(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0099166
(43) 공개일자 2024년06월28일

(51) 국제특허분류(Int. Cl.)
C12N 9/22 (2006.01) *C12N 15/10* (2017.01)
C12N 15/113 (2010.01) *C12N 9/12* (2006.01)
(52) CPC특허분류
C12N 9/22 (2013.01)
C12N 15/102 (2013.01)
(21) 출원번호 10-2024-7011697
(22) 출원일자(국제) 2022년09월07일
심사청구일자 없음
(85) 번역문제출일자 2024년04월08일
(86) 국제출원번호 PCT/US2022/076024
(87) 국제공개번호 WO 2023/039407
국제공개일자 2023년03월16일
(30) 우선권주장
63/241,931 2021년09월08일 미국(US)

(71) 출원인
플래그쉽 파이어니어링 이노베이션스 브이아이, 엘엘씨
미국 02142 매사추세츠주 캠브리지 55 캠브리지 파크웨이 8 플로어
(72) 발명자
알트슐러, 로버트 찰스
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
보트머, 앤 헬렌
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
(뒷면에 계속)
(74) 대리인
양영준, 이상남

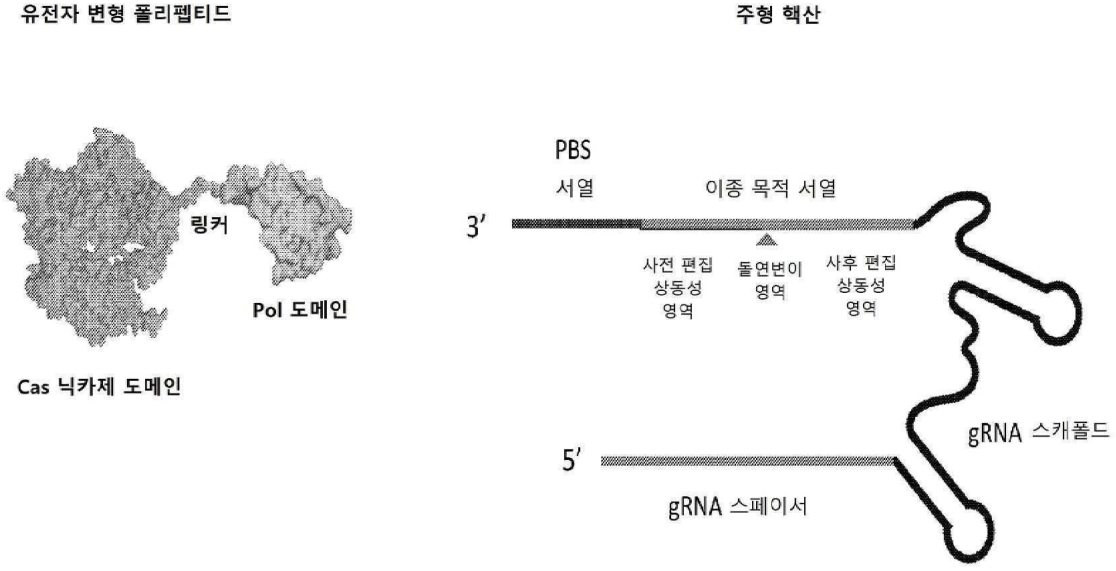
전체 청구항 수 : 총 46 항

(54) 발명의 명칭 **게놈을 조절하기 위한 방법 및 조성물**

(57) 요 약

표적 게놈을 조절하기 위한 방법 및 조성물이 개시되어 있다.

대 표 도



(52) CPC특허분류
C12N 15/113 (2013.01)
C12N 9/1252 (2013.01)
C07K 2319/09 (2013.01)
C12N 2310/20 (2017.05)
(72) 발명자
시토리크, 로버트 제임스
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
코타-라무시노, 세실리아 지오반나 실비아
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
김, 규식
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
코틀라, 랜디 미셸
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
쿠마르, 산딥
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
레이, 아나냐
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
로케트, 나타니엘
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
루벤스, 제이콥 로젠블룸
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
살로몬, 윌리엄 에드워드
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
스타인버그, 배럿 에단
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
왕, 지 준
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
아킨시, 얼신
미국 02142 매사추세츠 캠브리지 캠브리지 파크웨이 55 스위트 800이
두시스, 아타나시오스 디미트리
미국 매사추세츠

명 세 서

청구범위

청구항 1

유전자 변형 폴리펩티드로서,

Cas 도메인(예를 들어, Cas 닉카제 도메인, 예를 들어 Cas9 닉카제 도메인);

표 1 또는 표 23의 중합효소(Pol) 도메인, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열(여기서, Pol 도메인은 Cas 도메인의 C-말단임); 및

Pol 도메인과 Cas 도메인 사이에 배치된 링커

를 포함하는, 유전자 변형 폴리펩티드.

청구항 2

제1항에 있어서, 링커는 표 6의 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 3

제1항 또는 제2항에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 적어도 90% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 적어도 95% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 적어도 98% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 적어도 99% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 100% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 링커는 표 6의 링커 서열에 대하여 적어도 90% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 링커는 표 6의 링커 서열에 대하여 적어도 95% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 10

제1항 내지 제9항 중 어느 한 항에 있어서, 링커는 표 6의 링커 서열에 대하여 적어도 97% 동일성을 갖는 서열

을 갖는, 유전자 변형 폴리펩티드.

청구항 11

제1항 내지 제10항 중 어느 한 항에 있어서, 링커는 표 6의 링커 서열에 대하여 100% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

청구항 12

제1항 내지 제11항 중 어느 한 항에 있어서, Cas 도메인은 표 4의 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 98%, 또는 99% 동일성을 갖는 서열을 포함하는, 유전자 변형 폴리펩티드.

청구항 13

제1항 내지 제12항 중 어느 한 항에 있어서, Cas 도메인은 Cas 닉카제 도메인인, 유전자 변형 폴리펩티드.

청구항 14

제1항 내지 제13항 중 어느 한 항에 있어서, Cas 도메인은 Cas9 닉카제 도메인인, 유전자 변형 폴리펩티드.

청구항 15

제1항 내지 제14항 중 어느 한 항에 있어서, Cas 도메인은 N863A 돌연변이를 포함하는, 유전자 변형 폴리펩티드.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서, NLS를 포함하는, 예를 들어 유전자 변형 폴리펩티드는 2개의 NLS를 포함하는, 유전자 변형 폴리펩티드.

청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서, Cas9 도메인의 NLS N-말단을 포함하는, 유전자 변형 폴리펩티드.

청구항 18

제1항 내지 제17항 중 어느 한 항에 있어서, Pol 도메인의 NLS C-말단을 포함하는, 유전자 변형 폴리펩티드.

청구항 19

제1항 내지 제18항 중 어느 한 항에 있어서, Cas9 도메인의 N-말단인 제1 NLS, 및 Pol 도메인의 C-말단인 제2 NLS를 포함하는, 유전자 변형 폴리펩티드.

청구항 20

제1항 내지 제19항 중 어느 한 항의 유전자 변형 폴리펩티드를 인코딩하는 핵산(예를 들어, DNA 또는 RNA, 예를 들어 mRNA).

청구항 21

제1항 내지 제19항 중 어느 한 항의 유전자 변형 폴리펩티드 또는 제20항의 핵산을 포함하는 세포.

청구항 22

시스템으로서,

i) 제1항 내지 제19항 중 어느 한 항의 유전자 변형 폴리펩티드, 및

ii) a) 표적 핵산 서열의 일부에 상보적인 gRNA 스페이서;

b) 유전자 변형 폴리펩티드의 Cas 도메인에 결합하는 gRNA 스캐폴드;

c) 이종 목적 서열; 및

d) 프라이머 결합 부위 서열(PBS 서열)

을 포함하는 주형 핵산(예를 들어, 주형 RNA)

을 포함하는, 시스템.

청구항 23

제22항에 있어서, 주형 핵산은 RNA를 포함하는, 시스템.

청구항 24

제22항 또는 제23항에 있어서, 주형 핵산은 DNA를 포함하는, 시스템.

청구항 25

제22항에 있어서, 주형 핵산은 DNA 및 RNA를 포함하는, 시스템.

청구항 26

제22항 내지 제25항 중 어느 한 항에 있어서, 주형 핵산은 표 24에 열거된 바와 같은 핵산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함하는, 시스템.

청구항 27

제26항에 있어서, 주형 핵산의 gRNA 스페이서는 표 24에 열거된 바와 같은 스페이서 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함하는, 시스템.

청구항 28

제26항 또는 제27항에 있어서, 주형 핵산의 gRNA 스캐폴드는 표 24에 열거된 바와 같은 스캐폴드 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함하는, 시스템.

청구항 29

제26항 내지 제28항 중 어느 한 항에 있어서, 주형 핵산의 PBS 서열은 표 24에 열거된 바와 같은 PBS 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함하는, 시스템.

청구항 30

제26항에 있어서, 주형 핵산 서열은, 예를 들어 5'에서 3' 순서로, 다음 중 하나 이상(예를 들어, 1, 2, 또는 3개 전부)을 포함하는, 시스템:

(i) 표 24에 열거된 바와 같은 스페이서 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열;

(ii) 표 24에 열거된 바와 같은 스캐폴드 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열; 및/또는

(iii) 표 24에 열거된 바와 같은 PBS 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열.

청구항 31

제26항 내지 제30항 중 어느 한 항에 있어서, 주형 핵산은 표 24에 열거된 바와 같은 전체 주형 분자 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 핵산 서열을 포함하는, 시스템.

청구항 32

제22항 내지 제31항 중 어느 한 항에 있어서, gRNA 스페이서 및 gRNA 스캐폴드는 RNA를 포함하는, 시스템.

청구항 33

제22항 내지 제32항 중 어느 한 항에 있어서, 이종 목적 서열은 DNA를 포함하고 PBS 서열은 RNA를 포함하는, 시스템.

청구항 34

제22항 내지 제32항 중 어느 한 항에 있어서, 이종 목적 서열 및 PBS 서열은 DNA를 포함하는, 시스템.

청구항 35

제22항 내지 제34항 중 어느 한 항에 있어서, 유전자 변형 폴리펩티드는 표 23에 열거된 바와 같은 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 시스템.

