



등록특허 10-2802134



(19) 대한민국특허청(KR)

(12) 등록특허공보(B1)

(45) 공고일자 2025년05월07일

(11) 등록번호 10-2802134

(24) 등록일자 2025년04월24일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) A61K 38/00 (2006.01)
A61P 7/06 (2006.01) C12N 15/10 (2017.01)
C12N 15/113 (2010.01) C12N 15/52 (2006.01)
C12N 15/87 (2006.01)

(52) CPC특허분류
C12N 9/22 (2013.01)
A61K 38/00 (2013.01)

(21) 출원번호 10-2021-7037743

(22) 출원일자(국제) 2020년04월30일

심사청구일자 2023년04월28일

(85) 번역문제출일자 2021년11월18일

(65) 공개번호 10-2022-0012245

(43) 공개일자 2022년02월03일

(86) 국제출원번호 PCT/US2020/030782

(87) 국제공개번호 WO 2020/223514

국제공개일자 2020년11월05일

(30) 우선권주장

62/841,046 2019년04월30일 미국(US)

(뒷면에 계속)

(56) 선행기술조사문헌

W02015134812 A1*

*는 심사관에 의하여 인용된 문헌

(73) 특허권자

에넨도바이오 인코포레이티드

미국, 텔라웨어 19805, 윌밍턴, 스위트 403-비,
1013 센터 로드

(72) 발명자

바람 데이비드

이스라엘, 65255 텔 아비브, 5 심타 하타보 스트리트

이즈하 리오

이스라엘, 6910414 텔 아비브, 8 바트 스트리트
(뒷면에 계속)

(74) 대리인

특허법인한얼

전체 청구항 수 : 총 16 항

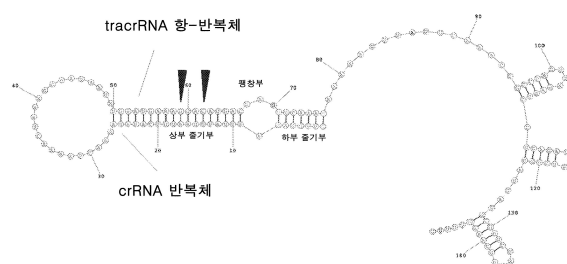
심사관 : 이형곤

(54) 발명의 명칭 신규한 OMNI-50 CRISPR 뉴클레아제

(57) 요약

본 발명은 서열번호: 3의 아미노산 서열과 적어도 95% 동일성을 갖는 서열을 포함하는 CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자를 포함하는 비-자연 발생 조성물을 제공한다.

대표도



(52) CPC특허분류

A61P 7/06 (2018.01)
C12N 15/102 (2013.01)
C12N 15/113 (2013.01)
C12N 15/52 (2013.01)
C12N 15/87 (2013.01)
C12N 2310/20 (2017.05)

(72) 발명자

헤르만 아사엘

이스라엘, 7420903 네스-지오나, 3 쉬자프 스트리트

록카 리앗

이스라엘, 7525941 리손 레지온, 7 하바알 쉘 토브

마르바크-바 나다브

이스라엘, 76466 레호보트, 하케렘 9/14

메론 누릿

이스라엘, 5258302 라맛 간, 하쇼샤님 블러바드 3

조지슨 조셉

이스라엘, 7628912 레호보트, 4 부스타나이 스트리트

(30) 우선권주장

62/897,806	2019년09월09일	미국(US)
62/931,630	2019년11월06일	미국(US)
62/959,672	2020년01월10일	미국(US)
62/991,285	2020년03월18일	미국(US)

명세서

청구범위

청구항 1

서열번호: 3의 아미노산 서열을 포함하는 클러스터링된 정규 배치 단편 회문 반복체(Clustered Regularly Interspaced Short Palindromic Repeats; CRISPR) 뉴클레아제를 포함하는, 비-자연 발생 조성물로서,

조성물이

a) CRISPR RNA(crRNA) 분자 및 트랜스활성화 CRISPR RNA(tracrRNA) 분자

(여기서, crRNA 분자, tracrRNA 분자, 및 CRISPR 뉴클레아제는 함께 자연 발생하지 않음); 또는

b) 단일 가이드 RNA(sgRNA) 분자

를 추가로 포함하고,

여기서, crRNA 분자 또는 sgRNA 분자는 표적 영역 내의 서열에 상보적인 가이드 서열 부분을 포함하며; 그리고

(i) crRNA 분자가 GUUUGAGAG의 서열을 포함하는 반복 서열 부분을 추가로 포함하거나;

(ii) tracrRNA 분자가 서열번호: 41 내지 43, 및 149 내지 154로부터 선택된 하나 이상의 서열을 추가로 포함하거나;

(iii) sgRNA 분자가 GUUUGAGAG의 서열을 포함하는 반복 서열 부분 또는 서열번호: 41 내지 43, 및 149 내지 154로부터 선택된 하나 이상의 서열을 추가로 포함하는,

조성물.

청구항 2

삭제

청구항 3

제1항에 있어서, 가이드 서열 부분은 21개 또는 22개 뉴클레오티드로 이루어진, 조성물.

청구항 4

삭제

청구항 5

삭제

청구항 6

제1항에 있어서, crRNA 분자가 GUUUGAGAG의 서열을 포함하는 crRNA 반복 서열 부분을 포함하는, 조성물.

청구항 7

제1항에 있어서, tracrRNA 분자가 서열번호: 41 내지 43, 및 149 내지 154로부터 선택된 하나 이상의 서열을 포함하는, 조성물.

청구항 8

삭제

청구항 9

삭제

청구항 10

제1항에 있어서, 표적 영역이 NGG, NAG, 및 NGA로부터 선택된 프로토스페이스 인접 모티프(Protospacer Adjacent Motif; PAM) 위치에 인접한, 조성물.

청구항 11

제1항에 있어서, CRISPR 뉴클레아제가 핵 편재 서열(nuclear localization sequence; NLS)을 추가로 포함하는, 조성물.

청구항 12

삭제

청구항 13

삭제

청구항 14

제1항에 있어서, 공여체 주형(donor template) 분자를 추가로 포함하는, 조성물.

청구항 15

삭제

청구항 16

삭제

청구항 17

제1항에 있어서, crRNA 분자, tracrRNA 분자, crRNA 분자와 tracrRNA 분자 모두, 또는 sgRNA 분자를 인코딩하는 DNA 분자를 추가로 포함하는, 조성물.

청구항 18

서열번호: 3의 아미노산 서열을 포함하는 CRISPR 뉴클레아제를 포함하는, 비-자연 발생 조성물로서, 조성물이

(i) 뉴클레아제-결합 RNA 뉴클레오티드 서열 부분; 및

(ii) 표적 DNA 서열 내의 서열에 상보적인 17개 내지 24개 뉴클레오티드의 서열로 이루어진 가이드 서열 부분을 포함하는 단일 가이드 RNA(sgRNA) 분자를 추가로 포함하고,

여기서, CRISPR 뉴클레아제가 sgRNA 분자와 복합체화하여 표적 DNA 서열과 혼성화할 수 있는 복합체를 형성할 수 있고;

표적 DNA 서열이 NGG, NAG, 및 NGA로부터 선택된 프로토스페이스 인접 모티프(PAM) 위치에 인접하며;

sgRNA 분자가 서열번호: 37, 38, 40, 44, 45, 87, 또는 88로부터 선택된 하나 이상의 서열을 포함하는, 조성물.

청구항 19

삭제

청구항 20

삭제

청구항 21

삭제

청구항 22

제18항에 있어서, sgRNA 분자가 50개 내지 200개 염기의 길이를 갖는, 조성물.

청구항 23

세포의 게놈 내 표적 부위의 뉴클레오티드 서열을 표적화하거나 변형하는 시험관내 방법으로서, 서열번호: 3의 아미노산 서열을 포함하는 클러스터링된 정규 배치 단편 회문 반복체(CRISPR) 뉴클레아제를 인코딩하는 핵산 분자를 세포 내로 도입하는 단계를 포함하고,

a) crRNA 분자 및 tracrRNA 분자; 또는

b) sgRNA 분자

를 세포 내로 도입하는 단계를 추가로 포함하고,

여기서, crRNA 분자 또는 sgRNA 분자는 표적 영역 내의 서열에 상보적인 가이드 서열 부분을 포함하며; 그리고

(i) crRNA 분자가 GUUUGAGAG의 서열을 포함하는 반복 서열 부분을 추가로 포함하거나;

(ii) tracrRNA 분자가 서열번호: 41 내지 43, 및 149 내지 154로부터 선택된 하나 이상의 서열을 추가로 포함하거나;

(iii) sgRNA 분자가 GUUUGAGAG의 서열을 포함하는 반복 서열 부분 또는 서열번호: 41 내지 43, 및 149 내지 154로부터 선택된 하나 이상의 서열을 포함하는,

방법.

청구항 24

삭제

청구항 25

제23항에 있어서, 공여체 주형 분자를 세포 내로 도입하는 단계를 추가로 포함하는, 방법.

청구항 26

삭제

청구항 27

포유동물 세포의 게놈 내 표적 부위의 뉴클레오티드 서열을 표적화하거나 변형하는 시험관내 방법으로서, 제18항의 조성물을 세포 내로 도입하는 단계를 포함하는, 방법.

청구항 28

제27항에 있어서, 가이드 서열 부분은 21개 또는 22개 뉴클레오티드로 이루어진, 방법.

청구항 29

삭제

청구항 30

제27항에 있어서, 공여체 주형 분자를 세포 내로 도입하는 단계를 추가로 포함하는, 방법.

청구항 31

삭제

청구항 32

삭제

청구항 33

삭제

청구항 34

삭제

청구항 35

삭제

청구항 36

삭제

청구항 37

삭제

청구항 38

삭제

청구항 39

제1항의 조성물을 포함하는, 돌연변이 장애를 갖는 대상체를 치료하기 위한 약학 조성물.

청구항 40

삭제

청구항 41

삭제

청구항 42

삭제

청구항 43

삭제

청구항 44

삭제

청구항 45

삭제

청구항 46

삭제

청구항 47

삭제

청구항 48

삭제

청구항 49

삭제

청구항 50

삭제

청구항 51

삭제

청구항 52

삭제

청구항 53

삭제

청구항 54

삭제

청구항 55

삭제

청구항 56

삭제

청구항 57

삭제

청구항 58

삭제

청구항 59

삭제

발명의 설명

기술 분야

- [0001] 본 출원은 2020년 3월 18일에 출원된 U.S. 가출원 번호 62/991,285, 2020년 1월 10일에 출원된 62/959,672, 2019년 11월 6일에 출원된 62/931,630, 2019년 9월 9일에 출원된 62/897,806, 및 2019년 4월 30일에 출원된 62/841,046의 이익을 청구하며, 그 내용이 본원에 참조로 포함된다.
- [0002] 본 출원을 통해, 괄호에 나타내는 것을 포함하여, 다양한 공보가 참조된다. 이의 전문이 본 출원에서 언급되는 모든 공보의 개시는 본 발명에 관련된 분야 및 본 발명에서 채택될 수 있는 당분야에서의 특성의 추가적인 설명을 제공하기 위해 본 출원 내에 참조로 본원에 포함된다.
- [0003] 서열 목록에 대한 참조
- [0004] 본 출원은 "200430_91116-A-PCT_SequenceListing_AWG.txt"로 명명된 파일로 제시되는 뉴클레오타이드 서열을 참

조로 포함하며, 이는 186킬로바이트 크기이고, MS-Windows와 호환성 운영 시스템을 갖는, IBM-PC 머신 포맷으로 2020년 4월 29일에 생성되었고, 본 출원의 일환으로 2020년 4월 30일에 출원된 텍스트 파일에 포함된다.

[0005] **발명의 분야**

[0006] 본 발명은 특히, 게놈 편집을 위한 조성물 및 방법에 관한 것이다.

배경 기술

[0007] 박테리아 및 고세균 적응 면역성의 CRISPR(클러스터링된 정규 배치 단편 회문나선 반복체, Clustered Regularly Interspaced Short Palindromic Repeat) 시스템은 단백질 조성 및 게놈 유전자위 구조의 커다란 다양성을 나타낸다. CRISPR 시스템은 연구 및 게놈 조작을 위해 중요한 도구가 되었다. 그럼에도 불구하고, CRISPR 시스템의 여러 상세사항은 결정되지 않았고 CRISPR 뉴클레아제의 적용 가능성은 서열 특이성 요건, 발현, 또는 전달 난제에 의해 제한될 수 있다. 상이한 CRISPR 뉴클레아제는 크기, PAM 부위, 온-표적 활성화, 특이성, 절단 패턴(예로 무딘(blunt), 엇갈린 말단), 및 절단 후 인델(indel) 형성의 돌출 패턴과 같은 다양한 특징을 갖는다. 상이한 특징 세트가 상이한 적용에 유용할 수 있다. 예를 들어, 일부 CRISPR 뉴클레아제는 다른 CRISPR 뉴클레아제가 PAM 부위의 제한으로 인해 표적화할 수 없는 특정 게놈 유전자위를 표적화할 수 있다. 또한, 현재 사용 중인 일부 CRISPR 뉴클레아제는 사전-면역성을 나타내며, 이는 생체내 적용 가능성을 제한할 수 있다. 문헌(Charlesworth 등, Nature Medicine (2019) 및 Wagner 등, Nature Medicine (2019))을 참고한다. 따라서, 신규한 CRISPR 뉴클레아제의 발견, 조작, 및 개선이 중요하다.

발명의 내용

[0008] 발명의 요약

[0009] 게놈 조작, 후성적 조작, 게놈 표적화, 세포의 게놈 편집, 및/또는 시험관내 진단을 위해 이용될 수 있는 조성물 및 방법이 본원에서 개시된다.

[0010] 개시된 조성물은 게놈 DNA 서열을 변형하기 위해 이용될 수 있다. 본원에서 사용된 바와 같이, 게놈 DNA는 관심 세포 또는 세포들에 존재하는 선형 및/또는 염색체 DNA 및/또는 플라스미드 또는 다른 염색체의 DNA 서열을 나타낸다. 일부 구현예에서, 관심 세포는 진행 세포이다. 일부 구현예에서, 관심 세포는 원핵 세포이다. 일부 구현예에서, 방법은 게놈 DNA 서열 내 사전-결정된 표적 부위에서 이중-가닥 절단(DSB)을 생성하여, 게놈 내 표적 부위(들)에서 DNA 서열의 돌연변이, 삽입, 및/또는 결실을 초래한다.

[0011] 따라서, 일부 구현예에서, 조성물은 클러스터링된 정규 배치 단편 회문나선 반복체(CRISPR) 뉴클레아제를 포함한다. 일부 구현예에서, CRISPR 뉴클레아제는 CRISPR-연관 단백질이다.

[0012] 일부 구현예에서, 조성물은 예자키엘라 페루엔시스(Ezakiella peruensis) 균주 M6.X2로부터 유래되는 CRISPR 뉴클레아제와 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85% 동일성을 갖는 클러스터링된 정규 배치 단편 회문나선 반복체(CRISPR) 뉴클레아제를 포함한다. 각각의 가능성은 별도 구현예를 나타낸다.

[0013] **OMNI-50 뉴클레아제**

[0014] 본 발명의 구현예는, 표 1에 제공된 바와 같이, "OMNI-50" 뉴클레아제로 명명되는 CRISPR 뉴클레아제를 제공한다.

[0015] 본 발명은 세포 내로 (i) 서열번호: 3의 아미노산 서열과 적어도 95% 동일성을 갖는 CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열로서, 서열번호: 12 또는 13의 핵산 서열과 적어도 95% 동일성을 갖는, 서열을 포함하는 핵산 분자를 포함하는 조성물 및 (ii) DNA-표적화 RNA 분자, 또는 표적 DNA 내 서열과 상보적인 뉴클레오티드 서열을 포함하는, DNA-표적화 RNA 분자를 인코딩하는 DNA 폴리뉴클레오티드를 도입하는 단계를 포함하는, 포유류 세포의 게놈 내 표적 부위에서 뉴클레오티드 서열을 변형하는 방법을 제공한다.

[0016] 본 발명은 또한 하기:

[0017] a) 직접 반복 서열에 연결된 가이드 서열 부분을 포함하는 하나 이상의 RNA 분자로서, 가이드 서열이 표적 서열과 혼성화할 수 있는, 분자, 또는 하나 이상의 RNA 분자를 인코딩하는 하나 이상의 뉴클레오티드 서열; 및

[0018] b) 서열번호: 3의 아미노산 서열과 적어도 95% 동일성을 갖는 아미노산 서열을 포함하는 CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자

- [0019] 를 포함하는 CRISPR 연관 시스템을 포함하는 비-자연 발생 조성물을 제공하며, 하나 이상의 RNA 분자는 표적 서열과 혼성화하고, 표적 서열은 프로토스페이서 인접 모티프(PAM)의 3'이고, 하나 이상의 RNA 분자는 RNA-가이드 뉴클레아제와 복합체를 형성한다.
- [0020] 본 발명은 또한 하기:
- [0021] a) 서열번호: 3의 아미노산 서열과 적어도 95% 동일성을 갖는 서열을 포함하는 CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자; 및
- [0022] b) 하기 중 적어도 하나를 포함하는, 하나 이상의 RNA 분자, 또는 하나 이상의 RNA 분자를 인코딩하는 하나 이상의 DNA 폴리뉴클레오티드:
- [0023] i) CRISPR 뉴클레아제와 상호작용할 수 있는/이에 결합할 수 있는 뉴클레아제-결합 RNA 뉴클레오티드 서열; 및
- [0024] ii) 표적 DNA 서열 내 서열과 상보적인 서열을 포함하는 DNA-표적화 RNA 뉴클레오티드 서열
- [0025] 을 포함하는 비-자연 발생 조성물을 제공하며, CRISPR 뉴클레아제는 하나 이상의 RNA 분자와 복합체화하여 표적 DNA 서열과 혼성화할 수 있는 복합체를 형성할 수 있다.

도면의 간단한 설명

- [0026] **도 1a:** 각각의 뉴클레아제에 대한 sgRNA의 설계에서 가능한 요소의 확인을 위해 사용되는 전체 듀플렉스 RNA 요소(crRNA:tracrRNA 키메라)의 예측된 이차 구조의 일례. **도 1b~1c:** 단일 뉴클레아제와 함께 사용하기 위해 설계된, 2개의 상이한 sgRNA, V1(**도 1b**) 및 V2(**도 1c**)의 영역 간 구조에서의 변화 일례. 상이한 위치에서의 상부 줄기부에서 듀플렉스를 단축함으로써, crRNA 및 tracrRNA가 테트라-루프 'gaaa'와 연결되어, sgRNA 스캐폴드를 생성하였다.
- 도 2a:** 무세포 시험관내 TXTL 시스템에서 OMNI-50 sgRNA V1에 대한 8 bp 서열을 따라 모든 가능한 PAM 위치의 축합된 4N 윈도우 라이브러리. 결실 검정 결과에 기반하여 PAM 부위에 대해 생성된 서열 모티프. 활성은 2개의 가장 크게 결실된 서열의 평균에 기반하여 추정하였고 1 - 결실 스코어로 계산되었다. **도 2b:** OMNI-50 sgRNA V2에 대한 8 bp 서열을 따라 모든 가능한 PAM 위치에 대해 생성된 서열 모티프.
- 도 3: 포유류 세포에서 OMNI-50의 발현.** OMNI-50 또는 SpCas9 뉴클레아제를 Hek293T 세포에서 일시적으로 전달 감염시켰다. 세포를 수확하고 72 h제에 용해시키고, 용해액을 사용하여 HA-태그에 대한 항체를 사용하는 웨스턴 블롯에 의해 포유류 세포에서 OMNI-50 발현을 시험하였다. SpCas9-HA를 동일한 방식으로 전달감염시키고 양성 대조군으로 작용하였다. GAPDH를 사용하여 로딩량을 표준화하였다.
- 도 4a: 인간 세포에서의 내인성 정확도.** OMNI-50 또는 SpCas9 뉴클레아제를 sgRNA 발현 플라스미드와 함께 DNA 전달감염에 의해 포유류 세포 시스템에서 발현시켰다. 세포 용해액을 부위 특이적 게놈 DNA 증폭 및 NGS를 위해 사용하였다. 인텔의 백분율을 섹션 vii, ELANeg35_OMNI-50 또는 ELANeg62_OMNI-50을 사용하는 HeLa 세포주에서의 표적 대 오프-표적 편집에 기재된 바와 같이 측정하고 분석하였다. 두 경우 모두에서 게놈 온 및 오프 표적 서열이 아래 차트에 주지되며, PAM 서열은 밑줄쳐 표시한다. 각각의 실험은 3회의 독립적 반복을 나타낸다. **도 4b:** OMNI-50 또는 SpCas9 표적화 ELANeg35의 RNP 도입에 용해, 부위 특이적 DNA 증폭 및 NGS가 뒤따랐다. 온 및 오프 표적 서열 둘 다의 편집 수준을 두 번째 시스템에 나타낸다.
- 도 5a ~ 도 5d: RNP로서의 OMNI-50 활성 검정.** OMNI-50 뉴클레아제를 과발현시키고 정제하였다. 정제된 단백질을 합성 sgRNA와 복합체화하여 RNP를 형성하였다. 시험관내 검정을 위해(**도 5a 및 도 5b**) RNP를 대응하는 표적 및 PAM 서열(표 5에 기재됨)을 함유하는 선형 DNA 주형과 인큐베이션하였다. 선형 주형의 절단에 의해 활성을 확인하였다. 생체내 검정을 위해(**도 5c 및 도 5d**), U2OS 세포를 RNP로 전기천공하고 활성을 NGS에 의한 인텔 빈도의 측정에 의해 결정하였다. **도 5a:** 상이한 스페이서 길이(17~23 bp)의 가이드 35(표 10)를 이용한 OMNI-50 RNP의 활성 검정. **도 5b:** 스페이서 길이 20 ~ 23 nt를 갖는 감소하는 양의 RNP(4 pmol, 2 pmol, 1.2 pmol, 0.6 pmol 및 0.2 pmol)를 100 ng DNA 표적 주형과 인큐베이션하였다. **도 5c:** U2OS 세포에서 RNP로서의 OMNI-50에 대한 활성 검정: 스페이서 길이 17 ~ 23 bp를 갖는 RNP를 U2OS 세포주 내로 전기천공하고 편집 수준(인텔)을 NGS에 의해 측정하였다. **도 5d:** U2OS 세포에서 RNP로서의 OMNI-50에 대한 활성 검정: ELANE g35 sgRNA V1~V4를 이용한 RNP를 U2OS 세포주 내로 전기천공하고 편집 수준(인텔)을 NGS에 의해 측정하였다.
- 도 6: iPSC에서 RNP로서의 OMNI-50에 대한 활성 검정.** 스페이서 길이 17 ~ 23 nt를 갖는 RNP(표 10)를 iPSC 세

포주 내로 전기천공하고 편집 수준(인텔)을 NGS에 의해 측정하였다.

도 7. 내인성 포유류 세포 맥락에서의 OMNI-50 뉴클레아제 활성. OMNI-50 뉴클레아제를 sgRNA 발현 플라스미드와 함께 DNA 전달감염에 의해 포유류 세포 시스템에서 발현시켰다. 세포 용해액을 부위 특이적 게놈 DNA 증폭 및 NGS를 위해 사용하였다. 인텔의 백분율을 측정하고 분석하여 편집 수준을 결정하였다. 가이드 RNA 없이 OMNI-50 뉴클레아제로 전달감염된 세포는 비교 및 배경 결정을 위한 음성 대조군으로 작용하였다. 상이한 게놈 위치에서의 편집 수준을 나타낸다.

도 8: OMNI-50 뉴클레아제의 도식적 표시. OMNI-50 뉴클레아제는 도메인 A-F로 도식에 나타낸, 몇몇 기능적 도메인을 포함한다. 도메인 A는 3개의 하위도메인, A1, A2, 및 A3을 포함하며, 흰색 상자로 도식에 나타낸다. 도메인 B는 가로선으로 도식에 나타낸다. 도메인 C는 3개의 하위도메인, C1, C2, 및 C3을 포함하며, 연한색 상자로 도식에 나타낸다. 도메인 D는 사선으로 도식에 나타낸다. 도메인 E는 진한색 상자로 도식에 나타낸다. 도메인 F는 점으로 표시한 상자로 도식에 나타낸다.

발명을 실시하기 위한 구체적인 내용

- [0027] 본 발명의 일부 양태에 따르면, 개시되는 조성물은 클러스터링된 정규 배치 단편 회문나선 반복체(CRISPR) 뉴클레아제 및/또는 이를 인코딩하는 서열을 포함하는 핵산 분자를 포함한다.
- [0028] 일부 구현예에서, CRISPR 뉴클레아제는 서열번호: 3에 나타낸 바와 같은 CRISPR 뉴클레아제와 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 또는 82% 아미노산 서열 동일성을 갖는 아미노산 서열을 포함한다. 하나의 구현예에서 CRISPR 뉴클레아제를 인코딩하는 서열은 서열번호: 11~13으로 구성되는 군으로부터 선택되는 핵산 서열과 적어도 95% 동일성을 갖는다.
- [0029] 일부 구현예에서, CRISPR 뉴클레아제는 예자키엘라 페루엔시스 균주 M6.X2로부터 유래되는 CRISPR 뉴클레아제와 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 82%, 81%, 80%, 75% 아미노산 서열 동일성을 갖는 아미노산 서열을 포함한다. 각각의 가능성은 별도 구현예를 나타낸다.
- [0030] 본 발명의 일부 양태에 따르면, 개시되는 조성물은 CRISPR 뉴클레아제 또는 변이체 CRISPR 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 포함하는 DNA 작제물 또는 벡터 시스템을 포함한다. 일부 구현예에서, CRISPR 뉴클레아제 또는 변이체 CRISPR 뉴클레아제를 인코딩하는 뉴클레오티드 서열은 관심 세포에서 작동 가능한 프로모터에 작동 가능하게 연결된다. 일부 구현예에서, 관심 세포는 진핵 세포이다. 일부 구현예에서 관심 세포는 포유류 세포이다. 일부 구현예에서, 조작된 CRISPR 뉴클레아제를 인코딩하는 핵산 서열은 특정 유기체로부터의 세포에서 사용하기 위해 최적화된 코돈이다. 일부 구현예에서, 뉴클레아제를 인코딩하는 핵산 서열은 대장균을 위해 최적화된 코돈이다. 일부 구현예에서, 뉴클레아제를 인코딩하는 핵산 서열은 진핵 세포를 위해 최적화된 코돈이다. 일부 구현예에서, 뉴클레아제를 인코딩하는 핵산 서열은 포유류 세포를 위해 최적화된 코돈이다.
- [0031] 일부 구현예에서, 조성물은 서열번호: 3과 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90% 동일성을 갖는 CRISPR 효소를 인코딩하는 폴리뉴클레오티드에 작동 가능하게 연결된 이중성 프로모터를 포함하는 재조합 핵산을 포함한다. 각각의 가능성은 별도 구현예를 나타낸다.
- [0032] 조성물의 하나의 구현예에서, CRISPR 뉴클레아제는 서열번호: 3에 나타낸 아미노산 서열과 적어도 75%, 80%, 85, 90%, 95%, 또는 97% 동일성을 갖거나 CRISPR 뉴클레아제를 인코딩하는 서열은 서열번호: 11, 12, 및 13으로 구성되는 군으로부터 선택되는 뉴클레오티드 서열과 적어도 75%, 80%, 85, 90%, 95%, 또는 97% 서열 동일성을 갖는다.
- [0033] 일부 구현예에 따르면, 서열번호: 3의 아미노산 서열과 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80% 동일성을 갖는 서열을 포함하는 CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자를 포함하는 조작된 또는 비-자연 발생 조성물이 제공된다. 각각의 가능성은 별도 구현예를 나타낸다.
- [0034] 하나의 구현예에서, CRISPR 뉴클레아제는 조작되거나 비-자연 발생이다. CRISPR 뉴클레아제는 또한 재조합일 수 있다. 이러한 CRISPR 뉴클레아제는 여러 원천으로부터의 유전 물질을 함께 가져오기 위해 실험실 방법(분자 클로닝)을 사용하여 생산되어, 다른 경우 생물학적 유기체에서 확인되지 않았을 서열을 생성한다.
- [0035] 하나의 구현예에서, 본 발명의 CRISPR 뉴클레아제는 하나 이상의 RNA 분자와 복합체화되는 경우 SpCas9 뉴클레

아제에 비해 표적 부위에 대해 증가된 특이성을 나타낸다.

- [0036] 하나의 구현예에서, 본 발명의 CRISPR 뉴클레아제 및 하나 이상의 RNA 분자의 복합체는 SpCas9 뉴클레아제에 비해 적어도 유지되는 표적 부위의 온-표적 편집 활성 및 감소된 오프-표적 활성을 나타낸다.
- [0037] 하나의 구현예에서, CRISPR 뉴클레아제는 DNA-표적화 RNA 분자(gRNA)와 상호작용할 수 있는 RNA-결합 부분 및 부위-유도 효소 활성을 나타내는 활성 부분을 추가로 포함한다.
- [0038] 하나의 구현예에서, 조성물은 DNA-표적화 RNA 분자 또는 DNA-표적화 RNA 분자를 인코딩하는 DNA 폴리뉴클레오티드를 추가로 포함하며, 여기서 DNA-표적화 RNA 분자는 표적 영역 내 서열과 상보적인 뉴클레오티드 서열을 포함하고, DNA-표적화 RNA 분자 및 CRISPR 뉴클레아제는 함께 자연 발생하지 않는다.
- [0039] 하나의 구현예에서, DNA-표적화 RNA 분자는 서열 GUUUGAGAG를 포함하는 crRNA 반복 서열을 포함한다.
- [0040] 하나의 구현예에서, DNA-표적화 RNA 분자는 서열번호: 41~43 및 서열번호: 149~154로부터 선택되는 하나 이상의 서열을 포함하는 tracrRNA 서열을 포함한다.
- [0041] 하나의 구현예에서, DNA-표적화 RNA 분자는 CRISPR 뉴클레아제와 복합체를 형성할 수 있는 뉴클레오티드 서열을 추가로 포함한다.
- [0042] 본 발명은 또한 하기:
- [0043] a) 직접 반복 서열에 연결된 가이드 서열 부분을 포함하는 하나 이상의 RNA 분자로서, 가이드 서열이 표적 서열과 혼성화할 수 있는, 분자, 또는 하나 이상의 RNA 분자를 인코딩하는 하나 이상의 뉴클레오티드 서열; 및
- [0044] b) 서열번호: 3의 아미노산 서열과 적어도 95% 동일성을 갖는 아미노산 서열을 포함하는 CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자;
- [0045] 를 포함하는 CRISPR 연관 시스템을 포함하는 비-자연 발생 조성물을 제공하며, 하나 이상의 RNA 분자는 표적 서열과 혼성화하고, 표적 서열은 프로토스페이서 인접 모티프(PAM)의 3'이고, 하나 이상의 RNA 분자는 RNA-가이드 뉴클레아제와 복합체를 형성한다.
- [0046] 하나의 구현예에서, 조성물은 CRISPR 뉴클레아제(tracrRNA)와 복합체를 형성할 수 있는 뉴클레오티드 서열을 포함하는 RNA 분자 또는 CRISPR 뉴클레아제와 복합체를 형성할 수 있는 RNA 분자를 인코딩하는 서열을 포함하는 DNA 폴리뉴클레오티드를 추가로 포함한다.
- [0047] 하나의 구현예에서, 조성물은 상동성 유도 복구(HDR)를 위한 공여체 주형을 추가로 포함한다.
- [0048] 하나의 구현예에서, 조성물은 세포의 게놈 내 표적 영역을 편집할 수 있다.
- [0049] 조성물의 하나의 구현예에서 CRISPR 뉴클레아제는 서열번호: 3과 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80% 동일성을 가지며, DNA-표적화 RNA 분자 내 CRISPR 뉴클레아제와 복합체를 형성할 수 있는 뉴클레오티드 서열은 서열번호: 37~45, 87~88, 149~154, 및 GUUUGAGAG로부터 선택되는 서열을 포함한다.
- [0050] 일부 구현예에 따르면, 하기:
- [0051] (a) RNA-결합 부분; 및
- [0052] 부위-유도 효소 활성을 나타내는 활성 부분
- [0053] 을 포함하는, CRISPR 뉴클레아제, 또는 CRISPR 뉴클레아제를 인코딩하는 폴리뉴클레오티드로서, CRISPR 뉴클레아제가 서열번호: 3과 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80% 동일성을 갖는, 뉴클레아제, 또는 폴리뉴클레오티드; 및
- [0054] (b) 하기:
- [0055] i) 표적 DNA 서열 내 서열과 상보적인 뉴클레오티드 서열을 포함하는, DNA-표적화 RNA 서열; 및
- [0056] ii) CRISPR 뉴클레아제의 RNA-결합 부분과 상호작용할 수 있는, 단백질-결합 RNA 서열
- [0057] 을 포함하는 하나 이상의 RNA 분자 또는 하나 이상의 RNA 분자를 인코딩하는 DNA 폴리뉴클레오티드
- [0058] 를 포함하는 비-자연 발생 조성물이 제공되며, DNA 표적화 RNA 서열 및 CRISPR 뉴클레아제는 함께 자연 발생하

지 않는다. 각각의 가능성은 별도 구현예를 나타낸다.

- [0059] 일부 구현예에서, DNA-표적화 RNA 서열 및 단백질-결합 RNA 서열을 포함하는 단일 RNA 분자가 제공되며, 여기서 RNA 분자는 CRISPR 뉴클레아제와 복합체를 형성할 수 있고 DNA 표적화 모듈로 작용할 수 있다. 일부 구현예에서, RNA 분자는 최대 1000 염기, 900 염기, 800 염기, 700 염기, 600 염기, 500 염기, 400 염기, 300 염기, 200 염기, 100 염기, 50 염기의 길이를 갖는다. 각각의 가능성은 별도 구현예를 나타낸다. 일부 구현예에서, DNA-표적화 RNA 서열을 포함하는 제1 RNA 분자 및 단백질-결합 RNA 서열을 포함하는 제2 RNA 분자는 염기쌍 형성에 의해 상호작용하거나 대안적으로 함께 융합되어 CRISPR 뉴클레아제와 복합체를 형성하고 DNA 표적화 모듈로 작용하는 하나 이상의 RNA 분자를 형성한다.
- [0060] 일부 구현예에서, CRISPR 뉴클레아제는 서열번호: 3과 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80% 동일성을 가지며, RNA 분자는 서열번호: 37~45, 87~88, 149~154 및 GUUUGAGAG로부터 선택되는 서열을 포함한다.
- [0061] 본 발명은 또한 하기:
- [0062] a) 서열번호: 3의 아미노산 서열과 적어도 95% 동일성을 갖는 서열을 포함하는 CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자; 및
- [0063] b) 하기:
- [0064] i) CRISPR 뉴클레아제와 상호작용할 수 있는/이에 결합할 수 있는 뉴클레아제-결합 RNA 뉴클레오티드 서열; 및
- [0065] ii) 표적 DNA 서열 내 서열과 상보적인 서열을 포함하는 DNA-표적화 RNA 뉴클레오티드 서열
- [0066] 중 적어도 하나를 포함하는, 하나 이상의 RNA 분자, 또는 하나 이상의 RNA 분자를 인코딩하는 하나 이상의 DNA 폴리뉴클레오티드
- [0067] 를 포함하는 비-자연 발생 조성물을 제공하며, 여기서 CRISPR 뉴클레아제는 하나 이상의 RNA 분자와 복합체화하여 표적 DNA 서열과 혼성화할 수 있는 복합체를 형성할 수 있다.
- [0068] 하나의 구현예에서, CRISPR 뉴클레아제 및 하나 이상의 RNA 분자는 표적 DNA 서열의 절단을 실시하기 위해 표적 DNA 서열에 결합할 수 있는 CRISPR 복합체를 형성한다.
- [0069] 하나의 구현예에서, CRISPR 뉴클레아제 및 하나 이상의 RNA 분자 중 적어도 하나는 함께 자연 발생하지 않는다.
- [0070] 하나의 구현예에서:
- [0071] a) CRISPR 뉴클레아제는 RNA-결합 부분 및 부위-유도 효소 활성을 나타내는 활성 부분을 포함하며;
- [0072] b) DNA-표적화 RNA 뉴클레오티드 서열은 표적 DNA 서열 내 서열과 상보적인 뉴클레오티드 서열을 포함하고;
- [0073] c) 뉴클레아제-결합 RNA 뉴클레오티드 서열은 CRISPR 뉴클레아제의 RNA-결합 부분과 상호작용하는 서열을 포함한다.
- [0074] 하나의 구현예에서, 뉴클레아제-결합 RNA 뉴클레오티드 서열 및 DNA-표적화 RNA 뉴클레오티드 서열은 단일 가이드 RNA 분자(sgRNA) 상에 있으며, 여기서 sgRNA 분자는 CRISPR 뉴클레아제와 복합체를 형성하고 DNA 표적화 모듈로 작용할 수 있다.
- [0075] 하나의 구현예에서, 뉴클레아제-결합 RNA 뉴클레오티드 서열은 제1 RNA 분자 상에 있으며 DNA-표적화 RNA 뉴클레오티드 서열은 단일 가이드 RNA 분자 상에 있고, 여기서 제1 및 제2 RNA 서열은 염기쌍 형성에 의해 상호작용하거나 함께 융합되어 CRISPR 뉴클레아제와 복합체를 형성하고 표적화 모듈로 작용하는 하나 이상의 RNA 분자 또는 sgRNA를 형성한다.
- [0076] 하나의 구현예에서, sgRNA는 최대 1000 염기, 900 염기, 800 염기, 700 염기, 600 염기, 500 염기, 400 염기, 300 염기, 200 염기, 100 염기, 50 염기의 길이를 갖는다.
- [0077] 하나의 구현예에서, 조성물은 상동성 유도 복구(HDR)를 위한 공여체 주형을 추가로 포함한다.
- [0078] 일부 구현예에서, (a) CRISPR 뉴클레아제는 서열번호: 3과 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80% 동일성을 갖거나, (b) CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자는

서열번호: 11, 12, 또는 13에 나타난 핵산 서열과 적어도 95% 서열 동일성의 서열을 포함하고, PAM은 NGG이다. 적합한 PAM 서열의 비제한적 예에는 GGG, AGG, 및 TGG가 포함된다. 상기 구현예에서, DNA-표적화 RNA 분자 내 CRISPR 뉴클레아제와 복합체를 형성할 수 있는 뉴클레오티드 서열은 서열번호: 37~45, 87~88, 149~154 및 GUUGAGAG로부터 선택되는 서열을 포함한다.

- [0079] 일부 구현예에서, CRISPR 뉴클레아제는 NAG 또는 NGA의 서열을 갖는 PAM을 이용한다.
- [0080] 하나의 구현예에서, CRISPR 뉴클레아제는 야생형 CRISPR 뉴클레아제의 아미노산 서열 대비 1~10, 10~20, 20~30, 30~40, 40~50, 50~60, 60~70, 70~80, 80~90, 90~100, 100~110, 110~120, 120~130, 130~140, 또는 140~150 아미노산 치환, 결실, 및/또는 삽입을 포함한다.
- [0081] 하나의 구현예에서, CRISPR 뉴클레아제는 야생형 CRISPR 뉴클레아제 대비 적어도 2%, 5%, 7%, 10%, 15%, 20%, 25%, 30%, 또는 35% 증가된 특이성을 나타낸다.
- [0082] 하나의 구현예에서, CRISPR 뉴클레아제는 적어도 2%, 5%, 7%, 10%, 15%, 20%, 25%, 30%, 또는 35% 증가된 활성을 나타낸다.
- [0083] 하나의 구현예에서, CRISPR 뉴클레아제는 야생형 CRISPR 뉴클레아제 대비 변경된 PAM 특이성을 갖는다.
- [0084] 하나의 구현예에서, CRISPR 뉴클레아제는 비-자연 발생이다.
- [0085] 하나의 구현예에서, CRISPR 뉴클레아제는 조작된 것이며 비천연 또는 합성 아미노산을 포함한다.
- [0086] 하나의 구현예에서, CRISPR 뉴클레아제는 조작된 것이며 핵 편재 서열(NLS), 세포 침투 펩티드 서열, 및/또는 친화도 태그 중 하나 이상을 포함한다.
- [0087] 하나의 구현예에서, CRISPR 뉴클레아제는 진핵 세포의 핵 내에서 검출 가능한 양으로 CRISPR 뉴클레아제를 포함하는 CRISPR 복합체의 축적을 유도하기 충분한 강도의 하나 이상의 핵 편재 서열을 포함한다.
- [0088] 본 발명은 또한 본 발명의 임의의 조성물을 세포 내로 도입하는 단계를 포함하는 무세포 시스템 또는 세포의 게놈 내 표적 부위에서 뉴클레오티드 서열을 변형하는 방법을 제공한다.
- [0089] 하나의 구현예에서, 세포는 진핵 세포이다.
- [0090] 또 다른 구현예에서, 세포는 원핵 세포이다.
- [0091] 일부 구현예에서, 하나 이상의 RNA 분자는 RNA 뉴클레아제(tracrRNA)와 복합체를 형성할 수 있는 뉴클레오티드 분자를 포함하는 RNA 서열 또는 CRISPR 뉴클레아제와 복합체를 형성할 수 있는 뉴클레오티드 서열을 포함하는 RNA 분자를 인코딩하는 DNA 폴리뉴클레오티드를 추가로 포함한다.
- [0092] 하나의 구현예에서, CRISPR 뉴클레아제는 아미노-말단에서 또는 근처에서 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 NLS, 카복시-말단에서 또는 근처에서 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 NLS, 또는 아미노-말단에서 또는 근처에서 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 NLS 및 카복시-말단에서 또는 근처에서 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개 이상의 NLS의 조합을 포함한다. 하나의 구현예에서 1~4개 NLS가 CRISPR 뉴클레아제와 융합된다. 하나의 구현예에서, NLS는 CRISPR 뉴클레아제의 개방-해독틀(ORF) 내에 위치한다.
- [0093] 발현된 단백질질의 아미노-말단에서 또는 근처에서, 카복시-말단에서 또는 근처에서, 또는 ORF 내에서 NLS와 융합시키는 방법은 당분야에 잘 알려져 있다. 일례로서, NLS를 CRISPR 뉴클레아제의 아미노-말단으로 융합시키기 위해, NLS의 핵산 서열은 NLS-융합 CRISPR 뉴클레아제를 인코딩하는 핵산 상에서 CRISPR 뉴클레아제의 시작 코돈 직후에 배치된다. 반대로, NLS를 CRISPR 뉴클레아제의 카복시-말단으로 융합시키기 위해, NLS의 핵산 서열은 CRISPR 뉴클레아제의 마지막 아미노산을 인코딩하는 코돈 뒤에 및 정지 코돈 앞에 배치된다.
- [0094] CRISPR 뉴클레아제의 ORF를 따른 임의의 위치에서의 NLS, 세포 침투 펩티드 서열, 및/또는 친화도 태그의 임의의 조합이 본 발명에서 고려된다.
- [0095] 본원에서 제공되는 CRISPR 뉴클레아제의 아미노산 서열 및 핵산 서열에는 CRISPR 뉴클레아제의 인접한 아미노산 또는 핵산 서열을 단속하기 위해 삽입된 NLS 및/또는 TAG가 포함될 수 있다.
- [0096] 하나의 구현예에서, 하나 이상의 NLS는 일렬 반복된다.
- [0097] 하나의 구현예에서, 하나 이상의 NLS는 NLS의 가장 가까운 아미노산이 N- 또는 C-말단으로부터 폴리펩티드쇄를 따라 약 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50개 이상의 아미노산 이내인 경우 N- 또는 C-말단의 인근인

것으로 간주된다.

- [0098] 논의된 바와 같이, CRISPR 뉴클레아제는 핵 편재 서열(NLS), 세포 침투 펩티드 서열, 및/또는 친화도 태그 중 하나 이상을 포함하도록 조작될 수 있다.
- [0099] 하나의 구현예에서, CRISPR 뉴클레아제는 하나 이상의 RNA 분자와 복합체화된 경우 야생형 CRISPR 뉴클레아제 대비 표적 부위에 대해 증가된 특이성을 나타낸다.
- [0100] 하나의 구현예에서, CRISPR 뉴클레아제 및 하나 이상의 RNA 분자의 복합체는 야생형 CRISPR 뉴클레아제 대비 표적 부위의 적어도 유지된 온-표적 편집 활성 및 감소된 오프-표적 활성을 나타낸다.
- [0101] 하나의 구현예에서, 조성물은 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 뉴클레오티드산 분자에 작동 가능하게 연결된 이중성 프로모터를 포함하는 재조합 핵산 분자를 추가로 포함한다.
- [0102] 하나의 구현예에서, CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자는 비-자연 발생이거나 조작된 것이다.
- [0103] 본 발명은 또한 본 발명의 임의의 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자를 포함하는 벡터 시스템을 포함하는 비-자연 발생 또는 조작된 조성물을 제공한다.
- [0104] 본 발명은 또한 대상체의 게놈 내 표적 부위에서 뉴클레오티드 서열의 변형을 포함하는, 게놈 돌연변이와 연관된 질환을 겪는 대상체의 치료를 위한 본 발명의 임의의 조성물의 용도를 제공한다.
- [0105] 본 발명은 세포 내로 (i) 서열번호: 3의 아미노산 서열과 적어도 95% 동일성을 갖는 CRISPR 뉴클레아제 또는 CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자를 포함하는 조성물로서, 서열이 서열번호: 12 또는 13으로 구성되는 군으로부터 선택되는 핵산 서열과 적어도 95% 동일성을 갖는, 조성물 및 (ii) DNA-표적화 RNA 분자, 또는 표적 DNA 내 서열과 상보적인 뉴클레오티드 서열을 포함하는, DNA-표적화 RNA 분자를 인코딩하는 DNA 폴리뉴클레오티드를 도입하는 단계를 포함하는, 포유류 세포의 게놈 내 표적 부위에서 뉴클레오티드 서열을 변형하는 방법을 제공한다.
- [0106] 일부 구현예에서, 방법은 생체의 수행된다. 일부 구현예에서, 방법은 생체내 수행된다. 일부 구현예에서, 방법의 일부 단계는 생체의 수행되며 일부 단계는 생체내 수행된다. 일부 구현예에서 포유류 세포는 인간 세포이다.
- [0107] 하나의 구현예에서, 방법은 세포 내로 (iii) 뉴클레아제-결합 RNA 서열을 포함하는 RNA 분자 또는 CRISPR 뉴클레아제와 상호작용하는 뉴클레아제-결합 RNA를 포함하는 RNA 분자를 인코딩하는 DNA 폴리뉴클레오티드를 도입하는 단계를 추가로 포함한다.
- [0108] 하나의 구현예에서, DNA 표적화 RNA 분자는 CRISPR 뉴클레아제와 활성 복합체를 형성하기 적합한 crRNA 분자이다.
- [0109] 하나의 구현예에서, 뉴클레아제-결합 RNA 서열을 포함하는 RNA 분자는 CRISPR 뉴클레아제와 활성 복합체를 형성하기 적합한 tracrRNA 분자이다.
- [0110] 하나의 구현예에서, DNA-표적화 RNA 분자 및 뉴클레아제-결합 RNA 서열을 포함하는 RNA 분자는 단일 가이드 RNA 분자 형태로 융합된다.
- [0111] 하나의 구현예에서, 방법은 세포 내로 (iv) 프로토스페이서 서열과 상보적인 서열을 포함하는 RNA 분자를 도입하는 단계를 추가로 포함한다.
- [0112] 하나의 구현예에서, CRISPR 뉴클레아제는 하나 이상의 RNA 분자와 복합체를 형성하며 프로토스페이서 인접 모티프(PAM)의 3'에서 이중 가닥 절단을 실시한다.
- [0113] 하나의 구현예에서, CRISPR 뉴클레아제는 하나 이상의 RNA 분자와 복합체를 형성하며 프로토스페이서 인접 모티프(PAM)의 5'에서 이중 가닥 절단을 실시한다.
- [0114] 일부 구현예에서, (a) CRISPR 뉴클레아제는 서열번호: 3과 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 85%, 80% 동일성을 갖거나, (b) CRISPR 뉴클레아제를 인코딩하는 서열을 포함하는 핵산 분자는 서열번호: 11, 12, 또는 13에 나타난 핵산 서열과 적어도 95% 서열 동일성의 서열을 포함하고, PAM은 NGG이다. 적합한 PAM 서열의 비제한적 예는 GGG, AGG, 및 TGG가 포함된다. 상기 구현예에서, DNA-표적화 RNA 분자에서 CRISPR 뉴클레아제와 복합체를 형성할 수 있는 뉴클레오티드 서열은 서열번호: 37~45, 87~88, 149~154 및

GUUUGAGAG로부터 선택되는 서열을 포함한다.

- [0115] 일부 구현예에서, CRISPR 뉴클레아제는 NAG 또는 NGA의 서열을 갖는 PAM을 이용한다.
- [0116] 본원에서 기재된 임의의 방법의 하나의 구현예에서, 방법은 대상체의 게놈 내 표적 부위에서 뉴클레오티드 서열을 변형하는 단계를 포함하는, 게놈 돌연변이와 연관된 질환을 겪는 대상체를 치료하기 위한 것이다.
- [0117] 하나의 구현예에서, 방법은 먼저 게놈 돌연변이와 연관된 질환을 겪는 대상체를 선택하는 단계 및 대상체로부터 세포를 수득하는 단계를 포함한다.
- [0118] 본 발명은 또한 본원에서 기재된 임의의 방법에 의해 수득되는 변형된 세포 또는 세포들을 제공한다. 하나의 구현예에서 이들 변형된 세포 또는 세포들은 자손 세포를 생성할 수 있다. 하나의 구현예에서 이들 변형된 세포 또는 세포들은 이식 후 자손 세포를 생성할 수 있다.
- [0119] 본 발명은 또한 이들 변형된 세포 및 약학적으로 허용 가능한 담체를 포함하는 조성물을 제공한다. 또한, 세포를 약학적으로 허용 가능한 담체와 혼합하는 단계를 포함하는, 이를 제조하는 시험관내 또는 생체의 방법이 제공된다.
- [0120] DNA-표적화 RNA 분자
- [0121] 본 발명의 구현예에서, DNA-표적화 RNA 서열은 가이드 서열 부분을 포함한다. RNA 분자의 "가이드 서열 부분"은 특정 표적 DNA 서열에 혼성화할 수 있는 뉴클레오티드 서열을 나타내며, 예로, 가이드 서열 부분은 가이드 서열 부분의 길이에 따라 표적화되는 DNA 서열과 전체 상보적인 뉴클레오티드 서열을 갖는다. 일부 구현예에서, 가이드 서열 부분은 17, 18, 19, 20, 21, 22, 23, 또는 24개 뉴클레오티드 길이, 또는 대략 17~24, 18~22, 19~22, 18~20, 17~20, 또는 21~22개 뉴클레오티드 길이이다. 가이드 서열 부분의 전체 길이는 가이드 서열 부분의 길이에 따라 표적화되는 DNA 서열과 전체 상보적이다. 가이드 서열 부분은 CRISPR 복합체의 DNA 표적화 부분으로 작용하는 가이드 서열 부분을 갖는 CRISPR 뉴클레아제와 복합체를 형성할 수 있는 RNA 분자의 일부일 수 있다. 가이드 서열 부분을 갖는 RNA 분자가 CRISPR 분자와 동시간대에 존재하는 경우, RNA 분자는 CRISPR 뉴클레아제를 특정 표적 DNA 서열로 표적화할 수 있다. 각각의 가능성은 별도 구현예를 나타낸다. RNA 분자는 임의의 요망되는 서열을 표적화하도록 맞춤 설계될 수 있다.
- [0122] 본 발명의 구현예에서, CRISPR 뉴클레아제는 20개 이하의 뉴클레오티드, 및/또는 24개 이상의 뉴클레오티드를 갖는 가이드 서열 부분을 포함하는 RNA 분자와 함께 사용되는 경우의 그 절단 활성에 비해, 21~23개 뉴클레오티드를 갖는 가이드 서열 부분을 포함하는 RNA 분자와 함께 사용되는 경우에 더 큰 절단 활성을 갖는다. 본 발명의 구현예에서, CRISPR 뉴클레아제는 20개 이하의 뉴클레오티드, 및/또는 23개 이상의 뉴클레오티드를 갖는 가이드 서열 부분을 포함하는 RNA 분자와 함께 사용되는 경우의 그 절단 활성에 비해, 21~22개 뉴클레오티드를 갖는 가이드 서열 부분을 포함하는 RNA 분자와 함께 사용되는 경우에 더 큰 절단 활성을 갖는다. 하나의 구현예에서, CRISPR 뉴클레아제는 22개 뉴클레오티드를 갖는 가이드 서열 부분을 포함하는 RNA 분자와 함께 사용되는 경우에 그 가장 큰 절단 활성을 갖는다.
- [0123] 하나의 구현예에서, 이러한 CRISPR 뉴클레아제는 서열번호: 3에 나타난 아미노산 서열과 적어도 95% 동일성을 갖거나 CRISPR 뉴클레아제를 인코딩하는 서열은 서열번호: 11~13 중 어느 하나와 적어도 95% 서열 동일성을 갖는다. 하나의 구현예에서, 이러한 CRISPR 뉴클레아제는 서열번호: 3에 나타난 아미노산 서열과 적어도 100%, 99.5%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 또는 82% 동일성을 갖거나 CRISPR 뉴클레아제를 인코딩하는 서열은 서열번호: 11~13 중 어느 하나와 적어도 100%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 또는 82% 서열 동일성을 갖는다.
- [0124] CRISPR 뉴클레아제의 특징적인 표적화된 뉴클레아제 활성은 그 특이적 도메인의 다양한 기능에 의해 부여된다. 본 출원에서 OMNI-50 도메인은 도 8 OMNI-50 도식에서 제시된 바와 같이 도메인 A, 도메인 B, 도메인 C, 도메인 D, 도메인 E, 및 도메인 F로 정의된다.
- [0125] 본원에서 사용된 바와 같이, 도메인 A는 3개의 하위도메인: 하위도메인 A1, 하위도메인 A2, 및 하위도메인 A3을 포함한다. 본원에서 사용된 바와 같이, 하위도메인 A1은 서열번호: 3의 1~10 내의 아미노산 위치에서 시작되며 45~55 내의 아미노산 위치에서 끝나고; 하위도메인 A2는 서열번호: 3의 736~746 내의 아미노산 위치에서 시작되며 784~794 내의 아미노산 위치에서 끝나고; 하위도메인 A3은 서열번호: 3의 957~967 내의 아미노산 위치에서 시작되며 1091~1101 내의 아미노산 위치에서 끝난다. 스미스-워터만(Smith-Waterman) 알고리즘을 사용하여 생성

된 국소 정렬의 선호되는 분석에 기반하여, 하나의 구현예에서 하위도메인 A1은 서열번호: 3의 아미노산 1 내지 50으로 확인되었으며, 하위도메인 A2는 서열번호: 3의 아미노산 741 내지 789로 확인되었고, 하위도메인 A3은 서열번호: 3의 아미노산 962 내지 1096으로 확인되었다.

[0126] 본원에서 사용된 바와 같이, 도메인 B는 서열번호: 3의 46~56 내의 아미노산 위치에서 시작되며 78~88 내의 아미노산 위치에서 끝난다. 스미스-워터만 알고리즘을 사용하여 생성된 국소 정렬의 선호되는 분석에 기반하여, 하나의 구현예에서 도메인 B는 서열번호: 3의 아미노산 51 내지 83으로 확인되었다.

