다. 또 다른 구체예에서, 세포는 줄기 세포일 수 있다. 적합한 줄기 세포의 무제한적 실례는 섹션 (IV)(g)에서 제공된다.

[0171] 본 발명은 최소한 하나의 변형된 염색체 서열을 포함하는 유전적으로 변형된 비-인간 배아 역시 제공한다. 염색체 서열은 본원에서 설명된 방법을 이용하여, 상기 본원에서 (동물에서 염색체 서열 변형을 설명하는 단락에서) 설명된 바와 같이 배아에서 변형될 수 있다. 한 구체예에서, 배아는 관심되는 동물 종의 비-인간 수태된 단세포 단계 배아이다. 단세포 배아를 비롯하여, 예시적인 포유류 배아는 제한 없이, 생쥐, 쥐, 햄스터, 설치류, 토끼, 고양이, 개, 양, 돼지, 소, 말, 그리고 영장류 배아를 포함한다.

[0173] **정의**

[0174] 달리 정의되지 않으면, 본원에서 이용된 모든 기술 용어와 과학 용어는 본 발명이 속하는 당해 분야의 평균적 기술자에 의해 통상적으로 이해되는 바와 동일한 의미를 갖는다. 다음 참고문헌은 본 발명에서 이용된 많은 용어의 일반적 정의를 당업자에게 제공한다: Singleton et al., *Dictionary of Microbiology and Molecular Biology* (2nd ed. 1994); *The Cambridge Dictionary of Science and Technology* (Walker ed., 1988); *The Glossary of Genetics*, 5th Ed., R. Rieger et al. (eds.), Springer Verlag (1991); 그리고 Hale & Marham, *The Harper Collins Dictionary of Biology* (1991). 본원에서 이용된 바와 같이, 다음 용어는 달리 특정되지 않으면, 그들에 생득된 의미를 갖는다.

[0175] 본 발명의 원소 또는 이의 바람직한 구체예(들)를 소개할 때, 관사 "a", "an", "the" 및 "상기"는 이들 원소 중에서 하나 또는 그 이상이 있다는 것을 의미하는 것으로 의도된다. 용어 "포함하는", "포함하는" 및 "갖는"은 포괄적인 것으로 의도되고, 그리고 열거된 원소 이외에 추가 원소가 있을 수 있다는 것을 의미한다.

[0176] 본원에서 이용된 바와 같이, 용어 "내인성 서열"은 세포에 선천적인 염색체 서열을 지칭한다.

[0177] 용어 "외인성"은 본원에서 이용된 바와 같이, 세포에 선천적이지 않은 서열, 또는 세포의 유전체 내에 선천적

위치가 상이한 염색체 위치에 있는 염색체 서열을 지칭한다.

- [0178] "유전자"는 본원에서 이용된 바와 같이, 유전자 산물을 인코딩하는 DNA 영역 (엑손과 인트론 포함)뿐만 아니라 이런 조절 서열이 코딩 서열 및/또는 전사된 서열에 인접하는 지에 상관없이 유전자 산물의 생산을 조절하는 모든 DNA 영역을 지칭한다. 따라서, 유전자는 프로모터 서열, 종결인자, 번역 조절 서열, 예를 들면, 리보솜 결합 부위 및 내부 리보솜 유입 부위, 인핸서, 사일런서, 인슐레이터, 경계 원소, 복제 기원, 매트릭스 부착 부위, 그리고 좌위 제어 영역을 포함하지만, 이들에 반드시 한정되지는 않는다.
- [0179] 용어 "이종유래"는 관심되는 세포에 내인성이거나 또는 선천적이지 않은 실체를 지칭한다. 가령, 이종단백질은 외인성 공급원, 예를 들면, 외인성으로 도입된 핵산 서열로부터 유래되거나 또는 이것으로부터 최초 유래되었던 단백질을 지칭한다. 일부 경우에, 이종 단백질은 관심되는 세포에 의해 정상적으로 생산되지 않는다.
- [0180] 용어 "핵산"과 "폴리뉴클레오티드"는 선형 또는 환상 입체형태에서, 그리고 단일- 또는 이중 가닥 형태에서 데옥시리보뉴클레오티드 또는 리보뉴클레오티드 중합체를 지칭한다. 본 발명의 목적으로, 이들 용어는 중합체의 길이에 대하여 한정하는 것으로 해석되지 않는다. 이들 용어는 자연 뉴클레오티드의 공지된 유사체뿐만 아니라 염기, 당 및/또는 인산염 모이어티 (가령, 포스포로티오에이트 중추)에서 변형되는 뉴클레오티드를 포괄할 수 있다. 일반적으로, 특정 뉴클레오티드의 유사체는 동일한 염기-짝짓기 특이성을 갖는다; 즉, A의 유사체는 T와 염기쌍을 이룰 것이다.
- [0181] 용어 "뉴클레오티드"는 데옥시리보뉴클레오티드 또는 리보뉴클레오티드를 지칭한다. 뉴클레오티드는 표준 뉴클레오티드 (즉, 아데노신, 구아노신, 시티딘, 티미딘, 그리고 우리딘) 또는 뉴클레오티드 유사체일 수 있다. 뉴클레오티드 유사체는 변형된 퓨린 또는 피리미딘 염기 또는 변형된 리보오스 모이어티를 갖는 뉴클레오티드를 지칭한다. 뉴클레오티드 유사체는 자연적으로 발생하는 뉴클레오티드 (가령, 이노신) 또는 비자연적으로 발생하는 뉴클레오티드일 수 있다. 뉴클레오티드의 당 또는 염기 모이어티 상에서 변형의 무제한적 실례는 아세틸 기, 아미노 기, 카르복실 기, 카르복시메틸 기, 히드록실 기, 메틸 기, 포스포릴 기, 그리고 티올 기의 부가 (또는 제거)뿐만 아니라 염기의 탄소와 질소 원자의 다른 원자 (가령, 7-데아자 퓨린)로 치환을 포함한다. 뉴클레오티드 유사체는 또한, 디데옥시 뉴클레오티드, 2'-O-메틸 뉴클레오티드, 잠금된 핵산 (LNA), 펩티드 핵산 (PNA), 그리고 모르폴리노를 포함한다.
- [0182] 용어 "폴리펩티드"와 "단백질"은 아미노산 잔기의 중합체를 지칭하기 위해 교체가능하게 이용된다.
- [0183] 핵산과 아미노산 서열 동일성을 결정하기 위한 기술은 당분야에서 공지된다. 전형적으로, 이런 기술은 유전자의 경우에 mRNA의 뉴클레오티드 서열을 결정하고 및/또는 이에 따른 인코딩된 아미노산 서열을 결정하고, 그리고 이들 서열을 두 번째 뉴클레오티드 또는 아미노산 서열과 비교하는 것을 포함한다. 유전체학 서열 역시 이러한 방식으로 결정되고 비교될 수 있다. 일반적으로, 동일성은 각각, 2개의 폴리뉴클레오티드 또는 폴리펩티드 서열의 정확한 뉴클레오티드-대-뉴클레오티드 또는 아미노산-대-아미노산 상응을 지칭한다. 2개 또는 그 이상의 서열 (폴리뉴클레오티드 또는 아미노산)은 그들의 퍼센트 동일성을 결정함으로써 비교될 수 있다. 핵산 또는 아미노산 서열인지에 상관없이, 두 서열의 퍼센트 동일성은 더욱 짧은 서열의 길이에 의해 나뉘셈되고 100에 의해 곱셈된, 두 정렬된 서열 사이에 정확한 매치의 숫자이다. 핵산 서열에 대한 근사 정렬은 Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981)의 국부 상동성 알고리즘에 의해 제공된다. 이러한 알고리즘은 Dayhoff, *Atlas of 단백질 서열s and Structure*, M. O. Dayhoff ed., 5 suppl. 3:353-358, National Biomedical Research Foundation, Washington, D.C., USA에 의해 개발되고, 그리고 Gribskov, *Nucl. Acids Res.* 14(6):6745-6763 (1986)에 의해 정규화된 채점 매트릭스를 이용함으로써 아미노산 서열에 적용될 수 있다. 서열의 퍼센트 동일성을 결정하기 위한 이러한 알고리즘의 예시적인 실행은 Genetics Computer Group (Madison, Wis.)에 의해 "BestFit" 유용성 애플리케이션에서 제공된다. 서열 사이에 퍼센트 동일성 또는 유사성을 계산하기 위한 다른 적합한 프로그램은 당분야에서 전반적으로 공지되어 있다, 예로서, 다른 정렬 프로그램은 디폴트 파라미터에서 이용된 BLAST이다. 가령, BLASTN과 BLASTP는 다음 디폴트 파라미터를 이용하여 이용될 수 있다: 유전자 코드 = 표준; 필터 = 없음; 가닥 = 둘 모두; 컷오프 = 60; 예상 = 10; 매트릭스 = BLOSUM62; 설명 = 50 개 서열; 분류 = 높은 점수; 데이터베이스 = 비다중, GenBank+EMBL+DDBJ+PDB+GenBank CDS translations+Swiss 단백질+Spupdate+PIR. 이들 프로그램의 상세는 GenBank 웹사이트 상에서 발견될 수 있다.
- [0184] 다양한 변화가 발명의 범위를 벗어나지 않으면서 상기 설명된 세포와 방법에서 만들어질 수 있었기 때문에, 상기 설명에서 및 아래 제공된 실시예에서 내포된 모든 물질은 예시적인 것으로 해석되고 제한하는 의미로 해석되지 않는 것으로 의도된다.

[0186] **실시예**

[0187] 다음 실시예는 본 발명의 일정한 양상을 예증한다.

[0189] **실례 1: 포유류 발현을 위한 Cas9 유전자의 변형**

[0190] 스트렙토코쿠스 피오게네스 (스트렙토코쿠스 피오게네스) 균주 MGAS15252 (수탁 번호 YP_005388840.1)로부터 Cas9 유전자는 포유류 세포에서 이의 번역을 증강하기 위해 호모사피엔스 코돈 선호로 최적화되었다. Cas9 유전자는 또한, 상기 단백질을 포유류 세포의 핵 내로 표적화하기 위해 C 말단에서 핵 국지화 신호 PKKKRKV (서열 번호:1)를 부가함으로써 변형되었다. 표 1은 변형된 Cas9 아미노산 서열을 제공하는데, 핵 국지화 서열은 밑줄 그어진다. 표 2는 코돈 최적화된, 변형된 Cas9 DNA 서열을 제공한다.

표 1

[0191] 변형된 Cas9 아미노산 서열

QEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVL SMPQVNI VKKTEVQTGGFSKESILPKRNSDKL IARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGGKSKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDL I IKLPKYSLFELENGRKR MLASAGELQKGNELALPSKYVNFYLYLASHYEKLGSPEDNEQKQLFVEQHKHYLDEI IEQISEFSKRVI LADANLDKVL SAYNKHRDKPIREQ AENI IHLFTLTNLGAPAAFYFDTTIDRKRYTSTKEVLDATLIHQSI TGLYETRIDLSQLGGDPKKKRKV (서열 번호:9)

표 2

[0192] 최적화된 Cas9 DNA 서열 (5'-3')

CCCCAAGAAAAAGCGCAAAGTG (서열 번호:10)

[0193] 변형된 Cas9 DNA 서열은 포유류 세포에서 구성적 발현을 위해 시토메갈로바이러스 (CMV) 프로모터의 제어 하에 배치되었다. 변형된 Cas9 DNA 서열은 또한, T7 RNA 중합효소로 시험관내 mRNA 합성을 위해 T7 프로모터의 제어 하에 배치되었다. 시험관내 RNA 전사는 MessageMAX T7 ARCA-캡핑된 Message 전사 키트 및 T7 mScript 표준 mRNA 생산 시스템 (Cellscript)을 이용함으로써 수행되었다. **실시예 2: Cas9 표적화**

[0194] 아데노 연관된 바이러스 통합 부위 1 (AAVS1) 좌위가 Cas9-매개된 인간 유전체 변형을 위한 표적으로서 이용되었다. 인간 AAVS1 좌위는 단백질 포스파타아제 1, 조절 아단위 12C (PPP1R12C)의 인트론 1 (4427 bp)에서 위치된다. 표 3은 PPP1R12C의 첫 번째 엑손 (음영된 회색) 및 첫 번째 인트론을 제공한다. 인트론 내에 밑줄 그어진 서열은 표적화된 변형 부위 (즉, AAVS1 좌위)이다.