청구항 36

제22항 내지 제35항 중 어느 한 항에 있어서, 유전자 변형 폴리펩티드는 nCas9-UL-PolΘ_L, nCas9-UL-PolΘ_M, nCas9-UL-PolΘ_4x0q, nCas9-FL-PolΘ_M, 또는 nCas9-FL-PolΘ_4x0q 중 어느 하나의 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 시스템.

청구항 37

제22항 내지 제36항 중 어느 한 항에 있어서, 유전자 변형 폴리펩티드는 표 23에 열거된 바와 같은 유전자 변형 폴리펩티드의 Cas 도메인 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 시스템.

청구항 38

제22항 내지 제37항 중 어느 한 항에 있어서, 유전자 변형 폴리펩티드는 nCas9-UL-PolΘ_L, nCas9-UL-PolΘ_M, nCas9-UL-PolΘ_4x0q, nCas9-FL-PolΘ_M, 또는 nCas9-FL-PolΘ_4x0q 중 어느 하나의 Cas 도메인 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 시스템.

청구항 39

제22항 내지 제38항 중 어느 한 항에 있어서, 유전자 변형 폴리펩티드는 표 23에 열거된 바와 같은 유전자 변형 폴리펩티드의 Pol 도메인 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 시스템.

청구항 40

제22항 내지 제39항 중 어느 한 항에 있어서, 유전자 변형 폴리펩티드는 nCas9-UL-PolΘ_L, nCas9-UL-PolΘ_M, nCas9-UL-PolΘ_4x0q, nCas9-FL-PolΘ_M, 또는 nCas9-FL-PolΘ_4x0q 중 어느 하나의 Pol 도메인 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 시스템.

청구항 41

제22항 내지 제40항 중 어느 한 항에 있어서, 유전자 변형 폴리펩티드는 표 23에 열거된 바와 같은 유전자 변형 폴리펩티드의 Cas 도메인 아미노산 서열 및 Pol 도메인 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 시스템.

청구항 42

제22항 내지 제41항 중 어느 한 항에 있어서, 유전자 변형 폴리펩티드는 nCas9-UL- Polθ_L, nCas9-UL- Polθ_M, nCas9-UL- Polθ_4x0q, nCas9-FL- Polθ_M, 또는 nCas9-FL- Polθ_4x0q 중 어느 하나의 Cas 도메인 아미노산 서열 및 Pol 도메인 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하는, 시스템.

청구항 43

세포(예를 들어, 인간 세포)에서 표적 핵산을 변형시키는 방법으로서, 세포를 제22항 내지 제42항 중 어느 한 항의 시스템, 또는 이를 인코딩하는 핵산과 접촉시키는 단계를 포함하며, 이에 의해 표적 핵산을 변형시키는, 방법.

청구항 44

유전적 결함과 연관된 질환 또는 병태를 갖는 대상체를 치료하는 방법으로서,

대상체에게 제1항 내지 제43항 중 어느 한 항의 시스템, 폴리펩티드, 주형 RNA 또는 이를 인코딩하는 DNA를 투여하여, 이에 의해 유전적 결합과 연관된 질환 또는 병태를 갖는 대상체를 치료하는 단계

를 포함하는 방법.

청구항 45

제44항에 있어서, 유전적 결함과 연관된 질환 또는 병태는 표 12 내지 15 중 임의의 것에 열거된 적응증이고/이거나, 유전적 결함은 표 12 내지 15 중 임의의 것에 열거된 유전자의 결함인, 방법.

청구항 46

제44항 또는 제45항에 있어서, 대상체는 인간 환자인, 방법.

발명의 설명

기 술 분 야

[0001] **관련 출원**

[0002] 본 출원은 2021년 9월 8일에 출원된 미국 일련번호 63/241,931에 대한 우선권을 주장하며, 이들 각각의 전체 내용은 본 명세서에 참조로 포함된다.

[0003] **서열 목록**

[0004] 본 출원은 XML 서식으로 전자적으로 제출된 서열 목록을 포함하며, 이는 본 명세서에 전체적으로 참고로 포함된다. 2021년 9월 8일에 생성된 상기 XML 사본은 파일명이 V2065-7020WO_SL.xml이며, 크기가 11,288,576 바이트이다.

배 경 기 술

[0005] 관심 핵산을 게놈으로 통합하는 것은, 삽입 사건을 촉진하는 특수화된 단백질의 부재 하에, 낮은 빈도 및 적은 부위 특이성으로 발생한다. CRISPR/Cas9와 같은 기존의 여러 접근법은 숙주 복구 경로에 의존하는 작은 편집체에 더 적합하고, 더 긴 서열의 통합에는 덜 효과적이다. Cre/loxP와 같은 기존의 다른 접근법은, loxP 부위를 게놈으로 삽입하는 제1 단계, 및 이어서 관심 서열을 loxP 부위로 삽입하는 제2 단계가 필요하다. 게놈에서 관심 서열을 삽입, 변경 또는 결실시키기 위한 개선된 조성물(예를 들어, 단백질 및 핵산) 및 방법이 당업계에 필요하다.

발명의 내용

[0006] 본 개시내용은, 생체 내 또는 시험관 내에서, 숙주 세포, 조직 또는 대상체의 하나 이상의 위치에서 게놈을 변경시키기 위한 새로운 조성물, 시스템, 및 방법에 관한 것이다. 구체적으로, 본 발명은 숙주 게놈에서 관심 서열을 삽입, 변경, 또는 결실시키기 위한 조성물, 시스템 및 방법을 특징으로 한다. 예를 들어, 본 개시내용은 유전자 활성을 조절할(예를 들어, 관심 서열을 삽입하거나 변경시키거나, 결실시킬) 수 있는 시스템 및 하나 이상의 그러한 시스템을 투여하여 뉴클레오티드에서 게놈 서열을 변경시켜 질환을 유발하는 병원성 돌연변이를 수

정함으로써 질환을 치료하는 방법을 제공한다.

[0007] 본 조성물 또는 방법의 특징은 다음 열거된 구현예들 중 하나 이상을 포함할 수 있다.

[0008] 1. 유전자 변형 폴리펩티드로서,

[0009] 표적 핵산 서열에 결합하는 DNA 결합 도메인(DBD) 및

[0010] 표 1 또는 23의 중합효소(Pol) 도메인, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열;

[0011] (여기서, DBD는 Pol 도메인에 이종성임); 및

[0012] Pol 도메인과 DBD 사이에 배치된 링커

[0013] 를 포함하는, 유전자 변형 폴리펩티드.

[0014] 2. 유전자 변형 폴리펩티드로서,

[0015] Cas 도메인(예를 들어, Cas 닉카제 도메인, 예를 들어 Cas9 닉카제 도메인);

[0016] 표 1 또는 23의 중합효소(Pol) 도메인, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열(여기서, Pol 도메인은 Cas 도메인의 C-말단임); 및

[0017] Pol 도메인과 Cas 도메인 사이에 배치된 링커

[0018] 를 포함하는, 유전자 변형 폴리펩티드.

[0019] 3. 구현예 1 또는 2에 있어서, 링커는 표 6의 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0020] 4. 선행하는 구현예에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 적어도 90% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0021] 5. 선행하는 구현예 중 어느 하나에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 적어도 95% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0022] 6. 선행하는 구현예 중 어느 하나에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 적어도 98% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0023] 7. 선행하는 구현예 중 어느 하나에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 적어도 99% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0024] 8. 선행하는 구현예 중 어느 하나에 있어서, Pol 도메인은 표 1 또는 23의 Pol 도메인에 대하여 100% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0025] 9. 선행하는 구현예 중 어느 하나에 있어서, 링커는 표 6의 링커 서열에 대하여 적어도 90% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0026] 10. 선행하는 구현예 중 어느 하나에 있어서, 링커는 표 6의 링커 서열에 대하여 적어도 95% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0027] 11. 선행하는 구현예 중 어느 하나에 있어서, 링커는 표 6의 링커 서열에 대하여 적어도 97% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0028] 12. 선행하는 구현예 중 어느 하나에 있어서, 링커는 표 6의 링커 서열에 대하여 100% 동일성을 갖는 서열을 갖는, 유전자 변형 폴리펩티드.

[0029] 13. 선행하는 구현예 중 어느 하나에 있어서, Cas 도메인은 표 4의 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 98%, 또는 99% 동일성을 갖는 서열을 포함하는, 유전자 변형 폴리펩티드.

[0030] 14. 선행하는 구현예 중 어느 하나에 있어서, Cas 도메인은 Cas 닉카제 도메인인, 유전자 변형 폴리펩티드.

[0031] 15. 선행하는 구현예 중 어느 하나에 있어서, Cas 도메인은 Cas9 닉카제 도메인인, 유전자 변형 폴리펩티드.

[0032] 16. 선행하는 구현예 중 어느 하나에 있어서, Cas 도메인은 N863A 돌연변이를 포함하는, 유전자 변형 폴리펩티드.

[0033] 17. 선행하는 구현예 중 어느 하나에 있어서, NLS를 포함하는, 예를 들어 유전자 변형 폴리펩티드는 2개의 NLS를 포함하는, 유전자 변형 폴리펩티드.

[0034] 18. 선행하는 구현예 중 어느 하나에 있어서, Cas9 도메인의 NLS N-말단을 포함하는, 유전자 변형 폴리펩티드.

[0035] 19. 선행하는 구현예 중 어느 하나에 있어서, Pol 도메인의 NLS C-말단을 포함하는, 유전자 변형 폴리펩티드.

[0036] 20. 선행하는 구현예 중 어느 하나에 있어서, Cas9 도메인의 N-말단인 제1 NLS 및 Pol 도메인의 C-말단인 제2 NLS를 포함하는, 유전자 변형 폴리펩티드.

[0037] 21. 선행하는 구현예 중 어느 하나의 유전자 변형 폴리펩티드를 인코딩하는 핵산(예를 들어, DNA 또는 RNA, 예를 들어 mRNA).

[0038] 22. 구현예 1 내지 20 중 어느 하나의 유전자 변형 폴리펩티드 또는 구현예 21의 핵산을 포함하는 세포.