[0127] 본원에서 사용된 바와 같이, 도메인 C는 3개의 하위도메인: 하위도메인 C1, 하위도메인 C2, 및 하위도메인 C3, 또는 대안적으로 2개의 하위도메인: 하위도메인 Ca 및 하위도메인 Cb를 포함한다. 본원에서 사용된 바와 같이, 하위도메인 C1은 서열번호: 3의 79~89 내의 아미노산 위치에서 시작되며 155~165 내의 아미노산 위치에서 끝나고; 하위도메인 C2는 서열번호: 3의 156~166 내의 아미노산 위치에서 시작되며 294~304 내의 아미노산 위치에서 끝나고; 하위도메인 C3은 서열번호: 3의 295~305 내의 아미노산 위치에서 시작되며 732~742 내의 아미노산 위치에서 끝난다. 스미스-워터만 알고리즘을 사용하여 생성된 국소 정렬의 선호되는 분석에 기반하여, 하나의 구현예에서 하위도메인 C1은 서열번호: 3의 아미노산 84~160으로 확인되었으며, 하위도메인 C2는 서열번호: 3의 아미노산 161~299로 확인되었고, 하위도메인 C3은 서열번호: 3의 아미노산 300~737로 확인되었다. 본원에서 사용된 바와 같이, 하위도메인 Ca는 서열번호: 3의 79~89 내의 아미노산 위치에서 시작되며 473~483 내의 아미노산 위치에서 끝나고; 하위도메인 Cb는 서열번호: 3의 474~484 내의 아미노산 위치에서 시작되며 732~742 내의 아미노산 위치에서 끝난다. 스미스-워터만 알고리즘을 사용하여 생성된 국소 정렬의 분석에 기반하여, 하나의 구현예에서 하위도메인 Ca는 서열번호: 3의 아미노산 84~478로 확인되었으며, 하위도메인 Cb는 서열번호: 3의 아미노산 479~737로 확인되었다.

[0128] 본원에서 사용된 바와 같이, 도메인 D는 서열번호: 3의 785~795 내의 아미노산 위치에서 시작되며 956~966 내의 아미노산 위치에서 끝난다. 스미스-워터만 알고리즘을 사용하여 생성된 국소 정렬의 선호되는 분석에 기반하여, 하나의 구현예에서 도메인 D는 서열번호: 3의 아미노산 790 내지 961로 확인되었다.

[0129] 본원에서 사용된 바와 같이, 도메인 E는 서열번호: 3의 1092~1102 내의 아미노산 위치에서 시작되며 1191~1201 내의 아미노산 위치에서 끝난다. 스미스-워터만 알고리즘을 사용하여 생성된 국소 정렬의 선호되는 분석에 기반하여, 하나의 구현예에서 도메인 E는 서열번호: 3의 아미노산 1097 내지 1196으로 확인되었다.

[0130] 본원에서 사용된 바와 같이, 도메인 F는 서열번호: 3의 1192~1202 내의 아미노산 위치에서 시작되며 1360~1370 내의 아미노산 위치에서 끝난다. 스미스-워터만 알고리즘을 사용하여 생성된 국소 정렬의 선호되는 분석에 기반하여, 하나의 구현예에서 도메인 F는 서열번호: 3의 아미노산 1197 내지 1370으로 확인되었다.

[0131] 각각의 OMNI-50 뉴클레아제 도메인의 활성이 본원에 기재되며, 각각의 도메인 활성은 뉴클레아제의 유리한 특성의 양태를 제공한다.

[0132] 구체적으로, OMNI-50 도메인 A는 DNA 가닥 절단에 참여하는 뉴클레아제 활성 부위를 함유한다. 도메인 A는 표적화 RNA 분자가 DNA 표적 부위에서 결합하는 DNA 가닥을 절단한다.

[0133] 도메인 B는 표적 DNA 부위에 대한 OMNI-50 결합 시 DNA 절단 활성의 개시에 연루된다.

[0134] 도메인 C는 표적화 RNA 분자에 결합하고 표적 부위 인식을 위한 특이성 제공에 참여한다. 도메인 C는 하위도메인 C1, 하위도메인 C2, 및 하위도메인 C3을 포함하며, 이는 각각 도메인 C 활성의 특이적인 기능적 양태에 참여한다. 예를 들어, C3은 DNA 표적 부위의 감지에 연루되며; C2는 뉴클레아제 도메인(예로 도메인 D)의 활성화 조절에 연루되고; C1은 표적 부위에서 뉴클레아제 도메인의 잠금에 연루된다. 따라서, 도메인 C는 오프-표적 서열의 절단 제어에 참여한다.

[0135] 도메인 D는 DNA 가닥 절단에 참여하는 뉴클레아제 활성 부위를 함유한다. 도메인 D는 DNA 표적 부위에서 결합하는 표적화 RNA 분자에 의해 변위되는 DNA 가닥을 절단한다.

[0136] 도메인 E는 토포이소머라제 도메인과 구조적으로 유사하다.

[0137] 도메인 F는 PAM 부위 조사 및 인식 양태를 포함하는, PAM 부위 특이성의 제공에 연루된다.

[0138] 다른 CRISPR 뉴클레아제 도메인 및 이의 일반적 기능의 추가 설명은, 특히 본원에 참조로 포함되는, 문헌(Mir 등, ACS Chem. Biol. (2019), Palermo 등, Quarterly Reviews of Biophysics (2018), Jiang and Doudna, Annual Review of Biophysics (2017), Nishimasu 등, Cell (2014) 및 Nishimasu 등, Cell (2015))에서 확인될

수 있다.

- [0139] 본 발명의 하나의 양태에서, OMNI-50 도메인 또는 하위도메인과 유사성을 갖는 아미노산 서열은 펩티드가 OMNI-50 도메인 또는 하위도메인 활성의 유리한 특성을 나타내도록, 비-자연 발생 펩티드, 예로 CRISPR 뉴클레아제의 설계 및 제조에서 이용될 수 있다.
- [0140] 하나의 구현예에서, 이러한 펩티드, 예로 CRISPR 뉴클레아제는 OMNI-50 뉴클레아제의 도메인 A 또는 그 3개의 하위 도메인 중 어느 하나, 도메인 B, 도메인 C 또는 그 3개의 하위 도메인 중 어느 하나, 도메인 D, 도메인 E, 또는 도메인 F 중 적어도 하나의 아미노산 서열과 적어도 100%, 99.5%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 또는 82% 동일성을 갖는 아미노산 서열을 포함한다. 하나의 구현예에서, 펩티드는 OMNI-50 뉴클레아제의 도메인 A 또는 그 3개의 하위 도메인 중 어느 하나, 도메인 B, 도메인 C 또는 그 3개의 하위 도메인 중 어느 하나, 도메인 D, 도메인 E, 또는 도메인 F 중 적어도 하나의 아미노산 서열과 적어도 100%, 99.5%, 99%, 98%, 97%, 96%, 95%, 94%, 93%, 92%, 91%, 90%, 89%, 88%, 87%, 86%, 85%, 84%, 83%, 또는 82% 동일성을 갖는 펩티드 아미노산 서열 외부에서 전장 OMNI-50 아미노산 서열(서열번호: 3)에 비해 광범위한 아미노산 가변성을 나타낸다. 하나의 구현예에서, 펩티드는 2개의 도메인 서열 간에 개재 아미노산 서열을 포함한다. 하나의 구현예에서, 개재 아미노산 서열은 1~10, 10~20, 20~40, 40~50 또는 최대 100개 아미노산 길이이다. 하나의 구현예에서, 개재 서열은 링커 서열이다.
- [0141] 본 발명의 하나의 양태에서, 펩티드에서 본원에 기재되는 OMNI-50 뉴클레아제의 도메인 중 어느 하나를 인코딩하는 아미노산 서열은 원래의 OMNI-50 도메인 서열 대비 하나 이상의 아미노산 치환을 포함할 수 있다. 아미노산 치환은 보존적 치환, 즉 원래의 아미노산과 유사한 화학적 특성을 갖는 아미노산으로의 치환일 수 있다. 예를 들어, 양으로 하전된 아미노산은 대안적인 양으로 하전된 아미노산으로 치환될 수 있으며, 예로 아르기닌 잔기는 라이신 잔기로 치환될 수 있고, 또는 극성 아미노산은 상이한 극성 아미노산으로 치환될 수 있다. 보존적 치환은 보다 관용 가능하며, OMNI-50 뉴클레아제의 도메인 중 어느 하나를 인코딩하는 아미노산 서열은 이러한 치환을 10%만큼 함유할 수 있다. 아미노산 치환은 급격한 치환, 즉 원래의 아미노산과 상이한 화학적 특성을 갖는 아미노산으로의 치환일 수 있다. 예를 들어, 양으로 하전된 아미노산은 음으로 하전된 아미노산으로 치환될 수 있으며, 예로 아르기닌 잔기는 글루탐산 잔기로 치환될 수 있고, 또는 극성 아미노산은 비-극성 아미노산으로 치환될 수 있다. 아미노산 치환은 반-보존적 치환일 수 있거나, 아미노산 치환은 임의의 다른 아미노산으로 일 수 있다. 치환은 원래의 OMNI-50 도메인 기능 대비 활성을 변경할 수 있으며, 예로 촉매성 뉴클레아제 활성을 감소시킬 수 있다.
- [0142] 본 발명의 일부 양태에 따르면, 개시되는 조성물은 CRISPR 뉴클레아제를 포함하는 비-자연 발생 조성물을 포함하며, 여기서 CRISPR 뉴클레아제는 OMNI-50 뉴클레아제의 도메인 A, 도메인 B, 도메인 C, 도메인 D, 도메인 E, 또는 도메인 F 중 적어도 하나의 아미노산 서열에 대응하는 아미노산 서열을 포함한다. 본 발명의 일부 구현예에서, CRISPR 뉴클레아제는 적어도 1, 적어도 2, 적어도 3, 적어도 4, 또는 적어도 5 아미노산 서열을 포함하며, 여기서 각각의 아미노산 서열은 OMNI-50 뉴클레아제의 도메인 A, 도메인 B, 도메인 C, 도메인 D, 도메인 E, 또는 도메인 F의 아미노산 서열 중 어느 하나에 대응한다. 따라서, CRISPR 뉴클레아제에는 OMNI-50 뉴클레아제의 임의의 도메인 A, 도메인 B, 도메인 C, 도메인 D, 도메인 E, 또는 도메인 F에 대응하는 아미노산 서열의 임의의 조합이 포함될 수 있다.
- [0143] 일부 구현예에서, CRISPR 뉴클레아제는 하기 중 적어도 하나를 포함하는 도메인 A를 포함한다:
- [0144] a) 서열번호: 3의 아미노산 1 내지 50과 적어도 97% 서열 동일성을 갖는 하위도메인 A1;
- [0145] b) 서열번호: 3의 아미노산 741 내지 789와 적어도 97% 서열 동일성을 갖는 하위도메인 A2; 또는
- [0146] c) 서열번호: 3의 아미노산 962 내지 1096과 적어도 97% 서열 동일성을 갖는 하위도메인 A3.
- [0147] 일부 구현예에서, CRISPR 뉴클레아제는 서열번호: 3의 아미노산 51 내지 83과 적어도 97% 서열 동일성을 갖는 도메인 B를 포함한다.
- [0148] 일부 구현예에서, CRISPR 뉴클레아제는 하기 중 적어도 하나를 포함하는 도메인 C를 포함한다:
- [0149] a) 서열번호: 3의 아미노산 84 내지 160과 적어도 97% 서열 동일성을 갖는 하위도메인 C1;
- [0150] b) 서열번호: 3의 아미노산 161 내지 299와 적어도 97% 서열 동일성을 갖는 하위도메인 C2; 또는
- [0151] c) 서열번호: 3의 아미노산 300 내지 737과 적어도 97% 서열 동일성을 갖는 하위도메인 C3.

- [0152] 일부 구현예에서, CRISPR 뉴클레아제는 하기 중 적어도 하나를 포함하는 도메인 C를 포함한다:
- [0153] a) 서열번호: 3의 아미노산 84 내지 478과 적어도 97% 서열 동일성을 갖는 하위도메인 Ca; 및
- [0154] b) 서열번호: 3의 아미노산 479 내지 737과 적어도 97% 서열 동일성을 갖는 하위도메인 Cb.
- [0155] 일부 구현예에서, 도메인 C는 서열번호: 3의 아미노산 84 내지 737과 적어도 97% 서열 동일성을 갖는다.
- [0156] 일부 구현예에서, CRISPR 뉴클레아제는 서열번호: 3의 아미노산 790 내지 961과 적어도 97% 서열 동일성을 갖는 도메인 D를 포함한다.
- [0157] 일부 구현예에서, CRISPR 뉴클레아제는 서열번호: 3의 아미노산 1097 내지 1196과 적어도 97% 서열 동일성을 갖는 도메인 E를 포함한다.
- [0158] 일부 구현예에서, CRISPR 뉴클레아제는 서열번호: 3의 아미노산 1197 내지 1370과 적어도 97% 서열 동일성을 갖는 도메인 F를 포함한다.
- [0159] 일부 구현예에서, CRISPR 뉴클레아제는 도메인 A, 도메인 B, 도메인 C, 도메인 D, 도메인 E, 및 도메인 F를 포함하며, 여기서
- [0160] a) 도메인 A는 하기:
- [0161] i. 서열번호: 3의 아미노산 1 내지 50과 적어도 97% 서열 동일성을 갖는 하위도메인 A1;
- [0162] ii. 서열번호: 3의 아미노산 741 내지 789와 적어도 97% 서열 동일성을 갖는 하위도메인 A2; 및
- [0163] iii. 서열번호: 3의 아미노산 962 내지 1096과 적어도 97% 서열 동일성을 갖는 하위도메인 A3
- [0164] 을 포함하고;
- [0165] b) 도메인 B는 서열번호: 3의 아미노산 51 내지 83과 적어도 97% 서열 동일성을 가지고;
- [0166] c) 도메인 C는 서열번호: 3의 아미노산 84 내지 737과 적어도 97% 서열 동일성을 가지고;
- [0167] d) 도메인 D는 서열번호: 3의 아미노산 790 내지 961과 적어도 97% 서열 동일성을 가지고;
- [0168] e) 도메인 E는 서열번호: 3의 아미노산 1097 내지 1196과 적어도 97% 서열 동일성을 가지고;
- [0169] f) 도메인 F는 서열번호: 3의 아미노산 1197 내지 1370과 적어도 97% 서열 동일성을 갖는다.
- [0170] 일부 구현예에서, CRISPR 뉴클레아제 서열은 적어도 100~250, 250~500, 500~1000, 또는 1000~2000개 아미노산 길이이다.
- [0171] 본 발명의 일부 양태에 따르면, 개시되는 조성물은 펩티드를 포함하는 비-자연 발생 조성물을 포함하며, 여기서 펩티드는 OMNI-50 뉴클레아제의 도메인 A, 도메인 B, 도메인 C, 도메인 D, 도메인 E, 또는 도메인 F 중 적어도 하나의 아미노산 서열과 적어도 97% 서열 동일성을 갖는 아미노산 서열을 포함한다.
- [0172] 일부 구현예에서, 도메인 A의 아미노산 서열은 하기 중 적어도 하나의 아미노산 서열을 포함한다:
- [0173] a) 서열번호: 3의 아미노산 1 내지 50과 적어도 97% 서열 동일성을 갖는 하위도메인 A1;
- [0174] b) 서열번호: 3의 아미노산 741 내지 789와 적어도 97% 서열 동일성을 갖는 하위도메인 A2; 또는
- [0175] c) 서열번호: 3의 아미노산 962 내지 1096과 적어도 97% 서열 동일성을 갖는 하위도메인 A3.
- [0176] 일부 구현예에서, 도메인 B의 아미노산 서열은 서열번호: 3의 아미노산 51 내지 83과 적어도 97% 서열 동일성을 갖는다.
- [0177] 일부 구현예에서, 도메인 C의 아미노산 서열은 하기 중 적어도 하나의 아미노산 서열을 포함한다:
- [0178] a) 서열번호: 3의 아미노산 84 내지 160과 적어도 97% 서열 동일성을 갖는 하위도메인 C1;
- [0179] b) 서열번호: 3의 아미노산 161 내지 299와 적어도 97% 서열 동일성을 갖는 하위도메인 C2; 또는
- [0180] c) 서열번호: 3의 아미노산 300 내지 737과 적어도 97% 서열 동일성을 갖는 하위도메인 C3.
- [0181] 일부 구현예에서, 도메인 C의 아미노산 서열은 하기 중 적어도 하나의 아미노산 서열을 포함한다:

- [0182] a) 서열번호: 3의 아미노산 84 내지 478과 적어도 97% 서열 동일성을 갖는 하위도메인 Ca; 및
- [0183] b) 서열번호: 3의 아미노산 479 내지 737과 적어도 97% 서열 동일성을 갖는 하위도메인 Cb.
- [0184] 일부 구현예에서, 도메인 C의 아미노산 서열은 서열번호: 3의 아미노산 84 내지 737과 적어도 97% 서열 동일성을 갖는다.
- [0185] 일부 구현예에서, 도메인 D의 아미노산 서열은 서열번호: 3의 아미노산 790 내지 961과 적어도 97% 서열 동일성을 갖는다.
- [0186] 일부 구현예에서, 도메인 E의 아미노산 서열은 서열번호: 3의 아미노산 1097 내지 1196과 적어도 97% 서열 동일성을 갖는다.
- [0187] 일부 구현예에서, 도메인 F의 아미노산 서열은 서열번호: 3의 아미노산 1197 내지 1370과 적어도 97% 서열 동일성을 갖는다.
- [0188] 일부 구현예에서, 아미노산 서열은 적어도 100~250, 250~500, 500~1000, 또는 1000~2000개 아미노산 길이이다.
- [0189] 본 발명의 일부 양태에 따르면, 개시되는 조성물은 OMNI-50 뉴클레아제의 도메인 A, 도메인 B, 도메인 C, 도메인 D, 도메인 E, 또는 도메인 F 중 적어도 하나의 아미노산 서열과 적어도 97% 서열 동일성을 갖는 아미노산 서열을 인코딩하는 폴리뉴클레오티드를 포함하는 비-자연 발생 조성물을 포함한다.
- [0190] 본 발명의 일부 양태에 따르면, 개시되는 조성물은 OMNI-50 뉴클레아제의 도메인 A, 도메인 B, 도메인 C, 도메인 D, 도메인 E, 또는 도메인 F 중 적어도 하나의 아미노산 서열과 적어도 97% 서열 동일성을 갖는 비-자연 발생 아미노산 서열을 포함한다.
- [0191] 본 발명의 일부 양태에 따르면, 개시되는 방법은 본원에서 기재되는 구현예 중 어느 하나의 조성물을 세포 내로 도입하는 단계를 포함하는, 무세포 시스템 또는 세포의 계층 내 표적 부위에서 뉴클레오티드 서열을 변형하는 방법을 포함한다.
- [0192] 일부 구현예에서, 세포는 진핵 세포, 바람직하게는 포유류 세포 또는 식물 세포이다.
- [0193] 본 발명의 일부 양태에 따르면, 개시되는 방법은 대상체의 계층 내 표적 부위에서 뉴클레오티드 서열을 변형하는 단계를 포함하는, 계층 돌연변이와 연관된 질환을 겪는 대상체의 치료를 위한 본원에서 기재되는 조성물 중 어느 하나의 사용을 포함한다.
- [0194] 본 발명의 일부 양태에 따르면, 개시되는 방법은 돌연변이 장애와 연관된 대립유전자로 본원에서 기재되는 조성물 중 어느 하나를 표적화하는 단계를 포함하는, 돌연변이 장애를 갖는 대상체를 치료하는 방법을 포함한다.
- [0195] 일부 구현예에서, 돌연변이 장애는 임의의 신생물, 연령-관련 황반 변성, 정신분열병, 신경학적, 신경변성, 또는 운동 장애, 취약 X 증후군, 시크리타제-관련 장애, 프리온-관련 장애, ALS, 중독, 자폐증, 알츠하이머병, 호중구감소증, 염증-관련 장애, 파킨슨병, 혈액 및 응고 질환 및 장애, 세포 조절이상 및 종양학 질환 및 장애, 염증 및 면역-관련 질환 및 장애, 대사성, 간, 신장 및 단백질 질환 및 장애, 근육 및 골격 질환 및 장애, 피부학적 질환 및 장애, 신경학적 및 신경세포 질환 및 장애, 및 눈 질환 및 장애로부터 선택되는 질환 또는 장애에 관한 것이다.
- [0196] 일부 구현예에서, 돌연변이 장애는 베타 지중해빈혈 또는 낫 적혈구 빈혈이다.
- [0197] 일부 구현예에서, 질환과 연관된 대립유전자는 BCL11A이다.
- [0198] 질환 및 치료법
- [0199] 본 발명의 특정 구현예는 유전자 편집, 치료 방법, 또는 치료법의 형태로서 질환 또는 장애와 연관된 특정 유전적 유전자위로 뉴클레아제를 표적화한다. 예를 들어, 유전자의 편집 또는 녹아웃을 유도하기 위해, 본원에서 개시되는 신규한 뉴클레아제는 맞춤 설계된 가이드 RNA 분자를 사용하여 유전자의 병리적 돌연변이체 대립유전자로 특이적으로 표적화될 수 있다. 가이드 RNA 분자는 바람직하게는 먼저 뉴클레아제의 PAM 요건을 고려함으로써 설계되며, 이는 본원에서 나타난 바와 같이 유전자 편집이 수행되는 시스템에 또한 의존한다. 예를 들어, 표적 부위로 OMNI-50 뉴클레아제를 표적화하도록 설계되는 가이드 RNA 분자는 OMNI-50 PAM 서열 "NGG" 인근 영역과 상보적인 스페이서 영역을 함유하도록 설계된다. 가이드 RNA 분자는 추가로 바람직하게는 뉴클레아제의 특이적 활성을 증가시키고 오프-표적 효과를 감소시키기 위해 충분하고 바람직하게는 최적의 길이의 스페이서 영역(즉

표적 대립유전자와 상보성을 갖는 가이드 RNA 분자의 영역)을 함유하도록 설계된다. 예를 들어, OMNI-50 뉴클레아제를 표적화하도록 설계되는 가이드 RNA 분자는 높은 온-표적 절단 활성을 위해 22 nt 스페이서를 함유하도록 설계될 수 있다.

- [0200] 비제한적 예로서, 가이드 RNA 분자는 뉴클레아제에 의해 유도되는 DNA 손상 시 비-상동 말단 연결(NHEJ) 경로가 유도되고 틀이동 돌연변이의 도입에 의해 돌연변이체 대립유전자의 침묵화를 야기하도록, 돌연변이체 대립유전자의 특정 영역, 예로 시작 코돈 근처로 뉴클레아제를 표적화하도록 설계될 수 있다. 가이드 RNA 분자 설계에 대한 상기 접근은 우성 음성 돌연변이의 효과를 변경하고 이에 따라 대상체를 치료하는 데 특히 유용하다. 별도의 비제한적 예로서, 가이드 RNA 분자는 뉴클레아제에 의해 유도되는 DNA 손상 시 상동성 유도 복구(HDR) 경로가 유도되고 돌연변이체 대립유전자의 주형 매개 교정을 야기하도록, 돌연변이된 대립유전자의 특정 병리적 돌연변이를 표적화하도록 설계될 수 있다. 가이드 RNA 분자 설계에 대한 상기 접근은 돌연변이된 대립유전자의 흡배수체부족 효과를 변경하고 이에 따라 대상체를 치료하는 데 특히 유용하다.
- [0201] 질환 또는 장애를 치료하기 위한 변경에 대해 표적화될 수 있는 특정 유전자의 비제한적 예가 아래에서 본원에 제시된다. 돌연변이 장애를 유도하는 특정 질환-연관 유전자 및 돌연변이는 문헌에 기재되어 있다. 이러한 돌연변이는 질환 연관된 유전자의 대립유전자로 CRISPR 조성물을 표적화하도록 DNA-표적화 RNA 분자를 설계하기 위해 사용될 수 있고, 여기서 CRISPR 조성물은 DNA 손상을 유도하고 DNA 복구 경로를 유도하여 대립유전자를 변경하고 이에 따라 돌연변이 장애를 치료한다.
- [0202] ELANE 유전자에서의 돌연변이는 호중구감소증과 연관된다. 이에 따라, 비제한적으로, ELANE를 표적화하는 본 발명의 구현에는 호중구감소증을 겪는 대상체를 치료하는 방법에서 사용될 수 있다.
- [0203] CXCR4는 인간 면역결핍 바이러스 1형(HIV-1) 감염을 위한 공동-수용체이다. 따라서, 비제한적으로, CXCR4를 표적화하는 본 발명의 구현에는 HIV-1을 겪는 대상체를 치료하거나 대상체에서 HIV-1 감염에 대한 내성을 부여하는 방법에서 사용될 수 있다.
- [0204] 프로그래밍된 세포사 단백질 1(PD-1) 손상은 종양 세포의 CAR-T 세포 매개된 사멸을 증강시키며 PD-1은 다른 암 치료법에서의 표적일 수 있다. 따라서, 비제한적으로, PD-1을 표적화하는 본 발명의 구현에는 암을 겪는 대상체를 치료하는 방법에서 사용될 수 있다. 하나의 구현예에서, 치료는 PD-1 결핍으로 본 발명에 따라 변형된 T 세포를 이용한 CAR-T 세포 치료법이다.
- [0205] 또한, BCL11A는 헤모글로빈 생산의 억제에서 역할을 담당하는 유전자이다. 글로빈 생산은 BCL11A를 억제함으로써 지중해빈혈 또는 낫 적혈구 빈혈과 같은 질환을 치료하기 위해 증가될 수 있다. 예를 들어, PCT 국제 공개 번호 WO 2017/077394A2; U.S. 공개 번호 US2011/0182867A1; Humbert 등 Sci. Transl. Med. (2019); 및 Canver 등 Nature (2015)을 참고한다. 따라서, 비제한적으로, BCL11A의 증강제를 표적화하는 본 발명의 구현에는 베타 지중해빈혈 또는 낫 적혈구 빈혈을 겪는 대상체를 치료하는 방법에서 사용될 수 있다.
- [0206] 본 발명의 구현에는 또한 아래의 표 A 또는 표 B에 기재된 임의의 질환 또는 장애를 연구하거나, 변경하거나, 치료하기 위한 임의의 질환-연관 유전자의 표적화를 위해 사용될 수 있다. 실제로, 유전적 유전자위와의 연관된-임의의 질환은 적절한 질환-연관 유전자를 표적화하기 위해 본원에서 개시되는 뉴클레아제, 예를 들어, U.S. 공개 번호 2018/0282762A1 및 유럽 특허 번호 EP3079726B1에 기재된 것들을 사용함으로써 연구되거나, 변경되거나, 치료될 수 있다.

[0207] [표 A]

질환, 장애 및 이의 연관된 유전자

질환 / 장애	유전자(들)
신생물	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4; Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF; HIF1a; HIF3a; Met; HRG; Bcl2; PPAR 알파; PPAR 감마; WT1(윌름 종양); FGF 수용체 패밀리 구성원(5개 구성원: 1, 2, 3, 4, 5); CDKN2a; APC; RB(망막모세포종); MEN1; VHL; BRCA1; BRCA2; AR(안드로겐 수용체); TSG101; IGF; IGF 수용체; Igf1(4개 변이체); gf2(3개 변이체); Igf 1 수용체; Igf 2 수용체; Bax; Bcl2; 카스파제 패밀리(9개 구성원: 1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
연령-관련 황반 변성	Abcr; Ccl2; Cc2; cp(ceruloplasmin); Timp3; cathepsinD; Vldlr; Ccr2
정신분열병	Neuregulin1(Nrg1); Erb4(Neuregulin에 대한 수용체); Complexin1(Cplx1); Tph1 트립토판 하이드록실라제; Tph2 트립토판 하이드록실라제 2; Neurexin 1; GSK3; GSK3a; GSK3b
신경학적, 신경변성, 및 운동 장애	5-HTT(Slc6a4); COMT; DRD(Drd1a); SLC6A3; DAOA; DTNBP1; Dao(Dao1)
트리뉴클레오티드 반복 장애	HTT(헌팅턴 Dx); SBMA/SMAX1/AR(케네디 Dx); FXN/X25(프리드리히 실조); ATX3(마차도-조셉(Machado-Joseph' s) Dx); ATXN1 및 ATXN2(척수소뇌 실조); DMPK(근긴장 디스트로피); Atrophin-1 및 Atn1(DRPLA Dx); CBP(Creb-BP-전반적 불안정성); VLDLR(알츠하이머); Atxn7; Atxn10
취약 X 증후군	FMR2; FXR1; FXR2; mGLUR5
시크리타제 관련 장애	APH-1(알파 및 베타); Presenilin(Psen1); nicastrin(Ncstn); PEN-2
기타	Nos1; Parp1; Nat1; Nat2

[0208]

질환 / 장애	유전자(들)
프리온 관련 장애	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF(VEGF-a; VEGF-b; VEGF-c)
중독	Prkce(알코올); Drd2; Drd4; ABAT(알코올); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Grial(알코올)
자폐증	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; 취약 X(FMR2(AFF2); FXR1; FXR2; Mglur5)
알츠하이머병	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterin; PS1; SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28(Aqp1, Aquaporin 1); Uchl1; Uchl3; APP
염증	IL-10; IL-1(IL-1a; IL-1b); IL-13; IL-17(IL-17a(CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); IL-23; Cx3cr1; ptpn22; TNFa; IBD에 있어서 NOD2/CARD15; IL-6; IL-12(IL-12a; IL-12b); CTLA4; Cx3cl1
파킨슨병	α -Synuclein; DJ-1; LRRK2; Parkin; PINK1

[0209]

[0210] [표 B]

질환, 장애 및 이의 연관된 유전자

질환 분류	질환 및 연관된 유전자
혈액 및 응고 질환 및 장애	빈혈(CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); 무표지 림프구 증후군(TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), 출혈 장애(TBXA2R, P2RX1, P2X1); 인자 H 및 인자 H-유사 1(HF1, CFH, HUS); 인자 V 및 인자 VIII(MCFD2); 인자 VII 결핍(F7); 인자 X 결핍(F10); 인자 XI 결핍(F11); 인자 XII 결핍(F12, HAF); 인자 XIIIa 결핍(F13A1, F13A); 인자 XIIIb 결핍(F13B); 판코니 빈혈(FANCA, FACA, FAL, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCI, PHF9, FANCL, FANCM, KIAA1596); 식혈세포성 림프조직구증 장애(PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); 혈우병 A(F8, F8C, HEMA); 혈우병 B(F9, HEMB), 출혈성 장애(PI, ATT, F5); 백혈구 결핍 및 장애(ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); 낮 적혈구 빈혈(HBB); 지중해빈혈(HBA2, HBB, HBD, LCRB, HBA1)
세포 조절이상 및 종양학 질환 및 장애	B-세포 비-호지킨 림프종(BCL7A, BCL7); 백혈병(TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOXA4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMP5, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN)

[0211]

질환 분류	질환 및 연관된 유전자
염증 및 면역 관련 질환 및 장애	AIDS(KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); 자가면역 림프증식성 증후군(TNFRSF6, APT1, FAS, CD95, ALPS1A); 복합 면역결핍(IL2RG, SCIDX1, SCIDX, IMD4); HIV-1(CCL5, SCYA5, D17S136E, TCP228), HIV 감수성 또는 감염(IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCR5(CCR5)); 면역결핍(CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TAC1); 염증(IL-10, IL-1(IL-1a, IL-1b), IL-13, IL-17(IL-17a(CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), IL-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12(IL-12a, IL-12b), CTLA4, Cx3cl1); 중증 복합 면역결핍(SCID)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4)
대사성, 간, 신장 및 단백질 질환 및 장애	아밀로이드 신경병증(TTR, PALB); 아밀로이드증(APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); 경화(KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); 낭성 섬유증(CFTR, ABCC7, CF, MRP7); 글리코젠 저장 질환(SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); 간 샘종, 142330(TCF1, HNF1A, MODY3), 간 부전, 조기 개시, 및 신경학적 장애(SCOD1, SCO1), 간 리파제 결핍(LIPC), 간모세포종, 암 및 암종(CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; 수질 낭성 신장 질환(UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); 폐닐케톤뇨증(PAH, PKU1, QDPR, DHPR, PTS); 다낭성 신장 및 간 질환(FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63)
근육 / 골격 질환 및 장애	베커 근 디스트로피(DMD, BMD, MYF6), 뤼시엔느 근 디스트로피(DMD, BMD); 에머리-드레이푸스 근 디스트로피(LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); 얼굴어깨위팔 근 디스트로피(FSHMD1A, FSHD1A); 근 디스트로피(FKRP,

[0212]

질환 분류	질환 및 연관된 유전자
	MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); 뼈화석증(LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); 근 위축(VAFB, VAFB, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1)
피부학적 질환 및 장애	백색증(TYR, OCA2, TYRP1, SLC45A2, LYST), 외배엽 형성이상(EDAR, EDARADD, WNT10A), 에틀러-단로스 증후군(COL5A1, COL5A2, COL1A1, COL1A2, COL3A1, TNXB, ADAMTS2, PLOD1, FKBP14), 비늘증-연관 장애(FLG, STS, TGM1, ALOXE3/ALOX12B, KRT1, KRT10, ABCA12, KRT2, GJB2, TGM1, ABCA12, CYP4F22, ALOXE3, CERS3, NSHDL, EBP, MBTPS2, GJB2, SPINK5, AGHD5, PHYH, PEX7, ALDH3A2, ERCC2, ERCC3, GFT2H5, GBA), 색소 실조증(IKBKG, NEMO), 결절 경화증(TSC1, TSC2), 조기 노화 증후군(POLR3A, PYCR1, LMNA, POLD1, WRN, DMPK)
신경학적 및 신경세포 질환 및 장애	ALS(SOD1, ALS2, STX, FUS, TARDBP, VEGF(VEGF-a, VEGF-b, VEGF-c); 알츠하이머병(APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLA2, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); 자폐증(Mecp2, BZRAP1, MDGA2, Sema5A, Neurexin 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); 취약 X 증후군(FMR2, FXR1, FXR2, mGLUR5); 헌팅턴병 및 헌팅턴병 유사 장애(HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); 파킨슨병(NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); 레트 증후군(MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79,

[0213]

질환 분류	질환 및 연관된 유전자
	x-Synuclein, DJ-1); 정신분열병(Neuregulin1(Nrg1), Erb4(Neuregulin에 대한 수용체), Complexin1(Cplx1), Tph1 트립토판 하이드록실라제, Tph2, 트립토판 하이드록실라제 2, Neurexin 1, GSK3, GSK3a, GSK3b, 5-HTT(Slc6a4), COMT, DRD(Drd1a), SLC6A3, DAOA, DTNBP1, Dao(Dao1)); 시크리타제 관련 장애(APH-1(알파 및 베타), Presenilin(Psen1), nicastrin,(Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); 트리뉴클레오티드 반복 장애(HTT(헌팅틴 Dx), SBMA/SMAX1/AR(케네디 Dx), FXN/X25(프리드리히 실조), ATX3(마차도-조셉 Dx), ATXN1 및 ATXN2(척수소뇌 실조), DMPK(근긴장 디스트로피), Atrophin-1 및 Atn1(DRPLA Dx), CBP(Creb-BP-전반적 불안정성), VLDLR(알츠하이머), Atxn7, Atxn10)
눈 질환 및 장애	연령-관련 황반 변성(Abcr, Ccl2, Cc2, cp(ceruloplasmin), Timp3, cathepsinD, Vldlr, Ccr2); 백내장(CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP40, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP40, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); 각막 혼탁 및 디스트로피(APOA1, TGFB1, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); 선천성 편평 각막(KERA, CNA2); 녹내장(MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); 레베르 선천 흑암시(CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); 황반 디스트로피(ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2)

[0214]

[0215]

달리 정의되지 않는 한, 본원에서 사용되는 모든 기술적 및/또는 과학적 용어는 본 발명이 관련되는 분야의 당업자에 의해 일반적으로 이해되는 바와 동일한 의미를 갖는다. 본원에서 기재되는 것들과 유사하거나 동등한 방법 및 물질이 본 발명의 구현예의 실시 또는 시험에서 사용될 수 있지만, 예시적인 방법 및/또는 물질이 후술된다. 상충 시, 정의를 포함하여, 특허 명세서가 우선할 것이다. 또한, 물질, 방법, 및 실시에는 단지 예시적인 것이며, 반드시 제한하려는 것이 아니다.

[0216]

달리 언급되지 않는 한 논의에서, 본 발명의 하나의 구현예의 특성 또는 특성들의 특징적 조건 또는 관계를 수식하는 "실질적으로" 및 "약"과 같은 형용사는 조건 또는 특징이 의도되는 적용을 위한 구현예의 작동을 위해 허용 가능한 관용성 내에서 정의됨을 의미하는 것으로 이해된다. 달리 나타내지 않는 한, 명세서 및 청구범위에서의 단어 "또는"은 배타적이기보다는 포괄적인 "또는"으로 간주되며, 이것이 결합시키는 항목의 적어도 하나 및 임의의 조합을 시사한다.

[0217]

상기 및 본원에서 다른 곳에서 사용되는 바와 같은 용어 하나의("a" 및 "an")는 "하나 이상의" 열거되는 성분을 나타내는 것으로 이해되어야 한다. 당업자에게는 구체적으로 달리 언급되지 않는 한 단수의 사용에 복수가 포함됨이 명확할 것이다. 따라서, 용어 하나("a," "an") 및 "적어도 하나"는 본 출원에서 상호 교환적으로 사용된다.

[0218]

본 발명의 교시를 더 잘 이해하고 달리 나타내지 않는 한, 어떠한 방식으로든 교시 범위를 제한하지 않는 목적으로, 양, 백분율 또는 비율을 표시하는 수치, 및 명세서 및 청구범위에서 사용되는 다른 수치 값은 모든 경우에서 용어 "약"에 의해 수식되는 것으로 이해되어야 한다. 따라서, 반대로 나타내지 않는 한, 하기 명세서 및 첨부되는 청구범위에서 나타내는 수치 파라미터는 수득하고자 하는 요망되는 특성에 따라 변할 수 있는 근사치

이다. 최소한, 각각의 수치 파라미터는 적어도 보고되는 유효 숫자의 수치의 견지에서 그리고 일반적인 반올림 기법을 적용하여 고려되어야 한다.

[0219] 수치 범위가 본원에서 인용되는 경우, 본 발명은 달리 언급되지 않는 한, 상한 및 하한 사이 및 이를 포함하는 각각의 정수를 고려하는 것으로 이해된다.

[0220] 본 출원의 기재 및 청구범위에서, 각각의 동사 "포함한다", "포함된다" 및 "갖는다" 그리고 이의 접합부는 동사의 목적어 또는 목적어들이 반드시 동사의 목적어 또는 목적어들의 성분, 요소 또는 부분의 전체 기재가 아님을 나타내기 위해 사용된다. 본원에서 사용된 바와 같은 다른 용어는 이의 당분야에서 널리 알려진 의미에 의해 정의되는 것임을 의미한다.

[0221] 용어 "폴리뉴클레오티드", "뉴클레오티드", "뉴클레오티드 서열", "핵산" 및 "올리고뉴클레오티드"는 상호 교환적으로 사용된다. 이들은 데옥시리보뉴클레오티드 또는 리보뉴클레오티드, 또는 이의 유사체인, 임의의 길이의 뉴클레오티드의 중합체 형태를 나타낸다. 폴리뉴클레오티드는 임의의 3차원 구조를 가질 수 있고, 알려지지거나 알려지지 않은, 임의의 기능을 수행할 수 있다. 하기는 폴리뉴클레오티드의 비제한적 예이다: 유전자 또는 유전자 단편의 코딩 또는 비-코딩 영역, 결합 분석으로부터 정의된 유전자위들(유전자위), 엑손, 인 아이언(in Iron), 메신저 RNA(mRNA), 전달 RNA, 리보솜 RNA, 짧은 간섭 RNA(siRNA), 짧은-헤어핀 RNA(shRNA), 마이크로-RNA(miRNA), 리보자임, cDNA, 재조합 폴리뉴클레오티드, 분기형 폴리뉴클레오티드, 플라스미드, 벡터, 임의의 서열의 단리된 DNA, 임의의 서열의 단리된 RNA, 핵산 탐침, 및 프라이머. 폴리뉴클레오티드는 메틸화된 뉴클레오티드 및 뉴클레오티드 유사체와 같은, 하나 이상의 변형된 뉴클레오티드를 포함할 수 있다. 존재하는 경우, 뉴클레오티드 구조에 대한 변형은 중합체의 어셈블리 전에 또는 후에 부여될 수 있다. 뉴클레오티드의 서열은 비-뉴클레오티드 성분에 의해 단속될 수 있다. 폴리뉴클레오티드는 표지화 성분과의 접합에 의한 것과 같이, 중합 후 추가 변형될 수 있다.

[0222] 용어 "뉴클레오티드 유사체" 또는 "변형된 뉴클레오티드"는 뉴클레오시드(예로, 시토신(C), 티민(T) 또는 우라실(U), 아데닌(A) 또는 구아닌(G))의 질소성 염기 내에 또는 상에, 뉴클레오시드의 당 모이어티(예로, 리보스, 데옥시리보스, 변형된 리보스, 변형된 데옥시리보스, 6'-원 당 유사체, 또는 개방쇄 당 유사체), 또는 포스페이트 내에 또는 상에, 하나 이상의 화학적 변형(예로, 치환)을 함유하는 뉴클레오티드를 나타낸다. 본원에서 기재되는 각각의 RNA 서열은 하나 이상의 뉴클레오티드 유사체를 포함할 수 있다.

[0223] 본원에서 사용된 바와 같이, 하기 뉴클레오티드 식별자가 언급되는 뉴클레오티드 염기(들)를 나타내기 위해 사용된다:

뉴클레오티드 언급	나타내는 염기(들)			
A	A			
C		C		
G			G	
T				T
W	A			T
S		C	G	
M	A	C		
K			G	T
R	A		G	
Y		C		T
B		C	G	T
D	A		G	T
H	A	C		T
V	A	C	G	
N	A	C	G	T

[0224]

[0225] 본원에서 사용된 바와 같이, 용어 "표적화 서열" 또는 "표적화 분자"는 특이적 표적 서열에 혼성화할 수 있는 뉴클레오티드 서열을 포함하는 뉴클레오티드 서열 또는 분자를 나타내며, 예로, 표적화 서열은 표적화 서열의 길이에 따라 표적화되는 서열과 적어도 부분적으로 상보적인 뉴클레오티드 서열을 갖는다. 표적화 서열 또는 표적화 분자는 CRISPR 복합체의 표적화 부분으로 작용하는 표적화 서열을 갖는 CRISPR 뉴클레아제와 복합체를 형성할 수 있는 표적화 RNA 분자의 일부일 수 있다. 표적화 서열을 갖는 분자가 CRISPR 분자와 동시시간대에 존재하

는 경우, RNA 분자는 특정 표적 서열로 CRISPR 뉴클레아제를 표적화할 수 있다. 각각의 가능성은 별도 구현예를 나타낸다. 표적화 RNA 분자는 임의의 요망되는 서열을 표적화하도록 맞춤 설계될 수 있다.

[0226] 본원에서 사용되는 용어 "표적화하다"는 표적화되는 뉴클레오티드 서열을 갖는 핵산과 표적화 서열 또는 표적화 분자의 우선적 혼성화를 나타낸다. 용어 "표적화하다"는 표적화되는 뉴클레오티드 서열을 갖는 핵산의 우선적 표적화가 존재하지만, 온-표적 혼성화에 부가하여 비의도적인 오프-표적 혼성화가 또한 일어날 수 있는, 가변적인 혼성화 효율을 포괄함이 이해된다. RNA 분자가 서열을 표적화하는 경우, RNA 분자 및 CRISPR 뉴클레아제 분자의 복합체가 뉴클레아제 활성화에 대한 서열을 표적화함이 이해된다.

[0227] 복수의 세포에 존재하는 DNA 서열을 표적화하는 맥락에서, 표적화는 하나 이상의 세포 내 서열과 RNA 분자의 가이드 서열 부분의 혼성화를 포괄하며, 또한 복수의 세포에서 모든 세포보다 적은 세포 내 표적 서열과 RNA 분자의 혼성화를 포괄하는 것으로 이해된다. 따라서, RNA 분자가 복수의 세포 내 서열을 표적화하는 경우, RNA 분자 및 CRISPR 뉴클레아제의 복합체는 하나 이상의 세포 내 표적 서열과 혼성화하는 것으로 이해되며, 또한 모든 세포보다 적은 세포 내 표적 서열과 혼성화할 수 있는 것으로 이해된다. 따라서, RNA 분자 및 CRISPR 뉴클레아제의 복합체는 하나 이상의 세포 내 표적 서열과의 혼성화에 관해 이중 가닥 절단을 도입하며 또한 모든 세포보다 적은 세포 내 표적 서열과의 혼성화에 관해 이중 가닥 절단을 도입할 수 있는 것으로 이해된다. 본원에서 사용되는 용어 "변형된 세포"는 이중 가닥 절단이 표적 서열과의 혼성화, 즉 온-표적 혼성화의 결과로서 RNA 분자 및 CRISPR 뉴클레아제의 복합체에 의해 영향받는 세포를 나타낸다.

[0228] 본원에서 사용되는 용어 "야생형"은 당업자에 의해 이해되는 당분야의 용어이며 이것이 돌연변이체 또는 변이체 형태와 구별되어 자연에서 발생하는 바와 같은 유기체, 균주, 유전자 또는 특징의 전형적 형태를 의미한다. 따라서, 본원에서 사용된 바와 같이, 아미노산 또는 뉴클레오티드의 서열이 야생형 서열을 나타내는 경우, 변이체는, 예로, 치환, 결실, 삽입을 포함하는, 그 서열의 변이체를 나타낸다. 본 발명의 구현예에서, 조작된 CRISPR 뉴클레아제는 표 1에 나타난 임의의 CRISPR 뉴클레아제의 CRISPR 뉴클레아제에 비해 적어도 하나의 아미노산 변형(예로, 치환, 결실, 및/또는 삽입)을 포함하는 변이체 CRISPR 뉴클레아제이다.

[0229] 용어 "비-자연 발생" 또는 "조작된"은 상호 교환적으로 사용되며 인간 조작을 시사한다. 이 용어는 핵산 분자 또는 폴리펩티드를 나타내는 경우, 핵산 분자 또는 폴리펩티드에 이들이 자연적으로 연관되고 자연에서 확인되는 바와 같은 적어도 하나의 다른 성분이 적어도 실질적으로 없음을 의미할 수 있다.

[0230] 본원에서 사용되는 용어 "아미노산"에는 글리신 및 D 또는 L 둘 다의, 광학 이성질체, 및 아미노산 유사체 및 펩티드모사체를 포함하는 천연 및/또는 비천연 또는 합성 아미노산이 포함된다.

[0231] 본원에서 사용된 바와 같이, "게놈 DNA"는 관심 세포 또는 세포들에 존재하는 선형 및/또는 염색체 DNA 및/또는 플라스미드 또는 다른 염색체외 DNA 서열을 나타낸다. 일부 구현예에서, 관심 세포는 진핵 세포이다. 일부 구현예에서, 관심 세포는 원핵 세포이다. 일부 구현예에서, 방법은 게놈 DNA 서열 내의 사전-결정된 표적 부위에서 이중-가닥 절단(DSB)을 생성하여, 게놈 내 표적 부위(들)에서 DNA 서열의 돌연변이, 삽입, 및/또는 결실을 초래한다.

[0232] "진핵" 세포에는 비제한적으로 진균 세포(예컨대 효모), 식물 세포, 동물 세포, 포유류 세포 및 인간 세포가 포함된다.

[0233] 본원에서 사용되는 용어 "뉴클레아제"는 핵산의 뉴클레오티드 서브유닛 간 포스포디에스테르 결합을 절단할 수 있는 효소를 나타낸다. 뉴클레아제는 천연 원천으로부터 단리되거나 유래될 수 있다. 천연 원천은 임의의 살아 있는 유기체일 수 있다. 대안적으로, 뉴클레아제는 포스포디에스테르 결합 절단 활성을 보유하는 변형된 또는 합성 단백질일 수 있다.

[0234] 본원에서 사용되는 용어 "PAM"은 표적화된 DNA 서열의 인근에 위치하며 CRISPR 뉴클레아제에 의해 인식되는 표적 DNA의 뉴클레오티드 서열을 나타낸다. PAM 서열은 뉴클레아제 정체에 따라 상이할 수 있다.

[0235] 본원에서 사용되는 용어 "돌연변이 장애" 또는 "돌연변이 질환"은 돌연변이에 의해 유도되는 유전자의 기능이상과 관련되는 임의의 장애 또는 질환을 나타낸다. 돌연변이 장애로 발현되는 기능이상 유전자는 그 대립유전자 중 적어도 하나에 돌연변이를 함유하며 "질환-연관 유전자"로 나타낸다. 돌연변이는 질환-연관 유전자의 임의의 부분에, 예를 들어, 조절, 코딩, 또는 비-코딩 부분에 있을 수 있다. 돌연변이는 치환, 삽입, 또는 결실과 같은 임의의 클래스의 돌연변이일 수 있다. 질환-연관 유전자의 돌연변이는 열성, 우성 음성, 기능-획득, 기능-상실, 또는 유전자 산물의 홀배수체부족을 야기하는 돌연변이와 같은 임의의 유형의 돌연변이의 기전에 따라 장애 또

는 질환으로 발현될 수 있다.

- [0236] 당업자는 본 발명의 구현예가 프로토스페이서 인접 모티프(PAM) 옆의 관심 표적 게놈 DNA 서열과 연합하기 위해 서와 같이, 뉴클레아제, 예로 CRISPR 뉴클레아제와 복합체화할 수 있는 RNA 분자를 개시함을 이해할 것이다. 이어서 뉴클레아제는 표적 DNA의 절단을 매개하여 프로토스페이서 내에서 이중-가닥 절단을 생성한다.
- [0237] 본 발명의 구현예에서, CRISPR 뉴클레아제 및 표적화 분자는 표적 DNA 서열에 결합하여 표적 DNA 서열의 절단을 실시하는 CRISPR 복합체를 형성한다. CRISPR 뉴클레아제는 추가적인, 별도의 tracrRNA 분자 없이, CRISPR 뉴클레아제 및 RNA 분자를 포함하는 CRISPR 복합체를 형성할 수 있다. 대안적으로, CRISPR 뉴클레아제는 CRISPR 뉴클레아제, RNA 분자, 및 tracrRNA 분자 간에 CRISPR 복합체를 형성할 수 있다.
- [0238] 용어 "단백질 결합 서열" 또는 "뉴클레아제 결합 서열"은 CRISPR 뉴클레아제와 결합하여 CRISPR 복합체를 형성할 수 있는 서열을 나타낸다. 당업자는 CRISPR 뉴클레아제와 결합하여 CRISPR 복합체를 형성할 수 있는 tracrRNA가 단백질 또는 뉴클레아제 결합 서열을 포함함을 이해할 것이다.
- [0239] CRISPR 뉴클레아제의 "RNA 결합 부분"은 RNA 분자에 결합하여 CRISPR 복합체를 형성할 수 있는 CRISPR 뉴클레아제의 부분, 예로 tracrRNA 분자의 뉴클레아제 결합 서열을 나타낸다. CRISPR 뉴클레아제의 "활성 부분" 또는 "활성이 있는 부분"은 DNA 분자에서, 예를 들어 DNA-표적화 RNA 분자와의 복합체로 있는 경우, 이중 가닥 절단을 실시하는 CRISPR 뉴클레아제의 부분을 나타낸다.
- [0240] RNA 분자는 염기쌍 형성을 통해 tracrRNA와 혼성화하고 CRISPR 복합체의 형성을 촉진하기 위해 tracrRNA 분자와 충분히 상보적인 서열을 포함할 수 있다(U.S. 특허 번호 8,906,616 참고). 본 발명의 구현예에서, RNA 분자는 tracr 메이트(mate) 서열을 갖는 부분을 추가로 포함할 수 있다.
- [0241] 본 발명의 구현예에서, 표적화 분자는 tracrRNA 분자의 서열을 추가로 포함할 수 있다. 이러한 구현예는 함께 단일 가이드 RNA(sgRNA)를 형성하는, RNA 분자의 가이드 부분(gRNA 또는 crRNA) 및 트랜스-활성화 crRNA(tracrRNA)의 합성 융합물로서 설계될 수 있다(Jinek 등, Science (2012) 참고). 본 발명의 구현예는 또한 별도의 tracrRNA 분자 및 가이드 서열 부분을 포함하는 별도의 RNA 분자를 이용하여 CRISPR 복합체를 형성할 수 있다. 이러한 구현예에서 tracrRNA 분자는 염기쌍 형성을 통해 RNA 분자와 혼성화할 수 있고 본원에서 기재되는 본 발명의 특정 적용에서 유리할 수 있다.
- [0242] 본 발명의 구현예에서 RNA 분자는 RNA 분자의 구조를 추가 정의할 수 있는 "교통반점(nexus)" 영역 및/또는 "헤어핀" 영역을 포함할 수 있다(Briner 등, Molecular Cell (2014) 참고).
- [0243] 본원에서 사용되는 용어 "직접 반복 서열"은 뉴클레오티드 서열의 특정 아미노산 서열의 2개 이상의 반복을 나타낸다.
- [0244] 본원에서 사용된 바와 같이, CRISPR 뉴클레아제"와 상호작용" 또는 "에 결합"할 수 있는 RNA 서열 또는 분자는 RNA 서열 또는 분자가 CRISPR 뉴클레아제와 CRISPR 복합체를 형성하는 능력을 나타낸다.
- [0245] 본원에서 사용되는 용어 "작동 가능하게 연결된"은 이의 의도되는 방식으로 기능하도록 허용하는 2개의 서열 또는 분자 간 관계(즉, 융합, 혼성화)를 나타낸다. 본 발명의 구현예에서, RNA 분자가 프로모터에 작동 가능하게 연결되는 경우, RNA 분자 및 프로모터는 둘 다 이의 의도되는 방식으로 기능하는 것이 허용된다.
- [0246] 본원에서 사용되는 용어 "이중성 프로모터"는 촉진되는 분자 또는 경로와 함께 자연 발생하지 않는 프로모터를 나타낸다.
- [0247] 본원에서 사용된 바와 같이, 서열 또는 분자는 분자의 서열 간 X%의 염기 또는 아미노산이 동일하고 동일한 상대 위치에 있는 경우, 또 다른 서열 또는 분자와 X% "서열 동일성"을 갖는다. 예를 들어, 제2 뉴클레오티드 서열과 적어도 95% 서열 동일성을 갖는 제1 뉴클레오티드 서열은 동일한 상대 위치에서, 다른 서열과 동일한, 적어도 95%의 염기를 가질 것이다.
- [0248] 핵 편재 서열
- [0249] 용어 "핵 편재 서열" 및 "NLS"는 이것이 핵막 장벽을 통해 세포의 세포질로부터 연관되는 단백질의 수송을 유도하는 아미노산 서열/펩티드를 나타내기 위해 상호 교환적으로 사용된다. 용어 "NLS"는 특정 펩티드의 핵 편재 서열뿐만 아니라 핵막 장벽을 통해 세포질 폴리펩티드의 전위를 유도할 수 있는 이의 유도체를 포괄하려는 것이다. NLS는 폴리펩티드의 N-말단, C-말단, 또는 N- 및 C-말단 둘 다에 부착되는 경우, 폴리펩티드의 핵 전위를 유도할 수 있다. 또한, 폴리펩티드의 아미노산 서열을 따라 무작위 배치된 아미노산 측쇄에 대해 그 N- 또는 C-

말단에 의해 커플링된 NLS를 갖는 폴리펩티드가 전위될 것이다. 전형적으로, NLS는 단백질 표면 상에 노출된 양으로 하전된 라이신 또는 아르기닌의 하나 이상의 짧은 서열로 구성되지만, 다른 유형의 NLS도 알려져 있다. NLS의 비제한적 예에는 하기로부터 유래되는 NLS 서열이 포함된다: SV40 바이러스 큰 T-항원, 핵플라스민(nucleoplasmin), c-myc, hRNP1 M9 NLS, 임포틴(importin)-알파로부터의 IBB 도메인, 근종 T 단백질, 인간 p53, 마우스 c-abl IV, 인플루엔자 vims NS1, 간염 바이러스 델타 항원, 마우스 Mx1 단백질, 인간 폴리(ADP-리보스) 폴리머라제, 및 스테로이드 호르몬 수용체 (인간)글루코코르티코이드. 이러한 NLS 서열은 서열번호: 69~84로 기재된다.

