



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0065382  
(43) 공개일자 2019년06월11일

- |  |   |
|--|---|
| <p>(51) 국제특허분류(Int. Cl.)<br/> <i>C12N 9/22</i> (2006.01) <i>C12N 15/85</i> (2006.01)<br/> <i>C12N 5/0783</i> (2010.01) <i>C12N 5/0789</i> (2010.01)</p> <p>(52) CPC특허분류<br/> <i>C12N 9/22</i> (2013.01)<br/> <i>C12N 15/85</i> (2013.01)</p> <p>(21) 출원번호 10-2019-7013234<br/> (22) 출원일자(국제) 2017년10월11일<br/> 심사청구일자 없음<br/> (85) 번역문제출일자 2019년05월08일<br/> (86) 국제출원번호 PCT/US2017/056178<br/> (87) 국제공개번호 WO 2018/071565<br/> 국제공개일자 2018년04월19일<br/> (30) 우선권주장<br/> 62/406,689 2016년10월11일 미국(US)<br/> 62/414,266 2016년10월28일 미국(US)</p> | <p>(71) 출원인<br/> <b>블루버드 바이오, 인코포레이티드.</b><br/> 미국, 매사추세츠 02142, 캄브리지, 60 비니 스트리트</p> <p>(72) 발명자<br/> <b>자르주르 조던</b><br/> 미국, 워싱턴 98146, 시애틀, 11014 37 애비뉴 사우스웨스트<br/> <b>포그슨 마크</b><br/> 미국, 워싱턴 98045, 노스 벤드, 3122 사우스이스트 16 스트리트</p> <p>(74) 대리인<br/> <b>손민</b></p> |
|--|---|

전체 청구항 수 : 총 42 항

(54) 발명의 명칭 **TCR α 호밍 엔도뉴클레아제 변이체**

**(57) 요약**

본 개시내용은 TCR α 유전자를 편집하기 위한 개선된 게놈 편집 조성물 및 방법을 제공한다.

**대표도** - 도1

	$n_{11}$	$n_{10}$	$n_9$	$n_8$	$n_7$	$n_6$	$n_5$	$n_4$	$n_3$	$n_2$	$n_1$	$p_1$	$p_2$	$p_3$	$p_4$	$p_5$	$p_6$	$p_7$	$p_8$	$p_9$	$p_{10}$	$p_{11}$
TCRa	T	G	T	C	T	G	C	C	T	A	T	T	C	A	C	C	G	A	T	T	T	T
KAT2B	A	G	T	A	T	G	C	C	T	A	T	T	C	T	A	C	A	A	G	A	A	A

(52) CPC특허분류

**C12N 5/0636** (2013.01)

**C12N 5/0646** (2013.01)

**C12N 5/0647** (2013.01)

**C12Y 301/11002** (2013.01)

**C07K 2319/00** (2013.01)

**C07K 2319/50** (2013.01)

**C07K 2319/81** (2013.01)

**C12N 2510/00** (2013.01)

---

## 명세서

### 청구범위

#### 청구항 1

인간 T 세포 수용체 알파(TCR α) 유전자 내 표적 부위를 절단하는 I-OnuI 호밍 엔도뉴클레아제(homing endonuclease: HE) 변이체를 포함하는 폴리펩타이드로서, 상기 변이체는 서열번호 1 내지 5 중 임의의 하나, 또는 이의 생물학적으로 활성인 단편의 아미노산 치환: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E 및 V238E를 포함하는, 폴리펩타이드.

#### 청구항 2

인간 T 세포 수용체 알파(TCR α) 유전자 내 표적 부위를 절단하는 I-OnuI 호밍 엔도뉴클레아제(HE) 변이체를 포함하는 폴리펩타이드로서, 상기 변이체는 서열번호 1 내지 5 중 임의의 하나, 또는 이의 생물학적으로 활성인 단편의 아미노산 치환: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, E178D, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E 및 V238E를 포함하는, 폴리펩타이드.

#### 청구항 3

제1항 또는 제2항에 있어서, 상기 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 1, 2, 3, 4, 5, 6, 7 또는 8개의 N-말단 아미노산을 결여하는, 폴리펩타이드.

#### 청구항 4

제3항에 있어서, 상기 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 4개의 N-말단 아미노산을 결여하는, 폴리펩타이드.

#### 청구항 5

제3항에 있어서, 상기 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 8개의 N-말단 아미노산을 결여하는, 폴리펩타이드.

#### 청구항 6

제1항 또는 제2항에 있어서, 상기 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 1, 2, 3, 4 또는 5개의 C-말단 아미노산을 결여하는, 폴리펩타이드.

#### 청구항 7

제6항에 있어서, 상기 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 C-말단 아미노산을 결여하는, 폴리펩타이드.

#### 청구항 8

제6항에 있어서, 상기 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 2개의 C-말단 아미노산을 결여하는, 폴리펩타이드.

#### 청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 I-OnuI HE 변이체는 서열번호 7 내지 8 중 임의의 하나에 제시된 아미노산 서열에 대해 적어도 99% 동일한 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함하는, 폴리펩타이드.

#### 청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서, 상기 HE 변이체는 서열번호 7에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함하는, 폴리펩타이드.

#### 청구항 11

제1항 내지 제9항 중 어느 한 항에 있어서, 상기 HE 변이체는 서열번호 8에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함하는, 폴리펩타이드.

#### 청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서, 상기 폴리펩타이드는 서열번호 17에 제시된 폴리뉴클레오타이드 서열에 결합하는, 폴리펩타이드.

#### 청구항 13

제1항 내지 제12항 중 어느 한 항에 있어서, DNA 결합 도메인을 더 포함하는, 폴리펩타이드.

#### 청구항 14

제13항에 있어서, 상기 DNA 결합 도메인은 TALE DNA 결합 도메인 및 아연 핑거 DNA 결합 도메인으로 이루어진 군으로부터 선택되는, 폴리펩타이드.

#### 청구항 15

제14항에 있어서, 상기 TALE DNA 결합 도메인은 약 9.5개의 TALE 반복 단위, 약 10.5개의 TALE 반복 단위, 약 11.5개의 TALE 반복 단위, 약 12.5개의 TALE 반복 단위, 약 13.5개의 TALE 반복 단위, 약 14.5개의 TALE 반복 단위 또는 약 15.5개의 TALE 반복 단위를 포함하는, 폴리펩타이드.

#### 청구항 16

제14항 또는 제15항에 있어서, 상기 TALE DNA 결합 도메인은 11.5개의 TALE 반복 단위를 포함하고, 서열번호 19에 제시된 폴리뉴클레오타이드 서열에 결합하는, 폴리펩타이드.

#### 청구항 17

제14항 또는 제15항에 있어서, 상기 TALE DNA 결합 도메인은 10.5개의 TALE 반복 단위를 포함하고, 서열번호 18에 제시된 폴리뉴클레오타이드 서열에 결합하는, 폴리펩타이드.

#### 청구항 18

제14항 내지 제17항 중 어느 한 항에 있어서, 상기 폴리펩타이드는 서열번호 20에 제시된 폴리뉴클레오타이드 서열에 결합하고, 이를 절단하는, 폴리펩타이드.

#### 청구항 19

제14항에 있어서, 상기 아연 핑거 DNA 결합 도메인은 2, 3, 4, 5, 6, 7 또는 8개의 아연 핑거 모티프를 포함하는, 폴리펩타이드.

#### 청구항 20

제1항 내지 제19항 중 어느 한 항에 있어서, 펩타이드 링커 및 말단-가공 효소(end-processing enzyme) 또는 이의 생물학적으로 활성인 단편을 더 포함하는, 폴리펩타이드.

#### 청구항 21

제1항 내지 제20항 중 어느 한 항에 있어서, 바이러스 자기-절단성 2A 펩타이드 및 말단-가공 효소 또는 이의 생물학적으로 활성인 단편을 더 포함하는, 폴리펩타이드.

#### 청구항 22

제20항 또는 제21항에 있어서, 상기 말단-가공 효소 또는 이의 생물학적으로 활성인 단편은 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제, 5' 플랩 엔도뉴클레아제, 헬리카제 또는 주형-독립적 DNA 중합효소 활성을 갖는, 폴리펩타이드.

#### 청구항 23

제20항 내지 제22항 중 어느 한 항에 있어서, 상기 말단-가공 효소는 Trex2 또는 이의 생물학적으로 활성인 단편을 포함하는, 폴리펩타이드.

#### 청구항 24

제1항 내지 제23항 중 어느 한 항에 있어서, 상기 폴리펩타이드는 서열번호 10 내지 12 중 임의의 하나에 제시된 아미노산 서열 또는 이의 생물학적으로 활성인 단편을 포함하는, 폴리펩타이드.

#### 청구항 25

제24항에 있어서, 상기 폴리펩타이드는 서열번호 10에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함하는, 폴리펩타이드.

#### 청구항 26

제24항에 있어서, 상기 폴리펩타이드는 서열번호 11에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함하는, 폴리펩타이드.

#### 청구항 27

제24항에 있어서, 상기 폴리펩타이드는 서열번호 12에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함하는, 폴리펩타이드.

#### 청구항 28

제1항 내지 제27항 중 어느 한 항의 폴리펩타이드를 암호화하는 폴리뉴클레오타이드.

#### 청구항 29

제1항 내지 제27항 중 어느 한 항의 폴리펩타이드를 암호화하는 mRNA.

#### 청구항 30

제1항 내지 제27항 중 어느 한 항의 폴리펩타이드를 암호화하는 cDNA.

#### 청구항 31

제1항 내지 제27항 중 어느 한 항의 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 포함하는 벡터.

#### 청구항 32

제1항 내지 제27항 중 어느 한 항의 폴리펩타이드를 포함하는 세포.

#### 청구항 33

제1항 내지 제27항 중 어느 한 항의 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 포함하는 세포.

#### 청구항 34

제27항의 벡터를 포함하는 세포.

#### 청구항 35

제1항 내지 제27항 중 어느 한 항의 폴리펩타이드에 의해 편집된 세포.

#### 청구항 36

제32항 내지 제35항 중 어느 한 항에 있어서, 상기 세포는 조혈 세포인, 세포.

#### 청구항 37

제32항 내지 제36항 중 어느 한 항에 있어서, 상기 세포는 T 세포인, 세포.

#### 청구항 38

제32항 내지 제37항 중 어느 한 항에 있어서, 상기 세포는  $CD3^+$ ,  $CD4^+$  및/또는  $CD8^+$  세포인, 세포.

#### 청구항 39

제32항 내지 제38항 중 어느 한 항에 있어서, 상기 세포는 면역 효과기 세포인, 세포.

#### 청구항 40

제32항 내지 제39항 중 어느 한 항에 있어서, 상기 세포는 세포독성 T 림프구(CTL), 종양 침윤성 림프구(TIL) 또는 헬퍼 T 세포인, 세포.

#### 청구항 41

제32항 내지 제39항 중 어느 한 항에 있어서, 상기 세포는 자연 살해(NK) 세포 또는 자연 살해 T(NKT) 세포인, 세포.

#### 청구항 42

제32항 내지 제41항 중 어느 한 항에 있어서, 상기 세포의 공급원은 말초혈액 단핵세포, 골수, 림프절 조직, 제대혈, 흉선 조직, 감염 부위로부터의 조직, 복수, 흉수, 비장 조직 또는 종양인, 세포.

### 발명의 설명

### 기술 분야

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 미국 특허법(35 U.S.C. § 119(e)) 하에서 2016년 10월 28일자로 출원된 미국 가출원 특허 제 62/414,266호, 및 2016년 10월 11일자로 출원된 미국 가출원 특허 제 62/406,689호의 우선권을 주장하며, 이들 각각은 그의 전문이 본 명세서에 참고로 포함된다.

[0003] 서열목록에 관한 언급

[0004] 본 출원과 관련된 서열목록은 종이 사본 대신에 텍스트 형식으로 제공되며, 본원에서 본 명세서 내에 참고로 포함된다. 서열목록을 함유하는 텍스트 파일명은 BLBD\_077\_02W0\_ST25.txt이다. 텍스트 파일은 114KB이며, 2017년 10월 11일자로 생성되었고, 본 명세서의 출원과 동시에 EFS-Web을 통해 전자적으로 제출 중에 있다.

[0005] 기술분야

[0006] 본 개시내용은 개선된 계층 편집 폴리펩타이드 및 조성물에 관한 것이다. 더 구체적으로, 본 개시내용은 인간 T 세포 수용체 알파(T cell receptor alpha: TCR α) 유전자를 편집하는 데 유용한 뉴클레아제 변이체 및 조성물에 관한 것이다.

### 배경 기술

[0007] 암의 전세계적 부담은 1975 내지 2000에 2배가 되었다. 암은 전세계적으로 이환율 및 사망률의 두 번째 주된 원인이며, 2012년에 대략 1천 4백 십만건의 새로운 사례 및 8백 2십만건의 암관련 사망이 있었다. 가장 통상적인 암은 유방암, 폐 및 기관지암, 전립선암, 결장 및 직장암, 방광암, 피부의 흑색종, 비호지킨 림프종, 갑상선암, 신장 및 신우암, 자궁내막암, 백혈병 및 췌장암이다. 새로운 암 사례의 수는 다음 20년 내에 2천 2백만건까지 상승할 것으로 추정된다.

[0008] 면역계는 인간암을 검출하고 이와 싸우는데 중요한 역할을 한다. 다수의 형질전환된 세포는 면역 보조에 의해 빠르게 검출되며, 클론 발현된 T 세포 수용체(TCR)를 통한 항원-특이적 T 세포의 활성화를 통해 파괴된다. 따라서, 암은 질환을 지속적으로 억제하고 제거하기 위한 필요한 항-종양 반응을 시작하기 위한 면역계의 부전인 면역학적 장애로 고려될 수 있다. 암과 더 효과적으로 싸우기 위해, 최근 몇 십년에 걸쳐 진행된 특정 면역요법 개입은 T 세포 면역을 향상시키는 데 특별하게 중점을 두었다. 이들 치료는 질환 관해의 산발적 사례만을 수득하였고, 실질적인 전반적 성공을 갖지 않았다.

[0009] 더 최근에, T 세포의 단리, 변형, 확장 및 재주입에 기반한 적응 세포 면역요법 전략은 초기 임상 시험에서 연구되고 시험되었다. T 세포는 종종 그들의 선택적 인식 및 강력한 효과기 메커니즘에 기인하여 암에 대한 선택의 효과기 세포가 되었다. 이들 치료는 혼합된 비율의 성장을 나타내었지만, 소수의 환자는 T 세포 기반 면역요법에 대한 아직까지 실현되지 않은 잠재력을 강조하는 지속 가능한 관해를 경험하였다.

[0010] 따라서, 다양한 형태의 암 및 다른 면역 장애를 치료하기 위한 더 효과적이고, 표적화되며, 더 안전하고 지속적인 요법에 대한 필요가 있다. 추가로, 높은 효율로 내인성 TCR 유전자를 정확하게 그리고 반복 가능하게 파괴할 수 있는 방법 및 조성물에 대한 필요가 있다. 대부분의 암에 대한 오늘날의 표준 치료는 이들 기준 중 일부 또는 모두에 미치지 못한다.

### 발명의 내용

[0011] 본 개시내용을 일반적으로, 부분적으로, 인간 TCR  $\alpha$  유전자의 엑손 1 내 표적 부위를 절단하는 호밍 엔도뉴클레아제(homing endonuclease: HE) 변이체 및 메가TAL, HE 변이체 및 메가TAL을 포함하는 조성물, 및 이를 이용하는 방법에 관한 것이다.

[0012] 다양한 실시형태에서, 본 개시내용은 부분적으로 인간 T 세포 수용체 알파(TCR  $\alpha$ ) 유전자에서 표적 부위를 절단하는 I-OnuI 호밍 엔도뉴클레아제(HE) 변이체를 포함하는 폴리펩타이드를 상정하되, 변이체는 서열번호 1 내지 5 중 임의의 하나 또는 이의 생물학적으로 활성인 단편의 아미노산 치환: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E, 및 V238E를 포함한다.

[0013] 다양한 실시형태에서, 본 개시내용은, 부분적으로, 인간 T 세포 수용체 알파(TCR  $\alpha$ ) 유전자에서 표적 부위를 절단하는 I-OnuI 호밍 엔도뉴클레아제(HE) 변이체를 포함하는 폴리펩타이드를 상정하되, 변이체는 서열번호 1 내지 5 중 임의의 하나 또는 이의 생물학적으로 활성인 단편의 아미노산 치환: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, E178D, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E, 및 V238E를 포함한다.

[0014] 특정 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 1, 2, 3, 4, 5, 6, 7 또는 8개의 N-말단 아미노산을 결여한다.

[0015] 특정 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 4개의 N-말단 아미노산을 결여한다.

[0016] 추가 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 8개의 N-말단 아미노산을 결여한다.

[0017] 일부 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 1, 2, 3, 4 또는 5개의 C-말단 아미노산을 결여한다.

[0018] 추가적인 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 C-말단 아미노산을 결여한다.

[0019] 특정 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 I-OnuI HE에 비해 2개의 C-말단 아미노산을 결여한다.

[0020] 특정 실시형태에서, I-OnuI HE 변이체는 서열번호 7 내지 8 중 임의의 하나에 제시된 아미노산 서열 또는 이의 생물학적으로 활성인 단편에 대해 적어도 99% 동일한 아미노산 서열을 포함한다.

- [0021] 추가적인 실시형태에서, HE 변이체는 서열번호 7에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0022] 특정 실시형태에서, HE 변이체는 서열번호 8에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0023] 일부 실시형태에서, 폴리펩타이드는 서열번호 17에 제시된 폴리뉴클레오타이드 서열에 결합한다.
- [0024] 특정 실시형태에서, 폴리펩타이드는 DNA 결합 도메인을 추가로 포함한다.
- [0025] 추가 실시형태에서, DNA 결합 도메인은 TALE DNA 결합 도메인 및 아연 핑거 DNA 결합 도메인으로 이루어진 군으로부터 선택된다.
- [0026] 추가적인 실시형태에서, TALE DNA 결합 도메인은 약 9.5개의 TALE 반복 단위, 약 10.5개의 TALE 반복 단위, 약 11.5개의 TALE 반복 단위, 약 12.5개의 TALE 반복 단위, 약 13.5개의 TALE 반복 단위, 약 14.5개의 TALE 반복 단위 또는 약 15.5개의 TALE 반복 단위를 포함한다.
- [0027] 일부 실시형태에서, TALE DNA 결합 도메인은 11.5개의 TALE 반복 단위를 포함하고, 서열번호 19에 제시된 폴리뉴클레오타이드 서열에 결합한다.
- [0028] 특정 실시형태에서, TALE DNA 결합 도메인은 10.5개의 TALE 반복 단위를 포함하고, 서열번호 18에 제시된 폴리뉴클레오타이드 서열에 결합한다.
- [0029] 특정 실시형태에서, TALE DNA 결합 도메인은 서열번호 20에 제시된 폴리뉴클레오타이드 서열에 결합한다.
- [0030] 특정 실시형태에서, 아연 핑거 DNA 결합 도메인은 2, 3, 4, 5, 6, 7 또는 8개의 아연 핑거 모티프를 포함한다.
- [0031] 추가 실시형태에서, 폴리펩타이드는 추가로 펩타이드 링커 및 말단-가공 효소 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0032] 특정 실시형태에서, 폴리펩타이드는 추가로 바이러스 자기-절단성 2A 펩타이드 펩타이드 및 말단-가공 효소 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0033] 일부 실시형태에서, 말단-가공 효소 또는 이의 생물학적으로 활성인 단편은 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제, 5' 플랩 엔도뉴클레아제, 헬리카제 또는 주형-독립적 DNA 중합효소 활성을 가진다.
- [0034] 특정 실시형태에서, 말단-가공 효소는 Trex2 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0035] 특정 실시형태에서, 폴리펩타이드는 서열번호 10 내지 12 중 임의의 하나에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0036] 특정 실시형태에서, 폴리펩타이드는 서열번호 10에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0037] 추가적인 실시형태에서, 폴리펩타이드는 서열번호 11에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0038] 특정 실시형태에서, 폴리펩타이드는 서열번호 12에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함한다.
- [0039] 다양한 실시형태에서, 본 명세서에 상정된 폴리펩타이드를 암호화하는 폴리뉴클레오타이드가 제공된다.
- [0040] 특정 실시형태에서, 본 개시내용은, 부분적으로, 본 명세서에 상정된 mRNA가 제공된다.
- [0041] 특정 실시형태에서, mRNA는 서열번호 16 내지 19 중 임의의 하나에 제시된 서열을 포함한다.
- [0042] 일부 실시형태에서, 본 명세서에 상정된 폴리펩타이드를 암호화하는 cDNA가 제공된다.
- [0043] 특정 실시형태에서, 본 명세서에 상정된 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 포함하는 벡터가 제공된다.
- [0044] 특정 실시형태에서, 본 명세서에 상정된 폴리펩타이드를 포함하는 세포가 제공된다.
- [0045] 추가 실시형태에서, 본 명세서에 상정된 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 포함하는 세포가 제



공된다.

- [0046] 추가적인 실시형태에서, 본 명세서에 상정된 백터를 포함하는 세포가 제공된다.
- [0047] 특정 실시형태에서, 본 명세서에 상정된 폴리펩타이드에 의해 세포가 제공된다.
- [0048] 특정 실시형태에서, 세포는 조혈세포이다.
- [0049] 일부 실시형태에서, 세포는 T 세포이다.
- [0050] 특정 실시형태에서, 세포는 CD3+, CD4+ 및/또는 CD8+ 세포이다.
- [0051] 특정 실시형태에서, 세포는 면역 효과기 세포이다.
- [0052] 추가 실시형태에서, 세포는 세포독성 T 림프구(cytotoxic T lymphocyte: CTL), 종양 침윤성 림프구(tumor infiltrating lymphocyte: TIL), 또는 헬퍼 T 세포이다.
- [0053] 특정 실시형태에서, 세포는 자연 살해(natural killer: NK) 세포 또는 자연 살해 T(natural killer T: NKT) 세포이다.
- [0054] 특정 실시형태에서, 세포의 공급원은 말초혈액 단핵세포, 골수, 림프절 조직, 체대혈, 흉선 조직, 감염 부위로 부터의 조직, 복수, 흉수, 비장 조직, 또는 종양이다.

### 도면의 간단한 설명

- [0055] **도 1**은 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 대해 설계된 I-OnuI HE 변이체에 대한 표적 상의 그리고 표적이탈 부위(서열번호 17 및 21)의 비교를 도시한 도면.
- 도 2**는 "CCG" 트라이뉴클레오타이드 서열(서열번호 17 및 67)에 고도로 선택적인 모듈 HE 변이체로부터 얻어진 아미노산 세트를 1세대 TCR  $\alpha$  HE 변이체(서열번호 6)에 접합시켜 2세대 TCR  $\alpha$  HE 변이체(TCR  $\alpha$  2.0)를 생성하는 전략을 도시한 도면. TCR  $\alpha$  2.0 HE 변이체는 목적으로 하는 기질 위치에서 "CCG"에 대해 높은 선택성을 보유한다.
- 도 3**은 기질 위치 p456에서 1세대 TCR  $\alpha$  HE 변이체 및 TCR  $\alpha$  2.0 HE 변이체에 대해 향상된 선택성 프로파일을 도시한 도면. 패널 (A)는 64개의 가능한 뉴클레오타이드 조합에 대한 효모 표면 디스플레이 절단 플롯을 나타내고, 패널 (B)는 이들 활성의 히트맵을 나타낸다.
- 도 4**는 관독으로서 CD3 표면 발현을 이용하는 TCR  $\alpha$  좌위에서의 유전자 편집 효율의 대표적인 데이터를 도시한 도면. 1세대 TCR  $\alpha$  메가TAL (A)(서열번호 6)은 TCR  $\alpha$  2.0 메가TAL(서열번호 7)로 처리한 T 세포에 비해 인간 1차 T 세포에서 증가된 유전자 편집 효율을 나타낸다.
- 도 5**는 TCR  $\alpha$  2.1 메가TAL (C)(서열번호 8)를 생성하기 위한 TCR  $\alpha$  2.0 메가TAL (B)에서의 표적화된 돌연변이(E178D)의 도입은 1세대 TCR  $\alpha$  메가TAL (A)의 표적 상의 절단 활성을 회복시키는 한편, TCR  $\alpha$  2.0 메가TAL의 표적 상의 선택성을 보유한다는 것을 도시한 도면.
- 도 6**은 생물정보학적으로 유래된 추정적 표적이탈 부위를 이용하는 효모-표면-디스플레이 프로파일링에 의한 TCR  $\alpha$  2.1 메가TAL의 증가된 선택성을 도시한 도면.
- 도 7**은 메가TAL 처리된 1차 T 세포의 게놈 DNA로부터 생성된 앰플리콘의 심층 서열분석에 의한 TCR  $\alpha$  2.1 메가TAL의 증가된 선택성을 도시한 도면.
- 도 8**은 TALE 반복부를 10.5 내지 11.5개로 증가시키기 위한 TCR  $\alpha$  2.1 메가TAL에 대한 추가적인 TALE RVD의 현수는(A) 1차 T 세포에서 TCR  $\alpha$  2.1 메가TAL의 표적 상의 활성(서열번호 68 및 69)을 추가로 증가시킨다(B)는 것을 도시한 도면.
- 도 9**는 상이한 잔토모나스와 구성원으로부터 유래된 첨부된 합리적으로 설계된 최소 N-말단 TAL 도메인은 잔토모나스 오리자에(X. oryzae) N-말단 TAL 도메인을 함유하는 1세대 TCR  $\alpha$  메가TAL에 비교하여 활성의 유사한 또는, 일부 예에서, 보통의 증가를 수득한다는 것을 도시한 도면.

### 발명을 실시하기 위한 구체적인 내용

- [0056] 서열 식별자의 간단한 설명

- [0057] 서열번호 1은 야생형 I-OnuI LAGLIDADG 호밍 엔도뉴클레아제(LHE)의 아미노산 서열이다.
- [0058] 서열번호 2는 야생형 I-OnuI LHE의 아미노산 서열이다.
- [0059] 서열번호 3은 야생형 I-OnuI LHE의 생물학적으로 활성인 단편의 아미노산 서열이다.
- [0060] 서열번호 4는 야생형 I-OnuI LHE의 생물학적으로 활성인 단편의 아미노산 서열이다.
- [0061] 서열번호 5는 야생형 I-OnuI LHE의 생물학적으로 활성인 단편의 아미노산 서열이다.
- [0062] 서열번호 6은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하도록 재프로그래밍된 I-OnuI LHE 변이체의 아미노산 서열이다.
- [0063] 서열번호 7은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하도록 재프로그래밍된 I-OnuI LHE 변이체의 아미노산 서열이다.
- [0064] 서열번호 8은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하도록 재프로그래밍된 I-OnuI LHE 변이체의 아미노산 서열이다.
- [0065] 서열번호 9는 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하는 메가TAL의 아미노산 서열이다.
- [0066] 서열번호 10은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하는 메가TAL의 아미노산 서열이다.
- [0067] 서열번호 11은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하는 메가TAL의 아미노산 서열이다.
- [0068] 서열번호 12는 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하는 메가TAL의 아미노산 서열이다.
- [0069] 서열번호 13은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하는 메가TAL의 mRNA 서열이다.
- [0070] 서열번호 14는 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하는 메가TAL의 mRNA 서열이다.
- [0071] 서열번호 15는 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하는 메가TAL의 mRNA 서열이다.
- [0072] 서열번호 16은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 표적 부위에 결합하고, 이를 절단하는 메가TAL의 mRNA 서열이다.
- [0073] 서열번호 17은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 I-OnuI LHE 변이체 표적 부위이다.
- [0074] 서열번호 18은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 10.5 RVD TALE DNA 결합 도메인 표적 부위이다.
- [0075] 서열번호 19는 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 내 11.5 RVD TALE DNA 결합 도메인 표적 부위이다.
- [0076] 서열번호 20은 인간 TCR  $\alpha$  유전자 불변 영역의 엑손 1 메가TAL 표적 부위이다.
- [0077] 서열번호 21은 서열번호 17을 표적화하는 TCR  $\alpha$ 에 대한 인간 KAT2B 유전자 내 표적이탈 부위이다.
- [0078] 서열번호 22 내지 24는 NTD 변이체 메가TAL을 암호화하는 아미노산 서열이다.
- [0079] 서열번호 25 내지 27은 NTD 변이체 메가TAL을 암호화하는 mRNA 서열이다.
- [0080] 서열번호 28은 뮤린 Trex2를 암호화하는 아미노산 서열이다.
- [0081] 서열번호 29는 뮤린 Trex2를 암호화하는 mRNA 서열이다.
- [0082] 서열번호 30 내지 40은 다양한 링커의 아미노산 서열을 제시한다.
- [0083] 서열번호 41 내지 65는 프로테아제 절단 부위 및 자기-절단 폴리펩타이드 절단 부위의 아미노산 서열을 제시한다.

[0084] 앞서 언급한 서열에서, X는 존재한다면, 임의의 아미노산 또는 아미노산의 부재를 지칭한다.

## [0085] A.개요

[0086] 본 개시내용은 일반적으로, 부분적으로, 개선된 게놈 편집 조성물 및 이의 사용 방법에 관한 것이다. 게놈 편집된 면역 효과기 세포는 바람직하지 않은 자가면역 반응의 감소된 위험에 기인하는 개선된 안전성, 더 예측 가능한 치료 유전자 발현을 이용하는 정확하게 표적화된 요법, 종양 미세환경 및 증가된 효능의 증가된 지속능력을 포함하지만, 이들로 제한되지 않는 기존의 세포기반 면역요법에 비해 수많은 이점을 제공한다. 임의의 특정 이론으로 구속되는 일 없이, 다양한 실시형태에 상정된 게놈 편집 조성물은 호밍 엔도뉴클레아제 및 T 세포 수용체 알파(TCR α) 유전자를 표적화하는 메가TAL을 포함하고, 개선된 안전성 및 증가된 표적 부위 특이성, 선택성 및 촉매적 활성을 위해 조작된다.

[0087] 특정 실시형태에서, 하나 이상의 TCR α 대립유전자의 변형은 TCR α 대립유전자(들)의 발현이 없거나 또는 실질적으로 없고, TCR α 대립유전자(들)의 발현을 감소시키거나/시키거나 TCR α 대립유전자(들)의 하나 이상의 기능을 손상시키거나, 실질적으로 손상시키거나, 제거하거나 또는 TCR α 대립유전자(들)를 비기능성으로 제공한다. 특정 실시형태에서, TCR α 기능은 세포 표면에 대한 CD3 동원(recruiting), MHC 의존적 인식 및 항원의 결합, TCR α β 신호전달의 활성화를 포함하지만, 이들로 제한되지 않는다.

[0088] 다양한 실시형태에서 상정된 게놈 편집 조성물 및 방법은 인간 T 세포 수용체 알파(TCR α) 유전자 내 표적 부위에 결합하고 이를 절단하도록 설계된 뉴클레아제 변이체를 포함한다. 특정 실시형태에서 상정된 뉴클레아제 변이체는 폴리뉴클레오타이드 주형, 예를 들어, 공여자 수선 주형의 부재 하에 비상동성 말단 결합(NHEJ)에 의해, 또는 공여자 수선 주형의 존재 하에 상동 직접 수선(HDR), 즉, 상동성 재조합에 의해 수선될 수 있는 표적 폴리뉴클레오타이드 서열에서 이중-가닥 파손을 도입하기 위해 사용될 수 있다. 특정 실시형태에서 상정된 뉴클레아제 변이체는 공여자 수선 주형의 존재 하에 세포의 염기-절단-수선(base-excision-repair: BER) 기작 또는 상동성 재조합을 이용하여 수선될 수 있는 단일-가닥 DNA 파손을 생성하는 틈내기효소로서 설계될 수 있다. NHEJ는 유전자 기능을 파괴하는 작은 삽입 및 결실의 제형을 빈번하게 초래하는 오류-유발 과정이다. 상동성 재조합은 수선을 위한 주형으로서 상동성 DNA를 필요로 하며, 표적 부위에 축적하는 영역에 대해 상동성을 보유하는 서열에 의해 측면 중 하나 상에 축적된, 표적 부위에서 목적으로 하는 서열을 함유하는 공여자 DNA의 도입에 의해 구체화되는 제한 없는 다양한 변형을 생성하도록 영향력을 가질 수 있다.

[0089] 하나의 바람직한 실시형태에서, 본 명세서에 상정된 게놈 편집 조성물은 인간 TCR α 유전자를 표적화하는 호밍 엔도뉴클레아제 변이체 또는 메가TAL을 포함한다.

[0090] 다양한 실시형태에서, 본 명세서에 상정된 호밍 엔도뉴클레아제 변이체 또는 메가TAL을 표적화하는 TCR α 유전자는 TCR α 유전자를 표적화하는 기존의 호밍 엔도뉴클레아제 변이체 또는 메가TAL에 비해 표적 부위에 대해 증가된 결합 부위 선택성 또는 특이성을 가진다.

[0091] 다양한 실시형태에서, 본 명세서에 상정된 호밍 엔도뉴클레아제 변이체 또는 메가TAL을 표적화하는 TCR α 유전자는 표적 부위에 대해 개선된 결합 부위 선택성 또는 특이성을 갖는 한편, TCR α 유전자를 표적화하는 호밍 엔도뉴클레아제 변이체 또는 메가TAL에 비해 높은 촉매적 활성을 보유한다.

[0092] 하나의 바람직한 실시형태에서, 본 명세서에 상정된 게놈 편집 조성물은 호밍 엔도뉴클레아제 변이체 또는 메가TAL 및 말단-가공 효소, 예를 들어, Trex2를 포함한다.

[0093] 특정 실시형태의 실행은, 달리 반대로 구체적으로 나타내지 않는 한, 화학, 생화학, 유기 화학, 분자 생물학, 미생물학, 재조합 DNA 기법, 유전자 면역 및 당업계 기술 내의 세포 생물학의 통상적인 방법을 사용할 것이며, 이 중 다수는 예시의 목적을 위해 이하에 기재한다. 이러한 기법은 문헌에서 완전하게 설명된다. 예를 들어, 문헌[Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (3rd Edition, 2001); Sambrook, et al., *Molecular Cloning: A Laboratory Manual* (2nd Edition, 1989); Maniatis et al., *Molecular Cloning: A Laboratory Manual* (1982); Ausubel et al., *Current Protocols in Molecular Biology* (John Wiley and Sons, updated July 2008); *Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology*, Greene Pub. Associates and Wiley-Interscience; Glover, *DNA Cloning: A Practical Approach*, vol. I & II (IRL Press, Oxford, 1985); Anand, *Techniques for the Analysis of Complex Genomes*, (Academic Press, New York, 1992); *Transcription and Translation* (B. Hames & S. Higgins, Eds., 1984); Perbal, *A Practical Guide to Molecular Cloning* (1984); Harlow and Lane, *Antibodies*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1998) *Current Protocols in*

*Immunology* Q. E. Coligan, A. M. Kruisbeek, D. H. Margulies, E. M. Shevach and W. Strober, eds., 1991); *Annual Review of Immunology*]; 뿐만 아니라 문헌[*Advances in Immunology*]과 같은 저널 내의 모노그래프를 참조.

[0094] **B. 정의**

[0095] 달리 정의되지 않는 한, 본 명세서에서 사용되는 모든 기술적 및 과학적 용어는 본 발명이 속하는 기술분야의 당업자에 의해 통상적으로 이해되는 것과 동일한 의미를 가진다. 본 명세서에 기재된 것과 유사한 또는 동등한 임의의 방법 및 물질이 특정 실시형태의 실행 또는 시험에서 사용될 수 있지만, 조성물, 방법 및 물질의 바람직한 실시형태가 본 명세서에 기재된다. 본 개시내용의 목적을 위해, 다음의 용어를 이하에 정의한다.

[0096] 단수의 항목은 본 명세서에서 항목의 문법적 대상의 하나 이상(즉, 적어도 하나, 또는 하나 이상)을 지칭하기 위해 사용된다. 예로서, "요소"는 하나의 요소 또는 하나 이상의 요소를 의미한다.

[0097] 대안의(예를 들어, "또는")의 사용은 대안 중 하나, 둘 다, 또는 이들의 임의의 조합을 의미하는 것으로 이해되어야 한다.

[0098] 용어 "및/또는"은 대안 중 하나, 또는 둘 다를 의미하는 것으로 이해되어야 한다.

[0099] 본 명세서에서 사용되는 용어 "약" 또는 "대략"은 기준 양, 수준, 값, 숫자, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이에 대해 15%, 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2% 또는 1%만큼 변하는 기준 양, 수준, 값, 숫자, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이를 지칭한다. 일 실시형태에서, 용어 "약" 또는 "대략"은 기준 양, 수준, 값, 숫자, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이에 대해  $\pm 15\%$ ,  $\pm 10\%$ ,  $\pm 9\%$ ,  $\pm 8\%$ ,  $\pm 7\%$ ,  $\pm 6\%$ ,  $\pm 5\%$ ,  $\pm 4\%$ ,  $\pm 3\%$ ,  $\pm 2\%$  또는  $\pm 1\%$ 의 기준 양, 수준, 값, 숫자, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이의 범위를 지칭한다.

[0100] 일 실시형태에서, 범위, 예를 들어, 1 내지 5, 약 1 내지 5, 또는 약 1 내지 약 5는 범위에 의해 포함되는 각각의 수치적 값을 지칭한다. 예를 들어, 하나의 비제한적 그리고 단지 예시적인 실시형태에서, 범위 "1 내지 5"는 표현 1, 2, 3, 4, 5; 또는 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 또는 5.0; 또는 1.0, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2.0, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3.0, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9 또는 5.0과 동등하다.

[0101] 본 명세서에서 사용되는 용어 "실질적으로"는 기준 양, 수준, 값, 숫자, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이에 비해 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 이상인 기준 양, 수준, 값, 숫자, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이를 지칭한다. 일 실시형태에서, "실질적으로 동일한"은 기준 양, 수준, 값, 숫자, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이와 거의 동일한 효과, 예를 들어, 생리적 효과를 생산하는 기준 양, 수준, 값, 숫자, 빈도, 백분율, 치수, 크기, 양, 중량 또는 길이를 지칭한다.

[0102] 본 명세서 전체적으로, 문맥에서 달리 요구되지 않는 한, 단어 "포함하다(comprise)", "포함하다(comprises)" 및 "포함하는(comprising)"은 단계들 또는 요소들의 언급된 단계 또는 요소 또는 그룹의 포함하지만, 임의의 다른 단계 또는 요소 또는 단계들의 제외는 아니라는 것을 나타내는 것으로 이해될 것이다. "이루어진"은 어구 "이루어진"에 따르는 것을 포함하지만, 이것으로 제한되지 않음을 의미한다. 따라서, 어구 "이루어진"은 열거된 요소가 필요하거나 또는 필수라는 것, 다른 요소가 존재하지 않을 수도 있다는 것을 나타낸다. "본질적으로 이루어진"은 어구 다음에 열거된 임의의 요소를 포함하며, 열거된 요소에 대해 본 개시내용에서 구체화된 활성 또는 작용을 방해하지 않거나 또는 기여하지 않는 다른 요소로 제한되지 않음을 의미한다. 따라서, 어구 "본질적으로 이루어진"은 열거된 요소가 필요하거나 또는 필수이지만, 열거된 요소의 활성 또는 작용에 실질적으로 영향을 미치는 다른 요소가 존재하지 않는다는 것을 나타낸다.