[0039] 23. 시스템으로서,

[0040] i) 구현예 1 내지 20 중 어느 하나의 유전자 변형 폴리펩티드, 및

[0041] ii) a) 표적 핵산 서열의 일부에 상보적인 gRNA 스페이서;

[0042] b) 유전자 변형 폴리펩티드의 Cas 도메인에 결합하는 gRNA 스캐폴드;

[0043] c) 이종 목적 서열; 및

[0044] d) 프라이머 결합 부위 서열(PBS 서열)

[0045] 을 포함하는 주형 핵산(예를 들어, 주형 RNA)

[0046] 을 포함하는, 시스템.

[0047] 24. 구현예 23에 있어서, 주형 핵산은 RNA를 포함하는, 시스템.

[0048] 25. 구현예 23 또는 24에 있어서, 주형 핵산은 DNA를 포함하는, 시스템.

[0049] 26. 구현예 23 내지 25 중 어느 하나에 있어서, gRNA 스페이서 및 gRNA 스캐폴드는 RNA를 포함하는, 시스템.

[0050] 27. 구현예 23 내지 26 중 어느 하나에 있어서, 이종 목적 서열은 DNA를 포함하고 PBS 서열은 RNA를 포함하는, 시스템.

[0051] 28. 구현예 23 내지 26 중 어느 하나에 있어서, 이종 목적 서열 및 PBS 서열은 DNA를 포함하는, 시스템.

[0052] 29. 세포(예를 들어, 인간 세포)에서 표적 핵산을 변형시키는 방법으로서, 세포를 구현예 23 내지 28 중 어느 하나의 시스템, 또는 이를 인코딩하는 핵산과 접촉시키는 단계를 포함하며, 이에 의해 표적 핵산을 변형시키는, 방법.

[0053] 30. 유전적 결함과 연관된 질환 또는 병태를 갖는 대상체를 치료하는 방법으로서,

[0054] 대상체에게 선행하는 구현예 중 어느 하나의 시스템, 폴리펩티드, 주형 RNA 또는 이를 인코딩하는 DNA를 투여하여, 이에 의해 유전적 결합과 연관된 질환 또는 병태를 갖는 대상체를 치료하는 단계

[0055] 를 포함하는 방법.

[0056] 31. 구현예 30에 있어서, 유전적 결함과 연관된 질환 또는 병태는 표 12 내지 15 중 임의의 것에 열거된 적응증이고/이거나, 유전적 결함은 표 12 내지 15 중 임의의 것에 열거된 유전자의 결함인, 방법.

[0057] 32. 구현예 30 또는 31에 있어서, 대상체는 인간 환자인, 방법.

[0058] 일 양태에서, 본 개시내용은 (a) 표적 프라이밍된 역전사가 가능한 유전자 변형 폴리펩티드를 인코딩하는 핵산으로서, 폴리펩티드는 (i) 중합효소(Pol) 도메인 및 (ii) DNA에 결합하고 엔도뉴클레아제 활성을 갖는 Cas9 닉

카제를 포함하는, 핵산, 및 (b) (i) 표적 유전자의 제1 부분에 상보적인 gRNA 스페이서, (ii) 폴리펩티드에 결합하는 gRNA 스캐폴드, (iii) 유전자를 변형시키는 돌연변이 영역을 포함하는 이종 목적 서열, 및 (iv) 주형 RNA의 3' 말단에서 표적 DNA 가닥에 대하여 100% 상동성의 적어도 3, 4, 5, 6, 7 또는 8개의 염기를 포함하는 프라이머 결합 부위(PBS) 서열을 포함하는 동일한 가닥에 리보뉴클레오티드 및 데옥시리보뉴클레오티드 잔기를 둘 다 갖는 주형 RNA, DNA, 또는 하이브리드를 포함하는 유전자를 변형시키기 위한 시스템에 관한 것이다.

[0059] gRNA 스페이서는 주형 RNA의 5' 말단에서 표적 DNA에 대하여 100% 상동성의 적어도 15개 염기를 포함할 수 있다. 주형 RNA는 표적 DNA 가닥에 대하여 적어도 80% 상동성의 적어도 5개 염기를 포함하는 PBS 서열을 추가로 포함할 수 있다. 주형 RNA는 하나 이상의 화학적 변형을 포함할 수 있다.

[0060] 유전자 변형 폴리펩티드의 도메인은 펩티드 링커에 의해 연결될 수 있다. 폴리펩티드는 하나 이상의 펩티드 링커를 포함할 수 있다. 유전자 변형 폴리펩티드는 핵 국재화 신호를 추가로 포함할 수 있다. 폴리펩티드는 하나 초과의 핵 국재화 신호, 예를 들어 다중 인접 핵 국재화 신호, 또는 폴리펩티드의 상이한 영역에서의 하나 이상의 핵 국재화 신호, 예를 들어 폴리펩티드의 N-말단에서의 하나 이상의 핵 국재화 신호 및 폴리펩티드의 C-말단에서의 하나 이상의 핵 국재화 신호를 포함할 수 있다. 유전자 변형 폴리펩티드를 인코딩하는 핵산은 하나 초과의 인테인 도메인을 인코딩할 수 있다.

[0061] 표적 세포에 시스템을 도입하면 적어도 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 500, 또는 1000개 염기쌍의 외인성 DNA의 삽입이 초래될 수 있다. 표적 세포에 시스템을 도입하면 결실이 초래될 수 있으며, 여기서 결실은 삽입의 상류 또는 하류에 있는 게놈 DNA의 2, 3, 4, 5, 10, 50, 또는 100개 염기쌍 미만이다. 표적 세포에 시스템을 도입하면 치환, 예를 들어 1, 2, 또는 3개 뉴클레오티드, 예를 들어 연속 뉴클레오티드의 치환이 초래될 수 있다.

[0062] 이종 목적 서열은 적어도 5, 10, 25, 50, 100, 150, 200, 250, 300, 400, 500, 600, 또는 700개 염기쌍일 수 있다.

[0063] 일 양태에서, 본 개시내용은 상기 기재된 시스템 및 약제학적으로 허용가능한 부형제 또는 담체를 포함하는 약제학적 조성물에 관한 것이며, 여기서 약제학적으로 허용가능한 부형제 또는 담체는 플라스미드 벡터, 바이러스 벡터, 소포, 및 지질 나노입자로 이루어진 군으로부터 선택된다. 일 양태에서, 본 개시내용은, 예를 들어 상기 기재된 시스템이 2가지 별개의 부형제 또는 담체, 예를 들어 2개의 지질 나노입자, 2개의 바이러스 벡터, 또는 1개의 지질 나노입자 및 1개의 바이러스 벡터에 의해 전달되는 경우, 상기 기재된 시스템 및 다수의 약제학적으로 허용가능한 부형제 또는 담체를 포함하는 약제학적 조성물에 관한 것이며, 여기서 약제학적으로 허용가능한 부형제 또는 담체는 플라스미드 벡터, 바이러스 벡터, 소포, 및 지질 나노입자로 이루어진 군으로부터 선택된다. 바이러스 벡터는 아데노-연관 바이러스(AAV)일 수 있다.

[0064] 일 양태에서, 본 개시내용은 상기 기재된 시스템을 포함하는 숙주 세포(예를 들어, 포유동물 세포, 예를 들어 인간 세포)에 관한 것이다.

[0065] 일 양태에서, 본 개시내용은 세포, 조직 또는 대상체에서 인간 유전자의 돌연변이를 수정하는 방법에 관한 것이며, 방법은 상기 기재된 시스템을 세포, 조직 또는 대상체에게 투여하는 단계를 포함한다. 시스템은 생체 내, 시험관 내, 생체 외, 또는 제 위치(*in situ*)에서 도입될 수 있다. (a)의 핵산은 숙주 세포의 게놈 내로 통합될 수 있다. 일부 구현예에서, (a)의 핵산은 숙주 세포의 게놈 내로 통합되지 않는다. 일부 구현예에서, 이종 목적 서열은 숙주 세포 게놈의 하나의 표적 부위에만 삽입된다. 이종 목적 서열은 숙주 세포 게놈의 2개 이상의 표적 부위에서, 예를 들어 2개의 상동 염색체의 동일한 대응 부위에서 또는 동일하거나 상이한 염색체의 2개의 상이한 부위에서 삽입될 수 있다. 이종 목적 서열은 포유동물 폴리펩티드, 또는 이의 단편 또는 변이체를 인코딩할 수 있다. 시스템의 구성요소는 1, 2, 3, 4개, 또는 그 이상의 별개의 핵산 분자로 전달될 수 있다. 시스템은 전기천공에 의해, 또는 플라스미드 벡터, 바이러스 벡터, 소포, 및 지질 나노입자로부터 선택되는 적어도 하나의 비히클을 사용함으로써 숙주 세포 내로 도입될 수 있다.

도면의 간단한 설명

[0066] **도 1**은 본 명세서에 기재된 바와 같은 유전자 변형 시스템을 도시한다. 왼쪽 다이어그램은 Cas 닉카제 도메인(예를 들어, spCas9 N863A) 및 링커에 의해 연결된 Pol 도메인을 포함하는 유전자 변형 폴리펩티드를 나타낸다. 오른쪽 다이어그램은 5'에서 3'으로 gRNA 스페이서, gRNA 스캐폴드, 이종 목적 서열, 및 프라이머 결합 부위 서열(PBS 서열)을 포함하는 주형 핵산을 나타낸다. 이종 목적 서열은 표적 부위에 대해 하나 이상의 서열 차이를 포함하는 돌연변이 영역을 포함할 수 있다. 이종 목적 서열은 또한 돌연변이 영역에 측접하는, 사전 편집 상동

성 영역 및 사후 편집 상동성 영역을 포함할 수 있다. 이론에 구속되고자 함이 없이, 주형 핵산의 gRNA 스페이서는 게놈에서 표적 부위의 제2 가닥에 결합하고, 주형 핵산의 gRNA 스캐폴드는 유전자 변형 폴리펩티드에 결합하여, 예를 들어 게놈에서 유전자 변형 폴리펩티드를 표적 부위에 국재화하는 것으로 생각된다. 유전자 변형 폴리펩티드의 Cas 도메인은, 표적 부위(예를 들어, 표적 부위의 제1 가닥)에 닉을 생성하여, 예를 들어, 표적 부위의 제1 가닥 상에서 변경되는 부위에 인접한 서열에 PBS 서열이 결합하는 것을 가능하게 하는 것으로 생각된다. 유전자 변형 폴리펩티드의 Pol 도메인은 주형 핵산의 PBS 서열을 프라이머로, 그리고 주형 핵산의 이종 목적 서열을 주형으로 포함하는 상보적 서열에 결합되는 표적 부위의 제1 가닥을 사용하여, 예를 들어 이종 목적 서열에 상보적인 서열을 중합하는 것으로 생각된다. 이론에 구속되고자 함이 없이, 이후에 DNA 중합이 사전 편집 상동성 영역, 그 다음 돌연변이 영역, 그 다음 사후 편집 상동성 영역을 통해 진행할 수 있으며, 이에 의해 이종 목적 서열에 의해 특정된 돌연변이를 포함하는 DNA 가닥을 생산한다.