[0250] 전달

[0251] 본원에서 기재되는 CRISPR 뉴클레아제 또는 CRISPR 조성물은 단백질, DNA 분자, RNA 분자, 리보핵단백질(RNP), 핵산 벡터, 또는 이의 임의의 조합으로 전달될 수 있다. 일부 구현예에서, RNA 분자는 화학적 변형을 포함한다. 적합한 화학적 변형의 비제한적 예에는 2'-O-메틸(M), 2'-O-에틸, 3' 포스포로티오에이트(PS) 또는 2'-O-메틸, 3' 티오PACE(MSP), 슈도유리딘, 및 1-메틸 슈도-유리딘이 포함된다. 각각의 가능성은 본 발명의 별도 구현예를 나타낸다.

[0252] 본원에서 기재되는 CRISPR 뉴클레아제 및/또는 이를 인코딩하는 폴리뉴클레오티드, 및 선택적으로 추가적인 단백질(예로, ZFP, TALEN, 전사 인자, 제한 효소) 및/또는 가이드 RNA와 같은 뉴클레오티드 분자는 임의의 적합한 수단에 의해 표적 세포로 전달될 수 있다. 표적 세포는 임의의 환경에서의, 예로 단리되거나 단리되지 않은, 배양 중에 유지된, 시험관내, 생체외, 생체내 또는 식물내, 임의의 유형의 세포, 예로, 진핵 또는 원핵 세포일 수 있다.

[0253] 일부 구현예에서, 전달될 조성물에는 뉴클레아제의 mRNA 및 가이드의 RNA가 포함된다. 일부 구현예에서, 전달될 조성물에는 뉴클레아제의 mRNA, 가이드의 RNA 및 공여체 주형이 포함된다. 일부 구현예에서, 전달될 조성물에는 CRISPR 뉴클레아제 및 가이드 RNA가 포함된다. 일부 구현예에서, 전달될 조성물에는 CRISPR 뉴클레아제, 가이드 RNA 및 예를 들어, 상동성 유도 복구를 통한 유전자 편집을 위한 공여체 주형이 포함된다. 일부 구현예에서, 전달될 조성물에는 뉴클레아제의 mRNA, DNA-표적화 RNA 및 tracrRNA가 포함된다. 일부 구현예에서, 전달될 조성물에는 뉴클레아제의 mRNA, DNA-표적화 RNA 및 tracrRNA 그리고 공여체 주형이 포함된다. 일부 구현예에서, 전달될 조성물에는 CRISPR 뉴클레아제 DNA-표적화 RNA 및 tracrRNA가 포함된다. 일부 구현예에서, 전달될 조성물에는 CRISPR 뉴클레아제, DNA-표적화 RNA 및 tracrRNA 그리고 예를 들어, 상동성 유도 복구를 통한 유전자 편집을 위한 공여체 주형이 포함된다.

[0254] 임의의 적합한 바이러스 벡터 시스템이 RNA 조성물을 전달하기 위해 사용될 수 있다. 통상적인 바이러스 및 비-바이러스 기반 유전자 전달 방법이 세포(예로, 포유류 세포, 식물 세포 등) 및 표적 조직에 핵산 및/또는 CRISPR 뉴클레아제를 도입하기 위해 사용될 수 있다. 이러한 방법은 또한 시험관내 세포에 인코딩 핵산 및/또는 CRISPR 뉴클레아제 단백질을 투여하기 위해 사용될 수 있다. 특정 구현예에서, 핵산 및/또는 CRISPR 뉴클레아제는 생체내 또는 생체외 유전자 치료법 용도를 위해 투여된다. 비-바이러스 벡터 전달 시스템에는 네이키드 핵산, 및 리포솜 또는 폴록사머와 같은 전달 비히클과 복합체화된 핵산이 포함된다. 유전자 치료법 절차의 검토를 위해, 문헌(Anderson, Science (1992); Nabel and Felgner, TIBTECH (1993); Mitani and Caskey, TIBTECH (1993); Dillon, TIBTECH (1993); Miller, Nature (1992); Van Brunt, Biotechnology (1988); Vigne 등, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer and Perricaudet, British Medical Bulletin (1995); Haddada 등, Current Topics in Microbiology and Immunology (1995); 및 Yu 등, Gene Therapy 1:13-26 (1994))을 참고한다.

[0255] 핵산 및/또는 단백질의 비-바이러스 전달 방법에는 전기천공, 지질감염, 마이크로주입, 바이오리스틱(biolistics), 입자 총 가속화, 바이로솜, 리포솜, 면역리포솜, 다중양이온 또는 지질:핵산 콘주게이트, 인공 비리온, 및 핵산의 제제-증강 흡수가 포함되거나 박테리아 또는 바이러스(예로, 아그로박테리움(Agrobacterium), 리조비움 종(Rhizobium sp.) NGR234, 시노리조보이움멜리로티(Sinorhizobiummeliloti), 메조리조비움 로티(Mesorhizobium loti), 담배 모자이크 바이러스, 감자 바이러스 X, 콜리플라워 모자이크 바이러스 및 카사바 배인 모자이크 바이러스에 의해 식물 세포로 전달될 수 있다. 예로, 문헌(Chung 등 Trends Plant Sci. (2006))을 참고한다. 예로, Sonitron 2000 시스템(Rich-Mar)을 사용하는 초음파천공이 또한 핵산의 전달을 위해 사용될 수 있다. 단백질 및/또는 핵산의 양이온성-지질 매개 전달이 또한 생체내 또는 시험관내 전달 방법으로서 고려된다. 문헌(Zuris 등, Nat. Biotechnol. (2015), Coelho 등, N. Engl. J. Med. (2013); Judge 등, Mol. Ther. (2006); 및 Basha 등, Mol. Ther. (2011))을 참고한다.

- [0256] 추가적인 예시적 핵산 전달 시스템에는 Amaxa® Biosystems(Cologne, Germany), Maxcyte, Inc.(Rockville, Md.), BTX Molecular Delivery Systems(Holliston, Mass.) 및 Copernicus Therapeutics Inc.,(예를 들어 U.S. 특허 번호 6,008,336 참고)에 의해 제공되는 것들이 포함된다. 지질감염은 예로, U.S. 특허 번호 5,049,386, U.S. 특허 번호 4,946,787; 및 U.S. 특허 번호 4,897,355)에 기재되어 있으며 지질감염 시약은 상업적으로 판매된다(예로, Transfectam.TM., Lipofectin.TM. 및 Lipofectamine.TM. RNAiMAX). 폴리뉴클레오티드의 효율적인 수용체-인식 지질감염에 적합한 양이온성 및 중성 지질에는 PCT 국제 공개 번호 WO/1991/017424 및 WO/1991/016024에 개시된 것들이 포함된다. 전달은 세포(생체외 투여) 또는 표적 조직(생체내 투여)에 대한 것일 수 있다.
- [0257] 면역지질 복합체와 같은 표적화된 리포솜을 포함하는, 지질:핵산 복합체의 제조는 당업자에게 잘 알려져 있다(예로, Crystal, Science (1995); Blaese 등, Cancer Gene Ther. (1995); Behr 등, Bioconjugate Chem. (1994); Remy 등, Bioconjugate Chem. (1994); Gao and Huang, Gene Therapy (1995); Ahmad and Allen, Cancer Res., (1992); U.S. 특허 번호 4,186,183; 4,217,344; 4,235,871; 4,261,975; 4,485,054; 4,501,728; 4,774,085; 4,837,028; 및 4,946,787 참고).
- [0258] 추가적인 전달 방법에는 EnGeneIC 전달 비히클(EDV) 내로 전달될 핵산 패키징의 사용이 포함된다. 이들 EDV는 이중특이적 항체를 사용하여 표적 조직으로 특이적으로 전달되며 여기서 항체의 하나의 아암(arm)은 표적 조직에 대한 특이성을 가지며 다른 아암은 EDV에 대한 특이성을 갖는다. 항체는 EDV를 표적 세포 표면으로 가져간 후 EDV가 세포내이입에 의해 세포 내로 들어간다. 일단 세포 안에 들어가면, 내용물이 방출된다(MacDiamid 등, Nature Biotechnology (2009) 참고).
- [0259] 핵산의 전달을 위한 RNA 또는 DNA 바이러스 기반 시스템의 사용은 신체내 특정 세포로의 바이러스 표적화 및 핵으로의 바이러스 페이로드 수송을 위해 고도로 진화된 공정의 장점을 취한다. 바이러스 벡터는 환자에 직접 투여될 수 있거나(생체내) 이들은 시험관내 세포를 처리하기 위해 사용될 수 있고 변형된 세포가 환자에 투여된다(생체외). 핵산의 전달을 위한 통상적인 바이러스 기반 시스템에는 비제한적으로 유전자 전달을 위한 재조합 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-연관, 백시니아 및 단순 헤르페스 바이러스 벡터가 포함된다. 그러나, RNA 바이러스가 본원에서 기재되는 RNA 조성물의 전달을 위해 바람직하다. 추가적으로, 높은 형질도입 효율이 여러 상이한 세포 유형 및 표적 조직에서 관찰되었다. 본 발명의 핵산은 비-통합 렌티바이러스에 의해 전달될 수 있다. 선택적으로, 렌티바이러스를 이용한 RNA 전달이 이용된다. 선택적으로 렌티바이러스에는 뉴클레아제의 mRNA, 가이드의 RNA가 포함된다. 선택적으로 렌티바이러스에는 뉴클레아제의 mRNA, 가이드의 RNA 및 공여체 주형이 포함된다. 선택적으로, 렌티바이러스에는 뉴클레아제 단백질, 가이드 RNA가 포함된다. 선택적으로, 렌티바이러스에는 뉴클레아제 단백질, 가이드 RNA 및/또는 예를 들어, 상동성 유도 복구를 통한 유전자 편집을 위한 공여체 주형이 포함된다. 선택적으로 렌티바이러스에는 뉴클레아제의 mRNA, DNA-표적화 RNA, 및 tracrRNA가 포함된다. 선택적으로 렌티바이러스에는 뉴클레아제의 mRNA, DNA-표적화 RNA, 및 tracrRNA, 그리고 공여체 주형이 포함된다. 선택적으로, 렌티바이러스에는 뉴클레아제 단백질, DNA-표적화 RNA, 및 tracrRNA가 포함된다. 선택적으로, 렌티바이러스에는 뉴클레아제 단백질, DNA-표적화 RNA, 및 tracrRNA, 그리고 예를 들어, 상동성 유도 복구를 통한 유전자 편집을 위한 공여체 주형이 포함된다.
- [0260] 상기 언급된 바와 같이, 본원에서 기재되는 조성물은 비-통합 렌티바이러스 입자 방법, 예로 LentiFlash® 시스템을 사용하여 표적 세포로 전달될 수 있다. 이러한 방법은 표적 세포로의 RNA의 전달이 표적 세포 내부에서 본원에서 기재되는 조성물의 어셈블리를 초래하도록, 표적 세포 내로 mRNA 또는 다른 유형의 RNA를 전달하기 위해 사용될 수 있다. 또한 PCT 국제 공개 번호 WO2013/014537, WO2014/016690, WO2016185125, WO2017194902, 및 WO2017194903을 참고한다.
- [0261] 레트로바이러스의 친화성은 외래 외피 단백질을 포함시킴으로써 변경되어, 표적 세포의 잠재적 표적 집단을 확장시킬 수 있다. 렌티바이러스 벡터는 비-분열 세포를 형질도입시키거나 감염시킬 수 있는 레트로바이러스 벡터이며 전형적으로 높은 바이러스 역가를 생산한다. 레트로바이러스 유전자 전달 시스템의 선택은 표적 조직에 의존한다. 레트로바이러스 벡터는 최대 6~10 kb의 외래 서열에 대한 패키징능을 갖는 시스-작용 장형 말단 반복체로 이루어진다. 최소 시스-작용 LTR이 벡터의 복제 및 패키징을 위해 충분하며, 이어서 이는 표적 세포 내로 치료 유전자를 통합시켜 영구적인 트랜스유전자 발현을 제공하기 위해 사용된다. 널리 사용되는 레트로바이러스 벡터에는 쥘과 백혈병 바이러스(MuLV), 긴팔 원숭이 백혈병 바이러스(GaLV), 유인원 면역결핍 바이러스(SIV), 인간 면역결핍 바이러스(HIV), 및 이의 조합에 기반한 것들이 포함된다(예로, Buchscher Panganiban, J. Virol. (1992); Johann 등, J. Virol. (1992); Sommerfelt 등, Virol. (1990); Wilson 등, J. Virol. (1989);

Miller 등, J. Virol. (1991); PCT 국제 공개 번호 WO/1994/026877A1 참고).

- [0262] 적어도 6개의 바이러스 벡터 접근이 현재 임상 시험에서 유전자 전달을 위해 이용 가능하며, 이는 형질도입 재제를 생성하기 위해 헬퍼 세포주 내로 삽입된 유전자에 의한 결합 벡터의 보완이 연루되는 접근을 이용한다.
- [0263] pLASN 및 MFG-S는 임상 시험에서 사용된 레트로바이러스 벡터의 예들이다(Dunbar 등, Blood (1995); Kohn 등, Nat. Med. (1995); Malech 등, PNAS (1997)). PA317/pLASN은 유전자 치료법 시험에서 사용된 최초의 치료 벡터였다(Blaese 등, Science (1995)). 50% 이상의 형질도입 효율이 MFG-S 패키징된 벡터에서 관찰되었다(Ellem 등, Immunol Immunother. (1997); Dranoff 등, Hum. Gene Ther. (1997)).
- [0264] 패키징 세포는 숙주 세포를 감염시킬 수 있는 바이러스 입자를 형성하기 위해 사용된다. 이러한 세포에는 아데노바이러스, AAV를 패키징하는 293 세포, 및 레트로바이러스를 패키징하는 psi.2 세포 또는 PA317 세포가 포함된다. 유전자 치료법에서 사용되는 바이러스 벡터는 보통 바이러스 입자 내로 핵산 벡터를 패키징하는 생산체 세포주에 의해 생성된다. 벡터는 전형적으로 패키징 및 숙주 내로의 후속 통합(적용 가능한 경우)을 위해 요구되는 최소 바이러스 서열을 함유하며, 다른 바이러스 서열은 발현될 단백질을 인코딩하는 발현 카세트에 의해 대체된다. 누락된 바이러스 기능은 패키징 세포주에 의해 트랜스로 공급된다. 예를 들어, 유전자 치료법에서 사용되는 AAV 벡터는 전형적으로 패키징 및 숙주 게놈 내로의 통합을 위해 요구되는 AAV 게놈으로부터의 역방위 말단 반복체(ITR) 서열만을 보유한다. 바이러스 DNA는 세포주에서 패키징되며, 이는 다른 AAV 유전자, 즉 rep 및 cap을 인코딩하지만, ITR 서열은 없는 헬퍼 플라스미드를 함유한다. 세포주는 또한 헬퍼로서 아데노바이러스로 감염된다. 헬퍼 바이러스는 헬퍼 플라스미드로부터 AAV 벡터의 복제 및 AAV 유전자의 발현을 촉진한다. 헬퍼 플라스미드는 ITR 서열의 부재로 인해 유의미한 양으로 패키징되지 않는다. 아데노바이러스로의 오염은, 예로, 아데노바이러스가 AAV보다 더 민감한 열 처리에 의해 감소될 수 있다. 추가적으로, AAV는 배칼로바이러스 시스템을 사용하여 임상 규모로 생산될 수 있다(U.S. 특허 번호 7,479,554 참고).
- [0265] 여러 유전자 치료법 적용에서, 유전자 치료법 벡터는 특정 조직 유형으로 고도의 특이성을 가지고 전달되는 것이 요망된다. 따라서, 바이러스 벡터는 리간드를 바이러스의 외부 표면 상의 바이러스 피막 단백질과의 융합 단백질로 발현함으로써 주어진 세포 유형에 대해 특이성을 갖도록 변형될 수 있다. 리간드는 관심 세포 유형의 세포 상에 존재하는 것으로 알려진 수용체에 대해 친화도를 갖도록 선택된다. 예를 들어, 문헌(Han 등, Proc. Natl. Acad. Sci. USA (1995))에서는 몰로니 췌장 백혈병 바이러스가 gp70에 융합된 인간 heregulin을 발현하도록 변형될 수 있고, 재조합 바이러스가 인간 표피 성장 인자 수용체를 발현하는 특정 인간 유방암 세포를 감염시킨다는 것을 보고하였다. 상기 원리는 다른 바이러스-표적 세포 페어로 연장될 수 있고, 여기서 표적 세포는 수용체를 발현하고 바이러스는 세포-표면 수용체에 대한 리간드를 포함하는 융합 단백질을 발현한다. 예를 들어, 필라멘트성 파지는 실질적으로 모든 선택된 세포성 수용체에 대해 특이적 결합 친화도를 갖는 항체 단편(예로, FAB 또는 Fv)을 디스플레이하도록 조작될 수 있다. 상기 기재는 주로 바이러스 벡터에 적용되지만, 동일한 원리가 비-바이러스 벡터에 적용될 수 있다. 이러한 벡터는 특정 표적 세포에 의한 섭취를 선호하는 특정 섭취 서열을 함유하도록 조작될 수 있다.
- [0266] 유전자 치료법 벡터는 후술되는 바와 같이, 전형적으로 전신 투여(예로, 정맥내, 복강내, 근육내, 진피하, 또는 두개내 주입) 또는 국소 적용에 의해, 개별 환자에 대한 투여에 의해 생체내 전달될 수 있다. 대안적으로, 벡터는 개별 환자로부터 외식된 세포(예로, 림프구, 골수 흡입물, 조직 생검)와 같은 생체의 세포 또는 범용 공여체 조혈 줄기 세포로 전달되고, 보통 벡터를 포함한 세포에 대한 선택 후에, 환자 내로의 세포의 재임플란트가 뒤따를 수 있다. 일부 구현예에서, 생체내 및 생체의 mRNA의 전달, 및 RNP 전달이 이용될 수 있다.
- [0267] 진단, 연구, 또는 유전자 치료법을 위한 생체의 세포 전달감염(예로, 숙주 유기체 내로의 전달감염된 세포의 재-주입을 통한)은 당업자에게 잘 알려져 있다. 바람직한 구현예에서, 세포는 대상 유기체로부터 단리되고, RNA 조성물로 전달감염되고, 대상 유기체(예로, 환자) 내로 다시 재-주입된다. 생체의 전달감염에 적합한 다양한 세포 유형은 당업자에게 잘 알려져 있다(예로, 환자로부터 세포를 어떻게 단리하고 배양하는지의 논의에 대해, Freshney, "Culture of Animal Cells, A Manual of Basic Technique and Specialized Applications (6th edition, 2010)) 및 여기서 인용된 참고문헌 참고).
- [0268] 적합한 세포에는 비제한적으로 진핵 및 원핵 세포 및/또는 세포주가 포함된다. 이러한 세포 또는 이러한 세포로부터 생성된 세포주의 비제한적 예에는 COS, CHO(예로, CHO--S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NSO, SP2/0-g14, HeLa, HEK293(예로, HEK293-F, HEK293-H, HEK293-T), 및 perC6 세포, 임의의 식물 세포(분화 또는 미분화)뿐만 아니라 스포도테라푸기페르다(Spodopterafugiperda(Sf))와 같은 곤충 세포, 또는 사카로마이세스(Saccharomyces), 피치아(Pichia) 및 스킨

조사카로마이세스(Schizosaccharomyces)와 같은 진균 세포가 포함된다. 특정 구현예에서, 세포주는 CHO-K1, MDCK 또는 HEK293 세포주이다. 추가적으로, 일차 세포가 단리되고 뉴클레아제(예로 ZFN 또는 TALEN) 또는 뉴클레아제 시스템(예로 CRISPR)으로의 처리 후 치료받을 대상체 내로의 재도입을 위해 생체의 사용될 수 있다. 적합한 일차 세포에는 말초혈 단핵 세포(PBMC), 및 비제한적으로 CD4+ T 세포 또는 CD8+ T 세포와 같은 다른 혈액 세포 하위세트가 포함된다. 적합한 세포에는 또한 예로서, 배아 줄기 세포, 유도된 다능성 줄기 세포, 조혈 줄기 세포(CD34+), 신경세포 줄기 세포 및 중간엽 줄기 세포와 같은 줄기 세포가 포함된다.

[0269] 하나의 구현예에서, 줄기 세포는 세포 전달감염 및 유전자 치료법을 위한 생체의 절차에서 사용된다. 줄기 세포를 사용하는 장점은 이들이 시험관내 다른 세포 유형으로 분화될 수 있거나 이들이 골수에 이식될 포유류(예컨대 세포의 공여체) 내로 도입될 수 있다는 것이다. 시험관내 CD34+ 세포를 GM-CSF, IFN-감마 및 TNF-알파와 같은 사이토카인을 사용하여 임상적으로 중요한 면역 세포 유형으로 분화시키는 방법이 알려져 있다(비제한적 예로서, Inaba 등, J. Exp. Med. (1992) 참고).

[0270] 줄기 세포는 알려진 방법을 사용하여 형질도입 및 분화를 위해 단리된다. 예를 들어, 줄기 세포는 CD4+ 및 CD8+(T 세포), CD45+(범 B 세포), GR-1(과립구), 및 Iad(분화 항원 제시 세포)와 같이, 원치 않는 세포에 결합하는 항체로 골수 세포를 팬닝함으로써 골수 세포로부터 단리된다(비제한적 예로서, Inaba 등, J. Exp. Med. (1992) 참고). 변형된 줄기 세포가 또한 일부 구현예에서 사용될 수 있다.

[0271] 특히, 본원에서 기재되는 CRISPR 뉴클레아제는 유사분열-후 세포 또는 활발히 분열 중이 아닌 임의의 세포, 예로, 정지된 세포에서 게놈 편집에 적합할 수 있다. 본 발명의 CRISPR 뉴클레아제를 사용하여 편집될 수 있는 유사분열-후 세포의 예에는 비제한적으로 근육세포, 심근세포, 간세포, 골세포 및 신경세포가 포함된다.

[0272] 치료 RNA 조성물을 함유하는 벡터(예로, 레트로바이러스, 리조좀 등)가 또한 생체내 세포의 형질도입을 위해 유기체에 직접 투여될 수 있다. 대안적으로, 네이키드 RNA 또는 mRNA가 투여될 수 있다. 투여는 비제한적으로 주사, 주입, 국소 적용 및 전기천공을 포함하는, 분자를 혈액 또는 조직 세포와의 궁극적 접촉으로 도입하기 위해 보통 사용되는 임의의 경로에 의한다. 이러한 핵산을 투여하는 적합한 방법이 이용 가능하며 당업자에게 잘 알려져 있고, 특정 조성물을 투여하기 위해 하나를 초과하는 경로가 사용될 수 있지만, 특정 경로가 종종 또 다른 경로보다 더 즉각적이고 더 효과적인 반응을 제공할 수 있다.

[0273] 면역 세포(예로, T-세포) 내로의 트랜스유전자의 도입에 적합한 벡터에는 비-통합 렌티바이러스 벡터가 포함된다. 예를 들어, U.S. 특허 공개 번호 2009/0117617을 참고한다.

[0274] 약학적으로 허용 가능한 담체는 부분적으로 투여되는 특정 조성물뿐만 아니라 조성물을 투여하기 위해 사용되는 특정 방법에 의해 결정된다. 따라서, 후술되는 바와 같이, 약학 조성물의 매우 다양한 적합한 제형물이 이용 가능하다(예로, Remington's Pharmaceutical Sciences, 17th ed., 1989 참고).

[0275] 상동 재조합에 의한 DNA 복구

[0276] 용어 "상동성-유도 복구" 또는 "HDR"은, 예를 들어, DNA에서 이중-가닥 및 단일-가닥 절단의 복구 동안, 세포 내 복구 DNA 손상의 복구를 위한 기전을 나타낸다. HDR은 뉴클레오타이드 서열 상동성을 필요로 하며 이중-가닥 또는 단일 절단이 일어난 서열(예로, DNA 표적 서열)을 복구하기 위해 "핵산 주형"을 사용한다(핵산 주형 또는 공여체 주형이 본원에서 상호 교환적으로 사용됨). 이는, 예를 들어, 핵산 주형으로부터 DNA 표적 서열로의 유전 정보의 전달을 초래한다. HDR은 핵산 주형 서열이 DNA 표적 서열과 상이하고 핵산 주형 폴리뉴클레오타이드 또는 올리고뉴클레오타이드의 일부 또는 전부가 DNA 표적 서열 내로 포함되는 경우, DNA 표적 서열의 변경(예로, 삽입, 결실, 돌연변이)을 초래할 수 있다. 일부 구현예에서, 전체 핵산 주형 폴리뉴클레오타이드, 핵산 주형 폴리뉴클레오타이드의 일부, 또는 핵산 주형의 카피가 DNA 표적 서열 부위에서 통합된다.

[0277] 용어 "핵산 주형" 및 "공여체"는 게놈 내로 삽입되거나 카피되는 뉴클레오타이드 서열을 나타낸다. 핵산 주형은 예로, 하나 이상의 뉴클레오타이드의, 뉴클레오타이드 서열을 포함하며, 이는 표적 핵산에서 주형에 부가될 것이거나 변화의 주형이 될 것이거나 표적 서열을 변형하기 위해 사용될 수 있다. 핵산 주형 서열은 임의의 길이, 예를 들어 2 내지 10,000개 뉴클레오타이드 길이(또는 그 사이 또는 초과)의 임의의 정수 값), 바람직하게는 약 100 내지 1,000개 뉴클레오타이드 길이(또는 그 사이의 임의의 정수), 보다 바람직하게는 약 200 내지 500개 뉴클레오타이드 길이일 수 있다. 핵산 주형은 단일 가닥 핵산, 이중 가닥 핵산일 수 있다. 일부 구현예에서, 핵산 주형은 예로, 하나 이상의 뉴클레오타이드의, 뉴클레오타이드 서열을 포함하며, 이는 표적 핵산, 예로, 표적 위치의 야생형 서열에 대응한다. 일부 구현예에서, 핵산 주형은 예로, 하나 이상의 리보뉴클레오타이드의, 리보뉴클레오타이드 서열을 포함하며, 이는 표적 핵산, 예로, 표적 위치의 야생형 서열에 대응한다. 일부 구현예에서, 핵산 주형은 변

형된 리보뉴클레오타이드를 포함한다.

- [0278] 예를 들어, 돌연변이체 유전자의 교정을 위한 또는 야생형 유전자의 증가된 발현을 위한, 외인성 서열("공여체 서열", 공여체 주형" 또는 "공여체"로도 불림)의 삽입이 또한 수행될 수 있다. 공여체 서열이 전형적으로 이것이 배치되는 게놈 서열과 동일하지 않음은 용이하게 자명할 것이다. 공여체 서열은 관심 위치에서 효율적인 HDR을 허용하기 위해 상동성인 2개 영역이 측면에 있는 비-상동 영역을 함유할 수 있다. 추가적으로, 공여체 서열은 세포성 염색질 내 관심 영역과 상동성이 아닌 서열을 함유하는 벡터 분자를 포함할 수 있다. 공여체 분자는 세포성 염색질과 상동성인 몇몇, 불연속 영역을 함유할 수 있다. 예를 들어, 보통 관심 영역에 존재하지 않는 서열의 표적화된 삽입을 위해, 상기 서열은 공여체 핵산 분자에 존재하고 관심 영역 내 서열과 상동성인 영역이 측면에 있을 수 있다.
- [0279] 공여체 폴리뉴클레오타이드는 DNA 또는 RNA, 단일-가닥 및/또는 이중-가닥일 수 있고 선형 또는 원형 형태로 세포 내에 도입될 수 있다. 예로, U.S. 특허 공개 번호 2010/0047805; 2011/0281361; 2011/0207221; 및 2019/0330620을 참고한다. 선형 형태로 도입되는 경우, 공여체 서열의 말단은 당업자에게 알려진 방법에 의해 보호될 수 있다(예로, 핵산말단 분해로부터). 예를 들어, 하나 이상의 디옥시뉴클레오타이드 잔기가 선형 분자의 3' 말단에 부가되고/되거나 자가-상보성 올리고뉴클레오타이드가 하나의 또는 양 말단으로 결합된다. 예를 들어, 문헌(Chang and Wilson, Proc. Natl. Acad. Sci. USA (1987); Nehls 등, Science (1996))을 참고한다. 분해로부터 외인성 폴리뉴클레오타이드를 보호하기 위한 추가적인 방법에는 비제한적으로 말단 아미노기(들)의 부가 및 예를 들어, 포스포로티오에이트, 포스포로아미테이트, 및 O-메틸 리보스 또는 데옥시리보스 잔기와 같은 변형된 뉴클레오타이드간 결합의 사용이 포함된다.
- [0280] 따라서, 복구를 위해 공여체 주형을 사용하는 본 발명의 구현에는 선형 또는 원형 형태로 세포 내에 도입될 수 있는 DNA 또는 RNA, 단일-가닥 및/또는 이중-가닥 공여체 주형을 사용할 수 있다. 본 발명의 구현예에서 유전자-편집 조성물은 (1) 복구 전에 유전자에서 이중 가닥 절단에 영향을 미치기 위한 가이드 서열을 포함하는 RNA 분자 및 (2) 복구를 위한 공여체 RNA 주형을 포함하며, 가이드 서열을 포함하는 RNA 분자는 제1 RNA 분자이고 공여체 RNA 주형은 제2 RNA 분자이다. 일부 구현예에서, 가이드 RNA 분자 및 주형 RNA 분자는 단일 분자의 일부로서 연결된다.
- [0281] 공여체 서열은 또한 올리고뉴클레오타이드일 수 있고 내인성 서열의 유전자 교정 또는 표적화된 변경을 위해 사용될 수 있다. 올리고뉴클레오타이드는 벡터 상에서 세포로 도입될 수 있거나, 세포 내로 전기천공될 수 있거나, 당 분야에 알려진 다른 방법을 통해 도입될 수 있다. 올리고뉴클레오타이드는 내인성 유전자 내의 돌연변이된 서열(예로, 베타 글로빈 내 낫(sickle) 돌연변이)을 '교정'하기 위해 사용될 수 있거나 내인성 유전자위 내로 요망되는 목적을 갖는 서열을 삽입하기 위해 사용될 수 있다.
- [0282] 폴리뉴클레오타이드는, 예를 들어, 복제 기원, 프로모터 및 항생제 내성을 인코딩하는 유전자와 같은 추가적인 서열을 갖는 벡터 분자의 일부로서 세포 내로 도입될 수 있다. 또한, 공여체 폴리뉴클레오타이드는 네이키드 핵산으로, 리포좀 또는 폴록사머와 같은 제제와 복합체화된 핵산으로 도입될 수 있거나, 재조합 바이러스(예로, 아데노바이러스, AAV, 헤르페스바이러스, 레트로바이러스, 렌티바이러스 및 인테그라제 결합 렌티바이러스(IDLV))에 의해 전달될 수 있다.
- [0283] 공여체는 일반적으로 그 발현이 통합 부위에서 내인성 프로모터, 즉 공여체가 삽입되는 내인성 유전자의 발현을 유도하는 프로모터에 의해 유도되도록 삽입된다. 그러나, 공여체가 프로모터 및/또는 인핸서, 예를 들어 구성적 프로모터 또는 유도성 또는 조직 특이적 프로모터를 포함할 수 있음이 자명할 것이다.
- [0284] 공여체 분자는 내인성 유전자의 전부, 일부가 발현되거나 전혀 발현되지 않도록 내인성 유전자 내로 삽입될 수 있다. 예를 들어, 본원에서 기재된 바와 같은 트랜스유전자는 내인성 서열의 일부(트랜스유전자에 대해 N-말단 및/또는 C-말단)가 발현되거나 전혀 발현되지 않도록, 예를 들어 트랜스유전자와의 융합물로서, 내인성 유전자위 내로 삽입될 수 있다. 다른 구현예에서, 트랜스유전자(예로, 내인성 유전자와 같은 추가적인 코딩 서열을 갖거나 갖지 않음)는 임의의 내인성 유전자위, 예를 들어 세이프-하버(safe-harbor) 유전자위, 예를 들어 CCR5 유전자, CXCR4 유전자, PPP1R12c(AAVS1로도 알려짐) 유전자, 알부민 유전자 또는 Rosa 유전자 내로 통합된다. 예로, U.S. 특허 번호 7,951,925 및 8,110,379; U.S. 공개 번호 2008/0159996; 20100/0218264; 2010/0291048; 2012/0017290; 2011/0265198; 2013/0137104; 2013/0122591; 2013/0177983 및 2013/0177960 및 U.S. 가출원 번호 61/823,689를 참고한다).
- [0285] 내인성 서열(내인성 또는 트랜스유전자의 일부)이 트랜스유전자와 함께 발현되는 경우, 내인성 서열은 전장 서

열(야생형 또는 돌연변이체) 또는 부분적 서열일 수 있다. 바람직하게는 내인성 서열은 기능적이다. 이들 전장 또는 부분적 서열의 기능의 비제한적 예에는 트랜스유전자(예로, 치료 유전자)에 의해 발현되는 폴리펩티드의 혈청 반감기 증가 및/또는 담체로서의 작용이 포함된다.

- [0286] 또한, 발현을 위해 요구되지는 않지만, 외인성 서열에는 또한 전사 또는 번역 조절 서열, 예를 들어, 프로모터, 인핸서, 인슐레이터(insulator), 내부 리보솜 진입 부위, 2A 펩티드를 인코딩하는 서열 및/또는 폴리아데닐화 신호가 포함될 수 있다.
- [0287] 특정 구현예에서, 공여체 분자는 단백질을 인코딩하는 유전자(예로, 세포에 또는 개체에 부재하는 단백질을 인코딩하는 서열 또는 단백질을 인코딩하는 유전자의 대안적 버전을 코딩함), 조절 서열 및/또는 마이크로RNA 또는 siRNA와 같은 구조적 핵산을 인코딩하는 서열로 구성되는 군으로부터 선택되는 서열을 포함한다.
- [0288] 상기 구현예에 있어서, 본원에서 개시되는 각각의 구현예는 각각의 다른 개시되는 구현예에 적용 가능한 것으로 고려된다. 예를 들어, 본 발명의 임의의 RNA 분자 또는 조성물이 본 발명의 임의의 방법에서 이용될 수 있음이 이해된다.
- [0289] 본원에서 사용된 바와 같이, 모든 제목은 단순히 구성을 위한 것이며 어떠한 방식으로든 본 개시를 제한하려는 것이 아니다. 임의의 개별 섹션의 내용은 모든 섹션에 동일하게 적용 가능하다.
- [0290] 본 발명의 추가적인 목적, 장점, 및 신규한 특성은 하기 실시예의 조사 시 당업자에게 자명해질 것이며, 이는 제한하려는 것이 아니다. 추가적으로, 본원에서 상기 묘사된 바와 같으며 아래의 청구범위 섹션에서 청구된 바와 같은 본 발명의 각각의 다양한 구현예 및 양태는 하기 실시예에서 실험적 지지를 구한다.
- [0291] 명확성을 위해 별도 구현예의 맥락에서 기재되는, 본 발명의 특정 특성이 또한 단일 구현예에서 조합되어 제공될 수 있음이 이해된다. 반대로, 간결성을 위해, 단일 구현예의 맥락에서 기재되는 본 발명의 다양한 특성은 또한 별도로 또는 임의의 적합한 하위-조합으로 또는 본 발명의 임의의 다른 기재된 구현예에서 적합한 바와 같이 제공될 수 있다. 다양한 구현예의 맥락에서 기재되는 특정 특성은 구현예가 이들 요소 없이 작동되지 않는 경우가 아닌 한, 이들 구현예의 필수적 특성으로 간주되어서는 안 된다.
- [0292] 일반적으로, 본원에서 사용되는 명명, 및 본 발명에서 이용되는 실험실 절차에는 분자, 생화학, 미생물학 및 재조합 DNA 기법이 포함된다. 이러한 기법은 문헌에서 자세히 설명되어 있다. 예를 들어, 모두 참조로 포함되는 Sambrook 등, "Molecular Cloning: A laboratory Manual" (1989); Ausubel, R. M. (Ed.), "Current Protocols in Molecular Biology" Volumes I-III (1994); Ausubel 등, "Current Protocols in Molecular Biology", John Wiley and Sons, Baltimore, Maryland (1989); Perbal, "A Practical Guide to Molecular Cloning", John Wiley & Sons, New York (1988); Watson 등, "Recombinant DNA", Scientific American Books, New York; Birren 등 (Eds.), "Genome Analysis: A Laboratory Manual Series", Vols.1-4, Cold Spring Harbor Laboratory Press, New York (1998); U.S. 특허 번호 4,666,828; 4,683,202; 4,801,531; 5,192,659 및 5,272,057에 나타난 방법론; Cellis, J. E. (Ed.), "Cell Biology: A Laboratory Handbook", Volumes I-III (1994); Freshney, "Culture of Animal Cells - A Manual of Basic Technique" Third Edition, Wiley-Liss, N. Y. (1994); Coligan J. E. (Ed.), "Current Protocols in Immunology" Volumes I-III (1994); Stites 등 (Eds.), "Basic and Clinical Immunology" (8th Edition), Appleton & Lange, Norwalk, CT (1994); Mishell and Shiigi (Eds.), "Strategies for Protein Purification and Characterization - A Laboratory Course Manual" CSHL Press (1996); Clokie and Kropinski (Eds.), "Bacteriophage Methods and Protocols", Volume 1: Isolation, Characterization, and Interactions (2009)을 참고한다. 다른 일반 참고문헌은 본 문헌을 통해 제공된다.
- [0293] 본 발명의 보다 완전한 이해를 촉진하기 위해 실시예가 아래에서 제공된다. 하기 실시예는 본 발명의 제조 및 실시의 예시적 방식을 예시한다. 그러나, 본 발명의 범위는 이들 실시예에 개시된 특정 구현예에 제한되지 않으며, 이는 단지 예시 목적을 위한 것이다.
- [0294] **실험 상세사항**
- [0295] 본 발명의 보다 완전한 이해를 촉진하기 위해 실시예가 아래에 제공된다. 하기 실시예는 본 발명을 제조 및 실시의 예시적 방식을 예시한다. 그러나, 본 발명의 범위는 이들 실시예에 개시되는 특정 구현예에 제한되지 않으며, 이는 단지 예시 목적을 위한 것이다.
- [0296] CRISPR 반복체(crRNA), 트랜스활성화 crRNA(tracrRNA), 뉴클레아제 폴리펩티드, 및 PAM 서열은 환경 샘플의 상

이한 메타게놈 서열 데이터베이스로부터 예측하였다. CRISPR 반복체, tracrRNA 서열, 및 뉴클레아제 폴리펩티드 서열이 예측된 박테리아 종/균주를 표 1에 제공한다.

[0297] **OMNI-50 뉴클레아제 폴리펩티드의 작제**

[0298] OMNI-50 뉴클레아제 폴리펩티드의 작제를 위해, OMNI-50 뉴클레아제의 개방 해독틀을 인간 세포주 발현을 위해 코돈 최적화하였다. 최적화된 ORF를 박테리아 플라스미드 pb-NNC 내로 및 포유류 플라스미드 pmOMNI 내로 클로닝하였다(표 4).

[0299] **sgRNA의 예측 및 작제**

[0300] OMNI-50 뉴클레아제에 있어서, sgRNA를 뉴클레아제가 확인된 박테리아 게놈 내 CRISPR 반복체 어레이 서열(crRNA) 및 트랜스-활성화 crRNA(tracrRNA)의 검출에 의해 예측하였다. 천연 성숙-전 crRNA 및 tracrRNA 서열을 컴퓨터내에서 테트라-루프 'gaaa'와 연결하고 듀플렉스의 이차 구조 요소를 RNA 이차 구조 예측 도구를 사용함으로써 예측하였다.

[0301] 전체 듀플렉스 RNA 요소(즉 crRNA-tracrRNA 키메라)의 예측된 이차 구조를 OMNI-50 뉴클레아제에 대해 다양한 버전을 갖는 sgRNA의 설계를 위한 가능한 tracr 서열의 확인을 위해 사용하였다(예를 들어, 도 1a 참고). 상이한 위치에서 상부 줄기부에서의 듀플렉스를 단축함으로써, crRNA 및 tracrRNA를 테트라-루프 'gaaa'와 연결함으로써, 가능한 sgRNA 스캐폴드를 생성하였다(예를 들어, 도 1b 참고; OMNI-50 sgRNA 설계를 표 2에 기재함). OMNI-50에 대해 가능한 설계된 스캐폴드의 적어도 2개 버전을 합성하고 22 nt 범용 고유 스페이스 서열(T2, 서열번호: 56)의 하류에 연결하고, 구성적 프로모터 하의 박테리아 발현 플라스미드 내로 및 U6 프로모터 하의 포유류 발현 플라스미드 내로 클로닝하였다(각각 pbGuide 및 pmGuide, 표 4).

[0302] 잠재적인 전사 및 구조적 제약을 극복하고 인간 세포 환경 맥락에서 sgRNA 스캐폴드의 유연성을 평가하기 위해, 몇몇 버전의 sgRNA를 시험하였다. 각각의 경우에서, 변형은 가능한 sgRNA의 뉴클레오티드 서열에서의 작은 변이를 나타낸다(도 1c, 표 2)

[0303] T1 – GGTGCGGTTTACCAGGGTGTCTG (서열번호: 55)

[0304] T2 – GGAAGAGCAGAGCCTTGCTCTC (서열번호: 56)

[0305] **TXTL에 의한 시험관내 결실 검증**

[0306] 시험관내 PAM 서열의 결실을 Maxwell 등, Methods (2018)에 따라 추적하였다. 간략하게, T7 프로모터 하에 OMNI-50 뉴클레아제 및 sgRNA를 발현하는 선형 DNA를 T7 폴리머라제를 발현하는 선형 작제물과 함께 TXTL 혼합물(Arbor Bioscience)에 첨가하였다. TXTL 혼합물에 의한 RNA 발현 및 단백질 번역은 RNP 복합체의 형성을 초래한다. 선형 DNA를 사용하였으므로, 분해로부터 DNA를 보호하기 위해 RecBCD 억제제인 Chi6 서열을 첨가하였다. sgRNA 스페이스를 잠재적 PAM 서열의 8N 무작위화 세트가 측면에 있는 표적화 프로토스페이스를 함유하는 플라스미드 라이브러리(pbPOS T2 라이브러리, 표 4)를 표적화하도록 설계한다. 절단된 라이브러리 및 비-표적화 gRNA를 발현하는 대조군 라이브러리(T1) 모두에 대해 필요한 어댑터 및 지표를 부가하기 위해 PCR 사용 시 고처리량 서열분석에 의해 라이브러리로부터의 PAM 서열의 결실을 측정하였다. 심층 서열분석 후, 시험관내 시스템에 의한 기능적 DNA 절단을 시사하는 OMNI 뉴클레아제에 의한 대조군에서의 이의 발생 대비 동일한 PAM 서열을 갖는 결실된 서열의 분율에 의해 시험관내 활성을 확인하였다(도 2, 표 3). OMNI-50을 2개의 sgRNA 버전(V1 및 V2)으로 시험하였다. 두 경우 모두에서, 분석으로부터 NGG의 명확한 PAM이 추정되었다(도 2). 일부 활성이 또한 NAG 및 NGA PAM 서열로 관찰되었다.

[0307] **포유류 시스템에서의 PAM 라이브러리**

[0308] PAM 서열 선호는 뉴클레아제의 내재적 특성으로 간주되므로, 어느 정도까지는, 세포성 환경, 게놈 조성, 및 게놈 크기에 의해 영향받을 수 있다. 인간 세포성 환경은 각각의 이들 특성에 관해 박테리아 환경과 크게 상이하므로, 인간 세포성 맥락에서 PAM 선호에서의 잠재적 차이를 해결하기 위해 "미세 조정" 단계를 도입하였다. 상기 목적을 위해, PAM 라이브러리를 인간 세포주에서 작제하였다. 상기 검증에서, PAM 라이브러리를 불변 표적 서열에 이어 6N의 신장물로서 바이러스 벡터를 사용하여 세포로 도입하였다(표 4 참고). OMNI-50 및 라이브러리 불변 표적 부위를 표적화하는 sgRNA의 도입 시, NGS 분석을 사용하여 편집된 서열 및 이와 연관된 PAM을 확인하였다. 이어서 농축된 편집된 서열을 사용하여 PAM 공통부를 정의하였다. 상기 방법론을 적용하여 포유류 세포에서 OMNI-50 뉴클레아제의 최적화된 PAM 요건을 결정한다(표 3, "포유류 요건"). OMNI-50 PAM은 시험관내 TXTL에

서 확인된 것과 동일한 것으로 확인되었다.

[0309] 포유류 세포에서 최적화된 DNA 서열에 의해 코딩되는 OMNI-50 뉴클레아제의 발현

[0310] 첫 번째로, 포유류 세포에서 OMNI-50을 인코딩하는 각각의 최적화된 DNA 서열의 발현을 검증하였다. 상기 목적을 위해, P2A 펩티드에 의해 mCherry에 연결된 HA-태그 OMNI-50 뉴클레아제 또는 스트렙토코커스 피오게네스 (Streptococcus Pyogenes) Cas9(SpCas9)를 코딩하는 발현 벡터(pmOMNI, 표 4)를 Jet-optimus™ 전달감염 시약 (polyplus-전달감염)을 사용하여 Hek293T 세포 내로 도입하였다. P2A 펩티드는 OMNI 뉴클레아제 및 mCherry가 발현 시 분리되도록 세포에서 재조합 단백질의 전달을 유도할 수 있는 자가-절단 펩티드이다. mCherry는 발현 벡터로부터 OMNI의 전사 효율의 표시인자로서 작용한다. OMNI-50 단백질의 발현은 항-HA 항체를 사용하여 웨스턴 블롯 검정에 의해 확인하였다(도 3).

[0311] 내인성 게놈 표적에 대한 인간 세포 내 활성

[0312] OMNI-50을 또한 인간 세포 내 특정 게놈 위치 상에서 편집을 촉진하는 그 능력에 대해 검증하였다. 상기 목적을 위해, OMNI-P2A-mCherry 발현 벡터(pmOMNI, 표 4)를 인간 게놈 내 특정 위치를 표적화하도록 설계된 sgRNA(pmGuide, 표 4)와 함께 HeLa 세포 내로 전달감염시켰다. 72 h째에, 세포를 수확하였다. 마커로서 mCherry 형광을 사용하여 FACS에 의해 전달감염 효율의 정량을 위해 세포의 절반을 사용하였다. 세포의 나머지 절반을 용해시키고, 이의 게놈 DNA를 사용하여 대응하는 추정 게놈 표적을 PCR 증폭시켰다. 앰플리콘으로 NGS를 거치고 생성 서열을 사용하여 각각의 표적 부위에서 편집 이벤트의 백분율을 계산하였다. 절단 부위 주위의 짧은 삽입 또는 결실(인델)은 뉴클레아제-유도 DNA 절단 후 DNA 말단 복구의 전형적인 결과이다. 편집%의 계산은 각각의 앰플리콘 내 인델-함유 서열의 분율로부터 추정하였다. 모든 편집 값을 각각의 실험에 대해 수득된 전달 감염 및 번역 효율에 대해 표준화하고 mCherry 발현 세포의 백분율로부터 추정하였다. 표준화 값은 발현된 뉴클레아제의 세포 집단 내 유효 편집 수준을 나타낸다.

[0313] OMNI-50의 게놈 활성을 각각 상이한 게놈 위치를 표적화하도록 설계된 11개의 고유 sgRNA 패널을 사용하여 평가하였다. 이들 실험의 결과를 표 6에 요약한다. 표(6열, "편집%")에서 알 수 있듯이, OMNI-50은 시험된 모든 표적 부위에서 음성 대조군(9열, "음성 대조군에서의 편집%")에 비해 높고 유의미한 편집 수준을 나타낸다. OMNI-50은 시험된 11/11 부위에서 높고 유의미한 편집 수준을 나타낸다.

[0314] 인간 세포에서의 내인성 정확도

[0315] 뉴클레아제의 내인성 정확도는 그 절단 특이성의 측정이다. 고-정확도 뉴클레아제는 의도되지 않는 표적의("오프-표적") 최소 절단 또는 절단 부재를 가지며 의도되는 표적 상의("온-표적") 절단을 촉진하는 뉴클레아제이다. CRISPR 뉴클레아제에 있어서, 표적은 가이드 RNA의 스페이서 요소에 대한 서열 상보성에 기반하여 획득된다. 오프-표적화는 스페이서 서열 및 의도되지 않는 표적 간 유사성으로부터 초래된다. 인간 세포 내 게놈 수준에서의 OMNI-50의 내인성 정확도를 온-표적 영역 및 사전-검증된 오프-표적 영역 모두에 대한 PCR 증폭, NGS, 및 인텔 분석 후, 상기 섹션에 기재된 바와 같이 활성 검정을 수행함으로써 측정하였다. OMNI-50에 대한 내인성 정확도의 측정을 도 4a에 나타낸다. 상기 예에서, OMNI-50 정확도를 독립적으로 2개의 가이드 RNA를 사용하여 측정하였고, 각각의 경우 SpCas9의 나란한 측정을 참조를 위해 제공한다. 첫 번째 부위를 chr19 상의 ELANE 유전자 상류의 정의된 온-표적 부위 및 Chr15 상의 오프-표적 부위를 갖는 ELANE g35 gRNA를 사용하여 표적화하였다(표 6). 도 4a에서 알 수 있듯이, OMNI-50에 의해 수득된 온-표적/-오프-표적 편집 효율비는 41:0인 반면 SpCas9 온/오프 비는 6.8:1이다(각각 40.9%/0%; 18.6%/2.7%). 두 번째 부위를 ELANE g62 gRNA에 의해 표적화하였다(표 6). 상기 gRNA 스페이서 서열은 chr19 상의 ELANE 유전자에 정의된 온-표적 부위 및 Chr1 상의 오프-표적 부위를 갖는다. 상기 경우에서, OMNI-50에 의해 수득된 온/오프 비는 SpCas9에 의해 수득된 1.7:1 비에 비해 72:1이었다(각각 38.9%/0.6%; 43.1%/25.8%). 이들 결과는 OMNI-50이 이들 특정 gRNA를 사용하는 SpCas9와 대조적으로 유의미하게 더 높은 내인성 정확도를 가짐을 실증한다. 내인성 정확도를 이후 U2OS 세포주 내로의 RNP 전기전공에 의해 제2 시스템에서 시험하였다(도 4b). ELANE g35에 있어서, OMNI-50에 의해 수득된 온/오프 표적 편집 효율비는 9:1이었던 반면 SpCas9 온/오프 비는 1:1이다(각각 91%/10%; 93%/91%). 이들 2개의 별도 시스템에서, OMNI-50 정확도가 SpCas9보다 우수하였다.

[0316] 가이드-seq 비편향 분석 방법을 사용하는 오프-표적 평가

[0317] OMNI-50의 특이성을 추가 평가하기 위해, 오프-표적 수를 가이드-seq를 사용하여 몇몇 부위에 걸쳐 시험하였다. SpCas9에 대한 오프-표적 카운트는 수 개부터 수백 개로 부위에 걸쳐 변화한 반면, OMNI-50 오프-표적 카운트는 시험된 모든 부위에서 20개 미만이었다. SpCas9 또는 OMNI-50을 사용하여 10개 초과 판독을 갖는 부위에 대해

확인된 오프-표적의 수 비교는 OMNI-50의 높은 특이성을 시사한다. 시험된 6개 부위 중 5개에서, SpCas9 오프-표적의 수는 OMNI-50에 비해 상당히(2 내지 20-배) 더 큰 반면, 6개 부위 중 1개에서만 오프-표적 카운트가 두 뉴클레아제 간에 필적한다(표 9).

[0318] OMNI-50 단백질의 정제

[0319] OMNI-50 개방 해독들을 박테리아 발현 플라스미드(T7-NLS-OMNI-NLS-HA-His-태그, pET9a, 표 4) 내로 클로닝하고 C43 세포(Lucigen)에서 발현시켰다. 세포를 log상 중간까지 Terrific 배양액 중에 성장시킨 후 온도를 18℃로 낮췄다. 발현을 16~20 h 동안 1 mM IPTG로 0.6 OD에서 유도한 후 세포를 수확하고 -80℃에서 냉동하였다. 세포 페이스트를 EDTA-비함유 완전 프로테아제 억제제 카테일 세트 III(Calbiochem)이 보충된 용해 완충액(50 mM NaH₂PO₄, 300 mM NaCl, 10 mM 이미다졸 pH 8.0, 1 mM TCEP) 중에 재현탁하였다. 세포를 초음파분쇄를 사용하여 용해시키고 청정화된 용해액을 Ni-NTA 수지와 인큐베이션하였다. 수지를 중력 칼럼 상으로 로딩하고, 세척 완충액(50 mM NaH₂PO₄, 300 mM NaCl, 50 mM 이미다졸 pH 8.0, 1 mM TCEP)으로 세척하고, OMNI-50 단백질을 100~500 mM 이미다졸이 보충된 세척 완충액으로 용출하였다. OMNI-50 단백질을 함유하는 분획을 폴링하고, 농축하고, 센트리콘(Amicon Ultra 15 ml 100 K, Merck) 상으로 로딩하고, GF 완충액(50 mM 트리스-HCl pH 7.5, 500 mM NaCl, 10% 글리세롤, 0.4 M 아르기닌)으로 완충액을 교환하였다. 농축된 OMNI-50 단백질을 50 mM 트리스-HCl pH 7.5, 500 mM NaCl, 10% 글리세롤, 0.4 M 아르기닌으로 HiLoad 16/600 Superdex 200 pg-SEC, AKTA Pure(GE Healthcare Life Sciences) 상에서 SEC에 의해 추가 정제하였다. OMNI-50 단백질을 함유하는 분획을 폴링하고, 농축하고, 10 mM 트리스-HCl, pH 7.5, 150 mM NaCl, 10% 글리세롤 및 1 mM TCEP의 최종 보관 완충액으로 센트리콘(Amicon Ultra 15 ml 100 K, Merck) 상으로 로딩하였다. 정제된 OMNI-50 단백질을 10 mg/ml 원액으로 농축하고, 액체 질소 중에 급속-냉동하고, -80℃에서 보관하였다.