표 3

PPP1R12C의 첫 번째 엑손과인트론 (5'-3')

```

GCGGGCGGGCGGTGCGATGTCCGGAGAGGATGGCCCGGGCGGCTGGCCCGGGG
GCGGCGGGCGGGCGGCTGCCCGGGAGCGGCGACGGGAGCAGCTGCGGCAGTGGG
GGGCGCGGGCGGGCGCCGAGCCTGGCCCCGGAGAGCGCCGCGCCCGCACCGT
CCGCTTCGAGCGCGCCGCCGAGTTCCTGGCGGCCTGTGCGGGCGGGCAGCTGG
ACGAGGCGCGTCTGATGCTGCGCGCCGCCGACCCTGGCCCCGGCGCCGAGCTC
GACCCCGCCGCGCCGCCGCCGCCGCCGCCGCGCCGTGCTGGACTCCACCAACGCCGA
CGGTATCAGCGCCCTGCACCAGGTCAGCGCCCCCGCCCGGCGTCTCCCGGGGC
CAGGTCCACCCCTGCTGCGCCACCTGGGGCATCCTCCTTCCCCGTTGCCAGTCT
CGATCCGCCCCCGTTCGTTCTGGCCCTGGGCTTTGCCACCCTATGCTGACACCCCG
TCCCAGTCCCCCTTACCATTCCCCTTCGACCACCCCACTCCGAATTGGAGCCGCT
TCAACTGGCCCTGGGCTTAGCCACTCTGTGCTGACCACTCTGCCCCAGGCCTCCT
TACCATTCCCCTTCGACCTACTCTTCCGCATTGGAGTCGCTTAACTGGCCCTG
GCTTTGGCAGCCTGTGCTGACCCATGCAGTCCTCCTTACCATCCCTCCCTCGACTT
CCCCTCTTCCGATGTTGAGCCCCCTCAGCCGGTCTGGACTTTGTCTCCTTCCCTG
CCCTGCCCTCTCCTGAACCTGAGCCAGCTCCCATAGCTCAGTCTGGTCTATCTGCC
TGCCCTGGCCATTGCACTTTGCGCTGCCCTCCTCTCGCCCCCGAGTGCCCTTG
CTGTGCCGCCGGAACCTCTGCCCTCTAACGCTGCCGTCTCTCTCCTGAGTCCGGAC
CACTTTGAGCTCTACTGGCTTCTGCGCCGCTCTGGCCCACTGTTTCCCCTTCCA
GGCAGGTCTGCTTTCTGACCTGCATTCTCTCCCCTGGGCCTGTGCCGCTTTCT
GTCTGCAGCTTGTGGCCTGGGTCACCTCTACGGCTGGCCCAGATCCTTCCCTGCC
GCCTCCTTCAGGTTCCGTCTTCCCTCCACTCCCTCTTCCCCTTGCTCTGTGCTGT
TGCTGCCCAAGGATGCTCTTCCGGAGCACTTCTTCTCGGCGCTGCACCACGTG
ATGTCTCTGAGCGGATCCTCCCCGTGTCTGGGTCTCTCCGGGCATCTCTCTC
CCTCACCCAACCCCATGCCGTCTTCACTCGCTGGGTTCCCTTTTCTTCTCTCT
GGGGCCTGTGCCATCTCTCGTTTCTTAGGATGGCCTTCTCCGACGGATGTCTCCCT
TGCGTCCCGCCTCCCCTTCTTGTAGGCCTGCATCATCACCGTTTTTCTGGACAACC
CCAAAGTACCCCGTCTCCCTGGCTTAGCCACCTCTCCATCCTCTTGCTTTCTTTG
CCTGGACACCCCGTTCTCCTGTGGATTCCGGGTCACCTCTCACTCCTTTCATTTGGG
CAGCTCCCCTACCCCTTACCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCCCTGT
CATGGCATCTTCCAGGGGTCCGAGAGCTCAGCTAGTCTTCTTCTCCAACCCGGG
CCCCTATGTCCACTTCAGGACAGCATGTTTGTGCCTCCAGGGATCCTGTGTCCCC
GAGCTGGGACCACCTTATATCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGT
ACTTTTATCTGTCCCCTCCACCCACAGTGGGGCCACTAGGGACAGGATTGGTGA
CAGAAAAGCCCCATCCTTAGGCCTCCTCCTTCTAGTCTCCTGATATTGGGTCTAA
CCCCACCTCCTGTTAGGCAGATTCTTATCTGGTGACACACCCCATTTCTGGA
GCCATCTCTCCTTGCCAGAACCTCTAAGGTTTGTACGATGGAGCCAGAGAGG
ATCCTGGGAGGGAGAGCTTGGCAGGGGGTGGGAGGGAAGGGGGGGATGCGTGA
CCTGCCCGGTTCTCAGTGGCCACCCTGCGCTACCCTCTCCAGAACCTGAGCTGC
TCTGACGCGGCCGTCTGGTGCCTTCACTGATCCTGGTGCTGCAGCTTCTTACA
CTTCCCAAGAGGAGAAGCAGTTTGGAAAAACAAAATCAGAATAAGTTGGTCTGAG
TTCTAACTTTGGCTCTTCACTTTCTAGTCCCCAATTTATATTGTTCTCCGTGCGT
CAGTTTTACCTGTGAGATAAGGCCAGTAGCCAGCCCCGTCTGGCAGGGCTGTGG
TGAGGAGGGGGGTGTCCGTGTGGAAAACCTCCTTTGTGAGAATGGTGCGTCTAG
    
```

[0195]

```

GTGTTACCAGGTCGTGGCCGCCTCTACTCCCTTTCTCTTTCTCCATCCTTCTTTCC
TTAAAGAGTCCCCAGTGCTATCTGGGACATATTCCTCCGCCAGAGCAGGGTCCC
GCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGTTTGTAGTCCTTGGCAAGCCCAGG
AGAGGGCGCTCAGGCTTCCCTGTCCCCCTTCCCTCGTCCACCATCTCATGCCCTGG
CTCTCCTGCCCTTCCCTACAGGGGTTCTTGCTCTGCTCTTCAGACTGAGCCCC
GTTCCCCTGCATCCCCGTTCCCCTGCATCCCCCTTCCCCTGCATCCCCCAGAGGC
CCCAGGCCACCTACTTGGCCTGGACCCACGAGAGGCCACCCAGCCCTGTCTA
CCAGGCTGCCTTTTGGGTGGATTCTCCTCCAAGTGTGGGGTGAAGTGTGGCAA
CTCACTCTTCGGGGTATCCCAGGAGGCCTGGAGCATTGGGGTGGGCTGGGGTTC
AGAGAGGAGGGATTCCCTTCTCAGGTTACGTGGCCAAGAAGCAGGGGAGCTGGG
TTTGGGTCAGGCTCTGGGTGTGGGGTGACCAGCTTATGCTGTTTGGCCAGGACAGC
CTAGTTTGTAGACTGAAACCCTCAGTCCTAGGAAAACAGGGATGGTTGGTCACTGT
CTCTGGGTGACTCTTGATTCCCGGCCAGTTTCTCCACCTGGGGCTGTGTTTCTCGT
CCTGCATCCTTCTCCAGGCAGGTCCCCAAGCATCGCCCCCTGCTGTGGCTGTTCC
CAAGTCTTAGGGTACCCCAAGTGGGTTTATCAACCACTTGGTGAGGCTGGTACC
CTGCCCCATTCTGACCCCAATTGCCTTAGTGGCTAGGGGGTTGGGGGCTAGA
GTAGGAGGGGCTGGAGCCAGGATTCTTAGGGCTGAACAGAGAAGAGCTGGGGGC
CTGGGCTCCTGGGTTTGTAGAGAGGAGGGGCTGGGGCCTGGACTCCTGGGTCCGA
GGGAGGAGGGGCTGGGGCCTGGACTCCTGGGTCTGAGGGTGGAGGGACTGGGG
GCCTGGACTCCTGGGTCCGAGGGAGGAGGGGCTGGGGCCTGGACTCGTGGGTCT
TGAGGGAGGAGGGGCTGGGGCCTGGACTTCTGGGTCTTAGGGAGGCGGGGCT
GGGCCTGGACCCCTGGGTCTGAATGGGGAGAGGCTGGGGCCTGGACTCCTTCA
TCTGAGGGCGAAGGGCTGGGGCCTGGCCTCCTGGGTGAATGGGGAGGGGTTG
GGCCTGGACTCTGGAGTCCCTGGTGGCCAGGCCTCAGGCATCTTTCACAGGGATG
CCTGTACTGGGCAGGTCCTTGAAAGGGAAAGGCCATTGCTCTCCTTGGCCCCCT
CCCCTATCGCCATGACAAGTGGGTGGAAATAAACGAGCCGAGTTCATCCCGTTCC
CAGGGCACGTGCGGCCCTTACAGCCGAGTTTCCATGACCTCATGCTCTTGGC
CCTCGTAGCTCCCTCCCGCCTCCTCCAGATGGGCAGCTTGGAGAGGTGAGGGAC
TTGGGGGTAATTTATCCCGTGGATCTAGGAGTTTAGCTTCACTCCTTCCCTCAGCT
CCAGTTCAGGTCCTCCGAGCCCACCCAGTGTCCACAAGGCCTGGGGCAAGTCCCT
CCTCCGACCCCTGGACTTCGGGCTTTGTCCCCCAAGTTTGGACCCCTAAGGG
AAGAATGAGAAACGGTGGCCCGTGTGACGCCCTGGCTGCAGGGCCCCGTGCAGA
GGGGGCTCAGTGAAGTGGAGTGTGACAGCCTGGGGCCAGGCACACAGGTGTG
CAGCTGTCTCACCCCTCTGGGAGTCCCGCCAGGCCCTGAGTCTGTCCAGCAC
AGGGTGGCCTTCTCCACCCTGCATAGCCCTGGGCCACGGCTTCGTTCTGCAG
AGTATCTGCTGGGGTGGTTTCCGAGCTTACCCTTGAAGGACCTGGCTGGGTTT
AAGCAGGAGGGGCTGGGGCCAGGACTCCTGGCTCTGAAGGAGGAGGGGCTG
GAACCTCTTCCCTAGTCTGAGCACTGGAAGCGCCACCTGTGGGTGGTACGGGG
GTTTTGCCGTGTCTAACAGGTACCATGTGGGGTTCGCGACCCAGATGAGAAGCC
CCCTCCCTTCCCCGTTCACTTCTGTTTGCAGATAGCCAGGAGTCTTTCGTGGTT
TCCACTGAGCACTGAAGGCCTGGCCGGCCTGACCACTGGGCAACCAGGCGTATC
TTAAACAGCCAGTGGCCAGAGGCTGTTGGGTCATTTTCCCCACTGTCTTAGCACC
GTGTCCCTGGATCTGTTTTCTGGTCCCTCTGGAGTCCCGACTTGTGGGACAC
CGTGGCTGGGGTAGGTGCGGCTGACGGCTGTTTCCACCCCCAG (서열 번호: 11)
    
```

[0196]

[0197]

Cas9 유도 RNA는 인간 AAVS1 좌위를 표적으로 하도록 설계되었다. 표적 인식 서열 (즉, 표적 서열의 비코딩 가닥에 상보적인 서열) 및 프로토스페이서 서열을 포함하는 42개 뉴클레오티드 RNA (본원에서 "crRNA" 서열로 지칭됨) (5'에서 3'); crRNA의 3' 서열에 상보성을 갖는 5' 서열 및 추가 헤어핀 서열을 포함하는 85개 뉴클레오티드 RNA (본원에서 "tracrRNA" 서열로 지칭됨); 그리고 crRNA의 뉴클레오티드 1-32, GAAA 루프 및 tracrRNA의 뉴클레오티드 19-45를 포함하는 키메라 RNA가 제조되었다. crRNA는 Sigma-Aldrich에 의해 화학적으로 합성되었다. tracrRNA 및 키메라 RNA는 T7-Scribe 표준 RNA IVT 키트 (Cellscript)를 이용하여, T7 RNA 중합효소로 시험관내 전사에 의해 합성되었다. 키메라 RNA 코딩 서열 역시 인간 세포에서 생체내 전사를 위해 인간 U6 프로모터의 제어 하에 배치되었다. 표 4는 유도 RNA의 서열을 제공한다.