도 2는 인간 DNA 중합효소 세타의 예시적인 절단을 나타내는 다이어그램이다.

도 3a 및 3b는 HEK293 세포(도 3a) 및 U2OS 세포(도 3b)에서 표시된 주형 핵산 분자에 대한 Cas-Pol 유전자 변형 폴리펩티드에 의한 편집 활성을 나타내는 일련의 그래프이다.

발명을 실시하기 위한 구체적인 내용

[0067] ***정의***

[0068] 본 명세서에 사용된 바와 같은 용어 "발현 카세트"는 본 발명의 핵산 분자의 발현에 충분한 핵산 요소를 포함하는 핵산 작제물을 지칭한다.

[0069] 본 명세서에 사용된 바와 같은 "gRNA 스페이서"는 표적 핵산에 대해 상보성을 갖고 gRNA 스캐폴드와 함께 Cas 단백질을 표적 핵산에 표적화할 수 있는 핵산의 일부를 지칭한다.

[0070] 본 명세서에 사용된 바와 같은 "gRNA 스캐폴드"는 Cas 단백질에 결합할 수 있고 gRNA 스페이서와 함께 Cas 단백질을 표적 핵산에 표적화할 수 있는 핵산의 일부를 지칭한다. 일부 구현예에서, gRNA 스캐폴드는 crRNA 서열, 테트라루프, 및 tracrRNA 서열을 포함한다.

[0071] 본 명세서에 사용된 바와 같은 "유전자 변형 폴리펩티드"는 (예를 들어, 포유동물 숙주 세포, 예컨대 숙주 세포의 게놈 DNA 분자에서) 핵산 서열(예를 들어, 주형 핵산에 제공된 서열)을 표적 DNA 분자에 통합할 수 있는, 중합효소 또는 레트로바이러스 역전사효소를 포함하는 폴리펩티드, 또는 중합효소 또는 레트로바이러스 역전사효소에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 아미노산 서열 동일성을 갖는 아미노산 서열을 포함하는 폴리펩티드를 지칭한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 숙주 기구에 실질적으로 의존하지 않으면서 서열을 통합할 수 있다. 일부 구현예에서, 유전자 변형 폴리펩티드는 서열을 게놈의 무작위 위치에 통합하고, 일부 구현예에서, 유전자 변형 폴리펩티드는 서열을 특정 표적 부위에 통합한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 종합적으로, 1) 주형 핵산의 결합, 2) 표적 DNA 분자의 결합, 및 3) 표적 DNA 내로 주형 핵산의 적어도 일부의 통합을 용이하게 하는 하나 이상의 도메인을 포함한다. 유전자 변형 폴리펩티드는 자연 발생 폴리펩티드뿐만 아니라, 예를 들어 자연 발생 서열에 대해 하나 이상의 아미노산 치환을 갖는, 상기한 조작된 변이체를 둘 다 포함한다. 유전자 변형 폴리펩티드는 또한, 예를 들어 이종성 하위도메인 또는 다른 치환된 도메인의 대체 또는 융합을 통해, 다른 야생형 도메인의 이종성 융합(또는 다른 접합체)뿐만 아니라 변형된 도메인의 융합을 통해서든 간에, 예를 들어 상기 언급된 도메인 중 하나 이상이 서로 이종성인 경우, 이종성 작제물을 포함한다. 본 명세서에 제공된 방법에서 사용될 수 있는 예시적인 유전자 변형 폴리펩티드, 및 이를 포함하는 시스템 및 이를 사용하는 방법은, 예를 들어 PCT/US2021/020948에 기재되어 있으며, 이는 레트로바이러스 역전사효소를 포함하는 유전자 변형 폴리펩티드에 대하여 본 명세서에 참고로 포함된다. 일부 구현예에서, 유전자 변형 폴리펩티드는 서열을 유전자에 통합시킨다. 일부 구현예에서, 유전자 변형 폴리펩티드는 서열을 유전자 외부의 서열에 통합시킨다. 본 명세서에 사용된 바와 같은 "유전자 변형 시스템"은 유전자 변형 폴리펩티드 및 주형 핵산을 포함하는 시스템을 지칭한다.

[0072] 본 명세서에 사용된 바와 같은 용어 "도메인"은 생체분자의 특정된 기능에 기여하는 생체분자의 구조를 지칭한다. 도메인은 생체분자의 인접(contiguous) 영역(예를 들어, 인접 서열) 또는 별개의 비인접 영역(예를 들어, 비인접 서열)을 포함할 수 있다. 단백질 도메인의 예는 엔도뉴클레아제 도메인, DNA 결합 도메인, 중합효소(Pol) 도메인, 동원 도메인, 역전사 도메인을 포함하지만 이에 제한되지 않고; 핵산의 도메인의 예는 조절 도메인, 예컨대 전사 인자 결합 도메인이다. 일부 구현예에서, 도메인(예를 들어, Cas 도메인)은 2개 이상의 더 작

은 도메인(예를 들어, DNA 결합 도메인 및 엔도뉴클레아제 도메인)을 포함할 수 있다.

[0073] 본 명세서에 사용된 바와 같은 용어 "외인성"은, 생체분자(예컨대, 핵산 서열 또는 폴리펩티드)와 관련하여 사용되는 경우, 그 생체분자가 인간의 손에 의해 숙주 게놈, 세포 또는 유기체 내로 도입되었음을 의미한다. 예를 들어, 재조합 DNA 기법 또는 기타 방법을 사용하여 기존 게놈, 세포, 조직 또는 대상체로 첨가됨에 따른 핵산은 기존 핵산 서열, 세포 조직 또는 대상체에 대해 외인성이다.

[0074] 본 명세서에 사용된 바와 같은 "제1 가닥" 및 "제2 가닥"은 표적 DNA의 개별 DNA 가닥을 설명하기 위해 사용되며, Pol 도메인이 중합을 개시하는 가닥, 예를 들어 표적 프라이밍된 합성이 개시된 위치를 기준으로 2개의 DNA 가닥을 구분한다. 제1 가닥은 Pol 도메인이 중합을 개시하는, 예를 들어 표적 프라이밍된 합성이 개시되는 표적 DNA의 가닥을 지칭한다. 제2 가닥은 표적 DNA의 다른 가닥을 지칭한다. 제1 및 제2 가닥 명칭은 다른 양태에서는 표적 부위 DNA 가닥을 설명하지 않으며; 예를 들어, 일부 구현예에서, 제1 및 제2 가닥은 본 명세서에 기재된 폴리펩티드에 의해 닉이 생기지만, '제1' 및 '제2' 가닥이라는 명칭은 그러한 닉이 발생하는 순서와는 관계가 없다.

[0075] "게놈 세이프 하버 부위"(GSH 부위)는, 예를 들어 삽입된 유전 요소가 숙주 세포 또는 유기체에 위험을 제기하는 숙주 게놈에서의 현저한 변경을 일으키지 않도록, 새로운 유전 물질의 통합을 수용할 수 있는 숙주 게놈 내 부위이다. GSH 부위는 다음 기준 중 일반적으로 1, 2, 3, 4, 5, 6, 7, 8 또는 9개를 충족한다: (i) 암-관련 유전자로부터 >300kb 떨어져 있음; (ii) miRNA/기타 기능성 소형 RNA로부터 >300kb 떨어져 있음; (iii) 5' 유전자 말단으로부터 >50kb 떨어져 있음; (iv) 복제 원점으로부터 >50kb 떨어져 있음; (v) 초보존된(ultraconservered) 요소로부터 >50kb 떨어져 있음; (vi) 낮은 전사 활성을 가짐(즉, mRNA +/- 25 kb 없음); (vii) 카피 수 가변 영역에 없음; (viii) 개방형 염색질에 존재함; 및/또는 (ix) 인간 게놈에 하나의 카피가 있는 유일한 것임. 이들 기준의 일부 또는 전부를 충족시키는 인간 게놈 내 GSH 부위의 예는, (i) 염색체 19 상의 AAV 바이러스의 통합의 자연 발생 부위인 아데노-연관 바이러스 부위 1(AAVS1); (ii) HIV-1 공동수용체로 알려진 케모카인 수용체 유전자인 케모카인(C-C 모티프) 수용체 5(CCR5) 유전자; (iii) 마우스 Rosa26 유전자좌의 인간 병렬상동체(ortholog); (iv) 리보솜 DNA("rDNA") 유전자자리를 포함한다. 추가의 GSH 부위가 알려져 있고, 예를 들어 문헌[Pellenz et al. epub August 20, 2018 (https://doi.org/10.1101/396390)]에 기재되어 있다.