[0320] RNP 활성 검정에 의한 가이드 최적화

[0321] OMNI-50의 합성 sgRNA를 3' 및 5' 말단에서 3개의 2'-O-메틸 3'-포스포로티오에이트와 함께 합성하였다(Agilent). 가이드 35의 상이한 스페이서 길이(17~23 nt)를 갖는 OMNI-50 RNP의 활성 검정이 본원에서 기재된다(표 5, 도 5a). 간략하게, 4 pmol의 OMNI-50 뉴클레아제를 6 pmol의 합성 가이드와 혼합하였다. 실온에서 10분 인큐베이션 후, RNP 복합체를 100 ng의 온-표적 주형과 반응시켰다. 22 nt 이상의 스페이서만 온-표적 주형의 거의 전체 절단을 나타낸다. 스페이서 길이 20 - 23 nt를 갖는 감소량의 RNP(4, 2, 1.2, 0.6 및 0.2 pmol)를 100 ng의 DNA 표적 주형(도 5b)과 반응시켰다. 22 nt 이상인 길이의 스페이서는 더 낮은 RNP 농도에서도 더 우수한 절단 활성을 나타낸다.

[0322] 스페이서 길이 최적화를 또한 포유류 세포 맥락에서 수행하였다. 100 uM 뉴클레아제를 상이한 스페이서 길이(17~23 nt, 표 5)를 갖는 120 uM의 합성 가이드 및 100 uM Cas9 전기천공 증강제(IDT)와 혼합함으로써 RNP를 어셈블리하였다. 실온에서 10분 인큐베이션 후, RNP 복합체를 200,000 사전-세척 U2OS, iPSC, 또는 HSC 세포와 혼합하고 제조업체의 프로토콜에 따라, 각각 DN100 또는 CA137 프로그램으로 Lonza SE 또는 P3 세포주 4D-Nucleofector™ X 키트를 사용하여 전기천공하였다. 72 h에 세포를 용해시키고 이의 게놈 DNA를 PCR 반응에서 사용하여 대응하는 추정 게놈 표적을 증폭하였다. 앰플리콘으로 NGS를 거친 후 생성 서열을 사용하여 편집 이벤트의 백분율을 계산하였다. 도 5c, 도 6, 및 표 10에서 알 수 있듯이, 17~19 nt의 스페이서는 낮은 편집 수준을 나타내며, 20 nt 스페이서는 중간 편집 수준을 나타내고, 21~23 nt의 스페이서는 가장 높은 편집 수준을 나타낸다.

[0323] U2OS 세포주를 사용하여, 상이한 tracer RNA 서열 변이를 시험하였다(표 2). 상이한 sgRNA 버전을 20 nt 스페이서로 시험하였다. 도 5d에서 알 수 있듯이, sgRNA V1, V2, 또는 V3을 사용하는 RNP 어셈블리는 유사한 편집 수준을 초래한다. 그러나, sgRNA V4를 사용하는 RNP 어셈블리는 유의미하게 더 높은 편집 수준을 초래한다.

[0324] 5개 게놈 부위에 걸쳐 21 nt 및 22 nt 스페이서를 사용하여 HSC에서 수득된 결과 비교는 22 nt 스페이서가 효율적인 편집을 위해 약간 선호됨을 제시한다(도 6 및 표 10).

[0325] RNP로서의 OMNI-50의 활성

[0326] 포유류 세포에서 RNP로서의 OMNI-50 단백질의 활성을 먼저 U2OS 세포주에서 시험하였고, 이후 3개의 일차 세포 시스템: iPSC, HSC, 및 T 세포에서 시험하였다. 표 7에서 알 수 있듯이, 편집이 모든 시스템에서 관찰되었다.

[0327] OMNI-50을 2개 유전자 상에서 T-세포 내 편집 활성에 대해 시험하였다(별첨 표 7). OMNI-50을 TRAC를 표적화하는 34개의 가이드 및 B2M을 표적화하는 26개의 가이드로 시험하였다. 시험된 TRAC 가이드의 64%(22/34)는 활성

인 것으로 확인되었으며, 편집 수준은 5% 내지 84% 범위였다. 유사하게, B2M 가이드의 57%는 활성이었고, 편집 수준은 5% 내지 61% 범위였다. 이들 결과를 별첨 표 7에 요약한다.

[0328] 각각의 19개 가이드 레퍼토리에서 TRAC 및 B2M 유전자 둘 다에서 높은 편집이 관찰되었다. 멀티플렉스화 및 추가 최적화에 대한 잠재력을 고려하면, 적절한 전략으로 OMNI-50에 의한 두 유전자의 전체 녹-아웃이 가능하다.

[0329] U2OS 세포, iPSC 및 HSC에서, ELANE 유전자를 표적화하는 가이드를 OMNI-50으로 시험하였다. 시험된 모든 5개 가이드는 U2OS 세포 및 HSC 둘 다에서 22% 초과 편집을 나타내었다. iPSC에서는 ELANE g35만 53%의 편집 수준으로 시험되었다. 상기 결과는 다른 시스템으로 수득된 결과에 비해 더 낮다.

[0330] 멀티플렉스화

[0331] 또한 2개 RNP 집단을 혼합하고 혼합물을 일차 T 세포 내로 전기천공함으로써 OMNI-50을 멀티플렉스 편집에 대해 시험하였다. gRNA #32를 TRAC에 대해 사용하였고, gRNA #15를 B2M에 대해 사용하였다(스페이서 서열은 표 8에 기재함). 72 h째에 세포를 수확하고 NGS에 의한 편집에 대해 시험하였다. TRAC 유전자는 50% 편집으로 측정되었고 B2M 유전자는 25% 편집으로 측정되었다. 이들 결과는 멀티플렉스 시험과 나란히 수행된 단일 RNP를 이용한 편집 수준과 유사하였다(표 8).

표 1

OMNI-50 뉴클레아제 서열

원천 유기체	OMNI-50 아미노산 서열의 서열번호	OMNI-50 을 인코딩하는 DNA 서열의 서열번호	인간 세포에서 OMNI-50 을 인코딩하기 위해 최적화된 DNA 서열 코돈의 서열번호
에자키엘라 페루엔시스(<i>Ezakiella peruensis</i>) 균주 M6.X2	3	11	12, 13

[0332]

[0333] 표 1. OMNI-50 뉴클레아제 서열: 표 1은 OMNI-50 뉴클레아제가 동정된 유기체, 그 단백질 서열, 그 DNA 서열, 및 그 인간 최적화된 DNA 서열(들)을 기재한다.

표 2

OMNI-50 가이드 서열

최소 crRNA:tracrRNA 듀플렉스	crRNA(반복 체)	GUUUGAGAG
	tracrRNA(항- 반복체)	CGAGUCAAU(서열번호: 149)
crRNA:tracrRNA 듀플렉스 V1	crRNA(반복 체)	GUUUGAGAGUUAUG(서열번호: 37)
	tracrRNA(항- 반복체)	CAUGACGAGUCAAU(서열번호: 38)
crRNA:tracrRNA 듀플렉스 V2	crRNA(반복 체)	GUUUGAGAGUUAUGUAA(서열번호: 39)
	tracrRNA(항- 반복체)	UUACAUGACGAGUCAAU(서열번호: 40)
TracrRNA 서열	TracrRNA 부분 1	AAAAUUUAUUCAAACC(서열번호: 150)
	TracrRNA 부분 2	GCCUAUUUAUAGGC(서열번호: 42)
	TracrRNA 부분 3	CGCAGAGUUCUGC(서열번호: 151)
	TracrRNA 부분 4	AUUUAUGCUUGCUAUUGCAAGCUUUUU(서열번호: 152)
	전체 tracrRNA V1	CAUGACGAGUCAAUAAAAUUUAUCAAACCGCCUUAUUUAAGGCCGAGAUUCUGCAU UAUGCUUGCUAUUGCAAGCUUUUU(서열번호: 153)
	전체 tracrRNA V2	UUACAUGACGAGUCAAUAAAAUUUAUCAAACCGCCUUAUUUAAGGCCGAGAUUCUG CAUUAUGCUUGCUAUUGCAAGCUUUUU(서열번호: 154)
sgRNA 버전	sgRNA V1	GUUUGAGAGUUAUGgaaaCAUGACGAGUCAAUAAAAUUUAUCAAACCGCCUUAUUUAAG GCCGAGAUUCUGCAUUAUGCUUGCUAUUGCAAGCUUUUU(서열번호: 44)
	sgRNA V2	GUUUGAGAGUUAUGUAgaaaaUUAUGACGAGUCAAUAAAAUUUAUCAAACCGCCUUAU UAUUAAGGCCGAGAUUCUGCAUUAUGCUUGCUAUUGCAAGCUUUUU(서열번호: 45)
기타 sgRNA 최적화	sgRNA V3	GUUUGAGAGUUAUGgaaaACAUGACGAGUCAAUAAAAUUUAUCAAACCGCCUUAUUUA AGGCCGAGAUUCUGCAUUAUGCUUGCUAUUGCAAGCUUUUU(서열번호: 87)
	sgRNA V4	GUUUGAGAGUUAUGUAgaaaaUACAUGACGAGUCAAUAAAAUUUAUCAAACCGCCUUAUU UAAGGCCGAGAUUCUGCAUUAUGCUUGCUAUUGCAAGCUUUUU(서열번호: 88)

[0334]

표 3

OMNI-50 PAM 서열

TXTL 결실	PAM 일반	NGG
	PAM 특이적	NGG
	활성(1 - 결실 스코어)*	0.98
	sgRNA	V1, V2
포유류 정련	PAM 포유류	NGG

[0335]

[0336] *결실 스코어 - 2개의 가장 많이 결실된 부위로부터의 비의 평균

표 4

플라스미드 및 작제물

플라스미드	목적	요소	예
pbNNC-2	박테리아 시스템에서의 OMNI 폴리펩티드 발현	T7 프로모터 HA 태그-링커-OMNI ORF(인간 최적화됨)-T7 종결인자	pbNNC2 OMNI-50
pbGuide T1/T2	박테리아 시스템에서의 OMNI sgRNA 발현	J23119 프로모터-T1/T2 스페이서 sgRNA 스캐폴드-rrnB T1 종결인자	pbGuide OMNI-50 T2 sgRNA V2
pbPOS T2 라이브러리	박테리아/TXTL 결실 검정	T2 프로토스페이서-8N PAM 라이브러리-클로람페니콜 아세틸트랜스퍼라제	pbPOS T2 라이브러리
pET9a	OMNI 단백질의 발현 및 정제	T7 프로모터-SV40 NLS-OMNI ORF(인간 최적화됨)-HA-SV40 NLS-8 His-태그-T7 종결인자	pET9a OMNI-50-His 태그
pmOMNI	포유류 시스템에서의 OMNI 폴리펩티드 발현	CMV 프로모터-코작-SV40 NLS-OMNI ORF(인간 최적화됨)-HA-SV40 NLS-P2A-mCherry-Bgh 폴리(A) 신호	pmOMNI OMNI-50
pmGuide 내인성 부위	포유류 시스템에서의 OMNI sgRNA 발현	U6 프로모터-내인성 스페이서 sgRNA 스캐폴드	pmGuide OMNI-50 CXCR4 sgRNA V3
pPM3L3.1	포유류 세포에서의 PAM 라이브러리에 대한 바이러스 벡터	LTR-HIV-1 ψ -CMV 프로모터-T2-PAM 라이브러리(6N)-GFP-SV40 프로모터-블라스토시딘 S 테아미나제-LTR	pPM3L3.1

[0337]

[0338] [표 4 별첨]

작제물 요소의 상세사항

요소	단백질 서열	DNA 서열
HA 태그	서열번호: 63	서열번호: 64
NLS	서열번호: 65	서열번호: 66
P2A	서열번호: 85	서열번호: 86
mCherry	서열번호: 67	서열번호: 68

[0339]

표 5

합성 sgRNA(스페이서 및 스캐폴드)

명칭	050_ELANE_V2_g35_23	050_ELANE_V2_g35_22	050_ELANE_V2_g35_21	050_ELANE_V2_g35_20	050_ELANE_V2_g35_19
스페이서	UgcAGUCCGG GCUGGAGCG GGU(서열번호: 112)	gcAGUCCGGG CUGGAGCGG GU(서열번호: 116)	cAGUCCGGG UGGAGCGGG U(서열번호: 118)	AGUCCGGGCU GGGAGCGGGU (서열번호: 120)	GUCCGGGCUG GGAGCGGGU(서열번호: 122)
스캐폴드	gUUUGAGAGU UAUGUAAgaa aUUACAUGAC GAGUUCAAAU AAAAAUUUU UCAAACCGCC UAUUUAUAGG CCGCAGAUU UCUGCAUUAU GCUUGCUAUU GCAAGCUUUU UU(서열번호: 45)	gUUUGAGAGU UAUGUAAgaa aUUACAUGAC GAGUUCAAAU AAAAAUUUU UCAAACCGCC UAUUUAUAGG CCGCAGAUU UCUGCAUUAU GCUUGCUAUU GCAAGCUUUU UU(서열번호: 45)	gUUUGAGAGU UAUGUAAgaa aUUACAUGAC GAGUUCAAAU AAAAAUUUU UCAAACCGCC UAUUUAUAGG CCGCAGAUU UCUGCAUUAU GCUUGCUAUU GCAAGCUUUU UU(서열번호: 45)	gUUUGAGAGU UAUGUAAgaa aUUACAUGAC GAGUUCAAAU AAAAAUUUU UCAAACCGCC UAUUUAUAGG CCGCAGAUU UCUGCAUUAU GCUUGCUAUU GCAAGCUUUU UU(서열번호: 45)	gUUUGAGAGU UAUGUAAgaa aUUACAUGAC GAGUUCAAAU AAAAAUUUU UCAAACCGCC UAUUUAUAGG CCGCAGAUU UCUGCAUUAU GCUUGCUAUU GCAAGCUUUU UU(서열번호: 45)
버전	V2	V2	V2	V2	V2
전체 sgRNA 서열	UgcAGUCCGG GCUGGAGCG GGUgUUUGAG AGUUAUGUAA gaaaUUACAU GACGAGUUA AAUAAAAUU UAUCAAACC GCCUAUUUAU	gcAGUCCGGG CUGGAGCGG GUgUUUGAGA GUUAUGUAAg aaaUUACAUG ACGAGUUA AUAAAAUUU AUCAAACCG CCUAUUUAUA	cAGUCCGGG UGGAGCGGG UgUUUGAGAG UAUGUAAgaa aaUUACAUGA CGAGUCAA UAAAAUUUA UCAAACCGC CUAUUAUAG	AGUCCGGGCU GGGAGCGGGU gUUUGAGAGU UAUGUAAgaa aUUACAUGAC GAGUUCAAAU AAAAAUUUU UCAAACCGCC UAUUUAUAGG	GUCCGGGCUG GGAGCGGGUg UUUGAGAGUU AUGUAAgaaa UUACAUGACG AGUUCAAUA AAAAAUUUU CAAACCGCCU AUUUUAUAGG

[0340]

	AGGCCGAGA UGUUCUGCAU UAUGCUUGCU AUUGCAAGCU UUUUU(서열번호: 113)	GGCCGAGAU GUUCUGCAU AUGCUUGCUA UUGCAAGCUU UUUU(서열번호: 117)	GCCGAGAU UUCUGCAUUA UGCUUGCUA UGCAAGCUU UUU(서열번호: 119)	CCGAGAU UCUGCAUUA GCUUGCUAU GCAAGCUUU UU(서열번호: 121)	CGAGAU CUGCAUUA CUUGCUAU CAAGCUUUU U(서열번호: 123)
프로토스페이서(PAM은 진핵세포 표적)	CTGTTGCTGC AGTCCGGGCT GGGAGCGGT GGGAGCAGA GGG(서열번호: 114)	CTGTTGCTGC AGTCCGGGCT GGGAGCGGT GGGAGCAGA GGG(서열번호: 114)	CTGTTGCTGC AGTCCGGGCT GGGAGCGGT GGGAGCAGA GGG(서열번호: 114)	CTGTTGCTGC AGTCCGGGCT GGGAGCGGT GGGAGCAGA GGG(서열번호: 114)	CTGTTGCTGC AGTCCGGGCT GGGAGCGGT GGGAGCAGA GGG(서열번호: 114)
프로토스페이서(PAM은 진핵세포 표적)-오프 표적	GTTAAGAgac AGTCCaGGCT GGGAGCaGGT GGGAGAGGA GGG(서열번호: 115)	GTTAAGAgac AGTCCaGGCT GGGAGCaGGT GGGAGAGGA GGG(서열번호: 115)	GTTAAGAgac AGTCCaGGCT GGGAGCaGGT GGGAGAGGA GGG(서열번호: 115)	GTTAAGAgac AGTCCaGGCT GGGAGCaGGT GGGAGAGGA GGG(서열번호: 115)	GTTAAGAgac AGTCCaGGCT GGGAGCaGGT GGGAGAGGA GGG(서열번호: 115)

[0341]

합성 sgRNA(스페이서 및 스캐폴드)

명칭	050_ELAME_V2_g35_18	050_ELAME_V2_g35_17	050_ELAME_V3_g35_20	050_ELAME_V4_g35_20
스페이서	UCCGGGCUGGGAGCGGGU(서열번호: 124)	CCGGGCUGGGAGCGGU(서열번호: 126)	AGUCCGGGCUGGGAGCGGGU(서열번호: 120)	AGUCCGGGCUGGGAGCGGGU(서열번호: 120)
스캐폴드	gUUUGAGAGUUAUGUAaaaaUUACAUGACGAGUUCAAA UAAAAUUUAUUC AAACCGCCUAUU AUAGGCCGCAGAU GUUCUGCAUUAUG CUUGCUAUUGCAA GCUUUUUU(서열번호: 45)	gUUUGAGAGUUAUGUAaaaaUUACAUGACGAGUUCAAA UAAAAUUUAUUC AAACCGCCUAUU AUAGGCCGCAGAU GUUCUGCAUUAUG CUUGCUAUUGCAA GCUUUUUU(서열번호: 45)	gUUUGAGAGUUAUGUaaaaACAUGACGAGUUCAAAUA AAUUUAUUCAAAC CGCCUAUUUAUAG GCCGCAGAUUUG UGCAUUAUGCUUG CUUUGCAAGCUU UUUU(서열번호: 87)	gUUUGAGAGUUAUGUAaaaaUACAUGACGAGUUCAAAUA AAAAAUUUAUCAA ACCGCCUAUUUAU AGGCCGCAGAUU UCUGCAUUAUGCU UGCUAUUGCAAGC UUUUU(서열번호: 88)
버전	V2	V2	V3	V4
전체 sgRNA 서열	UCCGGGCUGGGAGCGGGUgUUUGAGAGUUAUGUAaaaaUUAACAUGACGAGUUCAAA UAAAAUUUAUUAUUCAAACCGCCUAUUUAUAGGCCGCAGAUUUGCUAUUGCAUUGCAAGCUUUUUU(서열번호: 125)	CCGGGCUGGGAGCGGUgUUUGAGAGUUAUGUAaaaaUUAACAUGACGAGUUCAAA UAAAAUUUAUUAUUCAAACCGCCUAUUUAUAGGCCGCAGAUUUGCUAUUGCAUUGCAAGCUUUUUU(서열번호: 127)	AGUCCGGGCUGGGAGCGGGUgUUUGAGUUAUGUaaaaACAUGACGAGUUCAAA UAAAAUUUAUUAUUCAAACCGCCUAUUUAUAGGCCGCAGAUUUGCUAUUGCAAGCUUUUUU(서열번호: 128)	AGUCCGGGCUGGGAGCGGGUgUUUGAGUUAUGUaaaaUACAUGACGAGUUCAAA UAAAAUUUAUUAUUCAAACCGCCUAUUUAUAGGCCGCAGAUUUGCUAUUGCAUUGCAAGCUUUUUU(서열번호: 129)
프로토스페이서(PAM)	CTGTTGCTGCAGTCCGGGCTGGGAGC	CTGTTGCTGCAGTCCGGGCTGGGAGC	CTGTTGCTGCAGTCCGGGCTGGGAGC	CTGTTGCTGCAGTCCGGGCTGGGAGC

[0342]

은 진하게 표시)-온 표적	GGGTGGGAGCAGAGGG(서열번호: 114)	GGGTGGGAGCAGAGGG(서열번호: 114)	GGGTGGGAGCAGAGGG(서열번호: 114)	GGGTGGGAGCAGAGGG(서열번호: 114)
프로토스페이서(PAM)은 진하게 표시)-오프 표적	GTTAAGAGaCAGTCCaGGCTGGGAGCaGGTGGGAGAGGAGGG(서열번호: 115)	GTTAAGAGaCAGTCCaGGCTGGGAGCaGGTGGGAGAGGAGGG(서열번호: 115)	GTTAAGAGaCAGTCCaGGCTGGGAGCaGGTGGGAGAGGAGGG(서열번호: 115)	GTTAAGAGaCAGTCCaGGCTGGGAGCaGGTGGGAGAGGAGGG(서열번호: 115)

[0343]

표 6

내인성 게놈 표적 상에서 인간 세포 내 OMNI-50의 활성

게놈 부위	대응하는 스페이서 명칭	스페이서 서열	3' (PAM 함유) 게놈 서열 (PAM은 전하게 표시)	인텔%	전달감염%	표준화 편집%	음성 대조군에서의 전달감염%	음성 대조군에서의 전달감염%	음성 대조군에서의 표준화 편집%
EMX1 부위 2	EMX1g1_OMNI150	UCUGUGAAU GUUAGACCC AU(서열번호: 97)	GGGAGCAG	44.1 8- 25.7 2			0.02		
EMX1 부위 3	EMX1g2_OMNI150	CCAUGGGAG CAGCUGGUC AG(서열번호: 98)	AGGGGACC	55.8 1					
CXCR4 부위 3	CXCR4g1_OMNI150	GCAAGAGAC CCACACACC GG(서열번호: 99)	AGGAGCGC	29.6 8- 32.1 4			0.18		
CXCR4 부위 4	CXCR4g2_OMNI150	ACACCGGAG GAGCGCCCG CU(서열번호: 100)	TGGGGGAG	42.1 3- 40.8 5			0.22		
PDCD1 부위 4	PDCD1g1_OMNI150	CGUCUGGGC GGUGCUACA AC(서열번호: 101)	TGGGCTGG	13.3 5- 8.7			0.06		
PDCD1 부위 5	PDCD1g2_OMNI150	CUACAACUG GGCUGGCGG CC(서열번호: 102)	AGGATGGT	17.6 3					
ELANE g35	ELANEg35_OMNI150	AGUCCGGGC UGGGAGCGG GU(서열번호: 103)	GGGGAGCA	40.9 2- 55.3 9			0.24	5.95	3,982225429
ELANE g58	ELANEg58_OMNI150	GCUGCGGGA AAGGGAUUC CC(서열번호: 104)	TGGGACTC	11.1 1	20.60	54.23	0.18	5.95	2,974553445
ELANE g38	ELANEg38_OMNI150	ACAGCGGGU GUAGACUCC GA(서열번호: 105)	GGGGGACG	9.90					

[0344]

게놈 부위	대응하는 스페이서 명칭	스페이서 서열	3' (PAM 함유) 게놈 서열 (PAM은 전하게 표시)	인텔%	전달감염%	표준화 편집%	음성 대조군에서의 전달감염%	음성 대조군에서의 전달감염%	음성 대조군에서의 표준화 편집%
ELANE g39	ELANEg39_OMNI150	CAGCGGGUG UAGACUCCG AG (서열번호: 106)	GGGGACGT	24.8 7					
ELANE g62	ELANEg62_OMNI150	GUCAAGCCC CAGAGGCCA CA (서열번호: 107)	GGGACAGA	38.8 7- 52.7 4			0.12	5.95	2.002503126

[0345]

[0346]

표 6. 포유류 세포에서 내인성 맥락에서의 뉴클레아제 활성: OMNI-50 뉴클레아제를 sgRNA 발현 플라스미드와 함께 DNA 전달감염에 의해 포유류 세포 시스템(HeLa)에서 발현하였다. 세포 용해액을 부위 특이적 게놈 DNA 증폭 및 NGS를 위해 사용하였다. 인텔의 백분율을 측정하고 분석하여 편집 수준을 결정하였다. 각각의 sgRNA는 tracrRNA(표 2 참고) 및 본원에서 상세히 나타낸 스페이서로 이루어진다. 3' 게놈 스페이서 서열은 OMNI-50 뉴클레아제에 관련된 PAM을 함유한다. 전달감염 효율(전달감염%)을 상술된 바와 같이, mCherry 신호의 유세포 측정 정량에 의해 측정하였다. 전달감염 효율을 사용하여 편집 수준을 표준화하였다(표준화 인텔%). 모든 시험은 3회씩 수행하였다. OMNI 뉴클레아제만(즉, 가이드 없이) 전달감염된 세포가 음성 대조군으로 작용하였다.

표 7

RNP 로서의 OMNI-50 활성

시스템	게놈 부위	대응하는 스페이서 명칭	스페이서 서열	인델%
일차 T 세포	TRAC	gRNA 1	TCTCTCAGCTGGTACACGGCA(서열번호 : 156)	18%
		gRNA 2	GCGTCATGAGCAGATTAAACC (서열번호: 157)	81%
		gRNA 3	TCTCGACCAGCTTGACATCAC (서열번호: 158)	10%
		gRNA 4	TTAAACCCGGCCACTTTCAGG (서열번호: 159)	46%
		gRNA 5	CTGTGC 태그 ACATGAGGTCTA (서열번호: 160)	26%
		gRNA 8	ACTTCAAGAGCAACAGTGCTG (서열번호: 161)	3%
		gRNA 9	AAGAGCAACAGTGCTGTGGCC (서열번호: 162)	13%
		gRNA 10	GCTGGGAAGAAGGTGCTTC (서열번호: 163)	7%
		gRNA 15	A 태그 GCAGACAGACTTGTCAC (서열번호: 164)	16%
		gRNA 17	태그 AGTCTCTCAGCTGGTACA (서열번호: 165)	23%
		gRNA 18	GTCTCTCAGCTGGTACACGGC (서열번호: 166)	5%
		gRNA 19	CAGCTGGTACACGGCAGGGTC (서열번호: 167)	11%
		gRNA 20	AGCTGGTACACGGCAGGGTCA (서열번호: 168)	13%
		gRNA 21	TACACGGCAGGGTCAGGGTTC (서열번호: 169)	19%
		gRNA 23	CTTTCAAAACCTGTCAGTGAT (서열번호: 170)	4%
		gRNA 25	TCCGAATCCTCCTCCTGAAAG (서열번호: 171)	21%
		gRNA 26	AATCCTCCTCCTGAAAGTGGC (서열번호: 172)	11%
		gRNA 27	ATCCTCCTCCTGAAAGTGGCC	9%

[0347]

			(서열번호: 173)	
		gRNA 29	CTGCTCATGACGCTGCGGCTG (서열번호: 174)	15%
		gRNA 30	AGATTAAACCGGCCACTTTC (서열번호: 175)	24%
		gRNA 31	AACCCGGCCACTTTCAGGAGG (서열번호: 176)	20%
		gRNA 32	GCCACTTTCAGGAGGAGGATT (서열번호: 177)	20%
	B2M	gRNA 1	TACTCTCTCTTCTGGCCTGG (서열번호: 178)	5%
		gRNA 2	GCATACTCATCTTTTCAGTG (서열번호: 179)	12%
		gRNA 3	CGCTACTCTCTTCTGGCC (서열번호: 180)	17%
		gRNA 4	GCGCGAGCACAGCTAAGGCCA (서열번호: 181)	64%
		gRNA 6	GCTCGGCTACTCTCTTTC (서열번호: 182)	9%
		gRNA 7	AGAG 태그 CGCGAGCACAGCTA (서열번호: 183)	61%
		gRNA 15	TCACAGCCCAAGA 태그 TTAAG (서열번호: 184)	45%
		gRNA 16	CACAGCCCAAGA 태그 TTAAGT (서열번호: 185)	42%
		gRNA 18	GACAAAGTCACATGGTTCACA (서열번호: 186)	43%
		gRNA 19	AAGTCACATGGTTCACGCGC (서열번호: 187)	8%
		gRNA 20	AGGCATACTCATCTTTTCAG (서열번호: 188)	37%
		gRNA 21	GGCATACTCATCTTTTCAGT (서열번호: 189)	33%
		gRNA 22	CATACTCATCTTTTCAGTGG (서열번호: 190)	20%
		gRNA 23	TCAGTAAGTCAACTCAATGT (서열번호: 191)	41%
		gRNA 26	ACGTGAGTAAACCTGAATCTT (서열번호: 192)	22%

[0348]

	ELANE	ELANEg35_	AGTCCGGGCTGGGAGCGGGT (서열번호: 193)	40.5%
U2OS	g35	OMNI-50		
	ELANE	ELANEg35_	AGTCCGGGCTGGGAGCGGGT (서열번호: 193)	95%
	g35	OMNI-50		
	ELANE	ELANEg38_	ACAGCGGGTG 태그 ACTCCGA (서열번호: 194)	35%
	g38	OMNI-50		
	ELANE	ELANEg39_	CAGCGGGTG 태그 ACTCCGAG (서열번호: 195)	75%
iPSC	g39	OMNI-50		
	ELANE	ELANEg58_	GCTGCGGAAAGGGATTCCC (서열번호: 196)	83%
	g58	OMNI-50		
	ELANE	ELANEg62_	GTCAAGCCCAGAGGCCACA (서열번호: 197)	86%
	g62	OMNI-50		
	ELANE	ELANEg35_	AGTCCGGGCTGGGAGCGGGT (서열번호: 193)	53%
HSC	g35	OMNI-50		
	ELANE	ELANEg35_	AGTCCGGGCTGGGAGCGGGT (서열번호: 193)	96%
	g35	OMNI-50		
	ELANE	ELANEg38_	ACAGCGGGTG 태그 ACTCCGA (서열번호: 194)	44%
	g38	OMNI-50		
	ELANE	ELANEg39_	CAGCGGGTG 태그 ACTCCGAG (서열번호: 195)	59%
	g39	OMNI-50		
	ELANE	ELANEg58_	GCTGCGGAAAGGGATTCCC (서열번호: 196)	22%
	g58	OMNI-50		
	ELANE	ELANEg62_	GTCAAGCCCAGAGGCCACA (서열번호: 197)	59%
	g62	OMNI-50		

[0349]

[0350]

표 7. RNP로서의 OMNI-50 활성: OMNI-50 RNP를 합성 sgRNA(Agilent)와 어셈블리하고 세포 내로 전기천공하였다. 몇몇 세포 유형을 다양한 sgRNA로 시험하였다. 세포성 시스템, 유전자 명칭, 및 스페이서 서열을 NGS에 의해 측정되는 편집 수준 다음에 나타낸다.

표 8

멀티플렉스화

유전자	부위	스페이서 서열	OMNI-50 편집		OMNI-50 STD	
			공여체 1	공여체 2	공여체 1	공여체 2
TRAC	gRNA 32	GCCACTTTCAGGAGG AGGATT(서열번호 : 177)	59.00	53.00	9.00	10.00
B2M	gRNA 15	TCACAGCCCAAGA 태그 TTAAG(서열 번호: 184)	42.00	44.00	13.00	19.00
TRAC + B2M	gRNA 32 + gRNA 15	TRAC 에 대한 시험	55.00	44.00	5.00	1.00
TRAC + B2M	gRNA 32 + gRNA 15	B2M 에 대한 시험	22.00	27.00	3.00	1.00

표 8. 일차 T 세포에서의 OMNI-50 멀티플렉스화: OMNI-50의 멀티플렉스화를 활성화된 일차 T 세포 내로 전기천공에 의해 수행하여, TRAC 또는 B2M 유전자를 표적화하거나, 조합 표적화하였다. 처음 2개 열은 5명의-공여체 은행으로부터 무작위 선택된 2명의 공여체 상에서 별도로 각각의 유전자를 나타낸다. 마지막 2개 열은 전기천공이 멀티플렉스로서 수행되었을 때 각각의 유전자에 대해 동일한 분석을 나타낸다. 편집 활성을 앰플리콘 기반 NGS 후의 인텔 카운트에 의해 결정하였다. 2회 결과의 표준 편차를 또한 나타낸다. TRAC gRNA만의 사용은 B2M 유전자 상에 효과를 갖지 않았고 반대도 마찬가지였다(나타내지 않음).

표 9

OMNI-50 오프-표적

가이드	SpCas9 #1	SpCas9 #2	OMNI-50 #1	OMNI-50 #2	SpCas9 오프-표적 편집	SpCas9 ODN 통합	OMNI-50 오프-표적 편집	OMNI-50 ODN 통합
ELANE g35	206	201	11	5	85%, 90%	58%, 62%	97%, 97%	37%, 38%
ELANE g58	51	92	18	13	86%, 86%	39%, 42%	88%, 83%	37%, 36%
ELANE g58_alt	67	N.A.	4	9	82%	34%	88%, 82%	34%, 39%
ELANE g62	17	12	12	15	88%, 90%	3%, 27%	89%, 89%	22%, 17%
ELANE g62_alt	18	13	5	9	1%, 2%	N.A.	0%, 0%	N.A.
TRAC g32	10	9	5	5	93%, 81%	14%	51%, 74%	31%, 26%

표 9. 비편향 생화학 검정(가이드 seq)에 의한 OMNI-50 오프-표적 분석: SpCas9 또는 OMNI-50 뉴클레아제의 오프-표적 부위 카운트를 2회 반복으로 나타낸다. 상기 분석에 있어서, 10회 이상 판독을 갖는 증폭 부위만을 분석하였으며, 더 낮은 판독 횟수를 갖는 부위는 배경 노이즈를 감소시키기 위해 폐기하였다. 앰플리콘 기반 NGS 후 인텔 카운트에 의해 결정된 오프-표적 부위에서의 편집 수준뿐만 아니라 ODN 통합을 또한 나타낸다.

표 10

OMNI-50 스페이서 최적화

	U2OS 세포주				HSC				iPSC			
	편집% ELANE g35	편집% g35 오프-표적	STD ELANE g35	STD 오프-표적	편집% ELANE g35	편집% g35 오프-표적	STD ELANE g35	STD 오프-표적	편집% ELANE g35	편집% g35 오프-표적	STD ELANE g35	STD 오프-표적
17bp	0.54	0.20	0.28	0.04	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
18bp	0.59	0.15	0.30	0.01	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.	N.A.
19bp	3.20	0.21	0.29	0.07	11.54	0.00	3.24	0.00	0.41	0.00	0.19	0.00
20bp	43.77	0.21	12.77	0.02	26.47	0.00	3.23	0.00	6.63	0.00	0.68	0.00
21bp	91.50	0.37	4.66	0.03	74.86	0.00	1.70	0.00	48.15	0.00	0.16	0.00
22bp	90.87	10.50	3.63	5.55	89.10	0.10	1.14	0.11	52.80	3.60	3.77	0.27
23bp	75.81	7.59	8.35	4.25	85.86	0.10	1.69	0.27	51.16	2.50	7.40	0.15

[0356] 표 10. OMNI-50 스페이서 최적화. RNP를 상이한 길이의 sgRNA를 이용하여 OMNI-50 단백질에 대해 어셈블리하였다. RNP를 U2OS, HSC, 및 iPSC 세포 내로 전기천공하고, 앰플리콘 기반 NGS 후 인텔 카운트에 의해 활성을 결정하였다. OMNI-50을 2회 반복으로 ELANE g35 상에서 시험하였다(표준 편차를 나타냄). 표 10 별첨은 21 nt 대 22 nt 스페이서의 상세한 비교가 HSC 내 5개의 상이한 게놈 부위에 걸쳐 수행되었음을 나타낸다.

[0357] [표 10 - 별첨]

HSC에서 21 nt 대 22 nt 스페이서의 비교

	21 nt 편집%	22 nt 편집%	21 nt STD	22 nt STD
ELANE g35	71.04	96.32	3.23	1.43
ELANE g38	9.79	43.89	1.26	0.53
ELANE g39	19.30	58.87	1.20	1.07
ELANE g58	11.02	21.86	2.83	0.40
ELANE g62	26.86	58.70	0.23	0.70

[0358]

[0359] 참고문헌

- [0360] 1. Ahmad and Allen (1992) "Antibody-mediated Specific Binging and Cytotoxicity of Liposome-entrapped Doxorubicin to Lung Cancer Cells *in Vitro*", Cancer Research 52:4817-20.
- [0361] 2. Anderson (1992) "Human gene therapy", Science 256:808-13.
- [0362] 3. Basha 등 (2011) "Influence of Cationic Lipid Composition on Gene Silencing Properties of Lipid Nanoparticle Formulations of siRNA in Antigen-Presenting Cells", Mol. Ther. 19(12):2186-200.
- [0363] 4. Behr (1994) "Gene transfer with synthetic cationic amphiphiles: Prospects for gene therapy", Bioconjugate Chem 5:382-89.
- [0364] 5. Blaese 등 (1995) "Vectors in cancer therapy: how will they deliver", Cancer Gene Ther. 2:291-97.
- [0365] 6. Blaese 등 (1995) "T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years", Science 270(5235):475-80.
- [0366] 7. Briner 등 (2014) "Guide RNA functional modules direct Cas9 activity and orthognality", Molecular Cell 56:333-39.
- [0367] 8. Buchschacher and Panganiban (1992) "Human immunodeficiency virus vectors for inducible expression of foreign genes", J. Virol. 66:2731-39.
- [0368] 9. Burstein 등 (2017) "New CRISPR-Cas systems from uncultivated microbes", Nature 542:237-41.
- [0369] 10. Canver 등, (2015) "BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis", Nature Vol. 527, Pgs. 192-214.
- [0370] 11. Chang and Wilson (1987) "Modification of DNA ends can decrease end-joining relative to homologous recombination in mammalian cells", Proc. Natl. Acad. Sci. USA 84:4959-4963.
- [0371] 12. Charlesworth 등 (2019) "Identification of preexisting adaptive immunity to Cas9 proteins in humans", Nature Medicine, 25(2), 249.
- [0372] 13. Chung 등 (2006) "Agrobacterium is not alone: gene transfer to plants by viruses and other bacteria", Trends Plant Sci. 11(1):1-4.
- [0373] 14. Coelho 등 (2013) "Safety and efficacy of RNAi therapy for transthyretin amyloidosis" N. Engl. J. Med. 369, 819-829.
- [0374] 15. Crystal (1995) "Transfer of genes to humans: early lessons and obstacles to success", Science 270(5235):404-10.
- [0375] 16. Dillon (1993) "Regulation gene expression in gene therapy" Trends in Biotechnology 11(5):167-173.

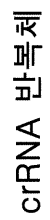
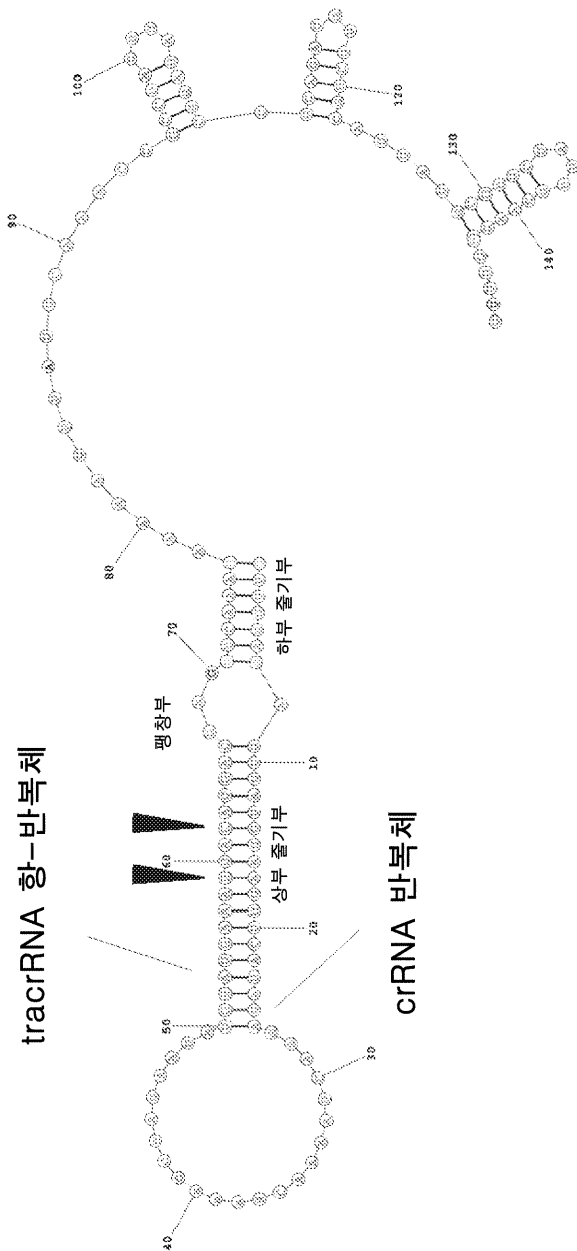
- [0376] 17. Dranoff 등 (1997) "A phase I study of vaccination with autologous, irradiated melanoma cells engineered to secrete human granulocyte macrophage colony stimulating factor", *Hum. Gene Ther.* 8(1):111-23.
- [0377] 18. Dunbar 등 (1995) "Retrovirally marked CD34-enriched peripheral blood and bone marrow cells contribute to long-term engraftment after autologous transplantation", *Blood* 85:3048-57.
- [0378] 19. Ellem 등 (1997) "A case report: immune responses and clinical course of the first human use of granulocyte/macrophage-colony-stimulating-factor-transduced autologous melanoma cells for immunotherapy", *Cancer Immunol Immunother* 44:10-20.
- [0379] 20. Gao and Huang (1995) "Cationic liposome-mediated gene transfer" *Gene Ther.* 2(10):710-22.
- [0380] 21. Haddada 등 (1995) "Gene Therapy Using Adenovirus Vectors", in: *The Molecular Repertoire of Adenoviruses III: Biology and Pathogenesis*, ed. Doerfler and Böhlm, pp. 297-306.
- [0381] 22. Han 등 (1995) "Ligand-directed retro-viral targeting of human breast cancer cells", *Proc. Natl. Acad. Sci. USA* 92(21):9747-51.
- [0382] 23. Humbert 등, (2019) "Therapeutically relevant engraftment of a CRISPR-Cas9-edited HSC-enriched population with HbF reactivation in nonhuman primates", *Sci. Trans. Med.*, Vol. 11, Pgs. 1-13.
- [0383] 24. Inaba 등 (1992) "Generation of large numbers of dendritic cells from mouse bone marrow cultures supplemented with granulocyte/macrophage colony-stimulating factor", *J Exp Med.* 176(6):1693-702.
- [0384] 25. Jiang and Doudna (2017) "CRISPR-Cas9 Structures and Mechanisms", *Annual Review of Biophysics* 46:505-29.
- [0385] 26. Jinek 등 (2012) "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity", *Science* 337(6096):816-21.
- [0386] 27. Johan 등 (1992) "GLVR1, a receptor for gibbon ape leukemia virus, is homologous to a phosphate permease of *Neurospora crassa* and is expressed at high levels in the brain and thymus", *J Virol* 66(3):1635-40.
- [0387] 28. Judge 등 (2006) "Design of noninflammatory synthetic siRNA mediating potent gene silencing in vivo", *Mol Ther.* 13(3):494-505.
- [0388] 29. Kohn 등 (1995) "Engraftment of gene-modified umbilical cord blood cells in neonates with adenosine deaminase deficiency", *Nature Medicine* 1:1017-23.
- [0389] 30. Kremer and Perricaudet (1995) "Adenovirus and adeno-associated virus mediated gene transfer", *Br. Med. Bull.* 51(1):31-44.
- [0390] 31. Macdiarmid 등 (2009) "Sequential treatment of drug-resistant tumors with targeted minicells containing siRNA or a cytotoxic drug", *Nat Biotechnol.* 27(7):643-51.
- [0391] 32. Malech 등 (1997) "Prolonged production of NADPH oxidase-corrected granulocytes after gene therapy of chronic granulomatous disease", *PNAS* 94(22):12133-38.
- [0392] 33. Maxwell 등 (2018) "A detailed cell-free transcription-translation-based assay to decipher CRISPR protospacer adjacent motifs", *Methods* 14348-57
- [0393] 34. Miller 등 (1991) "Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus", *J Virol.* 65(5):2220-24.
- [0394] 35. Miller (1992) "Human gene therapy comes of age", *Nature* 357:455-60.
- [0395] 36. Mir 등 (2019) "Type II-C CRISPR-Cas9 Biology, Mechanism and Application", *ACS Chem. Biol.* 13(2):357-365.
- [0396] 37. Mitani and Caskey (1993) "Delivering therapeutic genes - matching approach and application",

Trends in Biotechnology 11(5):162-66.

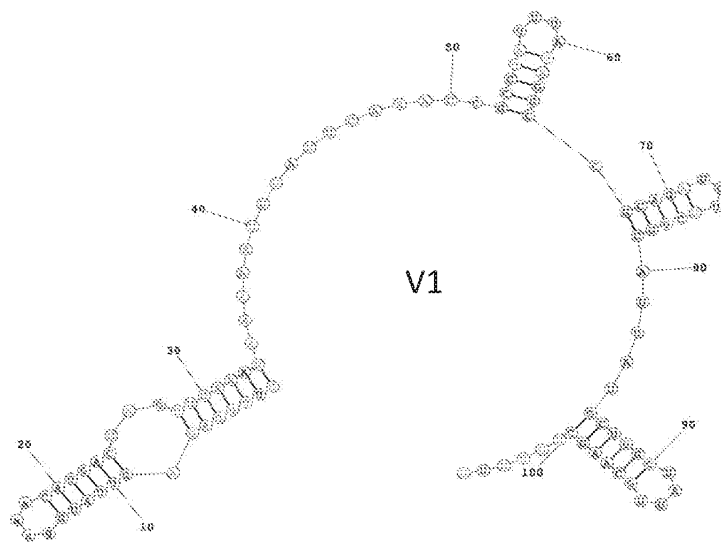
- [0397] 38. Nabel and Felgner (1993) "Direct gene transfer for immunotherapy and immunization", Trends in Biotechnology 11(5):211-15.
- [0398] 39. Nehls 등 (1996) "Two genetically separable steps in the differentiation of thymic epithelium" Science 272:886-889.
- [0399] 40. Nishimasu 등 "Crystal structure of Cas9 in complex with guide RNA and target DNA" (2014) Cell 156(5):935-49.
- [0400] 41. Nishimasu 등 (2015) "Crystal Structure of Staphylococcus aureus Cas9" Cell 162(5):1113-26.
- [0401] 42. Palermo 등 (2018) "Key role of the REC lobe during CRISPR-Cas9 activation by 'sensing', 'regulating', and 'locking' the catalytic HNH domain" Quarterly Reviews of Biophysics 51, e9, 1-11.
- [0402] 43. Remy 등 (1994) "Gene Transfer with a Series of Lipophilic DNA-Binding Molecules", Bioconjugate Chem. 5(6):647-54.
- [0403] 44. Sentmanat 등 (2018) "A Survey of Validation Strategies for CRISPR-Cas9 Editing", Scientific Reports 8:888, doi:10.1038/s41598-018-19441-8.
- [0404] 45. Sommerfelt 등 (1990) "Localization of the receptor gene for type D simian retroviruses on human chromosome 19", J. Virol. 64(12):6214-20.
- [0405] 46. Van Brunt (1988) "Molecular framing: transgenic animals as bioactors" Biotechnology 6:1149-54.
- [0406] 47. Vigne 등 (1995) "Third-generation adenovectors for gene therapy", Restorative Neurology and Neuroscience 8(1,2): 35-36.
- [0407] 48. Wagner 등 (2019) "High prevalence of Streptococcus pyogenes Cas9-reactive T cells within the adult human population" Nature Medicine, 25(2), 242
- [0408] 49. Wilson 등 (1989) "Formation of infectious hybrid virion with gibbon ape leukemia virus and human T-cell leukemia virus retroviral envelope glycoproteins and the *gag* and *pol* proteins of Moloney murine leukemia virus", J. Virol. 63:2374-78.
- [0409] 50. Yu 등 (1994) "Progress towards gene therapy for HIV infection", Gene Ther. 1(1):13-26.
- [0410] 51. Zetsche 등 (2015) "Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system" Cell 163(3):759-71.
- [0411] 52. Zuris 등 (2015) "Cationic lipid-mediated delivery of proteins enables efficient protein based genome editing in vitro and in vivo" Nat Biotechnol. 33(1):73-80.