[0103] 본 명세서 전체적으로 "일 실시형태", "실시형태", "특정 실시형태", "관련된 실시형태", "특정 실시형태", "추가적인 실시형태" 또는 "추가 실시형태" 또는 이들의 조합에 대한 언급은 실시형태와 관련하여 기재된 특정 특성, 구조 또는 특징이 적어도 일 실시형태에 포함된다는 것을 의미한다. 따라서, 본 명세서 전체적으로 다양한 곳에서 앞서 언급한 어구의 출현은 반드시 모두 동일한 실시형태에 관한 것은 아니다. 더 나아가, 특정 특성, 구조 또는 특징은 하나 이상의 실시형태에서 임의의 적합한 방식으로 조합될 수 있다. 또한 일 실시형태에서 특징의 긍정적 인용은 특정 실시형태에서의 특성을 제외하기 위한 기초로서 작용한다는 것이 이해된다.

[0104] 용어 "생체외"는 일반적으로 유기체 외부에서 일어나는 활동, 예컨대 유기체 외부의 인공 환경 내 살아있는 조직에서 또는 이런 조직 상에서(바람직하게는 천연 조건의 최소 변경) 행해지는 실험 또는 측정을 지칭한다. 특



정 실시형태에서, "생체외" 절차는 유기체로부터 취한 그리고 실험실 장치에서, 보통 멸균 조건 하에, 전형적으로 환경에 따라서 몇 시간 또는 약 24시간까지(48시간 또는 72시간을 포함) 배양 또는 조절한 살아있는 세포 또는 조직을 수반한다. 특정 실시형태에서, 이러한 조직 또는 세포는 수집되고 냉동될 수 있으며, 이후에 생체외 처리를 위해 해동될 수 있다. 살아있는 세포 또는 조직을 이용하여 며칠 이상 지속되는 조직 배양 실험 또는 절차는 전형적으로 "시험관내"(특정 실시형태에서, 이 용어가 생체외와 상호 호환적으로 사용될 수 있지만)가 되는 것으로 고려된다.

[0105] 용어 "생체내"는 일반적으로 유기체 내에서 생기는 활성을 지칭한다. 일 실시형태에서, 세포 계능은 생체내에서 조작되거나, 편집되거나 또는 변형된다.

[0106] "향상시키다" 또는 "촉진시키다" 또는 "증가시키다" 또는 "확장하다" 또는 "강화하다"는 일반적으로 비히클 또는 대조군 중 하나에 의해 야기되는 반응에 비해 더 큰 반응(즉, 생리적 반응)을 생산하거나, 유발하거나 또는 야기하기 위한 본 명세서에 상정된 뉴클레아제 변이체, 계능 편집 조성물 또는 계능 편집 세포의 능력을 지칭한다. 측정 가능한 반응은, 특히 당업계의 이해 및 본 명세서의 설명으로부터 분명한 조작된 TCR 또는 CAR 발현의 증가, HR 또는 HDR 효율의 증가, 결합 부위 선택성의 증가, 결합 부위 특이성의 증가, 표적에 대한 결합의 증가, 면역 효과기 세포 확장의 증가, 활성화, 지속성, 및/또는 암세포 사멸 능력의 증가를 포함할 수 있다. "증가된" 또는 "향상된" 양은 전형적으로 "통계적으로 유의한" 양이고, 비히클 또는 대조군에 의해 생산되는 반응의 1.1, 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30배 이상(예를 들어, 500, 1000배)(그 사이의 그리고 1 초과의 모든 정수 및 소수점을 포함, 예를 들어, 1.5, 1.6, 1.7, 1.8 등)인 증가를 포함할 수 있다.

[0107] "감소시키다" 또는 "낮추다" 또는 "줄이다" 또는 "줄어들게 하다" 또는 "없애다" 또는 "제거하다" 또는 "저해하다" 또는 "감쇠하다"는 일반적으로 비히클 또는 대조군 중 하나에 의해 야기되는 반응에 비해 더 적은 반응(즉, 생리적 반응)을 생산하거나, 유발하거나 또는 야기하기 위한 본 명세서에 상정된 뉴클레아제 변이체, 계능 편집 조성물 또는 계능 편집 세포의 능력을 지칭한다. 측정 가능한 반응은 내인성 TCR 발현 또는 기능의 감소, 표적 이탈의 감소, 면역효과기 세포 고갈과 관련된 발현의 감소 등을 포함할 수 있다. "감소된" 또는 "줄어든" 양은 전형적으로 "통계적으로 유의한" 양이며, 비히클 또는 대조군에 의해 생산된 반응(기준 반응)의 1.1, 1.2, 1.5, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 30배 이상(예를 들어, 500, 1000배)(그 사이의 그리고 1 초과의 모든 정수 및 소수점을 포함, 예를 들어, 1.5, 1.6, 1.7, 1.8 등)인 감소를 포함할 수 있다.

[0108] "유지하다" 또는 "보존하다" 또는 "유지" 또는 "변화 없음" 또는 "실질적인 변화 없음" 또는 "실질적인 감소 없음"은 일반적으로 비히클 또는 대조군에 의해 야기되는 반응에 비해 실질적으로 유사한 또는 비슷한 생리적 반응(즉, 하류 효과)를 생산하거나, 유발하거나 또는 야기하기 위한 본 명세서에 상정된 뉴클레아제 변이체, 계능 편집 조성물 또는 계능 편집 세포의 능력을 지칭한다. 비슷한 반응은 기준 반응과 상당히 상이하지 않은 또는 측정 가능하게 상이하지 않은 것이다.

[0109] 용어 "특이적 결합 친화도" 또는 "특이적으로 결합하다" 또는 "특이적으로 결합된" 또는 "특이적 결합" 또는 "특이적으로 표적화하다"는 배경 결합보다 더 큰 결합 친화도로 하나의 분자의 다른 분자에 대한 결합, 예를 들어, DNA에 대한 폴리펩타이드 결합의 DNA 결합 도메인을 기재한다. 결합 도메인이, 예를 들어, 약  $10^5 \text{ M}^{-1}$  이상의 친화도 또는  $K_d$ (즉, 1/M의 단위로 특정 결합 상호작용의 평형 결합 상수)로 표적 부위에 결합하거나 또는 회합된다면, 결합 도메인은 표적 부위에 "특이적으로 결합한다". 특정 실시형태에서, 결합 도메인은 약  $10^6 \text{ M}^{-1}$ ,  $10^7 \text{ M}^{-1}$ ,  $10^8 \text{ M}^{-1}$ ,  $10^9 \text{ M}^{-1}$ ,  $10^{10} \text{ M}^{-1}$ ,  $10^{11} \text{ M}^{-1}$ ,  $10^{12} \text{ M}^{-1}$  또는  $10^{13} \text{ M}^{-1}$  이상의  $K_d$ 로 표적 부위에 결합한다. "고친화도" 결합 도메인은 적어도  $10^7 \text{ M}^{-1}$ , 적어도  $10^8 \text{ M}^{-1}$ , 적어도  $10^9 \text{ M}^{-1}$ , 적어도  $10^{10} \text{ M}^{-1}$ , 적어도  $10^{11} \text{ M}^{-1}$ , 적어도  $10^{12} \text{ M}^{-1}$ , 적어도  $10^{13} \text{ M}^{-1}$  이상의  $K_d$ 를 갖는 해당 결합 도메인을 지칭한다.

[0110] 대안적으로, 친화도는 M 단위(예를 들어,  $10^{-5} \text{ M}$  내지  $10^{-13} \text{ M}$  이하)를 갖는 특정 결합 상호작용의 평형 해리 상수( $K_d$ )로서 정의될 수 있다. 특정 실시형태에서 상정된 DNA 표적 부위에 대한 하나 이상의 DNA 결합 도메인을 포함하는 뉴클레아제 변이체의 친화도는 통상적인 기법, 예를 들어, 효모 세포 표면 디스플레이를 이용하거나, 또는 결합 회합 또는 표지된 리간드를 이용하는 변위 분석에 의해 용이하게 결정될 수 있다.

[0111] 일 실시형태에서, 특이적 결합의 친화도는 배경 결합보다 약 2배 초과, 배경 결합보다 약 5배 초과, 배경 결합보다 약 10배 초과, 배경 결합보다 약 20배 초과, 배경 결합보다 약 50배 초과, 배경 결합보다 약 100배 초과

또는 배경 결합보다 약 1000배 초과 또는 그 이상이다.

- [0112] 용어 "선택적으로 결합한다" 또는 "선택적으로 결합된" 또는 "선택적으로 결합하는" 또는 "선택적으로 표적화한다"는 복수의 표적이탈 분자의 존재 하의 표적 분자(표적 상의 결합)에 대한 하나의 분자의 우선적인 결합을 설명한다. 특정 실시형태에서, HE 또는 메가TAL은 HE보다 약 5, 10, 15, 20, 25, 50, 100 또는 1000배 더 빈번하게 표적 상의 DNA 결합 부위에 선택적으로 결합하거나 또는 메가TAL은 표적이탈 DNA 표적 결합 부위에 결합한다.
- [0113] "표적 상"은 표적 부위 서열을 지칭한다.
- [0114] "표적이탈"은 표적 부위 서열과 유사하지만 동일하지 않은 서열을 지칭한다.
- [0115] "표적 부위" 또는 "표적 서열"은 결합 및/또는 절단이 존재하는 충분한 조건 하에서, 결합 분자에 결합하고/하거나 절단하는 핵산의 일부를 정하는 염색체 또는 염색체와 핵산 서열이다. 표적 부위 또는 표적 서열의 한 가닥만을 언급하는 폴리뉴클레오타이드 서열 또는 서열번호를 지칭할 때, 뉴클레아제 변이체에 의해 결합되고/되거나 절단되는 표적 부위 또는 표적 서열이 이중 가닥이고 기준 서열 및 그의 상보체를 포함한다는 것이 이해될 것이다. 바람직한 실시형태에서, 표적 부위는 인간 TCR  $\alpha$  유전자 내 서열이다.
- [0116] "유전자"는, 이러한 조절 서열이 암호 및/또는 전사된 서열에 인접하든 아니든 유전자 산물을 암호화하는 DNA 영역뿐만 아니라 유전자 산물의 생산을 조절하는 모든 DNA 영역을 지칭한다. 유전자는 프로모터 서열, 인핸서, 사일런서, 절연체, 경계 요소, 종결자, 폴리아데닐화 서열, 전사후 반응 요소, 번역 조절 서열, 예컨대 리보솜 결합 부위 및 내부 리보솜 유입 부위, 복제기점, 기질 부착 부위 및 좌위 제어 영역을 포함하지만, 이들로 제한되지 않는다.
- [0117] 본 명세서에서 사용되는 용어 "유전자 조작된" 또는 "유전자 변형된"은 DNA 또는 RNA 형태의 유전자와 물질의 세포 내 총 유전자 물질에 대한 염색체 또는 염색체의 첨가를 지칭한다. 유전자 변형은 세포 계능에서 특정 부위로 표적화 또는 비표적화될 수 있다. 일 실시형태에서, 유전자 변형은 부위 특이적이다. 일 실시형태에서, 유전자 변형은 부위 특이적이지 않다.
- [0118] 본 명세서에서 사용되는 용어 "계능 편집"은 유전자 또는 유전자 산물의 발현을 회복시키고/시키거나, 수정하고/하거나 파괴하고/하거나 변형시키는 세포 계능 내 표적 부위에서의 유전자 물질의 치환, 결실 및/또는 도입을 지칭한다. 특정 실시형태에서 상정된 계능 편집은 세포 계능에서, 선택적으로 공여자 수선 주형의 존재 하에 표적 부위에서 또는 표적 부위 근위에서 DNA 손상을 생성하기 위해 세포 내로 하나 이상의 뉴클레아제 변이체를 도입하는 것을 포함한다.
- [0119] 본 명세서에서 사용되는 용어 "양"은 임상 결과를 포함하는 유리한 또는 목적으로 하는 예방적 또는 치료적 결과를 달성하는 데 충분한 뉴클레아제 변이체, 계능 편집 조성물, 또는 계능 편집 세포의 "효과적인 양" 또는 "유효량"을 지칭한다.
- [0120] "예방적 유효량"은 목적으로 하는 예방적 결과를 달성하는 데 충분한 뉴클레아제 변이체, 계능 편집 조성물 또는 계능 편집 세포의 양을 지칭한다. 전형적으로, 만드지는 아니지만, 예방적 용량은 질환 전에 또는 초기에 대상체에서 사용되기 때문에, 예방적 유효량은 치료적 유효량보다 적다.
- [0121] 뉴클레아제 변이체, 계능 편집 조성물 또는 계능 편집 세포의 "치료적 유효량"은 개체의 질환 상태, 연령, 성별 및 체중 및 개체에서 목적으로 하는 반응을 유발하는 능력과 같은 인자에 따라 다를 수 있다. 치료적 유효량은 또한 임의의 독성 또는 유해한 효과가 치료적으로 유리한 효과보다 더 큰 것이다. 용어 "치료적 유효량"은 대상체(예를 들어, 환자)를 "치료"하는 데 효과적인 양을 포함한다. 치료적 양이 표시될 때, 특정 실시형태에서 상정되는 투여될 조성물의 정확한 양은 본 명세서를 고려하여 그리고 환자(대상체)의 연령, 체중, 종양 크기, 감염 또는 전이 정도 및 병태에서 개개 차이를 고려하여 결정될 수 있다.
- [0122] **C. 뉴클레아제 변이체**
- [0123] 본 명세서의 특정 실시형태에서 상정되는 뉴클레아제 변이체는 TCR  $\alpha$  유전자 내 표적 부위를 계능 편집하는 데 적합하고, 하나 이상의 DNA 결합 도메인 및 하나 이상의 DNA 절단 도메인(예를 들어, 하나 이상의 엔도뉴클레아제 및/또는 엑소뉴클레아제 도메인), 및 선택적으로 본 명세서에 상정된 하나 이상의 링커를 포함한다. 용어 "재프로그래밍된 뉴클레아제", "조작된 뉴클레아제" 또는 "뉴클레아제 변이체"는 상호 호환적으로 사용되며, 하나 이상의 DNA 결합 도메인 및 하나 이상의 DNA 절단 도메인을 포함하는 뉴클레아제를 지칭하되, 상기 뉴클레아제는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 이중 가닥 DNA 표적 서열에 결합하고 절단하도록 모 또는 천

연 유래 뉴클레아제로부터 설계되고/되거나 변형되었다.

- [0124] 특정 실시형태에서, 뉴클레아제 변이체는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 표적 서열, 바람직하게는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 서열번호 17에서, 더 바람직하게는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 서열번호 17에서 서열 "ATCC"에 결합하고, 절단한다.
- [0125] 뉴클레아제 변이체는 천연 유래 뉴클레아제로부터 또는 이전의 뉴클레아제 변이체로부터 설계되고/되거나 변형될 수 있다. 특정 실시형태에서 상정된 뉴클레아제 변이체는 하나 이상의 추가적인 작용성 도메인, 예를 들어, 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제(예를 들어, Trex2), 5' 플랩 엔도뉴클레아제, 헬리카제, 주형-의존적 DNA 중합효소 또는 주형-독립적 DNA 중합효소 활성을 나타내는 말단-가공 효소의 말단-가공 효소 도메인을 추가로 포함할 수 있다.
- [0126] TCR  $\alpha$  유전자 내 표적 서열에 결합하고 절단하는 뉴클레아제 변이체의 예시적 예는 호밍 엔도뉴클레아제(메가뉴클레아제) 변이체 및 메가TAL를 포함하지만, 이들로 제한되지 않는다.
- [0127] **1. 호밍 엔도뉴클레아제(메가뉴클레아제) 변이체**
- [0128] 다양한 실시형태에서, 호밍 엔도뉴클레아제 또는 메가뉴클레아제는 TCR  $\alpha$  유전자 내 표적 부위에서 이중-가닥 파손(DSB)을 도입하도록 재프로그래밍된다. 특정 실시형태에서, 호밍 엔도뉴클레아제 변이체는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1에서, 바람직하게는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 서열번호 17에서, 그리고 더 바람직하게는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 서열번호 17에서의 서열 "ATTC"에서 이중 가닥 파손을 도입한다. "호밍 엔도뉴클레아제" 및 "메가뉴클레아제"는 상호 교환적으로 사용되며, 12 내지 45개의 염기쌍 절단 부위를 인식하고, 서열 및 구조 모티프에 기반하여 5개의 패밀리(LAGLIDADG, GIY-YIG, HNH, His-Cys 박스 및 PD-(D/E)XK)로 통상적으로 그룹화되는 천연 유래 호밍 엔도뉴클레아제를 지칭한다.
- [0129] "기준 호밍 엔도뉴클레아제" 또는 "기준 메가뉴클레아제"는 천연에서 발견되는 야생형 호밍 엔도뉴클레아제 또는 호밍 엔도뉴클레아제를 지칭한다. 일 실시형태에서, "기준 호밍 엔도뉴클레아제"는 기본적 활성을 증가시키도록 변형된 야생형 호밍 엔도뉴클레아제를 지칭한다.
- [0130] "조작된 호밍 엔도뉴클레아제", "재프로그래밍된 호밍 엔도뉴클레아제", "호밍 엔도뉴클레아제 변이체", "조작된 메가뉴클레아제", "재프로그래밍된 메가뉴클레아제" 또는 "메가뉴클레아제 변이체"는 하나 이상의 DNA 결합 도메인 및 하나 이상의 DNA 절단 도메인을 포함하는 호밍 엔도뉴클레아제를 지칭하되, 호밍 엔도뉴클레아제는 TCR  $\alpha$  유전자 내 DNA 표적 서열에 결합하고 이를 절단하도록 모 또는 천연 유래 호밍 엔도뉴클레아제로부터 설계되고/되거나 변형되었다. 호밍 엔도뉴클레아제 변이체는 천연 유래 호밍 엔도뉴클레아제로부터 또는 다른 호밍 엔도뉴클레아제 변이체로부터 설계되고/되거나 변형될 수 있다. 특정 실시형태에서 상정된 호밍 엔도뉴클레아제 변이체는 하나 이상의 추가적인 작용성 도메인, 예를 들어, 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제(예를 들어, Trex2), 5' 플랩 엔도뉴클레아제, 헬리카제, 주형-의존적 DNA 중합효소 또는 주형-독립적 DNA 중합효소 활성을 나타내는 말단-가공 효소의 말단-가공 효소 도메인을 추가로 포함할 수 있다.
- [0131] 특정 실시형태에서, 본 명세서에 상정된 호밍 엔도뉴클레아제 변이체는 기존의 호밍 엔도뉴클레아제, 예를 들어, 서열번호 6에 비해 표적 부위에 대한 증가된 선택성을 포함한다. 특정 실시형태에서, 본 명세서에 상정된 호밍 엔도뉴클레아제는 표적 부위에 대한 증가된 선택성을 포함하는 한천, 기존의 호밍 엔도뉴클레아제, 예를 들어, 서열번호 6에 비해 촉매적 활성을 보유한다.
- [0132] 호밍 엔도뉴클레아제(HE) 변이체는 천연에서 존재하지 않으며, 재조합 DNA 기법에 의해 또는 무작위 돌연변이 유발에 의해 얻을 수 있다. HE 변이체는 하나 이상의 아미노산 변경을 만듦으로써, 예를 들어, 천연 유래 HE 또는 HE 변이체에서 하나 이상의 아미노산을 돌연변이, 치환, 첨가 또는 결실시킴으로써 얻을 수 있다. 특정 실시형태에서, HE 변이체는 DNA 인식 계면에 대한 하나 이상의 아미노산 변경을 포함한다.
- [0133] 특정 실시형태에서 상정된 HE 변이체는 하나 이상의 추가적인 링커 및/또는 추가적인 작용성 도메인, 예를 들어, 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제(예를 들어, Trex2), 5' 플랩 엔도뉴클레아제, 헬리카제, 주형-의존적 DNA 중합효소 또는 주형-독립적 DNA 중합효소 활성을 나타내는 말단-가공 효소의 말단-가공 효소 도메인을 추가로 포함할 수 있다. 특정 실시형태에서, HE 변이체는 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제(예를 들어, Trex2), 5' 플랩 엔도뉴클레아제, 헬리카제, 주형-의존적 DNA 중합효소 또는 주형-독립적 DNA 중합효소 활성을 나타내는 말단-가공 효소에 의해 T 세포 내로 도입된다. HE 변이체 및 3'가공 효소는 별개로, 예를 들어, 상이한 벡터 또는 별개의 mRNA에서, 또는

함께, 예를 들어, 융합 단백질로서, 또는 바이러스 자기-절단성 펩타이드 또는 IRES 요소에 의해 분리된 다실효 로성 작제물에서 도입될 수 있다.

- [0134] "DNA 인식 계면"은 핵산 표적 염기와 상호작용하는 HE 아미노산 잔기뿐만 아니라 인접한 해당 잔기를 지칭한다. 각각의 HE에 대해, DNA 인식 계면은 측쇄 대 측쇄 및 측쇄 대 DNA 접촉의 광대한 네트워크를 포함하며, 이 중 대부분은 특정 핵산 표적 서열을 인식하기에 충분히 독특하다. 따라서, 특정 핵산 서열에 대응하는 DNA 인식 계면의 아미노산 서열은 상당히 다르며, 임의의 천연 또는 HE 변이체의 특징이다. 비제한적 예로서, 특정 실시형태에 상정된 HE 변이체는 HE 변이체의 라이브러리를 작제함으로써 유도될 수 있으며, 이때 천연 HE(또는 이전에 생성된 HE 변이체)의 DNA 인식 계면에서 국소화된 하나 이상의 아미노산 잔기가 변한다. 라이브러리는 절단 분석을 이용하여 각각의 예측된 TCR  $\alpha$  표적 부위에 대한 표적 절단 활성에 대해 선별될 수 있다(예를 들어, 문헌 [Jarjour *et al.*, 2009. *Nuc. Acids Res.* **37**(20): 6871-6880] 참조).
- [0135] LAGLIDADG 호밍 엔도뉴클레아제(LHE)는 호밍 엔도뉴클레아제의 가장 잘 연구된 패밀리이며, 고세균에서 그리고 녹조류 및 진균에서의 세포소기관 DNA에서 주로 암호화되고, 가장 높은 전반적인 DNA 인식 특이성을 나타낸다. LHE는 단백질 쇠 당 1 또는 2개의 LAGLIDADG 촉매적 모티프를 포함하며, 동중이량체 또는 단일체 단량체로서 각각 작용한다. LAGLIDADG 단백질의 구조적 연구는  $\alpha\beta\beta\alpha\beta\beta\alpha$  폴드에 의해 특성규명되는 고도로 보존된 코어 구조를 동정하였고(Stoddard 2005), LAGLIDADG 모티프는 이 폴드의 첫 번째 나선에 속한다. LHE의 고도로 효율적인 그리고 특이적 절단은 신규한, 고도로 특이적이고 선택적인 엔도뉴클레아제를 유도하기 위한 단백질 스캐폴드를 나타낸다. 그러나, 비천연 또는 비정규 표적 부위를 결합하고 절단하기 위한 LHE의 조작은 표적 부위에서 염기쌍 위치의 2/3까지에서, 적절한 LHE 스캐폴드의 선택, 표적 좌위의 시험, 추정적 표적 부위의 선택, 및 DNA 접점 및 절단 특이성을 변경하기 위한 LHE의 광대한 변경을 필요로 한다.
- [0136] LHE의 예시적인 예는 I-AabMI, I-AaeMI, I-AniI, I-ApaMI, I-CapIII, I-CapIV, I-CkaMI, I-CpaMI, I-CpaMII, I-CpaMIII, I-CpaMIV, I-CpaMV, I-CpaV, I-CraMI, I-EjeMI, I-GpeMI, I-GpiI, I-GzeMI, I-GzeMII, I-GzeMIII, I-HjeMI, I-LtrII, I-LtrI, I-LtrWI, I-MpeMI, I-MveMI, I-NcrII, I-NcrI, I-NcrMI, I-OheMI, I-OnuI, I-OsoMI, I-OsoMII, I-OsoMIII, I-OsoMIV, I-PanMI, I-PanMII, I-PanMIII, I-PnoMI, I-ScuMI, I-SmaMI, I-SscMI 및 I-Vdi141I을 포함하지만, 이들로 제한되지 않는다.
- [0137] 일 실시형태에서, 재프로그래밍된 LHE 또는 LHE 변이체는 I-OnuI 변이체이다. 예를 들어, 서열번호 7 내지 8 참조.
- [0138] 일 실시형태에서, TCR  $\alpha$  유전자를 표적화하는 재프로그래밍된 I-OnuI LHE 또는 I-OnuI 변이체는 천연 I-OnuI 또는 이의 생물학적으로 활성인 단편(서열번호 1 내지 5)으로부터 생성되었다. 바람직한 실시형태에서, 인간 TCR  $\alpha$  유전자를 표적화하는 재프로그래밍된 I-OnuI LHE 또는 I-OnuI 변이체는 존재하는 I-OnuI 변이체로부터 생성되었다.
- [0139] 일 실시형태에서, 인간 TCR  $\alpha$  유전자에 결합하고, 이를 절단하는 I-OnuI LHE는 서열번호 7 또는 서열번호 8에 제시된 바와 같은 I-OnuI LHE 변이체, 및 이들의 생물학적으로 활성인 단편의 DNA 인식 계면과 적어도 99% 서열 동일성을 포함한다.
- [0140] 특정 실시형태에서, 인간 TCR  $\alpha$  유전자에 결합하고 이를 절단하는 I-OnuI LHE 변이체는 서열번호 7 또는 서열번호 8에 제시된 바와 같은 I-OnuI, 이들의 생물학적으로 활성인 단편 및/또는 이들의 추가적인 변이체의 DNA 인식 계면에서 하나 이상의 아미노산 치환 또는 변형을 포함한다.
- [0141] 특정 실시형태에서, HE 변이체는 TCR  $\alpha$  엑손 1 불변 영역 표적 부위를 절단하고, 서열번호 1 내지 5 중 임의의 하나, 또는 이의 생물학적으로 활성인 단편의 다음의 아미노산 치환: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E 및 V238E를 포함한다.
- [0142] 특정 실시형태에서, HE 변이체는 TCR  $\alpha$  엑손 1 불변 영역 표적 부위를 절단하고, 서열번호 1 내지 5 중 임의의 하나, 또는 이의 생물학적으로 활성인 단편의 다음의 아미노산 치환: L26I, R28D, N32R, K34N, S35E, V37N, G38R, S40R, E42S, G44R, V68K, A70T, G73S, N75R, S78M, K80R, L138M, T143N, S159P, S176A, E178D, C180H, F182G, I186K, S188V, S190G, K191T, L192A, G193K, Q195Y, Q197G, V199R, S201A, T203S, K207R, Y223S, K225R, S233R, D236E 및 V238E를 포함한다.
- [0143] 특정 실시형태에서, 인간 TCR  $\alpha$  유전자에 결합하고 이를 절단하는 I-OnuI LHE 변이체는 서열번호 7 또는 서열번호



호 8 중 임의의 하나에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편에 대해 적어도 99% 동일한 아미노산 서열을 포함한다.

[0144] 특정 실시형태에서, I-OnuI LHE 변이체는 서열번호 7 내지 8 중 임의의 하나, 또는 이의 생물학적으로 활성인 단편에 제시된 아미노산 서열을 포함한다.

[0145] 특정 실시형태에서, I-OnuI LHE 변이체는 서열번호 7에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함한다.

[0146] 특정 실시형태에서, I-OnuI LHE 변이체는 서열번호 8에 제시된 아미노산 서열, 또는 이의 생물학적으로 활성인 단편을 포함한다.

## [0147] 2. 메가TAL

[0148] 다양한 실시형태에서, 호밍 엔도뉴클레아제 변이체를 포함하는 메가TLA는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 표적 부위에서 이중-가닥 파손(DSB)을 도입하도록 재프로그래밍된다. 특정 실시형태에서, 메가TAL은 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1에서, 바람직하게는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 서열번호 17에서, 더 바람직하게는 인간 TCR  $\alpha$  유전자의 불변 영역의 엑손 1 내 서열번호 17에서 서열 "ATTC"에서 DSB를 도입한다. "메가TAL"은 TCR  $\alpha$  유전자 내 DNA 표적 서열에 결합하고, 이를 절단하며, 선택적으로 하나 이상의 링커 및/또는 추가적인 작용성 도메인, 예를 들어, 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제(예를 들어, Trex2), 5' 플랩 엔도뉴클레아제, 헬리카제 또는 주형-독립적 DNA 중합효소 활성을 나타내는 말단-가공 효소의 말단-가공 효소 도메인을 포함하는, TALE DNA 결합 도메인 및 호밍 엔도뉴클레아제 변이체를 포함하는 폴리펩타이드를 지칭한다.

[0149] 특정 실시형태에서, 메가TAL은 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제(예를 들어, Trex2), 5' 플랩 엔도뉴클레아제, 헬리카제, 주형-의존적 DNA 중합효소 또는 주형-독립적 DNA 중합효소 활성을 나타내는 말단-가공 효소를 따라서 세포 내로 도입된다. 메가TAL 및 3'가공 효소는 별개로, 예를 들어, 상이한 벡터 또는 별개의 mRNA에서, 또는 함께, 예를 들어, 융합 단백질로서, 또는 바이러스 자기-절단성 펩타이드 또는 IRES 요소에 의해 분리된 다시트로성 작제물에서 도입될 수 있다.

[0150] "TALE DNA 결합 도메인"은 식물 전사체를 조작하기 위해 식물 전사 활성체를 모방하는 전사 활성체-유사 효과기(TALE 또는 TAL-효과기)의 DNA 결합 부분이다(예를 들어, 문헌[Kay *et al.*, 2007. *Science* 318:648-651] 참조). 특정 실시형태에서 상정된 TALE DNA 결합 도메인은 드노보(*de novo*)로 또는 천연 유래 TALE로부터, 예를 들어, AvrBs3 from 세균점무늬병(*Xanthomonas campestris* pv. *vesicatoria*), 잔토모나스 가드네리(*Xanthomonas gardneri*), 잔토모나스 트랜스루센스(*Xanthomonas translucens*), 잔토모나스 악소노포디스(*Xanthomonas axonopodis*), 잔토모나스 퍼포란스(*Xanthomonas perforans*), 잔토모나스 알팔파(*Xanthomonas alfalfa*), 잔토모나스 시트리(*Xanthomonas citri*), 잔토모나스 유베시카토리아(*Xanthomonas euvesicatoria*) 및 잔토모나스 오리자에(*Xanthomonas oryzae*) 및 랄스토니아 솔라나세아룸(*Ralstonia solanacearum*)으로부터의 brg11 및 hpx17로부터 조작된다. DNA 결합 도메인을 유도 및 설계하기 위한 TALE 단백질의 예시적인 예는 미국 특허 제9,017,967호 및 그의 참고문헌에 개시되어 있으며, 이들 모두는 본 명세서에 그들의 전문이 참고로 포함된다.

[0151] 특정 실시형태에서, 메가TAL은 TALE DNA 결합 도메인의 그의 대응하는 표적 DNA 서열에 대한 결합에 연루된 하나 이상의 반복 단위를 포함하는 TALE DNA 결합 도메인을 포함한다. 단일 "반복 단위"(또한 "반복부"로서 지칭됨)는 전형적으로 길이가 33 내지 35개 아미노산이다. 각각의 TALE DNA 결합 도메인 반복부 단위는 전형적으로 반복부의 위치 12 및/또는 13에서 반복부 가변 2-잔기(RVD)를 구성하는 1 또는 2개의 DNA-결합 잔기를 포함한다. 이들 TALE DNA 결합 도메인의 DNA 인식을 위한 천연(정규) 암호는 위치 12 및 13에서 HD 서열이 사이토신(C)에 대한 결합을 야기하고, NG가 T, NI 내지 A에 결합하며, NN은 G 또는 A에 결합하고, NG가 T에 결합하도록 결정되었다. 특정 실시형태에서, 비-정규(비전형적) RVD가 상정된다.

[0152] 특정 실시형태에서 상정된 특정 메가TAL에서 사용하기에 적합한 비정규 RVD의 예시적 예는 구아닌(G)의 인식을 위해 HH, KH, NH, NK, NQ, RH, RN, SS, NN, SN, KN; 아데닌(A)의 인식을 위해 NI, KI, RI, HI, SI; 티민(T) NG, HG, KG, RG; 사이토신(C)의 인식을 위해 RD, SD, HD, ND, KD, YG; A 또는 G의 인식을 위해 NV, HN; 및 A 또는 T 또는 G 또는 C의 인식을 위해 H\*, HA, KA, N\*, NA, NC, NS, RA, S\*을 포함하지만, 이들로 제한되지 않되, (\*)는 위치 13에서 아미노산이 없다는 것을 의미한다. 특정 실시형태에서 상정된 특정 메가TAL에서 사용하기에 적합한 RVD의 추가적인 예시적 예는 본 명세서에 전문이 참고로 포함된 미국 특허 제8,614,092호에 개시된 것을 추가로 포함한다.

- [0153] 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 3 내지 30개의 반복 단위를 포함하는 TALE DNA 결합 도메인을 포함한다. 특정 실시형태에서, 메가TAL은 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 또는 30개의 TALE DNA 결합 도메인 반복 단위를 포함한다. 바람직한 실시형태에서, 본 명세서에 상정된 메가TAL은 5 내지 15개의 반복 단위, 더 바람직하게는 7 내지 15개의 반복 단위, 더 바람직하게는 9 내지 15개의 반복 단위, 및 더 바람직하게는 9, 10, 11, 12, 13, 14 또는 15개의 반복 단위를 포함하는 TALE DNA 결합 도메인을 포함한다.
- [0154] 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 3 내지 30개의 반복 단위를 포함하는 TALE DNA 결합 도메인 및 한 세트의 TALE 반복 단위의 C-말단에 위치한 20개의 아미노산을 포함하는 추가적인 절단된 TALE 반복부 단위, 즉, 추가적인 C-말단 절반-TALE DNA 결합 도메인 반복부 단위(본 명세서의 다른 곳에 개시된 C-캡의 아미노산 -20 내지 -1, 이하 참조)를 포함한다. 따라서, 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 3.5 내지 30.5개의 반복 단위를 포함하는 TALE DNA 결합 도메인을 포함한다. 특정 실시형태에서, 메가TAL은 3.5, 4.5, 5.5, 6.5, 7.5, 8.5, 9.5, 10.5, 11.5, 12.5, 13.5, 14.5, 15.5, 16.5, 17.5, 18.5, 19.5, 20.5, 21.5, 22.5, 23.5, 24.5, 25.5, 26.5, 27.5, 28.5, 29.5 또는 30.5개의 TALE DNA 결합 도메인 반복 단위를 포함한다. 바람직한 실시형태에서, 본 명세서에 상정된 메가TAL은 5.5 내지 15.5개의 반복 단위, 더 바람직하게는 7.5 내지 15.5개의 반복 단위, 더 바람직하게는 9.5 내지 15.5개의 반복 단위, 및 더 바람직하게는 9.5, 10.5, 11.5, 12.5, 13.5, 14.5 또는 15.5개의 반복 단위를 포함하는 TALE DNA 결합 도메인을 포함한다.
- [0155] 특정 실시형태에서, 메가TAL은 "N-말단의 도메인(NTD)" 폴리펩타이드, 하나 이상의 TALE 반복부 도메인/단위, "C-말단의 도메인(CTD)" 폴리펩타이드 및 호밍 엔도뉴클레아제 변이체를 포함하는 TLA 효과기 구조를 포함한다. 일부 실시형태에서, NTD, TALE 반복부 및/또는 CTD 도메인은 동일한 종으로부터 유래된다. 다른 실시형태에서, NTD, TALE 반복부 및/또는 CTD 도메인 중 하나 이상은 상이한 종으로부터 유래된다.
- [0156] 본 명세서에서 사용되는 용어 "N-말단의 도메인(NTD)" 폴리펩타이드는 천연 유래 TALE DNA 결합 도메인의 N-말단 부분 또는 단편에 측접하는 서열을 지칭한다. NTD 서열은, 존재한다면, TALE DNA 결합 도메인 반복 단위가 DNA에 결합하는 능력을 보유하는 한, 임의의 길이를 가질 수 있다. 특정 실시형태에서, NTD 폴리펩타이드는 TALE DNA 결합 도메인에 대해 N-말단의 적어도 120개 내지 적어도 140개 이상의 아미노산을 포함한다(0은 가장 N-말단의 반복부 단위의 아미노산 1임). 특정 실시형태에서, NTD 폴리펩타이드는 TALE DNA 결합 도메인에 대해 N-말단의 적어도 약 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139 또는 적어도 140개의 아미노산을 포함한다. 일 실시형태에서, 본 명세서에 상정된 메가TAL은 잔토모나스 TALE 단백질의 적어도 약 아미노산 +1 내지 +122 내지 적어도 약 +1 내지 +137의 NTD 폴리펩타이드를 포함한다(0은 가장 N-말단의 반복부 단위의 아미노산 1이다). 특정 실시형태에서, NTD 폴리펩타이드는 잔토모나스 TALE 단백질의 TALE DNA 결합 도메인에 대해 N-말단의 적어도 약 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136 또는 137개의 아미노산을 포함한다. 일 실시형태에서, 본 명세서에 상정된 메가TAL은 랄스토니아 TALE 단백질의 적어도 아미노산 +1 내지 +121의 NTD 폴리펩타이드를 포함한다(0은 가장 N-말단의 반복부 단위의 아미노산 1이다). 특정 실시형태에서, NTD 폴리펩타이드는 랄스토니아 TALE 단백질의 TALE DNA 결합 도메인에 대해 N-말단의 적어도 약 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136 또는 137개의 아미노산을 포함한다.
- [0157] 본 명세서에서 사용되는 용어 "C-말단의 도메인(CTD)" 폴리펩타이드는 천연 유래 TALE DNA 결합 도메인의 C-말단 부분 또는 단편에 측접하는 서열을 지칭한다. CTD 서열은, 존재한다면, TALE DNA 결합 도메인 반복 단위가 DNA에 결합하는 능력을 보유하는 한, 임의의 길이를 가질 수 있다. 특정 실시형태에서, CTD 폴리펩타이드는 TALE DNA 결합 도메인의 마지막 완전한 반복부에 대한 C-말단의 적어도 20 내지 적어도 85개의 아미노산을 포함한다(처음 20개의 아미노산은 마지막 C-말단의 전체 반복부 단위에 대한 절반-반복부 단위 C-말단이다). 특정 실시형태에서, CTD 폴리펩타이드는 TALE DNA 결합 도메인의 마지막 완전한 반복부에 대해 C-말단의 적어도 약 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84 또는 적어도 85개의 아미노산을 포함한다. 일 실시형태에서, 본 명세서에 상정된 메가TAL은 잔토모나스 TALE 단백질의 적어도 약 아미노산 -20 내지 -1의 CTD 폴리펩타이드를 포함한다(-20은 마지막 C-말단의 전체 반복부 단위에 대한 절반-반복부 단위 C-말단의 아미노산 1이다). 특정 실시형태에서, CTD 폴리펩타이드는 잔토모나스 TALE 단백질의 TALE DNA 결합 도메인의 마지막 전체 반복부에 대해 C-말단의 적어도 약 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 또는 1개의 아미노산을 포함한다. 일 실시형태에서, 본 명세서에 상정된 메가TAL은 랄스토니아 TALE 단백질의

적어도 약 아미노산 -20 내지 -1의 CTD 폴리펩타이드를 포함한다(-20은 마지막 C-말단의 전체 반복부 단위에 대한 절반-반복부 단위 C-말단의 아미노산 1이다). 특정 실시형태에서, CTD 폴리펩타이드는 랄스토니아 TALE 단백질의 TALE DNA 결합 도메인의 마지막 전체 반복부에 대해 C-말단의 적어도 약 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2 또는 1개의 아미노산을 포함한다.