표 4

[0198]

유도 RNA

RNA	5'-3' 서열	서열 번호:
AAVS1-crRNA	ACCCACAGUGGGGCCACUAGUUUAGAGCUAUGCUGUUUUG	12
tracrRNA	GGAACCAUUCAAAACAGCAUAGCAAGUAAAAUAAGGCUAGUCCGUUAUCAACUUGAAAAA GUGGCACCGAGUCGGUGCUUUUUUU	13
키메라 RNA	ACCCACAGUGGGGCCACUAGUUUAGAGCUAGAAAUAAGCAAGUAAAAUAAGGCUAGUCC G	14

[0199]

실시예 3: 유전체 변형을 모니터링하기 위한 공여자 폴리뉴클레오티드의 제조 PPP1R12C의 N 말단 내로 GFP 단백질의 표적화된 통합이 Cas9-매개된 유전체 변형을 모니터링하는데 이용되었다. 상동성 제조법에 의한 통합을 매개하기 위해, 공여자 폴리뉴클레오티드가 제조되었다. AAVS1-GFP DNA 공여자는 5' (1185 bp) AAVS1 좌위 상동성 팔, RNA 스플라이싱 수용체, 터보 GFP 코딩 서열, 3' 전사 종결인자, 그리고 3' (1217 bp) AAVS1 좌위 상동성 팔을 내포하였다. 표 5는 RNA 스플라이싱 수용체 및 GFP 코딩 서열, 그 이후에 3' 전사 종결인자의 서열을 제공한다. 플라스미드 DNA는 GenElute 내독소-없는 플라스미드 Maxiprep 키트 (Sigma)를 이용함으로써 제조되었다.

표 5

[0200]

AAVS1-GFP DNA 공여자 서열에서 서열

	5'-3' 서열	서열 번호:
RNA 스플라이싱 수용체	CTGACCTCTTCTCTTCTCCACAG	15
GFP 코딩 서열과 전사 종결인자	GCCACCATGGACTACAAAGACGATGACGACAAGGTCGACTCTAGAGCTGCAGAG AGCGACGAGAGCGGCTGCCCGCCATGGAGATCGAGTGCCGCATCACCGGCACC CTGAACGGCGTGGAGTTCGAGCTGGTGGCGGCGAGAGGGCACCCCGAGCAG GGCCGCATGACCAACAAGATGAAGAGCACCAAGGGCGCCTGACCTTCAGCCCC TACCTGCTGAGCCACGTGATGGCTACGGCTTCTACACTTCGGCACCTACCCC AGCGGCTACGAGAACCCTTCTGCACGCCATCAACAACGGCGGCTACACCAAC ACCCGATCGAGAAGTACGAGGACGGCGGCGTGTGCACGTGAGCTTCAGTAC CGCTACGAGGCCGCGCTGATCGGCGACTTCAAGGTGATGGCACCGGCTTC CCCAGGACAGCGTGATCTTACCGACAAGATCGTCCGAGCAACGCCACCGTG GAGCACTGCACCCATGGGCGATAACGATCTGGATGGCAGCTTACCCCGCACC TTCAGCTCGCGCAGCGGCTACTACAGCTCCGTGGTGACGCCACATGCAC TTCAAGAGCGCCATCCACCCAGCATCCTGCAGAACGGGGCCCCATGTTCCGCC TTCCGCGCGTGGAGGAGGATCACAGCAACCCGAGCTGGGCATCGTGGAGTAC CAGCACGCCCTTCAAGACCCCGGATCGAGATGCCGGTGAAGAATGAAGATCTCTG TGCTTCTAGTTGCCAGCCATCTGTTGTTGCCCTCCCCCGTGCCTTCTTGA CCCTGGAAGGTGCCACTCCACTGTCCTTTCCTAATAAAATGAGGAAATTCAT CGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTGGGGTGGGGCAGGACA GCAAGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCT CTATGGACTCGAGGTTTAAACGTCGACGCGGCCGCT	16

[0201]

표적화된 유전자 통합은 PPP1R12C의 첫 107개 아미노산 및 터보 GFP 사이에 융합 단백질을 유발할 것이다. 예상된 융합 단백질은 PPP1R12C의 첫 번째 엑손 및 가공된 스플라이스 수용체 사이에 RNA 스플라이싱으로부터 PPP1R12C의 첫 107개 아미노산 잔기 (회색으로 강조됨)를 내포한다 (표 6 참조).

표 6

[0202]

PPP1R12C-GFP 융합 단백질의 예측된 아미노산 서열.

```
MSGEDGPAAGPGAAAAAARRRRERLQWGARAGAEPGGERARRTVRFERAAEFLLAACAGGDLDEARLMLRAADPGGAELDPAAPPPARAV
LDSTNADG I SALHQATMDYKDDDDKVD SRAAESDESGLPAMEIECRI TGLNLFVEFELVGGGEGTPEQGRMTNKMKSTKGALTFSPYLLSHVM
GYGFYHFGTYPSTYENPFLHAINNGGYTNTRIEKYEDGGV LHV SF SYRYEAGRVI GDFKVMGTGFPEDSVIFTDKIVRSNATVEHLHPMGDND
LDGSFTRFSLRDGGYSSVVD SHMHFKSAIHPSILQNGGPMFAFRRVEEDHSNTELGIVEYQHAFKTPDADAGEE (서열 번호: 17)
```

[0203]

실시예 4: Cas9-매개된 표적화된 통합형질감염이 인간 K562 세포에서 수행되었다. K562 세포주는 American Type Culture Collection (ATCC)으로부터 획득되고, 그리고 10% FBS와 2 mM L-글루타민으로 보충된 Iscove의 변형된 Dulbecco의 배지에서 성장되었다. 모든 배지와 보충물은 Sigma-Aldrich로부터 획득되었다. 배양액은 형질감염

하루 전에 분할되었다 (형질감염 이전 mL당 대략 0.5 백만 세포에서). 세포는 T-016 프로그램에 따라 뉴클레오펙터 (Lonza) 상에서 뉴클레오펙터 용액 V (Lonza)로 형질감염되었다. 각 뉴클레오펙션은 대략 0.6 백만 세포를 내포하였다. 형질감염 처리는 표 7에서 상술된다. 세포는 뉴클레오펙션 직후에 37 °C 및 5% CO₂에서 성장되었다.

표 7

형질감염 처리.

처리	변형된 Cas9	유도 RNA	공여자 서열
A	안티 리버스 Cap 유사체와 함께 전사된 Cas9 mRNA (10 µg)	미리 어닐링된 crRNA-tracrRNA 이중나선 (0.3 nmol)	AAVS1-GFP 플라스미드 DNA (10 µg)
B	안티 리버스 Cap 유사체와 함께 전사된 Cas9 mRNA (10 µg)	키메라 RNA (0.3 nmol)	AAVS1-GFP 플라스미드 DNA (10 µg)
C	전사후 캡핑 반응을 통해 캡핑된 Cas9 mRNA (10 µg)	키메라 RNA (0.3 nmol)	AAVS1-GFP 플라스미드 DNA (10 µg)
D	Cas9 플라스미드 DNA (10 µg)	U6-키메라 RNA 플라스미드 DNA (5 µg)	AAVS1-GFP 플라스미드 DNA (10 µg)
E	없음	없음	AAVS1-GFP 플라스미드 DNA (10 µg)
F	없음	없음	없음

[0204]

[0205]

형광-활성화된 세포 분류 (FACS)가 형질감염 후 4 일에 수행되었다. FACS 데이터는 **도면 4**에서 제공된다. 4가지 실험적 처리 (A-D) 각각에서 검출된 퍼센트 GFP는 대조 처리 (E, F)에서보다 보다 컸는데, 이것은 공여자 서열의 통합 및 융합 단백질의 발현을 확증한다. **실시예 5: 표적화된 통합의 PCR 확증**

[0206]

유전체학 DNA는 형질감염 후 12 일에 GenElute 포유류 유전체학 DNA Miniprep 키트 (Sigma)로, 형질감염된 세포로부터 추출되었다. 유전체학 DNA는 이후, AAVS1-GFP 플라스미드 공여자의 5' 상동성 팔 외측에 위치한 전방 프라이머 및 GFP의 5' 영역에서 위치한 후방 프라이머로 PCR 증폭되었다. 전방 프라이머는 5'-CCACTCTGTGCTGACCACTCT-3' (서열 번호:18)이고, 그리고 후방 프라이머는 5'- GCGGCACTCGATCTCCA-3' (서열 번호:19)이었다. 접합부 PCR로부터 예상된 단편 크기는 1388 bp이었다. 증폭은 다음 순환 조건을 이용하여, JumpStart Taq ReadyMix (Sigma)로 수행되었다: 초기 변성을 위한 98°C에서 2 분; 98°C에서 15 초, 62°C에서 30 초, 그리고 72°C에서 1분 30 초의 35회 주기; 그리고 72°C에서 5 분 동안 최종 신장. PCR 산물은 1% 아가로오스 겔 상에서 분해되었다.

[0207]

안티 리버스 Cap 유사체와 함께 전사된Cas9 mRNA 10 µg, 0.3 nmol의 미리 어닐링된 crRNA-tracrRNA 이중나선, 그리고 10 µg의 AAVS1-GFP 플라스미드 DNA로 형질감염된 세포는 예상된 크기의 PCR 산물을 전시하였다 (참조: 라인 A, **도면 5**).

[0209]

실시예 6: 생쥐 배아에서 Cas9-기초된 유전체 편집

[0210]

생쥐 Rosa26 좌위가 유전체 변형을 위해 표적화될 수 있다. 표 8은 잠재적 표적 부위가 굵은 글씨체로 표시되는 생쥐 Rosa26 서열의 부분을 제공한다. 각 표적 부위는 프로토스페이서를 포함한다.

표 8

생쥐 Rosa26 서열

[0211]

```
GAGCGGCTGCGGGCGGGTGAAGCAGTTCGACTTGAGTTGCTCAAGAGGGCGTGCTGAGCCAGACCTCCATCGCGACTCCGGGGAG
TGGAGGAAGGAGCGAGGGCTCAGTTGGGCTGTTTTGGAGGCAGGAAGCACTTGCTCTCCAAAGTCGCTGAGTTGTTATCAGTAAGGGAG
CTGCAGTGGAGTAGGCGGGGAGAAGGCCGACCTTCTCCGAGGGGGGAGGGGAGTGTGCAATACCTTTCTGGGAGTTCTGCTGCCTCC
TGGCTTCTGAGGACCCCTGGGCTGGGAGAATCCCTTCCCCCTTTCCTCGTGATCTGCAACTCCAGTCTTTCTAGAAGATGGGCGGGAG
TCTTCTGGGCAGGCTAAAGGCTAACCTGGTGTGTGGCGTGTCTGCAGGGGAATTGAACAGGTGTAATAATGGAGGGACAAGACTCCCA
CAGATTTTCGGTTTTGTCGGGAAGTTTTTAATAGGGCAAAATAAGGAAAATGGGAGGATAGGTAGTCATCTGGGGTTTTATGCAGCAAAACT
ACAGGTTATTATGCTGTGATCCGCCTCGGAGTATTTCCATCGAGGTAGATTAAGACATGCTCACCCGAGTTTTATACTCTCTGCTTGA
GATCCTTACTACAGTATGAAATTACAGTGTGCGGAGTTAGACTATGTAAGCAGAAATTTA (서열 번호:20)
```

[0212] 유도 RNA는 생쥐 Rosa26 좌위에서 각 표적 부위를 표적으로 하도록 설계되었다. 서열은 표 9에서 도시되는데, 각각 길이에서 42개 뉴클레오티드이고, 그리고 5' 영역이 표 8에서 제공되지 않은 가닥 (즉, 표 8에서 도시된 가닥에 상보적인 가닥)에 상보적이다.