도면

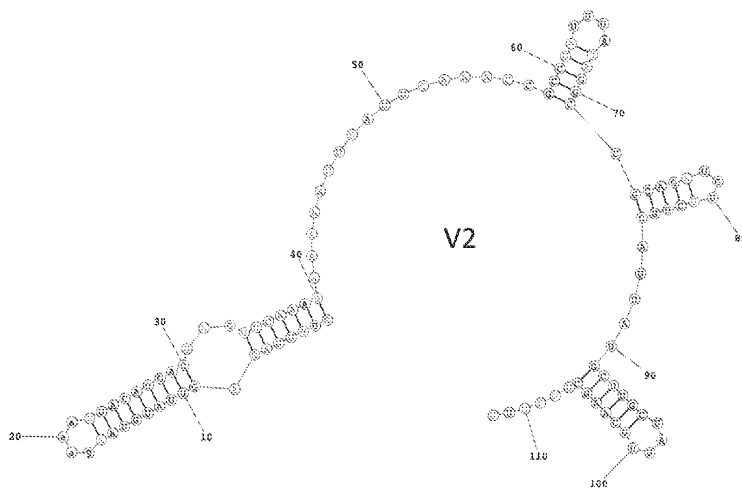
도면 1a



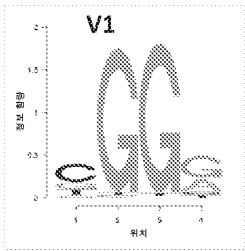
도면1b



도면1c

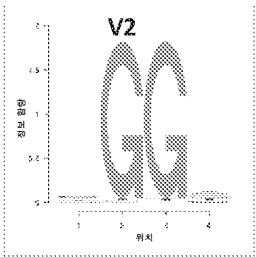


도면2a



위치 1-4								비
1	2	3	4	5	6	7	8	
C	G	G	G					0.01
C	G	G	A					0.02
	G	G	G	A				0.05
	G	G	A	G				0.05
	G	G	G	G				0.05
T	G	G	G					0.06
G	G	G	A					0.06
G	G	G	G					0.06
	G	G	G	C				0.06
	G	G	G	T				0.07
	G	G	A	T				0.08
T	G	G	A					0.11
	G	A	A	C				0.52
	G	G	C	C				0.55
	G	A	T	T				0.55
	G	G	A	G				0.57
	G	T	G	A				0.59
		A	T	A	G			0.59
	A	T	T	A				0.60
		T	C	G	G			0.61

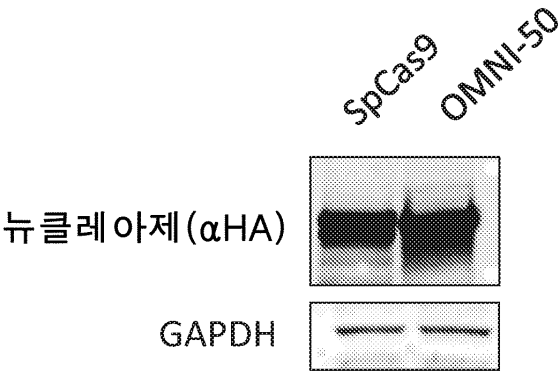
도면2b



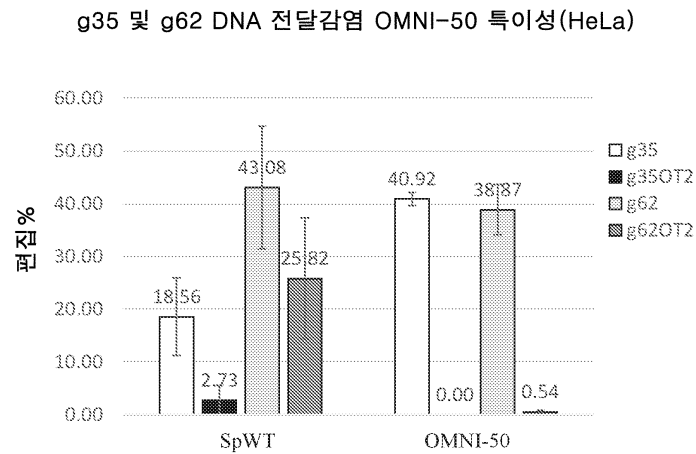
위치 1-4

1	2	3	4	5	6	7	8	비
C	G	G	A					0.02
	G	G	A	T				0.02
A	G	G	A					0.03
G	G	G	A					0.03
	G	G	G	T				0.03
	G	G	A	G				0.03
	G	G	A	C				0.03
C	G	G	T					0.03
	G	G	A	A				0.03
T	G	G	G					0.03
	G	G	T	C				0.03
C	G	G	G					0.04
	G	A	A	C				0.57
	G	G	C	C				0.59
	G	T	C	T				0.59
	G	A	T	T				0.60
	G	C	A	A				0.60
		A	G	G	G			0.60
	G	G	C	G				0.61

도면3



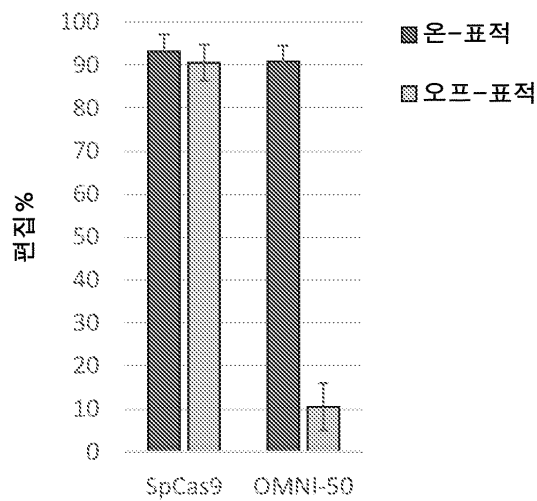
도면4a



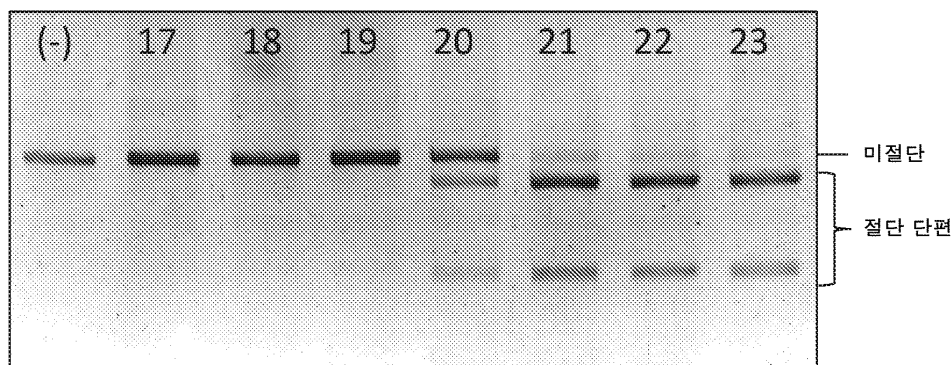
온-표적 g35 5' AGTCCGGGCTGGGAGCGGGTGGGGAGCA (서열번호: 130)
 온-표적 g62 5' GTCAAGCCCCAGAGGCCACAGGGACAGA (서열번호: 131)
 오프-표적 g35 5' AGTCCTGGCTGGGAGCAGGTGGGGAGAG (서열번호: 132)
 오프-표적 g62 5' GCCAAGCCTCAGAGGCCACAGGGCAGCA (서열번호: 133)

도면4b

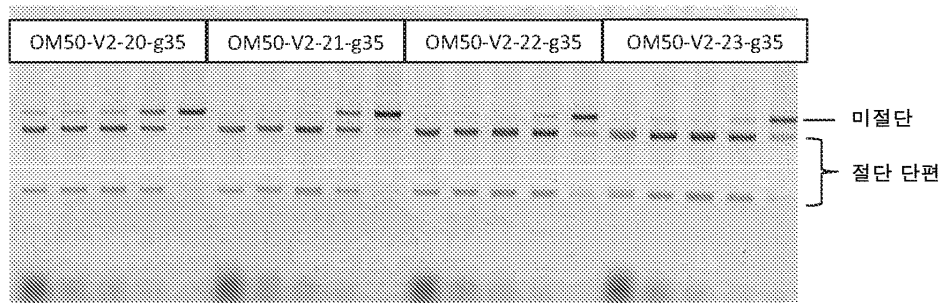
g35 RNP 전기천공에서의 OMNI-50 특이성(U2OS)



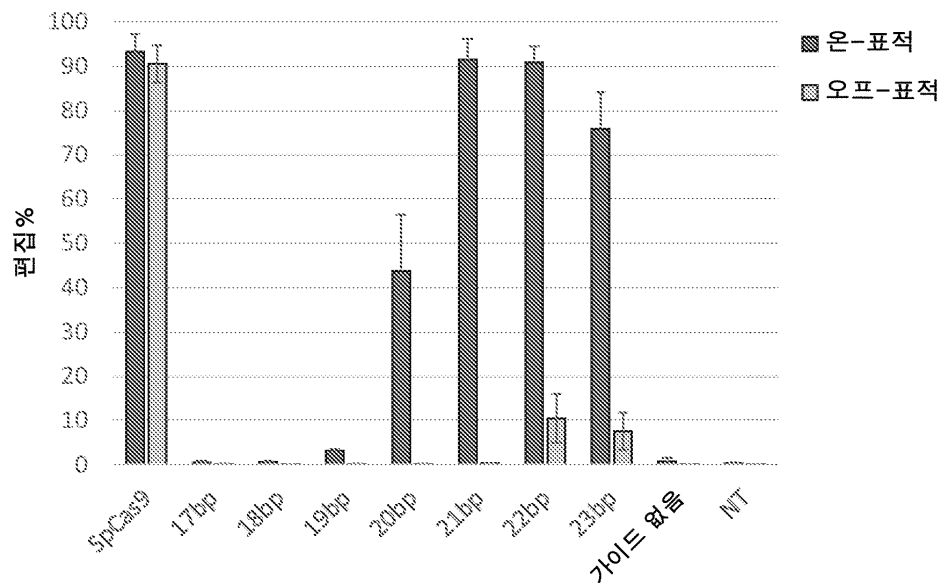
도면5a



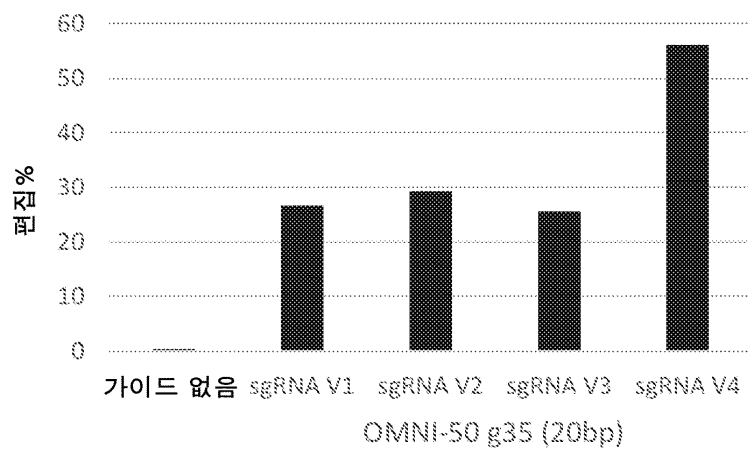
도면5b



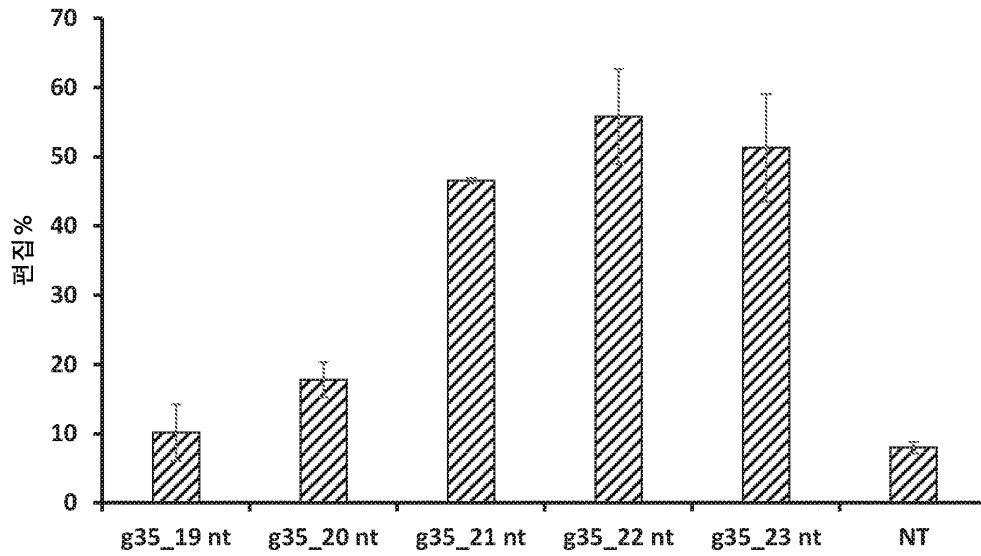
도면5c



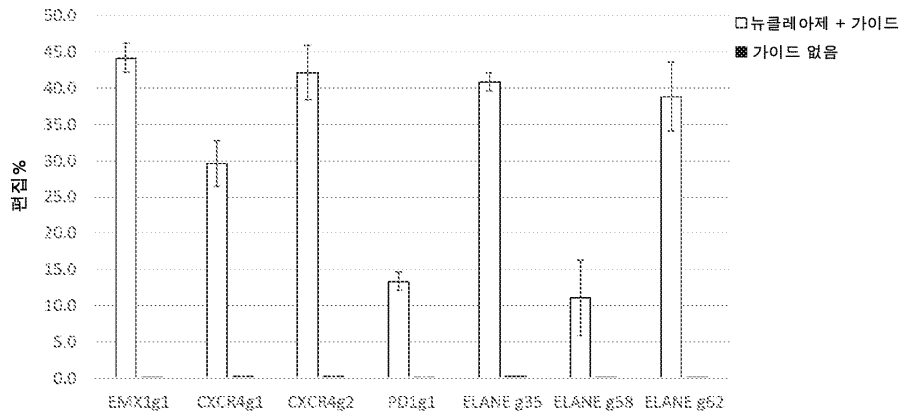
도면5d



도면6

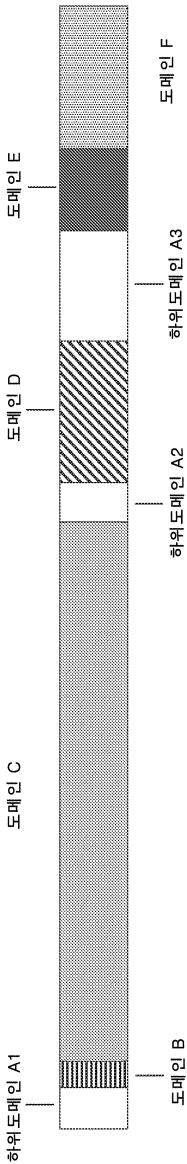


도면7



온 표적 EMX1: TCTGTGAATGTTAGACCCATGGGAGCAG	서열번호: 140
온 표적 CXCR4 1: GCAAGAGACCCACACACCGGAGGAGCGC	서열번호: 141
온 표적 CXCR4 2: ACACCGGAGGAGCGCCGCTTGGGGGAG	서열번호: 142
온 표적 PDCD1: CGTCTGGGCGGTGCTACAACCTGGGCTGG	서열번호: 143
온 표적 ELANE g35: AGTCCGGGCTGGGAGCGGGTGGGGAGCA	서열번호: 130
온 표적 ELANE g58: GCTGCGGGAAAGGGATTCCCTGGGACTC	서열번호: 144
온 표적 ELANE g62: GTCAAGCCCCAGAGGCCACAGGGACAGA	서열번호: 131

도면8



서열 목록

SEQUENCE LISTING

- <110> EmendoBio Inc.
Baram, David
Izhar, Lior
Herman, Asael
Rockah, Liat
Marbach Bar, Nadav
Meron, Nurit
Georgeson, Joseph
- <120> NOVEL OMNI-50 CRISPR NUCLEASE

<130> 91116-A-PCT/GJG/AWG
 <150> 62/991,285
 <151> 2020-03-18
 <150> 62/959,672
 <151> 2020-01-10
 <150> 62/931,630
 <151> 2019-11-06
 <150> 62/897,806
 <151> 2019-09-09
 <150> 62/841,016
 <151> 2019-04-30
 <160> 197
 <170> PatentIn version 3.5

<210> 1
 <211> 1100
 <212> PRT
 <213> Butyrivibrio sp. AC2005
 <400> 1

Met Gly Tyr Thr Ile Gly Leu Asp Leu Gly Val Ala Ser Leu Gly Trp
 1 5 10 15
 Ala Val Val Asn Asp Glu Tyr Glu Val Leu Glu Ser Cys Ser Asn Ile
 20 25 30
 Phe Pro Ala Ala Glu Ser Ala Asn Asn Val Glu Arg Arg Gly Phe Arg
 35 40 45
 Gln Gly Arg Arg Leu Ser Arg Arg Arg Arg Thr Arg Ile Ser Asp Phe
 50 55 60
 Arg Lys Leu Trp Glu Lys Ser Gly Phe Glu Val Pro Ser Asn Glu Leu
 65 70 75 80
 Asn Glu Val Leu Gln Tyr Arg Ile Lys Gly Met Asn Asp Lys Leu Ser
 85 90 95
 Glu Asp Glu Leu Tyr His Val Leu Leu Asn Ser Leu Lys His Arg Gly
 100 105 110

Ile Ser Tyr Leu Asp Asp Ala Asp Asp Glu Asn Ala Ser Gly Asp Tyr

115 120 125

Ala Ala Ser Ile Ala Tyr Asn Glu Asn Gln Leu Lys Thr Lys Leu Pro

130 135 140

Cys Glu Ile Gln Trp Glu Arg Tyr Lys Lys Tyr Gly Ala Tyr Arg Gly

145 150 155 160

Asn Ile Thr Ile Gln Glu Gly Gly Glu Pro Leu Thr Leu Arg Asn Val

165 170 175

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

180 185 190

Gln Ser Met Ser Asn Glu Lys Val Thr Lys Lys Phe Ile Asp Glu Tyr

195 200 205

Leu Lys Ile Phe Ser Arg Lys Arg Glu Tyr Tyr Ile Gly Pro Gly Asn

210 215 220

Lys Lys Ser Arg Thr Asp Tyr Gly Val Tyr Thr Thr Gln Lys Asn Glu

225 230 235 240

Asp Gly Thr Tyr His Thr Glu Gln Asn Leu Phe Asp Lys Leu Ile Gly

245 250 255

Lys Cys Ser Val Tyr Pro Asp Glu Arg Arg Ala Ala Gly Ala Thr Tyr

260 265 270

Thr Ala Gln Glu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Val Ile

275 280 285

Asp Gly Arg Lys Leu Asp Glu Gln Glu Lys Cys Gln Ile Val Asp Ala

290 295 300

Val Lys His Ala Lys Thr Val Asn Met Lys Asn Ile Ile Ala Lys Val

305 310 315 320

Ile Gly Thr Lys Ala Asn Ser Met Asn Met Thr Gly Ala Arg Ile Asp

325 330 335

Lys Asn Glu Lys Glu Ile Phe His Ser Phe Glu Ala Tyr Asn Lys Leu

340 345 350

Arg Lys Ala Leu Glu Glu Ile Asp Phe Asp Ile Glu Thr Leu Ser Thr

355 360 365
Asp Glu Leu Asp Ala Ile Gly Glu Val Leu Thr Leu Asn Thr Asp Arg

370 375 380
Lys Ser Ile Gln Asn Gly Leu Gln Glu Lys Arg Ile Val Val Pro Asp
385 390 395 400
Glu Val Arg Asp Val Leu Ile Ala Thr Arg Lys Arg Asn Gly Ser Leu
405 410 415
Phe Ser Lys Trp Gln Ser Phe Gly Ile Arg Ile Met Lys Glu Leu Ile
420 425 430
Pro Glu Leu Tyr Ala Gln Pro Lys Asn Gln Met Gln Leu Leu Thr Asp

435 440 445
Met Gly Val Phe Lys Thr Lys Asp Glu Arg Phe Val Glu Tyr Asp Lys
450 455 460
Ile Pro Ser Asp Leu Ile Thr Glu Glu Ile Tyr Asn Pro Val Val Ala
465 470 475 480
Lys Thr Val Arg Ile Thr Val Arg Val Leu Asn Ala Leu Ile Lys Lys
485 490 495
Tyr Gly Tyr Pro Asp Arg Val Val Ile Glu Met Pro Arg Asp Lys Asn

500 505 510
Ser Glu Glu Glu Lys Lys Arg Ile Ala Asp Phe Gln Lys Asn Asn Glu
515 520 525
Asn Glu Leu Gly Gly Ile Ile Lys Lys Val Lys Ser Glu Tyr Gly Ile
530 535 540
Glu Ile Thr Asp Ala Asp Phe Lys Asn His Ser Lys Leu Gly Leu Lys
545 550 555 560
Leu Arg Leu Trp Asn Glu Gln Asn Glu Thr Cys Pro Tyr Ser Gly Lys

565 570 575
His Ile Lys Ile Asp Asp Leu Leu Asn Asn Pro Asn Met Phe Glu Val
580 585 590
Asp His Ile Ile Pro Leu Ser Ile Ser Phe Asp Asp Ser Arg Ala Asn
595 600 605

Lys Val Leu Val Tyr Ala Ala Glu Asn Gln Asn Lys Gly Asn Arg Thr
 610 615 620
 Pro Met Ala Tyr Leu Ser Asn Val Asn Arg Glu Trp Asp Phe His Glu

 625 630 635 640
 Tyr Met Ser Phe Val Leu Ser Asn Tyr Lys Gly Thr Ile Tyr Gly Lys
 645 650 655
 Lys Arg Asp Asn Leu Leu Phe Ser Glu Asp Ile Tyr Lys Ile Asp Val
 660 665 670
 Leu Gln Gly Phe Ile Ser Arg Asn Ile Asn Asp Thr Arg Tyr Ala Ser
 675 680 685
 Lys Val Ile Leu Asn Ser Leu Gln Ser Phe Phe Gly Ser Lys Glu Cys

 690 695 700
 Asp Thr Lys Val Lys Val Val Arg Gly Thr Phe Thr His Gln Met Arg
 705 710 715 720
 Met Asn Leu Lys Ile Glu Lys Asn Arg Glu Glu Ser Tyr Val His His
 725 730 735
 Ala Val Asp Ala Met Leu Ile Ala Phe Ser Gln Met Gly Tyr Asp Ala
 740 745 750
 Tyr His Lys Leu Thr Glu Lys Tyr Ile Asp Tyr Glu His Gly Glu Phe

 755 760 765
 Val Asp Gln Lys Gly Tyr Glu Lys Leu Ile Glu Asn Asp Val Ala Tyr
 770 775 780
 Arg Glu Thr Thr Tyr Gln Asn Lys Trp Met Thr Ile Lys Lys Asn Ile
 785 790 795 800
 Glu Ile Ala Ala Glu Lys Asn Lys Tyr Trp Tyr Gln Val Asn Arg Lys
 805 810 815
 Ser Asn Arg Gly Leu Cys Asn Gln Thr Ile Tyr Gly Thr Arg Asn Leu

 820 825 830
 Asp Gly Lys Thr Val Lys Ile Ser Lys Leu Asp Ile Arg Thr Asp Asp
 835 840 845
 Gly Ile Lys Lys Phe Lys Gly Ile Val Glu Lys Gly Lys Leu Glu Arg

850	855	860	
Phe Leu Met Tyr Arg Asn Asp Pro Lys Thr Phe Glu Trp Leu Leu Gln			
865	870	875	880
Ile Tyr Lys Asp Tyr Ser Asp Ser Lys Asn Pro Phe Val Gln Tyr Glu			
	885	890	895
Ser Glu Thr Gly Asp Val Ile Lys Lys Val Ser Lys Thr Asn Asn Gly			
	900	905	910
Pro Lys Val Cys Glu Leu Arg Tyr Glu Asp Gly Glu Val Gly Ser Cys			
	915	920	925
Ile Asp Ile Ser His Lys Tyr Gly Tyr Lys Lys Gly Ser Lys Lys Val			
	930	935	940
Ile Leu Asp Ser Leu Asn Pro Tyr Arg Met Asp Val Tyr Tyr Asn Thr			
945	950	955	960
Lys Asp Asn Arg Tyr Tyr Phe Val Gly Val Lys Tyr Ser Asp Ile Lys			
	965	970	975
Cys Gln Gly Asp Ser Tyr Val Ile Asp Glu Asp Lys Tyr Ala Ala Ala			
	980	985	990
Leu Val Gln Glu Lys Ile Val Pro Glu Gly Lys Gly Arg Ser Asp Leu			
	995	1000	1005
Thr Glu Leu Gly Tyr Glu Phe Lys Leu Ser Phe Tyr Lys Asn Glu			
1010	1015	1020	
Ile Ile Glu Tyr Glu Lys Asp Gly Glu Ile Tyr Val Glu Arg Phe			
	1025	1030	1035
Leu Ser Arg Thr Met Pro Lys Val Ser Asn Tyr Ile Glu Thr Lys			
	1040	1045	1050
Pro Leu Glu Ala Ala Lys Phe Glu Lys Arg Asn Leu Val Gly Leu			
	1055	1060	1065
Ala Lys Thr Ser Arg Ile Arg Lys Ile Arg Val Asp Ile Leu Gly			
1070	1075	1080	
Asn Arg Tyr Leu Asn Ser Met Glu Asn Phe Asp Phe Val Val Gly			
	1085	1090	1095

His Lys

1100

<210> 2

<211> 1104

<212> PRT

<213> bacterium LF-3

<400> 2

Met Ser Arg Tyr Val Leu Gly Leu Asp Ile Gly Ile Thr Ser Val Gly

1 5 10 15

Tyr Gly Val Ile Asp Ile Asp Asn Asn Leu Phe Val Asp Tyr Gly Val

20 25 30

Arg Leu Phe Lys Glu Gly Thr Ala Ala Glu Asn Glu Thr Arg Arg Thr

35 40 45

Lys Arg Gly Ser Arg Arg Leu Lys Arg Arg Lys Ser Asn Arg Leu Asn

50 55 60

Asp Met Lys Asn Leu Leu Lys Glu Asn Asp Leu Tyr Phe Glu Asp Tyr

65 70 75 80

Arg Asn Tyr Asn Pro Tyr Glu Ile Arg Ala Lys Gly Leu Lys Glu Lys

85 90 95

Leu Leu Pro Glu Glu Leu Cys Thr Ala Ile Met His Ile Thr Lys Ser

100 105 110

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

115 120 125

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

130 135 140

Lys Tyr Ile Cys Glu Val Gln Leu Asp Arg Leu Asn Lys Asp His Lys

145 150 155 160

Val Arg Gly Thr Glu Asn Asn Phe Lys Thr Glu Asp Tyr Val Lys Glu

165 170 175

Leu Lys Glu Ile Leu Lys His Gln Asp Leu Asn Glu Glu Leu Cys Asp

180 185 190

Gln Ile Ile Glu Met Val Ser Arg Arg Arg Arg Tyr Asp Gln Gly Pro

195	200	205	
Gly Ser Glu Lys Ser Pro Thr Pro Tyr Gly Ser Tyr Arg Met Val Asp			
210	215	220	
Gly Val Leu Lys His Val Asn Leu Ile Asp Glu Met Arg Gly Arg Cys			
225	230	235	240
Ser Val Tyr Pro Asp Glu Phe Arg Ala Pro Lys Gln Ser Tyr Thr Ala			
245	250	255	
Glu Leu Phe Asn Leu Leu Asn Asp Leu Asn Asn Leu Thr Ile Lys Gly			
260	265	270	
Glu Lys Ile Thr Val Glu Glu Lys Glu Lys Val Val Ala Phe Val Asn			
275	280	285	
Glu Lys Gly Ser Ile Thr Val Lys Gln Leu Leu Lys Leu Leu Asp Ala			
290	295	300	
Gln Glu Asp Glu Val Thr Gly Phe Arg Ile Asp Lys Asn Asp Lys Pro			
305	310	315	320
Leu Ile Thr Glu Phe Lys Gly Tyr Ser Lys Val Leu Lys Val Phe Lys			
325	330	335	
Lys Tyr Asn Gln Gln Glu Leu Leu Glu Asp Lys Leu Ile Val Asp Gln			
340	345	350	
Val Ile Asp Ile Cys Thr Lys Ser Lys Gly Ile Asp Glu Arg Lys Lys			
355	360	365	
Asp Ile Lys Glu Leu Tyr Pro Glu Phe Asp Asn Glu Leu Ile Glu Glu			
370	375	380	
Leu Ala Ser Val Lys Gly Val Ser Ala Tyr His Ser Leu Ser Phe Lys			
385	390	395	400
Ala Met His Ile Ile Asn Lys Glu Met Leu Thr Thr Glu Met Asn Gln			
405	410	415	
Ile Gln Val Leu His Glu Ile Glu Met Phe Asp Lys Asn Arg Lys Ser			
420	425	430	
Leu Lys Gly Lys Lys Asn Ile Glu Pro Asp Glu Glu Ala Ile Leu Ser			
435	440	445	

Pro Val Ala Lys Arg Ala His Arg Glu Thr Phe Lys Val Ile Asn Ala
450 455 460

Leu Arg Lys Gln Tyr Gly Glu Phe Asp Ser Ile Val Ile Glu Met Thr
465 470 475 480

Arg Asp Lys Asn Ser Lys Glu Gln Val Lys Arg Ile Asn Asp Ser Gln
485 490 495

Lys Arg Phe Lys Ser Glu Asn Asp Arg Val Asp Gly Ile Ile Lys Asn
500 505 510

Ser Gly Ile Asp Pro Glu Arg Val Asn Gly Lys Thr Lys Thr Lys Ile
515 520 525

Arg Leu Tyr Leu Gln Gln Asp Cys Lys Thr Ala Tyr Thr Gln Gln Asp
530 535 540

Ile Asp Leu His Thr Leu Ile Phe Asp Asp Lys Ala Tyr Glu Ile Asp
545 550 555 560

His Ile Ile Pro Ile Ser Val Ser Leu Asp Asp Ser Leu Thr Asn Lys
565 570 575

Val Leu Ala Ser Arg Leu Glu Asn Gln Gln Lys Gly Asn Leu Thr Pro
580 585 590

Met Met Ala Tyr Leu Lys Gly Lys Phe Thr Gly Gly Asn Leu Glu Lys
595 600 605

Tyr Lys Leu Phe Val Ser Ser Asn Lys Asn Phe Asn Gly Lys Lys Arg
610 615 620

Asn Asn Leu Leu Thr Glu Gln Asp Ile Thr Lys Glu Asp Val Ala Arg
625 630 635 640

Lys Phe Ile Asn Arg Asn Leu Val Asp Thr Ser Tyr Ala Cys Arg Thr
645 650 655

Val Leu Asn Thr Leu Gln Arg Tyr Phe Lys Asp Asn Glu Ile Asp Thr
660 665 670

Lys Val His Thr Ile Arg Gly Gln Ser Thr Asn Ile Phe Arg Lys Arg
675 680 685

Ile Asn Leu Gln Lys Asp Arg Glu Gln Asp Tyr Phe His His Ala Ile

690	695	700	
Asp Ala Leu Ile Val Ala Ser Leu Lys Lys Met Asn Ile Val Asn Ser			
705	710	715	720
Tyr Leu Met His Tyr Asn Tyr Ser Asp Leu Tyr Asp Glu Glu Thr Gly			
	725	730	735
Glu Val Phe Asp Val Leu Pro Asp Lys Gln Phe Ile Asp Gln Arg Tyr			
	740	745	750
Ile Ser Phe Ile Ser Asp Leu Lys Asn Ile Tyr Gln Glu Ser Asn Gln			
	755	760	765
Tyr Asn Leu Gly Tyr Ile Thr Gln Glu Gln Met His Tyr Pro Leu Ile			
	770	775	780
Lys Val Ser His Lys Ile Asp Thr Lys Pro Asn Arg Lys Ile Ala Asp			
785	790	795	800
Glu Thr Ile Tyr Ser Thr Arg Asn Ile Glu Gly Gln Asp Met Leu Val			
	805	810	815
Glu Lys Ile Lys Asn Ile Tyr Asp Pro Lys Glu Lys Lys Ala Ile Glu			
	820	825	830
Leu Val Asn Asn Ile Ile Asn Asp Asp Thr Asp Lys Tyr Ile Met Lys			
	835	840	845
His Lys Asp Pro Gln Thr Phe Glu Lys Ile Lys Glu Val Val Leu Asn			
	850	855	860
His Phe Asn Asp Tyr Lys Asp Ser Lys Glu Tyr Tyr Val Ile Asp Lys			
865	870	875	880
Lys Gly Lys Tyr Ser Leu Lys Glu Glu Ser Pro Leu Thr Ser Tyr Tyr			
	885	890	895
Asn Glu Asn Gly Ala Ile Thr Lys Tyr Ser Lys Lys Asn Asn Gly Pro			
	900	905	910
Ala Ile Thr Ser Met Lys Phe Tyr Ser Glu Lys Leu Gly Asn His Leu			
	915	920	925
Ala Ile Thr Ser Asn Tyr Asn Thr Asn Asn Lys Lys Val Ile Leu Lys			
	930	935	940

Gln Ile Ser Pro Tyr Arg Thr Asp Phe Tyr Val Ser Pro Glu Gly Lys
 945 950 955 960
 Tyr Lys Phe Val Thr Val Arg Tyr Lys Asp Val Phe Tyr Lys Glu Thr
 965 970 975
 Ile His Lys Phe Val Ile Asp Glu Asn Trp Tyr His Glu Glu Lys Ile
 980 985 990

Lys Lys Gly Ile Leu Glu Asp Trp Lys Phe Val Cys Ser Met His Arg
 995 1000 1005
 Asp Glu Leu Ile Gly Leu Ile Lys Pro Glu Gly Lys Lys Phe Val
 1010 1015 1020
 Tyr Asp Ala Ser Ile Asn Gly Gly Gln Thr Gln Tyr His Asp Gly
 1025 1030 1035
 Lys His Tyr Glu Ile Leu Lys Phe Thr Ala Thr Asn Asp Glu Lys
 1040 1045 1050

Lys Arg Thr Phe Glu Val Lys Pro Ile Asn Thr Asn Cys Ser Lys
 1055 1060 1065
 Arg Leu Met Pro Ser Val Gly Pro Phe Ile Lys Ile Gln Lys Phe
 1070 1075 1080
 Ala Thr Asp Val Leu Gly Asn Ile Tyr Glu Val Lys Asp Asn Arg
 1085 1090 1095
 Leu Lys Leu Glu Phe Asp
 1100

<210> 3

<211> 1370

<212> PRT

<213> Ezakiella peruensis strain M6.X2

<400> 3

Met Thr Lys Val Lys Asp Tyr Tyr Ile Gly Leu Asp Ile Gly Thr Ser
 1 5 10 15
 Ser Val Gly Trp Ala Val Thr Asp Glu Ala Tyr Asn Val Leu Lys Phe
 20 25 30
 Asn Ser Lys Lys Met Trp Gly Val Arg Leu Phe Asp Asp Ala Lys Thr

35	40	45
Ala Glu Glu Arg Arg Gly Gln Arg Gly Ala Arg Arg Arg Leu Asp Arg		
50	55	60
Lys Lys Glu Arg Leu Ser Leu Leu Gln Asp Phe Phe Ala Glu Glu Val		
65	70	75
Ala Lys Val Asp Pro Asn Phe Phe Leu Arg Leu Asp Asn Ser Asp Leu		
85	90	95
Tyr Met Glu Asp Lys Asp Gln Lys Leu Lys Ser Lys Tyr Thr Leu Phe		
100	105	110
Asn Asp Lys Asp Phe Lys Asp Lys Asn Phe His Lys Lys Tyr Pro Thr		
115	120	125
Ile His His Leu Leu Met Asp Leu Ile Glu Asp Asp Ser Lys Lys Asp		
130	135	140
Ile Arg Leu Val Tyr Leu Ala Cys His Tyr Leu Leu Lys Asn Arg Gly		
145	150	155
His Phe Ile Phe Glu Gly Gln Lys Phe Asp Thr Lys Ser Ser Phe Glu		
165	170	175
Asn Ser Leu Asn Glu Leu Lys Val His Leu Asn Asp Glu Tyr Gly Leu		
180	185	190
Asp Leu Glu Phe Asp Asn Glu Asn Leu Ile Asn Ile Leu Thr Asp Pro		
195	200	205
Lys Leu Asn Lys Thr Ala Lys Lys Lys Glu Leu Lys Ser Val Ile Gly		
210	215	220
Asp Thr Lys Phe Leu Lys Ala Val Ser Ala Ile Met Ile Gly Ser Ser		
225	230	235
Gln Lys Leu Val Asp Leu Phe Glu Asn Pro Glu Asp Phe Asp Asp Ser		
245	250	255
Ala Ile Lys Ser Val Asp Phe Ser Thr Thr Ser Phe Asp Asp Lys Tyr		
260	265	270
Ser Asp Tyr Glu Leu Ala Leu Gly Asp Lys Ile Ala Leu Val Asn Ile		
275	280	285

Leu Lys Glu Ile Tyr Asp Ser Ser Ile Leu Glu Asn Leu Leu Lys Glu
 290 295 300
 Ala Asp Lys Ser Lys Asp Gly Asn Lys Tyr Ile Ser Asn Ala Phe Val
 305 310 315 320

 Lys Lys Tyr Asn Lys His Gly Gln Asp Leu Lys Glu Phe Lys Arg Leu
 325 330 335
 Val Arg Gln Tyr His Lys Ser Ala Tyr Phe Asp Ile Phe Arg Ser Glu
 340 345 350
 Lys Val Asn Asp Asn Tyr Val Ser Tyr Thr Lys Ser Ser Ile Ser Asn
 355 360 365
 Asn Lys Arg Val Lys Ala Asn Lys Phe Thr Asp Gln Glu Ala Phe Tyr
 370 375 380

 Lys Phe Ala Lys Lys His Leu Glu Thr Ile Lys Tyr Lys Ile Asn Lys
 385 390 395 400
 Val Asn Gly Ser Lys Ala Asp Leu Glu Leu Ile Asp Gly Met Leu Arg
 405 410 415
 Asp Met Glu Phe Lys Asn Phe Met Pro Lys Ile Lys Ser Ser Asp Asn
 420 425 430
 Gly Val Ile Pro Tyr Gln Leu Lys Leu Met Glu Leu Asn Lys Ile Leu
 435 440 445

 Glu Asn Gln Ser Lys His His Glu Phe Leu Asn Val Ser Asp Glu Tyr
 450 455 460
 Gly Ser Val Cys Asp Lys Ile Ala Ser Ile Met Glu Phe Arg Ile Pro
 465 470 475 480
 Tyr Tyr Val Gly Pro Leu Asn Pro Asn Ser Lys Tyr Ala Trp Ile Lys
 485 490 495
 Lys Gln Lys Asp Ser Glu Ile Thr Pro Trp Asn Phe Lys Asp Val Val
 500 505 510

 Asp Leu Asp Ser Ser Arg Glu Glu Phe Ile Asp Ser Leu Ile Gly Arg
 515 520 525
 Cys Thr Tyr Leu Lys Asp Glu Lys Val Leu Pro Lys Ala Ser Leu Leu

530	535	540	
Tyr Asn Glu Tyr Met Val Leu Asn Glu Leu Asn Asn Leu Lys Leu Asn			
545	550	555	560
Asp Leu Pro Ile Thr Glu Glu Met Lys Lys Lys Ile Phe Asp Gln Leu			
	565	570	575
Phe Lys Thr Arg Lys Lys Val Thr Leu Lys Ala Val Ala Asn Leu Leu			
	580	585	590
Lys Lys Glu Phe Asn Ile Asn Gly Glu Ile Leu Leu Ser Gly Thr Asp			
	595	600	605
Gly Asp Phe Lys Gln Gly Leu Asn Ser Tyr Asn Asp Phe Lys Ala Ile			
	610	615	620
Val Gly Asp Lys Val Asp Ser Asp Asp Tyr Arg Asp Lys Ile Glu Glu			
625	630	635	640
Ile Ile Lys Leu Ile Val Leu Tyr Gly Asp Asp Lys Ser Tyr Leu Gln			
	645	650	655
Lys Lys Ile Lys Ala Gly Tyr Gly Lys Tyr Phe Thr Asp Ser Glu Ile			
	660	665	670
Lys Lys Met Ala Gly Leu Asn Tyr Lys Asp Trp Gly Arg Leu Ser Lys			
	675	680	685
Lys Leu Leu Thr Gly Leu Glu Gly Ala Asn Lys Ile Thr Gly Glu Arg			
	690	695	700
Gly Ser Ile Ile His Phe Met Arg Glu Tyr Asn Leu Asn Leu Met Glu			
705	710	715	720
Leu Met Ser Ala Ser Phe Thr Phe Thr Glu Glu Ile Gln Lys Leu Asn			
	725	730	735
Pro Val Asp Asp Arg Lys Leu Ser Tyr Glu Met Val Asp Glu Leu Tyr			
	740	745	750
Leu Ser Pro Ser Val Lys Arg Met Leu Trp Gln Ser Leu Arg Ile Val			
	755	760	765
Asp Glu Ile Lys Asn Ile Met Gly Thr Asp Ser Lys Lys Ile Phe Ile			
	770	775	780

Glu Met Ala Arg Gly Lys Glu Glu Val Lys Ala Arg Lys Glu Ser Arg
785 790 795 800
Lys Asn Gln Leu Leu Lys Phe Tyr Lys Asp Gly Lys Lys Ala Phe Ile
805 810 815
Ser Glu Ile Gly Glu Glu Arg Tyr Ser Tyr Leu Leu Ser Glu Ile Glu
820 825 830

Gly Glu Glu Glu Asn Lys Phe Arg Trp Asp Asn Leu Tyr Leu Tyr Tyr
835 840 845
Thr Gln Leu Gly Arg Cys Met Tyr Ser Leu Glu Pro Ile Asp Ile Ser
850 855 860
Glu Leu Ser Ser Lys Asn Ile Tyr Asp Gln Asp His Ile Tyr Pro Lys
865 870 875 880
Ser Lys Ile Tyr Asp Asp Ser Ile Glu Asn Arg Val Leu Val Lys Lys
885 890 895

Asp Leu Asn Ser Lys Lys Gly Asn Ser Tyr Pro Ile Pro Asp Glu Ile
900 905 910
Leu Asn Lys Asn Cys Tyr Ala Tyr Trp Lys Ile Leu Tyr Asp Lys Gly
915 920 925
Leu Ile Gly Gln Lys Lys Tyr Thr Arg Leu Thr Arg Arg Thr Gly Phe
930 935 940
Thr Asp Asp Glu Leu Val Gln Phe Ile Ser Arg Gln Ile Val Glu Thr
945 950 955 960

Arg Gln Ala Thr Lys Glu Thr Ala Asn Leu Leu Lys Thr Ile Cys Lys
965 970 975
Asn Ser Glu Ile Val Tyr Ser Lys Ala Glu Asn Ala Ser Arg Phe Arg
980 985 990
Gln Glu Phe Asp Ile Val Lys Cys Arg Ala Val Asn Asp Leu His His
995 1000 1005
Met His Asp Ala Tyr Ile Asn Ile Ile Val Gly Asn Val Tyr Asn
1010 1015 1020

Thr Lys Phe Thr Lys Asp Pro Met Asn Phe Val Lys Lys Gln Glu

1025	1030	1035
Lys Ala Arg Ser Tyr Asn Leu	Glu Asn Met Phe Lys	Tyr Asp Val
1040	1045	1050
Lys Arg Gly Gly Tyr Thr Ala	Trp Ile Ala Asp Asp	Glu Lys Gly
1055	1060	1065
Thr Val Lys Asn Ala Ser Ile	Lys Arg Ile Arg Lys	Glu Leu Glu
1070	1075	1080
Gly Thr Asn Tyr Arg Phe Thr	Arg Met Asn Tyr Ile	Glu Ser Gly
1085	1090	1095
Ala Leu Phe Asn Ala Thr Leu	Gln Arg Lys Asn Lys	Gly Ser Arg
1100	1105	1110
Pro Leu Lys Asp Lys Gly Pro	Lys Ser Ser Ile Glu	Lys Tyr Gly
1115	1120	1125
Gly Tyr Thr Asn Ile Asn Lys	Ala Cys Phe Ala Val	Leu Asp Ile
1130	1135	1140
Lys Ser Lys Asn Lys Ile Glu	Arg Lys Leu Met Pro	Val Glu Arg
1145	1150	1155
Glu Ile Tyr Ala Lys Gln Lys	Asn Asp Lys Lys Leu	Ser Asp Glu
1160	1165	1170
Ile Phe Ser Lys Tyr Leu Lys	Asp Arg Phe Gly Ile	Glu Asp Tyr
1175	1180	1185
Arg Val Val Tyr Pro Val Val	Lys Met Arg Thr Leu	Leu Lys Ile
1190	1195	1200
Asp Gly Ser Tyr Tyr Phe Ile	Thr Gly Gly Ser Asp	Lys Thr Leu
1205	1210	1215
Glu Leu Arg Ser Ala Leu Gln	Leu Ile Leu Pro Lys	Lys Asn Glu
1220	1225	1230
Trp Ala Ile Lys Gln Ile Asp	Lys Ser Ser Glu Asn	Asp Tyr Leu
1235	1240	1245
Thr Ile Glu Arg Ile Gln Asp	Leu Thr Glu Glu Leu	Val Tyr Asn
1250	1255	1260

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

Asp Leu Lys Ser Ile Asp Leu Lys Ser Asp Tyr Gly Arg Leu Ser
 1325 1330 1335
 Ser Lys Thr Asn Asn Ile Gly Asn Tyr Gln Glu Phe Lys Ile Ile
 1340 1345 1350
 Asn Gln Ser Ile Thr Gly Leu Phe Glu Asn Glu Val Asp Leu Leu
 1355 1360 1365
 Lys Leu
 1370

<210> 4

<211> 1369

<212> PRT

<213> Clostridium sp. AF02-29

<400> 4

Met Lys Glu Lys Met Glu Tyr Tyr Leu Gly Leu Asp Met Gly Thr Asn

1 5 10 15
 Ser Val Gly Trp Ala Val Thr Asp Lys Glu Tyr Arg Leu Met Arg Ala
 20 25 30
 Lys Gly Lys Asp Leu Trp Gly Val Arg Leu Phe Glu Arg Ala Asn Thr
 35 40 45
 Ala Glu Glu Arg Arg Ala Tyr Arg Ile Asn Arg Arg Arg Arg Gln Arg
 50 55 60
 Glu Val Ala Arg Ile Gly Ile Leu Lys Glu Leu Phe Ala Asp Glu Ile
 65 70 75 80
 Ala Lys Val Asp Ala Asn Phe Phe Ala Arg Leu Asp Asp Ser Lys Tyr

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

Gln Tyr Asp His Lys Ala Tyr Asp Glu Met Phe Arg Ile Met Lys Asn
 340 345 350
 Gly Thr Tyr Ser Ala Tyr Val Gly Ser Val Asn Ser Phe Gly Lys Ile
 355 360 365
 Glu Arg Arg Thr Val Lys Thr Ser Arg Glu Glu Leu Leu Lys Asn Ile
 370 375 380
 Lys Lys Ile Leu Thr Gly Phe Pro Glu Asp Asp Ala Thr Val Gln Glu

 385 390 395 400
 Phe Leu Gly Lys Ile Asp Ser Asp Thr Leu Leu Gln Lys Gln Leu Thr
 405 410 415
 Ala Ser Asn Gly Val Ile Pro Asn Gln Val His Ala Lys Glu Met Lys
 420 425 430
 Val Ile Leu Lys Asn Ala Glu Lys Tyr Leu Pro Phe Leu Ser Glu Arg
 435 440 445
 Asp Glu Thr Gly Leu Ser Val Ser Glu Lys Ile Ile Ala Leu Phe Thr

 450 455 460
 Phe Thr Ile Pro Tyr Tyr Val Gly Pro Leu Gly Gln Gln His Leu Gly
 465 470 475 480
 Lys Glu Cys Ala His Gly Trp Val Glu Arg Lys Glu Lys Gly Thr Val
 485 490 495
 Tyr Pro Trp Asn Phe Glu Gln Lys Val Asp Leu Lys Ala Ser Ala Glu
 500 505 510
 His Phe Ile Glu Arg Met Val Lys His Cys Thr Tyr Leu Ser Asp Glu

 515 520 525
 Gln Ala Leu Pro Lys Gln Ser Leu Leu Tyr Glu Lys Phe Gln Val Leu
 530 535 540
 Asn Glu Leu Asn Asn Leu Lys Ile Arg Gly Glu Lys Ile Ser Val Glu
 545 550 555 560
 Leu Lys Gln Gln Ile Tyr Arg Asp Val Phe Glu His Thr Gly Lys Lys
 565 570 575
 Val Ser Met Lys Gln Leu Glu Asn Tyr Leu Lys Leu Asn Gly Leu Leu

580 585 590
 Glu Lys Asp Glu Lys Asp Ala Val Thr Gly Ile Asp Gly Gly Phe His
 595 600 605
 Ser Tyr Leu Ser Ser Leu Gly Lys Phe Ile Gly Ile Leu Gly Glu Glu
 610 615 620
 Ala His Tyr Gly Lys Asn Gln Asn Met Met Glu Lys Ile Val Phe Trp
 625 630 635 640
 Gly Thr Val Tyr Gly Gln Asp Lys Lys Phe Leu Arg Glu Arg Leu Ser

 645 650 655
 Glu Val Tyr Gly Asp Arg Leu Ser Lys Glu Gln Ile Arg Arg Ile Thr
 660 665 670
 Gly Met Lys Phe Glu Gly Trp Gly Arg Leu Ser Lys Glu Phe Leu Leu
 675 680 685
 Leu Glu Gly Ala Ser Arg Glu Glu Gly Glu Ile Arg Thr Leu Ile Arg
 690 695 700
 Ser Leu Trp Glu Thr Asn Glu Asn Leu Met Gly Leu Leu Ser Glu Arg

 705 710 715 720
 Tyr Thr Tyr Ser Glu Glu Val Arg Glu Lys Thr Leu Glu Cys Glu Lys
 725 730 735
 Ser Leu Ser Glu Trp Thr Ile Glu Asp Leu Glu Gly Met Tyr Leu Ser
 740 745 750
 Ala Pro Val Lys Arg Met Val Trp Gln Thr Leu Leu Ile Val Lys Glu
 755 760 765
 Leu Glu Lys Val Leu Gly Cys Ala Pro Arg Arg Ile Phe Val Glu Met

 770 775 780
 Ala Arg Glu Asp Ala Glu Lys Gly Arg Arg Thr Glu Ser Arg Lys Gln
 785 790 795 800
 Lys Leu Gln Asn Leu Tyr Lys Ala Ile Lys Lys Glu Glu Ile Asp Trp
 805 810 815
 Lys Lys Glu Ile Asp Glu Lys Thr Glu Gln Ala Phe Arg Ser Lys Lys
 820 825 830

Leu Tyr Leu Tyr Tyr Leu Gln Lys Gly Arg Cys Met Tyr Thr Gly Glu

835 840 845

Ser Ile Arg Phe Glu Asp Leu Met Asn Asp Asn Leu Tyr Asp Ile Asp

850 855 860

His Ile Tyr Pro Arg His Phe Val Lys Asp Asp Ser Leu Glu Gln Asn

865 870 875 880

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

885 890 895

Pro Ile Glu Ala Asp Ile Gln Lys Lys Met Ser Pro Phe Trp Lys Glu

900 905 910

Leu Lys Glu Arg Gly Phe Ile Ser Glu Glu Lys Tyr Met Arg Leu Thr

915 920 925

Arg Arg Tyr Gly Phe Ser Glu Glu Glu Lys Ala Gly Phe Ile Asn Arg

930 935 940

Gln Leu Val Glu Thr Arg Gln Gly Thr Lys Ser Ile Thr Glu Ile Leu

945 950 955 960

Gly Gln Ala Phe Pro Asp Val Asp Ile Ile Phe Ser Lys Ala Ser Asn

965 970 975

Val Ser Glu Phe Arg His Ile Tyr Gly Leu Tyr Lys Val Arg Ser Ile

980 985 990

Asn Asp Phe His His Ala His Asp Ala Tyr Leu Asn Ile Val Val Gly

995 1000 1005

Asn Thr Tyr His Val Lys Phe Thr Lys Asn Pro Leu Asn Phe Ile

1010 1015 1020

Arg Glu Ala Glu Lys Asn Pro Gln Asn Ala Glu Asn Lys Tyr Asn

1025 1030 1035

Met Asn Arg Met Phe Asp Trp Thr Val Lys Arg Gly Asn Glu Thr

1040 1045 1050

Ala Trp Ile Ala Ser Ser Asp Lys Glu Ala Gly Ser Ile Lys Ile

1055 1060 1065

Val Lys Ala Ile Leu Ala Lys Asn Thr Pro Leu Val Thr Lys Arg

1070	1075	1080
Cys Ala Glu Ala His Gly Gly	Ile Thr Arg Lys Ala	Thr Ile Trp
1085	1090	1095
Asn Lys Asn Lys Ala Ala Gly	Ser Gly Tyr Ile Pro	Val Lys Met
1100	1105	1110
Asn Asp Ala Arg Leu Leu Asp	Val Thr Lys Tyr Gly	Gly Leu Thr
1115	1120	1125
Ser Val Ser Ala Ser Gly Tyr	Thr Leu Leu Glu Tyr	Asp Val Lys
1130	1135	1140
Gly Lys Lys Ile Arg Ser Leu	Glu Ala Ile Pro Ile	Tyr Leu Gly
1145	1150	1155
Arg Val Ser Glu Leu Thr Asn	Glu Ala Ile Leu Lys	Tyr Phe Glu
1160	1165	1170
Lys Val Leu Ile Glu Glu Asn	Lys Gly Lys Glu Ile	Thr Glu Leu
1175	1180	1185
Arg Ile Cys Lys Lys Phe Ile	Pro Arg Glu Ser Leu	Val Arg Tyr
1190	1195	1200
Asn Gly Tyr Tyr Tyr Tyr Leu	Gly Gly Lys Ser Val	Glu Gln Ile
1205	1210	1215
Val Leu Lys Asn Ala Thr Gln	Met Ala Tyr Ser Glu	Glu Glu Thr
1220	1225	1230
Cys Tyr Ile Lys Lys Ile Glu	Lys Ala Ile Glu Lys	Thr Tyr Tyr
1235	1240	1245
Glu Glu Val Asp Lys Asn Lys	Asn Val Ile Leu Thr	Lys Thr Arg
1250	1255	1260
Asn Asn Ala Met Tyr Asp Lys	Phe Ile Ile Lys Tyr	Gln Asn Ser
1265	1270	1275
Ile Tyr Gln Asn Gln Ser Gly	Ala Met Lys Asn Ser	Ile Ile Gly
1280	1285	1290
Lys Arg Asn Glu Phe Leu Thr	Leu Ser Leu Glu Lys	Gln Cys Arg
1295	1300	1305

Ile Leu Lys Ala Leu Val Glu Tyr Phe Arg Thr Gly Asp Ile Ile

1310 1315 1320

Asp Leu Arg Glu Leu Gly Gly Ser Ser Gln Ala Gly Lys Val Ala

1325 1330 1335

Met Asn Lys Lys Ile Met Gly Ala Ser Glu Leu Val Leu Ile Ser

1340 1345 1350

Gln Ser Pro Thr Gly Leu Phe Gln Gln Glu Ile Asp Leu Leu Lys

1355 1360 1365

Ile

<210> 5

<211> 3303

<212> DNA

<213> Butyrivibrio sp. AC2005

<400> 5

atgggatata caataggact tgatcttggg gtggcttcat taggatgggc ttagtcaat 60

gatgaatatg aggtattaga atcatgctca aatatttttc ctgcagcaga atctgcaaat 120

aatgttgaag gacgaggctt taggcaggga agaaggttgt caaggcgtcg caggaccaga 180

attagtgtat tcagaaaact gtgggagaag agtggtttcg aggttccttc aaatgaattg 240

aacgagggtc ttcatgatag gattaaaggc atgaatgata aattatcaga agatgagctt 300

tatcatgttc ttttaaatag cctgaacat aggggaattt cgtatttga ttagtcagat 360

gatgaaatg catctgggga ttatgctgca agcattgctt ataacgaaa tcaattaaag 420

acaaaattgc ctgtgagat tcagtgggag cgtataaga aatatggtgc ttataggggg 480

aatattacta tccaagaagg tggggaaccg ctactctta gaaatgtatt cacaacaagt 540

gcgtatgaag aagaaattca gaagctatta gacgtacaat ctatgtcaaa tgagaaagta 600

acaaaaaagt ttattgatga atacttaaaa atcttttcaa gaaaaagaga atattatatt 660

gggccgggta acaaaaaatc cagaacagat tatggtgtat aactacaca aaaaaatgaa 720

gatggtactt atcactatga gcagaatcct ttgtataaat tgattggaaa gttagtgtta 780

tatcctgatg agagaagagc tgccggggct acttatactg cacaggaatt taatctttta 840

aatgatctga ataactttgt aattgatgga agaaaactag atgagcagga aaaatgtcag 900

attgttgatg ctgttaaaca tgctaaaacc gtcaatatga agaacattat tgcaaaagtc 960

attggaacaa aagcaaactc aatgaatatg accggcgcaa gaatagataa gaatgaaaaa	1020
gaaatttttc attcttttga ggcttataac aagttaagaa aagcactgga agaaatagat	1080
tttgatatag agactttgtc tacggatgag ttggatgcta taggagaagt gttgactctt	1140
aatactgacc gaaaatcaat tcaaacggga cttcaagaga aaagaatagt agttcctgat	1200
gaagtcaggg atgtgcttat cgcaaccagg aaaagaaatg gctcattatt tagcaaatgg	1260
cagtcatttg gtataagaat catgaaggaa ttgattcctg aattatatgc gcagcctaag	1320
aatcagatgc aactgcttac tgatatggga gtatttaaaa ctaaggatga gagatttggt	1380
gagtatgata agattccgtc tgatctaata acagaagaaa tctataatcc tgtggttgct	1440
aaaactgtaa ggattactgt cagagttttg aatgctctta ttaagaaata tggctatccg	1500
gatagagttg ttatagagat gccaaagatg aaaaactcag aagaagagaa aaagcgcata	1560
gcagattttc aaaaagaaca tgagaatgag ctgggtggaa taataaaaaa agtaaagtca	1620
gaatatggta ttgaaataac tgatgctgat ttaagaacc atagtaaaact tggacttaaa	1680
cttaggttgt ggaatgaaca gaatgaaca tgccttact cagggaaca tataaagatt	1740
gatgaccttt taaataatcc taatatgttt gaggtggatc atattatccc attatccatt	1800
tcatttgatg atagtagagc caataaagtg ttggtatagc ctgctgaaaa tcagaataag	1860
ggtaacagaa cgccaatggc atacctgtcc aatgttaata gagaatggga ttcccatgaa	1920
tacatgagtt ttgttcttag taattataag ggaacaatat atggtaagaa gagagataat	1980
cttttattct cagaggacat atataaaatt gatgttttac agggatttat tagcagaaat	2040
ataaatgata caagatatgc ttcaaaggta atacttaatt cattacagtc ttcttttggt	2100
tcaaaagagt gcgacacgaa ggtgaagggt gtttagaggaa cctttacaca tcagatgcga	2160
atgaatctaa agatagaaaa gaatagagag gagtcatatg tgcacatgc tgttgatgct	2220
atgcttatag ctttttctca aatggggtat gatgcatatc ataaacttac agagaagtat	2280
attgattatg aacatggcga attttagat cagaaaggct atgagaagct tattgaaat	2340
gatgtagcat atcgtgaac cacttatcaa aataagtgga tgactataaa gaaaaatata	2400
gaaatagcag ctgaaaagaa taaatactgg tatcaggtaa ataggaaaag caatagaggg	2460
ctttgcaacc agactattta tggtagcaga aatctggatg gcaagacagt aaagatcagc	2520
aaacttgata ttcggacaga tgatgggata aagaaattta aagggtatcg agaaaaaggt	2580
aaactagaac gcittttgat gtataggaat gatccaaaaa catttgaatg gctgcttcag	2640
atttataagg attattcaga ctccaaaaac ccatttgtcc aatatgaatc agagactggt	2700
gatgttatta agaaagtttc aaaaacgaat aatggaccaa aggtatgtga acttcgctat	2760
gaagatgggtg aggttggtag ctgtatcgat atttctcata agtatggata taaaaagggt	2820

agtaaaaagg taattctcga ttctttaaac ccttacagaa tggatgtata ttataacact 2880
aaggacaata ggtattatth ttgttggtgta aagtattcag acattaagt ccaaggtgat 2940
agctatgtaa tcgatgagga taaatagca gcagcactcg ttcaggaaaa aatagtgccg 3000
gaaggaaaag gaagaagtga cttacagag ctgtgttatg aatttaagct atcattttat 3060

aaaaatgaga taatagagta tgaaaaagat ggcgaaatat atgtagaaag atttttatcg 3120
cgaacaatgc caaaagttag caattatatt gaaactaagc cattggaagc tgcaaaattt 3180
gaaaaacgaa atttagtggg gtttagctaag actagcagaa taagaaaaat acgagtggat 3240
atacttggga atcgttatth aaatagtatg gaaaatttcg atttgttgtt gggacataaa 3300
taa 3303