[0158] 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 선택적으로 본 명세서의 다른 곳에 상정된 하나 이상의 링커 폴리펩타이드와 서로 결합된 표적 서열에 결합하도록 조작된 TALE DNA 결합 도메인, 표적 서열에 결합하고 절단하도록 재프로그래밍된 호밍 엔도뉴클레아제, 및 선택적으로 NTD 및/또는 CTD 폴리펩타이드를 포함하는 융합 폴리펩타이드를 포함한다. 임의의 특정 이론에 의해 결합되는 일 없이, TALE DNA 결합 도메인, 및 선택적으로 NTD 및/또는 CTD 폴리펩타이드를 포함하는 메가TAL은 호밍 엔도뉴클레아제 변이체에 추가로 융합된 링커 폴리펩타이드에 융합된다는 것이 상정된다. 따라서, TALE DNA 결합 도메인은 호밍 엔도뉴클레아제 변이체의 DNA 결합 도메인에 의해 결합된 표적 서열로부터 떨어져서 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 또는 15개의 뉴클레오타이드 내인 DNA 표적 서열에 결합한다. 이런 방법으로, 본 명세서에 상정된 메가TAL은 게놈 편집의 특이성, 선택성 및 효율을 증가시킨다.

[0159] 일 실시형태에서, 메가TAL은 재프로그래밍된 호밍 엔도뉴클레아제의 결합 부위 상류의 약 4, 5 또는 6개의 뉴클레오타이드, 바람직하게는 5 또는 6개의 뉴클레오타이드 내인 뉴클레오타이드 서열에 결합하는 호밍 엔도뉴클레아제 변이체 및 TALE DNA 결합 도메인을 포함한다.

[0160] 일 실시형태에서, 메가TAL은 호밍 엔도뉴클레아제 변이체에 의해 결합되고 절단되는 뉴클레오타이드 서열(서열 번호 17) 상류의 6개 뉴클레오타이드인 서열번호 18에 제시된 뉴클레오타이드 서열에 결합하는 호밍 엔도뉴클레아제 변이체 및 TALE DNA 결합 도메인을 포함한다. 바람직한 실시형태에서, 메가TAL 표적 서열은 서열번호 18이다.

[0161] 일 실시형태에서, 메가TAL은 호밍 엔도뉴클레아제 변이체에 의해 결합되고 절단되는 뉴클레오타이드 서열(서열 번호 17) 상류의 5개 뉴클레오타이드인 서열번호 19에 제시된 뉴클레오타이드 서열에 결합하는 호밍 엔도뉴클레아제 변이체 및 TALE DNA 결합 도메인을 포함한다. 바람직한 실시형태에서, 메가TAL 표적 서열은 서열번호 19이다.

[0162] 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 약 122개의 아미노산 내지 137개의 아미노산, 약 9.5, 약 10.5, 약 11.5, 약 12.5, 약 13.5, 약 14.5 또는 약 15.5개의 결합 반복 단위의 NTD, 약 20개의 아미노산 내지 약 85개의 아미노산의 CTD, 및 I-OnuI LHE 변이체를 포함한다. 특정 실시형태에서, NTD, DNA 결합 도메인 및 CTD 중 임의의 하나, 둘 또는 모두는 임의의 적합한 조합에서 동일한 종 또는 상이한 종으로부터 설계될 수 있다.

[0163] 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 서열번호 10 내지 12 중 임의의 하나에 제시된 아미노산 서열을 포함한다.

[0164] 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 서열번호 10에 제시된 아미노산 서열을 포함한다.

[0165] 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 서열번호 11에 제시된 아미노산 서열을 포함한다.

[0166] 특정 실시형태에서, 본 명세서에 상정된 메가TAL은 서열번호 12에 제시된 뉴클레오타이드에 결합하고 이를 절단한다.

[0167] 특정 실시형태에서, 메가TAL은 TALE DNA 결합 도메인을 포함하고, I-OnuI LHE 변이체는 서열번호 17 또는 서열번호 20에 제시된 뉴클레오타이드에 결합하고 이를 절단한다.

### [0168] 3. 말단-가공 효소

[0169] 특정 실시형태에 상정된 게놈 편집 조성물 및 방법은 뉴클레아제 변이체 및 말단-가공 효소의 하나 이상의 복제물을 이용하여 세포 게놈을 편집하는 것을 포함한다. 특정 실시형태에서, 단일 폴리뉴클레오타이드는 링커, 자기-절단성 펩타이드 서열, 예를 들어, 2A 서열에 의해, 또는 IRES 서열에 의해 분리되는 호밍 엔도뉴클레아제 변이체 및 말단-가공 효소를 암호화한다. 특정 실시형태에서, 게놈 편집 조성물은 뉴클레아제 변이체를 암호화하는 폴리뉴클레오타이드 및 말단-가공 효소를 암호화하는 별개의 폴리뉴클레오타이드를 포함한다. 특정 실시형태에서, 게놈 편집 조성물은 자기-절단성 펩타이드에 의해 분리되는 말단-가공 효소의 탠덤 복제물에 추가로 호밍 엔도뉴클레아제 변이체 말단-가공 효소 단일 폴리펩타이드 융합을 암호화하는 폴리뉴클레오타이드를 포함한다.

다.

[0170] 용어 "말단-가공 효소"는 폴리뉴클레오타이드 쇄의 노출된 말단을 변형시키는 효소를 지칭한다. 폴리뉴클레오타이드는 이중 가닥 DNA(dsDNA), 단일 가닥 DNA(ssDNA), RNA, DNA 및 RNA의 이중 가닥 혼성체, 및 합성 DNA(예를 들어, A, C, G 및 T 이외의 염기를 함유)일 수 있다. 말단-가공 효소는 하나 이상의 뉴클레오타이드를 첨가하고/하거나 하나 이상의 뉴클레오타이드를 제거하고/하거나 인산염기를 제거 또는 변형시키고/시키거나 하이드록실기를 제거 또는 변형시킴으로써 노출된 폴리뉴클레오타이드쇄를 변형시킬 수 있다. 말단-가공 효소는 엔도뉴클레아제 절단 부위에서, 또는 다른 화학적 또는 기계적 수단, 예컨대 전단(예를 들어, 미세한 게이지의 바늘을 통과, 가열, 음파처리, 미니 비드 텀블링, 및 네블라이징에 의해), 이온화 방사선, 자외선 방사, 산소 라디칼, 화학적 가수분해 및 화학요법에 의해 생성되는 말단에서 말단을 변형시킬 수 있다.

[0171] 특정 실시형태에서, 특정 실시형태에서 상정된 게놈 편집 조성물 및 방법은 호밍 엔도뉴클레아제 변이체 또는 메가TAL 및 DNA 말단-가공 효소를 이용하여 세포 게놈을 편집하는 것을 포함한다.

[0172] 용어 "DNA 말단-가공 효소"는 DNA의 노출된 말단을 변형시키는 효소를 지칭한다. DNA 말단-가공 효소는 평활 말단 또는 엇갈린 말단(5' 또는 3' 돌출부를 갖는 말단)을 변형시킬 수 있다. DNA 말단-가공 효소는 단일 가닥 또는 이중 가닥 DNA를 변형시킬 수 있다. DNA 말단-가공 효소는 엔도뉴클레아제 절단 부위에서 또는 다른 화학적 또는 기계적 수단, 예컨대 전단(예를 들어, 미세한 게이지의 바늘을 통과, 가열, 음파처리, 미니 비드 텀블링, 및 네블라이징에 의해), 이온화 방사선, 자외선 방사, 산소 라디칼, 화학적 가수분해 및 화학요법에 의해 생성되는 말단에서 말단을 변형시킬 수 있다. DNA 말단-가공 효소는 하나 이상의 뉴클레오타이드를 첨가하고/하거나 하나 이상의 뉴클레오타이드를 제거하고/하거나 인산염기를 제거 또는 변형시키고/시키거나 하이드록실기를 제거 또는 변형시킴으로써 노출된 DNA 말단을 변형시킬 수 있다.

[0173] 본 명세서에 상정된 특정 실시형태에서 사용하기에 적합한 DNA 말단-가공 효소의 예시적인 예는 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 3'-5' 엑소뉴클레아제, 5' 플랩 엔도뉴클레아제, 헬리카제, 포스파타제, 가수분해효소 및 주형-독립적 DNA 중합효소를 포함하지만, 이들로 제한되지 않는다.

[0174] 본 명세서에 상정된 특정 실시형태에서 사용하기에 적합한 DNA 말단-가공 효소의 추가적인 예시적 예는 Trex2, Trex1, 막관통 도메인이 없는 Trex1, Apollo, Artemis, DNA2, Exo1, ExoT, ExoIII, Fen1, Fan1, MreII, Rad2, Rad9, TdT(말단의 테옥시뉴클레오타이드 트랜스퍼라제), PNKP, RecE, RecJ, RecQ, 람다 엑소뉴클레아제, Sox, 백시니아 DNA 중합효소, 엑소뉴클레아제 I, 엑소뉴클레아제 III, 엑소뉴클레아제 VII, NDK1, NDK5, NDK7, NDK8, WRN, T7-엑소뉴클레아제 유전자 6, 조류 골수아세포종 바이러스 통합 단백질(IN), Bloom, Antarctic 포스파타제, 알칼리성 포스파타제, 폴리뉴클레오타이드 키나제(PNK), ApeI, Mung Bean 뉴클레아제, Hex1, TTRAP (TDP2), Sgs1, Sae2, CUP, Pol mu, Pol lambda, MUS81, EME1, EME2, SLX1, SLX4 및 UL-12를 포함하지만, 이들로 제한되지 않는다.

[0175] 특정 실시형태에서, 본 명세서에 상정된 세포 게놈을 편집하기 위한 게놈 편집 조성물 및 방법은 호밍 엔도뉴클레아제 변이체 또는 메가TAL 및 엑소뉴클레아제를 포함하는 폴리펩타이드를 포함한다. 용어 "엑소뉴클레아제"는 3' 또는 5' 말단 중 하나에서 포스포다이에스터 결합을 파손하는 가수분해 반응을 통해 폴리뉴클레오타이드 쇄의 말단에서 포스포다이에스터 결합을 절단하는 효소 또는 도메인을 지칭한다.

[0176] 본 명세서에 상정된 특정 실시형태에서 사용하기에 적합한 엑소뉴클레아제의 예시적 예는 hExoI, 효모 ExoI, 이콜라이(*E. coli*) ExoI, hTREX2, 마우스 TREX2, 래트 TREX2, hTREX1, 마우스 TREX1, 래트 TREX1, 및 래트 TREX1을 포함하지만, 이들로 제한되지 않는다.

[0177] 특정 실시형태에서, DNA 말단-가공 효소는 3' 내지 5' 엑소뉴클레아제, 바람직하게는 Trex 1 또는 Trex2, 더 바람직하게는 Trex2, 및 더욱더 바람직하게는 인간 또는 마우스 Trex2이다.

#### [0178] D. 폴리펩타이드

[0179] 호밍 엔도뉴클레아제 변이체, 메가TAL 및 융합 폴리펩타이드를 포함하지만, 이들로 제한되지 않는 다양한 폴리펩타이드가 본 명세서에서 상정된다. 바람직한 실시형태에서, 폴리펩타이드는 서열번호 1 내지 12 및 22 내지 24에 제시된 아미노산 서열을 포함한다. "폴리펩타이드", "폴리펩타이드 단편", "펩타이드" 및 "단백질"은, 달리 대조적으로 구체화되지 않는 한, 상호 호환적으로 그리고 통상적인 의미에 따라, 즉, 아미노산 서열로서 사용된다. 일 실시형태에서, "폴리펩타이드"는 융합 폴리펩타이드 및 다른 변이체를 포함한다. 폴리펩타이드는 임의의 다양한 잘 공지된 제조법 및/또는 합성 기법을 이용하여 제조될 수 있다. 폴리펩타이드는 특정 길이로 제한되지 않으며, 예를 들어, 그들은 전장 단백질 서열, 전장 단백질의 단편 또는 융합 단백질을 포함할 수 있고,



폴리펩타이드의 번역후 변형, 예를 들어, 글리코실화, 아세틸화, 포스포릴화 등뿐만 아니라 당업계에 공지된 다른 변형, 즉, 천연 유래와 비천연 유래 둘 다를 포함할 수 있다.

[0180] 본 명세서에서 사용된 바와 같은 "단리된 단백질", "단리된 펩타이드" 또는 "단리된 폴리펩타이드" 등은 세포 환경으로부터 그리고 세포의 다른 성분과의 결합으로부터의 펩타이드 또는 폴리펩타이드 분자의 시험관내 합성, 단리 및/또는 정제를 지칭하며, 즉, 이는 생체내 물질과 상당하게 관련되지 않는다.

[0181] 특정 실시형태에서 상정된 폴리펩타이드의 예시적 예는 호밍 엔도뉴클레아제 변이체, 메가TAL, 말단-가공 뉴클레아제, 융합 폴리펩타이드 및 이의 변이체를 포함하지만, 이들로 제한되지 않는다.

[0182] 폴리펩타이드는 "폴리펩타이드 변이체"를 포함한다. 폴리펩타이드 변이체는 하나 이상의 아미노산 치환, 결실, 첨가 및/또는 삽입에서 천연 유래 폴리펩타이드와 상이할 수 있다. 이러한 변이체는 천연 유래일 수 있거나 또는, 예를 들어, 상기 폴리펩타이드 서열의 하나 이상의 아미노산을 변형시킴으로써 합성에 의해 생성될 수 있다. 예를 들어, 특정 실시형태에서, 폴리펩타이드 내로 하나 이상의 치환, 결실, 첨가 및/또는 삽입을 도입함으로써 인간 TCR  $\alpha$  유전자 내 표적 부위에 결합하고, 이를 절단하는 호밍 엔도뉴클레아제, 메가TAL 등의 생물학적 특성을 개선시키는 것이 바람직할 수 있다. 특정 실시형태에서, 폴리펩타이드는 본 명세서에 상정된 임의의 기준 서열에 대해 적어도 약 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% 또는 99%의 아미노산 서열을 갖는 폴리펩타이드를 포함하며, 전형적으로 여기서 변이체는 기준 서열의 적어도 하나의 생물학적 활성을 유지한다.

[0183] 폴리펩타이드 변이체는 생물학적으로 활성인 "폴리펩타이드 단편"을 포함한다. 생물학적으로 활성인 폴리펩타이드 단편의 예시적 예는 DNA 결합 도메인, 뉴클레아제 도메인 등을 포함한다. 본 명세서에서 사용되는 용어 "생물학적으로 활성인 단편" 또는 "최소 생물학적으로 활성인 단편"은 천연 유래 폴리펩타이드 활성의 적어도 100%, 적어도 90%, 적어도 80%, 적어도 70%, 적어도 60%, 적어도 50%, 적어도 40%, 적어도 30%, 적어도 20%, 적어도 10% 또는 적어도 5%를 보유하는 폴리펩타이드 단편을 지칭한다. 바람직한 실시형태에서, 생물학적 활성은 표적 서열에 대한 결합 친화도 및/또는 절단 활성이다. 특정 실시형태에서, 폴리펩타이드 단편은 적어도 5 내지 약 1700개의 아미노산 길이의 아미노산 쇄를 포함할 수 있다. 특정 실시형태에서, 단편은 적어도 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700개 이상의 아미노산 길이라는 것이 인식될 것이다. 특정 실시형태에서, 폴리펩타이드는 호밍 엔도뉴클레아제 변이체의 생물학적으로 활성인 단편을 포함한다. 특정 실시형태에서, 본 명세서에 제시된 폴리펩타이드는 "X"로서 나타내는 하나 이상의 아미노산을 포함할 수 있다. "X"가 아미노산 서열 번호에 존재한다면, 임의의 아미노산을 지칭한다. 하나 이상의 "X" 잔기는 본 명세서에 상정된 특정 서열번호에 제시된 아미노산 서열의 N- 및 C-말단에 존재할 수 있다. "X" 아미노산이 존재하지 않는다면, 서열번호에 제시된 남아있는 아미노산 서열은 생물학적으로 활성인 단편이 고려될 수 있다.

[0184] 특정 실시형태에서, 폴리펩타이드는 호밍 엔도뉴클레아제 변이체의 생물학적으로 활성인 단편, 예를 들어, 서열 번호 6 내지 7, 또는 메가TAL(서열번호 10 내지 12)을 포함한다. 생물학적으로 활성인 단편은 N-말단의 절단 및/또는 C-말단의 절단을 포함할 수 있다. 특정 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 호밍 엔도뉴클레아제 서열에 비해 호밍 엔도뉴클레아제 변이체의 1, 2, 3, 4, 5, 6, 7 또는 8개의 N-말단 아미노산의 결실, 더 바람직하게는 대응하는 야생형 호밍 엔도뉴클레아제 서열에 비해 호밍 엔도뉴클레아제 변이체의 4개의 N-말단 아미노산의 결실을 결여하거나 또는 포함한다. 특정 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 호밍 엔도뉴클레아제 서열에 비해 호밍 엔도뉴클레아제 변이체의 1, 2, 3, 4 또는 5개의 C-말단 아미노산의 결실, 더 바람직하게는 대응하는 야생형 호밍 엔도뉴클레아제 서열에 비해 호밍 엔도뉴클레아제 변이체의 2개의 C-말단 아미노산의 결실을 결여하거나 또는 포함한다. 특정 바람직한 실시형태에서, 생물학적으로 활성인 단편은 대응하는 야생형 호밍 엔도뉴클레아제 서열에 비해 호밍 엔도뉴클레아제 변이체의 4개의 N-말단 아미노산 및 2개의 C-말단 아미노산의 결실을 결여하거나 또는 포함한다.

[0185] 특정 실시형태에서, I-OnuI 변이체는 다음의 N-말단 아미노산: M, A, Y, M, S, R, R, E의 1, 2, 3, 4, 5, 6, 7 또는 8개의 결실; 및/또는 다음의 1, 2, 3, 4 또는 5개의 C-말단 아미노산: R, G, S, F, V의 결실을 포함한다.

[0186] 특정 실시형태에서, I-OnuI 변이체는 다음의 N-말단 아미노산: M, A, Y, M, S, R, R, E의 1, 2, 3, 4, 5, 6, 7 또는 8개의 결실 또는 치환; 및/또는 다음의 1, 2, 3, 4 또는 5개의 C-말단 아미노산: R, G, S, F, V의 결실 또

는 치환을 포함한다.

- [0187] 특정 실시형태에서, I-OnuI 변이체는 다음의 N-말단 아미노산: M, A, Y, M, S, R, R, E의 1, 2, 3, 4, 5, 6, 7 또는 8개의 결실; 및/또는 다음의 1 또는 2개의 C-말단의 아미노산: F, V의 결실을 포함한다.
- [0188] 특정 실시형태에서, I-OnuI 변이체는 다음의 N-말단 아미노산: M, A, Y, M, S, R, R, E의 1, 2, 3, 4, 5, 6, 7 또는 8개의 결실 또는 치환; 및/또는 다음의 1 또는 2개의 C-말단의 아미노산: F, V의 결실 또는 치환을 포함한다.
- [0189] 상기 언급한 바와 같이, 폴리펩타이드는 아미노산 치환, 결실, 절단 및 삽입을 포함하는 다양한 방법으로 변경될 수 있다. 이러한 조작을 위한 방법은 일반적으로 당업계에 공지되어 있다. 예를 들어, 기존 폴리펩타이드의 아미노산 서열 변이체는 DNA에서의 돌연변이에 의해 제조될 수 있다. 돌연변이유발 및 뉴클레오타이드 서열 변경을 위한 방법은 당업계에 잘 공지되어 있다. 예를 들어, 문헌[Kunkel (1985, *Proc. Natl. Acad. Sci. USA*, 82: 488-492), Kunkel *et al.*, (1987, *Methods in Enzymol.*, 154: 367-382)], 미국 특허 제4,873,192호, 문헌[Watson, J. D. *et al.*, (*Molecular Biology of the Gene*, Fourth Edition, Benjamin/Cummings, Menlo Park, Calif., 1987)] 및 그에 인용된 참고문헌을 참조. 관심 대상의 단백질의 생물학적 활성에 영향을 미치지 않는 적절한 아미노산 치환에 대한 가이드는 문헌[Dayhoff *et al.*, (1978) *Atlas of Protein Sequence and Structure* (*Natl. Biomed. Res. Found.*, Washington, D.C.)]의 모델에서 찾을 수 있다.
- [0190] 특정 실시형태에서, 변이체는 하나 이상의 보존적 치환을 함유할 것이다. "보존적 치환"은 펩타이드 화학 분야의 당업자가 폴리펩타이드의 2차 구조 및 소수성 특성이 실질적으로 변하지 않는 것으로 예상하도록 아미노산이 유사한 특성을 갖는 다른 아미노산으로 치환되는 것이다. 변형은 특정 실시형태에서 상정된 폴리뉴클레오타이드 및 폴리펩타이드의 구조에서 만들어질 수 있으며, 폴리펩타이드는 적어도 폴리펩타이드를 포함하고, 바람직한 특징을 갖는 변이체 또는 유도체 폴리펩타이드를 암호화하는 작용성 분자를 여전히 얻는다. 동등한, 또는 심지어 개선된 변이체 폴리펩타이드를 생성하기 위해 폴리펩타이드의 아미노산 서열을 변경시키도록 요망될 때, 당업자는, 예를 들어, 표 1에 따라, 예를 들어 암호화 DNA 서열의 코돈 중 하나 이상을 변화시킬 수 있다.

[0191] [표 1]

# 아미노산 코돈

아미노산	1글자 암호	3글자 암호	코돈				
알라닌	A	Ala	GCA	GCC	GCG	GCU	
시스테인	C	Cys	UGC	UGU			
아스파르트산	D	Asp	GAC	GAU			
글루탐산	E	Glu	GAA	GAG			
페닐알라닌	F	Phe	UUC	UUU			
글리신	G	Gly	GGA	GGC	GGG	GGU	
히스티딘	H	His	CAC	CAU			
아이소류신	I	Iso	AUA	AUC	AUU		
라이신	K	Lys	AAA	AAG			
류신	L	Leu	UUA	UUG	CUA	CUC	CUG
메티오닌	M	Met	AUG				
아스파라긴	N	Asn	AAC	AAU			
프롤린	P	Pro	CCA	CCC	CCG	CCU	
글루타민	Q	Gln	CAA	CAG			
알기닌	R	Arg	AGA	AGG	CGA	CGC	CGG
세린	S	Ser	AGC	AGU	UCA	UCC	UCG
트레오닌	T	Thr	ACA	ACC	ACG	ACU	
발린	V	Val	GUA	GUC	GUG	GUU	
트립토판	W	Trp	UGG				
타이로신	Y	Tyr	UAC	UAU			

[0192]

[0193] 생물학적 활성을 없애는 일 없이 아미노산 잔기가 치환, 삽입 또는 결실되는 것을 결정하는 가이드는 당업계에 잘 공지된 컴퓨터 프로그램, 예컨대 DNASTAR, DNA Strider, Geneious, Mac Vector, 또는 Vector NTI 소프트웨어를 이용하여 찾을 수 있다. 바람직하게는, 본 명세서에 개시된 단백질 변이체에서 아미노산 변화는 보존적 아미노산 변화, 즉, 유사하게 하전된 또는 비하전된 아미노산의 치환이다. 보존적 아미노산 변화는 그들의 측쇄와 관련된 아미노산 패밀리 중 하나의 치환을 수반한다. 천연 유래 아미노산은 일반적으로 4가지 패밀리로 분류된다: 산성(아스파르트레이트, 글루타메이트), 염기성(라이신, 알기닌, 히스티딘), 비극성(알라닌, 발린, 류신, 아이소류신, 프롤린, 페닐알라닌, 메티오닌, 트립토판) 및 비하전 극성(글리신, 아스파라긴, 글루타민, 시스테인, 세린, 트레오닌, 타이로신) 아미노산. 페닐알라닌, 트립토판 및 타이로신은 때때로 공통적으로 방향족 아미노산으로서 분류된다. 펩타이드 또는 단백질에서, 아미노산의 적합한 보존적 치환은 당업자에게 공지되어 있으며, 일반적으로 얻어진 분자의 생물학적 활성을 변경시키는 일 없이 이루어질 수 있다. 당업자라면 일반적으로, 폴리펩타이드의 비필수 영역 내 단일 아미노산 치환이 생물학적 활성을 실질적으로 변경시키지 않는다는 것을 인식한다(예를 들어, 문헌[Watson *et al.* *Molecular Biology of the Gene*, 4th Edition, 1987, The Benjamin/Cummings Pub. Co., p.224] 참조).

[0194] 일 실시형태에서, 2 이상의 폴리펩타이드의 발현이 요망되는 경우, 그들을 암호화하는 폴리뉴클레오타이드 서열은 본 명세서의 다른 곳에서와 같이 IRES에 의해 분리될 수 있다.

[0195] 특정 실시형태에서 상정된 폴리펩타이드는 융합 폴리펩타이드를 포함한다. 특정 실시형태에서, 융합 폴리펩타이드 및 융합 폴리펩타이드를 암호화하는 폴리뉴클레오타이드가 제공된다. 융합 폴리펩타이드 및 융합 단백질은 적어도 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 폴리펩타이드 세그먼트를 갖는 폴리펩타이드를 지칭한다.

[0196] 다른 실시형태에서, 2 이상의 폴리펩타이드는 본 명세서의 다른 곳에 개시된 바와 같은 하나 이상의 자기-절단

성 폴리펩타이드 서열을 포함하는 융합 단백질로서 발현될 수 있다.

- [0197] 일 실시형태에서, 본 명세서에 상정된 융합 단백질은 하나 이상의 DNA 결합 도메인 및 하나 이상의 뉴클레아제, 및 하나 이상의 링커 및/또는 자기-절단성 폴리펩타이드를 포함한다.
- [0198] 일 실시형태에서, 본 명세서에 상정된 융합 단백질은 뉴클레아제 변이체; 링커 또는 자기-절단성 펩타이드; 및 5'-3' 엑소뉴클레아제, 5'-3' 알칼리성 엑소뉴클레아제, 및 3'-5' 엑소뉴클레아제(예를 들어, Trex2)를 포함하지만, 이들로 제한되지 않는, 말단-가공 효소를 포함한다.
- [0199] 융합 폴리펩타이드는 신호 펩타이드, 세포 침투성 펩타이드 도메인(CPP), DNA 결합 도메인, 뉴클레아제 도메인 등, 에피토프 태그(예를 들어, 말토스 결합 단백질("MBP"), 글루타티온 S 트랜스퍼라제(GST), HIS6, MYC, FLAG, V5, VSV-G 및 HA), 폴리펩타이드 링커 및 폴리펩타이드 절단 신호를 포함하지만, 이들로 제한되지 않는 하나 이상의 폴리펩타이드 도메인 또는 세그먼트를 포함할 수 있다. 융합 폴리펩타이드는, 그들이 또한 C-말단에서 C-말단으로, N-말단에서 N-말단으로, 또는 N-말단에서 C-말단으로 연결될 수 있지만, 전형적으로 C-말단에서 N-말단으로 연결된다. 특정 실시형태에서, 융합 단백질의 폴리펩타이드는 임의의 순서일 수 있다. 융합 폴리펩타이드 또는 융합 단백질은 또한, 융합 폴리펩타이드의 목적으로 하는 활성이 보존되는 한, 보존적으로 변형된 변이체, 다형성 변이체, 대립유전자, 돌연변이체, 하위서열 및 중간 상동체를 포함할 수 있다. 융합 폴리펩타이드는 화학적 합성 방법에 의해 또는 두 모이어티 사이의 화학적 결합에 의해 생산될 수 있거나 또는 일반적으로 다른 표준 기법을 이용하여 제조될 수 있다. 융합 폴리펩타이드를 포함하는 절찰된 DNA 서열은 본 명세서의 다른 곳에 개시된 바와 같은 적합한 전사 또는 번역 제어 요소에 작동 가능하게 연결된다.
- [0200] 융합 폴리펩타이드는 하나 이상의 폴리펩타이드 또는 폴리펩타이드 내의 도메인을 연결하기 위해 사용될 수 있는 링커를 선택적으로 포함할 수 있다. 펩타이드 링커 서열은 폴리펩타이드 도메인이 그들의 목적으로 하는 기능을 발휘하도록 각각의 폴리펩타이드가 그의 적절한 2차 및 3차 구조로 폴딩되는 것을 보장하는 데 충분한 거리만큼 임의의 2개 이상의 폴리펩타이드 성분을 분리시키는 데 사용될 수 있다. 이러한 펩타이드 링커 서열은 당업계의 표준 기법을 이용하여 융합 폴리펩타이드 내로 혼입된다. 적합한 펩타이드 링커 서열은 다음의 인자에 기반하여 선택될 수 있다: (1) 가요성의 연장된 입체배좌를 채택하는 그들의 능력; (2) 제1 및 제2 폴리펩타이드 상의 작용성 에피토프와 상호작용할 수 있는 2차 구조를 채택하는 그들의 불능; 및 (3) 폴리펩타이드 작용성 에피토프와 반응하는 소수성 또는 하전된 잔기의 결여. 바람직한 펩타이드 링커 서열은 Gly, Asn 및 Ser 잔기를 함유한다. 다른 근처의 중성 아미노산, 예컨대 Thr 및 Ala은 또한 링커 서열에서 사용될 수 있다. 링커로서 유용하게 사용될 수 있는 아미노산 서열은 문헌[Maratea *et al.*, *Gene* 40:39-46, 1985; Murphy *et al.*, *Proc. Natl. Acad. Sci. USA* 83:8258-8262, 1986]; 미국 특허 제4,935,233호 및 미국 특허 제4,751,180호에 개시된 것을 포함한다. 특정 융합 폴리펩타이드 세그먼트가 작용성 도메인을 분리시키고 입체 방해를 방지하기 위해 사용될 수 있는 비필수 N-말단 아미노산 영역을 함유할 때 링커 서열은 필요하지 않다. 바람직한 링커는 재조합 융합 단백질의 부분으로서 합성되는 전형적으로 가요성의 아미노산 하위서열이다. 링커 폴리펩타이드는 길이가 1 내지 200개의 아미노산, 길이가 1 내지 100개의 아미노산, 또는 길이가 1 내지 50개의 아미노산(그 사이의 모든 정수를 포함)일 수 있다.
- [0201] 예시적인 링커는 다음의 아미노산 서열을 포함하지만, 이들로 제한되지 않는다: 글리신 중합체( $(G)_n$ ); 글리신-세린 중합체( $(G_1S_1)_n$ )(여기서,  $n$ 은 적어도 1, 2, 3, 4 또는 5의 정수임); 글리신-알라닌 중합체; 알라닌-세린 중합체; GGG(서열번호 30); DGGGS(서열번호 31); TGEKP(서열번호 32)(예를 들어, 문헌[Liu *et al.*, *PNAS* 55:525-5530(1997)] 참조); GRRR(서열번호 33)(Pomerantz *et al.* 1995, 상기 참조);  $(GGGS)_n$ (여기서,  $n = 1, 2, 3, 4$  또는 5임)(서열번호 34)(Kim *et al.*, *PNAS* 93, 1156-1160(1996.)); EGKSSGSGSESKVD(서열번호 35)(Chaudhary *et al.*, 1990, *Proc. Natl. Acad. Sci. U.S.A.* 87:1066-1070); KESGVSSEQLAQFRSLD(서열번호 36)(Bird *et al.*, 1988, *Science* 242:423-426), GRRGGGS(서열번호 37); LRQDGERP(서열번호 38); LRQDGGGSERP(서열번호 39); LRQD(GGG)<sub>2</sub>ERP(서열번호 40). 대안적으로, 가요성 링커는 DNA-결합 부위와 펩타이드 그 자체를 둘 다 모델링할 수 있는 컴퓨터 프로그램을 이용하여(문헌[Desjarlais & Berg, *PNAS* 90:2256-2260(1993), *PNAS* 91:11099-11103 (1994)]) 또는 파지 디스플레이 방법에 의해 합리적으로 설계될 수 있다.
- [0202] 융합 폴리펩타이드는 본 명세서에 기재된 각각의 폴리펩타이드 사이에 또는 내인성 오픈 리딩 프레임과 공여자 수선 주형에 의해 암호화된 폴리펩타이드 사이에 폴리펩타이드 절단 신호를 추가로 포함할 수 있다. 추가로, 폴리펩타이드 절단 부위는 임의의 링커 펩타이드 서열 내로 들어갈 수 있다. 예시적인 폴리펩타이드 절단 신호는 폴리펩타이드 절단 인식 부위, 예컨대 프로테아제 절단 부위, 뉴클레아제 절단 부위(예를 들어, 희귀 제한 효소



인식 부위, 자기-절단성 리보자임 인식 부위) 및 자기-절단성 바이러스 올리고펩타이드를 포함한다(문헌 [deFelipe and Ryan, 2004. *Traffic*, 5(8); 616-26] 참조).

- [0203] 적합한 프로테아제 절단 부위 및 자기-절단성 펩타이드는 당업자에게 공지되어 있다(예를 들어, 문헌[Ryan *et al.*, 1997. *J. Gener. Virol.* 78, 699-722; Scymczak *et al.* (2004) *Nature Biotech.* 5, 589-594] 참조). 예시적인 프로테아제 절단 부위는 포티바이러스 NIa 프로테아제(예를 들어, 담배 식각 바이러스), 포티바이러스 HC 프로테아제, 포티바이러스 P1(P35) 프로테아제, 비오바이러스 NIa 프로테아제, 비오바이러스 RNA-2-암호화된 프로테아제, 아프타바이러스 L 프로테아제, 엔테로바이러스 2A 프로테아제, 리노바이러스 2A 프로테아제, 피코르나 3C 프로테아제, 코모바이러스 24K 프로테아제, 네포바이러스 24K 프로테아제, RTSV(벼 통그로 구형 바이러스(rice tungro spherical virus)) 3C-유사 프로테아제, PYVF(파스닙 황색 얼룩 바이러스(parsnip yellow fleck virus)) 3C-유사 프로테아제, 헤파린, 트롬빈, 인자 Xa 및 엔테로키나제의 절단 부위를 포함하지만, 이들로 제한되지 않는다. 그의 고절단 엄격성에 기인하여, TEV(담배 식각 바이러스) 프로테아제 절단 부위는 일 실시형태, 예를 들어, EXXYXQ(G/S)(서열번호 41), 예를 들어, ENLYFQG(서열번호 42) 및 ENLYFQS(서열번호 43)에서 바람직하되, X는 임의의 아미노산을 나타낸다(TEV에 의한 절단은 Q와 G 또는 Q와 S 사이에서 생김).
- [0204] 특정 실시형태에서, 자기-절단성 폴리펩타이드 부위는 2A 또는 2A-유사 부위, 서열 또는 도메인을 포함한다(Donnelly *et al.*, 2001. *J. Gen. Virol.* 82:1027-1041). 특정 실시형태에서, 바이러스 2A 펩타이드는 아프타바이러스 2A 펩타이드, 포티바이러스 2A 펩타이드, 또는 카디오바이러스 2A 펩타이드이다.
- [0205] 일 실시형태에서, 바이러스 2A 펩타이드는 구제역바이러스(FMDV) 2A 펩타이드, 마 비염 A 바이러스(ERAV) 2A 펩타이드, 토세아 아시그나 바이러스(Thosea asigna virus: TaV) 2A 펩타이드, 돼지 테스코바이러스-1(PTV-1) 2A 펩타이드, 테일로바이러스 2A 펩타이드 및 뇌심근염바이러스 2A 펩타이드로 이루어진 군으로부터 선택된다.
- [0206] 2A 부위의 예시적인 예를 표 2에 제공한다.

[0207] [표 2]

예시적인 2A 부위는 다음의 서열을 포함한다:

서열번호 44	GSGATNFSLLKQAGDVEENPGP
서열번호 45	ATNFSLLKQAGDVEENPGP
서열번호 46	LLKQAGDVEENPGP
서열번호 47	GSGEGRGSLTTCGDVEENPGP
서열번호 48	EGRGSLTTCGDVEENPGP
서열번호 49	LLTCGDVEENPGP
서열번호 50	GSGQCTNYALLKLAGDVESNPGP
서열번호 51	QCTNYALLKLAGDVESNPGP
서열번호 52	LLKLAGDVESNPGP
서열번호 53	GSGVKQTLNFDLLKLAGDVESNPGP
서열번호 54	VKQTLNFDLLKLAGDVESNPGP
서열번호 55	LLKLAGDVESNPGP
서열번호 56	LLNFDLLKLAGDVESNPGP
서열번호 57	TLNFDLLKLAGDVESNPGP
서열번호 58	LLKLAGDVESNPGP
서열번호 59	NFDLLKLAGDVESNPGP
서열번호 60	QLNFDLLKLAGDVESNPGP
서열번호 61	APVKQTLNFDLLKLAGDVESNPGP
서열번호 62	VTELLYRMKRAETCYCPRLAIHPTEARHKQKIVAPVKQT
서열번호 63	LNFDLLKLAGDVESNPGP
서열번호 64	LLAIHPTEARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP
서열번호 65	EARHKQKIVAPVKQTLNFDLLKLAGDVESNPGP

[0208]

[0209]

#### E. 폴리뉴클레오타이드

[0210]

특정 실시형태에서, 본 명세서에 상정된 하나 이상의 호밍 엔도뉴클레아제 변이체, 메가TAL, 말단-가공 효소 및 융합 폴리펩타이드를 암호화하는 폴리뉴클레오타이드가 제공된다. 본 명세서에서 사용되는 용어 "폴리뉴클레오타이드" 또는 "핵산"은 데옥시리보핵산(DNA), 리보핵산(RNA) 및 DNA/RNA 혼성체를 지칭한다. 폴리뉴클레오타이드는 단일 가닥 또는 이중 가닥이고 재조합, 합성이거나, 또는 단리될 수 있다. 폴리뉴클레오타이드는 전-전령 RNA(전-mRNA), 전령 RNA(mRNA), RNA, 짧은 간섭 RNA(siRNA), 짧은 헤어핀 RNA(shRNA), 마이크로RNA(miRNA), 리보자임, 게놈 RNA(gRNA), 플러스 가닥 RNA(RNA(+)), 마이너스 가닥 RNA(RNA(-)), tracrRNA, crRNA, 단일 가이드 RNA(sgRNA), 합성 RNA, 합성 mRNA, 게놈 DNA(gDNA), PCR 증폭된 DNA, 상보성 DNA(cDNA), 합성 DNA 또는 재조합 DNA를 포함하지만, 이들로 제한되지 않는다. 폴리뉴클레오타이드는 리보뉴클레오타이드 또는 데옥시뉴클레오타이드 또는 뉴클레오타이드 유형의 변형된 형태의 길이가 적어도 5, 적어도 10, 적어도 15, 적어도 20, 적어도 25, 적어도 30, 적어도 40, 적어도 50, 적어도 100, 적어도 200, 적어도 300, 적어도 400, 적어도 500, 적어도 1000, 적어도 5000, 적어도 10000, 또는 적어도 15000개 이상뿐만 아니라 모든 중간 길이인 뉴클레오타이드의 중합체 형태를 지칭한다. 이와 관련하여 "중간 길이"는 인용된 값 사이의 임의의 길이, 예컨대 6, 7, 8, 9 등, 101, 102, 103 등; 151, 152, 153 등; 201, 202, 203 등을 의미한다. 특정 실시형태에서, 폴리뉴클레오타이드 또는 변이체는 기준 서열에 대해 적어도 또는 약 50%, 55%, 60%, 65%, 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 가진다.