표 9

[0213] 생쥐 Rosa26 유도 RNA

RNA	5'-3' 서열	서열 번호:
mRosa26-crRNA-1	CUCCAGUCUUUCUAGAAGAUGUUUUAGAGCUAUGCUGUUUU G	21
mRosa26-crRNA-2	UGAACAGGUGUAAAAUUGGAGUUUUAGAGCUAUGCUGUUUU G	22
mRosa26-crRNA-3	UGUCGGGAAGUUUUUUAAUAGUUUUAGAGCUAUGCUGUUUU G	23

[0214] crRNA가 화학적으로 합성되고 tracrRNA에 미리 어닐링되었다 (서열 번호:13; 실시예 2 참조). 미리 어닐링된 crRNA / tracrRNA 및 변형된 Cas9 단백질을 인코딩하는 시험관내 전사된 mRNA (서열 번호: 9; 실시예 1 참조)는 수태된 생쥐 배아의 전핵 내로 미량주사될 수 있다. crRNA에 의한 표적 세트에 유도 시에, Cas9 단백질은 표적 부위를 개열하고, 그리고 결과의 이중 가닥 절단이 비상동성 말단 연결 (NHEJ) 복구 과정을 통해 복구될 수 있다. 주사된 배아는 37°C, 5% CO₂에서 하룻밤 동안 또는 최대 4 일 동안 배양되고, 그 이후에 유전자형분석될 수 있거나, 또는 주사된 배아는 생존 출산 동물이 유전자형분석될 수 있도록 수용자 암컷 생쥐 내로 이식될 수 있다. 생존 출산 동물로부터 시험관내-배양된 배아 또는 조직은 표준 방법을 이용하여, Rosa 좌위에서 Cas9-유발 돌연변이의 존재에 대해 스크리닝될 수 있다. 가령, 태아 또는 생존 출산 동물로부터 배아 또는 조직은 DNA 추출과 분석을 위해 수확될 수 있다. DNA는 표준 절차를 이용하여 단리될 수 있다. Rosa26 좌위의 표적화된 영역은 적절한 프라이머를 이용하여 PCR 증폭될 수 있다. NHEJ가 오류 가능하기 때문에, 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단의 복구 동안 일어날 수 있다. 돌연변이는 PCR-기초된 유전자형분석 방법, 예를 들면, Cel1-I 불일치 검정 및 DNA 염기서열결정을 이용하여 검출될 수 있다. **실시예 7: 생쥐 배아에서 Cas9-기초된 유전체 변형**

[0215] Rosa26 좌위는 상기 실시예 6에서 설명된 바와 같은 미리 어닐링된 crRNA / tracrRNA 및 변형된 Cas9를 인코딩하는 mRNA와 함께, 상기 섹션 (IV)(d)에서 상술된 바와 같이, 공여자 폴리뉴클레오티드를 공동 주사함으로써 생쥐 배아에서 변형될 수 있다. 생존 출산 동물로부터 시험관내-배양된 배아 또는 조직 (실시예 6에서 설명된 바와 같이)은 PCR-기초된 유전자형분석 방법, 예를 들면, RFLP 검정, 접합부 PCR, 그리고 DNA 염기서열결정을 이용하여, 변형된 Rosa26 좌위에 대해 스크리닝될 수 있다.

[0217] **실시예 8: 쥐 배아에서 Cas9-기초된 유전체 편집**

[0218] 쥐 Rosa26 좌위가 유전체 변형을 위해 표적화될 수 있다. 표 10은 잠재적 표적 부위가 굵은 글씨체로 표시되는 쥐 서열의 부분을 제공한다. 각 표적 부위는 프로토스페이서를 포함한다.

표 10

[0219] 쥐 Rosa26 서열

```
GGGATTCCTCCTTGAGTTGTGGCACTGAGGAACGTGCTGAACAAGACCTACATTGCACTCCAGGGAGTGGATGAAGGAGTTGGGGCTCAGTCG
GGTTGTATTGGAGACAAGAAGCACTTGCTCTCCAAAAGTCGGTTTTGAGTTATCATTAAGGGAGCTGCAGTGGAGTAGGGCGGAGAAAAGCCGC
ACCTTCTCAGGACGGGGAGGGAGGTGTGCAATACCTTTCTGGGAGTTCTCTGCTGCCTCTGTCTTCTGAGGACCGCCCTGGCCCTGGAA
GATTCCTTCCCCTTCTTCCCTCGTGATCTGCAACTGGAGTCTTTCTGGAAGATAGGGCGGAGTCTTCTGGGCAGGCTTAAAGGCTAACCTG
GTGCGTGGGGCGTTGCTCAGAGGAATTGAACAGGTGTAATAATTGGAGGGGCAAGACTTCCACAGATTTTCGATTGTGTGTTAAAGTATT
GTAATAGGGGCAAATAAGGAAATAGACTAGGCACCTACCTGGGGTTTTATGACAGAAAACACAGGTTATTATTGCTTGTGATCCGCCCTGG
AGAATTTTTCACCGAGGTAGATTGAAGACATGCCACCCAAATTTAATATTCTTCCACTTGCATCTTGTACAGTATGAAA (서열 번
호:24)
```

[0220] 유도 RNA는 쥐 Rosa26 좌위에서 각 표적 부위를 표적으로 하도록 설계되었다. 서열은 표 11에서 도시되는데, 각각 길이에서 42개 뉴클레오티드이고, 그리고 5' 영역이 표 10에서 제공되지 않은 가닥 (즉, 표 10에서 도시된

가닥에 상보적인 가닥)에 상보적이다.

표 11

쥐 Rosa26 유도 RNA

RNA	5'-3' 서열	서열 번호:
rRosa26-crRNA-1	AGGGGAAGGAAUCUCCAGUUUUAGAGCUAUGCUGUUU G	25
rRosa26-crRNA-2	UCUGCAACUGGAGUCUUUCUGUUUUAGAGCUAUGCUGUUU G	26
rRosa26-crRNA-3	AGGCGGGAGUCUUCUGGGCAGUUUUAGAGCUAUGCUGUUU G	27

[0221]

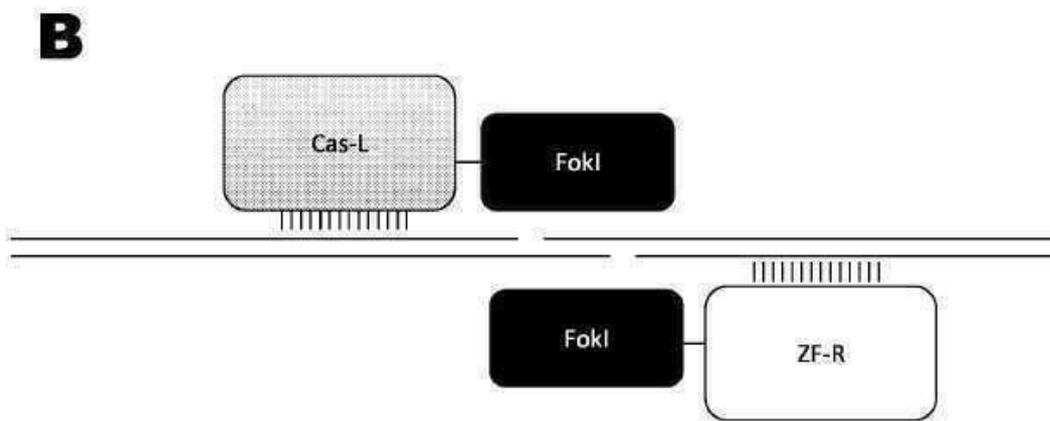
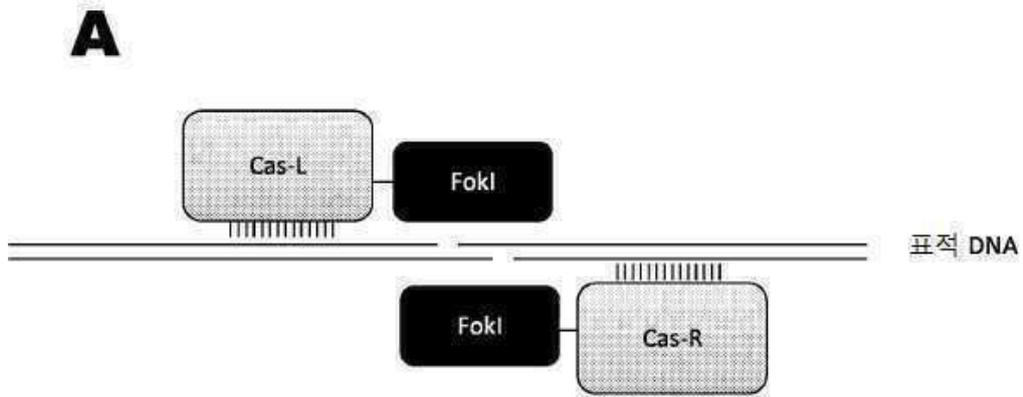
crRNA가 화학적으로 합성되고 tracrRNA에 미리 어닐링되었다 (서열 번호:13; 실시예 2 참조). 미리 어닐링된 crRNA / tracrRNA 및 변형된 Cas9 단백질을 인코딩하는 시험관내 전사된 mRNA (서열 번호. 9; 실시예 1 참조)는 수태된 쥐 배아의 전핵 내로 미량주사될 수 있다. crRNA에 의한 표적 세트에 유도 시에, Cas9 단백질은 표적 부위를 개열하고, 그리고 결과의 이중 가닥 절단이 비상동성 말단 연결 (NHEJ) 복구 과정을 통해 복구될 수 있다. 주사된 배아는 37°C, 5% CO₂에서 하룻밤 동안 또는 최대 4 일 동안 배양되고, 그 이후에 유전자형분석될 수 있거나, 또는 주사된 배아는 생존 출산 동물이 유전자형분석될 수 있도록 수용자 암컷 생쥐 내로 이식될 수 있다. 생존 출산 동물로부터 시험관내-배양된 배아 또는 조직은 표준 방법을 이용하여, Rosa 좌위에서 Cas9-유발 돌연변이의 존재에 대해 스크리닝될 수 있다. 가령, 태아 또는 생존 출산 동물로부터 배아 또는 조직은 DNA 추출과 분석을 위해 수확될 수 있다. DNA는 표준 절차를 이용하여 단리될 수 있다. Rosa26 좌위의 표적화된 영역은 적절한 프라이머를 이용하여 PCR 증폭될 수 있다. NHEJ가 오류 가능하기 때문에, 최소한 하나의 뉴클레오티드의 결실, 최소한 하나의 뉴클레오티드의 삽입, 최소한 하나의 뉴클레오티드의 치환, 또는 이들의 조합이 상기 절단의 복구 동안 일어날 수 있다. 돌연변이는 PCR-기초된 유전자형분석 방법, 예를 들면, Cel-I 불일치 검정 및 DNA 염기서열결정을 이용하여 검출될 수 있다. **실시예 9: 쥐 배아에서 Cas9-기초된 유전체 변형**

[0222]

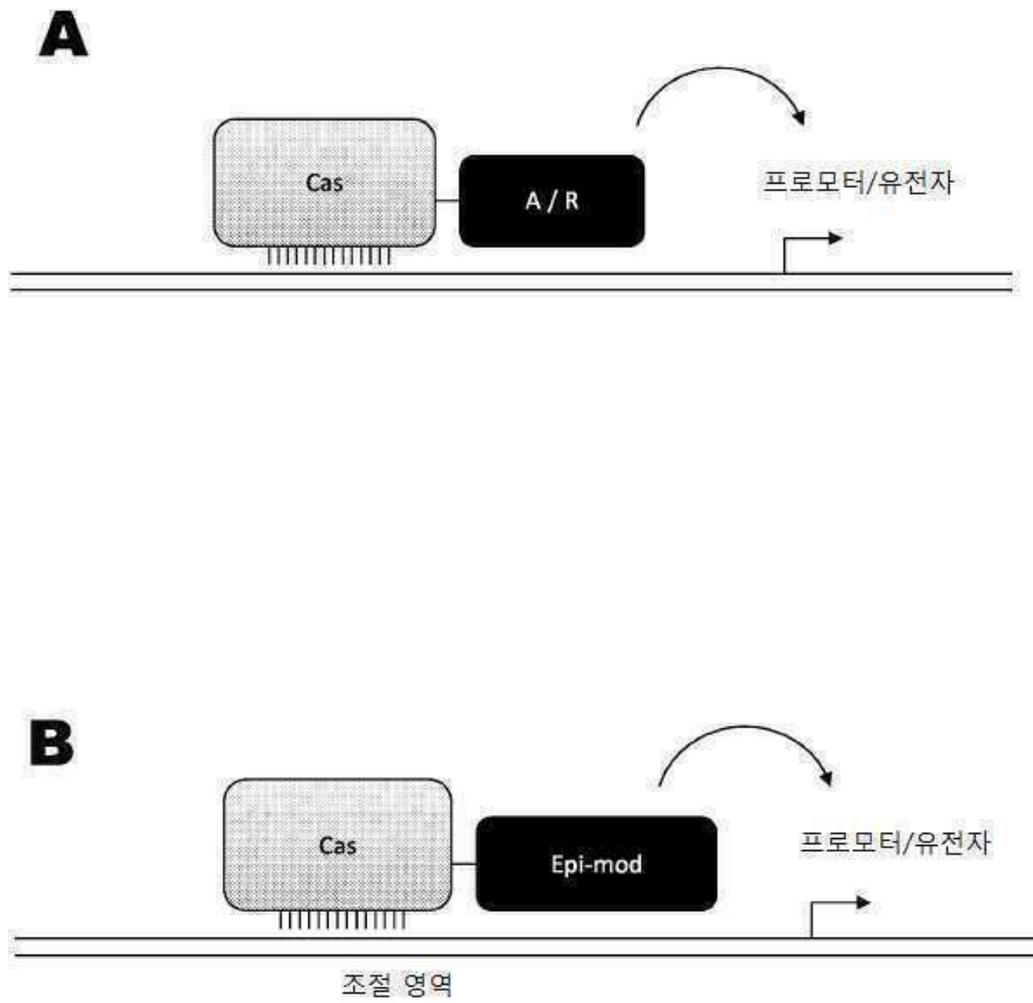
Rosa26 좌위는 상기 실시예 8에서 설명된 바와 같은 미리 어닐링된 crRNA / tracrRNA 및 변형된 Cas9를 인코딩하는 mRNA와 함께, 상기 섹션 (IV)(d)에서 상술된 바와 같이, 공여자 폴리뉴클레오티드를 공동 주사함으로써 쥐 배아에서 변형될 수 있다. 생존 출산 쥐로부터 시험관내-배양된 배아 또는 조직 (실시예 8에서 설명된 바와 같이)은 PCR-기초된 유전자형분석 방법, 예를 들면, RFLP 검정, 접합부 PCR, 그리고 DNA 염기서열결정을 이용하여, 변형된 Rosa26 좌위에 대해 스크리닝될 수 있다.