<210> 6
<211> 3297
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic
<400> 6

gggtacacca ttggcttga tttgggagt gcttcattgg gttgggcagt cgtgaacgac 60

gagtacgaag tgctcgagtc ttgtagcaac atcttccccg ccgcccagtc cgctaacaac 120
gtcagcgcaa gaggttccg ccaaggcagg cggttgtctc ggcgaggcg cactcgtata 180
agcgattttc gtaagctttg ggaaaagagc ggatttgaag tgcccagtaa cgagctgaat 240
gaagtcttcc aataccgat caaggggatg aacgacaagc tgagttagga cgaattgtac 300
cacgtgctgt tgaactcatt gaagcaccgg ggtatcagct acctggacga cgccgacgac 360
gagaacgcct caggtgacta cgccgcctct atcgctaca atgagaacca gttgaaaacc 420
aagctcccct gcgaaatcca atgggaaagg tacaagaagt acggggcgta ccgcggtaac 480

atcaccatac aggaggagg cgagccactg actctccgaa acgtgtttac gacgtctgct 540
tacgagaagg agatccagaa actcttggat gtgcagagta tgagtaacga aaaggtcacg 600
aagaaattca tcgacgagta tctgaagatt ttcagtcgca agaggagta ctacataggt 660
ccaggcaata agaagtcacg aaccgactac ggcttttata ccactcagaa gaacaggagc 720
ggcacctacc acacagaaca aaacctgttc gacaagctta tcggtaaatg ctccgtttac 780
cccgcagaaa ggcgcgagc ggggtgccaca tacacagccc aagagttcaa ctgctgaac 840
gacttgaaca acctcgttat cgacggcagg aagctggacg aacaagagaa gtgccaaatc 900

gtcagcgagg tgaagcacgc caagacggtt aacatgaaga atatcatcgc caaggtaatc 960

ggtactaagg cgaatagtat gaacatgaca ggggctagga ttgacaagaa cgagaaggag 1020
 atcttcaca gtttcgaagc gtacaataaa ctgaggaagg ctctcgagga gattgacttc 1080
 gacattgaaa ccctcagtac cgacgaactt gacgccatcg gggaagtcct gacactgaac 1140
 accgatagaa agagcatcca gaatgggttg caggaaaagc ggatcgtggt ccccgacgag 1200
 gtaagagatg tactgattgc cactcgtaag cgtaacggga gcctgttctc caagtggcaa 1260
 tctttcgga tccgtattat gaaagagctc atcccggagc tgtacgcca accaaagaac 1320

 caaatgcagt tgctgaccga catgggcgtc ttcaagacca aagacgaacg gttcgtggaa 1380
 tacgacaaaa tcccagtgta cctcatcacg gaagagatat acaaccccg tctcgccaag 1440
 accgtccgca tcaccgttcg cgtccttaac gcgctcatca agaagtacgg gtatcccgac 1500
 aggggtgtga tcgaaatgcc tcgtgacaag aatagtgagg aagaaaagaa aaggattgct 1560
 gacttcaga agaataacga aaacgaactg ggcgcatca tcaagaaggt caaaagttag 1620
 tacggcatcg agatcacga cgcagacttc aagaatcaca gcaagttggg tctcaagctg 1680
 cgactctgga acgagcaaaa cgagacttgi ccctatagcg gcaagcacat taaaatcgac 1740

 gatctgttga acaaccgaa catgttcgaa gtagaccaca tcattcccct ctcaatctec 1800
 ttcgacgact ctgcgctaa caagtcctg gtgtatgcag cagagaacca aaacaagga 1860
 aataggactc ccatggctta tttgagtaac gtcaaccgag agtgggactt tcacgagtat 1920
 atgtctttcg tgctgtcaaa ctacaaaggc actatctacg ggaagaaacg ggacaacctc 1980
 ttgttttccg aagatatcta caagatagac gtgctgcaag ggttcatctc ccggaacatc 2040
 aacgacacc gatacgcgag taaagtatt ctgaacagcc tgcaaagttt ctccgggtct 2100
 aaggaatgtg ataccaaagt caaagtggta cggggcactt tcacgcacca aatgagaatg 2160

 aacttgaaaa ttgagaagaa ccggaagaa agttacgtcc accacgcagt cgacgcaatg 2220
 ctgattgcct tcagccagat gggctacgac gcctaccaca agtcaccga gaaatacata 2280
 gactacgagc acggagagtt cgtggaccaa aagggatagc aaaagctgat cgagaacgac 2340
 gtcgcctaca gggaaacgac ctaccagaac aaatggatga caatcaagaa gaacattgag 2400
 atcgctgccg agaagaacaa gtattggtat caagtgaacc ggaagtcaaa caggggactg 2460
 tgtaatacaa ccatctacgg cactcgtaac ctgacggga aaaccgtgaa aatttctaag 2520
 ctcgacatcc gactgacga cggaatcaag aagttcaagg gtattgttga gaagggaag 2580

 cttagagat tccttatgta ccgtaacgac cctaagacct tcgagtggct cctgcaaatc 2640
 taaaagact actctgatag caagaatccc ttcgtgcagt acgagtcga aacaggtgac 2700
 gtgataaaga aggtaagcaa gacaaacaac ggcccaaag tctgcgagct gcgatacgag 2760
 gacggggaag tgggaagtgc cattgacata tcccacaaat acgggtacaa gaaaggcagc 2820

aagaaagtga tcctggacag cctgaatccc tatcgcatgg acgtgtacta caataccaaa	2880
gataacagat actacttcgt gggcggttaa tactctgata tcaaatgtca gggagactct	2940
tacgtgattg acgaagacaa gtatgctgct gccctggtag aagagaagat cgtacctgag	3000
gggaaggggc gcagcgatct cactgaactg ggctacgagt tcaactgtc tttctacaag	3060
aacgaaatta ttgaatacga gaaggacggg gagatctacg tcgagcgctt cctgtcaagg	3120
accatgcccc aggtctccaa ctacatcgag aaaaaacccc ttgaggccgc taagttcgag	3180
aagcggaacc tggtaggatt ggccaaaaca tcaaggattc gaaagattag agtcgacatt	3240
ctcggcaaca ggtatctgaa ctcaatggag aactttgact tcgtcgttgg tcacaag	3297
<210> 7	
<211> 3300	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 7	
atggggtaca ccattggctt ggatttggga gtggcttcat tgggttgggc agtcgtgaac	60
gacgagtacg aagtgtcga gtctttagc aacatcttcc ccgccgccga gtccgctaac	120
aacgtcgagc gaagagggtt ccgccaaggc aggcggttgt ctcggcgcag gcgcactcgt	180
ataagcgatt ttcgtaagct ttgggaaaag agcggatttg aagtccccag taacgagctg	240
aatgaagttc tccaataccg gatcaagggg atgaacgaca agctgagtga ggacgaattg	300
taccacgtgc tgttgaactc attgaagcac cggggtatca gctacctgga cgacccgac	360
gacgagaacg cctcaggtag ctacccgcc tctatcgct acaatgagaa ccagttgaaa	420
accaagctcc cctgcgaaat ccaatgggaa aggtacaaga agtacggggc gtaccgcggt	480
aacatcacca tacaggaggg aggcgagcca ctgactctcc gaaacgtgtt tacgacgtct	540
gcttacgaga aggagatcca gaaactcttg gatgtgcaga gtatgagtaa cgaaaaggtc	600
acgaagaaat tcacgacga gtatctgaag attttcagtc gcaagaggga gtactacata	660
ggtccaggca ataagaagtc acgaaccgac tacggcgttt ataccactca gaagaacgag	720
gacggcacct accacacaga aaaaaacctg ttgcacaagc ttatcggtaa atgctccgtt	780
taccccgacg aaaggcgcgc agcgggtgcc acatacacag cccaagagtt caacttgctg	840
aacgacttga acaacctcgt tatcgacggc aggaagctgg acgaacaaga gaagtgccaa	900
atcgtcgacg cggtagaaga cgccaagacg gttaacatga agaatacat cgccaaggta	960
atcggtagta aggcgaatag tatgaacatg acaggggcta ggattgacaa gaacgagaag	1020

gagatcttcc acagtttcga agcgtacaat aaactgagga aggcctctcga ggagattgac 1080
 ttcgacattg aaaccctcag taccgacgaa cttgacgcca tcggggaagt cctgacactg 1140
 aacaccgata gaaagagcat ccagaatggg ttgcaggaaa agcggatcgt ggtccccgac 1200
 gaggtaagag atgtactgat tgccactcgt aagcgtaacg ggagcctgtt ctccaagtgg 1260
 caatctttcg gaatccgtat tatgaaagag ctcatcccgg agctgtacgc ccaaccaaag 1320

 aaccaaatgc agttgctgac cgacatgggc gtcttcaaga ccaaagacga acggttcgtg 1380
 gaatacgaca aaatccccag tgacctcatc acggaagaga tatacaaccc cgttgctgcc 1440
 aagaccgtcc gcatcaccgt tcgcgtcctt aacgcgtca tcaagaagta cgggtatccc 1500
 gacagggtgg tgatcgaat gcctcgtgac aagaatagt aggaagaaaa gaaaaggatt 1560
 gctgacttcc agaagaataa cgaaaacgaa ctggcgaggc tcatcaagaa ggtcaaaagt 1620
 gagtacggca tcgagatcac cgacgcagac ttcaagaatc acagcaagtt gggtctcaag 1680
 ctgcgactct ggaacgagca aaacgagact tgtccctata gcggcaagca cattaaaatc 1740

 gacgatctgt tgaacaaccc gaacatgttc gaagtagacc acatcattcc cctctcaatc 1800
 tccttcgacg actctcgcgc taacaaggtc ctgggtgtatg cagcagagaa ccaaaacaaa 1860
 ggaaatagga ctcccatggc ttatttgagt aacgtcaacc gcgagtggga ctttcacgag 1920
 tatatgtctt tcgtgctgac aaactacaaa ggcaactatc acgggaagaa acgggacaac 1980
 ctcttgtttt ccgaagatat ctacaagata gacgtgctgc aagggttcat ctcccgaac 2040
 atcaacgaca cccgatacgc gagtaaagtg attctgaaca gcctgcaaag tttcttcggg 2100
 tctaaggaat gtgataccaa agtcaaagtg gtacggggca ctttcacgca ccaaatgaga 2160

 atgaacttga aaattgagaa gaaccgggaa gaaagttagc tccaccacgc agtcgacgca 2220
 atgctgattg ctttcagcca gatgggctac gacgcctacc acaagctcac cgagaaatac 2280
 atagactacg agcacggaga gttcgtggac caaaagggat acgaaaagct gatcgagaac 2340
 gacgtgcctt acagggaaac gacctaccag acaaatgga tgacaatcaa gaagaacatt 2400
 gagatcgctg ccgagaagaa caagtattgg tatcaagtga accggaagtc aaacagggga 2460
 ctgtgtaate aaaccatcta cggcactcgt aaccttgacg ggaaaaccgt gaaaatttct 2520
 aagctcgaca tccgcactga cgacggaatc aagaagtca agggatttgt tgagaagggc 2580

 aagcttgaga gattccttat gtaccgtaac gaccctaaga ccttcgagtg gctcctgcaa 2640
 atctacaaag actactctga tagcaagaat cccttcgtgc agtacgagtc cgaaacaggt 2700
 gacgtgataa agaaggtaag caagacaaac aacggcccca aagctcgcga gctgcgatac 2760
 gaggacgggg aagtgggaag ttgcattgac atatccaca aatacgggta caagaaaggc 2820
 agcaagaaag tgatcctgga cagcctgaat ccctatcgca tggacgtgta ctacaatacc 2880

aaagataaca gatactactt cgtgggcgtt aaatactctg atatcaaagtc tcaggagagac 2940
 tcttacgtga ttgacgaaga caagtatgct gctgccctgg tacaagagaa gatcgtacct 3000

 gaggggaagg ggcgcagcga tctcactgaa ctgggctacg agttcaaact gtctttctac 3060
 aagaacgaaa ttattgaata cgagaaggac ggggagatct acgtcgagcg cttcctgtca 3120
 aggaccatgc ccaaggtctc caactacatc gagacaaaac cccttgaggc cgctaagttc 3180
 gagaagcgga acctggtagg attggccaaa acatcaagga ttcgaaagat tagagtcgac 3240
 attctcggca acaggtatct gaactcaatg gagaactttg acttcgtcgt tggtcacaag 3300
 <210> 8
 <211> 3315
 <212> DNA
 <213> bacterium LF-3
 <400> 8
 atgagcagat atgtattagg attagatata ggaattactt ctgtagggta tgggtgaata 60

 gatattgata ataatttatt tgttgattat ggtgtaaggc ttttcaaaga aggaactgct 120
 gcagaaaatg aaacgcgaag aactaaaagg ggttcaagac gtttaaaaag aagaaaatct 180
 aatcgtttaa atgatatgaa aaatctttta aaggaaaatg acttataatt tgaagattat 240
 cgaaattata atccttatga gataagggtc aaaggattaa aagaaaagtt attgcctgaa 300
 gaactatgta cagcaattat gcatataaca aaatcaagag gaacaacttt agaagcactt 360
 gctgatgaaa gtcaagatga tgaaggaaca aaagctacac tttcaaaaaa tgctaaagaa 420
 ttaaattgatg gaaaatatat ttgtgaagtt caattggata gattaaataa ggatcataaa 480

 gtaagaggaa cggaaaataa tttcaaaaca gaagattatg tcaaagaact caaagaataa 540
 ttaaacacc aagattttaa tgaagaattg tgtgatcaaa ttattgaaat ggtttcaaga 600
 agaagacgtt atgatcaagg ccaggtagt gaaaaatcac caactcctta tggaagttat 660
 cgaatggtgg atggtgtttt aaaacatgtt aatttgattg atgaaatgcg tggaagatgt 720
 agtgtctatc cagatgaatt tagagcgctt aaacaatctt atacagcaga attatttaat 780
 ttgttaaatg atttaataa ttaacaatt aaaggtgaga aaataacagt tgaagaaaaa 840
 gaaaaggttg ttgcatttgt taatgaaaaa ggaagtatta cagtaaaaca attacttaaa 900

 ttattagatg ctcaagaaga tgaagttaca ggatttagaa ttgataaaaa tgataaacca 960
 ttaattacag aatttaagggt ttatagtaaa gttttaaaag tctttaaaaa atataaccaa 1020
 caagaattac tagaagataa attgattgtt gatcaagtta ttgacatatg tacaaaatca 1080
 aaaggtattg atgaaagaaa aaaagatatt aaagaattat atcctgaatt tgataatgag 1140

ttaattgaag aattagcttc agttaaggt gtttctgctt atcattcatt atcttttaaa	1200
gcaatgcata taatcaataa agaaatgctt acaacagaaa tgaatcaaat acaagttctt	1260
catgaaatag aaatgtttga taaaaataga aaatcattaa agggtaagaa aaatattgaa	1320
cctgatgaag aagctattct atctccagtt gctaaaagag cgcacgaga aacatttaaa	1380
gtcattaatg cgtaagaaa acaatatggc gaatttgata gtattgttat tgaaatgaca	1440
agagataaaa attcaaagga acaagtaaag cgaataaatg atagtcaaaa aagatttaaa	1500
agcgaaaatg atcgagtga tggaattatt aaaaattcag gtattgatcc agaaagagtt	1560
aatggaaaaa caaaaacgaa aattcgtctt tatttacaac aagattgtaa gacggcctat	1620
acacaacaag atattgattt acatacattg atttttgatg ataaagctta tgaaatagat	1680
catattattc caatatctgt ttcatggat gattctctta ctaataaagt attagcttct	1740
cgtttagaaa accaacaaaa aggtaatcta acaccaatga tggttattt aaagggaaaa	1800
tttacgggtg gtaatttga aaaatataa ttatttgtaa gtagtaataa aaattttaat	1860
ggtaaaaaaa gaaataattt acttactgaa caagatatta caaaagaaga ttagcaaga	1920
aagtttatca atcgtaattt agtagataca agctatgctt gtcgtacagt attaaatact	1980
ttgcaacgct attttaaga taatgaaata gatacaaaag ttcatactat tagaggacaa	2040
tcaaccaata ttttagaaa acgaataaat ttacaaaaag atagagagca agattatttt	2100
catcatgcaa tcgatgcatt gattgttgct tcgttaaaga aaatgaatat tgtcaattca	2160
tatttaatgc attacaacta tagtgattta tatgatgaag aaacagggga agtatttgat	2220
gttttacctg ataacaatt tattgatcaa agatatattt catttatctc tgatttaaaa	2280
aatatttatc aagaatcgaa tcaatataac tiagggtata ttaccaaga acaaatgcat	2340
tatccactta tcaaggtatc tcataaaata gatacaaaac caaataggaa aattgcggat	2400
gaaacaatat atagtacaag aaatatigaa ggacaagata tgctagtga aaaaataaaa	2460
aatatctatg atcctaaga aaagaaagca attgaacttg ttaataatat tattaatgat	2520
gatactgata agtacattat gaaacataaa gatccacaaa cttttgaaaa aataaaaaga	2580
gtggtattaa atcattttaa tgattataaa gattcaaaag aatattatgt aattgacaaa	2640
aaaggtaagt attcttttaa agaagaaagt cctttaacat catattataa tgaaaatgga	2700
gctattacta aatattctaa gaaaaataat ggaccagcaa ttacatcaat gaaattttac	2760
tctgaaaaac taggaaatca tttagcaatt acaagtaatt ataatacaaa taataaaaag	2820
gtaattttta aacaaataag cccatatcga acagactttt atgtatctcc tgaaggaaaa	2880
tataaatttg ttacagttag atataaagat gttttttata aagaacaat tcataaattt	2940

gtcatagatg aaaattggta tcatgaagaa aaaattaaaa aaggaattct agaagattgg 3000

aaatttgtat gttcaatgca tcgagatgaa cttattggac ttatcaaacc tgaaggtaaa 3060

aagtttgttt atgatgcttc aattaatggt ggtcaaacac aatatcatga tggtaaacat 3120

tatgaaatct tgaagtttac agcaacgaat gatgaaaaga aaagaacttt tgaagtaaaa 3180

ccgattaaca ctaactgctc aaaacgatta atgccatctg taggaccttt tattaaaatt 3240

caaaaatttg ctacggatgt tttaggaaat atatatgaag ttaaagataa tagattgaaa 3300

ttagagtctg attag 3315

<210> 9

<211> 3309

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 9

tctaggtacg tgttgggact ggacatcggc ataacttccg tgggctacgg ggttatcgac 60

atcgacaaca acctgttctg cgactacggg gtgagactgt ttaaggaagg cacagccgcg 120

gagaacgaga ccagacggac caagagaggg tcccgcgccc ttaagcgcag gaagagtaac 180

cgccttaacg acatgaagaa cctgctgaaa gagaacgac tgtacttcga ggactacaga 240

aactacaacc cgtacgaaat tcgagccaag gggttgaagg agaaacttct cccagaggag 300

ctgtgcaccg ctatcatgca catcactaag agtctgagg ctacctgga agccttggcc 360

gacgagtctc aggacgacga gggcaccaag gccaccctca gcaagaacgc gaaggagctt 420

aacgacggta agtacatctg cgaggtgcag ctggacaggt tgaacaaaga ccacaaggtc 480

cggggcactg agaacaactt taagaccgag gactacgtta aggaactgaa ggagatcctc 540

aagcatcagg acctgaacga ggagctctgc gaccagatca tcgagatggt atctcgtcgc 600

aggcggtacg accagggacc cggctctgag aagtccccc caccctacgg ttcttaccgg 660

atggtcgacg ggggtgtgaa gcacgtgaac ctgatcgacg agatgagggg ccgatgctcc 720

gtgtaccggg acgagttccg cgctccgaag cagagttaca ccgctgagct tttcaacctg 780

ctgaacgacc tcaacaacct cactatcaag ggagaaaaga ttacggtcga ggagaaggag 840

aaagtggctg ccttcgtgaa cgagaagggg tctatcactg ttaagcagct tctcaagctc 900

cttgacgcac aagaggacga ggtgaccggt ttccgcatcg acaagaacga caagcctctg 960

atcaccgagt tcaaaggata ctcaaagggt cttaaggtgt tcaagaagta caatcagcag 1020

gagcttctgg aagacaagct tatcgtggac caggtcatcg atatctgcac taagagcaag 1080

ggcatcgacg agaggaagaa ggacatcaag gagttgtacc cagagttcga caacgaactg	1140
atcgaggagt tggcaagcgt caagggcgtg tcagcatacc acagtctgag cttcaaggct	1200
atgcacatca ttaacaagga gatgctgacc accgagatga accagattca ggtcctgcac	1260
gagatcgaga tgttcgacaa gaaccgcaag agcttgaaag ggaagaagaa catcgagccc	1320
gacgaagagg ccatcctgtc ccccgtagcc aagcgggcac acccgagac cttcaagggtg	1380
atcaacgccc ttcgtaagca gtacggggag ttcgactcaa tcgtgatcga gatgacccgc	1440
gacaagaact ccaaagagca ggtgaaacgg atcaacgact ctcagaagcg tttcaagtca	1500
gagaacgaca gagtggacgg tatcatcaag aactctggaa tagaccccga gcgtgtcaac	1560
ggcaagacca agacaaagat acgcctctac ctgcagcagg actgcaaaac tgcgtacacc	1620
cagcaggaca tcgacctgca cactcttata ttcgacgaca aggcgtacga gatcgaccac	1680
ataatcccta tcagcgtcag tcttgacgac agtctgacca acaaggttct ggccctcaagg	1740
ctcgagaatc agcagaaggg gaacctcacc cctatgatgg cctacctcaa aggttaagttc	1800
actggcggaa acctggagaa gtacaagctg ttcgtgtcat ccaacaagaa cttcaacggc	1860
aagaagcgca acaacctgct gaccgagcag gacataacta aggaagacgt ggctcgaaaa	1920
ttcattaaca gaaacctggt ggacacatcc tacgcctgca gaaccgtctt gaacacactg	1980
cagaggtact tcaaggacaa cgagattgac actaaggtac acacaatccg aggccagagc	2040
acaaacatct tccgcaagcg cattaacctg cagaaggacc gcgaacagga ctacttcac	2100
cacgccattg acgccctgat cgtggccagt ctgaagaaga tgaacatcgt gaacagctac	2160
ctgatgcact ataattacag cgacttgtac gacgaggaga ctggcgaggt cttcgacgtg	2220
ctgcccgaca agcagttcat cgaccagcgg tacatctcct tcatttccga cctgaagaac	2280
atctaccagg agtccaacca gtacaatctg ggatacataa ctcaggagca gatgcactac	2340
ccgtgatta aagttagcca caagattgac accaagccca accggaagat agctgacgag	2400
actatctaca gcacccgcaa catcgagggc caggacatgt tggaggagaa gattaagaac	2460
atttacgacc ccaaggagaa gaaggccatc gagctggtga acaacataat caacgacgac	2520
accgacaaat atatcatgaa gcacaaggac cccagacct tcgagaagat caaagaggtc	2580
gtcctgaacc acttcaacga ctacaaggac tctaaagagt actacgtcat cgataagaag	2640
gggaaataca gcctgaaaga ggagagcccc ctgactagct actacaacga gaacggggcc	2700
ataacgaagt acagcaagaa gaacaacggg cccgtataa catccatgaa gttctatagc	2760
gagaagctcg gcaaccacct ggctatcact agcaactaca acacgaacaa caagaaagtg	2820
atcctcaage agatttcacc ctaccgtact gatttctacg tgagtccaga gggcaagtac	2880

aagttcgtga ccgtccggtg caaggacgtg ttctacaagg agaccatcca caagttcgtt 2940

atcgacgaga actggtatca cgaggagaag ataaagaagg gtatcctgga agactggaag 3000

ttcgtttgct ctatgcaccg ggacgagctg atcgggctga ttaagccaga gggcaagaaa 3060

ttcgtgtacg acgcgtccat caacggcggg cagactcagt accacgacgg caagcactac 3120

gagattctta aattcaccgc caccaacgac gagaagaaga ggaccttcga ggtgaagccc 3180

atcaatacaa attgtagtaa gaggttgatg ccttccgtcg gcccttcat caagatccag 3240

aagttcgcca ccgacgtcct ggggaacatc tacgaggtga aggacaacag gcttaagctg 3300

gaatttgac 3309

<210> 10

<211> 3312

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 10

atgtctaggt acgtgttggg actggacatc ggcatatctt ccgtgggcta cggggttatc 60

gacatcgaca acaacctgtt cgtcgactac ggggtgagac tgtttaagga aggcacagcc 120

gcggagaacg agaccagacg gaccaagaga gggccccgac gccttaagcg caggaagagt 180

aaccgcctta acgacatgaa gaacctgctg aaagagaacg atctgtactt cgaggactac 240

agaaactaca acccgtagca aattcgagcc aaggggttga aggagaaact tctcccagag 300

gagctgtgca ccgtatcat gcacatcact aagagtcgtg ggactaccct ggaagccttg 360

gccgacgagt ctgaggacga cgagggcacc aaggccaccc tcagcaagaa cgcgaaggag 420

cttaacgacg gtaagtacat ctgcgaggtg cagctggaca ggttgaacaa agaccacaag 480

gtccggggca ctgagaacaa ctttaagacc gaggactacg ttaaggaact gaaggagatc 540

ctcaagcacc aggacctgaa cgaggagctc tgcgaccaga tcatcgagat ggtatctcgt 600

cgcaggcggg acgaccaggg acccggtctt gagaagtcct ccacacccta cggttcttac 660

cggatggtcg acgggggtgt gaagcacgtg aacctgatcg acgagatgag gggccgatgc 720

tccgtgtacc cggacgagtt ccgcgtccg aagcagagtt acaccgctga gcttttcaac 780

ctgctgaacg acctcaacaa cctcactatc aaggagaaaa agattacggt cgaggagaag 840

gagaaagtgg tcgccttcgt gaacgagaag ggtcttatca ctgttaagca gcttctcaag 900

ctccttgacg cacaagagga cgaggtgacc ggtttccgca tcgacaagaa cgacaagcct 960

ctgatcaccg agttcaaagg atactcaaag gtgcttaagg tgttcaagaa gtacaatcag 1020

caggagcttc tggaagacaa gcttatcgtg gaccaggtca tcgatatctg cactaagagc	1080
aagggcatcg acgagaggaa gaaggacatc aaggagtgtg acccagagtt cgacaacgaa	1140
ctgatcgagg agttggcaag cgtcaagggc gtgtcagcat accacagtct gagcttcaag	1200
gctatgcaca tcattaacaa ggagatgctg accaccgaga tgaaccagat tcaggtcctg	1260
cacgagatcg agatgttcga caagaaccgc aagagcttga aagggaagaa gaacatcgag	1320
cccgaagcgaagg aggccatcct gtcccccgta gccaaagcggg cacaccgcga gaccttcaag	1380
gtgatcaacg cccttcgtaa gcagtacggg gagttcgact caatcgtgat cgagatgacc	1440
cgcgacaaga actccaaaga gcaggtgaaa cggatcaacg actctcagaa gcgtttcaag	1500
tcagagaacg acagagtgga cggatatcatc aagaactctg gaatagacc cgagcgtgtc	1560
aacggcaaga ccaagacaaa gatacgccctc tacctgcagc aggactgcaa aactgcgtac	1620
accacgcagg acatcgacct gcacactctt atattcgacg acaaggcgta cgagatcgac	1680
cacataatcc ctatcagcgt cagtcttgac gacagtctga ccaacaaggt tctggcctca	1740
aggctcgaga atcagcagaa ggggaacctc acccctatga tggcctacct caaaggtaaag	1800
ttcactggcg gaaacctgga gaagtacaag ctgttcgtgt catccaacaa gaacttcaac	1860
ggcaagaagc gcaacaacct gctgaccgag caggacataa ctaaggaaga cgtggctcga	1920
aaattcatta acagaaacct ggtggacaca tcctacgcct gcagaaccgt ctggaacaca	1980
ctgcagaggt acttcaagga caacgagatt gacactaagg tacacacaat ccgaggccag	2040
agcacaacaa tcttcgcaa gcgcattaac ctgcagaagg accgcgaaca ggactacttc	2100
caccacgcca ttgacgcct gatcgtggcc agtctgaaga agatgaacat cgtgaacagc	2160
tacctgatgc actataatta cagcgacttg tacgacgagg agactggcga ggtcttcgac	2220
gtgctgcccc acaagcagtt catcgaccag cggatcatct ccttcatttc cgacctgaag	2280
aacatctacc aggagtccaa ccagtacaat ctgggataca taactcagga gcagatgcac	2340
taccgctga ttaaagtcag ccacaagatt gacaccaagc ccaaccgaa gatagctgac	2400
gagactatct acagcaccgc caacatcgag ggccaggaca tgttggtgga gaagattaag	2460
aacatttacg accccaagga gaagaaggcc atcgagctgg tgaacaacat aatcaacgac	2520
gacaccgaca aatatatcat gaagcacaag gacccccaga ccttcgagaa gatcaaagag	2580
gtcgtcctga accattcaa cgactacaag gactctaaag agtactacgt catcgataag	2640
aaggggaaat acagcctgaa agaggagagc ccctgacta gctactaaa cgagaacggg	2700
gccataacga agtacagcaa gaagaacaac ggccccgcta taacatccat gaagttctat	2760
agcgagaagc tcggcaacca cctggctatc actagcaact acaacacgaa caacaagaaa	2820

gtgatcctca agcagatttc accctaccgt actgatttct acgtgagtcc agagggaag 2880

tacaagttcg tgaccgtccg gtacaaggac gtgttctaca aggagaccat ccacaagttc 2940

gttatcgacg agaactggta tcacgaggag aagataaaga agggatcctt ggaagactgg 3000

aagttcgttt gccttatgca ccgggacgag ctgacggggc tgattaagcc agagggaag 3060

aaattcgtgt acgacgcgtc catcaacggc ggacagactc agtaccacga cggcaagcac 3120

tacgagattc ttaaattcac cgccaccaac gacgagaaga agaggacctt cgaggtgaag 3180

cccatcaata caaattgtag taagaggttg atgccttccg tcggccccctt catcaagatc 3240

cagaagttcg ccaccgacgt cctggggaac atctacgagg tgaaggacaa caggcttaag 3300

ctggaatttg ac 3312

<210> 11

<211> 4113

<212> DNA

<213> Ezakiella peruensis strain M6.X2

<400> 11

atgacaaaag taaaagatta ttatatcgga cttgatatag gtacatcatc agttggctgg 60

gcagtaacag acgaggctta caatgttcta aaattcaact ccaagaagat gtggggagtt 120

cgtctttttg atgatgcaa aactgctgaa gaaagacgag ggcaaagagg ggccaggaga 180

agacttgacc gcaaaaaaga acgcttaagt ctcttgcaag atttttttgc agagggaagt 240

gctaaagtag atccaaattt ctttttgcgt ctagataaca gcgaccttta tatggaggac 300

aaagatcaaa agttaagtc caagtiacact ttatttaatg ataagattt taaagacaag 360

aacttcaca aaaaataatc gactatccac catctcctta tggacttgat tgaagatgat 420

agcaaaaaag atattagact gggtttattta gcttgccatt acttacttaa aaatcgtggc 480

cactttattt ttgaaggaca aaaatttgat acaaagagct cctttgaaaa ttctctaaat 540

gaattaaagg tccacttaaa tgatgaatac ggtcttgatc ttgagtttga taatgaaaat 600

ttgataaata tacttacaga tcctaagtta aacaagaccg caaaaaagaa agaacttaaa 660

agtgttattg gagatacaaa atttctaaag gcagtatctg ctattatgat tggtagctct 720

caaaagctag tagatctatt tgaaaatcct gaagactttg atgattcggc aatcaaatca 780

gtggattttt ctacgacgag ttttgatgat aaatatagcg attacgagtt agcccttggg 840

gataaaattg cctttgtaaa tatattaaaa gaaatctatg actcatctat acttgaaaat 900

ttattaaaag aagccgataa atcaaaagat ggcaataagt acatttctaa cgcctttgta 960

aaaaaatata acaagcatgg ccaggacctc aaggaattta agcgcctagt tagacagtac 1020

cataaatcag cctacttcga catcttttagg agtgaaaaag taaacgataa ctatgtttca	1080
tataccaagt caagtatatc caataacaag agagtgaagg cgaataagtt tacagaccaa	1140
gaagcttttt ataagtttgc taaaaagcac ctagaaacta taaaatacaa aattaataaa	1200
gttaatggta gcaaagctga ccttgaacta atagatggaa tgctaaggga tatggaattt	1260
aaaaatttca tgccaaagat aaaatcttct gataatggag ttatacctta tcaattgaaa	1320
cttatggagc taaataagat ccttgaaaac caatccaaac accatgaatt tttaaacgta	1380
tccgatgaat atggaagcgt ttgcgacaag attgcttcga ttatggaatt taggattcca	1440
tattatgttg ggccctttaa tcctaactca aaatatgctt ggattaagaa gcaaaaggac	1500
agcgaaatca cgccatggaa ttttaaagat gtagttgatt tggattcttc aagggaagag	1560
tttatagata gcttaattgg caggtgcaca tttttaaag atgaaaaagt tctaccaaag	1620
gcctcgcttc tctacaatga gtatatggtt ttaaatgaac tcaacaattt aaaattaaat	1680
gatcttctta ttactgaaga aatgaagaag aaaatcttcg atcaactctt taagaccagg	1740
aaaaaagtaa cattaaaggc tgtcgcta atctctcaaaa aagaatttaa tataaatgga	1800
gaaatcctat tgtccggcac agatggggat tttaaacaag ggctaaactc ttataacgat	1860
tttaaggcca ttgttgggga caaggttgac agcgacgact atagggataa aatcgaagaa	1920
attatcaagc taatcgtcct ctatggagat gacaaatctt acttgcaaaa gaaaataaag	1980
gcgggatacg gcaagtatit tacagattca gaaatcaaaa agatggctgg cctaaattat	2040
aaagactggg gcagattaag taaaaaacta ctacacaggt tagaaggcgc caataaaatt	2100
acaggcgaaa gaggatctat aatccatttt atgcgtgagt acaattttaa cttaatggaa	2160
ttaatgagcg ccagcttcac ttttacagag gaaattcaaa agttaaatcc agttgacgat	2220
agaaaactct cctatgagat ggttgatgag ctttatttat caccttcagt taagagaatg	2280
ttatggcaaa gtctaagaat agttgatgaa attaaaaata taatgggcac tgattccaag	2340
aaaatcttta ttgaaatggc caggggcaaa gaagaagtca aggctagaaa agaacttaga	2400
aaaaatcagc tctttaaatt ttacaaggat ggcaaaaaag cctttatatc agaaatcggc	2460
gaagaaagat atagctatct ttttaagtga atcgaaggag aagaggaaaa caaattcaga	2520
tgggacaatc tttatctcta ctacaccag cttggcaggt gtatgtatag tcttgagcca	2580
attgatattt cagaactctc atcgaaaaac atctatgacc aagaccacat ttatccaaag	2640
tcaaaaatct atgatgattc aattgaaaac agagtttttg ttaagaaaga tttaaatagc	2700
aagaaaggca attcataccc aataccggat gagattttta ataaaaattg ctatgcttat	2760

tggaaaattc tatatgacaa gggactaatt ggtcaaaaga aatataccag acttacacgt 2820

aggacaggat ttactgatga tgaacttgtc caatttatat ccaggcaaat agttgagacc 2880

aggcaggcta ccaaagaaac agcaaatctc ttaaaaacca tttgcaaaaa ttcagaaata 2940

gtttactcta aggcagaaaa tgctagcaga ttcagacagg aatttgatat agtaaaatgc 3000

cgtgcagtca atgacctcca ccacatgcat gacgcttata taaatataat cgttggcaat 3060

gtctacaata caaaatttac caaagacccc atgaactttg tcaaaaaaca agagaaagct 3120

agaagtata acttggaaaa catgtttaaa tatgacgtaa agcgcggggg ctatacagca 3180

tggatagcag acgatgaaaa aggcactgtt aaaaatgcta gcatcaagag aataagaaaa 3240

gaactagagg ggaccaacta cagatttact cgcatgaatt atatagaaag tggatgacta 3300

tttaatgcta ccttgcaaag aaaaaacaaa ggaagtcgcc ctctaaaaga taaggggcct 3360

aagagctcaa tagaaaaata tggatggatat actaatataa acaaggcttg ctttgtagtg 3420

ttggatatta aatcaaaaaa taaaatagaa agaaaattaa tgccagttga aagagaaata 3480

tacgctaagc aaaagaatga taaaaaattg agtgatgaaa tatttagcaa atatttgaaa 3540

gatagattcg gaattgaaga ttatagagta gtatatcctg tagtaaagat gagaactttg 3600

ttaaaaatag atggatctta ttattttata actggtggaa gtgacaaaac attggaatta 3660

aggagtgcac ttcaattaat attaccaaag aaaaatgaat gggcaataaa gcaaattgat 3720

aatccagtg agaattgatta cctaacaatt gaaaggatac aagatttaac ggaagaactt 3780

gtatacaata cgtttgatat aatagtgaat aaatttaaaa catctgtatt taaaaaatca 3840

tttttgaatt tattccaaga tgataaaatc gaaaatatag attttaaat caaatcaatg 3900

gattttaaag aaaagtgtaa aactctattg atgctagtaa aagccatcag agcttctggt 3960

gtacgccaag acttaaaatc tatagattta aaatcagact atggtagatt gagctccaag 4020

actaataata taggaaacta tcaagaattt aaatcataa accaatcaat tacaggcctc 4080

tttgaaaacg aagtggactt gttaaaatta tga 4113

<210> 12

<211> 4107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 12

accaaggtga aggactacta cataggcttg gacatcggca cctctagcgt cgggtgggcc 60

gtcaccgatg aagcctataa cgtgcttaag tttaatagca agaaaatgtg gggcgtgcgg 120

ctgttcgacg acgctaagac ggcagaggag cgtaggggcc agcgaggagc aagacgacgt	180
ctggatcgga agaaggagag actcagcctg ctgcaggact tcttcgccga agaggtagca	240
aaggtcgacc ccaacttctt cctcaggctg gacaattccg atctgtacat ggaagataag	300
gaccagaaac tgaagagcaa atatacactg ttcaacgaca aggacttcaa ggataagaat	360
tttcataaga agtaccacac aatacatcac ctgctgatgg atctgatcga ggacgacagt	420
aagaaggaca tccggctcgt ctacctggcc tgtcactatt tgctcaagaa caggggtcat	480
ttcatcttcg agggccagaa gttcgacact aaatcaagct tcgagaacag tttgaacgag	540
ctcaaagttc atttgaacga cgagtatgga ctggacctcg aatttgacaa cgagaacctg	600
attaacatct tgactgacc aaaactcaat aaaacggcca agaagaagga gctgaagtcc	660
gtaatcggcg acaccaagtt cctcaaagcc gtttcgcga taatgatcgg ctctagccag	720
aaactcgtcg acttggtcga gaaccccag gatttcgacg actctcgat aaagtccgtt	780
gacttctcaa ctacctctt cgacgacaag tactctgact atgaactcgc tctgggtgac	840
aagatcgctc tggtaacat ccttaaggaa atttacgata gctccatcct cgagaacctg	900
ctcaaagagg cagacaagtc taaggacggt aacaaatata tcagtaatgc attcgtgaag	960
aagtacaata aacacggaca agatctgaaa gagttcaaac gtctggtacg acaatatcac	1020
aagagtgcgt atttgatat ttccagatcc gagaaggtga atgacaatta cgtcagctac	1080
actaaaagct caattagcaa caataaacgc gtcaaagcaa acaagttcac tgatcaagag	1140
gccttctaca aattcgccaa gaaacatctg gagacaatca agtataagat caacaaggta	1200
aacggctcca aggcagatct ggagctgatt gacgggatgc tgcgggacat ggagttcaag	1260
aactttatgc caaaaattaa gtccagtac aacggggtga ttccatacca gctcaagctg	1320
atggaattga acaaaatact cgagaatcag tcaaagcatc acgagttcct caatgtcagc	1380
gacgagtacg gtcctgtgtg tgataaaatc gcatctatca tggagttccg tatccctac	1440
tacgtgggac ccctgaacct caatagcaag tacgcctgga tcaagaagca gaaagatagt	1500
gagattactc cctggaactt caaggacgtc gtggaccttg actccagcag agaggagttc	1560
attgactcac tgatcggacg ctgtacttac cttaaggacg agaaggtcct tcccaaagct	1620
tctttgctgt ataacgaata catggtgctg aacgagctga ataacctgaa gttgaacgac	1680
cttccatca ccgaggagat gaagaagaag atatttgacc agttgttcaa aacaagaaag	1740
aaggtcacc tttaagcggg ggcaaacctg ctgaagaagg agttcaacat caacggcgag	1800
attctgctct ctgggaccga cggtgacttc aagcagggt tgaactcata caatgacttc	1860
aaagctatcg tggcgataa agtcgattcc gatgattacc gggacaagat tgaggagatc	1920

attaaactga tagttcttta cggtagacgat aagagttacc ttcagaagaa gattaaagct	1980
gggtatggaa aatacttcac cgacagttag attaagaaaa tggcggggct gaactacaag	2040
gattggggaa ggctctcaaa gaagctgctg acgggactcg agggtgcaaa caagatcact	2100
ggagagcggg gctccattat tcacttcatg agggaaatata accttaatct gatggagctt	2160
atgtcagctt catttacgtt caccgaagag atacagaaac ttaaccccggt ggatgaccgc	2220
aagctgtcat acgaaatggt ggacgaactg tacctttctc ccagtgtgaa acggatgctc	2280
tggcagtcct tgcgcctcgt cgacgagata aagaacatca tgggaaccga cagtaagaag	2340
atcttcacgt agatggctcg gggttaaggaa gaggtgaaag cccgcaagga gtcaaggaag	2400
aaccaactgc tgaagticta taaagacgga aagaaggcat tcatcagcga gattggcgag	2460
gagaggtact cttacttctt ttctgagata gagggtagag aagagaataa gtttcgatgg	2520
gataacctgt acctttatta tactcaactg ggtcgtctga tgtactcttt ggaacctatc	2580
gacatatctg agctgtcttc aaagaatatt tacgatcagg atcatatcta ccccaaagc	2640
aagatttacg acgacagtat cgagaatagg gtgctggtga agaaggacct taactccaag	2700
aagggttaaca gctatcctat cccagacgaa atcctgaaca agaactgtta cgcctactgg	2760
aagatcctgt acgataaagg tcttatcggg cagaagaagt acactcggct gacccggaga	2820
actggttca cggacgacga gtcgttcag ttcacttcaa gacagatcgt ggaaactaga	2880
caagcaacaa aggagactgc taacctgctc aagacaatat gtaagaactc cgagatcgtg	2940
tattccaaag ccgagaacgc aagtcggttt aggcaagagt tcgacatcgt gaagtgtagg	3000
gcggtgaacg atcttcatca tatgcacgat gcctacatca acatcatagt ggggaacgtg	3060
tataacacca agttcacgaa ggacctatg aatttcgtaa agaagcagga aaaggcgcgg	3120
agctacaatc tcgagaatat gttcaagtac gatgtgaaac gtggcggata caccgttgg	3180
atcgccgatg acgagaaggg caccgtgaag aacgcgagta ttaaacgtat ccggaaggag	3240
ctggaaggca caaattatag gttcacaaga atgaactaca ttgagtctgg agcgttttc	3300
aacgccactc tcagcggaa gaataagggc tcagacccc tgaaggacaa aggcccga	3360
tcttccatcg agaagtacgg cggctacaca aacatcaata aagcctgttt cgctgttctt	3420
gacatcaagt ctaagaacaa gattgagagg aagctgatgc ccgtcgagcg tgagatctat	3480
gccaaacaga agaacgacaa gaagctgtcc gacgagattt tctcaaagta cctcaaggac	3540
cgatttggca tcgaggacta cagggttgtc taccagtggt tgaatatgcg cacactgctc	3600
aagatcgacg gcagctacta cttcatcaca ggcggttctg ataagaccct ggagttgcga	3660

tctgtctctgc agctgattct ccctaagaag aacgagtggg cgatcaaaca gatcgacaag	3720
tcttccgaaa acgactatct gacgatcgag cgtatccagg acctgaccga ggagctggtg	3780
tataacactt tcgacatcat cgtcaacaag ttcaagacca gtgtcttcaa gaagtctttc	3840
cttaacttgt ttcaggacga caagattgag aacattgact tcaagtttaa gtccatggac	3900
ttcaaggaga aatgcaagac acttctcatg ctggtcaagg cgattcgggc atccggcgtg	3960
aggcaggatc tcaagtccat cgacctcaag tctgattacg gacggctcag ttcaaagacc	4020
aacaacatcg gcaattacca ggagttcaag attattaatc agtccatcac tggactgttc	4080
 gagaatgagg tcgatctcct gaagctg	4107
<210> 13	
<211> 4110	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 13	
atgaccaagg tgaaggacta ctacataggc ttggacatcg gcacctctag cgtcgggtgg	60
gccgtcaccg atgaagccta taacgtgctt aagtttaata gcaagaaaat gtggggcgtg	120
cggctgttcg acgacgctaa gacggcagag gacgtaggg gccagcgagg agcaagacga	180
cgtctggatc ggaagaagga gagactcagc ctgctgcagg acttcttcgc cgaagaggta	240
gcaaaggctg accccaactt cttcttcagg ctggacaatt ccgatctgta catggaagat	300
 aaggaccaga aactgaaaag caaatataca ctgttcaacg acaaggactt caaggataag	360
aattttcata agaagtacc cacaatacat cacctgctga tggatctgat cgaggacgac	420
agtaagaagg acatccggct cgtctacctg gcctgtcact atttgc tcaa gaacaggggt	480
catttcatct tcgagggcc aagttcgc actaaatcaa gcttcgagaa cagtttgaac	540
gagctcaaag ttcatattga cgacgagtat ggactggacc tcgaatttga caacgagaac	600
ctgattaaca tcttgactga cccaaaactc aataaaacgg ccaagaagaa ggagctgaag	660
tccgtaatcg gcgacaccaa gttctcaaaa gccgtttccg cgataatgat cggctctagc	720
 cagaaactcg tcgacttgtt cgagaacccc gaggatttcg acgactctgc gataaagtc	780
gttgacttct caactacctc ttctgacgac aagtactctg actatgaact cgctctgggt	840
gacaagatcg ctctgg tcaa catccttaag gaaatttacg atagctccat cctcgagaac	900
ctgctcaaag aggcagacaa gtctaaggac ggtaacaaat atatcagtaa tgcattcgtg	960
aagaagtaca ataaacacgg acaagatctg aaagagtca aacgtctggt acgacaatat	1020
cacaagagtg cgtattttga tattttcaga tccgagaagg tgaatgacaa ttacgtcagc	1080

tacactaaaa gctcaattag caacaataaa cgcgtcaaag caaacaagtt cactgatcaa	1140
gaggccttct acaaattcgc caagaacat ctggagacaa tcaagtataa gatcaacaag	1200
gtaaacggct ccaaggcaga tctggagctg attgacggga tgctgcggga catggagttc	1260
aagaacttta tgcccaaat taagtccagt gacaacgggg tgattccata ccagctcaag	1320
ctgatggaat tgaacaaaat actcgagaat cagtcaaagc atcacgagtt cctcaatgtc	1380
agcgacgagt acggctccgt gtgtgataaa atcgcatcta tcatggagtt ccgtatcccc	1440
tactactggtg gacccctgaa cccaatagc aagtacgctt ggatcaagaa gcagaaagat	1500
agtgagatta ctccctggaa cttcaaggac gtcgtggacc ttgactccag cagagaggag	1560
ttcattgact cactgatcgg acgctgtact taccttaagg acgagaaggt ccttcccaaa	1620
gcttctttgc tgtataacga atacatggtg ctgaacgagc tgaataacct gaagttgaac	1680
gaccttccca tcaccgagga gatgaagaag aagatatattg accagttgtt caaaacaaga	1740
aagaaggcca cccttaaacg ggtggcaaac ctgctgaaga aggagttcaa catcaacggc	1800
gagattctgc tctctgggac cgacgggtgac ttcaagcagg gcttgaactc atacaatgac	1860
ttcaaagcta tcgtgggcga taaagtcgat tccgatgatt accgggacaa gattgaggag	1920
atcattaaac tgatagttct ttacggtgac gataagagtt accttcagaa gaagattaaa	1980
gctgggtatg gaaaatactt caccgacagt gagattaaga aaatggcggg gctgaactac	2040
aaggattggg gaaggctctc aaagaagctg ctgacgggac tcgagggtgc aaacaagatc	2100
actggagagc ggggctccat tattcacttc atgagggaat ataaccttaa tctgatggag	2160
cttatgtcag cttcatttac gttcaccgaa gagatacaga aacttaaccc cgtggatgac	2220
cgcaagctgt catacgaaat ggtggacgaa ctgtaccttt ctcccagtg gaaacggatg	2280
ctctggcagt ccttgcgcgt cgtcgacgag ataaagaaca tcatgggaac cgacagtaag	2340
aagattttca tcgagatggc tcggggtaag gaagaggtga aagcccga gtagtcaagg	2400
aagaaccaac tgctgaagtt ctataaagac ggaaagaagg cattcatcag cgagattggc	2460
gaggagaggt actcttactt gctttctgag atagagggtg aggaagagaa taagtttcga	2520
tgggataacc tgiaccttta ttatactcaa ctgggtcgct gcatgtactc tttggaacct	2580
atcgacatat ctgagctgtc ttcaaagaat atttacgac aggatcatat ctaccccaaa	2640
agcaagattt acgacgacag tatcgagaat aggggtgctgg tgaagaagga ccttaactcc	2700
aagaagggtg acagctatcc tatccagac gaaatcctga acaagaactg ttacgcctac	2760
tggaagatcc tgtacgataa aggtcttatc gggcagaaga agtacactcg gctgaccggg	2820

agaactggct tcacggacga cgagctcgtt cagttcatct caagacagat cgtggaaact 2880
 agacaagcaa caaaggagac tgctaacctg ctcaagacaa tatgtaagaa ctccgagatc 2940
 gtgtattcca aagccgagaa cgcaagtcgg tttaggcaag agttcgacat cgtgaagtgt 3000
 agggcggtga acgatcttca tcatatgcac gatgcctaca tcaacatcat agtggggaac 3060
 gtgtataaca ccaagttcac gaaggaccct atgaatttcg taaagaagca ggaaaaggcg 3120
 cggagctaca atctcgagaa tatgttcaag tacgatgtga aacgtggcgg atacaccgct 3180
 tggatcgccg atgacgagaa gggcaccgtg aagaacgca gtattaaacg tatccggaag 3240

gagctggaag gcacaaatta taggttcaca agaatgaact acattgagtc tggagcgctt 3300
 ttcaacgcca ctctccagcg gaagaataag ggctccagac ccctgaagga caaaggcccg 3360
 aaatcttcca tcgagaagta cggcggctac acaaacatca ataaagcctg tttcgctgtt 3420
 cttgacatca agtctaagaa caagattgag aggaagctga tgcccgtcga gcgtgagatc 3480
 tatgccaac agaagaacga caagaagctg tccgacgaga ttttctcaa gtacctcaag 3540
 gaccgatttg gcatcgagga ctacagggtt gtctaccag tggtgaaat gcgcacactg 3600
 ctcaagatcg acggcagcta ctacttcac acaggcgggt ctgataagac cctggagtgtg 3660

cgatctgctc tgcagctgat tctccctaag aagaacgagt gggcgatcaa acagatcgac 3720
 aagtcttccg aaaacgacta tctgacgac gagcgtatcc aggacctgac cgaggagctg 3780
 gtgtataaca ctttcgacat catcgtcaac aagttcaaga ccagtgtctt caagaagtct 3840
 ttccttaact tgtttcagga cgacaagatt gagaacattg acttcaagtt taagtccatg 3900
 gacttcaagg agaaatgcaa gacacttctc atgctggtca aggcgattcg ggcacccggc 3960
 gtgaggcagg atctcaagtc catcgacctc aagtctgatt acggacggct cagttcaaag 4020
 accaacaaca tcggcaatta ccaggagtgc aagattatta atcagtcct cactggactg 4080

ttcgagaatg aggtcgatct cctgaagctg 4110

<210> 14

<211> 4110

<212> DNA

<213> Clostridium sp. AF02-29

<400> 14

atgaaagaga aaatggaata ctatttaggt cttgacatgg gaaccaattc agtcggatgg 60
 gctgtaacag ataaagaata tcgtttgatg cgggcgaagg gaaaagattt gtggggagtt 120
 cgtttgttcg aacgtgctaa tacagctgaa gaacggcggg catataggat taaccgaaga 180
 agacgtcagc gggaagtagc tagaattgga attttaaaag aattatttgc cgatgaaatt 240

gcaaaagtgtg atgctaattt ttttgacgc ctggacgaca gtaaatatta tcttgatgat 300

agacaggaaa ataataaaca gaagtatgctg atatttgctg ataaagatta cacggacaaa 360

gagtatttta gccaatatca gacaattttt catctgagaa aagaactgat cctgtcagat 420

caacctcatg atgttcgtct tatttatctg gcgctgttaa atatgtttaa acatagaggc 480

cattttttta ataaaacgtt gggaacctcg gaatcattag aatcgttttt tgatatgtat 540

caaagattag ctgtatgtgc ggatggagag ggaatcaaac ttccagaaac ggtggattta 600

aagaaattag aacagatact tggagcacgt ggatgctcta gaaaggcaac attggaacat 660

atatctgaaa taatggggat taataaaaag aataaaccag tttatagcct catgcagatg 720

atatgtggac ttgatactaa aatgatagac ctttttgggc agaagattga tgaagaacac 780

aaaaaaatct ctctttcatt tcgaacgtcc aattatgaag aaatggcaga ggaagtcctg 840

aatacगतgaggatgtgc atttgaactt atattgacag caaaagaaat gcatgatgtt 900

ggcttgctgg cggagattat gaaaggatat tcataattgt cagaagcgcg ggtggctgtc 960

tatgaagagc atcgaaagga tttggctaaa ctgaaagccg tctttaaaca atatgaccat 1020

aaggcatatg atgaaatgtt tcgaatcatg aagaatggta cttatagtgc ctatgttggg 1080

agtgtaaata gtttcggtaa aatagagaga aggacagtaa aaacttcag agaagaattg 1140

ttaaaaaata taaagaaat tttaacaggg ttccagagg atgatgtac agtacaggaa 1200

tttttgggta agatagattc ggatacactt ctccaaaaac aactgacagc ttctaattgga 1260

gtgattccaa atcaggtaca tgcaaaagaa atgaaggtca ttttgaaaaa tgcagaaaaa 1320

taccttccat ttttaagtga aagagatgaa acaggattaa gtgtatcaga aaaaataata 1380

gctctgttta catttacgat cccgtactat gttggctctc ttgggcagca acatttagga 1440

aaggaatgtg cacatggctg ggtagagcga aaagaaaaag gtactgtgta tccatggaat 1500

tttgaacaaa aggttgattt aaaggcaagt gcagaacatt ttatagaaag aatggtaaaa 1560

cattgcacgt atttatctga tgagcaggca ttgccaaaac aatcattgtt gtatgaaaaa 1620

tttcaggtat tgaatgaatt aaacaattta aaaattcgag gagaaaaaat atcggtagaa 1680

ttaaaacagc agatatatcg ggatgtcttt gaacatactg ggaaaaaagt atcgatgaag 1740

cagttggaaa actatctgaa gttgaacggc ctgcttgaaa aagacgaaaa ggacgcagtt 1800

acaggaatag atggttggtt ccatagttaa ttgtcctctt taggaaaatt tataggaatt 1860

ttgggagaag aagctcatta tggtaaaaac cagaatatga tggaaaaaat tgtattttgg 1920

gggacagtat atggacagga taaaaaattc cttcggaac ggttaagtga agtttatgga 1980

gatagattgt caaaagagca gattcgtcgt attactggta tgaaatttga aggatgggga	2040
cgacttttcta aagaatttct tttactggag ggggcttcta gagaagaagg ggagattcgg	2100
acattgattc gttcattatg ggagacaaat gaaaatttga tggggctttt aagtgaacga	2160
tatacatata gcgaagaagt acgagaaaaa acgctagagt gtgagaagag cttttctgaa	2220
tggacgattg aagatttga aggaatgtat ctgtcagcac cggttaagcg catggtatgg	2280
cagactttgt taattgtaaa agagcttgaa aaggtgctgg gatgtgctcc acgacgtatt	2340
tttgtggaga tggcacgcga agatgcggag aaaggaagga gaacagaatc acgaaagcag	2400
aaattgcaga atctttataa agcaattaaa aaagaggaga tagactggaa aaaagagatt	2460
gatgaaaaaa cagagcaggc attccgcagt aaaaaattat atttgtatta cctgcagaag	2520
gggcgctgta tgtatacggg cgagtcctatt cgatttgaag atttgatgaa tgataattta	2580
tatgatatcg atcatattta tccgagacat tttgtgaagg atgatagttt agagcagaat	2640
ctggtgctgg taaagaagga aaaaaatgca cataaaagtg atgtatttcc gattgaggcg	2700
gatattcaga aaaagatgag tccgttctgg aaagaactga aagaaagagg ttttatatca	2760
gaagaaaagt atatgcgttt aacgaggagg tatggctttt cggaagagga aaaagcaggt	2820
tttatcaatc ggcaatttgt ggaaacaaga cagggaacaa agagtattac agagatattg	2880
ggacaagctt ttccagatgt ggatatcata ttttcaaaag cgtcgaatgt gtcggagttc	2940
agacatattt atggattgta taaggttcgc agtataaatg attttcatca tgcacatgac	3000
gcatacttaa atatagtgtt tggaaatacg tatcatgtga aatttacgaa aaatccgttg	3060
aattttattc gggaagcaga aaaaaatccg cagaatgcag aaaataaata caatatgaac	3120
cggatgttcg attggacagt aaaaagagga aatgaaacgg catggatagc aagttccgat	3180
aaagaagcag gtagtattaa aattgtaaag gctatatagg caaaaaatc accattagtt	3240
actaagaggt gtgcggaggc acatggagga ataaccagga aagcaactat ctggaataaa	3300
aacaaagctg ctggcagcgg atatattccg gttaagatga atgatgcaag acttttggat	3360
gttacaaaat atggcggttt gacatctgta tcagcttcgg ggtacacttt gctggaatat	3420
gatgtaaaag gtaaaaaaat aagaagtctt gaggcgattc caatttactt gggacgagtg	3480
tctgaattga caaatgagcg gattttgaaa tattttgaaa aagtgctgat agaggaaaat	3540
aagggaaaag aaataacaga acttcgcatt tgtaagaagt ttattccgag agagtcgtta	3600
gtgagatata atgggtatta ctattatctc ggaggaaaat cagtggaaaca gattgtattg	3660
aaaaatgcga cacagatggc atattcagaa gaagaaacat gttatataaa aaagatagaa	3720
aaagctatag aaaaaacata ttatgaagaa gtggacaaaa ataaaaatgt gattttgaca	3780
aaaactagaa acaatgcaat gtatgataaa ttattataa aatatcagaa ttcaatttat	3840