[0211]

특정 실시형태에서, 폴리뉴클레오타이드는 코돈-최적화될 수 있다. 본 명세서에서 사용되는 용어 "코돈-최적화

된"은 폴리펩타이드의 발현, 안정성 및/또는 활성을 증가시키기 위해 폴리펩타이드를 암호화하는 폴리뉴클레오타이드 내 코돈을 치환하는 것을 지칭한다. 코돈 최적화에 영향을 미치는 인자는 (i) 둘 이상의 유기체 사이의 코돈 편향의 변형 또는 유전자 또는 합성으로 작제된 편향 표, (ii) 유기체, 유전자 또는 유전자 세트 내에서 코돈 편향 정도의 변형, (iii) 맥락을 포함하는 코돈의 체계적 변형, (iv) 그들의 탈암호 tRNA에 따른 코돈의 변형, (v) 전반적으로 또는 삼중 중 하나의 위치에서 GC%에 따른 코돈의 변형, (vi) 기준 서열, 예를 들어 천연 유래 서열과의 유사성 정도의 변형, (vii) 코돈 빈도 컷오프에서의 변형, (viii) DNA 서열로부터 전사된 mRNA의 구조적 특성, (ix) 코돈 치환 세트의 설계에 기반한 DNA 서열의 기능에 관한 사전 지식, (x) 각각의 아미노산에 대한 코돈 세트의 체계적 변형, 및/또는 (xi) 비논리적 번역 개시 부위의 단리된 제거를 포함하지만, 이들로 제한되지 않는다.

[0212] 본 명세서에서 사용되는 용어 "뉴클레오타이드"는 포스포릴화된 당과의 N-글리코사이드 결합에서의 복소환식 질소 염기를 지칭한다. 뉴클레오타이드는 천연 염기 및 매우 다양한 기술-인식된 변형 염기를 포함하는 것으로 이해된다. 이러한 염기는 일반적으로 뉴클레오타이드 당 모이어티의 1' 위치에 위치된다. 뉴클레오타이드는 일반적으로 염기, 당 및 인산염기를 포함한다. 리보핵산(RNA)에서, 당은 리보스이며, 데옥시리보핵산(DNA)에서, 당은 데옥시리보스, 즉, 리보스에 존재하는 하이드록실기가 없는 당이다. 예시적인 천연 질소 염기는 퓨린, 아데노신(A) 및 구아니딘(G), 및 피리미딘, 사이티딘(C) 및 티미딘(T)(또는 RNA와 관련하여, 유라실(U))을 포함한다. 데옥시리보스의 C-1 원자는 피리미딘의 N-1 또는 퓨린의 N-9에 결합된다. 뉴클레오타이드는 보통 일, 이- 또는 삼인산염이다. 뉴클레오타이드는 당, 인산염 및/또는 염기 모이어티에서 비변형 또는 변형될 수 있다(또한 뉴클레오타이드 유사체, 뉴클레오타이드 유도체, 변형된 뉴클레오타이드, 비천연 뉴클레오타이드 및 비표준 뉴클레오타이드와 상호 호환적으로 지칭됨; 예를 들어, WO 92/07065 및 WO 93/15187). 변형된 핵산 염기의 예는 문헌[Limbach *et al.*, (1994, *Nucleic Acids Res.* 22, 2183-2196)]에 의해 요약된다.

[0213] 뉴클레오타이드는 또한 당의 C-5에 부착된 하이드록실기 상에서 생기는 에스터화에 의한 뉴클레오사이드의 인산염 에스터로서 간주될 수 있다. 본 명세서에서 사용되는 용어 "뉴클레오사이드"는 당과의 N-글리코사이드 결합에서의 복소환식 질소 염기를 지칭한다. 뉴클레오사이드는 천연 염기를 포함하고 잘 공지된 변형된 염기를 포함하는 것으로 당업계에서 인식된다. 이러한 염기는 일반적으로 뉴클레오사이드 당 모이어티의 1' 위치에 위치된다. 뉴클레오사이드는 일반적으로 염기 및 당기를 포함한다. 뉴클레오사이드는 비변형이거나 또는 당 및/또는 염기 모이어티에서 변형될 수 있다(또한 뉴클레오사이드 유사체, 뉴클레오사이드 유도체, 변형된 뉴클레오사이드, 비천연 뉴클레오사이드 또는 비표준 뉴클레오사이드로서 상호 호환적으로 지칭됨). 또한 상기 언급한 바와 같이, 변형된 핵산 염기의 예는 문헌[Limbach *et al.*, (1994, *Nucleic Acids Res.* 22, 2183-2196)]에 의해 요약된다.

[0214] 폴리뉴클레오타이드의 예시적 예는 서열번호 1 내지 12 및 22 내지 24를 암호화하는 폴리뉴클레오타이드, 및 서열번호 13 내지 16 및 25 내지 27에 제시된 폴리뉴클레오타이드 서열을 포함하지만, 이들로 제한되지 않는다.

[0215] 다양한 예시적 실시형태에서, 본 명세서에 상정된 폴리뉴클레오타이드는 호밍 엔도뉴클레아제 변이체, 메가TAL, 말단-가공 효소, 융합 폴리펩타이드, 및 발현 벡터, 바이러스 벡터, 및 본 명세서에 상정된 폴리뉴클레오타이드를 포함하는 전달 플라스미드를 암호화하는 폴리뉴클레오타이드를 포함하지만, 이들로 제한되지 않는다.

[0216] 본 명세서에서 사용되는 용어 "폴리뉴클레오타이드 변이체" 및 "변이체" 등은 기준 폴리뉴클레오타이드 서열과 실질적인 서열 동일성을 나타내는 폴리뉴클레오타이드 또는 본 명세서에서 이후에 나타내는 엄격한 조건 하에 기준 서열과 혼성화되는 폴리뉴클레오타이드를 지칭한다. 이들 용어는 또한 적어도 하나의 뉴클레오타이드의 첨가, 결실, 치환 또는 변형에 의해 기준 폴리뉴클레오타이드와 구별되는 폴리뉴클레오타이드를 포함한다. 따라서, 용어 "폴리뉴클레오타이드 변이체" 및 "변이체"는 하나 이상의 뉴클레오타이드가 첨가되거나 또는 결실되거나, 또는 변형되거나 또는 상이한 뉴클레오타이드로 대체된 폴리뉴클레오타이드를 포함한다. 이와 관련하여, 돌연변이, 첨가, 결실 및 치환을 포함하는 특정 변경은 기준 폴리뉴클레오타이드에 대해 이루어질 수 있고, 이에 의해 변경된 폴리뉴클레오타이드는 기준 폴리뉴클레오타이드의 생물학적 기능 또는 활성을 보유한다는 것이 잘 이해된다.

[0217] 일 실시형태에서, 폴리뉴클레오타이드는 엄격한 조건 하에 표적 핵산 서열에 혼성화하는 뉴클레오타이드 서열을 포함한다. "엄격한 조건" 하에 혼성화하는 것은 서로에 대해 적어도 60% 동일한 뉴클레오타이드 서열이 혼성화된 채로 남아있는 혼성화 프로토콜을 기재한다. 일반적으로, 엄격한 조건은 정해진 이온 강도 및 pH에서 특정 서열에 대해 열 용점(T<sub>m</sub>)보다 약 5°C 더 낮아지도록 선택된다. T<sub>m</sub>은 (정해진 이온 강도, pH 및 핵산 농도 하에) 표적 서열에 대해 상보성인 프로브의 50%가 평형 상태에서 표적 서열에 혼성화하는 온도이다. 표적 서열은 일반

적으로 과량으로 존재하기 때문에, T<sub>m</sub>에서, 프로브의 50%는 평형상태에서 점유된다.

- [0218] 설명 "서열 동일성" 또는, 예를 들어, 본 명세서에서 사용되는 바와 같이 "에 대해 50% 동일한 서열"을 포함하는 것은 서열이 비교창에 걸쳐 뉴클레오타이드 대 뉴클레오타이드 염기 기준 또는 아미노산 대 아미노산 기준과 동일한 정도를 지칭한다. 따라서, "서열 동일성의 백분율"은 비교창에 걸쳐 2개의 최적으로 정렬된 서열을 비교하는 단계, 매칭된 위치의 수를 수득하기 위해 동일한 핵산 염기(예를 들어, A, T, C, G, I) 또는 동일한 아미노산 잔기(예를 들어, Ala, Pro, Ser, Thr, Gly, Val, Leu, Ile, Phe, Tyr, Trp, Lys, Arg, His, Asp, Glu, Asn, Gln, Cys 및 Met)가 서열 둘 다에서 생기는 위치 수를 결정하는 단계, 매칭된 위치의 수를 비교창에서의 위치의 총수(즉, 창 크기)로 나누는 단계, 및 결과에 100을 곱하여서 서열 동일성의 백분율을 수득하는 단계에 의해 계산될 수 있다. 본 명세서에 기재된 임의의 기준 서열에 대해 적어도 약 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 또는 100% 서열 동일성을 갖는 뉴클레오타이드 및 폴리펩타이드가 포함되며, 전형적으로 폴리펩타이드 변이체는 기준 폴리펩타이드의 적어도 하나의 생물학적 활성을 유지한다.
- [0219] 2 이상의 폴리뉴클레오타이드 또는 폴리펩타이드 사이의 서열 관계를 기재하기 위해 사용되는 용어는 "기준 서열", "비교창", "서열 동일성", "서열 동일성의 백분율" 및 "실질적인 동일성"을 포함한다. "기준 서열"은 길이가 적어도 12개이지만, 빈번하게는 15 내지 18개 및 종종 적어도 25개의 단량체 단위(뉴클레오타이드 및 아미노산 잔기를 포함)이다. 두 폴리뉴클레오타이드는 (1) 두 폴리뉴클레오타이드 사이의 유사한 서열(즉, 완전한 폴리뉴클레오타이드 서열의 일부만), 및 (2) 두 폴리뉴클레오타이드 간에 다른 서열을 각각 포함할 수 있기 때문에, 2(이상의) 폴리뉴클레오타이드 사이의 서열 비교는 전형적으로 서열 유사성의 국소 영역을 확인하고 비교하기 위해 "비교창"에 걸쳐 두 폴리뉴클레오타이드의 서열을 비교함으로써 수행된다. "비교창"은 적어도 6개의 인접한 위치, 보통 약 50 내지 약 100, 더 보통으로는 약 100 내지 약 150개의 개념적 세그먼트를 지칭하며, 이때 서열은 두 서열이 최적으로 정렬된 후에 동일한 수의 인접한 위치의 기준 서열과 비교된다. 비교창은 두 서열의 최적의 정렬을 위해 기준 서열(첨가 또는 결실을 포함하지 않음)에 비해 약 20% 이하의 첨가 또는 결실(즉, 갭)을 포함할 수 있다. 비교창을 정렬하기 위한 서열의 최적의 정렬은 알고리즘의 컴퓨터화된 실행(위스콘신 제네틱스 소프트웨어 패키지 릴리즈 7.0에서 GAP, BESTFIT, FASTA 및 TFASTA, 미국 위스콘신주 메디슨 사이언스 드라이브 575에 소재한 제네틱스 컴퓨터 그룹(Genetics Computer Group))에 의해 또는 검사 및 선택된 임의의 다양한 방법에 의해 생성된 최적의 정렬(즉, 비교창에 걸쳐 가장 큰 백분율의 상동성을 초래)에 의해 수행될 수 있다. 또한, 예를 들어 문헌[Altschul *et al.*, 1997, *Nucl. Acids Res.* 25:3389]에 의해 개시되는 바와 같이 프로그램의 BLAST 패밀리가 언급될 수 있다. 서열 분석의 상세한 논의는 문헌[Unit 19.3 of Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley & Sons Inc., 1994-1998, Chapter 15]에서 찾을 수 있다.
- [0220] 본 명세서에서 사용되는 "단리된 폴리뉴클레오타이드"는 천연-유래 상태, 예를 들어, 단편에 정상적으로 인접한 서열로부터 제거된 DNA 단편에서 그것에 축적하는 서열로부터 정제된 폴리뉴클레오타이드를 지칭한다. 특정 실시형태에서, "단리된 폴리뉴클레오타이드"는 상보성 DNA(cDNA), 재조합 폴리뉴클레오타이드, 합성 폴리뉴클레오타이드 또는 천연에 존재하지 않고 사람의 손으로 만들어지는 다른 폴리뉴클레오타이드를 지칭한다.
- [0221] 다양한 실시형태에서, 폴리뉴클레오타이드는 호밍 엔도뉴클레아제 변이체, 메가TAL 및 말단-가공 효소를 포함하지만, 이들로 제한되지 않는 본 명세서에 상정된 폴리펩타이드를 암호화하는 mRNA를 포함한다. 특정 실시형태에서, mRNA는 캡, 하나 이상의 뉴클레오타이드 및 폴리(A) 꼬리를 포함한다.
- [0222] 본 명세서에서 사용되는 용어 "5' 캡" 또는 "5' 캡 구조" 또는 "5' 캡 모이어티"는 mRNA의 5' 말단에 혼입된 화학적 변형을 지칭한다. 5' 캡은 핵 유출, mRNA 안정성 및 번역에 연루된다.
- [0223] 특정 실시형태에서, 본 명세서에 상정된 mRNA는 말단의 구아노신 캡 잔기와 mRNA 분자의 5'-말단의 전사된 센스 뉴클레오타이드 사이의 5'-ppp-5'-삼인산염 결합을 포함하는 5' 캡을 포함한다. 이어서, 이 5'-구아닐레이트 캡은 메틸화되어 N7-메틸-구아닐레이트 잔기를 생성할 수 있다.
- [0224] 본 명세서에 상정된 mRNA 폴리뉴클레오타이드의 특정 실시형태에서 사용하기에 적합한 5' 캡의 예시적 예는 비메틸화된 5' 캡 유사체, 예를 들어, G(5')ppp(5')G, G(5')ppp(5')C, G(5')ppp(5')A; 메틸화된 5' 캡 유사체, 예를 들어, m<sup>7</sup>G(5')ppp(5')G, m<sup>7</sup>G(5')ppp(5')C, 및 m<sup>7</sup>G(5')ppp(5')A; 다이메틸화된 5' 캡 유사체, 예를 들어, m<sup>2,7</sup>G(5')ppp(5')G, m<sup>2,7</sup>G(5')ppp(5')C 및 m<sup>2,7</sup>G(5')ppp(5')A; 트라이메틸화된 5' 캡 유사체, 예를 들어, m<sup>2,2,7</sup>G(5')ppp(5')G, m<sup>2,2,7</sup>G(5')ppp(5')C, 및 m<sup>2,2,7</sup>G(5')ppp(5')A; 다이메틸화된 대칭적 5' 캡 유사체, 예를 들



어,  $m^7G(5')pppm^7(5')G$ ,  $m^7G(5')pppm^7(5')C$ , 및  $m^7G(5')pppm^7(5')A$ ; 및 안티-리버스 5' 캡 유사체, 예를 들어, 안티-리버스 캡 유사체(ARCA) 캡, 표기된  $3'O-Me-m^7G(5')ppp(5')G$ ,  $2'O-Me-m^7G(5')ppp(5')G$ ,  $2'O-Me-m^7G(5')ppp(5')C$ ,  $2'O-Me-m^7G(5')ppp(5')A$ ,  $m^72'd(5')ppp(5')G$ ,  $m^72'd(5')ppp(5')C$ ,  $m^72'd(5')ppp(5')A$ ,  $3'O-Me-m^7G(5')ppp(5')C$ ,  $3'O-Me-m^7G(5')ppp(5')A$ ,  $m^73'd(5')ppp(5')G$ ,  $m^73'd(5')ppp(5')C$ ,  $m^73'd(5')ppp(5')A$  및 이들의 사인산염 유도체(예를 들어, 문헌[Jemielity *et al.*, RNA, 9: 1108-1122 (2003)] 참조)를 포함하지만, 이들로 제한되지 않는다.

[0225] 특정 실시형태에서, mRNA는 제1 전사된 뉴클레오타이드의 5'-말단에 삼인산염 브리지를 통해 연결된 7-메틸 구아닐레이트(" $m^7G$ ")인 5' 캡을 포함하여,  $m^7G(5')ppp(5')N$ 을 초래하며, 여기서 N은 임의의 뉴클레오타이드이다.

[0226] 일부 실시형태에서, mRNA는 5' 캡을 포함하되, 캡은 Cap0 구조(Cap0 구조는 염기 1 및 2에 부착된 리보스의 2'-O-메틸 잔기를 결여함), Cap1 구조(Cap1 구조는 염기 2에서 2'-O-메틸 잔기를 가짐), 또는 Cap2 구조(Cap2 구조는 염기 2와 3 둘 다에 부착된 2'-O-메틸 잔기를 가짐)이다.

[0227] 일 실시형태에서, mRNA는  $m^7G(5')ppp(5')G$  캡을 포함한다.

[0228] 일 실시형태에서, mRNA는 ARCA 캡을 포함한다.

[0229] 특정 실시형태에서, 본 명세서에 상정된 mRNA는 하나 이상의 변형된 뉴클레오타이드를 포함한다.

[0230] 일 실시형태에서, mRNA는 슈도유리딘, 피리딘-4-온 리보뉴클레오타이드, 5-아자-유리딘, 2-티오-5-아자-유리딘, 2-티오유리딘, 4-티오-슈도유리딘, 2-티오-슈도유리딘, 5-하이드록시유리딘, 3-메틸유리딘, 5-카복시메틸-유리딘, 1-카복시메틸-슈도유리딘, 5-프로핀일-유리딘, 1-프로핀일-슈도유리딘, 5-타우리노메틸유리딘, 1-타우리노메틸-슈도유리딘, 5-타우리노메틸-2-티오-유리딘, 1-타우리노메틸-4-티오-유리딘, 5-메틸-유리딘, 1-메틸-슈도유리딘, 4-티오-1-메틸-슈도유리딘, 2-티오-1-메틸-슈도유리딘, 1-메틸-1-테아자-슈도유리딘, 2-티오-1-메틸-1-테아자-슈도유리딘, 다이하이드로유리딘, 다이하이드로슈도유리딘, 2-티오-다이하이드로유리딘, 2-티오-다이하이드로슈도유리딘, 2-메톡시유리딘, 2-메톡시-4-티오-유리딘, 4-메톡시-슈도유리딘, 4-메톡시-2-티오-슈도유리딘, 5-아자-사이티딘, 슈도아이스사이티딘, 3-메틸-사이티딘, N4-아세틸사이티딘, 5-폼일사이티딘, N4-메틸사이티딘, 5-하이드록시메틸사이티딘, 1-메틸-슈도아이스사이티딘, 피롤로-사이티딘, 피롤로-슈도아이스사이티딘, 2-티오-사이티딘, 2-티오-5-메틸-사이티딘, 4-티오-슈도아이스사이티딘, 4-티오-1-메틸-슈도아이스사이티딘, 4-티오-1-메틸-1-테아자-슈도아이스사이티딘, 1-메틸-1-테아자-슈도아이스사이티딘, 제볼라린, 5-아자-제볼라린, 5-메틸-제볼라린, 5-아자-2-티오-제볼라린, 2-티오-제볼라린, 2-메톡시-사이티딘, 2-메톡시-5-메틸-사이티딘, 4-메톡시-슈도아이스사이티딘, 4-메톡시-1-메틸-슈도아이스사이티딘, 2-아미노퓨린, 2,6-다이아미노퓨린, 7-테아자-아데닌, 7-테아자-8-아자-아데닌, 7-테아자-2-아미노퓨린, 7-테아자-8-아자-2-아미노퓨린, 7-테아자-2,6-다이아미노퓨린, 7-테아자-8-아자-2,6-다이아미노퓨린, 1-메틸아데노신, N6-메틸아데노신, N6-아이스오텐틸아데노신, N6-(시스-하이드록시아이스오텐틸)아데노신, 2-메틸티오-N6-(시스-하이드록시아이스오텐틸)아데노신, N6-글리신일카바모일아데노신, N6-트레오닐카바모일아데노신, 2-메틸티오-N6-트레오닐 카바모일아데노신, N6,N6-다이메틸아데노신, 7-메틸아데닌, 2-메틸티오-아데닌, 2-메톡시-아데닌, 이노신, 1-메틸-이노신, 위오신, 위부토신, 7-테아자-구아노신, 7-테아자-8-아자-구아노신, 6-티오-구아노신, 6-티오-7-테아자-구아노신, 6-티오-7-테아자-8-아자-구아노신, 7-메틸-구아노신, 6-티오-7-메틸-구아노신, 7-메틸이노신, 6-메톡시-구아노신, 1-메틸구아노신, N2-메틸구아노신, N2,N2-다이메틸구아노신, 8-옥소-구아노신, 7-메틸-8-옥소-구아노신, 1-메틸-6-티오-구아노신, N2-메틸-6-티오-구아노신 및 N2,N2-다이메틸-6-티오-구아노신으로 이루어진 군으로부터 선택된 하나 이상의 변형된 뉴클레오타이드를 포함한다.

[0231] 일 실시형태에서, mRNA는 슈도유리딘, 피리딘-4-온 리보뉴클레오타이드, 5-아자-유리딘, 2-티오-5-아자-유리딘, 2-티오유리딘, 4-티오-슈도유리딘, 2-티오-슈도유리딘, 5-하이드록시유리딘, 3-메틸유리딘, 5-카복시메틸-유리딘, 1-카복시메틸-슈도유리딘, 5-프로핀일-유리딘, 1-프로핀일-슈도유리딘, 5-타우리노메틸유리딘, 1-타우리노메틸-슈도유리딘, 5-타우리노메틸-2-티오-유리딘, 1-타우리노메틸-4-티오-유리딘, 5-메틸-유리딘, 1-메틸-슈도유리딘, 4-티오-1-메틸-슈도유리딘, 2-티오-1-메틸-슈도유리딘, 1-메틸-1-테아자-슈도유리딘, 2-티오-1-메틸-1-테아자-슈도유리딘, 다이하이드로유리딘, 다이하이드로슈도유리딘, 2-티오-다이하이드로유리딘, 2-티오-다이하이드로슈도유리딘, 2-메톡시유리딘, 2-메톡시-4-티오-유리딘, 4-메톡시-슈도유리딘 및 4-메톡시-2-티오-슈도유리딘으로 이루어진 군으로부터 선택된 하나 이상의 변형된 뉴클레오타이드를 포함한다.

- [0232] 일 실시형태에서, mRNA는 5-아자-사이티딘, 슈도아이소사이티딘, 3-메틸-사이티딘, N4-아세틸사이티딘, 5-폼일 사이티딘, N4-메틸사이티딘, 5-하이드록시메틸사이티딘, 1-메틸-슈도아이소사이티딘, 피콜로-사이티딘, 피콜로-슈도아이소사이티딘, 2-티오-사이티딘, 2-티오-5-메틸-사이티딘, 4-티오-슈도아이소사이티딘, 4-티오-1-메틸-슈도아이소사이티딘, 4-티오-1-메틸-1-테아자-슈도아이소사이티딘, 1-메틸-1-테아자-슈도아이소사이티딘, 제볼라린, 5-아자-제볼라린, 5-메틸-제볼라린, 5-아자-2-티오-제볼라린, 2-티오-제볼라린, 2-메톡시-사이티딘, 2-메톡시-5-메틸-사이티딘, 4-메톡시-슈도아이소사이티딘 및 4-메톡시-1-메틸-슈도아이소사이티딘으로 이루어진 군으로부터 선택된 하나 이상의 변형된 뉴클레오타이드를 포함한다.
- [0233] 일 실시형태에서, mRNA는 2-아미노퓨린, 2,6-다이아미노퓨린, 7-테아자-아데닌, 7-테아자-8-아자-아데닌, 7-테아자-2-아미노퓨린, 7-테아자-8-아자-2-아미노퓨린, 7-테아자-2,6-다이아미노퓨린, 7-테아자-8-아자-2,6-다이아미노퓨린, 1-메틸아데노신, N6-메틸아데노신, N6-아이소펜텐일아데노신, N6-(시스-하이드록시아이소펜텐일)아데노신, 2-메틸티오-N6-(시스-하이드록시아이소펜텐일)아데노신, N6-글리신일카바모일아데노신, N6-트레오닐카바모일아데노신, 2-메틸티오-N6-트레오닐 카바모일아데노신, N6,N6-다이메틸아데노신, 7-메틸아데닌, 2-메틸티오-아데닌 및 2-메톡시-아데닌으로 이루어진 군으로부터 선택된 하나 이상의 변형된 뉴클레오타이드를 포함한다.
- [0234] 일 실시형태에서, mRNA는 이노신, 1-메틸-이노신, 위오신, 위부토신, 7-테아자-구아노신, 7-테아자-8-아자-구아노신, 6-티오-구아노신, 6-티오-7-테아자-구아노신, 6-티오-7-테아자-8-아자-구아노신, 7-메틸-구아노신, 6-티오-7-메틸-구아노신, 7-메틸이노신, 6-메톡시-구아노신, 1-메틸구아노신, N2-메틸구아노신, N2,N2-다이메틸구아노신, 8-옥소-구아노신, 7-메틸-8-옥소-구아노신, 1-메틸-6-티오-구아노신, N2-메틸-6-티오-구아노신 및 N2,N2-다이메틸-6-티오-구아노신으로 이루어진 군으로부터 선택된 하나 이상의 변형된 뉴클레오타이드를 포함한다.
- [0235] 일 실시형태에서, mRNA는 하나 이상의 슈도유리딘, 하나 이상의 5-메틸-사이토신, 및/또는 하나 이상의 5-메틸-사이티딘을 포함한다.
- [0236] 일 실시형태에서, mRNA는 하나 이상의 슈도유리딘을 포함한다.
- [0237] 일 실시형태에서, mRNA는 하나 이상의 5-메틸-사이티딘을 포함한다.
- [0238] 일 실시형태에서, mRNA는 하나 이상의 5-메틸-사이토신을 포함한다.
- [0239] 특정 실시형태에서, 본 명세서에 상정된 mRNA는 엑소뉴클레아제 분해로부터 mRNA를 보호하고, mRNA를 안정화시키며, 번역을 용이하게 하도록 돕는 폴리(A) 꼬리를 포함한다. 특정 실시형태에서, mRNA는 3' 폴리(A) 꼬리 구조를 포함한다.
- [0240] 특정 실시형태에서, 폴리(A) 꼬리의 길이는 적어도 약 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450 또는 적어도 약 500개 이상의 아데닌 뉴클레오타이드 또는 임의의 개재된 수의 아데닌 뉴클레오타이드이다. 특정 실시형태에서, 폴리(A) 꼬리의 길이는 적어도 약 125, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 202, 203, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 235, 236, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 248, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 또는 275개 이상의 아데닌 뉴클레오타이드이다.
- [0241] 특정 실시형태에서, 폴리(A) 꼬리의 길이는 약 10 내지 약 500개의 아데닌 뉴클레오타이드, 약 50 내지 약 500개의 아데닌 뉴클레오타이드, 약 100 내지 약 500개의 아데닌 뉴클레오타이드, 약 150 내지 약 500개의 아데닌 뉴클레오타이드, 약 200 내지 약 500개의 아데닌 뉴클레오타이드, 약 250 내지 약 500개의 아데닌 뉴클레오타이드, 약 300 내지 약 500개의 아데닌 뉴클레오타이드, 약 50 내지 약 450개의 아데닌 뉴클레오타이드, 약 50 내지 약 400개의 아데닌 뉴클레오타이드, 약 50 내지 약 350개의 아데닌 뉴클레오타이드, 약 100 내지 약 500개의 아데닌 뉴클레오타이드, 약 100 내지 약 450개의 아데닌 뉴클레오타이드, 약 100 내지 약 400개의 아데닌 뉴클레오타이드, 약 100 내지 약 350개의 아데닌 뉴클레오타이드, 약 100 내지 약 300개의 아데닌 뉴클레오타이드, 약 150 내지 약 500개의 아데닌 뉴클레오타이드, 약 150 내지 약 450개의 아데닌 뉴클레오타이드, 약 150 내지 약 400개의 아데닌 뉴클레오타이드, 약 150 내지 약 350개의 아데닌 뉴클레오타이드, 약 150 내지 약 300개의 아데닌 뉴클레오타이드, 약 150 내지 약 250개의 아데닌 뉴클레오타이드, 약 150 내지 약 200개의 아데닌 뉴클

레오타이드, 약 200 내지 약 500개의 아데닌 뉴클레오타이드, 약 200 내지 약 450개의 아데닌 뉴클레오타이드, 약 200 내지 약 400개의 아데닌 뉴클레오타이드, 약 200 내지 약 350개의 아데닌 뉴클레오타이드, 약 200 내지 약 300개의 아데닌 뉴클레오타이드, 약 250 내지 약 500개의 아데닌 뉴클레오타이드, 약 250 내지 약 450개의 아데닌 뉴클레오타이드, 약 250 내지 약 400개의 아데닌 뉴클레오타이드, 약 250 내지 약 350개의 아데닌 뉴클레오타이드, 또는 약 250 내지 약 300개의 아데닌 뉴클레오타이드 또는 아데닌 뉴클레오타이드의 임의의 개재 범위이다.

- [0242] 폴리뉴클레오타이드의 배향을 기재하는 용어는 하기를 포함한다: 5'(정상적으로는 유리 인산염기를 갖는 폴리뉴클레오타이드의 말단) 및 3'(정상적으로는 유리 하이드록실(OH) 기를 갖는 폴리뉴클레오타이드의 말단). 폴리뉴클레오타이드 서열은 5'에서 3' 배향 또는 3'에서 5' 배향으로 주석이 달릴 수 있다. DNA 및 mRNA에 대해, 5'에서 3' 가닥은 "센스", "플러스" 또는 "암호" 가닥으로 표기되는데, 그의 서열이 전-전령(전-mRNA)의 서열과 동일하기 때문이다[DNA에서 티민(T)을 대신하는 RNA에서의 유라실(U)을 제외함]. DNA 및 mRNA에 대해, RNA 중합효소에 의해 전사되는 가닥인 상보성 3'에서 5' 가닥은 "주형", "안티센스", "마이너스" 또는 "비암호화" 가닥으로 표기된다. 본 명세서에서 사용되는 용어 "역 배향"은 3'에서 5' 배향으로 기재된 5'에서 3' 서열 또는 5'에서 3' 배향으로 기재된 3'에서 5' 서열을 지칭한다.
- [0243] 용어 "상보적" 및 "상보성"은 염기쌍 규칙과 관련된 폴리뉴클레오타이드(즉, 뉴클레오타이드의 서열)를 지칭한다. 예를 들어, DNA 서열 5' A G T C A T G 3'의 상보성 가닥은 3' T C A G T A C 5'이다. 후자의 서열은 종종 좌측 상에서 5' 말단 및 우측 상에서 3' 말단과 역 상보체로서 기재된다, 5' C A T G A C T 3'. 역 상보체와 동일한 서열은 회문구조 서열인 것으로 언급된다. 상보성은 "부분적"일 수 있는데, 이때 핵산의 염기 중 일부만이 염기쌍 규칙에 따라 매칭된다. 또는, 핵산 사이의 "완전한" 또는 "충" 상보성이 있을 수 있다.
- [0244] 폴리뉴클레오타이드는 당염기에 공지되고 이용 가능한 임의의 다양한 잘 확립된 기법을 이용하여 제조되고/되거나 조작되고/되거나 발현되고/되거나 전달될 수 있다. 목적으로 하는 폴리펩타이드를 발현시키기 위해, 폴리펩타이드를 암호화하는 뉴클레오타이드 서열은 적절한 벡터 내로 삽입될 수 있다. 목적으로 하는 폴리펩타이드는 또한 폴리펩타이드를 암호화하는 mRNA를 세포 내로 전달함으로써 발현될 수 있다.
- [0245] 벡터의 예시적 예는 플라스미드, 자율적 복제 서열 및 전위요소, 예를 들어, 잠자는 미녀(Sleeping Beauty), 피기백(PiggyBac)을 포함하지만, 이들로 제한되지 않는다.
- [0246] 벡터의 추가적인 예시적 예는 플라스미드, 파지미드, 코스미드, 인공 염색체, 예컨대 효모 인공 염색체(yeast artificial chromosome: YAC), 박테리아 인공 염색체(bacterial artificial chromosome: BAC) 또는 P1-유래 인공 염색체(PAC), 박테리오파지, 예컨대 람다 파지 또는 M13 파지 및 동물 바이러스를 포함하지만, 이들로 제한되지 않는다.
- [0247] 벡터로서 유용한 바이러스의 예시적 예는 레트로바이러스(렌티바이러스 및 통합 결핍 렌티바이러스를 포함), 아데노바이러스, 아데노-연관 바이러스, 헤르페스바이러스(예를 들어, 단순 포진 바이러스), 폭스바이러스, 마를로바이러스, 유두종바이러스 및 과보바이러스(예를 들어, SV40)을 포함하지만, 이들로 제한되지 않는다.
- [0248] 발현 벡터의 예시적 예는 포유류 세포에서 발현을 위한 pCIneo 벡터(프로메가(Promega)); 포유류 세포에서 렌티바이러스-매개 유전자 전달 및 발현을 위한 pLenti4/V5-DEST(상표명), pLenti6/V5-DEST(상표명) 및 pLenti6.2/V5-GW/lacZ(인비트로젠(Invitrogen))를 포함하지만, 이들로 제한되지 않는다. 특정 실시형태에서, 본 명세서에 개시된 폴리펩타이드의 암호 서열은 포유류 세포에서 폴리펩타이드의 발현을 위해 이러한 발현 벡터 내로 결합될 수 있다.
- [0249] 특정 실시형태에서, 벡터는 에피솜 벡터 또는 염색체외로 유지되는 벡터이다. 본 명세서에서 사용되는 용어 "에피솜"은 숙주의 염색체 DNA 내로의 통합 없이 그리고 분할 숙주 세포로부터의 점진적 상실 없이 복제할 수 있는 벡터를 지칭하는데, 또한 이는 상기 벡터가 염색체외로 또는 에피솜으로 복제한다는 것을 의미한다.
- [0250] 특정 실시형태에서, 각각의 복수의 폴리펩타이드의 효율적인 번역을 달성하기 위해, 폴리뉴클레오타이드 서열은 자기-절단성 폴리펩타이드를 암호화하는 하나 이상의 IRES 서열 또는 폴리뉴클레오타이드 서열에 의해 분리될 수 있다.
- [0251] 본 명세서에서 사용되는, "내부 리보솜 유입 부위" 또는 "IRES"는 시스트론(단백질 암호화 영역)의 개시 코돈, 예컨대 ATG에 대한 직접 내부 리보솜 유입을 촉진시킴으로써 유전자의 캡-독립적 번역을 야기하는 요소를 지칭한다. 예를 들어, 문헌[Jackson *et al.*, 1990. *Trends Biochem Sci* 15(12):477-83] 및 Jackson and Kaminski. 1995. *RNA* 1(10):985-1000] 참조. 당업자에 의해 일반적으로 사용되는 IRES의 예는 미국 특허 제6,692,736호에

기재된 것을 포함한다. 당업계에 공지된 "IRES"의 추가적인 예는 피코르나 바이러스로부터 얻을 수 있는 IRES(Jackson *et al.*, 1990) 및 바이러스 또는 세포 mRNA 공급원, 예를 들어, 면역글로불린 중쇄 결합 단백질 (BiP), 혈관 내피 성장인자(VEGF)(Huez *et al.* 1998. *Mol. Cell. Biol.* 18(11):6178-6190), 섬유아세포 성장 인자 2(FGF-2), 및 인슐린-유사 성장 인자(IGFII), 번역 개시 인자 eIF4G 및 효모 전사 인자 TFIID 및 HAP4, 노바겐(Novagen)으로부터 상업적으로 입수 가능한 뇌심근염 바이러스(EMCV)(Duke *et al.*, 1992. *J. Virol* 66(3):1602-9)로부터 얻을 수 있는 IRES 및 VEGF IRES(Huez *et al.*, 1998. *Mol Cell Biol* 18(11):6178-90)를 포함하지만, 이들로 제한되지 않는다. IRES는 또한 피코르나비리대(Picornaviridae), 디시스트로비리대(Dicistroviridae) 및 플라리비리대(Flaviviridae) 종의 바이러스 게놈에서 그리고 HCV, 프렌드 뮈린 백혈병 바이러스(FrMLV) 및 몰로니 뮈린 백혈병 바이러스(MoMLV)에서 보고되었다.

[0252] 일 실시형태에서, 본 명세서에서 상정된 폴리뉴클레오타이드에서 사용되는 IRES는 EMCV IRES이다.

[0253] 특정 실시형태에서, 폴리뉴클레오타이드는 공통 코작 서열을 갖고 목적으로 하는 폴리펩타이드를 암호화하는 폴리뉴클레오타이드를 포함한다. 본 명세서에서 사용되는 용어 "코작 서열"은 리보솜에 대한 mRNA의 초기 결합을 크게 용이하게 하고 번역을 증가시키는 짧은 뉴클레오타이드 서열을 지칭한다. 공통 코작 서열은 (GCC)RCCATGG (서열번호 66)이며, 여기서 R은 퓨린(A 또는 G)이다(Kozak, 1986. *Cell*. 44(2):283-92, 및 Kozak, 1987. *Nucleic Acids Res.* 15(20):8125-48).

[0254] 이중성 핵산 전사체의 효율적인 종결 및 폴리아데닐화를 지시하는 요소는 이중성 유전자 발현을 증가시킨다.