도면

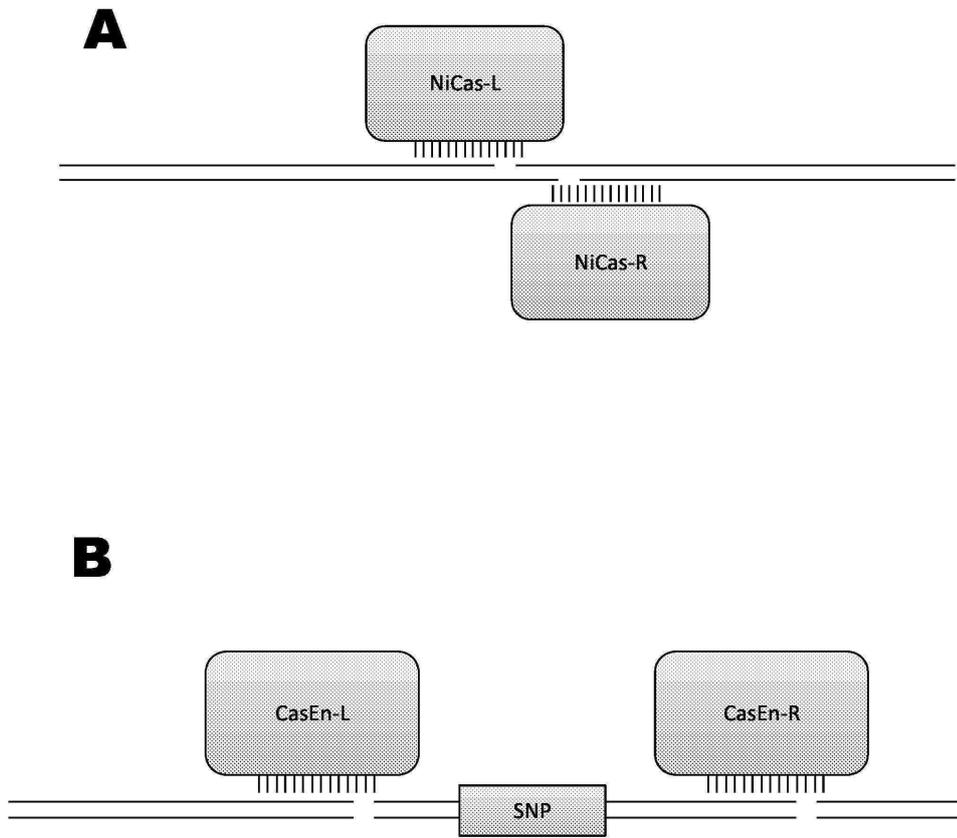
도면1



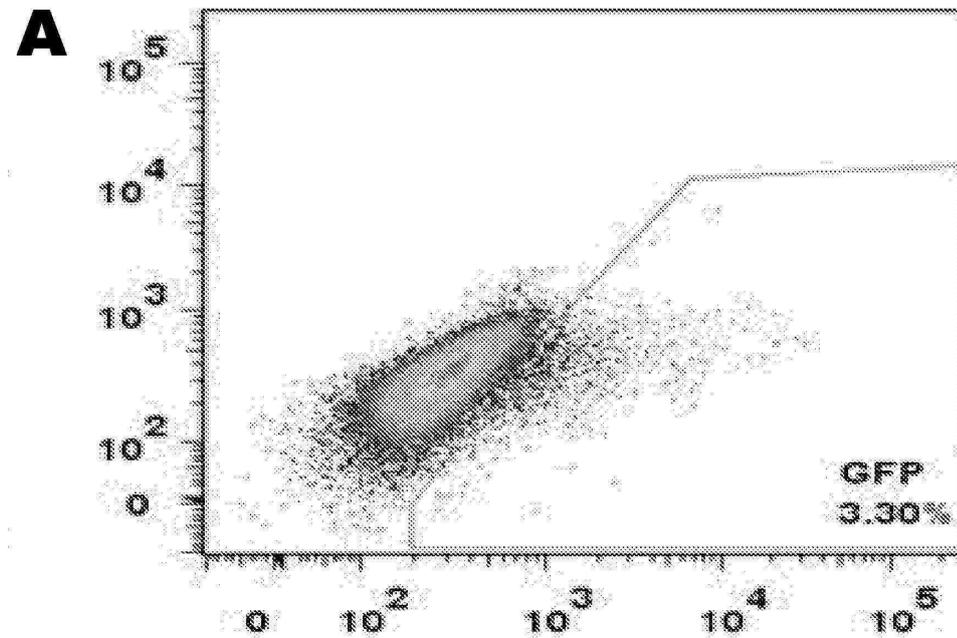
도면2



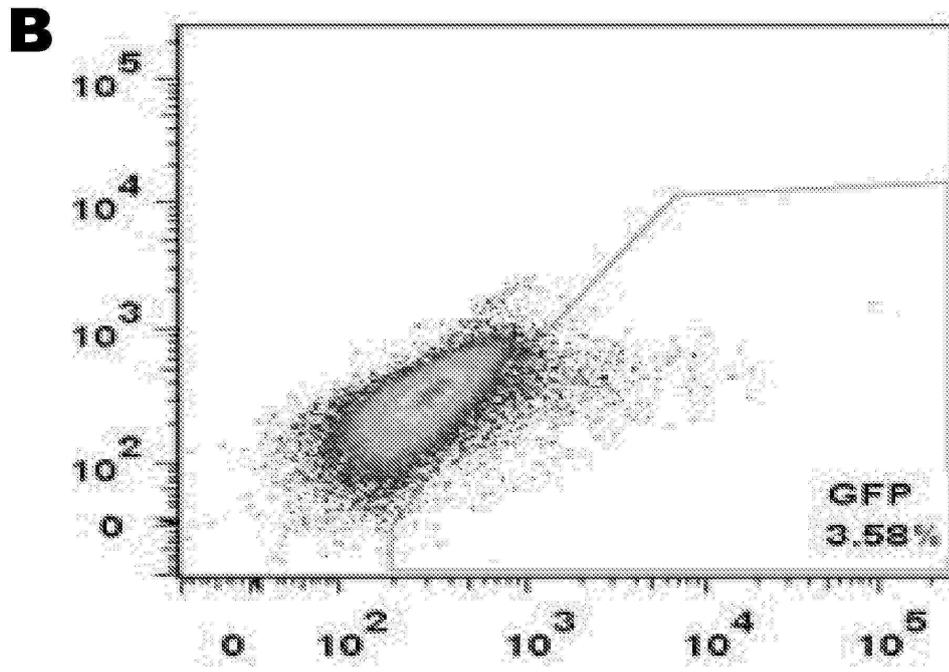
도면3



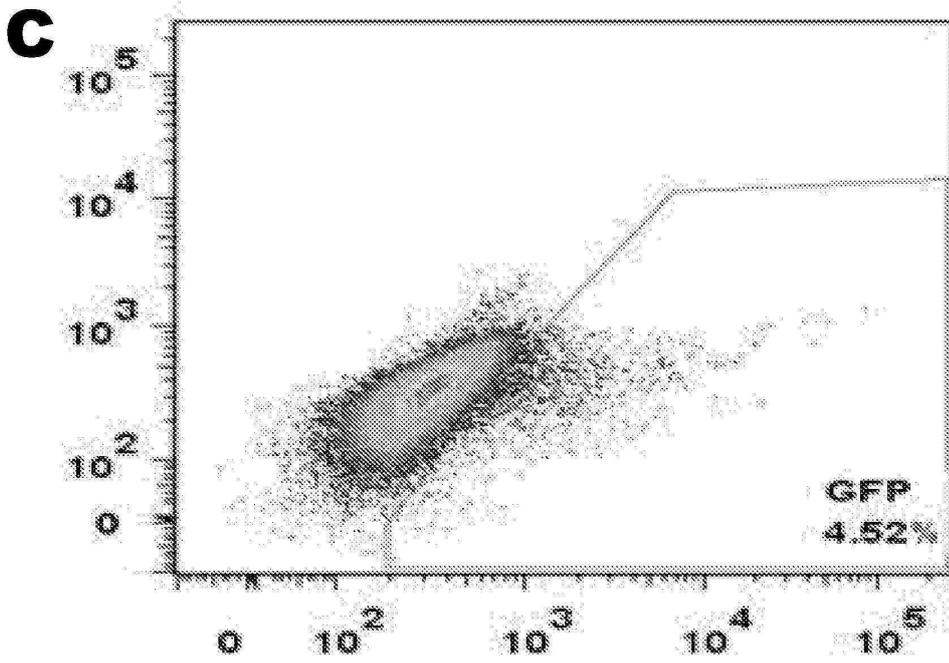
도면4a



도면4b

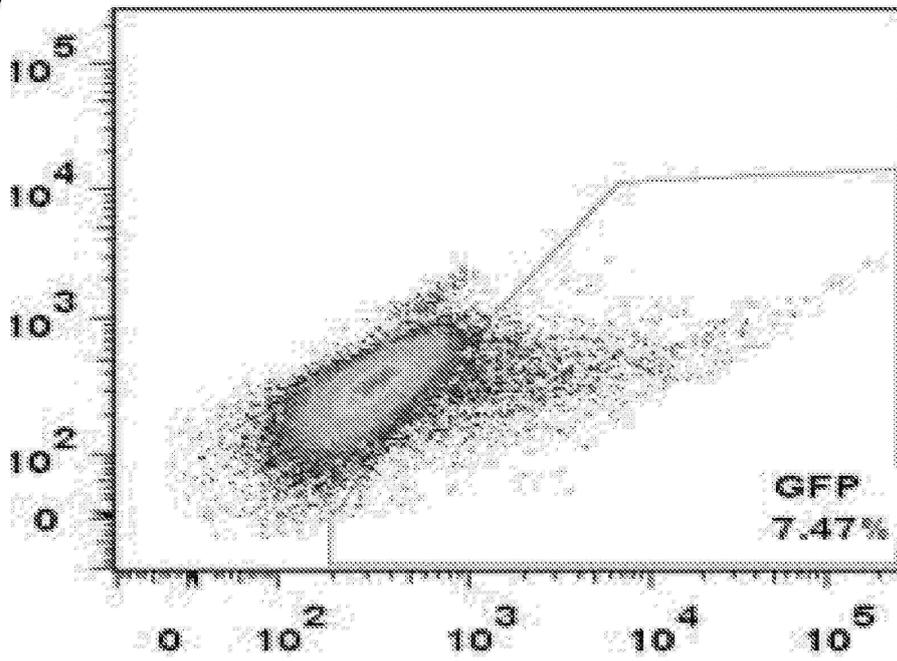


도면4c



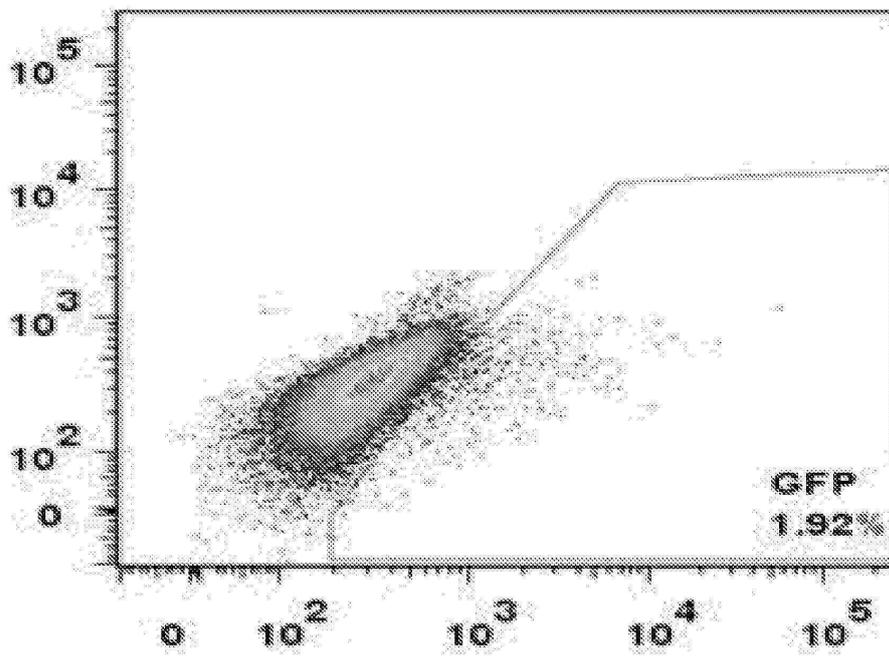
도면4d

D

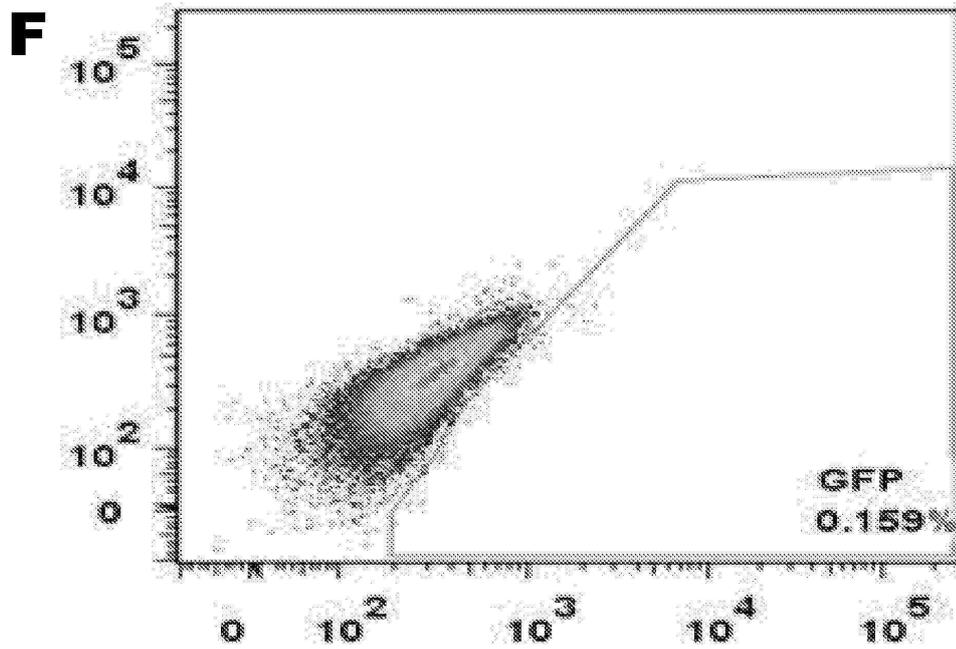


도면4e

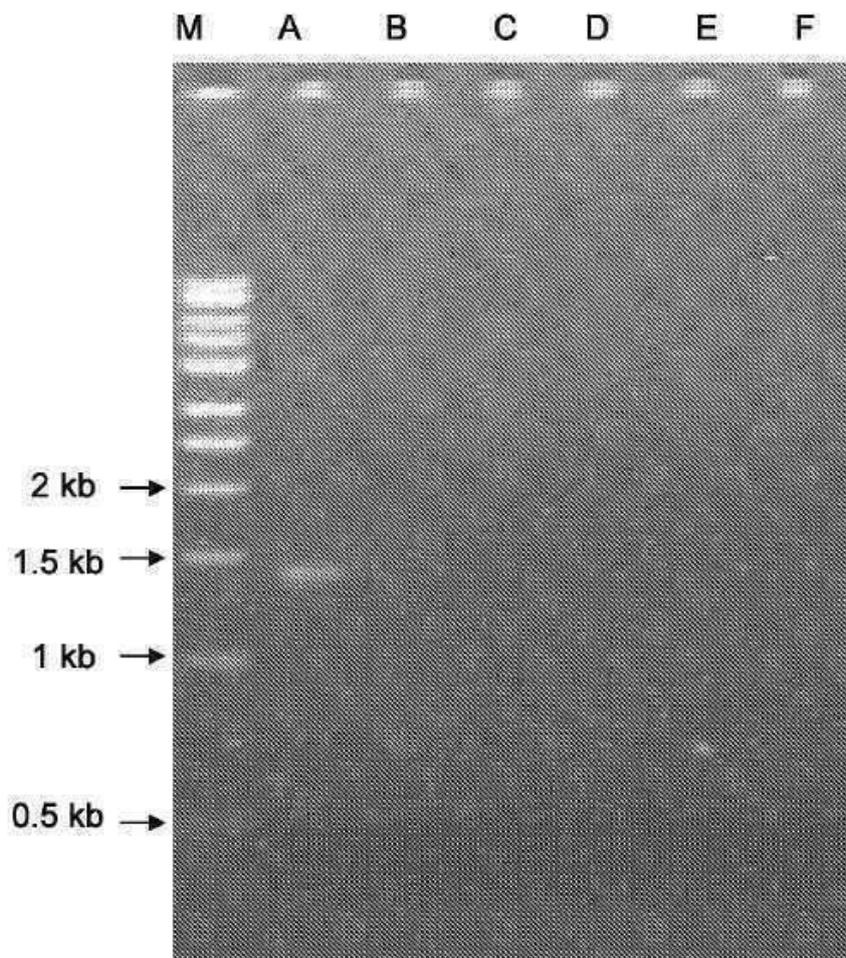
E



도면4f



도면5



서열목록

SEQUENCE LISTING

<110> SIGMA-ALDRICH CO. LLC

CHEN, Fuqiang

DAVIS, Gregory D.

KANG, Qiaohua

KNIGHT, Scott W.

<120> CRISPR-BASED GENOME MODIFICATION AND REGULATION

<130> 047497-465606

<150> US 61/734,256

<151> 2012-12-06

<150> US 61/758,624

<151> 2013-01-30

<150> US 61/761,046

<151> 2013-02-05

<150> US 61/794,422

<151> 2013-03-15

<160> 27

<170> PatentIn version 3.5

<210> 1

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 1

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 2

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 2

Pro Lys Lys Lys Arg Arg Val

1 5

<210> 3

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 3

Lys Arg Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys

1 5 10 15

<210> 4

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 4

Gly Arg Lys Lys Arg Arg Gln Arg Arg Arg Pro Pro Gln Pro Lys Lys

1 5 10 15

Lys Arg Lys Val

20

<210> 5

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 5

Pro Leu Ser Ser Ile Phe Ser Arg Ile Gly Asp Pro Pro Lys Lys Lys

1 5 10 15

Arg Lys Val

<210> 6

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 6

Gly Ala Leu Phe Leu Gly Trp Leu Gly Ala Ala Gly Ser Thr Met Gly

1 5 10 15

Ala Pro Lys Lys Lys Arg Lys Val

20

<210> 7

<211> 27

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 7

Gly Ala Leu Phe Leu Gly Phe Leu Gly Ala Ala Gly Ser Thr Met Gly

1 5 10 15

Ala Trp Ser Gln Pro Lys Lys Lys Arg Lys Val

20

25

<210> 8

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 8

Lys Glu Thr Trp Trp Glu Thr Trp Trp Thr Glu Trp Ser Gln Pro Lys

1 5 10 15

Lys Lys Arg Lys Val

20

<210> 9

<211> 1374

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 9

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

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Asp Tyr Lys Val Pro Ser Lys Lys Phe
 20 25 30
 Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile
 35 40 45
 Gly Ala Leu Leu Phe Gly Ser Gly Glu Thr Ala Glu Ala Thr Arg Leu
 50 55 60
 Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

 65 70 75 80
 Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
 85 90 95
 Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
 100 105 110
 His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
 115 120 125
 His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Ala Asp

 130 135 140
 Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
 145 150 155 160
 Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
 165 170 175
 Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Ile Tyr
 180 185 190
 Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Arg Val Asp Ala

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

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525
 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575
 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590
 Ala Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605
 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Gly Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640
 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 645 650 655
 Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670
 Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
 675 680 685
 Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
 690 695 700
 Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly His Ser Leu
 705 710 715 720
 His Glu Gln Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735
 Ile Leu Gln Thr Val Lys Ile Val Asp Glu Leu Val Lys Val Met Gly
 740 745 750
 His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln Thr

Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala Lys
 1010 1015 1020
 Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe Tyr
 1025 1030 1035
 Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala Asn
 1040 1045 1050
 Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu Thr
 1055 1060 1065
 Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val Arg
 1070 1075 1080
 Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr Glu
 1085 1090 1095
 Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys Arg
 1100 1105 1110
 Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro Lys
 1115 1120 1125
 Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val Leu
 1130 1135 1140
 Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys Ser
 1145 1150 1155
 Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser Phe
 1160 1165 1170
 Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys Glu
 1175 1180 1185
 Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu Phe
 1190 1195 1200
 Glu Leu Glu Asn Gly Arg Lys Arg Met Leu Ala Ser Ala Gly Glu
 1205 1210 1215
 Leu Gln Lys Gly Asn Glu Leu Ala Leu Pro Ser Lys Tyr Val Asn
 1220 1225 1230
 Phe Leu Tyr Leu Ala Ser His Tyr Glu Lys Leu Lys Gly Ser Pro

1235 1240 1245
 Glu Asp Asn Glu Gln Lys Gln Leu Phe Val Glu Gln His Lys His
 1250 1255 1260
 Tyr Leu Asp Glu Ile Ile Glu Gln Ile Ser Glu Phe Ser Lys Arg

 1265 1270 1275
 Val Ile Leu Ala Asp Ala Asn Leu Asp Lys Val Leu Ser Ala Tyr
 1280 1285 1290
 Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn Ile
 1295 1300 1305
 Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala Phe
 1310 1315 1320
 Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser Thr