cagaatcaaa gtggtgcaat gaaaaattct attattggaa aaagaaatga atttttaaca 3900
 ttatcattgg aaaagcagtg tagaatattg aaggcactag tagaatattt taggacagga 3960
 gatattattg atttgagaga attaggaggt agttcacagg caggaaaagt ggctatgaac 4020
 aagaaaatta tgggagcaag tgaattagtg ctaataagtc aatctccaac aggtttattc 4080

caacaagaga ttgatttact aaaaatatga 4110

<210> 15

<211> 4104

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 15

aaggaaaaga tggagtatta cctggggctg gatatgggca ctaacagcgt gggttgggcg 60
 gtgaccgaca aggagtaccg gctgatgagg gcaaaaggga aggacctgtg gggcgtagcg 120
 ctgtttgaga gagcgaacac tgcggaagag aggcgcgcct acagaatcaa tagacgacgg 180
 cggcaacgag aggttgcaag gatcgggtatc cttaaggaac tcttcgctga cgagatcgcc 240
 aaggtggacg caaacttctt cgccagactt gatgattcaa agtactacct ggacgaccgg 300

caagagaaca acaagcagaa atacgttatt ttcgccgaca aggactatac tgataaggaa 360
 tacttctccc agtaccaaac tatcttccat ctccggaagg agcttatact cagtgaccag 420
 ccacacgacg tgagactgat ctaccttgct cttctgaaca tgttcaagca ccggggacac 480
 ttcttgaaca agactctggg gacttccgag agtttggagt ctttcttcga catgtaccag 540
 cgactggcag tgtgcgcaga cggggaaggc attaatgtgc ccgagaccgt agaccttaag 600
 aagctcgagc aaatcctggg cgccggggga ttagcagga aagccaccct tgagcacatc 660
 agcgagatta tgggaatcaa caagaagaac aagcccgtct actcctgat gcaaatgatt 720

tgcggtctgg acaccaagat gatcgatctg ttcggacaaa agatcgacga ggagcataag 780
 aagataagcc tgtccttcag aactagcaac tacgaggaga tggccgaaga ggttagaaac 840
 acaattggcg acgacgcctt cgagctgatt ctactgccca aggagatgca cgacttcggg 900
 ctgttggctg aaatcatgaa ggggtactcc tacctgagcg aggctcgctg tgccgtgtac 960
 gaggaacacc ggaagacct ggccaagctc aaggcagtgt tcaagcagta cgatcacaaa 1020
 gcttacgacg agatgttcag gattatgaag aacgggacat actcagctta cgtagggtcc 1080
 gtgaactcct ttggcaagat cgaacgcaga accgtgaaga cctctcgca ggagcttctt 1140

aagaacatta agaagatcct gaccggtttc cccgaagacg acgcaactgt gcaagagttc 1200

ctcgggaaaa ttgactctga cacgctgcct cagaagcagt tgactgccag caacggcgta	1260
atccctaacc aagtcacgc gaaggagatg aaagtaatcc tgaagaacgc cgagaagtat	1320
ctgcctttcc tgtccgagag ggacgagact gggctctcag tctccgagaa gatcattgca	1380
ttgttcacgt tcaactattcc ttattacgtg ggacccctgg gtcaacagca cttggggaaa	1440
gagtgcgccc acggttgggt ggaaagaaag gagaagggga ccgtttaccc ctggaacttc	1500
gagcagaaag tcgaccttaa agcttccgct gagcacttca ttgagcgcat ggtgaagcac	1560
tgtacatacc tgtccgacga acaagctctg cccaagcaga gtctgctcta cgagaagttc	1620
caagtgttta acgagctcaa taacttgaag atcaggggagc agaagatcag tgtggagctg	1680
aagcaacaaa ttaccggga cgttttcgag cacaccggaa agaaggtttc aatgaaacaa	1740
ctggagaatt acttgaaact gaatgggctt ctggagaagg atgagaaaga tgccgtgacc	1800
gggatcgacg gcggatttca ctcatacctt tcttccttgg gcaagttcat cggcatcctc	1860
ggggaagagg cactacgg aaagaatcaa aacatgatgg agaagatcgt gtctctgggt	1920
acggtgtacg ggcaagacaa gaagtttctg cgggagcgtc tgtccgaggt gtacggcgac	1980
cggtgagca aggaacaaat cagaaggata acaggaatga agttcgaggg ctggggccgg	2040
ctctccaagg agttcctgct gctcgaagga gcaagtcggg aagagggcga aatccgcacc	2100
ctcatacgga gcctgtggga aacgaacgag aacctgatgg gactgctgtc agagagatac	2160
acttactcag aagaggtccg cgagaagact ctgcaatgag agaaatctct gtcagagtgg	2220
accatcgagg acctcgagg catgtacctt tccgccctg taaaacggat ggtctggcaa	2280
acctcttga tagtgaagga actggagaaa gtctctggct gcgccctcg aaggatcttc	2340
gttgaatgg ctagagagga cgcagaaaag ggtcgccgga ccgagtcgga caaacaaaag	2400
ctgcaaaacc tgtacaaggc tatcaagaag gaagaaattg attggaagaa ggaaatcgac	2460
gagaagaccg aacaagcctt taggagcaag aagctgtacc tgtactatct ccagaaagga	2520
cgatgcatgt acaccggaga aagcatccgc ttcgaggacc tcatgaacga caacttgtac	2580
gacatagacc acatctaccc ccggcacttc gttaaagacg actcccttga acaaacctc	2640
gttttggtta agaaggagaa gaacgctcac aagagcgacg tgttcccaat cgaagccgac	2700
atacaaaaga aaatgtctcc cttttggaag gagctcaagg agaggggatt catctctgag	2760
gagaaataca tgagactcac tcgaagatac gggttcagtg aggaagagaa ggctggattc	2820
attaacagac agctggtaga gaccgtcaa ggcacgaaat ctatcactga aatcctgggc	2880
caggccttcc ccgacgttga cataattttc tccaagcctt caaacgtttc agaatttcgg	2940
cacatctacg gcctctacaa agtgaggctt attaacgact tccaccacgc gcacgatgct	3000
tatctgaaca tcgtcgtagg caacacttac cagtttaagt tcacaaagaa cccctgaac	3060

ttcatccgcg aggccgagaa gaaccacaaa aacgccgaga acaagtataa catgaatcgc 3120
atgtttgact ggaccgtgaa gaggggcaac gagactgcct ggatcgccag cagtgacaaa 3180
gaggccggat ctatcaagat agtcaaagcg attcttgcca agaacacccc tcttgtgacc 3240

aaacggtgcg cagaagctca cggcggcatt actcgcaagg cgacaatttg gaacaagaat 3300
aaggccgagg gtcttgcta catcccagtg aaaatgaacg acgcccggt cctggacgtg 3360
accaagtacg ggggactgac ctcaagtgt gcgtccggct ataccctgct tgagtacgac 3420
gtgaagggga agaagattcg atccctggaa gctatcccca tctatcttgg gagagtcagt 3480
gagctcacta acgaagccat cctcaagtac ttcgagaagg ttcttatcga agagaacaaa 3540
gggaaggaga ttaccgagct ccgtatctgc aagaagtca taccctga aagcctcgtt 3600
cggtaacaag gatactatta ctacctgggc ggcaagtctg ttgagcaaat agtcctgaag 3660

aagccaccc aaatggctta ctccaggaa gagacttgct acatcaagaa aattgagaag 3720
gcaattgaga agacctacta cgaagaggtc gataagaaca agaacgtaat actgactaag 3780
acccgcaata acgcatgta cgacaagttc atcattaagt accaaaacag tatataccaa 3840
aaccagagcg gagccatgaa gaactcaatc atagggaaga ggaacgagtt cctgactctc 3900
agtctcgaga acaaatgccg catcctcaaa gctctgggtcg agtacttccg gaccggggac 3960
atcatagacc tgcgggagct cggcggatca agccaagcgg gcaaggctgc gatgaataag 4020
aagatcatgg gcgcgagcga gctggtcctg atttcacagt cccccaccgg gttgtttcag 4080

caggaaatcg acctgctgaa gatt 4104

<210> 16

<211> 4107

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 16

atgaaggaaa agatggagta ttacctgggg ctggatatgg gactaacag cgtgggttgg 60
gcggtgaccg acaaggagta ccggtgatg agggcaaaag ggaaggacct gtggggcgta 120
cggctgtttg agagagcga cactcggaag gagaggcgcg cctacagaat caatagacga 180
cggcggcaac gagaggttgc aaggatcggc atccttaagg aactcttcgc tgacgagatc 240
gccaaagtgg acgcaaactt cttcgccaga cttgatgatt caaagtacta cctggacgac 300

cggcaagaga acaacaagca gaaatacgt attttcgccg acaaggacta tactgataag 360
gaatacttct ccagtagca aactatcttc catctccgga aggagcttat actcagtgac 420

cagccacacg acgtgagact gatctacctt gctcttctga acatgttcaa gcaccgggga	480
cacttcttga acaagactct ggggacttcc gagagtttgg agtctttctt cgacatgtac	540
cagcgactgg cagtgtgctc agacggggaa ggcattaagt tgcccagac cgtagacctt	600
aagaagctcg agcaaatcct gggcgcccgg ggaatgtagca ggaaagccac ccttgagcac	660
atcagcgaga ttatgggaat caacaagaag aacaagcccg tctactcct gatgcaaatg	720
atttgcggtc tggacaccaa gatgatcgat ctgttcggac aaaagatcga cgaggagcat	780
aagaagataa gcctgtcctt cagaactagc aactacgagg agatggccga agaggttaga	840
aacacaattg gcgacgacgc cttcgagctg attctcactg ccaaggagat gcacgacttc	900
gggctgttgg ctgaaatcat gaagggttac tcctacctga gcgaggctcg cgttgccgtg	960
tacgaggaac accggaaaga cctggccaag ctcaaggcag tgttcaagca gtacgatcac	1020
aaagcttacg acgagatgtt caggattatg aagaacggga catactcagc ttacgtaggg	1080
tccgtgaact cctttggcaa gatcgaacgc agaaccgtga agacctctcg cgaggagctt	1140
cttaagaaca ttaagaagat cctgaccggt ttccccgaag acgacgcaac tgtgcaagag	1200
ttctctggga aaattgactc tgacacgctg cttcagaagc agttgactgc cagcaacggc	1260
glaatcccta accaagtcca cgcaaggag atgaaagtaa tcctgaagaa cgccgagaag	1320
tatctgcctt tctgtccga gagggacgag actgggctct cagtctccga gaagatcatt	1380
gcattgttca cgttactat tccttattac gtgggacccc tgggtcaaca gcacttgggg	1440
aaagagtgcg cccacggttg ggtggaaaga aaggagaagg ggaccgttta ccctggaac	1500
ttcgagcaga aagtcgacct taaagcttcc gctgagcact tcattgagcg catggtgaag	1560
cactgtacat acctgtccga cgaacaagct ctgcccgaagc agagtctgct ctacgagaag	1620
ttccaagtgc ttaacgagct caataacttg aagatcaggg gcgagaagat cagtgtggag	1680
ctgaagcaac aaatttaccg ggacgttttc gagcacaccg gaaagaaggt ttcaatgaaa	1740
caactggaga attacttgaa actgaatggg cttctggaga aggatgagaa agatgccgtg	1800
accgggatcg acggcggtt tcaactatac ctttcttccc tgggcaagtt catcggcac	1860
ctcggggaag aggcacacta cggaagaat caaaacatga tggagaagat cgtgttctgg	1920
ggtacggtgt acgggcaaga caagaagttt ctgcgggagc gtctgtccga ggtgtacggc	1980
gaccgctga gcaaggaaca aatcagaagg ataacaggaa tgaagtctga gggctggggc	2040
cggctctcca aggagttcct gctgctcgaa ggagcaagtc gggaagaggc cgaaatccgc	2100
accctcatac ggagcctgtg ggaaacgaac gagaacctga tgggactgct gtcagagaga	2160
tacacttact cagaagaggt ccgcgagaag actctcgaat gcgagaaatc tctgtcagag	2220
tggaccatcg aggacctga gggcatgtac ctttccgccc ctgtaaaacg gatggtctgg	2280

caaaccctct tgatagtga ggaactggag aaagtctctg gctgcgcccc tcgaaggatc	2340
ttcgttgaaa tggctagaga ggacgcagaa aagggtcgcc ggaccgagtc ccgcaaaaa	2400
aagctgcaaa acctgtacaa ggctatcaag aaggaagaaa ttgattggaa gaaggaaatc	2460
gacgagaaga ccgaacaagc ctttaggagc aagaagctgt acctgtacta tctccagaaa	2520
ggacgatgca tgtacaccgg agaaagcatc cgcttcgagg acctcatgaa cgacaacttg	2580
tacgacatag accacatcta cccccggcac ttcgttaaag acgactccct tgaacaaaac	2640
ctcgttttgg ttaagaagga gaagaacgct cacaagagcg acgtgttccc aatcgaagcc	2700
gacatacaaa agaaaatgtc tcccttttgg aaggagctca aggagagggg attcatctct	2760
gaggagaaat acatgagact cactcgaaga tacgggttca gtgaggaaga gaaggctgga	2820
ttcattaaca gacagctggg agagacccgt caaggcacga aatctatcac tgaaatcctg	2880
ggccaggcct tccccagct tgacataatt tttccaagg cttcaaacgt ttcagaattt	2940
cggcacatct acggcctcta caaagtgagg tctattaacg acttcacca cgcgcacgat	3000
gcttatctga acatcgtcgt aggcaacact taccacgtta agttcacaaa gaacccctg	3060
aacttcatcc gcgaggccga gaagaacceca caaacgccg agaacaagta taacatgaat	3120
cgcagtgttg actggaccgt gaagaggggc aacgagactg cctggatcgc cagcagtgac	3180
aaagaggccg gatctatcaa gatagtcaaa gcgattcttg ccaagaacac ccctcttgtg	3240
accaaacggt gcgcagaagc tcacggcggc attactcgca aggcgacaat ttggaacaag	3300
aataaggccg cgggttctgg ctacatcca gtgaaaatga acgacgccg gctcctggac	3360
gtgaccaagt acggcggact gacctcagt agtgcgtcgg gctataccct gcttgagtac	3420
gacgtgaagg ggaagaagat tcgatccctg gaagctatcc ccatctatct tgggagagtc	3480
agtgagctca ctaacgaagc catctcaag tacttcgaga aggttcttat cgaagagaac	3540
aaaggaagg agattaccga gctccgtatc tgcaagaagt tcatacccg tgaaagcctc	3600
gttcggtaca acggatacta ttactacctg ggcggaagt ctgttgagca aatagtcctg	3660
aagaacgcca cccaaatggc ttactccgag gaagagactt gctacatcaa gaaaattgag	3720
aaggcaattg agaagaccta ctacgaagag gtcgataaga acaagaacgt aatactgact	3780
aagaccgca ataacgcgat gtacgacaag ttcatcatta agtacaaaa cagtatatac	3840
caaaaccaga gcggagccat gaagaactca atcataggga agaggaacga gttcctgact	3900
ctcagtcctg agaacaatg ccgcatcctc aaagctctgg tcgagtactt ccggaccggg	3960
gacatcatag acctgcggga gctcggcgga tcaagccaag cgggcaaggt cgcgatgaat	4020

aagaagatca tgggcgcgag cgagctgggc ctgatttcac agtccccac cgggttggtt 4080

cagcaggaaa tcgacctgct gaagatt 4107

<210> 17

<211> 16

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 17

guuuuaguac cuagag 16

<210> 18

<211> 18

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 18

cuuuagaccu acuaaaaau 18

<210> 19

<211> 19

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 19

guuuuaguac cuagagaaa 19

<210> 20

<211> 21

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 20

uuucuuuaga ccuacuaaaa u 21

<210> 21

<211> 46

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 21

aaggcuuuau gccgagauua aaggauccg acgggcaucc uuuuuu 46

<210> 22

<211> 11

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 22

ggcuuuauagc c 11

<210> 23

<211> 23

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 23

aaggauccg acgggcaucc uu 23

<210> 24

<211> 84

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 24

guuuuaguac cuagaggaaa cuuuagaccu acuaaaauaa ggcuuuagc cgagauuaaa 60

ggaugccgac gggcauccuu uuuu 84

<210> 25

<211> 90

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 25

guuuuaguac cuagagaaag aaauuucuuu agaccuacua aaauaaggcu uuaugccgag 60

auuaaaggau gccgacgggc auccuuuuuu	90
<210> 26	
<211> 90	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 26	
guuuuaguac cuagagaaag aaauuucuuu agaccuacuu aaauaaggcu uuaugccgag	60
auuaaaggau gccgacgggc auccuuuuuu	90
<	
210> 27	
<211> 16	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 27	
guuuuguuac cauaug	16
<210> 28	
<211> 18	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 28	
uauaugaccu aacaaaac	18
<210> 29	
<211> 19	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 29	
guuuuguuac cauaugauu	19
<210> 30	
<211> 21	
<212> RNA	

<213> Artificial Sequence
 <220><223> Synthetic
 <400> 30
 auuuauauga ccuaacaaaa c 21
 <210> 31
 <211> 38
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 31
 aaggguuuau cccggacucg gcucuucgga gccuuuuu 38
 <210> 32
 <211> 11
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 32
 ggguuuaucc c 11

 <210> 33
 <211> 14
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 33
 ggcucuucgg agcc 14
 <210> 34
 <211> 76
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 34
 guuuuguuac cauauaggaaa uauaugaccu aacaaaacaa ggguuuaucc cggacucggc 60
 ucuucggagc cuuuuu 76
 <210> 35

<211> 82

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 35

guuuuguuac cauaugauug aaaaauuaua ugaccuaaca aaacaagggu uuaucccgga 60

cucggcucuu cggagccuuu uu 82

<210> 36

<211> 82

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 36

guuuaguuac cauaugauug aaaaauuaua ugaccuaacu aaacaagggu uuaucccgga 60

cucggcucuu cggagccuuu uu 82

<210> 37

<211> 14

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<

400> 37

guuugagagu uaug 14

<210> 38

<211> 16

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 38

caugacgagu ucaaau 16

<210> 39

<211> 17

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic
 <400> 39
 guuugagagu uauguaa 17
 <210> 40
 <211> 19
 <212> RNA
 <213> Artificial Sequence
 <
 220><223> Synthetic
 <400> 40
 uuacaugacg aguucaau 19
 <210> 41
 <211> 72
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 41
 aaaaauuuau ucaaaccgcc uauuuauagg ccgcagaugu ucugcauuau gcuugcuauu 60
 gcaagcuuuu uu 72
 <210> 42
 <211> 14
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 42
 gccuauuuau aggc 14
 <210> 43
 <211> 34
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 43
 gcagauguuc ugcauuauugc uugcuauugc aagc 34
 <210> 44

<211> 106
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 44
 guuugagagu uauggaaaca ugacgaguuc aaauaaaaau uuauucaaac cgccuauuuu 60
 uaggccgcag auguucugca uuaugcuugc uauugcaagc uuuuuu 106
 <210> 45
 <211> 112
 <212> RNA
 <213> Artificial Sequence
 <220><223>
 Synthetic
 <400> 45
 guuugagagu uauguaagaa auuacaugac gaguucaaaau aaaaauuuau ucaaaccgcc 60
 uauuuauagg ccgcagaugu ucugcauuau gcuugcuauu gcaagcuuuu uu 112
 <210> 46
 <211> 14
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 46
 guuugagaac caug 14
 <210> 47
 <211> 16
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 47
 cauggugagu gcaauu 16
 <210> 48
 <211> 17
 <212> RNA
 <213> Artificial Sequence

<220><223> Synthetic

<400> 48

guuugagaac cauguaa 17

<210> 49

<211> 19

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 49

uuacauggug agugcaau 19

<210> 50

<211> 64

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 50

aaggauuau cgaauuugua ugcccgcgau gugcggcaau aaaaaggcuc gaaagagucu 60

uuuu 64

<210> 51

<211> 64

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 51

aaggauuau cgaauuugua ugcccgcgau gugcggcaau aaaaaggcuc gaaagagucu 60

uuuu 64

<210> 52

<211> 46

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 52

uguaugcccg cauugugcgg caauaaaaag gcucgaaaga gucuuu 46

<

210> 53
 <211> 98
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 53
 guuugagaac cauggaaaca uggugagugc aaauaaggau uaucgaaa uguaugcccg 60
 cauugugcgg caauaaaaag gcucgaaaga gucuuuuu 98
 <210> 54
 <211> 104
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 54
 guuugagaac cauguaagaa auuacauggu gagugcaau aaggauuau cgaaaaugua 60
 ugcccgaau gugcggcaau aaaaaggcuc gaaagagucu uuuu 104
 <
 210> 55
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 55
 ggtgcggttc accagggtgt cg 22
 <210> 56
 <211> 22
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 56
 ggaagagcag agccttggtc tc 22
 <210> 57
 <211> 6670
 <212> DNA
 <213> Artificial Sequence

<220><223> Synthetic

<400> 57

gttgtgccac gcggttggga atgtaattca gctccgcat cgccgcttcc actttttccc	60
gcgttttcgc agaaacgtgg ctggcctggi tcaccacgcg ggaaacggtc tgataagaga	120
caccggcata ctctgcgaca tcgtataacg ttactggttt cacattcacc accctgaatt	180
gactctcttc cgggcgctat catgccatac cgcgaaaggt tttgcgcat tcgatgggtgt	240
cgggatctcg acgctaaatt aatacgactc actatagggg aattgtgagc ggataacaat	300
tccccgtag aaataatfff gtttaactaa agaggagaaa tttcatatgt acccatacga	360
tgtgccagat tacgctggca ccgagctcgg taccggctat acaattggcc tggatctggg	420
cgttgccctct ctiggctggg ccgtcgtgaa tgatgagtac gaggtgctgg aaagctgcag	480
caacatcttt cctgcgcgag agagcgccaa caacgtggaa agaagaggct tccggcaagg	540
cagacggctg agcagaagaa gaaggacccg gatcagcgac ttcagaaagc tgtgggagaa	600
gtccggcttc gaggtgcccc gcaatgagct gaatgagggt ctgcagtacc ggatcaaggg	660
catgaacgac aagctgagcg aggacgagct gtaccacgtg ctgctgaaca gcctgaagca	720
cagaggcatc agctacctgg acgacgccga tgatgagaac gcctctggcg attatgccgc	780
ctctatgcc tacaacgaga accagctgaa aacaaagctg ccctgcgaga tccagtggga	840
gagatacaag aagtacggcg cctaccgggg caacatcaca atccaagaag gcggcgagcc	900
cctgacactg agaaatgigt ttaccaccag cgcttacgag aaagagatcc agaaactgct	960
ggacgtgcag agcatgagca acgagaaaagt gaccaagaag ttcatcgacg agtacctcaa	1020
gatcttcagc cggaagagag agtactacat cgccctggc aacaagaagt ccagaaccga	1080
ctacggcgtg tacaccacac agaagaacga ggacggcacc taccacaccg agcagaacct	1140
gttcgataag ctgateggca agtgcagcgt gtaccctgat gagcgtagag ccgctggcgc	1200
cacatacaca gcccaagagt tcaacctgct gaacgatctg aacaacctgg tcatcgacgg	1260
ccggaagctg gacgagcaag agaagtgtca gatcgtggat gccgtgaagc acgccaagac	1320
cgtgaacatg aagaacatca ttgccaaagt gatcggcacc aaggccaaca gcatgaacat	1380
gaccggcgcc agaatcgaca agaattgagaa agaaatcttc cacagcttcg aggcctacaa	1440
caagctgcgg aaggccctgg aagagatcga ctctgacatc gagacactga gcaccgacga	1500
gctggatgcc attgagagg tgctgaccct gaacaccgac cggaagtcta tccagaacgg	1560
cctgcaagag aaacggatcg tgggtcccca tgaagtgcgg gatgtgctga tcgccaccag	1620
aaagagaaat ggcagcctgt tctccaagtg gcagagcttc ggcatcggga tcatgaagga	1680

actgatccca gagctgtacg cccagcctaa gaaccagatg cagctgctga ccgacatggg 1740

cggtgttcaag accaaggacg agagattcgt ggaatacgac aagatcccca gcgacctgat 1800

caccgaagag atctacaacc ccgtgggtggc caagacagtg cggatcaccg ttagagtgtt 1860

gaacgccctg atcaagaagt atggctaccc cgaccgggtc gtgatcgaga tgcccagaga 1920

taagaactcc gaggaagaga agaagcggat cgccgacttc cagaagaaca acgaaaacga 1980

gcttggcggc atcatcaaga aagtgaagtc cgagtacggc atcgagatca ccgacgccga 2040

ctttaagaac cacagcaagc tgggcctgaa gctgagactg tggaacgagc agaatgagac 2100

atgcccctac agcggcaagc acatcaagat cgacgacctg ctcaacaacc ccaacatgtt 2160

cgagggtggac cacatcatcc ctctgagcat cagcttcgac gacagcagag ccaacaaggt 2220

gctggtgtac gccgccgaaa accagaacaa gggcaacaga acccctatgg cctacctgag 2280

caacgtgaac agagagtggg acttccacga gtacatgagc ttctgtgtga gcaactacaa 2340

gggcaccatc tacggcaaga agcgggacaa tctgtgttcc tccgaggaca tctacaagat 2400

cgatgtgttg cagggttcca tctcccgaa catcaacgac accagatacg cctctaaagt 2460

gatcctgaac tccttcgaga gctttttcgg cagcaaagaa tgcgacacca aagtgaaggt 2520

cgtgcggggc accttcacac accagatgag gatgaacctg aagatcgaga agaaccggga 2580

agagtccctac gtgcaccacg ccgtggatgc tatgtgatt gccttcagcc agatgggcta 2640

cgacgcctac cacaactga ccgagaagta tatcgactac gagcacggcg agttcgtgga 2700

ccagaaggga tacgagaagc tgattgagaa cgacgtggcc tacagagaga caacctatca 2760

gaacaagtgg atgacatca agaagaatat cgagatcgcc gctgagaaaa acaagtactg 2820

glatcaagt aaatcggaagt ccaaccgggg cctgtgcaac cagaccatct atggcaccag 2880

aaacctggac ggcaaacccg tgaagatctc caagctggac atccggaccg acgacggcat 2940

caaaaagtgt aagggcacg tggaaaaggg caagctggaa cggttcctga tgtaccggaa 3000

cgaccccaag accttcgagt ggctgctgca gatctataag gactacagcg acagcaagaa 3060

ccccttcgtg cagtacgagt ctgagacagg cgacgtgatc aaaaaggtgt ccaagacaaa 3120

caacggcccc aaagtgtgag agctgagata cgaggatggc gaagtgggct cctgcatcga 3180

catcagccac aaataggct acaagaaggg cagcaagaaa gtcaccttg atctctgaa 3240

cccctaccg atggacgtgt actacaacac caaggacaac cggtactact tctgtggcgt 3300

gaagtactcc gacatcaagt gccagggcga cagctacgtg atcgacagg ataagtatgc 3360

cgccgctctg gtgcaagaaa agatcgtgcc agaaggcaag ggcagatccg atctgaccga 3420

gctgggctat gaggccaagc tgccttctca caagaacgag atcatcgagt acgagaagga 3480
 cggggagatc tacgtcgagc gggttctgtc cagaacaatg cctaaagtgt ccaactatat 3540
 cgagacaaag cccctggaag ccgccaagtt cgagaagaga aacctcgtgg gcctcgccaa 3600
 gacaagccgg atcagaaaga tcagagtggg catcctgggg aaccgctacc tgaacagcat 3660
 ggaaaacttc gacttcgtcg tgggccacaa gggatcctaa gcggccgcct agcataaccc 3720
 ctggggcct ctaaaccggg cttgaggggt tttttgacct aggctagggg atatatccg 3780
 ctctctcgct cactgactcg ctacgctcgg tcgttcgact gcggcgagcg gaaatggctt 3840

 acgaacgggg cggagatttc ctggaagatg ccaggaagat acttaacagg gaagtgagag 3900
 ggccgcggca aagccgtttt tccataggct ccgccccct gacaagcatc acgaaatctg 3960
 acgctcaaat cagtgggtggc gaaacccgac aggactataa agataccagg cgtttcccc 4020
 tggcgctcc ctctgctgct ctctgttcc tgcctttcgg ttaccggtg tcattccgct 4080
 gtatggccg cgtttgtctc attccacgcc tgacactcag ttccgggtag gcagttcgct 4140
 ccaagctgga ctgtatgcac gaacccccg ttcagtcga ccgtcgcc ttatccggtg 4200
 actatcgtct tgagtccaac ccggaagac atgcaaaagc accactggca gcagccactg 4260

 glaattgatt tagaggagt agtcttgaag tcatgcgcc gttaaggcta aactgaaagg 4320
 acaagttttg gtgactgcgc tctccaagc cagttacctc ggttcaaaga gttagtagct 4380
 cagagaacct tcgaaaaacc gccctgcaag gcggtttttt cgttttcaga gcaagagatt 4440
 acgcgcagac caaacgac tcaagaagat catcttatta atcagataaa atatttctag 4500
 atttcagtgc aatttatctc ttcaaatgta gcacctgaag tcagcccat acgatataag 4560
 ttgttactag tgcttgatt ctaccaaata aaaaacgccc ggcggaacc gagcgttctg 4620
 aacaaatcca gatggagttc tgaggtcatt actggatcta tcaacaggag tccaagcgag 4680

 ctgtaaaact tggctgaca gttaccaatg cttaatcagt gaggcaccta tctcagcat 4740
 ctgtctattt cgttcatcca tagttgcctg actccccgtc gtgtagataa ctacgatacg 4800
 ggagggtta ccatctggcc ccagtgtgc aatgataccg cgggagccac gtcaccggc 4860
 tccagattta tcagcaataa accagccagc cggaagggcc gagcgagaa gtggtcctgc 4920
 aactttatcc gcctccatcc agtctattaa ttgttgccgg gaagctagag taagtagttc 4980
 gccagttaat agtttgcga acgttgttgc cattgtaca ggcatcgtgg tgcacgctc 5040
 gtcgtttggt atggcttcat tcagctccgg ttcccaacga tcaaggcgag ttacatgatc 5100

 ccccatgttg tgcaaaaaag cggtagctc ctccggtcct ccgatcgttg tcagaagtaa 5160
 gtggccgca gtgttatcac tcatggttat ggcagcactg cataattctc ttactgtcat 5220
 gccatccgta agatgctttt ctgtgactgg tgagtactca accaagtcac tctgagaata 5280

gtgtatgcgg cgaccgagtt gctcttgccc ggcgtaata cgggataata ccgcgccaca 5340
tagcagaact ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag 5400
gatcttaccg ctgttgagat ccagttcgat gtaaccact cgtgcacca actgatcttc 5460
agcatctttt actttacca gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc 5520

aaaaaaggga ataaggcgca cacggaatg ttgaatactc atactcttcc tttttcaata 5580
ttattgaagc atttatcagg gttattgtct catgagcgga tacatatttg aatgtattta 5640
gaaaaataaa caaatagggg ttccgcgcac atttccccga aaagtgcac ctgacgtcct 5700
cgagtcccg tgcctaata gtgagctaac ttacattaat tgcgttgcgc tcaactcccc 5760
ctttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcgccaa cgcgcgggga 5820
gaggcggttt gcgtattggg cgccaggggtg gtttttcttt tcaccagtga cacgggcaac 5880
agctgattgc ctttcaccgc ctggccctga gagagttgca gcaagcggtc cacgctggtt 5940

tgccccagca ggcgaaaatc ctgtttgatg gtgggttaacg gcgggatata acatgagctg 6000
tcttcggtat cgtcgtatcc cactaccgag atgtccgcac caacgcgcag cccggactcg 6060
gtaatggcgc gcattgcgcc cagcgcctac tgatcgttgg caaccagcat cgcagtggga 6120
acgatgccct cattcagcat ttgcatggtt tgttgaaaac cggacatggc actccagtcg 6180
ccttcccgtt ccgctatcgg ctgaatttga ttgcgagtga gatatttatg ccagccagcc 6240
agacgcagac gcgccgagac agaacttaat gggcccgccta acagcgcgat ttgctggtga 6300
cccaatgcga ccagatgctc cagccccagt cgcgtaccgt cttcatggga gaaaataata 6360

ctgttgatgg gtgtctggtc agagacatca agaaataacg ccggaacatt agtgcaggca 6420
gcttcacag caatggcatc ctggcatcc agcggatagt taatgatcag cccactgacg 6480
cgttcgcga gaagatttg caccgccgt ttacagctt cgacgccgt tcgttctacc 6540
atcgacacca ccacgtggc acccagttga tcggcgcgag atttaatcgc cgcgacaatt 6600
tgcgacggcg cgtgcagggc cagactggag gtggcaacgc caatcagcaa cgactgtttg 6660
cccgccagtt 6670

<210> 58
<211> 2567
<212> DNA
<213> Artificial Sequence

<220><223> Synthetic
<400> 58
gcataaccaa gcctatgcct acagcatcca gggtagcggg gccgaggatg acgatgagcg 60

cattgttaga tticatacac ggtgcctgac tgcgttagca atttaactgt gataaactac 120
cgcattaaag ctatcgtatg ataagctgtc aacacatttc cccgaaaagt gccacctgac 180
gtcctcgagt cccgcataat cgaaatttga cagctagctc agtcctaggt ataatactag 240
tggaagagca gagccttggc ctcgtttttag tacctagaga aagaaatttc tttagaccta 300
ctaaaataag gctttatgcc gagattaaag gatgccgacg ggcatccttt tttgaattct 360
caaataaaac gaaaggctca gtcgaaagac tgggcctttc gttttatctg ttgtttgtcg 420

glgaacgctc tctgagtag gacaaatggt accccgcttc ctgctcact gactcgctac 480
gctcggtcgt tgcactgagg cgagcggaaa tggcttacga acggggcgga gatttcctgg 540
aagatgccag gaagatactt aacagggaag tgagagggcc gcggcaaagc cgtttttcca 600
taggctccgc cccctgaca agcatcacga aatctgacgc tcaaatcagt ggtggcgaaa 660
cccgacagga ctataaagat accaggcgtt tccccctggc ggctccctcg tgcgctctcc 720
tgttctgccc tticggttta ccggtgtcat tccgtgttta tggccgcgtt tgtctcattc 780
cacgcctgac actcagttcc gggtaggcag ttcgtccaa gctggactgt atgcacgaac 840

ccccgttca gtccgaccgc tgcgccttat ccggttaacta tcgtcttgag tccaaccggc 900
aaagacatgc aaaagcacca ctggcagcag ccactggtaa ttgatttaga ggagttagtc 960
ttgaagtcat gcgccgttta aggctaaact gaaaggacaa gttttggtga ctgcgctcct 1020
ccaagccagt tacctcggtt caaagagttg gtatgtcaga gaaccttcga aaaaccgccc 1080
tgcaaggcgg ttttttcgtt ttcagagcaa gagattacgc gcagaccaa acgatctcaa 1140
gaagatcatc ttattaatca gataaaatat ttctagattt cagtgcattt tatctcttca 1200
aatgtagcac ctgaagtcag ccccatagca tataagttgt tactagtgtc tggattctca 1260

ccaataaaaa acgccccggc gcaaccgagc gttctgaaca aatccagatg gagttctgag 1320
gtcattactg gatctatcaa caggagtcca agcgagaagg gttggtttgc gcattcacag 1380
ttctccgcaa gaattgattg gctccaattc ttggagtggg gaatccgtta gcgaggtgcc 1440
gccggcttcc attcaggctg aggtggcccg gctccatgca ccgcgacgca acgcggggag 1500
gcagacaagg tatagggcgg cgcctacaat ccatgccaac ccgttccatg tgctcgccga 1560
ggcggcataa atcgccgtga cgatcagcgg tccaatgac gaagttagc tggttaagagc 1620
cgcgagcgat ccttgaagct gtcctgatg gtcgtcatct acctgcctgg acagcatggc 1680

ctgcaacgcg ggcatcccga tgccgccgga agcgagaaga atcataatgg ggaaggccat 1740
ccagcctcgc gtcgcaacg ccagcaagac gtagccagc gcgtcggccg ccatgccggc 1800
gataatggcc tgcttctcgc cgaaacgttt ggtggcggga ccagtacga aggcttgagc 1860
gagggcgtgc aagattccga ataccgaag cgacaggccg atcatcgtcg cgctccagcg 1920

aaagcgggtcc tcgccgaaaa tgaccagag cgtgcccgc accgtccta cgagttgcat 1980
gataaagaag acagtataa gtgcggcgac gatagtcacg ccccgcgccc accggaagga 2040
gctgactggg ttgaaggctc tcaagggcat cggtcgacgc tctcccttat gcgactcctg 2100

cattaggaag cagcccagta gtaggttgag gccgttgagc accgccgccc caaggaatgg 2160
tgcatgcaag gagatggcgc ccaacagtcc cccggccacg gggcctgcc ccataccac 2220
gccgaaacaa gcgtcatga gcccgaagtg gcgagcccga tcttcccat cggtgatgtc 2280
ggcgatatag gcgccagcaa ccgcacctgt ggcgccggtg atgccggcca cgatgcgtcc 2340
ggcgtagagg atccacagga cgggtgtggt cgccatgac gcgtagtcca tagtggtcc 2400
aagtagcgaa gcgagcagga ctggcgggcg gccaaagcgg tcggacagt ctccgagaac 2460
gggtgcgcat agaaattgca tcaacgcata tagcgctagc agcacgcat agtgactggc 2520

gatgctgtcg gaatggacga tatcccgcaa gaggcccggc agtacccg 2567

<210> 59

<211> 5009

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (3040)..(3047)

<223> n is a, c, g, or t

<400> 59

tcgagctttt acactttatg cttccggctc gtatgttggt tggaattgtg agcggataac 60
aatttcacac atgattacgg attcaacgtc gtgactggta aaaccgggc gttacccaac 120
ttaatcgctt tgcagcacat ccccttttcg ccagcaggcg taataaggaa aggattcatg 180
tactatttga aaaacacaaa cttttggatg ttcggtttat tctttttctt ttactttttt 240

atcatgggag cctacttccc gtttttcccg atttggttac atgatataa ccatatcagc 300
aaaagtata cgggtattat ttttgccgtc atttctctgt tctcgtatt attccaaccg 360
ctgtttggtc tgctttctga caaactcggc ctacgcaaat acctgctgtg gattattacc 420
ggcatgttag tgatgtttgc gccgttcttt atttttatct tcgggccact gctgcagtac 480
aacattttag tagggctcat gtgttggtgtt atttatctag gctttagttt taacgccggt 540
gcgccagcag tagaggcatt tattgagaaa gtcagccggc gcagtaattt cgaatttggt 600
cgcgcgcgga tgtttggcag gtgttggttg gcgctggttg cctcgattgt cgggatcatg 660

ttcaccatta ataatacagtt tgttttctgg ctgggctctg gcagttgtct catcctcgcc	720
gttttactct ttttcgcaa aacggacgcg cctcaagtg ccacggttgc caatgcggta	780
ggtgccaacc attcggcatt tagccttaag ctggcactgg aactgttcag acagccaaaa	840
ctgtggtttt tgcactgta tgttattggc gtttccctca cctacgatgt ttttgaccaa	900
cagtttgcta atttctttac ttcgttcttt gctaccggtg aacagggtac ccgcgtattt	960
ggctacgtaa cgacaatggg cgaattactt aacgcctcga ttatgttctt tgcgccactg	1020
atcattaatc gcatcgggtg gaagaatgcc ctgctgctgg ctggcactat tatgtctgta	1080
cgtattattg gctcatcggt cgccacctca gcgctggaag tggttattct gaaaacgctg	1140
catatgtttg aagtaccgtt cctgctgggt ggctccttta aatatattac tagtcagttt	1200
gaagtgcgtt tttcagcgac gatttatctg gtcagtttca gtttcttta gcaactggcg	1260
atgattttta tgtctgtact ggcgggcaat atgtatgaaa gcataggttt ccaaggcgct	1320
tatctggtag tgggtctggt ggcgctgggc ttcacctaa tttccgtgtt cagccttagc	1380
ggccccggcc cgctttccct gctgcgtcgt cagggtgaatg aagtcgctta aaggcctcga	1440
tgcagctagc atgctaactt gattcgttac caattatgac aacttgacgg ctacatcatt	1500
cactttttct tcacaaccgg cacggaactc gctcgggctg gccccggtgc attttttaaa	1560
taccgcgag aaatagagtt gatcgtcaaa accaacttg cgaccgacgg tggcgatagg	1620
catccgggtg gtgtcaaaa gcagcttcgc ctggctgata cgttggctct cgcgccagct	1680
taagacgcta atccctaact gctggcggaa aagatgtgac agacgcgacg gcgacaagca	1740
aacatgctgt gcgacgtgg cgatatcaaa attgctgtct gccagggtgat cgctgatgta	1800
ctgacaagcc tcgctgaccc gattatccat cggtaggatgg agcgactcgt taatcgcttc	1860
catgcgccgc agtaacaatt gctcaagcag atttatcgcc agcagctccg aatagcgccc	1920
ttccccctgc cggcggttaa tgatttgcgc aaacaggtcg ctgaaatgcg gctggtagcg	1980
ttcatccggg cgaaagaacc ccgtattggc aaatattgac ggccagttaa gccattcatg	2040
ccagtaggcg cgcgacgaa agtaaaccca ctggtgatac cattcgcgag cctccggatg	2100
acgaccgtag tgatgaatct ctctggcgga gaacagcaaa atatcacccg gtcggcaaac	2160
aaattctcgt cctgatattt tcaccacccc ctgaccgca atggtgagat tgagaatata	2220
acctttcatt cccagcggtc ggtcgataaa aaaatcgaga taaccgttgg cctcaatcgg	2280
cgttaaaccc gccaccagat gggcattaaa cgagtatccc ggcagcaggg gatcattttg	2340
cgcttcagcc atacttttca tactcccccc attcagagaa gaaaccaatt gtccatattg	2400
catcagacat tgccgtcact gcgtctttta ctggctcttc tcgctaacca aaccggtaac	2460
cccgttattt aaaagcattc tgtaacaaag cgggaccaaa gccatgacaa aaacgcgtaa	2520

caaaagtgtc tataatcacg gcagaaaagt ccacattgat tatttgcacg gcgtcacact 2580
 ttgctatgcc atagcatttt tatccataag attagcggat cctacctgac gctttttatc 2640
 gcaactctct actgtttctc catacccggt tttttggggt agcgattgaa aacgatgcag 2700
 ttttaagggtt acacctataa aagagagagc cgttatcgtc tgtttgtgga tgtacagagt 2760

gatattattg acacgcccg gcgacggatg gtgatccccc tggccagtgc acgtctgctg 2820
 tcagataaag tctcccggtga actttacccg gtggtgcata tcggggatga aagctggcgc 2880
 atgatgacca ccgatatggc cagtgtgccg gtctccgtta tcggggaaga agtggctgat 2940
 ctgagccacc gcgaaaatga catcaaaaac gccattaacc tgatgttttg gggaatataa 3000
 tcttctagac atacaatgga agagcagagc cttggtctcn nnnnnnaag cttgatatcg 3060
 aattcctgca gcccggggga tcccatggta cgcgtgctag aggcacaaa taaaacgaaa 3120
 ggctcagtcg aaagactggg cctttcggtt tatctgttgt ttgtcggatga acgtctcct 3180

gagtaggaca aatccgccgc cctagacctt ggcttccggc tgcggcgagc ggtatcagct 3240
 cactcaaagg cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg 3300
 tgagcaaaag gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggctgttttc 3360
 cataggctcc gccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga 3420
 aaccgacag gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct 3480
 cctgttccga cctgcccgt tactcggtac ctgtccgcct ttctcccttc gggaagcgtg 3540
 gcgctttctc aatgctcacg ctgtaggtat ctcagttcgg ttaggtcgt tcgctccaag 3600

ctgggctgtg tgcacgaacc cccgttcag ccgaccgct gcgccttacc cggtaactat 3660
 cgtcttgagt ccaaccgggt aagacacgac ttatcgccac tggcagcagc cactggtaac 3720
 aggattagca gagcgaggta ttagggcgt gctacagagt tcttgaagtg gtggcctaac 3780
 tacggctaca ctagaaggac agtatttgggt atctgcgctc tgctgaagcc agttaccttc 3840
 ggaaaaagag ttggtagctc ttgatccggc aaacaaacca ccgctggtag cgggtgtttt 3900
 ttgtttgca agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc 3960
 ttttctacgg ggtctgacgc tcagtgggaac gaaaactcac gttaagggat ttgggtcatg 4020

actagtgtt ggattctcac caataaaaaa cggccggcgg caaccgagcg ttctgaacaa 4080
 atccagatgg agttctgagg tcattactgg atctatcaac aggagtccaa gcgagctcga 4140
 tatcaaatga cggccgccc tgccactcat cgcagtactg ttgtaattca ttaagcattc 4200
 tgccgacatg gaagccatca cagacggcat gatgaacctg aatcgccagc ggcacagca 4260
 ccttgcgcc ttgcgtataa tatttgccca tggtgaaaac gggggcgaag aagttgtcca 4320
 tattggccac gtttaaatca aaactgggtga aactcaccca gggattggct gagacgaaaa 4380

acatatcttc aataaacctt ttagggaaat aggccagggtt ttcaccgtaa cacgccacat 4440

cttgcaata tatgtgtaga aactgccgga aatcgtcgtg gtattcactc cagagcgatg 4500

aaaacgtttc agtttgcctc tggaaaacgg tgaacaagg gtgaacacta tcccatatca 4560

ccagctcacc gtctttcatt gccatcagga attccggatg agcattcctc aggcgggcaa 4620

gaatgtgaat aaaggccgga taaaacttgt gcttattttt ctttacggtc tttaaaaagg 4680

ccgtaatatc cagctgaacg gtctgggttat aggtacattg agcaactgac tgaaatgcct 4740

caaaatgttc tttacgatgc cattgggata tatcaacggg ggtatatcca gtgatttttt 4800

tctccatttt agcttcccta gctcctgaaa atctcgataa ctcaaaaaat acgcccggta 4860

gtgatcttat ttcattatgg tgaaagttag aacctcttac gtgccgatca acgtctcatt 4920

ttcgccagat atcgacgtct aagaacctat tattatcatg acattaacct ataaaaatag 4980

gcgtatcacg aggccttttc gtcttcacc 5009

<210> 60

<211> 6439

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 60

taatacgact cactataggg agaccacaac ggtttccctc tagagagaca ataacctga 60

taatgcttca ataatttga aaaaggaaga gtatgcctaa gaagaagaga aaggtgggta 120

ccaccaaggt gaaggactac tacataggct tggacatcgg cacctctagc gtcgggtggg 180

ccgtcaccga tgaagcctat aacgtgctta agtttaatag caagaaaatg tggggcgtgc 240

ggctgttcga cgacgctaag acggcagagg agcgtagggg ccagcgagga gcaagacgac 300

gtctggatcg gaagaaggag agactcagcc tgctgcagga cttcttcgcc gaagaggtag 360

caaaggtcga cccaacttc ttcctcaggc tggacaattc cgatctgtac atggaagata 420

aggaccagaa actgaaaagc aaatatacac tgttcaacga caaggacttc aaggataaga 480

attttcataa gaagtacccc acaatacatc acctgctgat ggatctgac gaggacgaca 540

gtaagaagga catccggctc gtctacctgg cctgtcacta tttgctcaag aacaggggtc 600

atttcatctt cgagggccag aagttcgaca cttaatcaag cttcgagaac agtttgaacg 660

agctcaaagt tcatttgaac gacgagtatg gactggacct cgaatttgac aacgagaacc 720

tgattaacat cttgactgac caaaaactca ataaaacggc caagaagaag gagctgaagt 780

ccgtaatcgg cgacaccaag ttcctcaaag ccgtttccgc gataatgatc ggctctagcc 840

agaaactcgt cgacttggtc gagaaccccg aggatttcga cgactctgcg ataaagtccg 900
 ttgactttctc aactacctct ttcgacgaca agtactctga ctatgaactc gctctgggtg 960
 acaagatcgc tctggtcaac atccttaagg aaatttacga tagctccatc ctcgagaacc 1020

 tgctcaaaga ggcagacaag tctaaggacg gtaacaaata tatcagtaat gcattcgtga 1080
 agaagtacaa taaacacgga caagatctga aagagttaa acgtctggta cgacaatata 1140
 acaagagtgc gtattttgat attttcagat ccgagaagggt gaatgacaat tacgtcagct 1200
 aactaaaag ctcaattagc aacaataaac gcgtcaaagc aaacaagttc actgatcaag 1260
 aggcctttcta caaattcgcc aagaaacatc tggagacaat caagtataag atcaacaagg 1320
 taaacggctc caaggcagat ctggagctga ttgacgggat gctgcgggac atggagtcca 1380
 agaactttat gcccaaaatt aagtccagtg acaacgggggt gattccatac cagctcaagc 1440

 tgatggaatt gaacaaaata ctcgagaatc agtcaaagca tcacgagttc ctcaatgtca 1500
 gcgacgagta cggctccgtg tgtgataaaa tcgcatctat catggagtcc cgtatccctt 1560
 actacgtggg acccctgaac cccaatagca agtacgcctg gatcaagaag cagaaagata 1620
 gtgagattac tccttggaac ttcaaggacg tcgtggacct cgactccagc agagaggagt 1680
 tcattgactc acigatcgga cgctgtactt accttaagga cgagaaggtc cttcccaaag 1740
 cttctttgct gtataacgaa tacatggtgc tgaacgagct gaataacctg aagttgaacg 1800
 accttcccat caccgaggag atgaagaaga agatatttga ccagttgttc aaaacaagaa 1860

 agaaggtcac ctttaaagcg gtggcaaacc tgctgaagaa ggagttcaac atcaacggcg 1920
 agattctgct ctctgggacc gacggtgact tcaagcaggg cttgaactca tacaatgact 1980
 tcaaagctat cgtgggcgat aaagtcgatt ccgatgatta ccgggacaag attgaggaga 2040
 tcattaaact gatagttctt tacggtgacg ataagagtta cttcagaag aagattaaag 2100
 ctgggtatgg aaaatacttc accgacagtg agattaagaa aatggcgggg ctgaactaca 2160
 aggattgggg aaggctctca aagaagctgc tgacgggact cgagggtgca aacaagatca 2220
 ctggagagcg gggtccatt attcacttca tgagggaata taaccttaat ctgatggagc 2280

 ttatgtcagc ttcatctacg ttaccgaag agatacagaa acttaacccc gtggatgacc 2340
 gcaagctgtc atacgaaatg gtggacgaac tgtaccttc tcccagtgtg aaacggatgc 2400
 tctggcagtc cctgcgcacg gtcgacgaga taaagaacat catgggaacc gacagtaaga 2460
 agattttcat cgagatggct cggggttaagg aagaggtgaa agcccgaag gagtcaagga 2520
 agaaccaact gctgaagttc tataaagacg gaaagaaggc attcatcagc gagattggcg 2580
 aggagaggta ctcttacttg ctttctgaga tagagggtga ggaagagaat aagtttcgat 2640

gggataacct gtacctttat tatactcaac tgggtcgctg catgtactct ttggaaccta 2700

tcgacatatc tgagctgtct tcaaagaata ttacgatca ggatcatatc taccccaaaa 2760

gcaagattta cgacgacagt atcgagaata ggggtgctggt gaagaaggac cttaactcca 2820

agaagggtaa cagctatcct atcccagacg aaatcctgaa caagaactgt tacgcctact 2880

ggaagatcct gtacgataaa ggtctttatcg ggcagaagaa gtacactcgg ctgacccgga 2940

gaactggctt cacggacgac gagctcgctt agttcatctc aagacagatc gtggaaacta 3000

gacaagcaac aaaggagact gctaacctgc tcaagacaat atgtaagaac tccgagatcg 3060

tgtattccaa agccgagaac gcaagtcggt ttaggcaaga gttcgacatc gtgaagtgt 3120

gggcggtgaa cgatcttcat catatgcacg atgcctacat caacatcata gtggggaacg 3180

tgtataacac caagttcacg aaggacccta tgaatttcgt aaagaagcag gaaaaggcgc 3240

ggagctacaa tctcgagaat atgttcaagt acgatgtgaa acgtggcgga tacaccgctt 3300

ggatcgccga tgacgagaag ggcaccgtga agaacgcgag tattaaacgt atccggaagg 3360

agctggaagg cacaattat aggttcacaa gaatgaacta cattgagtct ggagcgcttt 3420

tcaacgccac tctccagcgg aagaataagg gctccagacc cctgaaggac aaaggcccga 3480

aatcttccat cgagaaglac ggcggtctaca caaacatcaa taaagcctgt ttcgctgttc 3540

ttgacatcaa gtctaagaac aagattgaga ggaagctgat gcccgtcgag cgtgagatct 3600

atgccaaaca gaagaacgac aagaagctgt ccgacgagat tttctcaaag tacctcaagg 3660

accgatttgg catcgaggac tacagggttg tctaccaggt ggtgaaaatg cgcacactgc 3720

tcaagatcga cggcagctac tacttcatca caggcggttc tgataagacc ctggagtgtc 3780

gatctgctct gcagctgatt ctccctaaga agaacgagtg ggcgatcaaa cagatcgaca 3840

agtcttccga aaacgactat ctgacgatcg agcgtatcca ggacctgacc gaggagctgg 3900

tgtataacac tttcgacatc atcgtcaaca agttcaagac cagtgtcttc aagaagtctt 3960

tccttaactt gtttcaggac gacaagattg agaacattga cttcaagttt aagtcgatgg 4020

acttcaagga gaaatgcaag acacttctca tgcgtgtcaa ggcgattcgg gcatccggcg 4080

tgaggcagga tctcaagtc atcgacctca agtctgatta cggacggctc agttcaaaga 4140

ccaacaacat cggcaattac caggagtcca agattattaa tcagtcctac actggactgt 4200

tcgagaatga ggtcgatctc ctgaagctgg gatcctaccc atacgatgtt ccagattacg 4260

cggccgctcc aaaaaagaaa agaaaagttg cggttagcca tcatcacat caccatcatc 4320

attaaggctg ctaacaaagc ccgaaaggaa gctgagttgg ctgctgccac cgctgagcaa 4380

taactagcat aaccccttgg ggcctctaaa cgggtcttga ggggtttttt gctgaaagga	4440
ggaactatat ccggatatcc acaggacggg tgtggtcgcc atgatcgct agtcgatagt	4500
ggctccaagt agcgaagcga gcaggactgg gcggcgcca aagcggtcgg acagtgtcc	4560
gagaacgggt gcgcatagaa attgcatcaa cgcatatagc gctagcagca cgccatagt	4620
actggcgatg ctgtcggaat ggacgatatc ccgcaagagg cccggcagta ccggcataac	4680
caagcctatg cctacagcat ccagggtgac ggtgccgagg atgacgatga gcgcattgtt	4740
agatttcata cacgggtgcct gactgcgtta gcaatttaac tgtgataaac taccgatta	4800
aagcttatcg atgataagct gtcaaacatg agaattctta gaaaaactca tcgagcatca	4860
aatgaaactg caatttatc atatcaggat tatcaatacc atatttttga aaaagccgtt	4920
tctgtaatga aggagaaaac tcaccgaggc agttccatag gatggcaaga tcttggtatc	4980
ggtctcgcat tccgactcgt ccaacatcaa tacaacctat taatttcccc tcgtcaaaaa	5040
taaggttatc aagtgagaaa tcacatgag tgacgactga atccggtgag aatggcaaaa	5100
gcttatgcat ttctttccag acttggtcaa caggccagcc attacgctcg tcatcaaat	5160
cactcgcatc aaccaaaccg ttattcattc gtgattgcgc ctgagcgaga cgaaatacgc	5220
gatcgctgtt aaaaggacaa ttacaacag gaatcgaatg caaccggcgc aggaacactg	5280
ccagcgcatc aacaatatc tcacctgaat caggatattc ttctaatacc tggaatgctg	5340
ttttccggg gatcgcatg gtgagtaacc atgcatcatc aggagtacgg ataaaatgct	5400
tgatggtcgg aagaggcata aattccgtca gccagtttag tctgaccatc tcatctgtaa	5460
catcattggc aacgtacct ttgccatgtt tcagaaacaa ctctggcgca tcgggcttcc	5520
catacaatcg atagattgtc gcacctgatt gcccgacatt atcgcgagcc catttatacc	5580
catataaatc agcatccatg ttggaattta atcgcggcct cgagcaagac gtttcccggt	5640
gaatatggct cataacaccc ctgtatttac tgtttatgta agcagacagt tttattgttc	5700
atgacaaaaa tcccttaacg tgagttttcg ttccactgag cgtcagaccc cgtagaaaag	5760
atcaaaggat cttcttgaga tctttttttt ctgcgcgtaa tctgtgctt gcaaacaaaa	5820
aaaccaccgc taccagcggg ggtttgtttg ccggatcaag agctaccaac tctttttccg	5880
aaggtaaactg gcttcagcag agcgcagata ccaataactg tcttctagt gtagccgtag	5940
ttaggccacc acttcaagaa ctctgtagca ccgcctacat acctcgtct gtaaatcctg	6000
ttaccagtgg ctgtgccag tggcgataag tcgtgtctta ccgggttgga ctcaagacga	6060
tagttaccgg ataaggcgca gcggtcgggc tgaacggggg gttcgtgcac acagcccagc	6120
ttggagcgaa cgacctacac cgaactgaga tacctacagc gtgagctatg agaaagcgcc	6180
acgcttcccg aaggagaaaa ggcggacagg tatccggtaa gcggcagggt cggaacagga	6240

gagcgcacga gggagcttcc aggggggaaac gcctgggtatc tttatagtcc tgtcgggttt 6300
cgccacctct gacttgagcg tcgatttttg tgatgctcgt caggggggcg gaggcctatgg 6360
aaaaacgcca gcaacgcggc ctttttacgg ttcttggcct tttgctggcc ttttgctcac 6420
atgttcgata ccgcgaaat 6439