[0255] 특정 실시형태에서, 벡터는 발현될 폴리펩타이드를 암호화하는 폴리뉴클레오타이드의 폴리아데닐화 서열 3' 을 포함한다. 본 명세서에서 사용되는 용어 "폴리(A) 부위" 또는 "폴리(A) 서열"은 RNA 중합효소 II에 의한 초기 RNA 전사체의 종결과 폴리아데닐화를 둘 다 지시하는 DNA 서열을 의미한다. 폴리아데닐화 서열은 암호 서열의 3' 말단에 대한 폴리(A) 꼬리의 첨가에 의해 mRNA 안정성을 촉진시킬 수 있고, 따라서, 증가된 번역 효율에 기여한다. 절단 및 폴리아데닐화는 RNA 내 폴리(A) 서열에 의해 지시된다. 포유류 전-mRNA에 대한 코어 폴리(A) 서열은 절단-폴리아데닐화 부위에 축적하는 2개의 인식 요소를 가진다. 전형적으로 대부분의 비변이체 AAUAAA 핵사머는 U 또는 GU 잔기가 풍부한 더 가변적인 요소 상류의 20 내지 50개의 뉴클레오타이드에 놓인다. 초기 전 사체의 절단은 이들 두 요소 사이에 일어나며, 5' 절단 산물에 대한 250개까지의 아데노신의 첨가에 결합된다. 특정 실시형태에서, 코어 폴리(A) 서열은 합성 폴리(A) 서열(예를 들어, AATAAA, ATTAAA, AGTAAA)이다. 폴리(A) 서열의 예시적 예는 SV40 폴리(A) 서열, 소 성장 호르몬 폴리(A) 서열(BGHpA), 토끼 β-글로빈 폴리(A) 서열(rβgpA) 또는 당업계에 공지된 다른 적합한 이중성 또는 내인성 폴리(A) 서열을 포함하지만, 이들로 제한되지 않는다.

[0256] 특정 실시형태에서, 하나 이상의 뉴클레아제 변이체, 메가TAL, 말단-가공 효소 또는 융합 폴리펩타이드를 암호화하는 폴리뉴클레오타이드는 조혈세포, 예를 들어, T 세포 내로, 비바이러스 및 바이러스적 방법에 의해 도입될 수 있다. 특정 실시형태에서, 뉴클레아제 및/또는 공여자 수선 주형을 암호화하는 하나 이상의 폴리뉴클레오타이드의 전달은 동일한 방법에 의해 또는 상이한 방법에 의해, 그리고/또는 동일한 벡터에 의해 또는 상이한 벡터에 의해 제공될 수 있다.

[0257] 용어 "벡터"는 다른 핵산 분자를 전달하거나 또는 수송할 수 있는 핵산 분자를 지칭하기 위해 본 명세서에서 사용된다. 전달된 핵산은 일반적으로 벡터 핵산 분자에 연결되고, 예를 들어, 이에 삽입된다. 벡터는 세포 내 자 율 복제를 지시하는 서열을 포함할 수 있거나, 또는 숙주 세포 DNA 내로의 통합을 허용하기에 충분한 서열을 포함할 수 있다. 특정 실시형태에서, 비바이러스 벡터는 본 명세서에 상정된 하나 이상의 폴리뉴클레오타이드를 T 세포에 전달하는 데 사용된다.

[0258] 비바이러스 벡터의 예시적 예는 플라스미드(예를 들어, DNA 플라스미드 또는 RNA 플라스미드), 트랜스포존, 코스미드 및 박테리아 인공 염색체를 포함하지만, 이들로 제한되지 않는다.

[0259] 특정 실시형태에서 상정된 폴리뉴클레오타이드의 비바이러스 전달의 예시적 방법은 전기천공법, 초음파천공법, 리포펙션, 미량주사법, 유전자총법, 바이로솜, 리포솜, 면역리포솜, 나노입자, 다양이온 또는 지질:핵산 접합체, 네이키드 DNA, 인공 비리온, DEAE-텍스트란-매개 전달, 유전자 총 및 열-충격을 포함하지만, 이들로 제한되지 않는다.

[0260] 특정 실시형태에 상정된 특정 실시형태에서 사용하기에 적합한 폴리뉴클레오타이드 전달 시스템의 예시적 예는 아막사 바이오시스템즈(Amaxa Biosystems), 맥스사이트 인코포레이티드(Maxcyte, Inc.), BTX 분자물리 딜리버리 시스템즈(BTX Molecular Delivery Systems) 및 코페르니쿠스 세라퓨틱스 인코포레이티드(Copernicus



Therapeutics Inc.)에 의해 제공된 것을 포함하지만, 이들로 제한되지 않는다. 리포펙션 시약은 상업적으로 시판된다(예를 들어, 트랜스펙탐(Transfectam)(상표명) 및 리포펙틴(Lipofectin)(상표명)). 폴리뉴클레오타이드의 효율적인 수용체-인식 리포펙션에 적합한 양이온성 및 중성 지질은 문헌에 기재되어 있다. 예를 들어, 문헌[Liu et al. (2003) *Gene Therapy*. 10:180-187; 및 Balazs et al. (2011) *Journal of Drug Delivery*. 2011:1-12] 참조. 항체-표적화된, 박테리아 유래된, 살아있지 않은 나노세포-기반 전달이 또한 특정 실시형태에서 상정된다.

[0261] **F.세포**

[0262] 특정 실시형태에서 상정된 방법에 의해 제조된 게놈 편집된 세포는 TCR $\alpha$  유전자에서의 하나 이상의 유전자 편집을 포함하고, 암, GVHD, 감염성 질환, 자가면역 질환, 면역결핍 또는 이와 관련된 병태의 적어도 하나의 증상의 예방, 치료 또는 개선을 위한 개선된 세포 기반 치료를 제공한다. 임의의 특정 이론에 의해 구속되는 일 없이, 본 명세서에 상정된 방법에 의해 제조된 게놈 편집된 면역 효과기 세포는 증가된 개선된 안전성, 생체내 효능 및 지속능력을 포함하는, 우수한 특성을 지닌다.

[0263] 특정 실시형태에서 상정된 게놈 편집된 세포는 자가/자율("자기") 또는 비-자가("비-자기", 예를 들어, 동종이계, 동계 또는 이종성)일 수 있다. 본 명세서에서 사용되는 "자가"는 동일한 대상체로부터의 세포를 지칭한다. 본 명세서에서 사용되는 "동종이계"는 비교하는 세포와 유전적으로 상이한 동일한 종의 세포를 지칭한다. 본 명세서에서 사용되는 "동계"는 비교하는 세포와 유전적으로 동일한 상이한 대상체의 세포를 지칭한다. 본 명세서에서 사용되는 "이종성"은 비교하는 세포와 상이한 종의 세포를 지칭한다. 바람직한 실시형태에서, 세포는 포유류 대상체로부터 얻는다. 더 바람직한 실시형태에서, 세포는 영장류 대상체, 선택적으로는 비인간 영장류로부터 얻는다. 가장 바람직한 실시형태에서, 세포는 인간 대상체로부터 얻는다.

[0264] "단리된 세포"는 비-천연 유래 세포, 예를 들어, 생체내 조직 또는 인간으로부터 얻거나 또는 세포의 기질이 실질적으로 없는 천연에서 존재하지 않는 세포, 변형된 세포, 조작된 세포 등을 지칭한다.

[0265] 본 명세서에서 사용되는 용어 "세포의 집단"은 본 명세서의 다른 곳에 기재된 바와 같이 다수의 그리고/또는 상동성 또는 이종성 세포 유형의 조합을 구성할 수 있는 복수의 세포를 지칭한다. 예를 들어, T 세포의 형질도입을 위해, 세포의 집단은 말초 혈액으로부터 단리되거나 또는 얻어질 수 있다. 세포의 집단은 편집될 표적 세포 유형의 약 10%, 약 20%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%를 포함할 수 있다. 특정 실시형태에서, T 세포는 당업계에 공지된 방법을 이용하여 이종성 세포 집단으로부터 단리되거나 또는 정제될 수 있다.

[0266] 게놈이 본 명세서에 상정된 조성물 및 방법을 이용하여 편집될 수 있는 세포 유형의 예시적 예는 세포주, 1차 세포, 줄기 세포, 조상 세포, 및 분화된 세포, 및 이들의 혼합물을 포함하지만, 이들로 제한되지 않는다.

[0267] 바람직한 실시형태에서, 게놈 편집 조성물 및 방법은 조혈세포, 더 바람직하게는 면역 세포, 및 더욱더 바람직하게는 T 세포를 편집하기 위해 사용된다.

[0268] 용어 "T 세포" 또는 "T 림프구"는 당업계에 인식되어 있고, 흉선세포, 면역 효과기 세포, 조절 T 세포, 미경험 T 림프구, 미성숙 T 림프구, 성숙 T 림프구, 휴지 T 림프구 또는 활성화된 T 림프구를 포함하는 것으로 의도된다. T 세포는 T 헬퍼(Th) 세포, 예를 들어 T 헬퍼 1(Th1) 또는 T 헬퍼 2(Th2) 세포일 수 있다. T 세포는 헬퍼 T 세포(HTL; CD4<sup>+</sup> T 세포) CD4<sup>+</sup> T 세포, 세포독성 T 세포(CTL; CD8<sup>+</sup> T 세포), 중앙 침윤성 세포독성 T 세포(TIL; CD8<sup>+</sup> T 세포), CD4<sup>+</sup>CD8<sup>+</sup> T 세포, CD4<sup>+</sup>CD8<sup>-</sup> T 세포, 또는 T 세포의 임의의 다른 서브세트일 수 있다. 일 실시형태에서, T 세포는 면역 효과기 세포이다. 일 실시형태에서, T 세포는 NKT 세포이다. 특정 실시형태에서 사용하기에 적합한 T 세포의 다른 예시적 집단은 미경험 T 세포 및 기억 T 세포를 포함한다.

[0269] 다양한 실시형태에서, 게놈 편집된 세포는 본 명세서에 상정된 조성물 및 방법에 의해 편집된 TCR $\alpha$  유전자를 포함하는 면역 효과기 세포를 포함한다. "면역 효과기 세포"는 하나 이상의 효과기 기능(예를 들어, 세포독성 세포 사멸 활성, 사이토카인의 분비, ADCC 및/또는 CDC의 유도)을 갖는 면역계의 임의의 세포이다. 특정 실시형태에서 상정된 예시적인 면역 효과기 세포는 T 림프구, 특히 세포독성 T 세포(CTL; CD8<sup>+</sup> T 세포), TIL 및 헬퍼 T 세포(HTL; CD4<sup>+</sup> T 세포)이다. 일 실시형태에서, 면역 효과기 세포는 자연 살해(NK) 세포이다. 일 실시형태에서, 면역 효과기 세포는 자연 살해 T(NKT) 세포이다.

[0270] "강한 T 세포" 및 "어린 T 세포"는 특정 실시형태에서 상호호환적으로 사용되며, T 세포 표현형을 지칭하되, T

세포는 증식 및 분화의 동시 감소를 가능하게 한다. 특정 실시형태에서, 어린 T 세포는 "미경험 T 세포"의 표현형을 가진다. 특정 실시형태에서, 어린 T 세포는 다음의 생물학적 마커: CD62L, CCR7, CD28, CD27, CD122, CD127, CD197 및 CD38 중 하나 이상, 또는 모두를 포함한다. 일 실시형태에서, 어린 T 세포는 다음의 생물학적 마커: CD62L, CD127, CD197 및 CD38 중 하나 이상 또는 모두를 포함한다. 일 실시형태에서, 어린 T 세포는 CD57, CD244, CD160, PD-1, CTLA4 및 LAG3의 발현을 결여한다.

[0271] T 세포는 말초혈액 단핵세포, 골수, 림프절 조직, 제대혈, 흉선 조직, 감염 부위로부터의 조직, 복수, 흉수, 비장 조직 및 종양을 포함하지만, 이들로 제한되지 않는 다수의 공급원으로부터 얻을 수 있다.

[0272] 특정 실시형태에서, 세포의 집단은 호밍 엔도뉴클레아제 변이체 또는 본 명세서에 상정된 바와 같은 TCR  $\alpha$  유전자 불변 영역의 엑손 1을 표적화하는 메가TAL를 포함하는 면역 효과기 세포 또는 T 세포를 포함한다.

[0273] 특정 실시형태에서, 세포의 집단은 호밍 엔도뉴클레아제 변이체 또는 본 명세서에 상정된 바와 같은 TCR  $\alpha$  유전자 불변 영역의 엑손 1을 표적화하는 메가TAL폴리펩타이드를 포함하는 면역 효과기 세포 또는 T 세포를 포함한다.

[0274] 특정 실시형태에서, 세포의 집단은 호밍 엔도뉴클레아제 변이체 또는 본 명세서에 상정된 바와 같은 TCR  $\alpha$  유전자 불변 영역의 엑손 1을 표적화하는 메가TAL폴리펩타이드를 포함하는 벡터를 포함하는 면역 효과기 세포 또는 T 세포를 포함한다.

[0275] 특정 실시형태에서, 세포의 집단은 호밍 엔도뉴클레아제 변이체 또는 본 명세서에 상정된 바와 같은 TCR  $\alpha$  유전자 불변 영역의 엑손 1을 표적화하는 메가TAL폴리펩타이드를 암호화하는 mRNA를 포함하는 면역 효과기 세포 또는 T 세포를 포함한다.

#### [0276] G. 조성물 및 제형

[0277] 특정 실시형태에서 상정된 조성물은 본 명세서에 상정된 바와 같이 하나 이상의 폴리펩타이드, 폴리뉴클레오타이드, 이를 포함하는 벡터, 및 게놈 편집 조성물 및 게놈 편집 세포 조성물을 포함할 수 있다. 특정 실시형태에서 상정된 게놈 편집 조성물 및 방법은 세포 또는 세포 집단에서 인간 T 세포 수용체 알파(TCR  $\alpha$ ) 유전자 내 표적 부위를 편집하는 데 유용하다. 바람직한 실시형태에서, 게놈 편집 조성물은 조혈세포, 예를 들어, T 세포 또는 면역 효과기 세포에서 TCR  $\alpha$  유전자를 편집하는 데 사용될 수 있다.

[0278] 다양한 실시형태에서, 본 명세서에 상정된 조성물은 뉴클레아제 변이체, 및 선택적으로 말단-가공 효소, 예를 들어, 3'-5' 엑소뉴클레아제(Trex2)를 포함한다. 뉴클레아제 변이체는 상기 개시된 폴리뉴클레오타이드 전달 방법, 예를 들어, 전기천공법, 지질 나노입자 등을 통해 세포 내로 도입되는 mRNA 형태일 수 있다. 일 실시형태에서, 호밍 엔도뉴클레아제 변이체 또는 메가TAL을 암호화하는 mRNA, 및 선택적으로 3'-5' 엑소뉴클레아제를 포함하는 조성물은 상기 개시한 폴리뉴클레오타이드 전달 방법을 통해 세포에 도입된다. 조성물은 오류유발 NHEJ에 의해 게놈 편집된 세포 또는 게놈 편집된 세포의 집단을 생성하는 데 사용될 수 있다.

[0279] 다양한 실시형태에서, 본 명세서에 상정된 조성물은 공여자 수선 주형을 포함한다. 조성물은 뉴클레아제 변이체, 및 선택적으로 말단-가공 효소를 발현시키거나 또는 발현시킬 세포에 전달될 수 있다. 일 실시형태에서, 조성물은 호밍 엔도뉴클레아제 변이체 또는 메가TAL, 및 선택적으로 3'-5' 엑소뉴클레아제를 발현시키거나 또는 발현시킬 세포에 전달될 수 있다. 공여자 수선 주형의 존재 하에 유전자 편집 효소의 발현은 HDR에 의해 게놈 편집된 세포 또는 게놈 편집된 세포의 집단을 생성하는 데 사용될 수 있다.

[0280] 조성물은 약제학적 조성물을 포함하지만, 이들로 제한되지 않는다. "약제학적 조성물"은 단독으로 또는 한 가지 이상의 치료 양상과 조합하여 세포 또는 동물에게 투여를 위한 약제학적으로 허용 가능한 또는 생리적으로 허용 가능한 용액 중에서 제형화된 조성물을 지칭한다. 또한 원한다면, 조성물은 마찬가지로 다른 제제, 예컨대 사이토카인, 성장 인자, 호르몬, 소분자, 화학치료제, 프로드러그, 약물, 항체 또는 다른 다양한 약제학적 활성제와 조합하여 투여될 수 있다는 것이 이해될 것이다. 또한 조성물 중에 포함될 수 있는 다른 성분에 대한 제한은 사실상 없으며, 단, 추가적인 제제는 조성물에 유해하게 영향을 미치지 않는다.

[0281] 어구 "약제학적으로 허용 가능한"은 타당한 의학적 판단의 범위 내에서, 합리적인 유해/유익비에 비례하여 과도한 독성, 자극, 알레르기 반응 또는 다른 문제 또는 합병증 없이 인간 및 동물의 조직과 접촉하여 사용하기에 적합한 본 명세서에서 해당 화합물, 물질, 조성물 및/또는 투약 형태를 지칭하기 위해 사용된다.

[0282] 용어 "약제학적으로 허용 가능한 담체"는 치료 세포와 함께 투여되는 희석제, 보조제, 부형제 또는 비히클을 지칭한다. 약제학적 담체의 예시적 예는 멸균 액체, 예컨대 세포 배양 배지, 물 및 석유, 동물성, 식물성 또는 합

성 유래의 오일, 예컨대 땅콩유, 대두유, 광유, 참깨유 등일 수 있다. 식염수 용액 및 수성 텍스트로스와 글리세롤 용액은 또한 액체 담체로서, 특히 주사용 용액으로 사용될 수 있다. 특정 실시형태에서, 적합한 약제학적 부형제는 전분, 글루코스, 락토스, 수크로스, 젤라틴, 맥아, 쌀, 밀가루, 백악, 실리카겔, 스테아르산나트륨, 글리세롤 모노스테아레이트, 톨크, 염화나트륨, 건조탈지유, 글리세롤, 프로필렌, 글리콜, 물, 에탄올 등을 포함한다. 임의의 통상적인 배지 또는 제제가 활성 성분에 적합하지 않은 경우를 제외하고, 치료 조성물에서의 그의 용도가 상정된다. 보충적 활성 성분이 또한 조성물 내로 혼입될 수 있다.

[0283] 일 실시형태에서, 약제학적으로 허용 가능한 담체를 포함하는 조성물은 대상체에 대한 투여에 적합하다. 특정 실시형태에서, 담체를 포함하는 조성물은 비경구 투여, 예를 들어, 혈관내(정맥내 또는 동맥내), 복강내 또는 근육내 투여에 적합하다. 특정 실시형태에서, 약제학적으로 허용 가능한 담체를 포함하는 조성물은 심실내, 척수내 또는 척추강내에 적합하다. 약제학적으로 허용 가능한 담체 멸균 수용액, 세포 배양 배지 또는 분산물을 포함한다. 약제학적으로 활성인 물질에 대한 이러한 배지 및 제제의 사용은 당업계에 잘 공지되어 있다. 임의의 통상적인 배지 또는 제제가 형질도입 세포에 적합하지 않은 경우를 제외하고, 약제학적 조성물 중의 그의 용도가 상정된다.

[0284] 특정 실시형태에서, 본 명세서에서 상정되는 조성물은 유전자 변형된 T 세포 및 약제학적으로 허용 가능한 담체를 포함한다. 본 명세서에 상정된 세포 기반 조성물을 포함하는 조성물은 장용 또는 비경구 투여 방법에 의해 별개로 또는 목적으로 하는 치료 목적을 달성하기 위한 다른 적합한 화합물과 조합하여 투여될 수 있다.

[0285] 약제학적으로 허용 가능한 담체는 치료 중인 인간 대상체에 대한 투여에 적합하도록 충분히 고순도 그리고 충분히 낮은 독성을 가져야 한다. 추가로 조성물의 안정성을 유지하거나 또는 증가시켜야 한다. 약제학적으로 허용 가능한 담체는 액체 또는 고체일 수 있고, 조성물의 다른 성분과 조합될 때 목적으로 하는 벌크, 점조도 등을 제공하기 위해, 유념해둔 계획한 투여 방식으로 선택된다. 예를 들어, 약제학적으로 허용 가능한 담체는 결합제(예를 들어, 전호화된 메이즈 전분, 폴리비닐피롤리돈 또는 하이드록시프로필 메틸셀룰로스 등), 충전제(예를 들어, 락토스 및 다른 당, 미정질 셀룰로스, 펙틴, 젤라틴, 황산칼슘, 에틸 셀룰로스, 폴리아크릴레이트, 인산 수소칼슘 등), 윤활제(예를 들어, 스테아르산마그네슘, 톨크, 실리카, 콜로이드 이산화규소, 스테아르산, 금속 스테아레이트, 수소화된 식물성유, 옥수수 전분, 폴리에틸렌 글리콜, 벤조산나트륨, 아세트산나트륨 등), 봉쇄제(예를 들어, 전분, 글리콜산나트륨전분 등) 또는 습윤제(예를 들어, 라우릴황산나트륨 등)일 수 있지만, 이들로 제한되지 않는다. 본 명세서에 상정된 조성물에 대한 다른 적합한 약제학적으로 허용 가능한 담체는 물, 염 용액, 알코올, 폴리에틸렌 글리콜, 젤라틴, 아밀로스, 스테아르산마그네슘, 톨크, 규산, 점성 파라핀, 하이드록시메틸셀룰로스, 폴리비닐피롤리돈 등을 포함하지만, 이들로 제한되지 않는다.

[0286] 이러한 담체 용액은 또한 완충제, 희석제 및 다른 적합한 첨가제를 함유할 수 있다. 본 명세서에서 사용되는 용어 "완충제"는 화학적 구성이 pH의 상당한 변화없이 산 또는 염기를 중화시키는 용액 또는 액체를 지칭한다. 본 명세서에 상정된 완충제의 예는 둘베코 인산염 완충 식염수(PBS), 링거 용액, 수 중 5% 텍스트로스(D5W), 정상/생리식염수(0.9% NaCl)를 포함하지만, 이들로 제한되지 않는다.

[0287] 약제학적으로 허용 가능한 담체는 조성물의 pH를 약 7로 유지하기에 충분한 양으로 존재할 수 있다. 대안적으로, 조성물은 약 6.8 내지 약 7.4, 예를 들어, 6.8, 6.9, 7.0, 7.1, 7.2, 7.3 및 7.4 범위의 pH를 가진다. 또 다른 실시형태에서, 조성물은 약 7.4의 pH를 가진다.

[0288] 본 명세서에 상정된 조성물은 비독성의 약제학적으로 허용 가능한 배지를 포함할 수 있다. 조성물은 현탁액일 수 있다. 본 명세서에서 사용되는 용어 "현탁액"은 세포가 고체 지지체에 부착되지 않는 비부착 상태를 지칭한다. 예를 들어, 현탁액으로서 유지된 세포는 교반되거나 또는 휘저어질 수 있고, 지지체, 예컨대 배양 접시에 부착되지 않는다.

[0289] 특정 실시형태에서, 본 명세서에 상정된 조성물은 계능 편집된 T 세포가 정맥내(IV) 등에서 허용 가능한 액체 배지 또는 용액, 예를 들어, 식염수 또는 무혈청 배지 내에서 분산되는 현탁액 중에서 제형화된다. 허용 가능한 희석제는 물, 링거용액, 등장성 염화나트륨(식염수) 용액, 무혈청 세포 배양 배지 및 저온 저장에 적합한 배지, 예를 들어, Cryostor(등록상표) 배지를 포함하지만, 이들로 제한되지 않는다.

[0290] 특정 실시형태에서, 약제학적으로 허용 가능한 담체는 인간 또는 동물 유래의 천연 단백질이 실질적으로 없고, 계능 편집된 T 세포 집단을 포함하는 조성물을 저장하는 데 적합하다. 치료 조성물은 인간 환자에게 투여되도록 의도되고, 따라서 세포 배양 성분, 예컨대 소 혈청 알부민, 말 혈청 및 소태아 혈청이 실질적으로 없다.

[0291] 일부 실시형태에서, 조성물은 약제학적으로 허용 가능한 세포 배양 배지에서 제형화된다. 이러한 조성물은 인간

대상체에 대한 투여에 적합하다. 특정 실시형태에서, 약제학적으로 허용 가능한 세포 배양 배지는 무혈청 배지이다.

[0292] 무혈청 배지는 단순화된 그리고 더 잘 한정된 조성물, 감소된 정도의 오염물질, 감염제의 잠재적 공급원의 제거 및 보다 저비용을 포함하는, 혈청 함유 배지 이상의 몇몇 이점을 가진다. 다양한 실시형태에서, 무혈청 배지는 무동물이며, 선택적으로 무단백질일 수 있다. 선택적으로, 배지는 생약제학적으로 허용 가능한 재조합 단백질을 함유할 수 있다. "무 동물" 배지는 성분이 비동물 공급원으로부터 유래된 배지를 지칭한다. 재조합 단백질은 무동물 배지에서 천연 동물 단백질을 대체하며, 영양분은 합성, 식물 또는 미생물 공급원으로부터 얻어진다. "무 단백질" 배지는 대조적으로 단백질이 실질적으로 없는 것으로 정의된다.

[0293] 특정 조성물에서 사용되는 무혈청 배지의 예시적 예는 QBSF-60(퀄리티 바이올로지컬 인코포레이티드(Quality Biological, Inc.)), 스템프로-34(StemPro-34)(라이프 테크놀로지스(Life Technologies)) 및 X-VIVO 10을 포함하지만, 이들로 제한되지 않는다.

[0294] 바람직한 실시형태에서, 게놈 편집된 T 세포를 포함하는 조성물은 PlasmaLyte에서 제형화된다.

[0295] 다양한 실시형태에서, 게놈 편집된 T 세포를 포함하는 조성물은 동결보존 배지에서 제형화된다. 예를 들어, 동결보존제가 있는 동결보존 배지는 해동 후 높은 세포 생존도 결과를 유지하기 위해 사용될 수 있다. 특정 조성물에서 사용되는 동결보존 배지의 예시적 예는 CryoStor CS10, CryoStor CS5 및 CryoStor CS2를 포함하지만, 이들로 제한되지 않는다.

[0296] 일 실시형태에서, 조성물은 50:50 PlasmaLyte A 대 CryoStor CS10을 포함하는 용액에서 제형화된다.

[0297] 특정 실시형태에서, 조성물은 마이코플라스마, 내독소 및 미생물 오염물질이 실질적으로 없다. 내독소와 관련하여 "실질적으로 없는"은 생물제제에 대해 FDA에 의해 허용되는 것(5 EU/kg 체중/일의 총 내독소, 평균 70kg의 사람은 세포의 총 용량 당 350 EU임)보다 세포의 용량 당 내독소가 더 적다는 것을 의미한다. 특정 실시형태에서, 본 명세서에 상정된 형질도입된 조혈 줄기 또는 조상 세포를 포함하는 조성물은 약 0.5 EU/ml, 내지 약 5.0EU/ml, 또는 약 0.5EU/ml, 1.0EU/ml, 1.5EU/ml, 2.0EU/ml, 2.5EU/ml, 3.0EU/ml, 3.5EU/ml, 4.0EU/ml, 4.5 EU/ml, 또는 5.0 EU/ml를 함유한다.

[0298] 특정 실시형태에서, 하나 이상의 재프로그래밍된 뉴클레아제를 암호화하는 하나 이상의 mRNA, 및 선택적으로 말단-가공 효소를 포함하지만, 이들로 제한되지 않는 폴리뉴클레오타이드의 전달에 적합한 조성물 및 제형이 상정된다.

[0299] 생체의 전달을 위한 예시적인 제형은 또한 당업계에 공지된 다양한 형질감염제, 예컨대 인산칼슘, 전기천공법, 열 충격 및 다양한 리포솜 제형(즉, 지질-매개 형질감염)의 사용을 포함할 수 있다. 이하에 더 상세하게 기재하는 리포솜은 수성 유체의 분획을 붙잡는 지질 이중층이다. DNA는 양이온성 리포솜의 외면에 (그의 하전에 의해) 결합하고, 이들 리포솜은 세포막과 상호작용할 것이다.

[0300] 특정 실시형태에서, 약제학적으로 허용 가능한 담체 용액은, 예를 들어, 장용 및 비경구, 예를 들어, 혈관내, 정맥내, 동맥내, 골내, 심실내, 대뇌내, 두개내, 척수내, 척추강내 및 골수내 투여 및 제형을 포함하는 다양한 치료 요법에서 본 명세서에 기재된 특정 조성물을 이용하기 위한 적합한 투약 및 치료 요법의 개발과 같이, 당업자에게 잘 공지되어 있다. 본 명세서에 상정된 특정 실시형태는 다른 제형, 예컨대 약제학적 분야에 잘 공지되어 있고, 예를 들어, 본 명세서에 전문이 참고로 포함된 문헌[Remington: The Science and Practice of Pharmacy, volume I and volume II. 22<sup>nd</sup> Edition. Edited by Loyd V. Allen Jr. Philadelphia, PA: Pharmaceutical Press; 2012]에 기재되어 있는 것을 포함할 수 있다는 것은 당업자에 의해 이해될 것이다.

[0301] 본 명세서의 모든 간행물, 특허 출원 및 공개된 특허는 각각의 개개 간행물, 특허 출원 또는 공개된 특허가 구체적인 개별적으로 참고로 포함되도록 포함되는 것과 같이 본 명세서에 참고로 포함된다.

[0302] 앞서 언급한 실시형태는 이해의 분명함의 목적을 위해 예시 및 실시예로 일부 상세하게 기재하였지만, 특정 변화 및 변형이 첨부된 청구범위의 정신 또는 범주로부터 벗어나는 일 없이 이루어질 수 있다는 것은 본 명세서에 상정된 교시에 비추어 당업자에게 용이하게 분명할 것이다. 다음의 실시예는 단지 예로서 제한의 방법에 의하지 않고 제공한다. 당업자는 본질적으로 유사한 결과를 획득하기 위해 변화 또는 변형될 수 있는 다양한 중요하지 않은 매개변수를 용이하게 인식할 것이다.

[0303] 실시예



- [0304] 실시예 1
- [0305] 인간 TCR α 유전자를 표적화하는 I-OnuI 변이체의 선택성 개선
- [0306] 차선 선택성 영역을 인간 TCR α 유전자 불변 영역 내 엑손 1을 표적화하는 I-OnuI 변이체(서열번호 6)("본래의 TCR α I-OnuI 변이체")의 DNA 인식 계면 영역에서 동정하였다. 표적 부위 내 위치 p4, p5 및 p6은 염기쌍 치환에 대해 상대적으로 용인되는 것을 발견하였다. 뉴클레오타이드는 TCR α 표적 상의 부위(p456에서 CCG)(서열번호 17) 및 KAT2B 유전자의 인트로 6 내 빈번한 표적이탈 부위(p456에서 ACA, 문헌[Osborne et. al., 2015] 참조)(도 1, 서열번호 21)에서 상이하다. 이들 위치에서 향상된 선택성을 갖는 도메인을 포함하는 I-OnuI 변이체는 전반적인 선택성 특성뿐만 아니라 KAT2B 표적이탈 부위에서 활성의 선택적 감소(서열번호 21)를 개선시키도록 개발되었다.
- [0307] 천연 표적 부위와 복잡체화된 야생형 I-OnuI 단백질(예를 들어, 서열번호 1 내지 5)의 결정 구조의 연구는 위치 p456와 접촉하거나 또는 근위일 가능성이 있는 8개 아미노산 잔기의 서브세트를 동정하였다. 인간 TCR α 유전자를 표적화하는 다른 I-OnuI 변이체의 국소 분석은 본래의 TCR α 표적화 I-OnuI 변이체와의 비교에서 CCG에 대해 상당히 더 큰 선택성을 갖는 아미노산을 동정하였다. TCR α 표적화 I-OnuI 변이체에서 높은 CCG 삼중 선택성을 초래하는 아미노산을 본래의 TCR α I-OnuI 변이체에 삽입하여 TCR α 2.0 I-OnuI 변이체(서열번호 7)를 생성하였다(도 2). TCR α 2.0 I-OnuI 변이체 효소 활성을 효모 표면 디스플레이를 이용하여 p456 기질 위치에서 모두 64-뉴클레오타이드 조합에 대해 평가하였다(도 3). TCR α 2.0 I-OnuI 변이체는 상당히 더 양호한 기질 식별을 나타내는데, 이는 본래의 TCR α I-OnuI 변이체보다 훨씬 더 소수의 기질을 절단함으로써 표시된다.
- [0308] 실시예 2
- [0309] TCR α 2.0 메가TAL의 촉매 활성 개선
- [0310] TCR α 좌위의 붕괴는 세포 표면에 대한 CD3의 수송을 방지한다. CD3 발현은 형광 표지된 항-CD3 항체 및 유세포 분석을 이용하여 측정될 수 있다. 유전자 편집 효율은 CD3 형광에 반비례한다.
- [0311] 1차 인간 T 세포는 CD3 및 CD28에 의해 활성화되고, 메가TAL로서 형식화된 본래의 TCR α HE 변이체(예를 들어, 서열번호 13) 또는 메가TAL로서 형식화된 TCR α 2.0 HE 변이체(예를 들어, 서열번호 14) 중 하나를 암호화하는 시험관 내 전사된 mRNA로 전기천공시켰다. p456 기질 위치에서 상당히 향상된 선택성을 나타내지만, TCR α 2.0 메가TAL은 감소된 촉매 활성을 나타내었는데, 본래의 TCR α 메가TAL에 비해 증가된 CD3에 의해 나타내었다(도 4). 게놈 DNA를 메가TAL 처리 세포로부터 후속적으로 단리시키고, 표적 부위에서 분해에 의해 삽입결실의 추적(Tracking of Indels by DEcomposition: TIDE, 문헌[Brinkman et al., 2014] 참조)과 결합된 PCR을 이용하여 감소된 촉매적 활성을 확인하였다.
- [0312] 다른 I-OnuI 변이체에서 촉매적 활성이 증가된 단일 아미노산 치환인 E178D는 TCR α 2.0 메가TAL의 활성 부위에 삽입되어 TCR α 2.1 메가TAL(서열번호 8)을 생성하였다. TCR α 2.1 메가TAL(예를 들어, 서열번호 15)을 암호화하는 시험관내 전사된 mRNA를 1차 인간 T 세포에 전기천공시켰다. TCR α 2.1 메가TAL은 TCR α 2.0 메가TAL에 비해 CD3 Nickdown을 10배 증가시켰다. 도 5.
- [0313] 실시예 3
- [0314] TCR α 2.1 I-OnuI 변이체 및 TCR α 2.1 메가TAL은 향상과 표적 상과 표적이탈의 차이를 나타내었다
- [0315] 치료적 게놈 편집은 매우 낮은 빈도의 표적이탈 편집 사건에 기인하는 위험도를 운반한다. 게놈 편집 효소는 존재하는 방법을 이용하여 표적이탈 편집 사건에 대한 깊이를 특성규명할 수 있다(문헌[Tsai et al. 2016] 참조).
- [0316] TCR α 표적 서열과 비슷한 22bp 서열(서열번호 17)에 대한 인간 게놈을 스캔하는 생물정보학 알고리즘을 이용하여 잠재적 표적이탈 부위를 동정하였다. 생물정보학 연구로부터 돌려받은 상부 764개의 22-bp 서열을 효모 표면 디스플레이를 이용하여 본래의 TCR α HE 변이체 및 TCR α 2.1 HE 변이체에 의한 절단에 대해 개개로 분석하였다. TCR α 2.1 HE 변이체는 본래의 TCR α HE 변이체(도 6)에 비해 TCR α 표적 서열에 대해 증가된 선택성을 나타내었는데, 이는 실시예 1에서 얻은 결과와 일치된다.
- [0317] 연구를 추가로 알아내기 위한 764개의 기질 패널로부터의 상부 50개의 절단된 기질의 서열 정보를 이용하여 제2 생물정보학 연구를 수행하였다. 본래의 TCR α 및 TCR α 2.1 HE 변이체에 의한 절단을 위해 이들을 개개로 분석하였다. 정제된 TCR α 2.1 HE 변이체는 본래의 TCR α HE 변이체에 비해 향상된 선택성 프로파일을 나타내었다.
- [0318] 인테그라제 결핍 렌티바이러스 벡터(IDLV, 문헌[Gabriel 2011] 참조)를 사용하는 제2 분석을 사용하여 생체내에

서 생성된 추가적인 표적이탈 부위를 동정하였다. 간략하게, IDLV를 효소 처리 및 모의 처리 세포 내에서 이중 가닥 파손 부위에서 포획한다. 이어서, IDLV 통합 빈도 및 위치를 맵핑하고 두 집단 간을 비교할 수 있다. 모의 처리된 샘플에 비해 효소 처리된 샘플 내 특정 게놈 부위에서 증가된 IDLV 흡수 빈도는 효소 절단을 나타낸다. 이 방법을 이용하여 본래의 TCR  $\alpha$  메가TAL의 1차 표적이탈 부위를 KAT2B 유전자(서열번호 21, 문헌[Osborne et al. 2015])의 인트론 6에서 동정하였다.

[0319] 동정한 추정적 표적이탈 부위를 메가TAL 처리된 공여자 유래 1차 T 세포로부터 생성된 앰플리콘의 심층 서열분석에 의해 평가하였다(도 7). TCR  $\alpha$  2.1 메가TAL은 KAT2B 유전자 좌위에서 표적이탈 편집을 제거하였고, 본래의 TCR  $\alpha$  메가TAL에 비해 표적 상의 부위에 대해 향상된 선택성을 나타내고, 표적이탈 부위에서 감소된 활성을 나타낸다.

[0320] 실시예 4

[0321] TAL DNA 결합 도메인의 연장은 TCR  $\alpha$  게놈 편집을 증가시킨다

[0322] 메가TAL은 목적으로 하는 표적 서열에 대해 결합 활성을 증가시키거나 또는 처리하는 메가뉴클레아제 구조에 대한 TALE DNA 결합 도메인 융합을 포함한다(문헌[Boissel et al. 2013] 참조). TCR  $\alpha$  좌위에서 표적 상의 편집 효율을 추가로 증가시키기 위해, 10.5 내지 11.5개의 RVD 반복부로부터의 TALE DNA 결합 도메인 어레이의 총 길이를 증가시키기 위해 그리고 11.5 RVD TCR  $\alpha$  2.1 메가TAL(예를 들어, 서열번호 12)을 생성하기 위해 TALE DNA 결합 도메인의 C-말단 단부에 추가적인 TALE RVD를 현수하였다. 11.5개의 RVD TALE DNA 결합 도메인은 3' 말단 상에서 추가적인 뉴클레오타이드(서열번호 19)를 갖는 본래의 TALE 결합 부위(서열번호 18)를 인식한다.