[0076] 제2의 요소와 관련하여 제1 요소를 설명하기 위해 본 명세서에 사용된 바와 같이 용어 "이종"은, 제1 요소 및 제2 요소가 설명된 바와 같이 배치되어 자연에 존재하지 않음을 의미한다. 예를 들어, 이종 폴리펩티드, 핵산 분자, 작제물 또는 서열은 (a) 그가 발현되는 세포에서 천연이 아닌 폴리펩티드, 핵산 분자, 또는 폴리펩티드 또는 핵산 분자 서열의 일부, (b) 그의 천연 상태에 대해 변경 또는 돌연변이된 폴리펩티드, 핵산 분자, 또는 폴리펩티드 또는 핵산 분자의 일부, 또는 (c) 유사한 조건 하에서 천연 발현 수준과 비교하여 변경된 발현을 갖는 폴리펩티드 또는 핵산 분자를 지칭한다. 예를 들어, 이종 조절 서열(예를 들어, 프로모터, 인핸서)은 유전자 또는 핵산 분자의 발현을, 그 유전자 또는 핵산 분자가 자연에서 보통 발현되는 것과 상이한 방식으로 조절하는 데 사용될 수 있다. 또 다른 예에서, 폴리펩티드 또는 핵산 서열의 이종 도메인(예를 들어, 폴리펩티드의 DNA 결합 도메인 또는 폴리펩티드의 DNA 결합 도메인을 인코딩하는 핵산)은 다른 도메인에 대해 배치될 수 있거나, 폴리펩티드 또는 그의 인코딩 핵산의 다른 도메인 또는 일부에 대해 상이한 서열 또는 상이한 공급원으로부터의 것일 수 있다. 특정 구현예에서, 이종 핵산 분자는 천연 숙주 세포 게놈 내에 존재할 수 있지만, 변경된 발현 수준을 가질 수 있거나 또는 상이한 서열을 가질 수 있거나, 또는 둘 모두 가질 수 있다. 다른 구현예에서, 이종 핵산 분자는 숙주 세포 또는 숙주 게놈에 대해 내인성이 아닐 수 있지만, 대신 형질 전환(예를 들어, 형질감염(transfection), 전기천공)에 의해 숙주 세포 내로 도입되었을 수 있으며, 여기서 추가된 분자는 숙주 게놈 내로 통합될 수 있거나, 일시적으로(예를 들어, mRNA) 또는 한 세대 초과 동안 반안정적으로 염색체외 유전 물질(예를 들어, 에피좀 바이러스 벡터, 플라스미드 또는 기타 자기-복제 벡터)로 존재할 수 있다.

[0077] 본 명세서에 사용된 바와 같이, 표적 부위로의 서열의 "삽입"은, 예를 들어 편집되지 않은 표적 부위에 동족 위치가 없는 이종 목적 서열에 새로운 뉴클레오티드가 있는 경우, 표적 부위에 DNA 서열이 순 추가되는 것을 지칭한다. 일부 구현예에서, 표적 핵산 서열에 대한 PBS 서열 및 이종 목적 서열의 뉴클레오티드 정렬은 표적 핵산 서열에서 정렬 갭을 생성할 것이다.

[0078] 본 명세서에 사용된 바와 같은, 표적 부위에서 이종 목적 서열에 의해 생성된 "결실"은, 예를 들어 이종 목적 서열에 동족 위치가 없는 편집되지 않은 표적 부위에 뉴클레오티드가 있는 경우, 표적 부위에 DNA 서열이 순 결실되는 것을 지칭한다. 일부 구현예에서, 표적 핵산 서열에 대한 PBS 서열 및 이종 목적 서열의 뉴클레오티드 정렬은 PBS 서열 및 이종 목적 서열을 포함하는 분자에서 정렬 갭을 생성할 것이다.

[0079] 본 명세서에 사용된 바와 같은 용어 "역 말단 반복부" 또는 "ITR"은, 이들의 대칭성으로 인해 그렇게 명명된 AAV 바이러스 시스-요소를 지칭한다. 이들 요소는 AAV 게놈의 효율적인 증식을 촉진한다. ITR 기능을 위한 최소 요소는, Rep-결합 부위(RBS; AAV2의 경우 5'-GCGCGCTCGCTCGCTC-3') 및 말단 분해(resolution) 부위(TRS; AAV2의 경우 5'-AGTTGG-3')에 더하여, 헤어핀 형성을 가능하게 하는 가변 회문 서열인 것으로 가정된다. 본 발명에 따르면, ITR은 적어도 이들 3개의 요소(RBS, TRS 및 헤어핀 형성을 가능하게 하는 서열)을 포함한다. 또한, 본 발명에서 용어 "ITR"은 알려진 천연 AAV 혈청형의 ITR(예를 들어, 혈청형 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 또는 11 AAV의 ITR), 상이한 혈청형으로부터 유래된 ITR 요소들의 융합에 의해 형성된 키메라 ITR, 및 이의 기능성 변이체를 지칭한다. "기능성 변이체"는, 알려진 ITR과 적어도 80%, 85%, 90%, 바람직하게는 적어도 95%의 서열 동일성을 나타내어, Rep 단백질의 존재 하에 상기 ITR을 포함하는 서열의 증식을 가능하게 하는 서열을 지칭한다.

[0080] 본 명세서에 사용된 바와 같은 용어 "돌연변이 영역"은 표적 핵산의 상응하는 서열에 비해 하나 이상의 서열 차이를 갖는 주형 핵산의 영역을 지칭한다. 서열 차이는, 예를 들어 치환, 삽입, 프레임시프트, 또는 결실을 포함할 수 있다.

[0081] 용어 "돌연변이된"이 핵산 서열에 적용되는 경우, 이는 핵산 서열 내 뉴클레오티드가 참조(예를 들어, 천연) 핵산 서열과 비교하여 삽입, 결실 또는 변경됨을 의미한다. 유전자좌에서 단일 변경이 이루어질 수 있거나(점 돌연변이), 또는 단일 유전자좌에서 다수의 뉴클레오티드가 삽입, 결실, 또는 변경될 수 있다. 추가로, 핵산 서열 내의 임의의 수의 유전자자리에서 하나 이상의 변경이 이루어질 수 있다. 핵산 서열은 당업계에 알려진 임의의 방법에 의해 돌연변이될 수 있다.

[0082] "핵산 분자"는 제한 없이, 상보적 DNA("cDNA"), 게놈 DNA("gDNA"), 및 메신저 RNA("mRNA")를 포함하는 RNA 및 DNA 분자를 모두 지칭하며, 본 명세서에 기재된 바와 같은, RNA 주형과 같이 화학적으로 합성되거나 재조합적으로 생성된 것들과 같은 합성 핵산 분자를 또한 포함한다. 핵산 분자는 이중 가닥 또는 단일 가닥, 환형 또는 선형일 수 있다. 단일 가닥인 경우, 핵산 분자는 센스 가닥 또는 안티센스 가닥일 수 있다. 달리 나타내지 않는 한, 그리고 일반 형식 "서열번호"로 본 명세서에 기재된 모든 서열에 대한 예로서, "서열번호 1을 포함하는 핵산"은 적어도 일부가 (i) 서열번호 1의 서열, 또는 (ii) 서열번호 1에 상보적인 서열을 갖는 핵산을 지칭한다. 둘 사이의 선택은 서열번호 1이 사용되는 맥락에 따라 결정된다. 예를 들어, 핵산이 프로브로 사용되는 경우, 프로브가 원하는 표적에 상보적이어야 한다는 요건에 따라 둘 사이의 선택이 결정된다. 본 개시내용의 핵산 서열은 화학적 또는 생화학적으로 변형될 수 있거나, 또는 당업자에 의해 쉽게 이해될 것인 바와 같이 비-천연 또는 유도체화된 뉴클레오티드 염기를 함유할 수 있다. 그러한 변형은, 예를 들어 표지, 메틸화, 하나 이상의 자연에 존재하는 뉴클레오티드의 유사체로의 치환, 인터-뉴클레오티드 변형, 예컨대 비하전된 연결(예를 들어, 메틸 포스포네이트, 포스포트리에스테르, 포스포아미데이트, 카르바메이트 등), 하전된 연결(예를 들어, 포스포로티오에이트, 포스포로디티오에이트 등), 현수 모이어티(예를 들어, 폴리펩티드), 인터칼레이터(예를 들어, 아크리딘, 소랄렌 등), 킬레이터, 알킬화제, 및 변형된 연결(예를 들어, 알파 아노머 핵산 등)을 포함한다. 화학적으로 변형된 염기(예를 들어, 하기 표 9 참조), 백본(예를 들어, 하기 표 10 참조), 및 변형된 캡(예를 들어, 하기 표 11 참조)도 포함된다. 수소 결합 및 기타 화학적 상호작용을 통해 지정된 서열에 결합하는 능력 면에서 폴리뉴클레오티드를 모방한 합성 분자도 포함된다. 그러한 분자는 당업계에 알려져 있으며, 예를 들어 펩티드 결합이 분자의 백본에서 포스페이트 연결을 치환한 것들, 예를 들어 펩티드 핵산(PNA)을 포함한다. 기타 변형은, 예를 들어 리보스 고리가 가교 모이어티 또는 "잠금" 핵산(LNA)에서 발견되는 변형과 같은 기타 구조를 함유한 유사체를 포함할 수 있다. 다양한 구현예에서, 핵산은 추가의 유전자 요소, 예컨대 조직-특이적 발현-제어 서열(들)(예를 들어, 조직-특이적 프로모터 및 조직-특이적 마이크로RNA 인식 서열), 및 추가의 요소, 예컨대 역전 반복부(예를 들어, 역 말단 반복부, 예컨대 바이러스, 예를 들어 AAV ITR로부터의 요소 또는 이로부터 유래된 요소) 및 탠덤 반복부(tandem repeat), 역전 반복부/직접 반복부, 상동성 영역(표적 DNA에 대해 다양한 정도의 상동성을 갖는 절편), 비번역 영역(UTR)(5', 3', 또는 5' 및 3' UTR), 및 상기 이들의 다양한 조합과 작동적으로 연관된다. 본 발명에 의해 제공된 시스템의 핵산 요소는 다양한 토폴로지에 제공될 수 있으며, 이는 단일 가닥, 이중 가닥, 환형, 선형, 개방형 말단을 갖는 선형, 폐쇄형 말단을 갖는 선형, 및 이들의 특정 형태, 예컨대 도기본(doggybone) DNA(dbDNA), 폐쇄형 말단 DNA(ceDNA)를 포함한다.

[0083] 본 명세서에 사용된 바와 같은 "유전자 발현 단위"는 적어도 하나의 이펙터 서열에 작동가능하게 연결된 적어도 하나의 조절 핵산 서열을 포함하는 핵산 서열이다. 제1 핵산 서열이 제2 핵산 서열과 기능적 관계에 위치할 때, 제1 핵산 서열은 제2 핵산 서열과 작동가능하게 연결된다. 예를 들어, 프로모터 또는 인핸서가 코딩 서열의 전사 또는 발현에 영향을 미치는 경우, 프로모터 또는 인핸서는 코딩 서열에 작동가능하게 연결된다. 작동가능하

게 연결된 DNA 서열은 인접하거나 인접하지 않을 수 있다. 2개의 단백질-코딩 영역을 결합하는 데 필요한 경우, 작동가능하게 연결된 서열은 동일한 판독 프레임에 존재할 수 있다.