<210> 61

<211> 9542

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 61

ggtcgctgag tagtgcgca gcaaaattta agctacaaca aggcaaggct tgaccgacaa 60
ttgcatgaag aatctgctta gggttaggcg ttttgcgctg cttcgcatg tacgggccag 120
atatacgctg tgacattgat tattgactag ttattaatag taatcaatta cggggtcatt 180
agttcatagc ccatatatgg agttccgctg tacataactt acggtaaatg gcccgcttgg 240
ctgaccgccc aacgaccccc gcccatgac gtcaataatg acgtatgttc ccatagtaac 300
gccaataggg actttccatt gacgtcaatg ggtggagtat ttacggtaaa ctgccactt 360

ggcagtagat caagtgtatc atatccaag tacgccccct attgacgtca atgacggtaa 420
atggcccgcc tggcattatg ccagtagat gaccttatgg gactttccta cttggcagta 480
catctacgta ttagtcatcg ctattacat ggtgatgcgg ttttggcagt acatcaatgg 540
gcgtggatag cggtttgact cacggggatt tccaagcttc caccctattg acgtcaatgg 600
gagtttgttt tggcacaaa atcaacggga ctttcaaaa tgcgtaaca actccgcccc 660
attgacgcaa atgggcggtg ggctgtacg gtgggaggtc tatataagca gagctctctg 720
gtaactaga gaaccactg ctactggct tatcgaaatt aatagcactc actataggga 780

gaccaagct ggctagcgtt taaacttaag ctggccacca tgcctaagaa gaagagaaag 840
gtgggtaccg gctatacaat tggcctggat ctgggcgttg cctctcttgg ctgggccgtc 900
gtgaatgatg agttagcagg gctggaaagc tgcagcaaca tctttcctgc cgccgagagc 960
gccaacaacg tggaaagaag aggttccgg caaggcagac ggctgagcag aagaagaagg 1020
accggatca gcgacttcag aaagctgtgg gagaagtccg gcttcgaggt gccagcaat 1080
gagctgaatg aggtgctgca gtaccggatc aaggcatga acgacaagct gagcgaggac 1140
gagctgtacc acgtgctgct gaacagcctg aagcacagag gcatcagcta cctggacgac 1200

gccgatgatg agaacgcctc tggcgattat gccgcctcta tcgcctaca cgagaaccag 1260

ctgaaaacaa agctgccctg cgagatccag tgggagagat acaagaagta cggcgcttac	1320
cggggcaaca tcacaatcca agaaggcggc gagcccctga cactgagaaa tgtgtttacc	1380
accagcgcct acgagaaaga gatccagaaa ctgctggacg tgcagagcat gagcaacgag	1440
aaagtgacca agaagttcat cgacgagtac ctcaagatct tcagccggaa gagagagtac	1500
tacatcgccc ctggcaacaa gaagtccaga accgactacg gcgtgtacac cacacagaag	1560
aacgaggacg gcacctacca caccgagcag aacctgttcg ataagctgat cggcaagtgc	1620
agcgtgtacc ctgatgagcg tagagccgct ggcgccacat acacagccca agagttcaac	1680
ctgctgaacg atctgaacaa cctggtcatc gacggccgga agctggacga gcaagagaag	1740
tgtcagatcg tggatgccgt gaagcacgcc aagaccgtga acatgaagaa catcattgcc	1800
aaagtgatcg gcaccaaggc caacagcatg aacatgaccg gcgccagaat cgacaagaat	1860
gagaaagaaa tcttccacag cttcgaggcc tacaacaagc tgcggaaggc cctggaagag	1920
atcgacttcg acatcgagac actgagcacc gacgagctgg atgccattgg agagggtctg	1980
acctgaaca ccgaccggaa gtctatccag aacggcctgc aagagaaacg gatcgtggtg	2040
cccgatgaag tgcgggatgt gctgatcgcc accagaaaga gaaatggcag cctgttctcc	2100
aagtggcaga gcttcggcat ccgatcatg aaggaaactga tcccagagct gtacgccag	2160
cctaagaacc agatgcagct gctgaccgac atgggcgtgt tcaagaccaa ggacgagaga	2220
ttcgtggaat acgacaagat cccagcgac ctgatcacg aagagatcta caacccctg	2280
gtggccaaga cagtgcggat caccgttaga gtgctgaacg ccctgatcaa gaagtatggc	2340
taccccgacc gggtcgtgat cgagatgcc agagataaga actccgagga agagaagaag	2400
cggatcgccg acttccagaa gaacaacgaa aacgagcttg gcggcatcat caagaaagtg	2460
aagtccgagt acggcatcga gatcaccgac gccgacttta agaaccacag caagctgggc	2520
ctgaagctga gactgtggaa cgagcagaat gagacatgcc cctacagcg caagcacatc	2580
aagatcgacg acctgtcaa caacccaac atgttcgagg tggaccacat catccctctg	2640
agcatcagct tcgacgacg cagagccaac aaggtgctgg tgtacgccgc cgaaccag	2700
aacaaggga acagaacccc tatggcctac ctgagcaacg tgaacagaga gtgggacttc	2760
cacgagtaca tgagcttcgt gctgagcaac tacaaggga ccatctacgg caagaagcgg	2820
gacaatctgc tgttctccga ggacatctac aagatcgatg tgctgcaggg cttcatctcc	2880
cggaacatca acgacaccag atacgcctct aaagtgatcc tgaactccct gcagagcttt	2940
ttcggcagca aagaatgcga cacaaagtg aaggtcgtgc ggggcacctt cacacaccag	3000
atcgcgatga acctgaagat cgagaagaac cggaagagt cctacgtgca ccacgccgtg	3060
gatgctatgc tgattgcctt cagccagatg ggctacgacg cctaccacaa actgaccgag	3120

aagtatatcg actacgagca cggcgagttc gtggaccaga agggatacga gaagctgatt	3180
gagaacgacg tggcctacag agagacaacc tatcagaaca agtggatgac catcaagaag	3240
aatatcgaga tcgccgctga gaaaaacaag tactggtatc aagtgaatcg gaagtccaac	3300
cggggcctgt gcaaccagac catctatggc accagaaacc tggacggcaa aaccgtgaag	3360
atctccaagc tggacatccg gaccgacgac ggcatcaaaa agtttaaggg catcgtggaa	3420
aagggaagc tggaaagggt cctgatgtac cggaacgacc ccaagacctt cgagtggctg	3480
ctgcagatct ataaggacta cagcgacgac aagaaccctt tcgtgcagta cgagtctgag	3540
acaggcgacg tgatcaaaaa ggtgtccaag acaacaacg gcccacaaagt gtgcgagctg	3600
agatacgagg atggcgaagt gggctcctgc atcgacatca gccacaaata cggctacaag	3660
aagggcagca agaaagtcac cctggattct ctgaaccctt accggatgga cgtgtactac	3720
aacaccaagg acaaccggtc ctacttcgtg ggctgaagt actccgacat caagtcccag	3780
ggcgacagct acgtgatcga cgaggataag tatgcccgcc ctctgggtgca agaaaagatc	3840
gtgccagaag gcaaggcgag atccgatctg accgagctgg gctatgagtt caagctgtcc	3900
ttctacaaga acgagatcat cgagtacgag aaggacgggg agatctacgt cgagcggttc	3960
ctgtccagaa caatgcctaa agtgtccaac tatatcgaga caaagcccct ggaagccgcc	4020
aagttcgaga agagaaacct cgtgggcctc gccaagacaa gccggatcag aaagatcaga	4080
gtggacatcc tggggaaccg ctacctgaac agcatggaaa acttcgactt cgtcgtgggc	4140
cacaagggat cctaccata cgatgttcca gattacgcgg ccgctccaaa aaagaaaaga	4200
aaagtgaat tcggcgagcag cggcgccacc aacttcagcc tgctgaagca ggccggcgac	4260
gtggaggaga accccggccc catggtgagc aaggcgaggg aggataacat ggccatcatc	4320
aaggagtcca tgcgcttcaa ggtgcacatg gagggctccg tgaacggcca cgagttcgag	4380
atcgaggcg agggcgaggg ccgcccctac gagggcaccc agaccgcaa gctgaagtg	4440
accaagggtg gccccctgcc cttcgcttgg gacatcctgt cccctcagtt catgtacggc	4500
tccaaggcct acgtgaagca ccccgccgac atccccgact acttgaagct gtccttcccc	4560
gagggttcca agtgggagcg cgtgatgaac ttcgaggacg gcggcgtggt gaccgtgacc	4620
caggactcct ccttcgagga cggcgagttc atctacaagg tgaagctgcg cggcaccaac	4680
ttccctccg acggccccgt aatgcagaag aagaccatgg gctgggaggc ctctccgag	4740
cggatgtacc ccgaggacgg cgccctgaag ggcgagatca agcagaggct gaagctgaag	4800
gacggcggcc actacgacgc tgaggtcaag accacctaca aggccaagaa gcccgtgcag	4860
ctgcccggcg cctacaacgt caacatcaag ttggacatca cctcccacaa cgaggactac	4920

accatcgtgg aacagtacga acgcgccgag ggccgccact ccaccggcgg catggacgag	4980
ctgtacaagt agtcgagtc tagagggccc gtttaaacc gctgatcagc ctcgactgtg	5040
ccttctagtt gccagccatc tgttgtttgc cctcccccg tgccttctt gaccctggaa	5100
ggtgccactc cactgtcct ttcctaataa aatgaggaaa ttgcatcgca ttgtctgagt	5160
aggtgtcatt ctattctggg gggtaggggtg gggcaggaca gcaaggggga ggattgggaa	5220
gacaatagca ggcatgctgg ggatgcggtg ggctctatgg cttctgaggc ggaaagaacc	5280
agctggggct ctagggggta tccccacgcg cctgtagcg gcgcattaag cgcggcgggt	5340
gtggtggtta cgcgagcgt gaccgtaca ctgcccagcg ccctagcgcc cgctccttc	5400
gctttcttc cticctttct cgccacgttc gccggcttc cccgtcaagc tctaaatcgg	5460
gggtccctt tagggttccg atttagtgct ttacggcacc tcgaccccaa aaaacttgat	5520
tagggtagt gttcacgtag tgggccatcg cctgataga cggtttttcg cctttgacg	5580
ttggagtcca cgttctttaa tagtggactc ttgttcaaaa ctggaacaac actcaaccct	5640
atctcggctc attcttttga tttataaggg attttgccga tttcgcccta ttggttaaaa	5700
aatgagctga tttacaataa attaacgcg aattaattct gtggaatgtg tgtcagttag	5760
ggtgtggaaa gtccccagcg tccccagcag gcagaagtat gcaaagcatg catctcaatt	5820
agtcagcaac caggtgtgga aagtccccag gctccccagc aggcagaagt atgcaaagca	5880
tgcactcaa ttagtcagca accatagtcc cgcccctaac tccgccatc ccgccctaa	5940
ctccgccag tccgcccat tctccgcc ctaggtgact aattttttt atttatgcag	6000
aggccgagcg cgctctgcc tctgagctat tccagaagta gtgaggagcg ttttttgag	6060
gcctaggctt ttgcaaaaag ctccgggag ctgttatatc cattttcgga tctgatcaag	6120
agacaggatg aggatcgtt cgcatgattg aacaagatgg attgcacgca ggttctccg	6180
ccgcttgggt ggagagccta ttcggctatg actgggcaca acagacaatc ggctgctctg	6240
atgccgccgt gttccggctg tcagcgcagg ggcgccgggt tctttttgtc aagaccgacc	6300
tgtccggtgc cctgaatgaa ctgcaggacg aggcagcgcg gctatcgtgg ctggccacga	6360
cgggcgttcc ttgcgcagct gtgctcgacg ttgtactga agcgggaagg gactggctgc	6420
tattgggcga agtgccgggg caggatctcc tgcactca ccttgctcct gccagaaaag	6480
tatccatcat ggctgatgca atgcggcggc tgcatacgt tgatccggt acctgccat	6540
tcgaccacca agcgaacat cgcatcgagc gagcacgtac tcggatggaa gccggtcttg	6600
tcgatcagga tgatctggac gaagagcatc aggggctcgc gccagccgaa ctgttcgcca	6660

ggctcaaggc gcgcatgccc gacggcgagg atctcgtcgt gacccatggc gatgcctgct 6720
 tgccgaatat catggtggaa aatggccgct tttctggatt catcgactgt ggccggctgg 6780
 gtgtggcgga ccgctatcag gacatagcgt tggctacccg tgatattgct gaagagcttg 6840
 gcggcgaatg ggctgaccgc ttcctcgtgc tttacggtat cggcgctccc gattcgcagc 6900
 gcatcgcttt ctatcgcttt cttgacgagt tcttctgagc gggactctgg ggttcgaaat 6960
 gaccgaccaa gcgacgcccc acctgccatc acgagatttc gattccaccg ccgccttcta 7020
 tgaaaggttg ggcttcggaa tcgttttccg ggacgccggc tggatgatcc tccagcgcg 7080

 ggatctcatg ctggagtctt tcgcccaccc caacttgttt attgcagctt ataatggtta 7140
 caaataaagc aatagcatca caaatctcac aaataaagca tttttttcac tgcattctag 7200
 ttgtggtttg tccaaactca tcaatgtatc ttatcatgtc tgtataccgt cgacctctag 7260
 ctagagcttg gcgtaatcat ggtcatagct gtttctctgt tgaaattgtt atccgctcac 7320
 aattccacac aacatacgag ccggaagcat aaagtgtaaa gcctgggggtg cctaagagt 7380
 gagctaactc acattaattg cgttgcgtc actgccgct ttcagtcgg gaaacctgtc 7440
 gtgccagctg cattaatgaa tcggccaacg cgcggggaga ggcggtttgc gtattgggcg 7500

 ctcttccgct tcctcgctca ctgactcgtc gcgctcggtc gttcggctgc ggcgagcgg 7560
 atcagctcac tcaaaggcgg taatacggtt atccacagaa tcaggggata acgcaggaaa 7620
 gaacatgtga gcaaaaggcc agcaaaaggc caggaaccgt aaaaaggcgg cgttgcctggc 7680
 gtttttccat aggtctccgc cccctgacga gcatcacaaa aatcgacgt caagtcagag 7740
 gtggcgaaac ccgacaggac tataaagata ccaggcgttt ccccttgaa gctccctcgt 7800
 gcgctctcct gttccgacct tgccgcttac cggatactg tccgcctttc tcccttcggg 7860
 aagcgtggcg ctttctcata gctcacgtg taggtatctc agttcgggtg aggtcgttcg 7920

 ctccaagctg ggetgtgtgc acgaaccccc cgttcagccc gaccgctgcg ccttatccgg 7980
 taactatcgt cttagtcca acccgtaag acacgactta tcgccactgg cagcagccac 8040
 tggtaacagg attagcagag cgaggtatgt aggcgggtgt acagagttct tgaagtgtg 8100
 gcctaactac ggctacacta gaagaacagt atttggatc tgcgctctgc tgaagccagt 8160
 taccttcgga aaaagagttg gtagctcttg atccggcaaa caaaccaccg ctggtagcgg 8220
 tggttttttt gtttgcaagc agcagattac gcgcagaaaa aaaggatctc aagaagatcc 8280
 tttgatcttt tctacggggt ctgacgctca gtggaacgaa aactcacgtt aagggtttt 8340

 ggtcatgaga ttatcaaaaa ggatcttcac ctagatcctt ttaaattaaa aatgaagttt 8400
 taaatcaatc taaagtatat atgagtaaac ttggtctgac agttaccaat gcttaatcag 8460
 tgaggcacct atctcagcga tctgtctatt tcgttcatcc atagttgcct gactccccgt 8520

cgtgtagata actacgatac gggagggcctt accatctggc cccagtgtg caatgatacc 8580
 gcgagaccca cgctcaccgg ctccagattt atcagcaata aaccagccag ccggaagggc 8640
 cgagcgcaga agtggctcctg caactttatc cgcctccatc cagtctatta attgttgccg 8700
 ggaagctaga gtaagtagtt cgccagttaa tagtttgccg aacgttgttg ccattgctac 8760

aggcatcgtg gtgtcacgt cgctgttttg tatggcttca ttcagctccg gttcccaacg 8820
 atcaaggcga gttacatgat ccccatgtt gtgcaaaaaa gcggttagct ccttcggtec 8880
 tccgatcgtt gtcagaagta agttggccgc agtgttatca ctcatggta tggcagcact 8940
 gcataattct cttactgtca tgccatccgt aagatgtttt tctgtgactg gtgagtactc 9000
 aaccaagtca ttctgagaat agtgtatgcg gcgaccgagt tgctcttgcc cggcgtaaat 9060
 acgggataat accgcgccac atagcagaac tttaaaagtg ctcatcattg gaaaacgttc 9120
 ttcggggcga aaactctcaa ggatcttacc gctgttgaga tccagttcga tgtaaccac 9180

tcgtgcaccc aactgatctt cagcatcttt tactttcacc agcgtttctg ggtgagcaaa 9240
 aacaggaagg caaatgccg caaaaaagg aataaggcg acacggaat gttgaatact 9300
 catactcttc ctttttcaat attattgaag catttatcag ggttattgtc tcatgagcgg 9360
 atacatattt gaatgtattt agaaaaataa acaaataggg gttccgcgca catttccccg 9420
 aaaagtgcc cctgacgtcg acggatcggg agatctcccg atcccctatg gtgcactctc 9480
 agtacaatct gctctgatgc cgcatagtta agccagtatc tgctccctgc ttgtgtgttg 9540
 ga 9542

<210> 62

<211> 2726

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 62

gagggcctat ttcccatgat tccttcatat ttgcatatac gatacaaggc tgtagagag 60
 ataattagaa ttaatttgac tgtaaacaca aagatattag taaaaaac gtgacgtaga 120
 aagtaataat ttcttgggta gtttgagtt ttaaaattat gttttaaaat ggactatcat 180
 atgcttaccg taacttgaaa gtatttcgat ttcttggcct tatatatctt gtggaaagga 240
 cgaaacaccg ccaagtgata aacacgagga gtttaagtac ctagagaaag aaatttcttt 300
 agacctactt aaataaggct ttatgccgag attaaaggat gccgacgggc atcctttttt 360
 gaattctcaa ataaaacgaa aggctcagtc gaaagactgg gcctttcgtt ttatctgttg 420

tttgtcgggtg aacgctctcc tgagtaggac aaatgggtacc ccgcttcctc gctcactgac	480
tcgctacgct cggtcgttcg actgcggcga gcggaatgg cttacgaacg gggcggagat	540
ttcctggaag atgccaggaa gatacttaac agggaagtga gagggccgcg gcaaagccgt	600
ttttccatag gctccgcccc cctgacaagc atcacgaaat ctgacgctca aatcagtgg	660
ggcgaaccac gacaggacta taaagatacc aggcgtttcc ccctggcggc tccctcgtgc	720
gctctcctgt tcttgccctt cggtttaccg gtgtcatcc gctgttatgg ccgcgtttgt	780
ctcattccac gcctgacact cagttccggg taggcagttc gctccaagct ggactgtatg	840
cacgaacccc ccgttcagtc cgaccgctgc gccttatccg gtaactatcg tcttgagtc	900
aacccggaaa gacatgcaaa agcaccactg gcagcagcca ctggtaattg atttagagga	960
gttagtcttg aagtcatgcg ccggttaagg ctaaaactgaa aggacaagtt ttggtgactg	1020
cgctcctcca agccagttac ctcggttcaa agagttggta gctcagagaa ccttcgaaaa	1080
accgccctgc aaggcggttt ttctgttttc agagcaagag attacgcgca gacaaaaacg	1140
atctcaagaa gatcatctta ttaatcagat aaaatatctc tagatttcag tgcaatttat	1200
ctcttcaaat gtagcacctg aagtcagccc catacgatat aagttgttac tagtgcttgg	1260
attctcacca ataaaaaacg cccggcggca accgagcgtt ctgaacaaat ccagatggag	1320
ttctgaggtc attactggat ctatcaacag gagtccaagc gagaagggtt ggtttcgca	1380
ttcacagttc tccgaagaa ttgattggct ccaattcttg gagtggtgaa tccgttagcg	1440
aggtgcccgc ggcttcatt caggtcaggg tggccccgct ccatgcaccg cgacgcaacg	1500
cggggaggca gacaaggtat agggcggcgc ctacaatcca tgccaaccg ttccatgtgc	1560
tcgccgaggc ggcataaatc gccgtgacga tcagcgggtcc aatgatcgaa gttaggctgg	1620
taagagccgc gagcgatcct tgaagctgtc cctgatggtc gtcatctacc tgcttgaca	1680
gcatggcctg caacgcgggc atcccgatgc cgccggaagc gagaagaatc ataatgggga	1740
aggccatcca gcctcgctc gcgaacgcca gcaagacgta gcccagcgcg tcggccgcca	1800
tgccggcgat aatggcctgc ttctcgccga aacgtttggg ggccggacca gtgacgaagg	1860
cttgagcgag ggcgtgcaag attccgaata ccgcaagcga caggccgatc atcgtcgcg	1920
tccagcgaaa gcggtcctcg ccgaaatga cccagagcgc tgccggcacc tgtcctacga	1980
gttgcatgat aaagaagaca gtcataagtg cggcgacgat agtcatgccc cgcgcccacc	2040
ggaaggagct gactgggttg aaggctctca agggcatcgg tcgacgtctc cccttatgcg	2100
actctgcat taggaagcag cccagtagta ggttgaggcc gttgagcacc gccgccgcaa	2160
ggaatggctg atgcaaggag atggcgccca acagtcccc ggccacgggg cctgccacca	2220
taccacgcc gaacaagcg ctcatgagcc cgaagtggcg agcccgatct tccccatcg	2280

tgatgtcggc gatataggcg ccagcaaccg cacctgtggc gccggtgatg ccggccacga 2340
 tgcgtccggc gtagaggatc cacaggacgg gtgtggtcgc catgatcgcg tagtcgatag 2400
 tggctccaag tagcgaagcg agcaggactg ggcggcggcc aaagcggtcg gacagtgtc 2460

cgagaacggg tgcgcataga aattgcatca acgcatatag cgctagcagc acgcatagt 2520
 gactggcgat gctgtcggaa tggacgatat cccgcaagag gcccggcagt accggcataa 2580
 ccaagcctat gcctacagca tccagggtga cgggtccgag gatgacgatg agcgattgt 2640
 tagatttcat acacggtgcc tgactgcgtt agcaatttaa ctgtgataaa ctaccgcatt 2700
 aaagcttatt gatgataagc tgtcaa 2726

<210> 63

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 63

Tyr Pro Tyr Asp Val Pro Asp Tyr Ala

1 5

<210> 64

<211> 27

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 64

taccatacgt atgttcaga ttacgt 27

<210> 65

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 65

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 66

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 66

ccaaaaaaga aaagaaaagt t

21

<210> 67

<211> 236

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 67

Met Val Ser Lys Gly Glu Glu Asp Asn Met Ala Ile Ile Lys Glu Phe

1 5 10 15

Met Arg Phe Lys Val His Met Glu Gly Ser Val Asn Gly His Glu Phe

20 25 30

Glu Ile Glu Gly Glu Gly Glu Gly Arg Pro Tyr Glu Gly Thr Gln Thr

35 40 45

Ala Lys Leu Lys Val Thr Lys Gly Gly Pro Leu Pro Phe Ala Trp Asp

50 55 60

Ile Leu Ser Pro Gln Phe Met Tyr Gly Ser Lys Ala Tyr Val Lys His

65 70 75 80

Pro Ala Asp Ile Pro Asp Tyr Leu Lys Leu Ser Phe Pro Glu Gly Phe

85 90 95

Lys Trp Glu Arg Val Met Asn Phe Glu Asp Gly Gly Val Val Thr Val

100 105 110

Thr Gln Asp Ser Ser Leu Gln Asp Gly Glu Phe Ile Tyr Lys Val Lys

115 120 125

Leu Arg Gly Thr Asn Phe Pro Ser Asp Gly Pro Val Met Gln Lys Lys

130 135 140

Thr Met Gly Trp Glu Ala Ser Ser Glu Arg Met Tyr Pro Glu Asp Gly

145 150 155 160

Ala Leu Lys Gly Glu Ile Lys Gln Arg Leu Lys Leu Lys Asp Gly Gly

165 170 175
His Tyr Asp Ala Glu Val Lys Thr Thr Tyr Lys Ala Lys Lys Pro Val

180 185 190
Gln Leu Pro Gly Ala Tyr Asn Val Asn Ile Lys Leu Asp Ile Thr Ser

195 200 205
His Asn Glu Asp Tyr Thr Ile Val Glu Gln Tyr Glu Arg Ala Glu Gly

210 215 220
Arg His Ser Thr Gly Gly Met Asp Glu Leu Tyr Lys

225 230 235

<210> 68
<211> 711
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic

<400> 68

atggtgagca agggcgagga ggataacatg gccatcatca aggagtcat gcgcttcaag 60
gtgcacatgg agggctccgt gaacggccac gatttcgaga tcgagggcga gggcgagggc 120
cgccctacg agggcaccca gaccgccaag ctgaaggtga ccaagggtgg cccctgccc 180
ttgcctggg acatcctgtc cctcagttc atgtacggt ccaaggccta cgtgaagcac 240
cccgccgaca tccccgacta cttgaagctg tccttccccg agggcttcaa gtgggagcgc 300
gtgatgaact tcgaggacgg cggcgtggtg accgtgacc aggactcctc cctgcaggac 360
ggcgagttca tctacaaggt gaagctgcgc ggcaccaact tcccctccga cgccccgta 420

atgcagaaga agaccatggg ctgggaggcc tcctccgagc ggatgtacc cgaggacggc 480
gccctgaagg gcgagatcaa gcagaggctg aagctgaagg acggcggcca ctacacgct 540
gaggtcaaga ccacctaaa ggccaagaag cccgtgcagc tgcccggcgc ctacaacgtc 600
aacatcaagt tggacatcac ctcccacaac gaggactaca ccatcgtgga acagtacgaa 660
cgcgccgagg gccgccactc caccggcggc atggacgagc tgtacaagta g 711

<210> 69
<211> 7
<212> PRT
<213> Simian virus 40
<400> 69

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 70

<211> 16

<212> PRT

<213> Unknown

<220><223> Synthetic

<400> 70

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

1 5 10 15

<210> 71

<211> 7

<212> PRT

<213> Unknown

<220><223> Synthetic

<400> 71

Pro Ala Ala Arg Val Leu Asp

1 5

<210> 72

<211> 11

<212> PRT

<213> Unknown

<220><223> Synthetic

<400> 72

Arg Gln Arg Arg Asn Glu Leu Lys Arg Ser Pro

1 5 10

<210> 73

<211> 37

<212> PRT

<213> Homo sapiens

<400> 73

Asn Gln Ser Ser Asn Phe Gly Pro Met Gly Gly Asn Phe Gly Gly Arg

1 5 10 15

Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro Arg
 20 25 30

Asn Gln Gly Gly Tyr
 35

<210> 74

<211> 42

<212> PRT

<213> Unknown

<220><223> Synthetic

<400> 74

Arg Met Arg Ile Glx Phe Lys Asn Lys Gly Lys Asp Thr Ala Glu Leu

1 5 10 15

Arg Arg Arg Arg Val Glu Val Ser Val Glu Leu Arg Lys Ala Lys Lys
 20 25 30

Asp Glu Gln Ile Leu Lys Arg Arg Asn Val
 35 40

<210> 75

<211> 8

<212> PRT

<213> Unknown

<220><223> Synthetic

<400> 75

Val Ser Arg Lys Arg Pro Arg Pro

1 5

<210> 76

<211> 8

<212> PRT

<213> Unknown

<220><223> Synthetic

<400> 76

Pro Pro Lys Lys Ala Arg Glu Asp

1 5

<210> 77

<211> 7

<212> PRT

<213> Homo sapiens

<400> 77

Pro Gln Pro Lys Lys Pro Leu

1 5

<210> 78

<211> 9

<212> PRT

<213> Mus musculus

<400> 78

Ser Ala Ile Ile Lys Lys Lys Lys Met

1 5

<210> 79

<211> 5

<212> PRT

<213> Influenza virus

<400> 79

Asp Arg Leu Arg Arg

1 5

<210> 80

<211> 7

<212> PRT

<213> Influenza virus

<400> 80

Pro Lys Gln Lys Lys Arg Lys

1 5

<210> 81

<211> 10

<212> PRT

<213> Hepatitis delta virus

<400> 81

Arg Lys Leu Lys Lys Lys Ile Lys Lys Leu

1 5 10

<210> 82

<211> 10

<212> PRT

<213> Mus musculus

<400> 82

Arg Glu Lys Lys Lys Phe Leu Lys Arg Arg

1 5 10

<210> 83

<211> 19

<212> PRT

<213> Homo sapiens

<400> 83

Lys Arg Gly Asp Glu Val Asp Gly Val Asp Glu Val Ala Lys Lys Lys

1 5 10 15

Ser Lys Lys

<210> 84

<211> 17

<212> PRT

<213> Homo sapiens

<400> 84

Arg Lys Cys Leu Gln Ala Gly Met Asn Leu Glu Ala Arg Lys Thr Lys

1 5 10 15

Lys

<210> 85

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 85

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn

1 5 10 15

Pro Gly Pro

<210> 86

<211> 57

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 86

gccaccaact tcagcctgct gaagcaggcc ggcgacgtgg aggagaacct cggcccc 57

<210> 87

<211> 108

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 87

guuugagagu uaugugaaaa caugacgagu ucaaaaaaa auuuuuucaa accgccuauu 60

uauaggccgc agauguucug cauuuugcuu gcuauugcaa gcuuuuuu 108

<210> 88

<211> 110

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 88

guuugagagu uauguagaaa uacaugacga guucaaaaa aaauuuauuc aaaccgccua 60

uuuauaggcc gcagauguuc ugcauuuugc uugcuauugc aagcuuuuuu 110

<210> 89

<211> 20

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 89

ccaagugaua aacacgagga 20

<210> 90

<211> 21

<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 90	
gucaccucca augacuaggg u	21
<210> 91	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 91	
gccgccauug acagaggac	20
<210> 92	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 92	
aacugguacc gcaugagccc	20
<210> 93	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 93	
caucaggcuc ucagcucagc	20
<210> 94	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 94	
aggugccguu uguucauuuu	20

<210> 95
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 95
 ccaguuguag caccgccag 20
 <210> 96
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 96
 ucuccccagc ccugcucgug 20

 <210> 97
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 97
 ucugugaaug uuagaccgau 20
 <210> 98
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 98
 ccaugggagc agcuggucag 20
 <210> 99
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 99

gcaagagacc cacacaccgg	20
<210> 100	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 100	
acaccggagg agcgcccgcu	20
<210> 101	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 101	
cgucugggcg gugcuacaac	20
<210> 102	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 102	
cuacaacugg gcuggcggcc	20
<210> 103	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 103	
aguccgggcu gggagcgggu	20
<210> 104	
<211> 20	
<212> RNA	
<213> Artificial Sequence	

<220><223> Synthetic
 <400> 104
 gcugcgggaa agggauuccc 20
 <210> 105
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 105
 acagcgggug uagacuccga 20

 <210> 106
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 106
 cagcgggugu agacuccgag 20
 <210> 107
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 107
 gucaagcccc agaggccaca 20
 <210> 108
 <211> 20
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 108
 gccuggggcc ccuaaccua 20

 <210> 109
 <211> 20

<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 109	
auuuucugac acucccgccc	20
<210> 110	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 110	
auccuggccg ccagcccagu	20
<210> 111	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 111	
ggagagcuuc gugcuaaacu	20
<210> 112	
<211> 23	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 112	
ugcaguccgg gcugggagcg ggu	23
<210> 113	
<211> 135	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 113	
ugcaguccgg gcugggagcg gguguuugag aguuauguaa gaaauuacau gacgaguuca	60
aaauaaaauu uauucaaacc gccuauuuau aggccgcaga uguucugcau uaugcuugcu	120

auugcaagcu uuuuu	135
<210> 114	
<211> 43	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 114	
ctgttgctgc agtccgggct gggagcgggt ggggagcaga ggg	43
<210> 115	
<211> 43	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 115	
gttaagagac agtccaggct gggagcagggt ggggagagga ggg	43
<210> 116	
<211> 22	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 116	
gcaguccggg cugggagcgg gu	22
<210> 117	
<211> 134	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 117	
gcaguccggg cugggagcgg guguuugaga guuauguaag aaauuacaug acgagucaa	60
auaaaaauuu auucaaaaccg ccuaauuuaua ggccgcagau guucugcauu augcuugcua	120
uugcaagcuu uuuu	134
<210> 118	
<211> 21	

<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 118	
caguccgggc ugaggagcggg u	21
<210> 119	
<211> 133	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 119	
caguccgggc ugaggagcggg uguuugagag uuauguaaga aauuacauga cgagucaaaa	60
uaaaaauuua uucaaaccgc cuuuuuauag gccgcagaug uucugcauua ugcugcuau	120
ugcaagcuuu uuu	133
<210> 120	
<211> 20	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 120	
aguccgggcu gggagcgggu	20
<210> 121	
<211> 132	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 121	
aguccgggcu gggagcgggu guuugagagu uauguaagaa auuacaugac gaguucaaa	60
aaaaauuuau ucaaaccgcc uuuuuauagg ccgcagaugu ucugcauuau gcuugcuauu	120
gcaagcuuuu uu	132
<210> 122	
<211> 19	
<212> RNA	

<213> Artificial Sequence	
<220><223> Synthetic	
<400> 122	
guccgggcug ggagcgggu	19
<210> 123	
<211> 131	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 123	
guccgggcug ggagcgggug uuugagaguu auguaagaaa uuacaugacg aguucaaaaua	60
aaaauuuuu caaaccgccu auuuauaggc cgcagauguu cugcauuuug cuugcuauug	120
caagcuuuuu u	131
<210> 124	
<211> 18	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 124	
uccgggcugg gagcgggu	18
<210> 125	
<211> 130	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 125	
uccgggcugg gagcgggugu uugagaguua uguaagaaau uacaugacga guucaaaaua	60
aaaauuuuu aaaccgccua uuuaauaggc gcagauguuc ugcauuuugc uugcuauugc	120
aagcuuuuuu	130
<210> 126	
<211> 17	
<212> RNA	
<213> Artificial Sequence	

<220><223> Synthetic	
<400> 126	
ccgggcuggg agcgggu	17
<210> 127	
<211> 129	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 127	
ccgggcuggg agcggguguu ugagaguau guaagaaau acaugacgag uucaaaauaa	60
aaauuuauca aaccgccuau uuauaggccg cagauguucu gcauuauugcu ugcuaauugca	120
agcuuuuuu	129
<210> 128	
<211> 128	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 128	
aguccgggcu gggagcgggu guuugagagu uauguaaaa caugacgagu ucaaaauaaa	60
auuuuuucaa accgccuauu uauaggccgc agauguucug cauuaugcuu gcuaauugcaa	120
gcuuuuuu	128
<210> 129	
<211> 130	
<212> RNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 129	
aguccgggcu gggagcgggu guuugagagu uauguagaaa uacaugacga guucaaauaa	60
aaauuuauuc aaaccgccua uuauaggcc gcagauguuc ugcauuauugc uugcuauugc	120
aagcuuuuuu	130
<210> 130	
<211> 28	

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 130
 agtccgggct gggagcgggt ggggagca 28
 <210> 131
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 131
 gtcaagcccc agaggccaca gggacaga 28
 <210> 132
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 132
 agtcctggct gggagcaggt ggggagag 28

 <210> 133
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 133
 gccaagcctc agaggccaca gggcagca 28
 <210> 134
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 134
 ccaagtgata aacacgagga tggcaaga 28
 <210> 135

<211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 135
 aactggtacc gcatgagccc cagcaacc 28

<210> 136
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 136
 catcaggctc tcagctcagc ctgagtgt 28

<210> 137
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 137
 aggtgccgtt tgttcatttt ctgacact 28

<210> 138
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 138
 ccagttgtag caccgcccag acgactgg 28

<210> 139
 <211> 28
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 139

tctccccagc cctgctcgtg gtgaccga	28
<210> 140	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 140	
tctgtgaatg ttagacccat gggagcag	28
<210> 141	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 141	
gcaagagacc cacacaccgg aggagcgc	28
<210> 142	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 142	
acaccggagg agcgcccgt tgggggag	28
<210> 143	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 143	
cgtctgggcg gtgctacaac tgggctgg	28
<210> 144	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	

<400> 144	
gctgcgggaa agggattccc tgggactc	28
<210> 145	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 145	
gcctggggcc cctaacccta tgtagcct	28
<210> 146	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 146	
attttctgac actcccgccc aatatacc	28
<210> 147	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 147	
atcctggcgc ccagcccagt tgtagcac	28
<210> 148	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> Synthetic	
<400> 148	
ggagagcttc gtgctaaact ggtaccgc	28
<210> 149	
<211> 11	
<212> RNA	

<213> Artificial Sequence	
<220><223> tracrRNA	
<400> 149	
cgagucaaaa u	11
<210> 150	
<211> 17	
<212> RNA	
<213> Artificial Sequence	
<220><223> tracrRNA Portion 1	
<400> 150	
aaaaauuuau ucaaacc	17
<210> 151	
<211> 14	
<212> RNA	
<213> Artificial Sequence	
<220><223> tracrRNA Portion 3	
<400> 151	
cgcagauguu cugc	14
<210> 152	
<211> 27	
<212> RNA	
<213> Artificial Sequence	
<220><223> tracrRNA Portion 4	
<400> 152	
auuaugcuug cuauugcaag cuuuuuu	27
<210> 153	
<211> 88	
<212> RNA	
<213> Artificial Sequence	
<220><223> Full tracrRNA V1	
<400> 153	
caugacgagu ucaaaauaaaa auuuauucaa accgccuauu uauaggccgc agauguucug	60
cauuauugcuu gcuaauugcaa gcuuuuuu	88

<210> 154
 <211> 91
 <212> RNA
 <213> Artificial Sequence
 <220><223> Full tracrRNA V2
 <400> 154
 uuacaugacg aguucaaaaua aaaaauuuuuu caaacgccu auuuauaggc cgcagauguu 60
 cugcauuuug cuugcuauug caagcuuuuu u 91
 <210> 155
 <211> 4194
 <212> DNA
 <213> Artificial Sequence
 <220><223> pPML3.1
 <220><221> misc_feature
 <222> (2256)..(2261)
 <223> n is a, c, g, or t
 <
 400> 155
 gggctctctct ggtagacca gatctgagcc tgggagctct ctggctaact agggaaaccca 60
 ctgcttaagc ctcaataaag cttgccttga gtgcttcaag tagtgtgtgc ccgtctgttg 120
 tgtgactctg gtaactagag atccctcaga cctttttagt cagtgtggaa aatctctagc 180
 agtggcgccc gaacagggac ttgaaagcga aagggaacc agaggagctc tctcgacgca 240
 ggactcggct tgctgaagcg cgcacggcaa gaggcgaggg gcgcgactg gtgagtacgc 300
 caaaaatfff gactagcgga ggctagaagg agagagatgg gtgcgagagc gtcagtatta 360
 agcgggggag aattagatcg cgatgggaaa aaattcgggt aaggccaggg ggaaagaaaa 420
 aatataaatt aaaacatata gtatgggcaa gcaggagct agaacgattc gcagttaatc 480
 ctggcctgtt agaaacatca gaaggctgta gacaaatact gggacagcta caaccatccc 540
 ttcagacagg atcagaagaa cttagatcat tatataatac agtagcaacc ctctattgtg 600
 tgcacaaag gatagagata aaagacacca aggaagcttt agacaagata gaggaagagc 660
 aaaacaaaag taagaccacc gcacagcaag cggccgctga tcttcagacc tggaggagga 720
 gatatgaggg acaattggag aagtgaatta tataaatata aagtagtaaa aattgaacca 780
 ttaggagtag caccaccaa ggcaaagaga agagtgtgtc agagagaaaa aagagcagtg 840
 ggaataggag ctttgttctt tgggttcttg ggagcagcag gaagcactat gggcgagcgc 900

tcaatgacgc tgacggtaca ggccagacaa ttattgtctg gtatagtga gcagcagaac	960
aatttgctga gggctattga ggcgaacag catctgttg aactcacagt ctggggcatc	1020
aagcagctcc aggcaagaat cctggctgtg gaaagatacc taaaggatca acagctcctg	1080
gggatttggg gtigctctgg aaaactcatt tgcaccactg ctgtgccttg gaatgctagt	1140
tggagtaata aatctctgga acagatttgg aatcacacga cctggatgga gtgggacaga	1200
gaaattaaca attacacaag cttaatacac tccttaattg aagaatcgca aaaccagcaa	1260
gaaaagaatg aacaagaatt attggaatta gataaatggg caagtttgtg gaattggttt	1320
aacataacaa attggctgtg gtatataaaa ttattcataa tgatagtagg aggcttggtg	1380
ggtttaagaa tagtttttgc tgtactttct atagtgaata gagttaggca gggatattca	1440
ccattatcgt ttcagaccca cctcccaacc ccgaggggac ccgacaggcc cgaagggaata	1500
gaagaagaag gtggagagag agacagagac agatccattc gattagtga cggatctcga	1560
cggtatcgat aagcttggga gttccgcgtt acataactta cggtaaatgg cccgcctggc	1620
tgaccgcca acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg	1680
ccaatagga ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg	1740
gcagtacatc aagtgtatca tatgccaagt acgcccccta ttgacgtcaa tgacggtaaa	1800
tggccgcct ggcatatgc ccagtacatg accttatggg actttctac ttggcagtac	1860
atctacgtat tagtcatcgc tattaccatg gtgatgcggt tttggcagta catcaatggg	1920
cgtggatagc ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg	1980
agtttgttt ggacacaaaa tcaacgggac ttccaaaat gtcgtaacaa ctccgcccc	2040
ttgacgcaaa tggcggttag gcgtgtacgg tgggaggtct atataagcag agctcgttta	2100
gigaaccgtc agatcgctg gagacgcat ccacgtgtt ttgacctca tagaagacac	2160
cgactctaga ggatccacta gtccagtgtg gtggaattct gcagatatca aagcttgcca	2220
ccatgcatac aatggaagag cagagccttg gtctcnnnn ngcgggtctg gtggcgctag	2280
cgtgtccaag ggcgaggagc tgttcaccgg cgtggtgccc atcctggtgg agctggacgg	2340
cgacgtgaac ggccacaagt tcagcgtgag cggcgagggc gaaggggacg ctacttacgg	2400
caaactgact ctcaagtta tctgtactac cgggaagctc cctgtccctt ggcctacact	2460
ggtcacaact ctcacatatg ggttcagtgt cttcagcaga taccgccacc acatgaagca	2520
gcacgacttc ttcaagagcg ccatgcccga gggctacgtg caggagagaa ccatcttctt	2580
caaggacgac ggcaactaca agaccagagc tgaggtcaag tttgagggtg acaccctggt	2640
gaacagaatc gagctgaagg gcatcgactt caaggaggac ggcaacatcc tgggccacaa	2700
gctggagtac aactacaaca gccacaacgt gtacatcatg gctgataaac agaagaatgg	2760

gattaaggtg aacttcaaga tcagacacaa catcgaggac ggcagcgtgc agctggccga 2820
ccactaccag cagaacaccc ccatacggcga cggccccgtg ctgctgcccg acaaccacta 2880
cctgagcacc cagagcgctc tcagtaagga ccctaagtag aagagagacc acatggtgct 2940

gctggagtgc gtgaccgccg ccggcatcac cctgggcatg gacgagctgt acaagtgagg 3000
gcctaagtag ttggaatta attctgtgga atgtgtgtca gttagggtgt ggaaagtccc 3060
caggctcccc agcaggcaga agtatgcaaa gcatgcatct caattagtca gcaaccaggc 3120
gtggaaagtc ccaggctcc ccagcaggca gaagtatgca aagcatgcat ctcaattagt 3180
cagcaacat agtccccccc ctaactccgc ccataccgcc cctaactccg ccagttccg 3240
cccattctcc gcccattggc tgactaattt tttttattta tgcagaggcc gagggccgct 3300
ctgcctctga gctattccag aagtagtgag gaggtttttt tggaggccta ggcttttgca 3360

aaaagctccc gggagcttgc atattccattt tcggatctga tcagcacgtg ttgacaatta 3420
atcatcggca tagtatatcg gcatagtata atacgacaag gtgaggaact aaaccatggc 3480
caagcctttg tctcaagaag aatccaccct cattgaaaga gcaacggcta caatcaacag 3540
catcccatc tctgaagact acagcgctgc cagcgagct ctctctagcg acggccgcat 3600
cttactgtgt gtcaatglat atcattttac tgggggacct tgtgcagaac tcgtggtgct 3660
gggcactgct gctgctgcgg cagctggcaa cctgacttgt atcgtcgcga tcggaaatga 3720
gaacaggggc atcttgagcc cctgcggacg gtgccgacag gtgcttctcg atctgcatcc 3780

tgggatcaaa gccatagtag aggacagtga tggacagccg acggcagttg ggattcgtga 3840
attgctgccc tctggttatg tgtgggaggg ctaagcaca ttcgagctcg gtacctttaa 3900
gaccaatgac ttacaaggca gctgtagatc ttaggcactt tttaaaagaa aaggggggac 3960
tggaaaggct aattcactcc caacgaagac aagatctgct ttttgcctgt actgggtctc 4020
tctggttaga ccagatctga gcctgggagc tctctggcta actaggaac ccactgctta 4080
agcctcaata aagcttgctt tgagtgttc aagtagtgtg tgcccgtctg ttgtgtgact 4140
ctggtaacta gagatccctc agaccctttt agtcagtgtg gaaaatctct agca 4194

<210> 156

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> TRAC gRNA 1

<400> 156

tctctcagct ggtacacggc a

21

<210> 157
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 2
 <400> 157
 gcgtcatgag cagattaaac c 21
 <210> 158
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 3
 <400> 158
 tctcgaccag cttgacatca c 21

 <210> 159
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 4
 <400> 159
 ttaaaccg cg ccactttcag g 21
 <210> 160
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 5
 <400> 160
 ctgtgctaga catgaggtct a 21
 <210> 161
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 8
 <400> 161

acttcaagag caacagtgct g	21
<210> 162	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> TRAC gRNA 9	
<400> 162	
aagagcaaca gtgctgtggc c	21
<210> 163	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> TRAC gRNA 10	
<400> 163	
gctggggaag aaggtgtctt c	21
<210> 164	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> TRAC gRNA 15	
<400> 164	
ataggcagac agacttgta c	21
<210> 165	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> TRAC gRNA 17	
<400> 165	
tagagtctct cagctggta a	21
<210> 166	
<211> 21	
<212> DNA	
<213> Artificial Sequence	

<220><223> TRACR gRNA 18
 <400> 166
 gtctctcagc tggtagacagg c 21
 <210> 167
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 19
 <400> 167
 cagctggtac acggcagggt c 21

 <210> 168
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 20
 <400> 168
 agctggtaca cggcagggtc a 21
 <210> 169
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 21
 <400> 169
 tacacggcag ggtcagggtt c 21
 <210> 170
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 23
 <400> 170
 ctttcaaaac ctgtcagtga t 21

 <210> 171
 <211> 21

<212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 25
 <400> 171
 tccgaatcct cctcctgaaa g 21
 <210> 172
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 26
 <400> 172
 aatcctcctc ctgaaagtgg c 21
 <210> 173
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 27
 <400> 173
 atcctcctcc tgaaagtggc c 21

 <210> 174
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 29
 <400> 174
 ctgctcatga cgctgcggct g 21
 <210> 175
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 30
 <400> 175
 agattaaacc cggccacttt c 21
 <210> 176

<211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 31
 <400> 176
 aacccggcca ctttcaggag g 21

<210> 177
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> TRAC gRNA 32
 <400> 177
 gccactttca ggaggaggat t 21

<210> 178
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> B2M gRNA 1
 <400> 178
 tactctctct ttctggcctg g 21

<210> 179
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> B2M gRNA 2
 <400> 179
 gcatactcat ctttttcagt g 21

<210> 180
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> B2M gRNA 3
 <400> 180

cgctactctc tctttctggc c	21
<210> 181	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 4	
<400> 181	
gcgcgagcac agctaaggcc a	21
<210> 182	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 6	
<400> 182	
gctcgcgcta ctctctcttt c	21
<210> 183	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 7	
<400> 183	
agagtagcgc gagcacagct a	21
<210> 184	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 15	
<400> 184	
tcacagccca agatagttaa g	21
<210> 185	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 16	

<400> 185	
cacagcccaa gatagttaag t	21
<210> 186	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 18	
<400> 186	
gacaaagtca catggttcac a	21
<210> 187	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 19	
<400> 187	
aagtcacatg gttcacacgg c	21
<210> 188	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 20	
<400> 188	
aggcatactc atctttttca g	21
<210> 189	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> B2M gRNA 21	
<400> 189	
ggcatactca tctttttcag t	21
<210> 190	
<211> 21	
<212> DNA	

<213> Artificial Sequence

<220><223> B2M gRNA 22

<400> 190

cataactcatc tttttcagtg g 21

<210> 191

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> B2M gRNA 23

<400> 191

tcagtaagtc aacttcaatg t 21

<210> 192

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> B2M gRNA 26

<400> 192

acgtgagtaa acctgaatct t 21

<210> 193

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> ELANEg35_OMNI-50

<400> 193

agtccgggct gggagcgggt 20

<210> 194

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> ELANEg38_OMNI-50

<400> 194

acagcgggtg tagactccga 20

<210> 195

<211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> ELANEg39_OMNI-50
 <400> 195
 cagcgggtgt agactccgag 20
 <210> 196
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> ELANEg58_OMNI-50
 <400> 196
 gctgcgggaa agggattccc 20
 <210> 197
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220><223> ELANEg62_OMNI-50
 <400> 197
 gtcaagcccc agaggccaca 20