[0323] 인간 1차 T 세포를 10.5개의 RVD 및 11.5개의 RVD TCR  $\alpha$  2.1 메가TAL을 암호화하는 시험관내 전사된 mRNA로 전기천공시켰다. 11.5개의 RVD TCR  $\alpha$  2.1 메가TAL은 10.5 RVD TCR  $\alpha$  2.1 메가TAL에 의한 TCR  $\alpha$  편집에 비해 유세포분석을 이용하여 CD3 낙다운에 의해 측정하여, TCR  $\alpha$  편집을 30 내지 50%로 증가시켰다(도 8).

[0324] 실시예 5

[0325] 변이체 N-말단 TALE 구조

[0326] 메가TAL은 N-말단 도메인, DNA 결합 도메인, 및 C-말단 도메인을 포함하는 TAL 효과기(TALE) 구조에 융합된 호밍 엔도뉴클레아제를 포함하는 혼성 뉴클레아제이다. 일부 TALE 구조는 대략 136개의 아미노산의 최소 N-말단 영역('텔타-154'), 가변적 수의 34개의 아미노산 RVD-함유 반복부(절단된 20개 아미노산 반복부에 의해 종결), 및 대략 63개의 아미노산의 최소 C-말단 영역을 포함한다.

[0327] 앞의 실시예에 기재한 호밍 엔도뉴클레아제에 융합된 TALE 구조는 잔토모나스 오리자에 TALE 단백질로부터 유래된 서열을 포함한다. 잔토모나스 오리자에 TALE 도메인 서열에 추가로, 최소 N-말단 TALE 도메인(NTD) 영역은 잔토모나스과의 다른 구성원, 예컨대 잔토모나스 트랜스루센스(*Xanthomonas translucens*) 및 잔토모나스 시트리(*Xanthomonas citri*)로부터 유래될 수 있다. N-말단 TALE 도메인의 다중 서열 정렬은 잔토모나스 오리자가 각각 잔토모나스 트랜스루센스 및 잔토모나스 시트리와 85% 및 87% 서열 유사성을 공유한다는 것을 나타내었다. TAL 활성이 대안의 중 NTD를 이용하여 추가로 향상될 수 있는지 여부를 평가하기 위해, 내인성 잔토모나스 트랜스루센스 및 잔토모나스 시트리 서열 내 다음의 아미노산 위치를 합리적으로 치환하고 조합함으로써 혼성 NTD를 생성하였다: V152G, D153K, D165E, E166K, I167L, P169L, A173P, T174I, Q180E, M183I, H191Q, V205I, V207A, I224V, V227I, A236V, G249S, G257S, V271I(예를 들어, 서열번호 21); V152G, D153K, T170K, V171A, V183I, A200P, A236T, A239V, A244V, E245Q, G257S, L260V, V271I, (예를 들어, 서열번호 24); T170K, A200P, Q230H, W231C, S232G, A236T, A239V, A244V, E245Q, G257S(예를 들어, 서열번호 23). 혼성 서열(예를 들어, 서열번호 22, 23, 24)은 잔토모나스 오리자에 N-말단 TAL 도메인에 대해 80% 유사성을 나타내었다. 각각의 이들 최소 TALE 도메인은 본래의 TCR  $\alpha$  I-OnuI 변이체 메가TAL의 TALE DNA 결합 도메인의 N-말단에 개개로 접합시켰다. TALE N-말단 도메인(NTD) 변이체 또는 잔토모나스 오리자에 NTD를 포함하는 변이체 메가TAL을 암호화하는 시험관내 전사된 mRNA를 인간 1차 T 세포에 전기천공시켜 TCR  $\alpha$  좌위에서 편집률을 평가하였다. CD3 발현의 유세포분석은 잔토모나스 오리자에 NTD를 포함하는 본래의 TCR  $\alpha$  메가TAL에 비해 NTD 변이체 도메인을 포함하는 TCR  $\alpha$  메가TAL이 유사한 활성, 일부 경우에 약간 더 높은 활성을 나타내었다는 것을 입증하였다(도 9).

[0328] 일반적으로, 다음의 청구범위에서, 사용하는 용어는 본 명세서 및 청구범위에 개시된 구체적 실시형태로 청구범위를 제한하는 것으로 해석되어서는 안 되며, 이러한 청구범위가 자격을 부여하는 균등론의 완전한 범주와 함께 모든 가능한 실시형태를 포함하는 것으로 해석되어야 한다. 따라서, 청구범위는 본 개시내용에 의해 제한되지

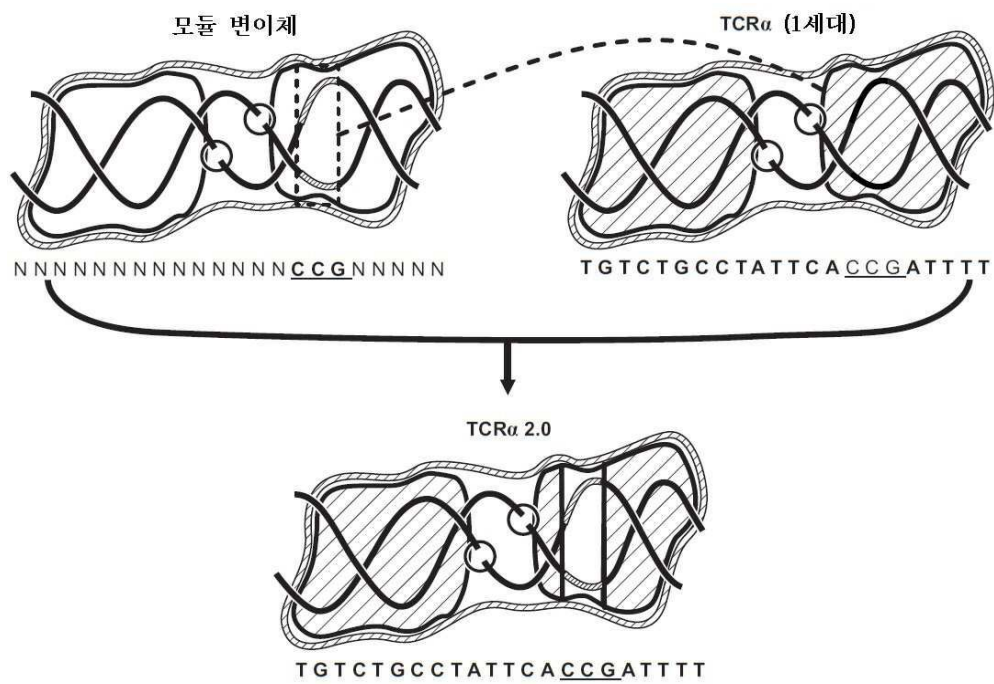
않는다.

도면

도면1

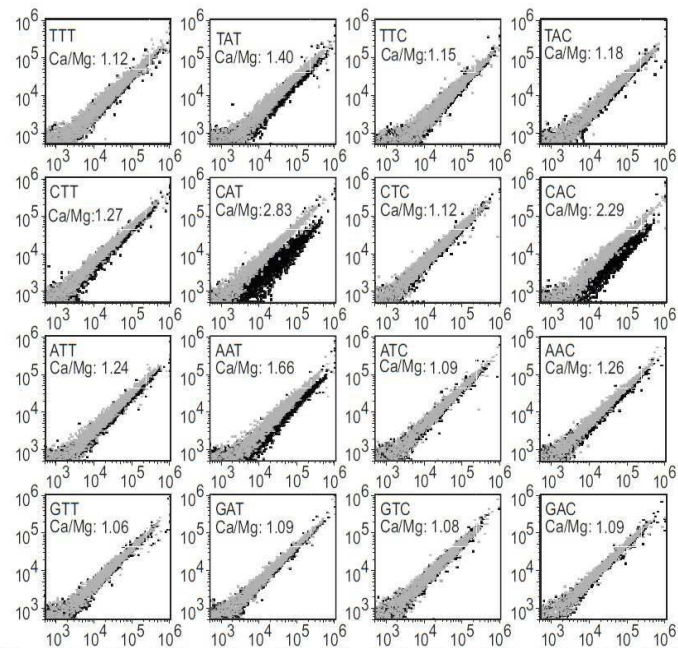
	n <sub>11</sub>	n <sub>10</sub>	n <sub>9</sub>	n <sub>8</sub>	n <sub>7</sub>	n <sub>6</sub>	n <sub>5</sub>	n <sub>4</sub>	n <sub>3</sub>	n <sub>2</sub>	n <sub>1</sub>	p <sub>1</sub>	p <sub>2</sub>	p <sub>3</sub>	p <sub>4</sub>	p <sub>5</sub>	p <sub>6</sub>	p <sub>7</sub>	p <sub>8</sub>	p <sub>9</sub>	p <sub>10</sub>	p <sub>11</sub>
TCRa	T	G	T	C	T	G	C	C	T	A	T	T	C	A	C	C	G	A	T	T	T	T
KAT2B	A	G	T	A	T	G	C	C	T	A	T	T	C	T	A	C	A	A	G	A	A	A

도면2

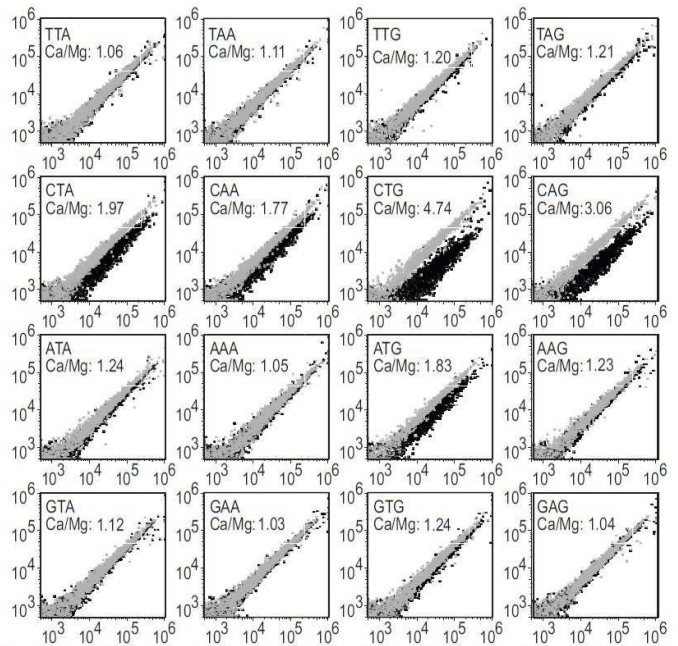


도면3aa

TCR $\alpha$  2.0

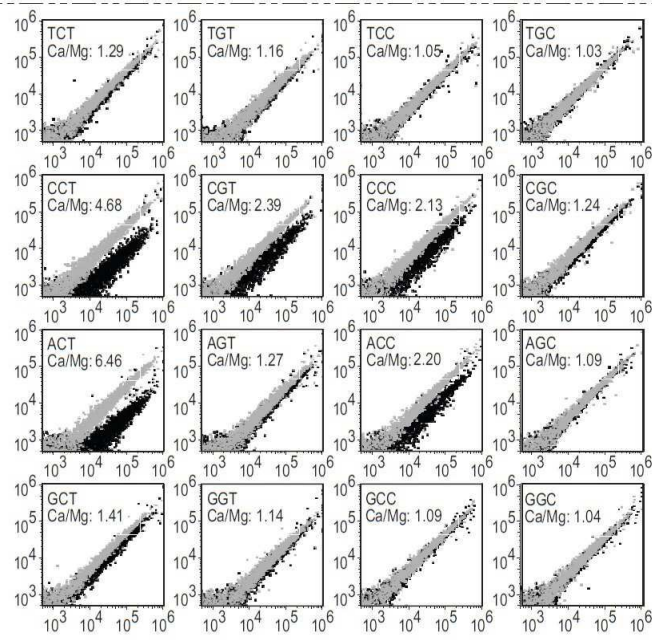


도면3ab

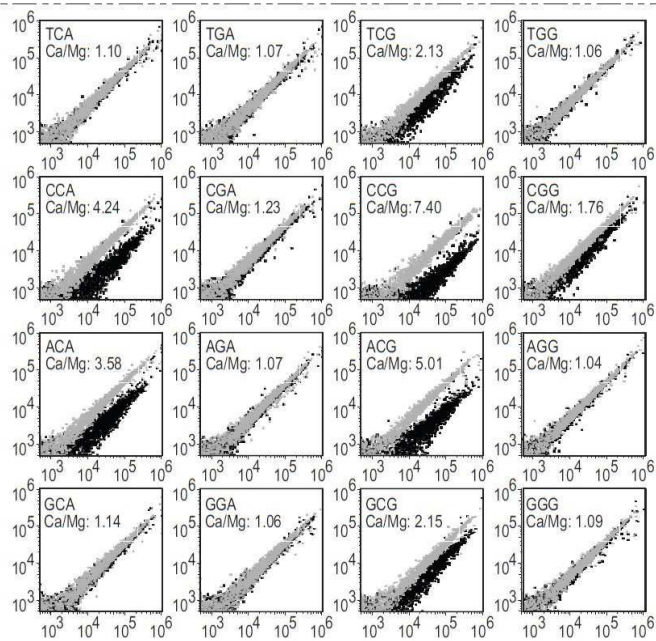




도면3ac

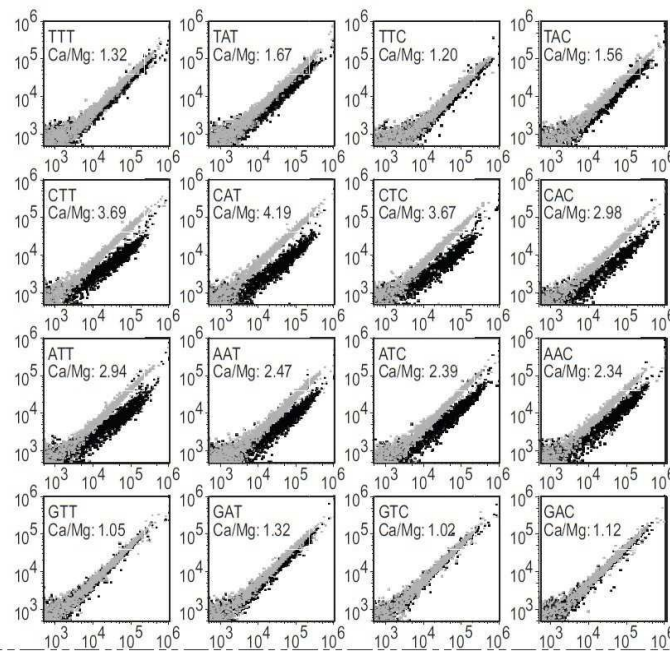


도면3ad

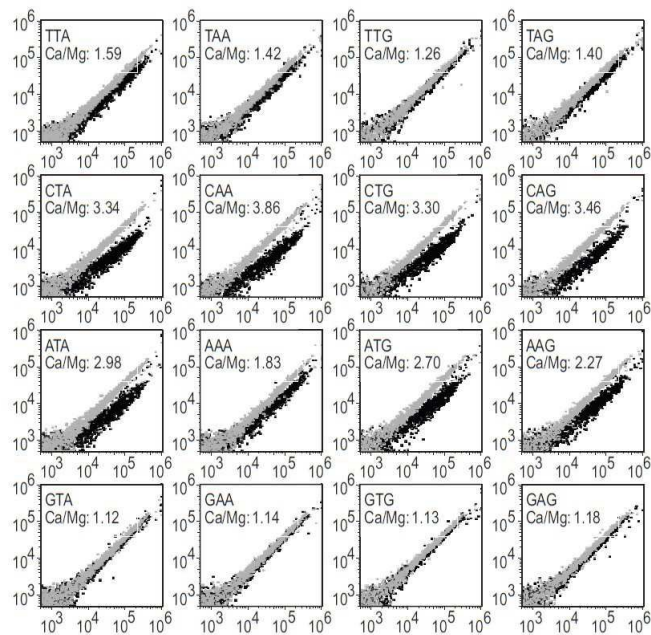


도면3ae

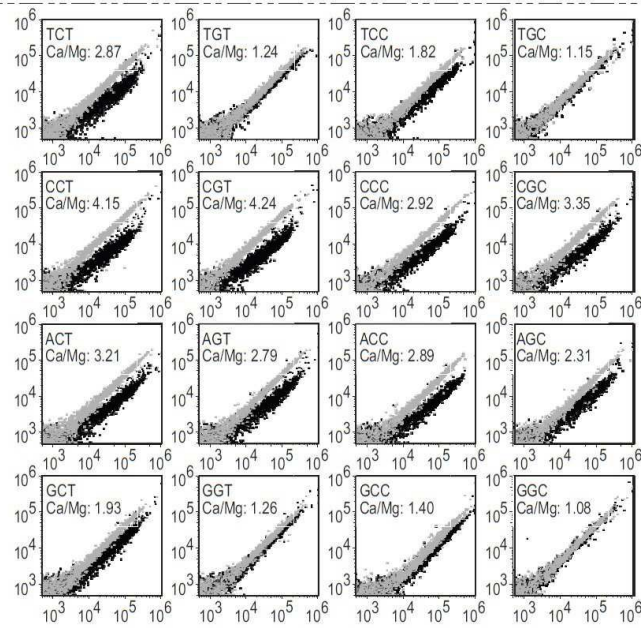
TCR $\alpha$  (1세대)



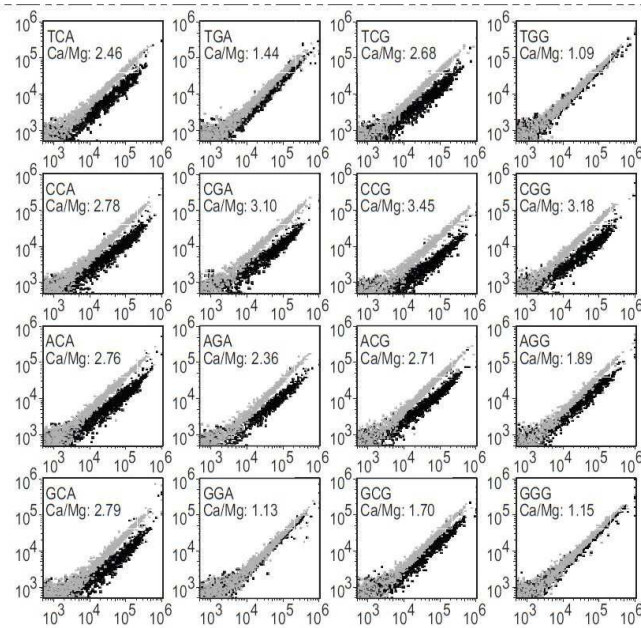
도면3af



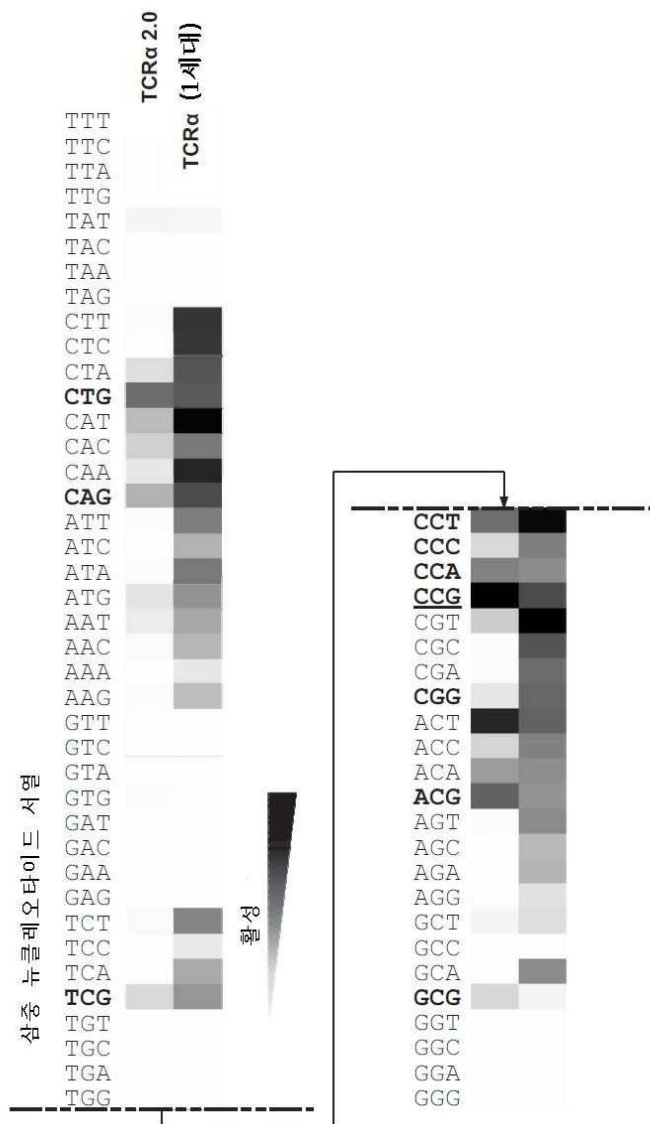
도면3ag



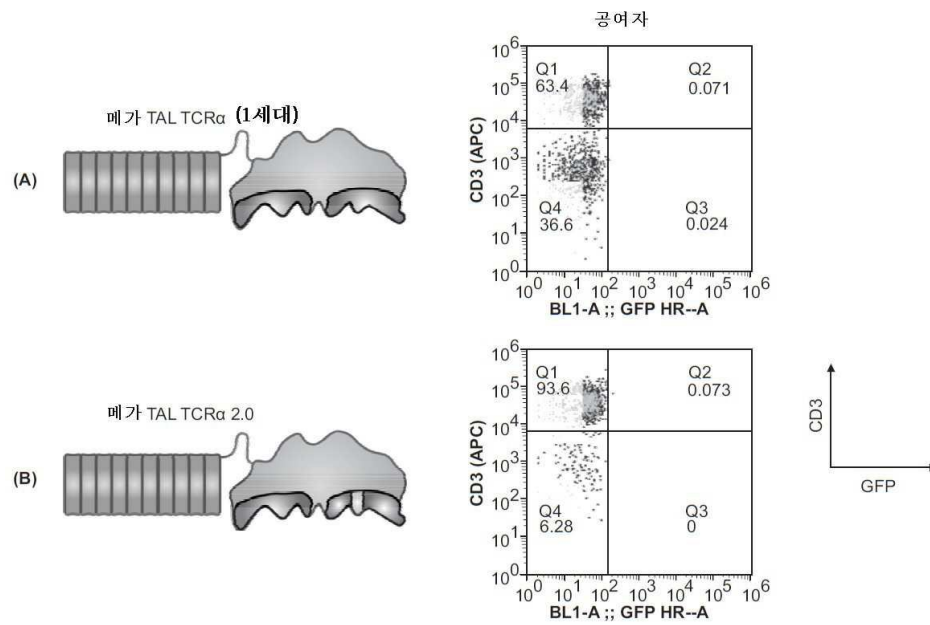
도면3ah



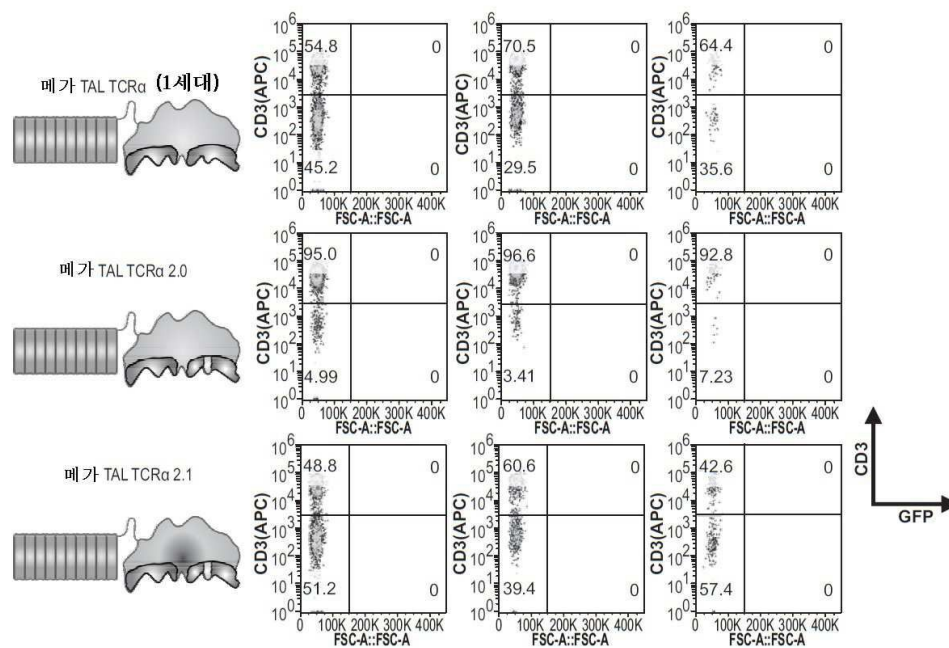
도면3b



도면4

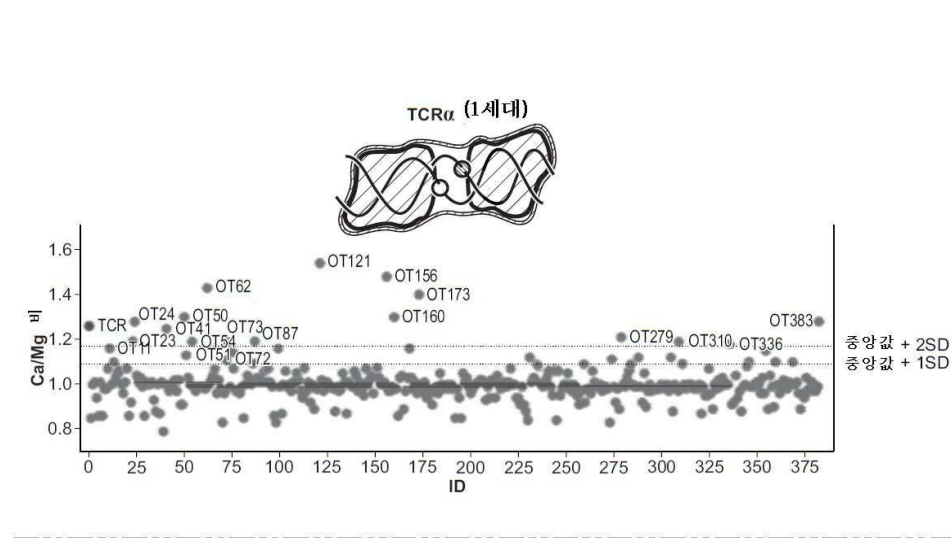


도면5

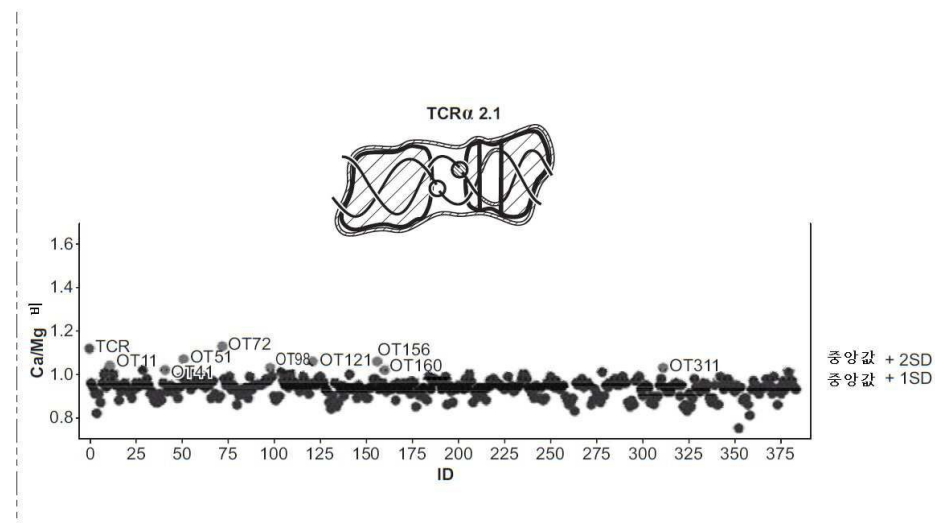




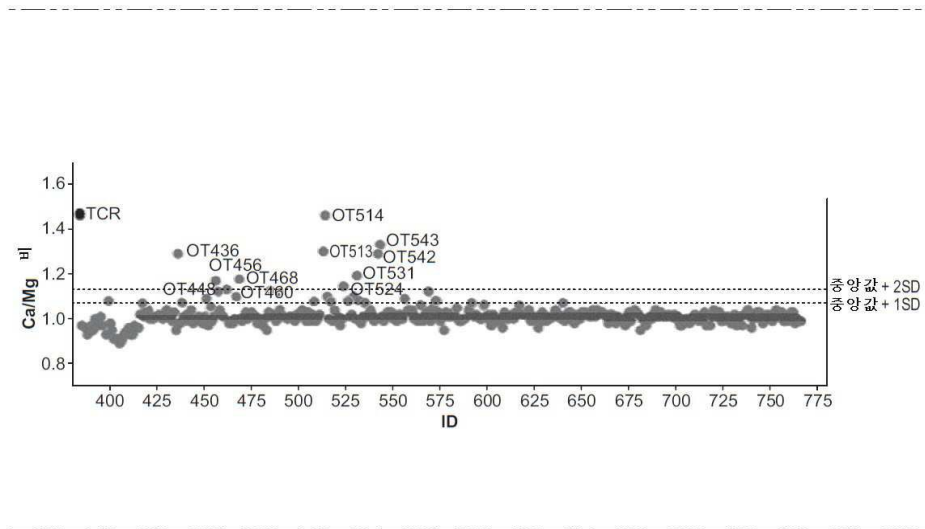
도면6a



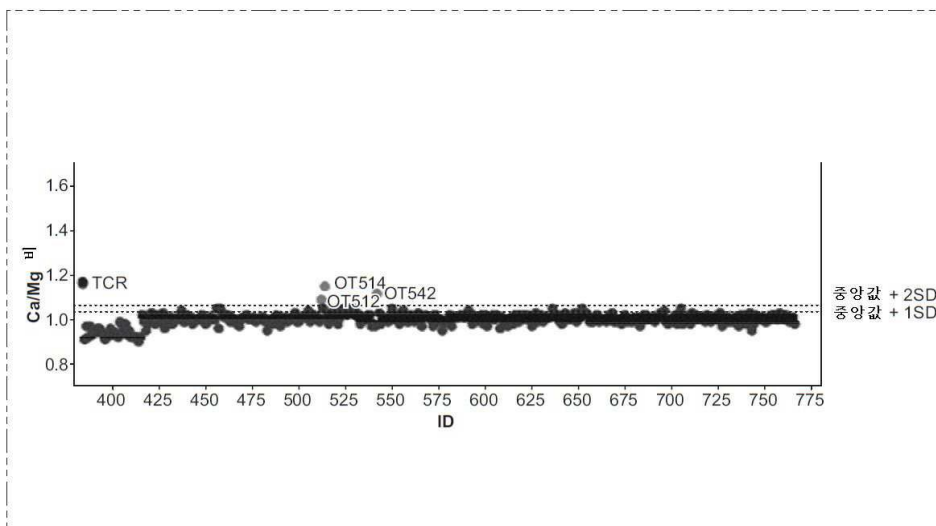
도면6b



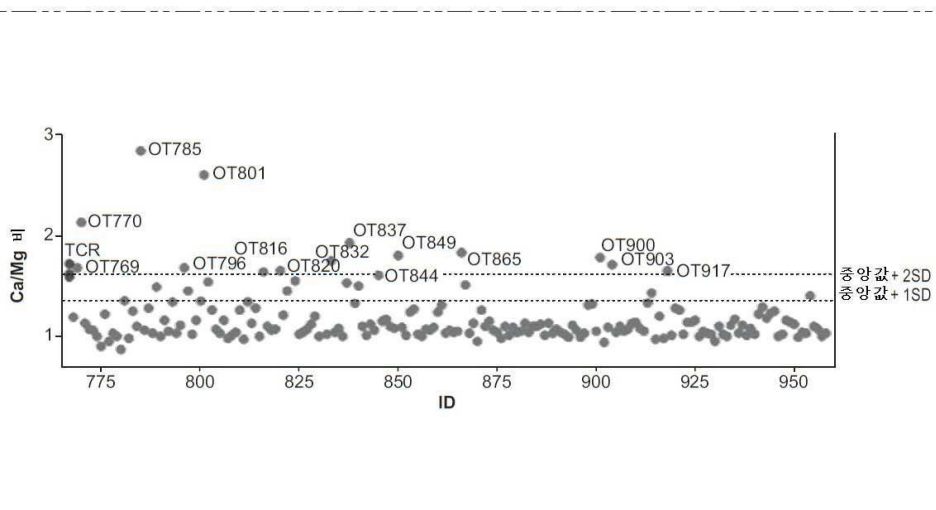
도면6c



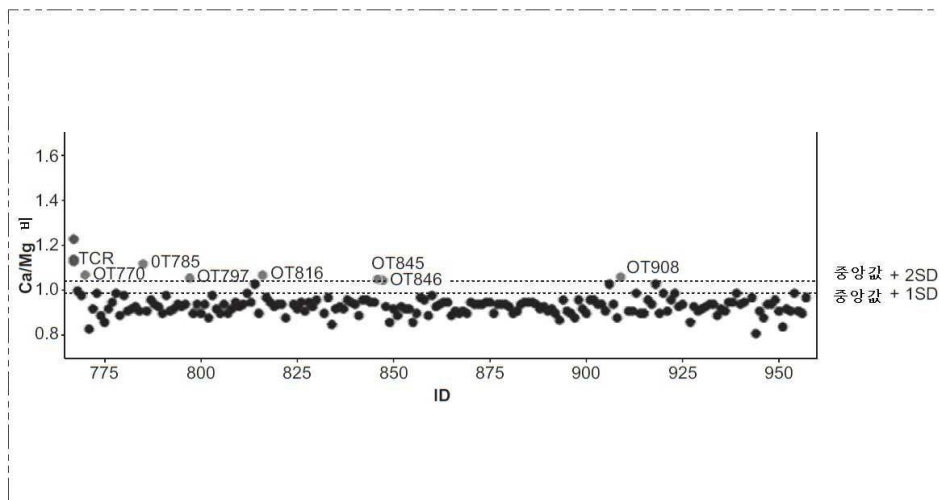
도면6d



도면6e

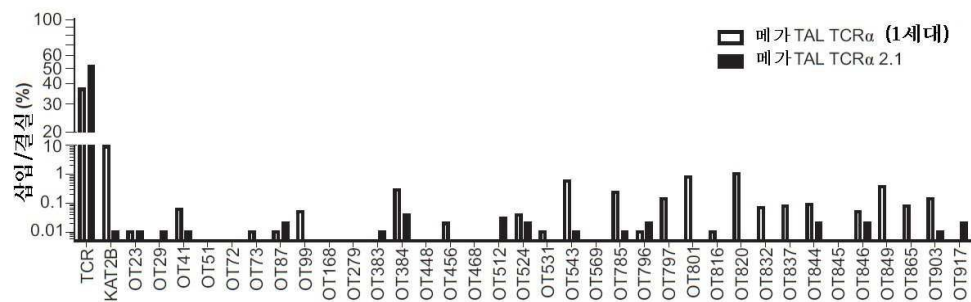


도면6f

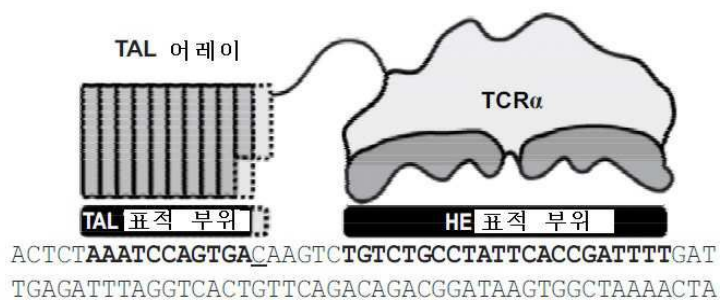


도면7

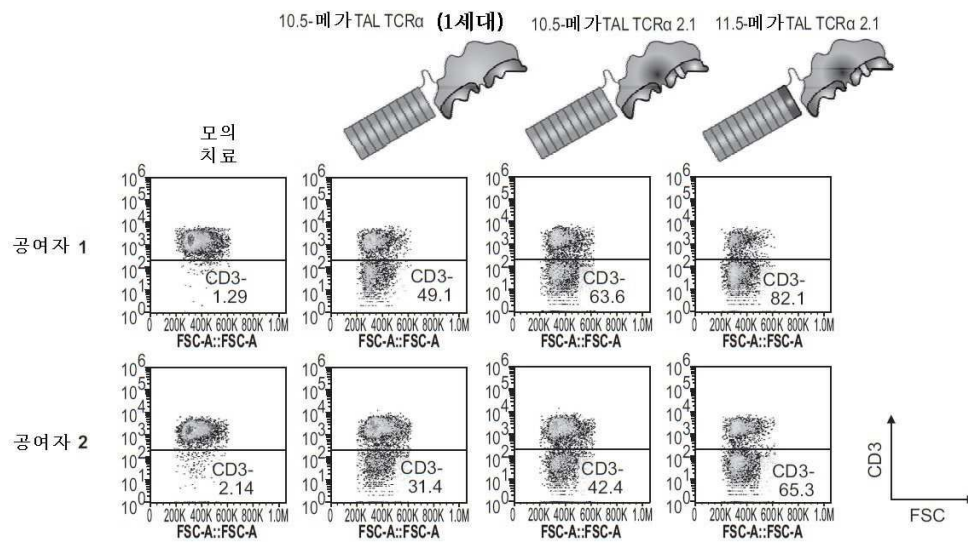
033016 TCR 메가TAL 표적이탈 앰플리콘 분석



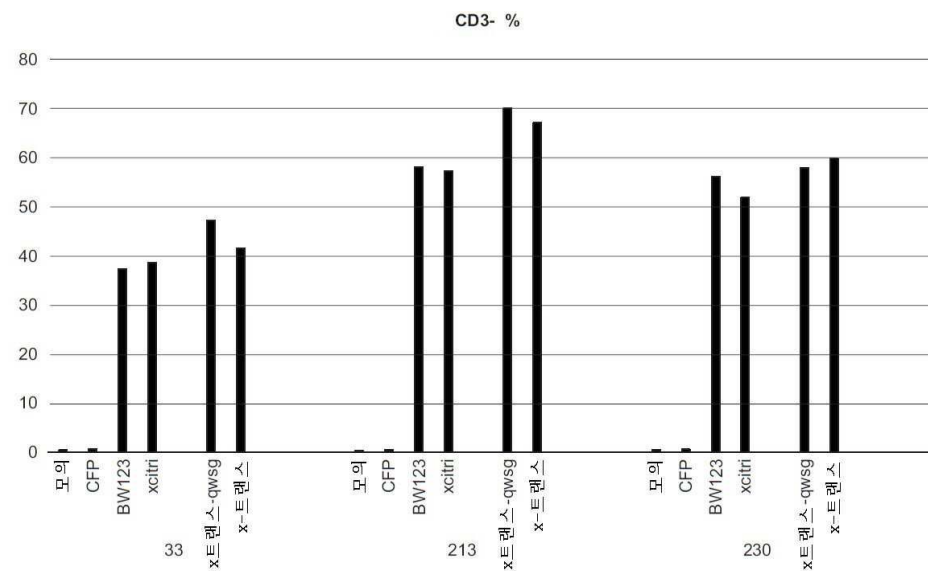
도면8a



도면8b



도면9



## 서열 목록

### SEQUENCE LISTING

<110> BLUEBIRD BIO, INC.