[0084] 본 명세서에 사용된 바와 같은 용어 "숙주 게놈" 또는 "숙주 세포"는, 단백질 및/또는 유전 물질이 도입된 세포 및/또는 그의 게놈을 지칭한다. 그러한 용어는 특정한 대상체 세포 및/또는 게놈뿐만 아니라, 그러한 세포의 자손 및/또는 그러한 세포의 자손의 게놈도 지칭하는 것으로 이해되어야 한다. 본 명세서에 사용된 바와 같은. 돌연변이 또는 환경적 영향으로 인해 후속 세대에서 특정 변형이 일어날 수 있기 때문에, 그러한 자손은 실제로 모세포와 동일하지 않을 수 있지만, 본 명세서에 사용된 바와 같은 용어 "숙주 세포"의 범주 내에 여전히 포함된다. 숙주 게놈 또는 숙주 세포는 배양물에서 성장된 단리된 세포 또는 세포주일 수 있거나, 또는 그러한 세포 또는 세포주로부터 단리된 게놈 물질, 또는 살아있는 조직 또는 유기체를 구성하는 숙주 세포 또는 숙주 게놈일 수 있다. 일부 경우에서, 숙주 세포는 예를 들어, 본 명세서에 기재된 바와 같은 동물 세포 또는 식물 세포일 수 있다. 특정 경우에, 숙주 세포는 포유동물 세포, 인간 세포, 조류 세포, 파충류 세포, 소 세포, 말 세포, 돼지 세포, 염소 세포, 양 세포, 닭 세포, 또는 칠면조 세포일 수 있다 특정 경우에, 숙주 세포는 옥수수 세포, 콩 세포, 밀 세포 또는 쌀 세포일 수 있다.

[0085] 본 명세서에 사용된 바와 같은 "작동적 연관"은 1) 프로모터 및 2) 이종 목적 서열과 같은 2개의 핵산 서열 사이의 기능적 관계를 설명하며, 이러한 예에서 프로모터와 이종 목적 서열(예를 들어, 관심 유전자)는, 적합한 조건 하에서 프로모터가 이종 목적 서열의 발현을 구동하도록 배향됨을 의미한다. 예를 들어, 프로모터 및 이종 목적 서열을 보유하는 주형 핵산은 단일 가닥, 예를 들어 (+) 또는 (-) 배향일 수 있다. 이 주형에서 프로모터와 이종 목적 서열 사이의 "작동적 연관"은 주형 핵산이 특정 상태에서 전사될 것인지 여부에 관계없이, 적합한 상태에 있을 때(예를 들어, 필요한 촉매 인자, 및 NPT 등의 존재 하에 (+) 배향에 있음) 정확하게 전사되는 것을 의미한다. 작동적 연관은 다른 조직-특이적 발현 제어 서열(예컨대 인핸서, 억제자(repressor) 및 마이크로RNA 인식 서열), IR/DR, ITR, UTR, 또는 상동성 영역 및 이종 목적 서열 또는 레트로바이러스 RT 도메인을 인코딩하는 서열을 포함하는 핵산의 다른 쌍에 유사하게 적용된다.

[0086] 본 명세서에 사용된 바와 같은 용어 "프라이머 결합 부위 서열" 또는 "PBS 서열"은 표적 핵산 서열에 포함된 영역에 결합할 수 있는 주형 핵산의 일부를 지칭한다. 일부 경우에, PBS 서열은 표적 핵산 서열에 포함된 영역에 대하여 100% 동일성을 갖는 적어도 3, 4, 5, 6, 7, 또는 8개 염기를 포함하는 핵산 서열이다. 일부 구현예에서 프라이머 영역은 표적 핵산 서열에 포함된 영역에 대하여 100% 동일성을 갖는 적어도 5, 6, 7, 8개 염기를 포함한다. 이론에 구속되고자 함이 없이, 일부 구현예에서 주형 핵산이 PBS 서열 및 이종 목적 서열을 포함하는 경우, PBS 서열은 표적 핵산 서열에 포함된 영역에 결합하여, Pol 도메인이 해당 영역을 DNA 중합을 위한 프라이머로서 사용할 수 있게 하고, 이종 목적 서열을 주형으로서 사용할 수 있게 한다.

[0087] 본 명세서에 사용된 바와 같은 "스템-루프 서열"은, 예를 들어 적어도 2개(예를 들어, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 염기 쌍을 포함하는 스템 및 적어도 3개(예를 들어, 4개)의 염기 쌍을 갖는 루프를 갖는 스템-루프를 형성하기에 충분한 자기-상보성을 갖는 핵산 서열(예를 들어, RNA 서열)을 지칭한다. 스템은 미스매치 또는 벌지를 포함할 수 있다.

[0088] 본 명세서에 사용된 바와 같은 "조직-특이적 발현-제어 서열"은 표적 조직에서 이종 목적 서열을 포함하는 전사체의 수준을, 조직-특이적 방식으로, 예를 들어 표적외 조직(들)에 비해 표적상 조직(들)에서 우선적으로, 증가 또는 감소시키는 핵산 요소를 의미한다. 일부 구현예에서, 조직-특이적 발현-제어 서열은 우선적으로 표적 조직에서 이종 목적 서열을 포함하는 전사체의 전사, 활성 또는 반감기를 조직-특이적 방식으로, 예를 들어 표적외 조직(들)에 비해 표적상 조직(들)에서 우선적으로, 구동 또는 억제한다. 예시적인 조직-특이적 발현-제어 서열은 조직-특이적 프로모터, 억제자, 인핸서, 또는 이의 조합, 및 조직-특이적 마이크로RNA 인식 서열을 포함한다. 조직 특이성은 표적상(주형 핵산의 발현 또는 활성이 바람직하거나, 허용가능한 조직(들)) 및 표적외(주형 핵산의 발현 또는 활성이 요구지 않거나, 허용가능하지 않은 조직(들))를 지칭한다. 예를 들어, 조직-특이적 프로모터는, 표적외 조직에 비해 표적상 조직에서 우선적으로 발현을 구동한다. 대조적으로, 조직-특이적 마이크로RNA 인식 서열에 결합하는 마이크로RNA는 표적상 조직에 비해, 표적외 조직에서 우선적으로 발현되어, 표적외 조직에서 주형 핵산의 발현을 감소시킨다. 따라서, 표적 조직과 같은 동일한 조직에 특이적인 프로모터 및 마이크로RNA 인식 서열은 그 조직 내 연관된 서열의 전사, 활성, 또는 반감기와 관련하여 대조적인 기능(각각 일치하는 발현 수준으로 촉진 및 억제함, 즉, 표적외 조직에서는 높은 마이크로RNA 수준이고 표적상 조직에서는 낮은 수준인 한편, 프로모터는 표적상 조직에서는 높은 발현을, 표적외 조직에서는 낮은 발현을 구동함)을 갖는다.

[0089] ***목차***

[0090] 1) 서론

[0091] 2) 유전자 변형 시스템

[0092] a) 유전자 변형 시스템의 폴리펩티드 구성요소

[0093] i) 라이팅 도메인

[0094] ii) 엔도뉴클레아제 도메인 및 DNA 결합 도메인

[0095] (1) Cas 도메인을 포함하는 유전자 변형 폴리펩티드

[0096] (2) TAL 이펙터 및 징크 핑거 뉴클레아제

[0097] iii) 링커

[0098] iv) 유전자 변형 시스템을 위한 국재화 서열

[0099] v) 유전자 변형 폴리펩티드 및 시스템의 진화된 변이체

[0100] vi) 인테인

[0101] vii) 추가 도메인

[0102] b) 주형 핵산

[0103] i) gRNA 스페이서 및 gRNA 스캐폴드

[0104] ii) 이종 목적 서열

[0105] iii) PBS 서열

[0106] iv) 예시적인 주형 서열

[0107] c) 유도성 활성을 갖는 gRNA

[0108] d) 유전자 변형 시스템의 환형 RNA 및 리보자임

[0109] e) 표적 핵산 부위

[0110] f) 제2 가닥 닉 생성

[0111] 3) 조성물 및 시스템의 생산

[0112] 4) 적용

[0113] a) 치료 적용

[0114] b) 식물에의 적용

[0115] 5) 투여 및 전달

[0116] a) 조직 특이적 활성/투여

[0117] i) 프로모터

[0118] ii) 마이크로RNA

[0119] b) 바이러스 벡터 및 이의 구성요소

[0120] c) AAV 투여

[0121] d) 지질 나노입자

[0122] 6) 키트, 제조 물품, 및 약제학적 조성물

[0123] 7) 화학, 제조, 및 제어(CMC)

[0124] ***서론***

[0125] 본 개시내용은 세포, 조직 또는 대상체, 예를 들어 생체 내, 또는 시험관 내에서, DNA 서열 내 하나 이상의 위치에서 DNA 서열을 표적화, 편집, 변형, 또는 조작(예를 들어, 이종 목적 서열을 포유동물 게놈의 표적 부위 내로 삽입)하기 위한 조성물, 시스템, 및 방법에 관한 것이다. 이종 목적 DNA 서열은, 예를 들어 치환, 결실, 삽입, 예를 들어 코딩 서열, 조절 서열 또는 유전자 발현 단위를 포함할 수 있다.

[0126] 더 구체적으로, 본 개시내용은, 예를 들어 관심 서열에/로부터 하나 이상의 뉴클레오티드를 삽입, 결실, 또는 치환함으로써, 관심 게놈 DNA 서열을 변경시키기 위한 Pol 중합효소(Pol) 기반 시스템을 제공한다.