 1325 1330 1335
 Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr Gly
 1340 1345 1350
 Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp Pro
 1355 1360 1365
 Lys Lys Lys Arg Lys Val
 1370
 <210> 10
 <211> 4122
 <212> DNA
 <213> Artificial Sequence
 <220><223> SYNTHESIZED
 <400> 10
 atggacaaga agtacagcat cggcctggac atcggcacca actctgtggg ctgggccgtg 60

 atcaccgacg actacaaggt gccagcaag aaattcaagg tgctgggcaa caccgaccgg 120
 cacagcatca agaagaacct gatcggcgcc ctgctgttcg gctctggcga aacagccgag 180
 gccaccggcg tgaagagaac cgccagaaga agatacacca gacggaagaa cgggatctgc 240
 tatctgcaag agatcttcag caacagatg gccaaggtgg acgacagctt cttccacaga 300
 ctggaagagt ccttcctggt ggaagaggat aagaagcacg agcggcaccc catcttcggc 360
 aacatcgtgg acgaggtggc ctaccacgag aagtaccca ccatctacca cctgagaaag 420

aagctggccg acagcaccca caaggccgac ctgagactga tctacctggc cctggcccac 480

atgatcaagt tccggggcca cttcctgatc gagggcgacc tgaaccccga caacagcgac 540
gtggacaage tgttcatcca gctggtgcag atctacaatc agctgttcga ggaaaacccc 600
atcaacgcca gcagagtgga cgccaaggcc atcctgagcg ccagactgag caagagcaga 660
cggctggaaa atctgatcg ccagctgccc ggcgagaagc ggaatggcct gttcggcaac 720
ctgattgccc tgagcctggg cctgaccccc aacttcaaga gcaacttcga cctggccgag 780
gatgccaaac tgcagctgag caaggacacc tacgacgacg acctggaaa cctgctggcc 840
cagatcggcg accagtacgc cgacctgttt ctggccgcca agaacctgtc cgacgccatc 900

ctgctgagcg acatcctgag agtgaacagc gagatcacca aggccccct gtccgcctct 960
atgatcaaga gatagcagca gcaccaccag gacctgacct tgctgaaagc tctcgtcggg 1020
cagcagctgc ctgagaagta caaagagatt ttcttcgacc agagcaagaa cggctacgcc 1080
ggctacatcg atggcggagc cagccaggaa gatttctaca agttcatcaa gccatcctg 1140
gaaaagatgg acggcaccga ggaactgtc gtgaagctga acagagagga cctgctcggg 1200
aagcagcgga ctttcgaaa cggcagcatc ccccaccaga tccacctggg agagctgac 1260
gccattctgc ggcggcagga agatttttac ccattcctga aggacaaccg ggaaaagatc 1320

gagaagatcc tgacctcag aatcccctac tacgtgggccc ctctggccag gggaaacagc 1380
agattcgctt ggatgaccag aaagagcgag gaaacatca ccccctggaa cttcaggaa 1440
tgggtggaca agggcgccc cgcccagagc ttcacgagc ggatgacaa cttcgataag 1500
aacctgccc aagagaaggt gctgcccag cacagcctgc tgtacgagta cttcacctg 1560
tacaacgagc tgaccaaagt gaaatcgtg accgagggaa tgcggaagcc cgccttctg 1620
agcggcgagc agaaaaagc catcgtggac ctgctgttca agaccaaccg gaaagtgacc 1680
gtgaagcagc tgaagagga ctacttcaag aaaatcgagt gcttcgacag cgtggaatc 1740

agcggcgtgg aagatcggtt caacgcctcc ctgggcgctt atcacgatct gctgaaaatt 1800
atcaaggaca aggacttctt ggacaatgag gaaaacgagg acattctgga agatatctg 1860
ctgacctga cactgtttga ggaccggggc atgatcgagg aacggctgaa aacctatgcc 1920
cacctgttcg acgacaaagt gatgaagcag ctgaagcggc ggagatacac cggctggggc 1980
aggctgagcc ggaagctgat caacggcatc cgggacaagc agtccggcaa gacaatcctg 2040
gatttctga agtccgacgg cttcggcaac agaaacttca tgcagctgat ccacgacgac 2100
agcctgacct ttaaagagga catccagaaa gccaggtgt cggccaggg acactctctg 2160

cacgagcaga tcgccaatct ggccggatcc cccgccatta agaagggcat cctgcagaca 2220
 gtgaagattg tggacgagct cgtgaaagt atgggccaca agcccagaaa catcgtgatc 2280
 gaaatggcca gagagaacca gaccaccag aaggacaga agaacagccg cgagagaatg 2340
 aagcggatcg aagagggcat caaagagctg ggcagccaga tcctgaaaga acaccccgctg 2400
 gaaaacaccc agctgcagaa cgagaagctg tacctgtact acctgcagaa tgggcgggat 2460
 atgtacgtgg accaggaact ggacatcaac cggctgtccg actacgatgt ggaccacatt 2520
 gtgccccagt ccttcatcaa ggacgactcc atcgataaca aagtgtgac tcggagcgac 2580

 aagaaccggg gcaagagcga caactgccc tccgaagagg tcgtgaagaa gatgaagaac 2640
 tactggcgc agctgctgaa tgccaagctg attaccaga ggaagtcca caatctgacc 2700
 aaggccgaga gaggcggcct gagcgaactg gataaggccg gcttcattaa gcggcagctg 2760
 gtggaaccc ggagatcac aaagcagctg gcacagatcc tggactccg gatgaacact 2820
 aagtacgacg agaacgaaa actgatccgg gaagtgaag tgatcacct gaagtccaag 2880
 ctggtgtccg acttcagaaa ggatttccag ttttcaaag tgcgcgagat caacaactac 2940
 caccacgcc acgagccta cctgaacgcc gtctgggaa ccgcctgat caaaaagtac 3000

 cctaagctgg aaagcgagtt cgtgtacggc gattacaagg tgtacgacgt gcggaagatg 3060
 atcgccaaga gcgagcagga aatcggaag gctaccgcca agtacttctt ctacagcaac 3120
 atcatgaact tttcaagc cgagatcaca ctggccaacg gcgagatcag aaagcggcct 3180
 ctgatcgaga caaacggcga aaccggggag atcgtgtggg ataaggccg ggattttgcc 3240
 acagtgcgga aagtgtgtc catgcccaca gtgaatatcg tgaaaaagac cgaggtgcag 3300
 accggcggct tcagcaaa gtctatcctg cccaagagga actccgaaa gctgatcgcc 3360
 agaaagaagg attgggacc taagaagtac ggcggctttg acagcccac cgtggcctac 3420

 tctgtctgg tgggtggcaa agtgaaaaag ggcaagtcca agaaactgaa gagtgtgaaa 3480
 gagctgctgg ggatcacat catgaaaga agcagcttcg agaagaatcc catcgacttt 3540
 ctggaagcca agggctaaa agaagtgaaa aaggacctga tcatcaagct gcctaagtac 3600
 tcctgttcg agctgaaaa cggccggaag cggatgctgg cttctgccg gaaactgcag 3660
 aaggaaacg agctggcct gccctcaaa tatgtgaact tcctgtacct ggccagccac 3720
 tatgagaagc tgaaggctc ccccgaggat aatgagcaga aacagctgtt tgtggaacag 3780
 cacaagcact acctggacga gatcatcgag cagattagcg agttctcaa gcgcgtgatc 3840

 ctggccgatg ccaacctgga caagtgctg agcgcctaca acaagcaccg ggataagccc 3900
 atcagagagc aggcgagaaa tatcatccac ctgtttacc tgaccaacct gggagcccct 3960
 gccgcctca agtacttga caccacatc gaccggaaga ggtacaccag caccaaagag 4020

gtgctggacg ccacctgat ccaccagagc atcaccggcc tgtacgagac acggatcgac 4080
ctgtctcagc tgggaggcga cccaagaaa aagcgcaaag tg 4122
<210> 11
<211> 4764
<212> DNA
<213> Homo sapiens
<400> 11
gcgggcgggc ggtgcgatgt ccggagagga tggcccggcg gctggcccgg gggcggcggc 60

ggcggctgcc cgggagcggc gacgggagca gctgcggcag tggggggcgc gggcggcgc 120
cgagcctggc cccggagagc gccgcgccg caccgtccgc ttcgagcgcg ccgccagatt 180
cctggcggcc tgtgcgggcg gcgacctgga cgaggcgcgt ctgatgctgc gcgccccga 240
ccctggcccc ggcgccgagc tcgaccccc gcgcgccccg cccgcccgcg ccgtgctgga 300
ctccaccaac gccgacggta tcagcgcctt gcaccaggtc agcgccccc gcccggcgtc 360
tccccgggccc aggtccacc tctgctgcgc cacctggggc atcctcttc cccgttgcca 420
gtctcgatcc gccccgtct tcttggccct gggtttgccc accctatgct gacacccct 480

cccagtcctt cttaccattc ccttcgacc acccacttc cgaattggag ccgcttcaac 540
tggccctggg cttagecact ctgtgctgac cactctgccc caggcctctt taccattccc 600
cttcgaccta ctctcttccg cattggagtc gctttaactg gccctggctt tggcagcctg 660
tgctgaccca tgcagtcctc cttaccatcc ctccctcgac tccccctctt ccgatgttga 720
gccccctcag ccggtcctgg actttgtctc ctccctgcc ctgcctctc ctgaacctga 780
gccagctccc atagctcagt ctggtctatc tgcttggccc tggccattgt cactttgcgc 840
tgccctctc tcgccccga gtgcccttgc tgtgccgccg gaactctgcc ctctaacct 900

gccgtctctc tctgagtc ccgaccattt gagctctact ggcttctgcg ccgctctgg 960
cccactgttt ccccttccca ggcaggctct gctttctctg acctgcattc tctcccctgg 1020
gcctgtgccg ctttctgtct gcagcttggt gcctgggtca cctctacggc tggcccagat 1080
ccttccctgc cgctccttc aggttccgtc ttctccact cctcttccc cttgctctct 1140
gctgtgttgc tgeccaagga tgccttttcc ggagcacttc cttctcggcg ctgcaccaag 1200
tgatgtcctc tgagcggate ctccccgtgt ctgggtctc tccgggcatc tctctccct 1260
cacccaacce catgccgtct tcaactcgtg ggttccctt tcttctct tctggggcct 1320

gtgccatctc tegtittctta ggatggcctt ctccgacgga tgtctccctt gcgtcccgcc 1380
tccccctctt gtaggcctgc atcatcaccg ttttctgga caacccaaa gtaccccgctc 1440

tccttgctt tagccacctc tccatcctct tgccttcttt gcctggacac cccgttctcc 1500
 tgtggattcg ggtcacctct cactcctttc atttgggcag ctcccctacc cccttacct 1560
 ctctagtctg tgetagctct tccagcccc tgcctatggca tcttccaggg gtccgagagc 1620
 tcagctagtc ttcttctcc aaccggggcc cctatgtcca cttcaggaca gcatgtttgc 1680
 tgcctccagg gatcctgtgt ccccgagctg ggaccacctt atattccag gcccgttaa 1740

 tgtggctctg gtctgggta cttttatctg tcccctccac cccacagtgg ggccactagg 1800
 gacaggattg gtgacagaaa agccccatcc ttaggcctcc tccttctag tctcctgata 1860
 ttgggtctaa cccccacctc ctgttaggca gattccttat ctggtgacac accccattt 1920
 cctggagcca tcctctctct tgccagaacc tctaaggttt gcttacgatg gagccagaga 1980
 ggatcctggg agggagagct tggcaggggg tgggaggaa ggggggatg cgtgacctgc 2040
 ccggttctca gtggccacc tgcctaccc tctcccagaa cctgagctgc tctgacgagg 2100
 ccgtctgggt cgtttactg atcctgggtgc tgcagcttcc ttacacttcc caagaggaga 2160

 agcagtttg aaaaacaaaa tcagaataag ttggtcctga gttctaactt tggctcttca 2220
 cctttctagt cccaattta tattgttctt ccgtgcgtca gttttactg tgagataagg 2280
 ccagtagcca gccccgtctt ggcagggctg ttggtgaggag gggggtgtcc gtgtgaaaa 2340
 ctccctttgt gagaatgggt cgtcttaggt gttcaccagg tctggccgc ctctactccc 2400
 tttctcttcc tccatcttc tttcttaaa gaggccccag tgcctatctg gacatatcc 2460
 tccgcccaga gcagggtccc gcttcctaa ggccctgctc tgggcttctg ggtttgagtc 2520
 cttggcaagc ccaggagagg cgctcaggct tccctgtccc ccttctctgt ccaccatctc 2580

 atgccccctg ctctctgcc ctttcctac aggggttctt ggctctgctc ttcagactga 2640
 gccccgttcc cctgcatccc cgttccccctg catccccctt cccctgcatc ccccagagc 2700
 cccaggccac ctacttgccc tggacccac gagagccac cccagccctg tetaccagc 2760
 tgccttttgg gtggattctc ctccaactgt ggggtgactg cttggcaaac tactcttctg 2820
 gggatccca ggaggcctgg agcattgggg tgggtgggg ttccagagagg agggattccc 2880
 ttctcaggtt acgtggccaa gaagcagggg agctgggttt gggctcagtc tgggtgtggg 2940
 gtgaccagct taigtgttt gccaggaca gcctagtttt agcactgaaa cctcagtc 3000

 taggaaaaca gggatggttg gtcactgtct ctgggtgact cttgattccc ggccagtttc 3060
 tccacctggg gctgtgttcc tegtctgca tccttctcca ggcaggctcc caagcatcgc 3120
 cccccgtctg tggtgttcc caagtctta gggtagccca cgtgggttta tcaaccactt 3180
 ggtgaggctg gtacctgcc cccattctct caccacaatt gccttagtgg ctagggggtt 3240
 gggggctaga gtaggagggg ctggagccag gattcttagg gctgaacaga gaagagctgg 3300

gggcctgggc tcctgggttt gagagaggag gggctggggc ctggactcct gggccgagg 3360
 gaggaggggc tggggcctgg actcctgggt ctgagggtgg agggactggg gcctggact 3420