<120> TCRα HOMING ENDONUCLEASE VARIANTS

<130> IPA190499-US

<150> US 62/406,689

<151> 2016-10-11

<150> US 62/414,266

<151> 2016-10-28

<160> 69

<170> PatentIn version 3.5

<210> 1

<211> 303

<212> PRT

<213> Ophiostoma novo-ulmi subsp. americana (mitochondrion)

<400> 1

Met Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr

1 5 10 15

Gly Phe Ala Asp Ala Glu Gly Ser Phe Leu Leu Arg Ile Arg Asn Asn

20 25 30

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

35 40 45

Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp

50 55 60

Lys Val Gly Val Ile Ala Asn Ser Gly Asp Asn Ala Val Ser Leu Lys

65 70 75 80

Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys

85 90 95

Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Met Leu Phe Lys Gln

100 105 110

Ala Phe Cys Val Met Glu Asn Lys Glu His Leu Lys Ile Asn Gly Ile

115 120 125

Lys Glu Leu Val Arg Ile Lys Ala Lys Leu Asn Trp Gly Leu Thr Asp

130 135 140

Glu Leu Lys Lys Ala Phe Pro Glu Ile Ile Ser Lys Glu Arg Ser Leu

145 150 155 160

Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser

165 170 175

Gly Glu Gly Cys Phe Phe Val Asn Leu Ile Lys Ser Lys Ser Lys Leu

180 185 190

Gly Val Gln Val Gln Leu Val Phe Ser Ile Thr Gln His Ile Lys Asp



195                                      200                                      205  
 Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile  
 210                                      215                                      220  
 Lys Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Asp Phe Val Val Thr  
 225                                      230                                      235                                      240  
 Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn  
 245                                      250                                      255  
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val  
 260                                      265                                      270  
  
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp  
 275                                      280                                      285  
 Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe  
 290                                      295                                      300  
 <210> 2  
 <211> 303  
 <212> PRT  
 <213> Ophiostoma novo-ulmi subsp. americana (mitochondrion)  
 <400> 2  
 Met Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr  
 1                                      5                                      10                                      15  
 Gly Phe Ala Asp Ala Glu Gly Ser Phe Leu Leu Arg Ile Arg Asn Asn  
  
 20                                      25                                      30  
 Asn Lys Ser Ser Val Gly Tyr Ser Thr Glu Leu Gly Phe Gln Ile Thr  
 35                                      40                                      45  
 Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp  
 50                                      55                                      60  
 Lys Val Gly Val Ile Ala Asn Ser Gly Asp Asn Ala Val Ser Leu Lys  
 65                                      70                                      75                                      80  
 Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys  
  
 85                                      90                                      95  
 Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln

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

<210> 3  
 <211> 303  
 <212> PRT  
 <213> Ophiostoma novo-ulmi subsp. americana (mitochondrion)  
 <220><221> MOD\_RES  
 <222> (1)..(3)  
 <223> Any amino acid or absent

<400> 3

Xaa Xaa Xaa Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr  
1 5 10 15  
Gly Phe Ala Asp Ala Glu Gly Ser Phe Leu Leu Arg Ile Arg Asn Asn  
20 25 30  
Asn Lys Ser Ser Val Gly Tyr Ser Thr Glu Leu Gly Phe Gln Ile Thr  
35 40 45  
Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp  
50 55 60  
Lys Val Gly Val Ile Ala Asn Ser Gly Asp Asn Ala Val Ser Leu Lys  
65 70 75 80  
Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys  
85 90 95  
Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln  
100 105 110  
Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile  
115 120 125  
Lys Glu Leu Val Arg Ile Lys Ala Lys Leu Asn Trp Gly Leu Thr Asp  
130 135 140  
Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Ser Leu  
145 150 155 160  
Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser  
165 170 175  
Gly Glu Gly Cys Phe Phe Val Asn Leu Ile Lys Ser Lys Ser Lys Leu  
180 185 190  
Gly Val Gln Val Gln Leu Val Phe Ser Ile Thr Gln His Ile Lys Asp  
195 200 205  
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile  
210 215 220  
Lys Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Asp Phe Val Val Thr  
225 230 235 240

Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn  
 245 250 255  
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val  
 260 265 270  
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp

275 280 285  
 Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe  
 290 295 300

<210> 4

<211> 303

<212> PRT

<213> Ophiostoma novo-ulmi subsp. americana (mitochondrion)

<220><221> MOD\_RES

<222> (1)..(4)

<223> Any amino acid or absent

<220><221> MOD\_RES

<222> (302)..(303)

<223> Any amino acid or absent

<400> 4

Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr

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

65 70 75 80  
 Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys  
 85 90 95  
 Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln

100                      105                      110  
 Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile  
 115                      120                      125  
 Lys Glu Leu Val Arg Ile Lys Ala Lys Leu Asn Trp Gly Leu Thr Asp  
  
 130                      135                      140  
 Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Ser Leu  
 145                      150                      155                      160  
 Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser  
 165                      170                      175  
 Gly Glu Gly Cys Phe Phe Val Asn Leu Ile Lys Ser Lys Ser Lys Leu  
 180                      185                      190  
 Gly Val Gln Val Gln Leu Val Phe Ser Ile Thr Gln His Ile Lys Asp  
  
 195                      200                      205  
 Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile  
 210                      215                      220  
 Lys Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Asp Phe Val Val Thr  
 225                      230                      235                      240  
 Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn  
 245                      250                      255  
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val  
  
 260                      265                      270  
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp  
 275                      280                      285  
 Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa  
 290                      295                      300  
 <210> 5  
 <211> 303  
 <212> PRT  
 <213> Ophiostoma novo-ulmi subsp. americana (mitochondrion)  
 <220><221> MOD\_RES  
 <222> (1)..(8)  
 <223> Any amino acid or absent



<220><221> MOD\_RES

<222> (302)..(303)

<223> Any amino acid or absent

<400> 5

Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Ser Ile Asn Pro Trp Ile Leu Thr

1 5 10 15

Gly Phe Ala Asp Ala Glu Gly Ser Phe Leu Leu Arg Ile Arg Asn Asn

20 25 30

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

35 40 45

Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp

50 55 60

Lys Val Gly Val Ile Ala Asn Ser Gly Asp Asn Ala Val Ser Leu Lys

65 70 75 80

Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys

85 90 95

Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln

100 105 110

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

115 120 125

Lys Glu Leu Val Arg Ile Lys Ala Lys Leu Asn Trp Gly Leu Thr Asp

130 135 140

Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Ser Leu

145 150 155 160

Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser

165 170 175

Gly Glu Gly Cys Phe Phe Val Asn Leu Ile Lys Ser Lys Ser Lys Leu

180 185 190

Gly Val Gln Val Gln Leu Val Phe Ser Ile Thr Gln His Ile Lys Asp

195 200 205

Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Tyr Ile

210 215 220  
Lys Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Asp Phe Val Val Thr  
225 230 235 240  
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn

245 250 255  
Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val  
260 265 270  
Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp  
275 280 285  
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa  
290 295 300

<210> 6

<211> 303

<212> PRT

<213> Artificial Sequence

<220><223> Synthesized I-OnuI variant

<400> 6

Met Ala Tyr Met Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr  
1 5 10 15  
Gly Phe Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg  
20 25 30  
Asn Asn Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr  
35 40 45  
Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp

50 55 60  
Lys Val Gly Lys Ile Thr Asn Ser Gly Asp Arg Ala Val Met Leu Arg  
65 70 75 80  
Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys  
85 90 95  
Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln  
100 105 110

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

115 120 125  
Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Thr Asp  
130 135 140  
Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu  
145 150 155 160  
Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser  
165 170 175  
Gly Asp Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Asn Ala

180 185 190  
Lys Val Tyr Val Gly Leu Arg Phe Ser Ile Ser Gln His Ile Arg Asp  
195 200 205  
Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile  
210 215 220  
Trp Glu Lys Asn Lys Ser Glu Phe Ser Trp Leu Glu Phe Val Val Thr  
225 230 235 240  
Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn

245 250 255  
Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val  
260 265 270  
Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp  
275 280 285  
Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe  
290 295 300

<210> 7

<211> 303

<212> PRT

<213> Artificial Sequence

<220><223> Synthesized I-OnuI LHE variant

<220><221> MOD\_RES

<222> (1)..(4)

<223> Any amino acid or absent

<220><221> MOD\_RES

<222> (302)..(303)

<223> Any amino acid or absent

<400> 7

Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr

1 5 10 15

Gly Phe Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg

20 25 30

Asn Asn Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr

35 40 45

Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp

50 55 60

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

65 70 75 80

Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys

85 90 95

Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln

100 105 110

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

115 120 125

Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp

130 135 140

Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu

145 150 155 160

Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ala

165 170 175

Gly Glu Gly His Phe Gly Val Asn Leu Lys Lys Val Lys Gly Thr Ala

180 185 190

Lys Val Tyr Val Gly Leu Arg Phe Ala Ile Ser Gln His Ile Arg Asp

195 200 205

Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile

210 215 220  
 Arg Glu Lys Asn Lys Ser Glu Phe Arg Trp Leu Glu Phe Glu Val Thr  
 225 230 235 240  
 Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn  
 245 250 255  
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val  
 260 265 270  
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp  
 275 280 285  
 Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Xaa Xaa

290 295 300  
 <210> 8  
 <211> 303  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Synthesized I-OnuI LHE variant  
 <220><221> MOD\_RES  
 <222> (1)..(4)  
 <223> Any amino acid or absent  
 <220><221> MOD\_RES  
 <222> (302)..(303)  
 <223> Any amino acid or absent  
 <400> 8

Xaa Xaa Xaa Xaa Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr  
 1 5 10 15  
 Gly Phe Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg

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



65					70				75				80			
Val	Thr	Arg	Phe	Glu	Asp	Leu	Lys	Val	Ile	Ile	Asp	His	Phe	Glu	Lys	
				85				90				95				
Tyr	Pro	Leu	Ile	Thr	Gln	Lys	Leu	Gly	Asp	Tyr	Lys	Leu	Phe	Lys	Gln	
				100				105				110				
Ala	Phe	Ser	Val	Met	Glu	Asn	Lys	Glu	His	Leu	Lys	Glu	Asn	Gly	Ile	
				115				120				125				
Lys	Glu	Leu	Val	Arg	Ile	Lys	Ala	Lys	Met	Asn	Trp	Gly	Leu	Asn	Asp	
				130				135				140				
Glu	Leu	Lys	Lys	Ala	Phe	Pro	Glu	Asn	Ile	Ser	Lys	Glu	Arg	Pro	Leu	
145				150				155				160				
Ile	Asn	Lys	Asn	Ile	Pro	Asn	Phe	Lys	Trp	Leu	Ala	Gly	Phe	Thr	Ala	
				165				170				175				
Gly	Asp	Gly	His	Phe	Gly	Val	Asn	Leu	Lys	Lys	Val	Lys	Gly	Thr	Ala	
				180				185				190				
Lys	Val	Tyr	Val	Gly	Leu	Arg	Phe	Ala	Ile	Ser	Gln	His	Ile	Arg	Asp	
				195				200				205				
Lys	Asn	Leu	Met	Asn	Ser	Leu	Ile	Thr	Tyr	Leu	Gly	Cys	Gly	Ser	Ile	
210				215				220								
Arg	Glu	Lys	Asn	Lys	Ser	Glu	Phe	Arg	Trp	Leu	Glu	Phe	Glu	Val	Thr	
225				230				235				240				
Lys	Phe	Ser	Asp	Ile	Asn	Asp	Lys	Ile	Ile	Pro	Val	Phe	Gln	Glu	Asn	
				245				250				255				
Thr	Leu	Ile	Gly	Val	Lys	Leu	Glu	Asp	Phe	Glu	Asp	Trp	Cys	Lys	Val	
				260				265				270				
Ala	Lys	Leu	Ile	Glu	Glu	Lys	Lys	His	Leu	Thr	Glu	Ser	Gly	Leu	Asp	
275				280				285								
Glu	Ile	Lys	Lys	Ile	Lys	Leu	Asn	Met	Asn	Lys	Gly	Arg	Xaa	Xaa		
290				295				300								
<210> 9																
<211> 877																

<212> PRT

<213> Artificial Sequence

<220><223> Synthesized I-OnuI variant megaTAL construct

<400> 9

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg

1 5 10 15

Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val

20 25 30

Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe

35 40 45

Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly

50 55 60

Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala

65 70 75 80

Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg

85 90 95

Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro

100 105 110

Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly

115 120 125

Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly

130 135 140

Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn

145 150 155 160

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

165 170 175

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

180 185 190

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

195 200 205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210 215 220  
 Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg  
 225 230 235 240  
 Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val  
 245 250 255  
 Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val  
 260 265 270  
 Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp  
  
 275 280 285  
 Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu  
 290 295 300  
 Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr  
 305 310 315 320  
 Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala  
 325 330 335  
 Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly  
  
 340 345 350  
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys  
 355 360 365  
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp  
 370 375 380  
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly  
 385 390 395 400  
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys  
  
 405 410 415  
 Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn  
 420 425 430  
 Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val  
 435 440 445  
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala  
 450 455 460

Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu

465 470 475 480

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

485 490 495

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

500 505 510

Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His

515 520 525

Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val

530 535 540

Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg

545 550 555 560

Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Val Gly

565 570 575

Gly Ser Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe

580 585 590

Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg Asn Asn

595 600 605

Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr Leu His

610 615 620

Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp Lys Val

625 630 635 640

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

645 650 655

Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro

660 665 670

Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe

675 680 685

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

690 695 700

Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Thr Asp Glu Leu

705                      710                      715                      720  
 Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn  
  
                          725                      730                      735  
 Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser Gly Asp  
                          740                      745                      750  
 Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Asn Ala Lys Val  
                          755                      760                      765  
 Tyr Val Gly Leu Arg Phe Ser Ile Ser Gln His Ile Arg Asp Lys Asn  
                          770                      775                      780  
 Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile Trp Glu  
  
 785                      790                      795                      800  
 Lys Asn Lys Ser Glu Phe Ser Trp Leu Glu Phe Val Val Thr Lys Phe  
                          805                      810                      815  
 Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu  
                          820                      825                      830  
 Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys  
                          835                      840                      845  
 Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile  
  
                          850                      855                      860  
 Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe  
 865                      870                      875  
 <210> 10  
 <211> 877  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Synthesized megaTAL construct  
 <400> 10  
 Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg  
 1                      5                      10                      15  
 Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val  
                          20                      25                      30



Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe  
 35 40 45  
 Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly  
 50 55 60  
 Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala  
 65 70 75 80  
 Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg  
 85 90 95  
  
 Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro  
 100 105 110  
 Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly  
 115 120 125  
 Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly  
 130 135 140  
 Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn  
 145 150 155 160  
  
 Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val  
 165 170 175  
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala  
 180 185 190  
 Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu  
 195 200 205  
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala  
 210 215 220  
  
 Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg  
 225 230 235 240  
 Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val  
 245 250 255  
 Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val  
 260 265 270  
 Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp

275                                      280                                      285  
 Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu  
 290                                      295                                      300  
 Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr  
 305                                      310                                      315                                      320  
 Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala  
 325                                      330                                      335  
 Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly  
 340                                      345                                      350  
  
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys  
 355                                      360                                      365  
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp  
 370                                      375                                      380  
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly  
 385                                      390                                      395                                      400  
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys  
 405                                      410                                      415  
  
 Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn  
 420                                      425                                      430  
 Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val  
 435                                      440                                      445  
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala  
 450                                      455                                      460  
 Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu  
 465                                      470                                      475                                      480  
  
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala  
 485                                      490                                      495  
 Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala  
 500                                      505                                      510  
 Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His  
 515                                      520                                      525

Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val  
 530 535 540

Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg  
 545 550 555 560

Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Val Gly  
 565 570 575

Gly Ser Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe  
 580 585 590

Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg Asn Asn  
 595 600 605

Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr Leu His  
 610 615 620

Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp Lys Val  
 625 630 635 640

Gly Lys Ile Thr Asn Ser Gly Asp Arg Ala Val Met Leu Arg Val Thr  
 645 650 655

Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro  
 660 665 670

Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe  
 675 680 685

Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu  
 690 695 700

Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Thr Asp Glu Leu  
 705 710 715 720

Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn  
 725 730 735

Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser Gly Asp  
 740 745 750

Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Asn Ala Lys Val  
 755 760 765

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

770                      775                      780  
 Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile Trp Glu  
 785                      790                      795                      800

Lys Asn Lys Ser Glu Phe Ser Trp Leu Glu Phe Val Val Thr Lys Phe  
                                  805                      810                      815  
 Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu  
                                  820                      825                      830  
 Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys  
                                  835                      840                      845  
 Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile  
                                  850                      855                      860

Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe  
 865                      870                      875

<210> 11

<211> 875

<212> PRT

<213> Artificial Sequence

<220><223> Synthesized megaTAL construct

<400> 11

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

                                 35                      40                      45  
 Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly  
                                  50                      55                      60  
 Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala  
 65                      70                      75                      80  
 Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg  
                                  85                      90                      95

Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro

100 105 110

Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly

115 120 125

Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly

130 135 140

Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn

145 150 155 160

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

165 170 175

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

180 185 190

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

195 200 205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210 215 220

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

225 230 235 240

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val

245 250 255

Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val

260 265 270

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp

275 280 285

Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu

290 295 300

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr

305 310 315 320

Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala

325 330 335

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly

340                      345                      350  
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys  
  
 355                      360                      365  
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp  
 370                      375                      380  
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly  
 385                      390                      395                      400  
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys  
 405                      410                      415  
 Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn  
  
 420                      425                      430  
 Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val  
 435                      440                      445  
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala  
 450                      455                      460  
 Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu  
 465                      470                      475                      480  
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala  
  
 485                      490                      495  
 Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala  
 500                      505                      510  
 Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His  
 515                      520                      525  
 Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val  
 530                      535                      540  
 Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg  
  
 545                      550                      555                      560  
 Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Val Gly  
 565                      570                      575  
 Gly Ser Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe  
 580                      585                      590



Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg Asn Asn  
595 600 605  
Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr Leu His  
610 615 620  
Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp Lys Val  
625 630 635 640  
Gly Lys Ile Thr Asn Ser Ser Asp Arg Ala Val Met Leu Arg Val Thr  
645 650 655  
Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro  
660 665 670  
Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe  
675 680 685  
Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu  
690 695 700  
Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp Glu Leu  
705 710 715 720  
Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn  
725 730 735  
Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ala Gly Asp  
740 745 750  
Gly His Phe Gly Val Asn Leu Lys Lys Val Lys Gly Thr Ala Lys Val  
755 760 765  
Tyr Val Gly Leu Arg Phe Ala Ile Ser Gln His Ile Arg Asp Lys Asn  
770 775 780  
Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile Arg Glu  
785 790 795 800  
Lys Asn Lys Ser Glu Phe Arg Trp Leu Glu Phe Glu Val Thr Lys Phe  
805 810 815  
Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu  
820 825 830  
Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys

835 840 845  
 Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile  
 850 855 860  
 Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg  
 865 870 875

<210> 12

<211> 909

<212> PRT

<213> Artificial Sequence

<220><223> Synthesized megaTAL construct

<400> 12

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Val Asp Leu Arg  
 1 5 10 15  
 Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Val  
 20 25 30  
 Arg Ser Thr Val Ala Gln His His Glu Ala Leu Val Gly His Gly Phe  
 35 40 45

Thr His Ala His Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly  
 50 55 60  
 Thr Val Ala Val Thr Tyr Gln His Ile Ile Thr Ala Leu Pro Glu Ala  
 65 70 75 80  
 Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg  
 85 90 95  
 Ala Leu Glu Ala Leu Leu Thr Asp Ala Gly Glu Leu Arg Gly Pro Pro  
 100 105 110

Leu Gln Leu Asp Thr Gly Gln Leu Val Lys Ile Ala Lys Arg Gly Gly  
 115 120 125  
 Val Thr Ala Met Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly  
 130 135 140  
 Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn  
 145 150 155 160  
 Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

165	170	175
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala		
180	185	190
Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu		
195	200	205
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala		
210	215	220
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg		
225	230	235
		240
Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val		
245	250	255
Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val		
260	265	270
Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp		
275	280	285
Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu		
290	295	300
Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr		
305	310	315
Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala		
325	330	335
Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly		
340	345	350
Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys		
355	360	365
Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp		
370	375	380
His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly		
385	390	395
Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys		
405	410	415

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn  
420 425 430

Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val  
435 440 445

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala  
450 455 460

Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu  
465 470 475 480

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala  
485 490 495

Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg  
500 505 510

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val  
515 520 525

Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu Ser Ile  
530 535 540

Val Ala Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn  
545 550 555 560

Asp His Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp  
565 570 575

Ala Val Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val  
580 585 590

Asn Arg Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg  
595 600 605

Val Gly Gly Ser Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr  
610 615 620

Gly Phe Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg  
625 630 635 640

Asn Asn Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr  
645 650 655

Leu His Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp

660                      665                      670  
 Lys Val Gly Lys Ile Thr Asn Ser Ser Asp Arg Ala Val Met Leu Arg  
 675                      680                      685  
  
 Val Thr Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys  
 690                      695                      700  
 Tyr Pro Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln  
 705                      710                      715                      720  
 Ala Phe Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile  
 725                      730                      735  
 Lys Glu Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Asn Asp  
 740                      745                      750  
  
 Glu Leu Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu  
 755                      760                      765  
 Ile Asn Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ala  
 770                      775                      780  
 Gly Asp Gly His Phe Gly Val Asn Leu Lys Lys Val Lys Gly Thr Ala  
 785                      790                      795                      800  
 Lys Val Tyr Val Gly Leu Arg Phe Ala Ile Ser Gln His Ile Arg Asp  
 805                      810                      815  
  
 Lys Asn Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile  
 820                      825                      830  
 Arg Glu Lys Asn Lys Ser Glu Phe Arg Trp Leu Glu Phe Glu Val Thr  
 835                      840                      845  
 Lys Phe Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn  
 850                      855                      860  
 Thr Leu Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val  
 865                      870                      875                      880  
  
 Ala Lys Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp  
 885                      890                      895  
 Glu Ile Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg  
 900                      905

<210> 13

<211> 2634

<212> RNA

<213> Artificial Sequence

<220><223> Synthesized I-OnuI variant megaTAL construct

<400> 13

augggauccg cgccaccuaa gaagaaacgc aaagucgugg aucuacgcac gcucggcuac	60
agucagcagc agcaagagaa gaucaaaccg aaggugcguu cgacaguggc gcagcaccac	120
gaggcacugg ugggccauagg guuuacacac gcgcacaucg uugcgcucag ccaacacccg	180
gcagcguuag ggaccgucgc ugucacguau cagcacauaa ucacggcguu gccagaggcg	240
acacacgaag acaucguugg cgucggcaaa cagugguccg gcgcacgcgc ccuggaggcc	300
uugcucacgg augcggggga guugagaggu ccgccguuac aguuggacac aggccaacuu	360
gugaagauug caaacgugg cggcgugacc gcaauggagg cagugcaugc aucgcgcaau	420
gcacugacgg gugccccu gaaccugacc ccggaccaag ugguggcuaug cgccagcaac	480
auuggcggca agcaagcguu cgaaacggug cagcggcugu ugccggugcu gugccaggac	540
cauggccuga ccccggaaca agugguggc uauccagca acauuggcgg caagcaagcg	600
cucgaaacgg ugcagcggcu guugccggug cugugccagg accauggccu gaccccggaac	660
caaugggugg cuaucgccag caacauuggc ggcaagcaag cgcucgaaac ggugcagcgg	720
cuguugccgg ugcugugcca ggaccauggc cugaccccg accaaguggu ggcuaucgcc	780
agcaacggug gcggcaagca agcgcucgaa acggugcagc ggcuuugcc ggugcugugc	840
caggaccaug gccugacucc ggaccaaug guggcuauug ccagccacga uggcggcaag	900
caagcgcucg aaacggugca gcggcuguug ccggugcugu gccaggacca uggccugacu	960
ccggaccaag ugguggcuaug cgccagccac gauggcggca agcaagcguu cgaaacggug	1020
cagcggcugu ugccggugcu gugccaggac cauggccuga ccccggaaca agugguggc	1080
aucgccagca acauuggcgg caagcaagcg cucgaaacgg ugcagcggcu guugccggug	1140
cugugccagg accauggccu gaccccggaac caaugggugg cuaucgccag caacauggc	1200
ggcaagcaag cgcucgaaac ggugcagcgg cuguugccgg ugcugugcca ggaccauggc	1260
cugaccccg accaaguggu ggcuaucgcc agcaacggug gcggcaagca agcgcucgaa	1320
acggugcagc ggcuuugcc ggugcugugc caggaccaug gccugacccc ggaccaaug	1380
guggcuauug ccagcaacaa uggcggcaag caagcgcucg aaacggugca gcggcuguug	1440
ccggugcugu gccaggacca uggccugacc ccggaccaag ugguggcuaug cgccagcaac	1500



auuggcggca agcaagcgcu cgaaagcauu guggcccagc ugagccggcc ugauccggcg 1560  
uuggccgcgu ugaccaacga ccaccucguc gccuuggccu gccucggcgg acguccugcc 1620  
auggaugcag ugaaaaaggg auugccgcac gcgccggaau ugaucagaag agucaaucgc 1680  
cguauuggcg aacgcacguc ccaucgcguu gcgauaucua gagugggagg aagcucgcgc 1740  
agagagucca ucaacccaug gauucugacu gguuucgcug augccgaagg aucauucuaa 1800

cuagacaucc gcaaccgaaa caacgaaagc aacagauacc gaacuucgcu gagauuccag 1860  
aucaccucg acaacaagga caaaucgauu cuggagaaua uccagucgac uuggaagguc 1920  
ggcaagauc caaacagcgg cgacagagcc gucaugcuga gggucacccg uuucgaagau 1980  
uugaaaguga uuaucgacca cuucgagaaa uauccgcuga uuaccagaa auugggcgau 2040  
uacaaguugu uuaaacaggc auucagcguc auggagaaca aagaacauu uaaggagaau 2100  
gggauuaagg agcucguacg aaucaaagcu aagaugaaau ggggucucac ugacgaauug 2160  
aaaaaagcau uuccagagaa cauuagcaaa gagcgccccc uuaucaauaa gaacauuccg 2220

aaauucaaau ggcuggcugg auucacauu ggugauggcu acuucggcgu gaaucaaaaa 2280  
aagguaaagg gcaacgcaaa gguauacgug ggacugagau ucucaaucuc acagcacauc 2340  
agagacaaga accugaugaa uucauuguaa acauaccuag gcugugguuc caucugggag 2400  
aagaacaagu cugaguucag uggcucgag uucgucguua ccaaaucag cgauaucaac 2460  
gacaagauc uuccgguaau ccaggaaaau acucugauug gcgucaaacu cgaggacuuu 2520  
gaagauuggu gcaagguugc caauugauc gaagagaaga aacaccugac cgaauccggu 2580  
uuggaugaga uuaagaaaau caagcugaac augaacaag gucgugucuu cuga 2634

<210> 14

<211> 2634

<212> RNA

<213> Artificial Sequence

<220><223> Synthesized megaTAL construct

<400> 14

augggaucg cgccaccuaa gaagaaacgc aaagucgugg aucuacgcac gcucggcuac 60  
agucagcagc agcaagagaa gaucaaaccg aaggugcguu cgacaguggc gcagaccac 120  
gaggcacugg ugggccaugg guuuacacac gcgcacaucg uugcgucag ccaacacccg 180  
gcagcguuag ggaccgucgc ugucacguau cagcacauaa ucacggcguu gccagaggcg 240  
acacacgaag acaucguugg gcucggcaaa cagugguccg gcgcacgcgc ccuggaggcc 300  
uugcucacgg augcggggga guugagaggu ccgccguuac aguuggacac aggccaacuu 360

gugaagauug caaaacgugg cggcgugacc gcaauggagg cagugcaugc aucgcgcaau 420  
 gcacugacgg gugcccccuc gaaccugacc ccggaccaag ugguggcuauc cgccagcaac 480  
 auuggcggca agcaagcgcu cgaaacggug cagcggcugu ugccggugcu gugccaggac 540  
 cauggccuga ccccggaaca agugguggcuc aucgccagca acauuggcgg caagcaagcg 600  
 cucgaaacgg ugcagcggcu guugccggug cugugccagg accauggccu gaccccggaac 660  
 caaugggugg cuaucgccag caacauuggc ggcaagcaag cgcucgaaac ggugcagcgg 720  
 cuguugccgg ugcugugcca ggaccauggc cugaccccg accaaguggu ggcuaucgcc 780  
  
 agcaacggug ggcgcaagca agcgcucgaa acggugcagc ggcuugucc ggugcugugc 840  
 caggaccaug gccugacucc ggaccaaug guggcuauuc ccagccacga uggcggcaag 900  
 caagcgcucg aaacggugca ggcgcuguug ccggugcugu gccaggacca ugccugacu 960  
 ccggaccaag ugguggcuauc cgccagccac gauggcggca agcaagcgcu cgaacggug 1020  
 cagcggcugu ugccggugcu gugccaggac cauggccuga ccccggaaca agugguggcuc 1080  
 aucgccagca acauuggcgg caagcaagcg cucgaaacgg ugcagcggcu guugccggug 1140  
 cugugccagg accauggccu gaccccggaac caaugggugg cuaucgccag caacaauaggc 1200  
  
 ggcaagcaag cgcucgaaac ggugcagcgg cuguugccgg ugcugugcca ggaccauggc 1260  
 cugaccccg accaaguggu ggcuaucgcc agcaacggug ggcgcaagca agcgcucgaa 1320  
 acggugcagc ggcuugucc ggugcugugc caggaccaug gccugacccc ggaccaaug 1380  
 guggcuauuc ccagcaacaa uggcggcaag caagcgcucg aaacggugca ggcgcuguug 1440  
 ccggugcugu gccaggacca ugccugacc ccggaccaag ugguggcuauc cgccagcaac 1500  
 auuggcggca agcaagcgcu cgaaagcauu guggcccagc ugagccggcc ugaucggcg 1560  
 uggcccgcu ugaccaacga ccaccucguc gccuuggccu gccucggcgg acguccugcc 1620  
  
 auggaugcag ugaanaagg auugccgcac ggcgggaau ugaucagaag agucaaucgc 1680  
 cguaauuggc aacgcacguc ccaucgcguu gcgaauucua gagugggagg aagcucgcgc 1740  
 agagagucca ucaaccuag gauucgacu gguuucguc augccgaagg acauucuaa 1800  
 cuagacaucc gcaaccgaaa caacgaaagc aacagauacc gaacuucgcu gagauuccag 1860  
 aucaccucg acaacaagga caaaucgauu cuggagaaua uccaguccac uuggaagguc 1920  
 ggcaagauc caaacagcag ugacagagcc guaucguga gggucacccg uuucgaagau 1980  
 uugaaaguga uuaucgacca cuucgagaaa uaucgcuga uuaccagaa auugggcgau 2040  
  
 uacaaguugu uuaaacaggc auucagcguc auggagaaua aagaacauuc uaaggagaau 2100  
 gggauuaagg agcucguacg aaucaaagcu aagaugaaau ggggucuaa ugacgaauug 2160  
 aaaaaagcau uuccagagaa cauuagcaaa gagcgcccc uuaucuaa gaacauuccg 2220

aauuucaaa ggucggcugg auucacagcu ggugaaggcc auuucggcgu gaaucuaaaa 2280  
 aagguaaagg gcaccgcaaa gguauacgug ggacugagau ucgcuaucuc acagcacauc 2340  
 agagacaaga accugaugaa uucauugaua acauaccuag gcugugguuc caucagagag 2400  
 aagaacaagu cugaguucag auggcucgag uucgaagua ccaaaucag cgauaucaac 2460  
  
 gacaagauca uuccgguaau ccaggaauu acucugauug gcgucaaacu cgaggacuuu 2520  
 gaagauuggu gcaagguugc caaaugauc gaagagaaga aacaccugac cgaaucgggu 2580  
 uuggaugaga uuaagaaaa caagcugaac augaacaag gucgugucuu cuga 2634  
 <210> 15  
 <211> 2628  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Synthesized megaTAL construct  
 <400> 15  
 augggauccg cgccaccuaa gaagaaacgc aaagucgugg aucuacgcac gcucggcuac 60  
 agucagcagc agcaagagaa gaucaaacgc aaggugcguu cgacaguggc gcagcaccac 120  
 gaggcacugg ugggccauug guuuacacac gcgcacaucg uugcgcucag ccaacacccg 180  
  
 gcagcguuag ggaccgucgc ugucacguau cagcacauaa ucacggcguu gccagaggcg 240  
 acacacgaag acaucguugg cgucggcaaa cagugguccg gcgcacgcgc ccuggaggcc 300  
 uugcucacgg augcggggga guugagaggu ccgccguuac aguuggacac aggccaaacu 360  
 gugaagauug caaaacgugg cggcgugacc gcaauggagg cagugcaugc aucgcgcaau 420  
 gcacugacgg gugcccccua gaaccugacc ccggaccaag ugguggcuaa cgccagcaac 480  
 auuggcggca agcaagcguu cgaaacggug cagcggcugu ugccggugcu gugccaggac 540  
 cauggccuga ccccggaaca agugguggcua aucgccagca acauuggcgg caagcaagcg 600  
  
 cucgaaacgg ugcagcggcu guugccggug cugugccagg accauggccu gaccccggaac 660  
 caaguggugg cuaucgccag caacauuggc ggcaagcaag gcucgaaac ggugcagcgg 720  
 cuguugccgg ugcugugcca ggaccauggc cugaccccgg accaaguggu ggcuaucgcc 780  
 agcaacggug gcggcaagca agcgcucgaa acggugcagc ggcuugucc ggugcuguc 840  
 caggaccaug gccugacucc ggaccaaug guggcuauac ccagccacga uggcggcaag 900  
 caagcgcucg aaacggugca gcggcuguug ccggugcugu gccaggacca ugccugacu 960  
 ccggaccaag ugguggcuaa cgccagccac gauggcggca agcaagcguu cgaaacggug 1020  
  
 cagcggcugu ugccggugcu gugccaggac cauggccuga ccccggaaca agugguggcua 1080

aucgccagca acauuggcgg caagcaagcg cucgaaacgg ugcagcggcu guugccggug 1140  
 cugugccagg accauggccu gaccccgac caaguggugg cuaucgccag caacauggc 1200  
 ggcaagcaag cgcucgaaac ggugcagcgg cuguugccgg ugcugugcca ggaccauggc 1260  
 cugaccccg accaaguggu ggcuauccg agcaacggug gcggcaagca agcgcucgaa 1320  
 acggugcagc ggcuuguucc ggugcugugc caggaccaug gccugacccc ggaccaaug 1380  
 guggcuauccg ccagcaacaa uggcggcaag caagcgcucg aaacggugca gcggcuguug 1440

ccggugcugu gccaggacca uggccugacc ccggaccaag ugguggcuau cgccagcaac 1500  
 auuggcggca agcaagcgu cgaagcauu guggcccagc ugagccggcc ugaucggcg 1560  
 uugcccgcu ugaccaacga ccaccucguc gccuuggccu gccucggcgg acguccugcc 1620  
 auggaugcag ugaanaaggg auugccgcac gcgccggaau ugaucagaag agucaaucgc 1680  
 cguauuggcg aacgcacguc ccaucgcuu gcgauaucua gagugggagg aagcucgcgc 1740  
 agagagucca ucaaccgaug gauucugacu gguuucgucg augccgaagg aucauucua 1800  
 cuagacaucc gcaaccgaaa caacgaaagc aacagauacc gaacuucgu gagauuccag 1860

aucaccucg acaacaagga caaaucgauu cuggagaaua uccaguccac uuggaagguc 1920  
 ggcaagauca caaacagcag ugacagagcc gucaugcuga gggucacccg uuucgaagau 1980  
 uugaaaguga uuaucgacca cuucgagaaa uauccgcuga uuaccagaa auugggcgau 2040  
 uacaaguugu uuaaacaggc auucagcgu auggagaaua aagaacauu uaaggagaau 2100  
 gggauuaagg agcucguacg aaucacagcu aagaugaaau ggggucucua ugacgaauug 2160  
 aaaaaagcau uuccagagaa cauugcaaaa gagcgcccc uuaucaaua gaacauuccg 2220  
 aauuucuuu ggcuggcugg auucacagcu ggugauggcc auuucggcgu gaaucuaaaa 2280

aagguaaagg gcaccgcaaa gguauacgug ggacugagau ucgcuaucuc acagcacauc 2340  
 agagacaaga accugaagaa uucauugaua acauaccuag gcugugguuc caucagagag 2400  
 aagaacaagu cugaguucag auggcucgag uucgaagua ccaauucag cgauaucaac 2460  
 gacaagauca uuccgguuuu ccaggaaaau acucugauug gcgucaaacu cgaggacuuu 2520  
 gaagauuggu gcaagguugc caauugauc gaagagaaga aacaccugac cgaauccggu 2580  
 uuggaugaga uuaagaaaau caagcugaac augaacaag gucguuga 2628