[0127] Cas9-관련 기능과 중합효소 기능의 융합체는 게놈 DNA의 변형을 구동하는 데 사용될 수 있다. 중합효소 기능은, 예를 들어 핵산 주형으로부터 DNA를 합성하는 DNA 중합효소일 수 있다. 핵산 주형은, 예를 들어 DNA 또는 RNA일 수 있다. RNA 주형을 사용할 수 있는 DNA 중합효소, 예를 들어 RNA-의존성 DNA 중합효소, 예를 들어 역전사효소의 경우, 유전자 변형 폴리펩티드 구성요소에 상기 기재된 바와 같은 주형 RNA가 제공될 수 있다. 그러한 일 예는, 주형으로서 DNA 또는 RNA를 사용하는 것으로 나타난 진핵생물 DNA 중합효소인 DNA 중합효소 Θ(POLQ에 의해 인코딩됨, 본 명세서에서 "POLQ" 또는 "PolΘ"로도 지칭될 수 있는 폴리펩티드 생성물)이다. Chandramouly et al. 2021 (DOI: 10.1126/sciadv.abf1771). 따라서 POLQ(또는 POLQ의 구성요소)에 (선택적으로 링커를 통해) 융합된 Cas9 기능은 (gRNA 스페이서를 통해) 변형을 위해 원하는 게놈 부위를 표적화하고, (gRNA 스캐폴드를 통해) Cas9 기능을 동원하며, (주형 RNA를 통해) DNA 합성을 프라이밍하고 주형으로 사용하는 주형 RNA를 이용하여 유기체 또는 세포에 투여될 때 게놈 변형을 위한 드라이버로 사용될 수 있다.

[0128] 본 개시내용은 부분적으로 유전자 변형 폴리펩티드 구성요소 및 주형 핵산(예를 들어, 주형 RNA) 구성요소를 포함하는 유전자 변형 시스템을 제공한다. 일부 구현예에서, 유전자 변형 시스템은 게놈 내 표적 부위로 변경을 도입하는 데 사용될 수 있다. 일부 구현예에서, 유전자 변형 폴리펩티드 구성요소는 라이팅 도메인(예를 들어, 역전사효소 도메인), DNA-결합 도메인, 및 엔도뉴클레아제 도메인(예를 들어, 닉카제 도메인)을 포함한다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 게놈 내 표적 부위에 결합하는(예를 들어, 표적 부위의 제2 가닥에 결합하는) 서열(예를 들어, gRNA 스페이서), 유전자 변형 폴리펩티드 구성요소에 결합하는 서열(예를 들어, gRNA 스캐폴드), 이종 목적 서열, 및 PBS 서열을 포함한다. 이론에 구속되고자 함이 없이, 주형 핵산(예를 들어, 주형 RNA)은 게놈 내 표적 부위의 제2 가닥에 결합하고, 유전자 변형 폴리펩티드 구성요소에 결합(예를 들어, 폴리펩티드 구성요소를 게놈 내 표적 부위에 국재화)하는 것으로 생각된다. 유전자 변형 폴리펩티드 구성요소의 엔도뉴클레아제(예를 들어, 닉카제)는, 표적 부위(예를 들어, 표적 부위의 제1 가닥)를 절단하여, 예를 들어, PBS 서열이 표적 부위의 제1 가닥 상에서 변경되는 부위에 인접한 서열에 결합하는 것을 가능하게 하는 것으로 생각된다. 폴리펩티드 구성요소의 라이팅 도메인(예를 들어, 역전사효소 도메인)은 주형 핵산의 PBS 서열을 프라이머로, 그리고 주형 핵산의 이종 목적 서열을 주형으로 포함하는 상보적 서열에 결합되는 표적 부위의 제1 가닥을 사용하여, 예를 들어 이종 목적 서열에 상보적인 서열을 중합하는 것으로 생각된다. 이론에 구속되고자 함이 없이, 적절한 이종 목적 서열의 선택은 표적 부위에서 하나 이상의 뉴클레오티드의 치환, 결실, 및/또는 삽입을 생성할 수 있는 것으로 생각된다.

[0129] ***유전자 변형 시스템***

[0130] 일부 구현예에서, 본 명세서에 기재된 유전자 변형 시스템은 다음을 포함한다: (A) 유전자 변형 폴리펩티드, 또는 유전자 변형 폴리펩티드를 인코딩하는 핵산(여기서, 유전자 변형 폴리펩티드는 (i) 역전사효소 도메인, 및 (x) DNA 결합 기능을 함유하는 엔도뉴클레아제 도메인 또는 (y) 엔도뉴클레아제 도메인 및 별도의 DNA 결합 도메인을 포함함); 및 (B) 주형 RNA. 일부 구현예에서, 유전자 변형 폴리펩티드는 숙주 기구에 실질적으로 의존하지 않으면서, (예를 들어, 포유동물 숙주 세포, 예컨대 숙주 세포의 게놈 DNA 분자에서) 주형 핵산 서열을 표적 DNA 분자에 통합할 수 있는 실질적으로 자율적인 단백질 기계로서 작용한다. 예를 들어, 유전자 변형 단백질은 DNA-결합 도메인, 역전사효소 도메인, 및 엔도뉴클레아제 도메인을 포함할 수 있다. 일부 구현예에서, DNA-결합 기능은 단백질을 DNA 서열로 지시하는 RNA 구성요소, 예를 들어 gRNA 스페이서를 포함할 수 있다. 다른 구현예에서, 유전자 변형 폴리펩티드는 역전사효소 도메인 및 엔도뉴클레아제 도메인을 포함할 수 있다. 유전자 변형 시스템의 RNA 주형 요소는 통상적으로 유전자 변형 폴리펩티드 요소에 대해 이종성이며 숙주 게놈에 삽입될(역전사될) 목적 서열을 제공한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 표적 프라이밍된 역전사가 가능하다. 일부 구현예에서, 유전자 변형 폴리펩티드는 제2 가닥 합성이 가능하다.

[0131] 일부 구현예에서 유전자 변형 시스템은 제2 폴리펩티드와 조합된다. 일부 구현예에서, 제2 폴리펩티드는 엔도뉴클레아제 도메인을 포함할 수 있다. 일부 구현예에서, 제2 폴리펩티드는 중합효소 도메인, 예를 들어 역전사효소 도메인을 포함할 수 있다. 일부 구현예에서, 제2 폴리펩티드는 DNA-의존성 DNA 중합효소 도메인을 포함할 수

있다. 일부 구현예에서, 제2 폴리펩티드는 예를 들어 제2 가닥 합성 또는 DNA 복구 해결에 기여함으로써 게놈 편집의 완료를 돕는다.

[0132] 기능성 유전자 변형 폴리펩티드는 관련없는 DNA 결합, 역전사, 및 엔도뉴클레아제 도메인으로 구성될 수 있다. 이러한 모듈 구조는 기능성 도메인들, 예를 들어 dCas9(DNA 결합), MMLV 역전사효소(역전사), FokI(엔도뉴클레아제)의 조합을 가능하게 한다. 일부 구현예에서, 다수의 기능성 도메인은 단일 단백질, 예를 들어 Cas9 또는 Cas9 닉카제(DNA 결합, 엔도뉴클레아제)로부터 발생할 수 있다.

[0133] 일부 구현예에서, 유전자 변형 폴리펩티드는 종합적으로, 1) 주형 핵산의 결합, 2) 표적 DNA 분자의 결합, 및 3) 표적 DNA 내로 주형 핵산의 적어도 일부의 통합을 용이하게 하는 하나 이상의 도메인을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 상응하는 자연 발생 서열에 대한 하나 이상의 아미노산 치환을 포함하는 조작된 폴리펩티드이다. 일부 구현예에서, 유전자 변형 폴리펩티드는, 예를 들어 이종성 하위도메인 또는 다른 치환된 도메인의 대체 또는 융합을 통해, 예를 들어 다른 야생형 도메인의 이종성 융합(또는 다른 접합체)뿐만 아니라 변형된 도메인의 융합을 통해, 서로에 대해 이종성인 2개 이상의 도메인을 포함한다. 예를 들어, 일부 구현예에서 다음 중 하나 이상이다: RT 도메인이 DBD에 대해 이종성이거나; DBD가 엔도뉴클레아제 도메인에 대해 이종성이거나; RT 도메인이 엔도뉴클레아제 도메인에 대해 이종성임.

[0134] 일부 구현예에서, 시스템에서 사용하기 위한 주형 RNA 분자는 5'에서 3'으로 (1) gRNA 스페이서; (2) gRNA 스캐폴드; (3) 이종 목적 서열 (4) 프라이머 결합 부위(PBS) 서열을 포함한다. 일부 구현예에서:

[0135] (1) gRNA 스페이서는 약 18 내지 22 nt, 예를 들어 20 nt의 gRNA 스페이서이다

[0136] (2) gRNA 스캐폴드는 주형을 Cas 도메인, 예를 들어 닉카제 Cas9 도메인과 회합시키기 위한 하나 이상의 헤어핀 루프, 예를 들어 1, 2, 또는 3개의 루프를 포함하는 gRNA 스캐폴드이다. 일부 구현예에서, gRNA 스캐폴드는 5'에서 3'으로 서열 GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGGACCGAGTCGGTCC(서열번호 4008)를 포함한다.

[0137] (3) 일부 구현예에서, 이종 목적 서열은, 예를 들어 길이가 7 내지 74, 예를 들어 10 내지 20, 20 내지 30, 30 내지 40, 40 내지 50, 50 내지 60, 60 내지 70, 또는 70 내지 80 nt 또는, 80 내지 90 nt이다. 일부 구현예에서, 서열의 첫 번째(가장 5') 염기는 C가 아니다.

[0138] (4) 일부 구현예에서, 닉 생성이 발생한 후 표적 프라이밍 서열에 결합하는 PBS 서열은, 예를 들어 3 내지 20 nt, 예를 들어 7 내지 15 nt, 예를 들어 12 내지 14 nt이다. 일부 구현예에서, PBS 서열은 40 내지 60% GC 함량을 갖는다.

[0139] 일부 구현예에서, 시스템과 회합된 제2 gRNA는 완전한 통합의 구동을 도울 수 있다. 일부 구현예에서, 제2 gRNA는 제1 가닥 닉으로부터 0 내지 200 nt 떨어진 위치, 예를 들어 제1 가닥 닉으로부터 0 내지 50, 50 내지 100, 100 내지 200 nt 떨어진 위치를 표적화할 수 있다. 일부 구현예에서, 제2 gRNA는 편집이 이루어진 후 그의 표적 서열에만 결합할 수 있으며, 예를 들어 gRNA는 이종 목적 서열에는 존재하지만, 최초 표적 서열에는 존재하지 않는 서열에 결합한다.

[0140] 일부 구현예에서, 본 명세서에 기재된 유전자 변형 시스템은 HEK293, K562, U2OS, 또는 HeLa 세포에서 편집을 수행하는 데 사용된다. 일부 구현예에서, 유전자 변형 시스템은 1차 세포, 예를 들어 E18.5 마우스 유래의 1차 피질 뉴런에서 편집을 수행하는 데 사용된다.