 cctgggtccg agggaggagg ggctggggcc tggactcgtg ggtctgaggg aggaggggct 3480
 gggggcctgg acttctgggt cttagggagg cggggctggg cctggacccc tgggtctgaa 3540
 tggggagagg ctgggggctt ggactccttc atctgagggc ggaagggctg gggcctggcc 3600
 tcctgggttg aatggggagg ggttgggctt ggactctgga gtccttggg cccaggcctc 3660
 aggcattttt cacagggatg cctgtactgg gcaggtcctt gaaagggaaa ggccatttgc 3720
 tctccttgc cccctccct atcgccatga caactgggtg gaaataaacg agccagttc 3780
 atcccgttcc cagggcactg gcggcccctt cacagcccga gtttccatga cctcatgctc 3840

 ttggccctcg tagctccctc ccgctcctc cagatgggca gctttggaga ggtgagggac 3900
 ttgggggta atttatccc tggatctagg agtttagctt cactccttcc tcagctccag 3960
 ttcaggtccc ggagcccacc cagtgtccac aaggcctggg gcaagtcctt cctccgacct 4020
 cctggacttc ggcttttgc ccccaagt tggaccctt aaggaagaa tgagaaacgg 4080
 tggcccgtgt cagcccctgg ctgcagggcc ccgtgcagag ggggcctcag tgaactggag 4140
 tgtgacagcc tggggcccag gcacacaggt gtgcagctgt ctcccctc tgggagtccc 4200
 gcccaggccc ctgagtctgt cccagcacag ggtggcctt ctcaccctg catagcctg 4260

 ggcccacggc ttcgttctg cagagtatct gctggggtgg tttccgagct tgacccttgg 4320
 aaggacctgg ctgggttaa ggcaggagg gctgggggcc aggactcctg gctctgaagg 4380
 aggaggggct ggaacctctt ccctagtctg agcactggaa gcgccacctg tgggtgtga 4440
 cgggggtttt gccgtgtcta acaggtacca tgtggggttc ccgcaccag atgagaagcc 4500
 ccctccctc cccgttact tctgtttgc agatagccag gactccttc gtggtttcca 4560
 ctgagcactg aaggcctggc cggcctgacc actgggcaac caggcgtatc ttaaacagcc 4620
 agtggccaga ggctgttggg tcattttccc cactgtccta gcaccgtgtc cctggatctg 4680

 ttttcgtggc tcctctgga gtcgccactt gctgggacac cgtggctggg gtaggtgagg 4740
 ctgacggctg tttcccacc ccag 4764

 <210> 12
 <211> 42
 <212> RNA
 <213> Artificial Sequence
 <220><223> SYNTHESIZED
 <400> 12

acccacagu ggggccacua guuuuagagc uaugcuguuu ug	42
<210> 13	
<211> 86	
<212> RNA	
<213> Artificial Sequence	
<220><223> SYNTHESIZED	
<400> 13	
ggaaccauuc aaaacagcau agcaaguuaa aaauaggcua guccguuauc aacuugaaaa	60
aguggcaccg agucggugcu uuuuuu	86
<210> 14	
<211> 62	
<212> RNA	
<213> Artificial Sequence	
<220><223> SYNTHESIZED	
<400> 14	
acccacagu ggggccacua guuuuagagc uagaaauagc aaguuaaaau aaggcuaguc	60
cg	62
<210> 15	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220><223> SYNTHESIZED	
<400> 15	
ctgaccttt ctcttctcc cacag	25
<210> 16	
<211> 1009	
<212> DNA	
<213> Artificial Sequence	
<220><223> SYNTHESIZED	
<400> 16	
gccacatgg actacaaaga cgatgacgac aaggtcgact ctagagctgc agagagcgac	60
gagagcggcc tgcccgccat ggagatcgag tgccgcatca ccggcacct gaacggcgtg	120
gagttcgagc tgggtggcgg cggagagggc acccccgagc agggccgcat gaccaacaag	180

atgaagacga ccaaaggcgc cctgacctic agcccctacc tgctgagcca cgtgatgggc 240
 tacggcttct accacttcgg cacctacccc agcggctacg agaaccctt cctgcacgcc 300
 atcaacaacg gcggctacac caacaccgcg atcgagaagt acgaggacgg cggcgtgctg 360

cacgtgagct tcagctaccg ctacgaggcc ggccgcgtga tcggcgactt caaggatgat 420
 ggcaccggct tccccgagga cagcgtgatc ttcaccgaca agatcgtccg cagcaacgcc 480
 accgtggage acctgcaccc catggcgcat aacgatctgg atggcagctt caccegcacc 540
 ttcagcctgc gcgacggcgg ctactacagc tccgtggtgg acagccacat gcacttcaag 600
 agcgcacatc accccagcat cctgcagaac gggggcccca tgttcgctt ccgccgcgtg 660
 gagggagatc acagcaacac cgagctgggc atcgtggagt accagcacgc cttcaagacc 720
 ccgatgcag atgccggtga agaataaga tctctgtgcc ttctagtgc cagccatctg 780

ttgtttgccc ctccccctg ccttcttga ccttggagg tgccactccc actgtccttt 840
 cctaataaaa tgaggaaatt gcatcgcat gtctgagtag gtgtcattct attctggggg 900
 gtgggtggg gcaggacagc aagggggagg attgggaaga caatagcagg catgctgggg 960
 atgcggtggg ctctatggac tcgaggttta aacgtcgacg cggccgcgt 1009

<210> 17

<211> 355

<212> PRT

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 17

Met Ser Gly Glu Asp Gly Pro Ala Ala Gly Pro Gly Ala Ala Ala Ala

1 5 10 15

Ala Ala Arg Glu Arg Arg Arg Glu Gln Leu Arg Gln Trp Gly Ala Arg

20 25 30

Ala Gly Ala Glu Pro Gly Pro Gly Glu Arg Arg Ala Arg Thr Val Arg

35 40 45

Phe Glu Arg Ala Ala Glu Phe Leu Ala Ala Cys Ala Gly Gly Asp Leu

50 55 60

Asp Glu Ala Arg Leu Met Leu Arg Ala Ala Asp Pro Gly Pro Gly Ala

65 70 75 80

Glu Leu Asp Pro Ala Ala Pro Pro Pro Ala Arg Ala Val Leu Asp Ser

	85	90	95
Thr Asn Ala Asp Gly Ile Ser Ala Leu His Gln Ala Thr Met Asp Tyr			
	100	105	110
Lys Asp Asp Asp Asp Lys Val Asp Ser Arg Ala Ala Glu Ser Asp Glu			
	115	120	125
Ser Gly Leu Pro Ala Met Glu Ile Glu Cys Arg Ile Thr Gly Thr Leu			
	130	135	140
Asn Gly Val Glu Phe Glu Leu Val Gly Gly Gly Glu Gly Thr Pro Glu			
145	150	155	160
Gln Gly Arg Met Thr Asn Lys Met Lys Ser Thr Lys Gly Ala Leu Thr			
	165	170	175
Phe Ser Pro Tyr Leu Leu Ser His Val Met Gly Tyr Gly Phe Tyr His			
	180	185	190
Phe Gly Thr Tyr Pro Ser Gly Tyr Glu Asn Pro Phe Leu His Ala Ile			
	195	200	205
Asn Asn Gly Gly Tyr Thr Asn Thr Arg Ile Glu Lys Tyr Glu Asp Gly			
	210	215	220
Gly Val Leu His Val Ser Phe Ser Tyr Arg Tyr Glu Ala Gly Arg Val			
225	230	235	240
Ile Gly Asp Phe Lys Val Met Gly Thr Gly Phe Pro Glu Asp Ser Val			
	245	250	255
Ile Phe Thr Asp Lys Ile Val Arg Ser Asn Ala Thr Val Glu His Leu			
	260	265	270
His Pro Met Gly Asp Asn Asp Leu Asp Gly Ser Phe Thr Arg Thr Phe			
	275	280	285
Ser Leu Arg Asp Gly Gly Tyr Tyr Ser Ser Val Val Asp Ser His Met			
	290	295	300
His Phe Lys Ser Ala Ile His Pro Ser Ile Leu Gln Asn Gly Gly Pro			
305	310	315	320
Met Phe Ala Phe Arg Arg Val Glu Glu Asp His Ser Asn Thr Glu Leu			
	325	330	335

Gly Ile Val Glu Tyr Gln His Ala Phe Lys Thr Pro Asp Ala Asp Ala

340 345 350

Gly Glu Glu

355

<210> 18

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 18

ccactctgtg ctgaccactc t 21

<210> 19

<211> 17

<212> DNA

<213> Artificial Sequence

<220><223> SYNTHESIZED

<400> 19

gcggcactcg atctcca 17

<210> 20

<211> 711

<212> DNA

<213> Mus musculus

<400> 20

gagcggctgc gggcggggtg caagcacgtt tccgacttga gttgcctcaa gaggggcgtg 60

ctgagccaga cctccatcgc gcactccggg gagtggaggg aaggagcgag ggctcagttg 120

ggctgttttg gaggcaggaa gcacttgctc tcccaaagtc gctctgagtt gttatcagta 180

agggagctgc agtggagtag gcggggagaa ggccgcaccc ttctccggag gggggagggg 240

agtgttgcaa tacctttctg ggagttctct gctgcctcct ggcttctgag gaccgcctg 300

ggcctgggag aatcccttcc ccctcttccc tcgtgatctg caactccagt ctttctagaa 360

gatgggcggg agtcttctgg gcaggcttaa aggctaacct ggtgtgtggg cgttgtcctg 420

caggggaatt gaacaggtgt aaaattggag ggacaagact tcccacagat tttcggtttt 480

gtcgggaagt tttttaatag gggcaaataa ggaaaatggg aggatagta gtcactctggg 540

gttttatgca gcaaaactac aggttattat tgcttggat cgcctcggga gtattttcca 600

tcgaggtaga ttaaagacat gctcacccga gttttatact ctcttgcttg agatccttac 660
 tacagtatga aattacagtg tcgcgagtta gactatgtaa gcagaatttt a 711
 <210> 21
 <211> 42
 <212> RNA
 <213> Artificial Sequence

<220><223> SYNTHESIZED
 <400> 21
 cuccagucuu ucuagaagau guuuuagagc uaugcuguuu ug 42
 <210> 22
 <211> 42
 <212> RNA
 <213> Artificial Sequence

<220><223> SYNTHESIZED
 <400> 22
 ugaacaggug uaaaauugga guuuuagagc uaugcuguuu ug 42
 <210> 23
 <211> 42
 <212> RNA
 <213> Artificial Sequence

<220><223> SYNTHESIZED
 <400> 23
 ugucgggaag uuuuuuaaau guuuuagagc uaugcuguuu ug 42
 <210> 24
 <211> 642
 <212> DNA

<213> Rattus rattus
 <400> 24
 gggattcctc cttagattgt ggcactgagg aacgtgctga acaagaccta cattgcactc 60
 cagggagtgg atgaaggagt tgggctcag tcgggttga ttggagaca gaagcacttg 120
 ctctcaaaa gtcggttga gttatcatta agggagetgc agtggagtag gcggagaaaa 180
 ggccgcacc ttctcaggac gggggagggg agtgttgcaa tacctttctg ggagtctct 240
 gctgcctcct gtcttctgag gaccgccctg ggcctggaag attcccttcc cccttcttcc 300

ctcgtgatct gcaactggag tctttctgga agataggcgg gagtcttctg ggcaggctta 360
aaggctaacc tggtagcgtgg ggcgttgtcc tgcagaggaa ttgaacaggt gtaaaattgg 420

aggggcaaga cttcccacag attttcgatt gtgttgtaa gtattgtaat aggggcaaat 480
aagggaata gactaggcac tcacctgggg ttttatgcag caaaactaca gttattatt 540
gcttgtgac cgcctggag aattttcac cgaggtagat tgaagacatg cccacccaaa 600
ttttaatatt cttccacttg cgatccttgc tacagtatga aa 642

<210> 25
<211> 42
<212> RNA
<213> Artificial Sequence
<220><223> SYNTHESIZED
<400> 25

agggggaagg gaauucuca guuuuagagc uaucuguuu ug 42
<210> 26
<211> 42

<212> RNA
<213> Artificial Sequence
<220><223> SYNTHESIZED
<400> 26

ucugcaacug gagucuucu guuuuagagc uaucuguuu ug 42
<210> 27
<211> 42
<212> RNA
<213> Artificial Sequence
<220><223> SYNTHESIZED
<400> 27

aggcgggagu cuucuggca guuuuagagc uaucuguuu ug 42