<210> 16

<211> 2733

<212> RNA

<213> Artificial Sequence

<220><223> Synthesized megaTAL construct

<400> 16

augggauccg cccuccgaa gaagaagcgg aaagucgugg accuccggac ccuggguuac	60
ucucagcagc agcaggagaa gaucaagccg aaggugcggu cgacuguggc ccagcaucac	120
gaggccucgg ugggacacgg cuucacccac gccacauug uggccugag ccagcacccg	180
gcagcgucgg gaaccguggc cgugaccuac cagcacauca uuacugcccu gccugaagcg	240
accacagagg auaucguggg cgucggaaag cagugguccg gagccagagc cuuggaggcu	300
cugcugacug acgccggaga gcugcggggc ccgcccugc aacuggauac cggccagcuc	360
gugaaaaucg ccaagagagg aggagugacc gccauggaag ccgugcaugc aucccgaau	420
gcacugacug gugcaccctu gaaccucacu ccugaccagg ucgucguau cgcaagcaac	480
aucggaggga aacaagcucu cgagacagug cagcgccucc ugccagugcu uugccaggac	540
cacggccuga cuccagacca gguggucgu auugcgucga acauuggagg gaagcaagcc	600
cuugaaaccg ugcagaggcu gcucccgug cugugccaag accauggacu caccgccgac	660
caaguggugg cuauugcuag caacauuggc gguaagcagg cgcuggagac aguccagcgg	720
cugcugccgg uguugugcca agaucacggu cuuaccccag accaagucgu ggcgauugcc	780
uccaacggug gcggcaagca agcacucgaa acuguccaga gacugcucc ugugcucugu	840
caagaccacg gguugacccc cgaccaagug guggccaucg ccucccauga uggaggaaag	900
caggccucg agacugucca gcgacugcuc cccguguugu gucaggauca uggauugacg	960
cccgaucagg ucguggccau ugccuccac gacgguggaa agcaagcgu ggaaacugug	1020
cagcgguugc ugccgguccu gugccaggac cacggacuga cuccggacca gguggucgcc	1080
aucgcaucca acauuggugg caagcaggcu cucgaaaccg uccaacgccu guugccggug	1140
cugugucagg aucauggacu gaccccgac caaguggugg cuaucgccuc caacaacggg	1200
ggcaaacagg ccuggaaac cgugcaacgc cugcugcccg uccucugcca ggaccacggu	1260
cugacccug accaggucgu cgcgaucgcg ucaaaugggg ggggcaaaca ggcucuggaa	1320
acggugcagc ggcuccuucc gguguuauugc caggaccacg ggcugacucc cgaccaggug	1380
guggcgauug ccucgaacaa cggaggcaaa caagccucg agacugugca gagacuccug	1440
cccugucugu gccaaagacca ugggcucacc ccugaucagg ugguggcaau gccucaaac	1500
aucggcgga agcaggcacu ggaaacugug cagagacucc ugcccugcu gugccaagac	1560
caugggcuga ccccgacca aguggggcu aucgccucc acgacggggg caaacaggcc	1620
cuggaaucca ucgucguca gcugagccgg ccugaccag cacucgccgc ccugaccaau	1680
gaccaucugg ucgccuggc cugccuggga ggcagaccgc cgauggacgc ggucaagaag	1740
ggucugccgc acgcccuga gcuuauucgg agagugaaca ggcgcaucgg ugaacgcacc	1800

ucccaucggg ucgcaaucuc uagagugggc ggaucgucc ggccgggaguc caucaauccu 1860  
 uggauccuga ccggcuucgc cgacgccgaa ggcuccuua uccuggacau caggaacagg 1920  
 aacaacgagu caaacaggua ccgcaccucc cuucgguucc agauuacucu gcacaacaag 1980  
 gauaagucca uccucgagaa cauccaguca accuggaaag ugggcaagau cacuaacucc 2040

ucggaccgcg cagugaugcu ccgggugacc cgcuucgagg accugaaggu gaucauugac 2100  
 cacuucgaga aguaccucu cauaaccag aagcugggag auuacaagcu guuuuagcag 2160  
 gcguucuccg ugauggagaa caaagaacac cuuaaggaga augggauuaa ggaacugguc 2220  
 cgcauaaagg ccaagaugaa cuggggacug aacgacgagu ugaaaaaggc auuuccugaa 2280  
 aacaucucca aggaacggcc gcucaucaac aagaacauuc ccaauucaa guggcuggcg 2340  
 gggauucacug ccggggacgg acacuucgga gugaaccuga agaaggugaa gggcaccgcc 2400  
 aagguguacg ugggccugcg guucgcgauc agccagcaca uccgggauaa gaaccugaug 2460

aacagccuca ucaccuaccu gggaugcggg agcauccggg agaagaaca gucagaauc 2520  
 cgauggcugg aauuugaagu gaccaaguuc uccgacauca acgacaagau cauccccgug 2580  
 uuccaggaga acaccuau uggagugaag cuggaggacu ucgaggacug gugcaaggug 2640  
 gccaaagcuca ucgaagagaa gaagcaccug accgaaagcg gccuggauga gauuaagaag 2700  
 auuaagcuca acaugaacaa gggaagauag uag 2733

<210> 17

<211> 22

<212> DNA

<213> Homo sapiens

<400> 17

tgtctgccta ttcaccgatt tt 22

<210> 18

<211> 11

<212> DNA

<213> Homo sapiens

<400> 18

aaatccagtg a 11

<210> 19

<211> 12

<212> DNA

<213> Homo sapiens



<400> 19  
 aaatccagtg ac 12  
 <210> 20  
 <211> 39  
 <212> DNA  
 <213> Homo sapiens  
 <400> 20  
 aaatccagtg acaagtcgtg ctgcctattc accgatttt 39  
 <210> 21  
 <211> 22  
 <212> DNA  
 <213> Homo sapiens  
 <400> 21  
 agtatgccta ttctacaaga aa 22  
 <210> 22  
 <211> 877  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Synthesized NTD variant megaTAL construct  
 <400> 22  
 Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Lys Leu Cys  
 1 5 10 15  
 Thr Leu Gly Tyr Ser Gln Gln Gln Gln Glu Lys Leu Lys Leu Lys Ala  
 20 25 30  
 Arg Pro Ile Val Ala Gln His His Glu Ala Leu Ile Gly His Gly Phe  
 35 40 45  
 Thr Arg Ala Gln Ile Val Ala Leu Ser Gln His Pro Ala Ala Leu Gly  
 50 55 60  
 Thr Ile Ala Ala Lys Tyr Gln Ala Met Ile Ala Ala Leu Pro Glu Ala  
 65 70 75 80  
 Thr His Glu Asp Val Val Gly Ile Gly Lys Gln Trp Ser Gly Ala Arg  
 85 90 95

Thr Leu Glu Val Leu Leu Thr Val Ser Gly Glu Leu Arg Ser Pro Pro

100 105 110

Leu Gln Leu Asp Thr Ser Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly

115 120 125

Val Thr Ala Ile Glu Ala Val His Ala Trp Arg Asn Ala Leu Thr Gly

130 135 140

Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn

145 150 155 160

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

165 170 175

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

180 185 190

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

195 200 205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210 215 220

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

225 230 235 240

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val

245 250 255

Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val

260 265 270

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp

275 280 285

Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu

290 295 300

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr

305 310 315 320

Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala

325 330 335

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly

340                      345                      350  
 Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys  
  
 355                      360                      365  
 Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp  
 370                      375                      380  
 His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly  
 385                      390                      395                      400  
 Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys  
 405                      410                      415  
 Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn  
  
 420                      425                      430  
 Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val  
 435                      440                      445  
 Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala  
 450                      455                      460  
 Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu  
 465                      470                      475                      480  
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala  
  
 485                      490                      495  
 Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala  
 500                      505                      510  
 Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His  
 515                      520                      525  
 Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val  
 530                      535                      540  
 Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg  
  
 545                      550                      555                      560  
 Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Val Gly  
 565                      570                      575  
 Gly Ser Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe  
 580                      585                      590

Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg Asn Asn  
595 600 605  
Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr Leu His  
610 615 620  
Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp Lys Val  
625 630 635 640  
Gly Lys Ile Thr Asn Ser Gly Asp Arg Ala Val Met Leu Arg Val Thr  
645 650 655  
Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro  
660 665 670  
Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe  
675 680 685  
Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu  
690 695 700  
Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Thr Asp Glu Leu  
705 710 715 720  
Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn  
725 730 735  
Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser Gly Asp  
740 745 750  
Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Asn Ala Lys Val  
755 760 765  
Tyr Val Gly Leu Arg Phe Ser Ile Ser Gln His Ile Arg Asp Lys Asn  
770 775 780  
Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile Trp Glu  
785 790 795 800  
Lys Asn Lys Ser Glu Phe Ser Trp Leu Glu Phe Val Val Thr Lys Phe  
805 810 815  
Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu  
820 825 830  
Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys

835 840 845  
 Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile  
 850 855 860  
 Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe

865 870 875  
 <210> 23  
 <211> 877  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Synthesized NTD variant megaTAL construct  
 <400> 23

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

35 40 45  
 Thr His Ala His Ile Val Glu Leu Ser Lys His Pro Pro Ala Leu Gly  
 50 55 60  
 Thr Ile Ala Ala Arg Tyr Ser Glu Met Ile Ala Ala Leu Pro Glu Ala  
 65 70 75 80  
 Thr His Glu Asp Ile Val Gly Val Gly Lys His Cys Ala Gly Ala Arg  
 85 90 95  
 Thr Leu Glu Val Leu Leu Met Val Val Gln Glu Leu Arg Ala Pro Pro

100 105 110  
 Leu Gln Leu Val Thr Ser Gln Leu Leu Lys Ile Ala Lys Arg Gly Gly  
 115 120 125  
 Val Thr Ala Val Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly  
 130 135 140  
 Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn  
 145 150 155 160

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

165 170 175

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

180 185 190

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

195 200 205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210 215 220

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

225 230 235 240

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val

245 250 255

Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val

260 265 270

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp

275 280 285

Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu

290 295 300

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr

305 310 315 320

Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala

325 330 335

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly

340 345 350

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys

355 360 365

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp

370 375 380

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly

385 390 395 400

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys



405 410 415  
Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn

420 425 430  
Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

435 440 445  
Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

450 455 460  
Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu

465 470 475 480  
Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

485 490 495  
Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala

500 505 510  
Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His

515 520 525  
Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val

530 535 540  
Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg

545 550 555 560  
Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Val Gly

565 570 575  
Gly Ser Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe

580 585 590  
Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg Asn Asn

595 600 605  
Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr Leu His

610 615 620  
Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp Lys Val

625 630 635 640  
Gly Lys Ile Thr Asn Ser Gly Asp Arg Ala Val Met Leu Arg Val Thr

645 650 655

Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro

660 665 670

Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe

675 680 685

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

690 695 700

Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Thr Asp Glu Leu

705 710 715 720

Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn

725 730 735

Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser Gly Asp

740 745 750

Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Asn Ala Lys Val

755 760 765

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

770 775 780

Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile Trp Glu

785 790 795 800

Lys Asn Lys Ser Glu Phe Ser Trp Leu Glu Phe Val Val Thr Lys Phe

805 810 815

Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu

820 825 830

Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys

835 840 845

Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile

850 855 860

Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe

865 870 875

<210> 24

<211> 877

<212> PRT

<213> Artificial Sequence

<220><223> Synthesized NTD variant megaTAL construct

<400> 24

Met Gly Ser Ala Pro Pro Lys Lys Lys Arg Lys Val Gly Lys Leu Ser

1 5 10 15

Thr Phe Gly Tyr Ser Gln Gln Gln Gln Glu Lys Ile Lys Pro Lys Ala

20 25 30

Arg Ser Thr Val Ala Gln His His Ala Ala Leu Ile Gly His Gly Phe

35 40 45

Thr His Ala His Ile Val Glu Leu Ser Lys His Pro Pro Ala Leu Gly

50 55 60

Thr Ile Ala Ala Arg Tyr Ser Glu Met Ile Ala Ala Leu Pro Glu Ala

65 70 75 80

Thr His Glu Asp Ile Val Gly Val Gly Lys Gln Trp Ser Gly Ala Arg

85 90 95

Thr Leu Glu Val Leu Leu Met Val Val Gln Glu Leu Arg Ala Pro Pro

100 105 110

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

115 120 125

Val Thr Ala Ile Glu Ala Val His Ala Ser Arg Asn Ala Leu Thr Gly

130 135 140

Ala Pro Leu Asn Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn

145 150 155 160

Ile Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

165 170 175

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

180 185 190

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

195 200 205

Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala

210 215 220

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

225 230 235 240

Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val

245 250 255

Val Ala Ile Ala Ser Asn Gly Gly Gly Lys Gln Ala Leu Glu Thr Val

260 265 270

Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp

275 280 285

Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala Leu Glu

290 295 300

Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly Leu Thr

305 310 315 320

Pro Asp Gln Val Val Ala Ile Ala Ser His Asp Gly Gly Lys Gln Ala

325 330 335

Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp His Gly

340 345 350

Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Ile Gly Gly Lys

355 360 365

Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys Gln Asp

370 375 380

His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn Asn Gly

385 390 395 400

Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val Leu Cys

405 410 415

Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala Ser Asn

420 425 430

Gly Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu Pro Val

435 440 445

Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala Ile Ala

450 455 460

Ser Asn Asn Gly Gly Lys Gln Ala Leu Glu Thr Val Gln Arg Leu Leu

465                      470                      475                      480  
 Pro Val Leu Cys Gln Asp His Gly Leu Thr Pro Asp Gln Val Val Ala  
  
                          485                      490                      495  
 Ile Ala Ser Asn Ile Gly Gly Lys Gln Ala Leu Glu Ser Ile Val Ala  
                          500                      505                      510  
 Gln Leu Ser Arg Pro Asp Pro Ala Leu Ala Ala Leu Thr Asn Asp His  
                          515                      520                      525  
 Leu Val Ala Leu Ala Cys Leu Gly Gly Arg Pro Ala Met Asp Ala Val  
                          530                      535                      540  
 Lys Lys Gly Leu Pro His Ala Pro Glu Leu Ile Arg Arg Val Asn Arg  
  
 545                      550                      555                      560  
 Arg Ile Gly Glu Arg Thr Ser His Arg Val Ala Ile Ser Arg Val Gly  
                          565                      570                      575  
 Gly Ser Ser Arg Arg Glu Ser Ile Asn Pro Trp Ile Leu Thr Gly Phe  
                          580                      585                      590  
 Ala Asp Ala Glu Gly Ser Phe Ile Leu Asp Ile Arg Asn Arg Asn Asn  
                          595                      600                      605  
 Glu Ser Asn Arg Tyr Arg Thr Ser Leu Arg Phe Gln Ile Thr Leu His  
  
                          610                      615                      620  
 Asn Lys Asp Lys Ser Ile Leu Glu Asn Ile Gln Ser Thr Trp Lys Val  
 625                      630                      635                      640  
 Gly Lys Ile Thr Asn Ser Gly Asp Arg Ala Val Met Leu Arg Val Thr  
                          645                      650                      655  
 Arg Phe Glu Asp Leu Lys Val Ile Ile Asp His Phe Glu Lys Tyr Pro  
                          660                      665                      670  
 Leu Ile Thr Gln Lys Leu Gly Asp Tyr Lys Leu Phe Lys Gln Ala Phe  
  
                          675                      680                      685  
 Ser Val Met Glu Asn Lys Glu His Leu Lys Glu Asn Gly Ile Lys Glu  
                          690                      695                      700  
 Leu Val Arg Ile Lys Ala Lys Met Asn Trp Gly Leu Thr Asp Glu Leu  
 705                      710                      715                      720

Lys Lys Ala Phe Pro Glu Asn Ile Ser Lys Glu Arg Pro Leu Ile Asn

725 730 735

Lys Asn Ile Pro Asn Phe Lys Trp Leu Ala Gly Phe Thr Ser Gly Asp

740 745 750

Gly Tyr Phe Gly Val Asn Leu Lys Lys Val Lys Gly Asn Ala Lys Val

755 760 765

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

770 775 780

Leu Met Asn Ser Leu Ile Thr Tyr Leu Gly Cys Gly Ser Ile Trp Glu

785 790 795 800

Lys Asn Lys Ser Glu Phe Ser Trp Leu Glu Phe Val Val Thr Lys Phe

805 810 815

Ser Asp Ile Asn Asp Lys Ile Ile Pro Val Phe Gln Glu Asn Thr Leu

820 825 830

Ile Gly Val Lys Leu Glu Asp Phe Glu Asp Trp Cys Lys Val Ala Lys

835 840 845

Leu Ile Glu Glu Lys Lys His Leu Thr Glu Ser Gly Leu Asp Glu Ile

850 855 860

Lys Lys Ile Lys Leu Asn Met Asn Lys Gly Arg Val Phe

865 870 875

<210> 25

<211> 2634

<212> RNA

<213> Artificial Sequence

<220><223> Synthesized NTD variant megaTAL construct

<400> 25

augggauccg cgccaccuaa gaagaaacgc aaagucggca agcugugcac ccugggcuac	60
agccagcagc agcaggagaa gcugaagcug aaggccagac ccaucguggc ccagcaccac	120
gaggcccuga ucggccacgg cuucaccaga gcccagaucg uggcccugag ccagcacccc	180
gccgcccugg gcaccaucgc cgccaaguac caggccauga ucgccgcccu gcccgaggcc	240
accacagagg acgugguggg caucggcaag caguggagcg gcgccagaac ccuggaggug	300

cuacugaccg ugagcggcga gcugagaagc cccccccugc agcuggacac cagccagcug 360  
 cugaagaucg ccaagagagg cggcgugacc gccaucgagg ccgugcacgc cuggcgcaau 420  
 gcacugacgg gugcccccu gaaccugacc ccggaccaag ugguggcuaug cgccagcaac 480  
 auuggcggca agcaagcgcu cgaaacggug cagcggcugu ugccggugcu gugccaggac 540  
 cauggccuga ccccggaaca agugguggcuc aucgccagca acauuggcgg caagcaagcg 600  
 cucgaaacgg ugcagcggcu guugccggug cugugccagg accauggccu gaccccggaac 660  
 caaugggugg cuaucgccag caacauuggc ggcaagcaag cgcucgaaac ggugcagcgg 720  
  
 cuguugccgg ugcugugcca ggaccauggc cugaccccgg accaaguggu ggcuaucgcc 780  
 agcaacggug gcggcaagca agcgcucgaa acggugcagc ggcuguugcc ggugcugugc 840  
 caggaccaug gccugacucc ggaccaagug guggcuauug ccagccacga uggcggcaag 900  
 caagcgcucg aaacggugca gcggcuguug ccggugcugu gccaggacca uggccugacu 960  
 ccggaccaag ugguggcuaug cgccagccac gauggcggca agcaagcgcu cgaaacggug 1020  
 cagcggcugu ugccggugcu gugccaggac caugccuga ccccggaaca agugguggcuc 1080  
 aucgccagca acauuggcgg caagcaagcg cucgaaacgg ugcagcggcu guugccggug 1140  
  
 cugugccagg accauggccu gaccccggaac caaugggugg cuaucgccag caacauggc 1200  
 ggcaagcaag cgcucgaaac ggugcagcgg cuguugccgg ugcugugcca ggaccauggc 1260  
 cugaccccgg accaaguggu ggcuaucgcc agcaacggug gcggcaagca agcgcucgaa 1320  
 acggugcagc ggcuguugcc ggugcugugc caggaccaug gccugacccc ggaccaagug 1380  
 guggcuauug ccagcaacaa uggcggcaag caagcgcucg aaacggugca gcggcuguug 1440  
 ccggugcugu gccaggacca uggccugacc ccggaccaag ugguggcuaug cgccagcaac 1500  
 auuggcggca agcaagcgcu cgaaagcauu guggcccagc ugagccggcc ugaucggcg 1560  
  
 uggcgcgcu ugaccaacga ccaccucguc gccuuggccu gccucggcgg acguccugcc 1620  
 auggaugcag ugaaaaaggg auugccgcac gcgccggaau ugaucagaag agucaaucgc 1680  
 cgauuuggcg aacgcacguc ccaucgcguu gcgauaucua gagugggagg aagcucgcgc 1740  
 agagagucca ucaaccgaug gauucugacu gguuucgcug augccgaagg aucauucua 1800  
 cuagacaucc gcaaccgaaa caacgaaagc aacagauacc gaacuucgcu gagauuccag 1860  
 aucaccucg acaacaagga caaaucauu cuggagaaua uccagucgac uuggaagguc 1920  
 ggcaagauca caaacagcgg cgacagagcc gucaugcuga gggucacccg uuucgaagau 1980  
  
 uugaaaguga uuaucgacca cuucgagaaa uauccgcuga uuaccagaa auugggcgau 2040  
 uacaaguugu uuaaacaggc auucagcgcug auggagaaca aagaacauu uaaggagaau 2100  
 gggauuaagg agcucguacg aaucaaagcu aagaugaaau ggggucucac ugacgaaug 2160



aaaaaagcau uuccagagaa cauuagcaaa gagcgccccc uuaucaauaa gaacauuccg 2220  
 aauuucaaaau ggcuggcugg auucacaucau ggugauggcu acuucggcgu gaaucuaaaa 2280  
 aagguaaagg gcaacgcaaa gguauacgug ggacugagau ucuaaauuc acagcacauc 2340  
 agagacaaga accugaugaa uucauugaua acauaccuag gcugugguuc caucugggag 2400  
  
 aagaacaagu cugaguucag uuggcucgag uucgucguua ccaaaauucag cgauaucaac 2460  
 gacaagauc uuccgguaau ccagaaaau acucugauug gcgucaaacu cgaggacuuu 2520  
 gaagauuggu gcaagguugc caaaauugauc gaagagaaga aacaccugac cgaaucgggu 2580  
 uuggaugaga uuaagaaaau caagcugaac augaacaag gucgugucuu cuga 2634  
  
 <210> 26  
 <211> 2634  
 <212> RNA  
 <213> Artificial Sequence  
 <220><223> Synthesized NTD variant megaTAL construct  
 <400> 26  
  
 augggauccg cgccaccuaa gaagaaacgc aaagucgugg accugagcac cuucggcuac 60  
  
 agccagcagc agcaggagaa gaucaagccc aaggugagaa gcaccguggc ccagcaccac 120  
 gccgcccugg ugggccacgg cuucacccac gccacaucg uggagcugag caagcaccac 180  
 cccgcccugg gcaccaucgc cgccagauac agcgagauga ucgccgccu gcccgaggcc 240  
 acccagcagg acaucguggg cgugggcaag cacugcgccg gcgccagaac ccuggaggug 300  
 cugcugaugg uggugcagga gcugagagcu ccaccacugc agcuggugac cagccagcug 360  
 cugaagaucg ccaagagagg cggcgugacc gccguggagg ccgugcacgc cagccgcaau 420  
 gcacugacgg gugcccccu gaaccugacc ccggaccaag ugguggcuauc cgccagcaac 480  
  
 auuggcggca agcaagcgu cgaaacggug cagcggcugu ugccggugcu gugccaggac 540  
 cauggccuga ccccggaaca agugguggcu aucgccagca acauuggcgg caagcaagcg 600  
 cucgaaacgg ugcagcggcu guugccggug cugugccagg accauggccu gaccccgac 660  
 caaguggugg cuaucgccag caacauuggc ggcaagcaag cgcucgaaac ggugcagcgg 720  
 cuguugccgg ugcugugcca ggaccauggc cugaccccgg accaaguggu ggcuaucgcc 780  
 agcaacggug gcggcaagca agcgucgaa acggugcagc ggcuugucc ggugcugugc 840  
 caggaccaug gccugacucc ggaccaagug guggcuauuc ccagccacga uggcggaag 900  
  
 caagcgcucg aaacggugca gcggcuguug ccggugcugu gccaggacca uggccugacu 960  
 ccggaccaag ugguggcuauc cgccagccac gauggcggca agcaagcgu cgaaacggug 1020

cagcggcugu ugccggugcu gugccaggac cauggccuga ccccgacga agugguggu	1080
aucgccagca acauuggcgg caagcaagcg cucgaaacgg ugcagcggcu guugccggug	1140
cugugccagg accauggccu gaccccgac caaguggug cuaucgccag caacauggc	1200
ggcaagcaag cgcugaaac ggugcagcgg cuguugccgg ugcugugcca ggaccauggc	1260
cugaccccg accaaguggu ggcuaucgcc agcaacggug gcggcaagca agcgcucgaa	1320
acggugcagc ggcuguugcc ggugcuguc caggaccaug gccugacccc ggaccaaug	1380
guggcuaucg ccagcaacaa uggcggcaag caagcgcucg aaacggugca gcggcuguug	1440
ccggugcugu gccaggacca uggccugacc ccggaccaag ugguggcua cgcagcaac	1500
auuggcggca agcaagcgu cgaaagcau uggcccagc ugagccggcc ugaucggcg	1560
uuggccgcu ugaccaacga ccaccuguc gccuuggccu gccucggcg acgucugcc	1620
auggaugcag ugaagaggg auugccgcac gcggcggaau ugaucagaag agucaaucg	1680
cguauuggcg aacgcacguc ccaucgcuu gcgauaucua gagugggagg aagcucgcg	1740
agagagucca ucaaccaug gauucugacu gguuucguc augccgaagg aucauucua	1800
cuagacaucc gcaaccgaaa caacgaaagc aacagauacc gaacuucgu gagauuccag	1860
aucaccugc acaacaagga caaaucgau cuggagaaua uccagucgac uuggaagguc	1920
ggcaagauca caaacagcgg cgacagagcc gucaugcuga gggucacccg uuucgaagau	1980
uugaaaguga uuaucgacca cuucgagaaa uauccgcuga uuaccagaa auugggcgau	2040
uacaaguugu uuaaacaggc auucagcgu auggagaaca aagaacauu uaaggagaau	2100
gggaauaagg agcucguacg aaucaaagcu aagaugaaau ggggucucac ugacgaauug	2160
aaaaagcau uuccagagaa cauagcaaa gagcgcccc uuaucaaua gaacauuccg	2220
aaauucaaau ggcuggcugg auucacauu ggugauggcu acuucggcg gaaucuaaaa	2280
aagguaaagg gcaacgcaaa gguauacgug ggacugagau ucuaaucuc acagcacauc	2340
agagacaaga accugaugaa uucauugua acauaccuag gcugugguu caucugggag	2400
aagaacaagu cugaguucag uuggcucgag uucgucguu ccaauucag cgauaucaac	2460
gacaagauca uuccgguuu ccaggaaaau acucugauug gcgucaacu cgaggacuuu	2520
gaagauuggu gcaagguugc caauugauc gaagagaaga aacaccugac cgaauccggu	2580
uuggaugaga uuaagaaaa caagcugaac augaacaag gucgugucu cuga	2634
<210> 27	
<211> 2634	
<212> RNA	
<213> Artificial Sequence	

<220><223> Synthesized NTD variant megaTAL construct

<400> 27

augggauccg cgccaccuaa gaagaaacgc aaagucggca agcugagcac cuucggcuac	60
agccagcagc agcaggagaa gaucaagccc aaggccagaa gcaccguggc ccagcaccac	120
gccgcccuga ucggccacgg cuucacccac gcccaucugc uggagcugag caagcacccc	180
ccccccucgg gcaccaucgc cgccagauac agcgagauga ucggcgcccu gcccagggcc	240
acccacgagg acaucguggg cgugggcaag caguggagcg gcgccagaac ccuggaggug	300
cugcugaugg uggugcagga gcugagagcu ccaccacugc agcuggugac cagccagcug	360
gugaagaucg ccaagagagg cggcgugacc gccaucgagg ccgugcacgc cagccgcaau	420
gcacugacgg gugcccccu gaaccugacc ccggaccaag ugguggcuau cgccagcaac	480
auuggcggca agcaagcgu cgaaacggug cagcggcugu ugccggugcu gugccaggac	540
cauggccuga ccccgaccac agugguggc cuagccagca acauuggcgg caagcaagcg	600
cucgaaacgg ugcagcggcu guugccggug cugugccagg accauggccu gaccccgac	660
caaugggugg cuaucgccag caacauuggc ggcaagcaag cgcucgaaac ggugcagcgg	720
cuguugccgg ugcugugcca ggaccauggc cugaccccg accaaguggu ggcuaucgcc	780
agcaacggug gcggcaagca agcgucgaa acggugcagc ggcuugucc ggugcugugc	840
caggaccaug gccugacucc ggaccaagug guggcuaucc ccagccacga uggcggaag	900
caagcgcucg aaacggugca gcggcuguug ccggugcugu gccaggacca uggccugacu	960
ccggaccaag ugguggcuau cgccagccac gauggcggca agcaagcgu cgaaacggug	1020
cagcggcugu ugccggugcu gugccaggac cauggccuga ccccgaccac agugguggc	1080
aucgccagca acauuggcgg caagcaagcg cucgaaacgg ugcagcggcu guugccggug	1140
cugugccagg accauggccu gaccccgac caaugggugg cuaucgccag caacauggc	1200
ggcaagcaag cgcucgaaac ggugcagcgg cuguugccgg ugcugugcca ggaccauggc	1260
cugaccccg accaaguggu ggcuaucgcc agcaacggug gcggcaagca agcgucgaa	1320
acggugcagc ggcuugucc ggugcugugc caggaccaug gccugacccc ggaccaagug	1380
guggcuaucc ccagcaacaa uggcggaag caagcgcucg aaacggugca gcggcuguug	1440
ccggugcugu gccaggacca uggccugacc ccggaccaag ugguggcuau cgccagcaac	1500
auuggcggca agcaagcgu cgaaagcauu guggcccagc ugagccggcc ugaucggcg	1560
uuggcccgcu ugaccaacga ccaccuguc gccuuggccu gccucggcg acguccugc	1620
auggaugcag ugaaaaaggg auugccgcac gcgccggaau ugaucagaag agucaaucgc	1680
cguauuggcg aacgcacguc ccaucgcguu gcgauaucua gagugggagg aagcucgcgc	1740

agagagucca ucaacccaug gauucugacu gguuucgcug augccgaagg aucauucuaa	1800
cuagacaucc gcaaccgaaa caacgaaagc aacagauacc gaacuucgcu gagauuccag	1860
aucacccugc acaacaagga caaaucgauu cuggagaaua uccagucgac uuggaagguc	1920
ggcaagauga caaacagcgg cgacagagcc gucaugcuga gggucacccg uuucgaagau	1980
uugaaaguga uuaucgacca cuucgagaaa uauccgcuga uuaccagaa auugggcgau	2040
uacaaguugu uuaaacaggc auucagcguc auggagaaca aagaacauu uaaggagaau	2100
gggaauaagg agcucguacg aaucaaagcu aagaugaauu ggggucucac ugacgaauug	2160
aaaaaagcau uuccagagaa cauuagcaaa gagcgccccc uuaucaauaa gaacauuccg	2220
aaauucaaau ggucggcugg auucacauu ggugauggcu acuucggcgu gaaucaaaaa	2280
aagguaaagg gcaacgcaaa gguauacgug ggacugagau ucuaaucuc acagcacauc	2340
agagacaaga accugaugaa uucauugaua acauaccuag gcugugguuc caucugggag	2400
aagaacaagu cugaguucag uuggcucgag uucgucguua ccaauucag cgauaucaac	2460
gacaagauga uuccgguaau ccaggaauu acucugauug gcgucaaacu cgaggacuuu	2520
gaagauggu gcaagguugc caauuugauc gaagagaaga aacaccugac cgaauccggu	2580
uuggaugaga uuaagaaaa caagcugaac augaacaag gucgugucu cuga	2634
<210> 28	
<211> 711	
<212> RNA	
<213> Mus musculus	
<400> 28	
augucugagc caccucgggc ugagaccuuu guauuccugg accuagaagc cacugggcuc	60
ccaaacaugg accugagau ugacagaua uccuuuuug cuguucacg cucuuccug	120
gagaaccag aacgggauga uucugguucc uuggugcugc cccguguucu ggacaagcuc	180
acacugugca ugugcccga gcgccccuuu acugccaagg ccagugagau uacugguuug	240
agcagcgaaa gccugaugca cugcgggaag gcugguuua auggcgcugu gguaaggaca	300
cugcagggcu uccuaagccg ccaggagggc ccaucugcc uuguggcca caaugcuuc	360
gaauaugacu ucccacugcu gugcacggag cuacaacguc ugggugcca ucugcccaa	420
gacacugucu gccuggacac acugccugca uugcggggcc uggaccguc ucacagccac	480
ggcaccaggg cucaaggccg caaaagcuac agccugcca gucucuucca ccgcuacuuc	540
caggcugaac ccagugcugc ccauucagca gaaggugaug ugcacaccu gcuucugauc	600
uuccugauc gucuccuga gcugcugcc ugggcagaug agcaggccc cagcugggcu	660

cauauugagc ccauguacgu gccaccugau gguccaagcc ucgaagccug a

711

<210> 29

<211> 236

<212> PRT

<213> Mus musculus

<400> 29

Met Ser Glu Pro Pro Arg Ala Glu Thr Phe Val Phe Leu Asp Leu Glu

1 5 10 15

Ala Thr Gly Leu Pro Asn Met Asp Pro Glu Ile Ala Glu Ile Ser Leu

20 25 30

Phe Ala Val His Arg Ser Ser Leu Glu Asn Pro Glu Arg Asp Asp Ser

35 40 45

Gly Ser Leu Val Leu Pro Arg Val Leu Asp Lys Leu Thr Leu Cys Met

50 55 60

Cys Pro Glu Arg Pro Phe Thr Ala Lys Ala Ser Glu Ile Thr Gly Leu

65 70 75 80

Ser Ser Glu Ser Leu Met His Cys Gly Lys Ala Gly Phe Asn Gly Ala

85 90 95

Val Val Arg Thr Leu Gln Gly Phe Leu Ser Arg Gln Glu Gly Pro Ile

100 105 110

Cys Leu Val Ala His Asn Gly Phe Asp Tyr Asp Phe Pro Leu Leu Cys

115 120 125

Thr Glu Leu Gln Arg Leu Gly Ala His Leu Pro Gln Asp Thr Val Cys

130 135 140

Leu Asp Thr Leu Pro Ala Leu Arg Gly Leu Asp Arg Ala His Ser His

145 150 155 160

Gly Thr Arg Ala Gln Gly Arg Lys Ser Tyr Ser Leu Ala Ser Leu Phe

165 170 175

His Arg Tyr Phe Gln Ala Glu Pro Ser Ala Ala His Ser Ala Glu Gly

180 185 190

Asp Val His Thr Leu Leu Leu Ile Phe Leu His Arg Ala Pro Glu Leu

Leu Ala Trp Ala Asp Glu Gln Ala Arg Ser Trp Ala His Ile Glu Pro

Met Tyr Val Pro Pro Asp Gly Pro Ser Leu Glu Ala

225                      230                      235

<210> 30

<211> 3

<212> PRT

### <213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 30

Gly Gly Gly

1

<210> 31

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

&lt;400&gt; 31

Asp Gly Gly Gly Ser

1 5

<210> 32

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 32

Thr Gly Glu Lys Pro

1 5

<210> 33

<211> 4

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 33

Gly Gly Arg Arg

1

<210> 34

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 34

Gly Gly Gly Gly Ser

1 5

<210> 35

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 35

Glu Gly Lys Ser Ser Gly Ser Gly Ser Glu Ser Lys Val Asp

1 5 10

<210> 36

<211> 18

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 36

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

1 5 10 15

Leu Asp

<210> 37

<211> 8

<212> PRT



<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 37

Gly Gly Arg Arg Gly Gly Gly Ser

1 5

<210> 38

<211> 9

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 38

Leu Arg Gln Arg Asp Gly Glu Arg Pro

1 5

<210> 39

<211> 12

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 39

Leu Arg Gln Lys Asp Gly Gly Gly Ser Glu Arg Pro

1 5 10

<210> 40

<211> 16

<212> PRT

<213> Artificial Sequence

<220><223> Exemplary linker sequence

<400> 40

Leu Arg Gln Lys Asp Gly Gly Gly Ser Gly Gly Gly Ser Glu Arg Pro

1 5 10 15

<210> 41

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Cleavage sequence by TEV protease

<220><221> misc\_feature

<222> (2)..(3)

<223> Xaa is any amino acid

<220><221> misc\_feature

<222> (5)..(5)

<223> Xaa is any amino acid

<220><221> MISC\_FEATURE

<222> (7)..(7)

<223> Xaa = Gly or Ser

<400> 41

Glu Xaa Xaa Tyr Xaa Gln Xaa

1 5

<210> 42

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Cleavage sequence by TEV protease

<400> 42

Glu Asn Leu Tyr Phe Gln Gly

1 5

<210> 43

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Cleavage sequence by TEV protease

<400> 43

Glu Asn Leu Tyr Phe Gln Ser

1 5

<210> 44

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 44

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

1 5 10 15

Glu Glu Asn Pro Gly Pro

20

<210> 45

<211> 19

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 45

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

1 5 10 15

Pro Gly Pro

<210> 46

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 46

Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn Pro Gly Pro

1 5 10

<210> 47

<211> 21

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 47

Gly Ser Gly Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu

1 5 10 15

Glu Asn Pro Gly Pro

20

<210> 48  
 <211> 18  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Self-cleaving polypeptide comprising 2A site  
 <400> 48  
 Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro  
 1 5 10 15  
 Gly Pro

<210> 49  
 <211> 13  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Self-cleaving polypeptide comprising 2A site

<400> 49  
 Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro Gly Pro  
 1 5 10

<210> 50  
 <211> 23  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Self-cleaving polypeptide comprising 2A site  
 <400> 50  
 Gly Ser Gly Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp  
 1 5 10 15  
 Val Glu Ser Asn Pro Gly Pro  
 20

<210> 51  
 <211> 20  
 <212> PRT  
 <213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 51

Gln Cys Thr Asn Tyr Ala Leu Leu Lys Leu Ala Gly Asp Val Glu Ser

1 5 10 15

Asn Pro Gly Pro

20

<210> 52

<211> 14

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 52

Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly Pro

1 5 10

<210> 53

<211> 25

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 53

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

1 5 10 15

Gly Asp Val Glu Ser Asn Pro Gly Pro

20 25

<210> 54

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 54

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

1 5 10 15

Glu Ser Asn Pro Gly Pro

20

<210> 55  
 <211> 14  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Self-cleaving polypeptide comprising 2A site  
 <400> 55

Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly Pro  
 1 5 10

<210> 56  
 <211> 19  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Self-cleaving polypeptide comprising 2A site  
 <400> 56

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

1 5 10 15

Pro Gly Pro

<210> 57  
 <211> 19  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Self-cleaving polypeptide comprising 2A site  
 <400> 57

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

1 5 10 15

Pro Gly Pro

<210> 58  
 <211> 14  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> Self-cleaving polypeptide comprising 2A site

<400> 58

Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly Pro

1 5 10

<210> 59

<211> 17

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 59

Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly

1 5 10 15

Pro

<210> 60

<211> 20

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 60

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

1 5 10 15

Asn Pro Gly Pro

20

<210> 61

<211> 24

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 61

Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly

1 5 10 15

Asp Val Glu Ser Asn Pro Gly Pro

20

<210> 62

<211> 40

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 62

Val Thr Glu Leu Leu Tyr Arg Met Lys Arg Ala Glu Thr Tyr Cys Pro

1 5 10 15

Arg Pro Leu Leu Ala Ile His Pro Thr Glu Ala Arg His Lys Gln Lys

20 25 30

Ile Val Ala Pro Val Lys Gln Thr

35 40

<210> 63

<211> 18

<212> PRT

<213>

> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 63

Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro

1 5 10 15

Gly Pro

<210> 64

<211> 40

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 64

Leu Leu Ala Ile His Pro Thr Glu Ala Arg His Lys Gln Lys Ile Val

1 5 10 15

Ala Pro Val Lys Gln Thr Leu Asn Phe Asp Leu Leu Lys Leu Ala Gly

20 25 30

Asp Val Glu Ser Asn Pro Gly Pro



35 40

<210> 65

<211> 33

<212> PRT

<213> Artificial Sequence

<220><223> Self-cleaving polypeptide comprising 2A site

<400> 65

Glu Ala Arg His Lys Gln Lys Ile Val Ala Pro Val Lys Gln Thr Leu

1 5 10 15

Asn Phe Asp Leu Leu Lys Leu Ala Gly Asp Val Glu Ser Asn Pro Gly

20 25 30

Pro

<210> 66

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Consensus Kozak sequence

<400> 66

gccrccatgg 10

<210> 67

<211> 22

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic construct

<220><221> misc\_feature

<222> (1)..(14)

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

<220><221> misc\_feature

<222> (18)..(22)

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

<400> 67

nnnnnnnnnn nnnnccgnnn nn 22

<210> 68

<211> 47

<212> DNA

<213> Homo sapiens

<400> 68

actctaaatc cagtgacaag tctgtctgcc tattcaccca ttttgat

47

<210> 69

<211> 47

<212> DNA

<213> Homo sapiens

<400> 69

atcaaaatcg gtgaataggc agacagactt gtcactggat ttagagt

47