[0141] 일부 구현예에서, 본 명세서에 기재된 바와 같은 유전자 변형 폴리펩티드는 MoMLV RT 서열 또는 이의 변이체를 포함하는 (예를 들어, 본 명세서에 기재된 바와 같은) 역전사효소 또는 RT 도메인을 포함한다. 구현예에서, MoMLV RT 서열은 D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, 및 K103L로부터 선택되는 하나 이상의 돌연변이를 포함한다. 구현예에서, MoMLV RT 서열은 선택적으로 T306K 및/또는 W313F를 추가로 포함하는 D200N, L603W, 및 T330P와 같은 돌연변이의 조합을 포함한다.

[0142] 일부 구현예에서, (예를 들어, 본 명세서에 기재된 바와 같은) 엔도뉴클레아제 도메인은, 예를 들어 H840A 돌연변이를 포함하는 nCAS9를 포함한다.

[0143] 일부 구현예에서, (예를 들어, 본 명세서에 기재된 바와 같은 시스템의) 이종 목적 서열은 길이가 약 1 내지 50, 50 내지 100, 100 내지 200, 200 내지 300, 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700,

700 내지 800, 800 내지 900, 900 내지 1000개, 또는 그 이상의 뉴클레오티드이다.

[0144] 일부 구현예에서, RT 및 엔도뉴클레아제 도메인은, 예를 들어 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSS(서열번호 4006)를 포함하는 가요성 링커에 의해 연결된다.

[0145] 일부 구현예에서, 엔도뉴클레아제 도메인은 RT 도메인에 대해 N-말단이다. 일부 구현예에서, 엔도뉴클레아제 도메인은 RT 도메인에 대해 C-말단이다.

[0146] 일부 구현예에서, 시스템은, 예를 들어 본 명세서에 기재된 바와 같은, TPRT에 의해 이종 목적 서열을 표적 부위로 포함시킨다.

[0147] 일부 구현예에서, 유전자 변형 폴리펩티드는 DNA 결합 도메인을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 RNA 결합 도메인을 포함한다. 일부 구현예에서, RNA 결합 도메인은 B-박스 단백질, MS2 코트 단백질, dCas, 또는 본 명세서 표의 서열 요소의 RNA 결합 도메인을 포함한다. 일부 구현예에서, RNA 결합 도메인은 참조 RNA 결합 도메인보다 더 큰 친화도로 주형 RNA에 결합할 수 있다.

[0148] 일부 구현예에서, 유전자 변형 시스템은 적어도 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 또는 100개의 뉴클레오티드(및 선택적으로 500, 400, 300, 200, 또는 100개 이하의 뉴클레오티드)의, 표적 부위 내로의 삽입을 생성할 수 있다. 일부 구현예에서, 유전자 변형 시스템은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 또는 100개의 뉴클레오티드(및 선택적으로 500, 400, 300, 200, 또는 100개 이하의 뉴클레오티드)의 표적 부위 내로의 삽입을 생성할 수 있다. 일부 구현예에서, 유전자 변형 시스템은 적어도 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 또는 10 킬로베이스(및 선택적으로 1, 5, 10, 또는 20 이하의 킬로베이스)의 표적 부위 내로의 삽입을 생성할 수 있다. 일부 구현예에서, 유전자 변형 시스템은 적어도 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 또는 200개의 뉴클레오티드(및 선택적으로 500, 400, 300, 또는 200개 이하의 뉴클레오티드)의 결실을 생성할 수 있다. 일부 구현예에서, 유전자 변형 시스템은 적어도 81, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 또는 200개의 뉴클레오티드(및 선택적으로 500, 400, 300, 또는 200개 이하의 뉴클레오티드)의 결실을 생성할 수 있다. 일부 구현예에서, 유전자 변형 시스템은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 또는 200개의 뉴클레오티드(및 선택적으로 500, 400, 300, 또는 200개 이하의 뉴클레오티드)의 결실을 생성할 수 있다. 일부 구현예에서, 유전자 변형 시스템은 적어도 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5 또는 10 킬로베이스(및 선택적으로 1, 5, 10, 또는 20 이하의 킬로베이스)의 결실을 생성할 수 있다. 일부 구현예에서, 유전자 변형 시스템은 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 또는 100개 또는 그 이상의 뉴클레오티드의 표적 부위 내로의 치환을 생성할 수 있다. 일부 구현예에서, 유전자 변형 시스템은 1 내지 2, 2 내지 3, 3 내지 4, 4 내지 5, 5 내지 10, 10 내지 15, 15 내지 20, 20 내지 30, 30 내지 40, 40 내지 50, 50 내지 60, 60 내지 70, 70 내지 80, 80 내지 90, 또는 90 내지 100개 뉴클레오티드의 표적 부위에서의 치환을 생성할 수 있다.

[0149] 일부 구현예에서, 치환은 전위(transition) 돌연변이이다. 일부 구현예에서, 치환은 전환 돌연변이이다. 일부 구현예에서, 치환은 아데닌을 티민으로, 아데닌을 구아닌으로, 아데닌을 시토신으로, 구아닌을 티민으로, 구아닌을 시토신으로, 구아닌을 아데닌으로, 티민을 시토신으로, 티민을 아데닌으로, 티민을 구아닌으로, 시토신을 아데닌으로, 시토신을 구아닌으로, 또는 시토신을 티민으로 전환한다.

[0150] 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 유전자의 발현(예를 들어, 전사 또는 번역)을 증가시키거나 감소시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 프로모터 또는 인핸서, 예를 들어 전사 인자에 결합하는 서열에서 서열을 변경, 추가, 또는 결실시킴으로써 유전자의 발현(예를 들어, 전사 또는 번역)을 증가시키거나 감소시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 유전자의 번역을 변경시키고(예를 들어, 아미노산 서열을 변경함), 시작 또는 정지 코돈을 삽입 또는 결실시키고, 유전자의 번역 프레임을 변경 또는 고정시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은, 예를 들어 스플라이스 어셉터 수여 또는 공여 부위를 삽입, 결실, 또는 변경시킴으로써, 유전자의 스플라이싱을 변경시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 전사체 또는 단백질 반감기를 변경시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 (예를 들어, 세포질에서 미토콘드리아로, 세포질에서 세포외 공간으로(예를 들어, 분비 태그를 추가함)) 세포 내 단백질 국재화를 변경시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 (예를 들어, 잘못 접힌 단백질의 축적을 방지하기 위해) 단백질 접힘을 변경시킨다(예를 들어, 개

선시킴). 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 유전자, 예를 들어 유전자에 의해 인코딩된 단백질의 활성을 변경, 증가, 감소시킨다.

[0151] 예시적인 유전자 변형 폴리펩티드, 및 이를 포함하는 시스템 및 이를 사용하는 방법은, 예를 들어 PCT/US2021/020948에 기재되어 있으며, 이는 그 안의 아미노산 및 핵산 서열을 포함하여 레트로바이러스 RT 도메인에 관하여 본 명세서에 참고로 포함된다.

[0152] 예시적인 유전자 변형 폴리펩티드 및 레트로바이러스 RT 도메인 서열은 또한, 예를 들어 2021년 3월 4일에 출원된 국제 출원 PCT/US21/20948, 예를 들어 그 안의 표 30, 표 31, 및 표 44에 기재되어 있으며; 전체 출원은, 예를 들어 상기 서열 및 표의 레트로바이러스 RT에 관하여 본 명세서에 참고로 포함된다. 따라서, 본 명세서에 기재된 유전자 변형 폴리펩티드는 본 단락에 언급된 표 중 임의의 것에 따른 아미노산 서열, 또는 이의 도메인(예를 들어, 레트로바이러스 RT 도메인), 또는 상기한 것 중 임의의 것의 기능성 단편 또는 변이체, 또는 이에 대하여 적어도 70%, 80%, 85%, 90%, 95%, 또는 99% 동일성을 갖는 아미노산 서열을 포함할 수 있다.

[0153] 일부 구현예에서, 본 명세서에 기재된 시스템 중 임의의 것에서 사용하기 위한 폴리펩티드는 다중 상동성 단백질의 정렬된 폴리펩티드 서열에 기반한 분자 재구성 또는 조상(ancestral) 재구성일 수 있다. 일부 구현예에서, 본 명세서에 기재된 시스템 중 임의의 것에서 사용하기 위한 역전사효소 도메인은 동일하거나 상이한 공급원 유래의 역전사효소 도메인의 정렬에 기반한 분자 재구성 또는 조상 재구성일 수 있거나 특정 잔기에서 변형될 수 있다. 당업자는, 본 명세서에 제공된 수탁 번호를 기반으로, 예를 들어 BLAST(Basic Local Alignment Search Tool) 또는 보존된 도메인 분석을 위한 CD-검색과 같은 일상적인 서열 분석 도구를 사용함으로써, 폴리펩티드 또는 핵산 서열을 정렬할 수 있다. 분자 재구성은, 예를 들어 문헌[Ivics et al., Cell 1997, 501 - 510 ; Wagstaff et al., Molecular Biology and Evolution 2013, 88-99]에 기재된 접근법을 사용하여, 서열 공통성에 기반하여 창출될 수 있다.

[0154] *유전자 변형 시스템의 폴리펩티드 구성요소*

[0155] 일부 구현예에서, 유전자 변형 폴리펩티드는 DNA 표적 부위 결합, 주형 핵산(예를 들어, RNA) 결합, DNA 표적 부위 절단, 및 주형 핵산(예를 들어, RNA) 라이팅, 예를 들어, 역전사의 기능을 갖는다. 일부 구현예에서, 각각의 기능은 별개의 도메인 내에 포함된다. 일부 구현예에서, 기능은 둘 이상의 도메인(예를 들어 둘 이상의 도메인이 함께, 기능성을 나타냄)에 기인할 수 있다. 일부 구현예에서, 둘 이상의 도메인은 동일하거나 유사한 기능을 가질 수 있다(예를 들어, 둘 이상의 도메인은, 예를 들어 2개의 상이한 DNA 서열에 대해, 각각 독립적으로 DNA-결합 기능성을 가짐). 다른 구현예에서, 하나 이상의 도메인은 하나 이상의 기능을 가능하게 할 수 있으며, 예를 들어 Cas9 도메인은 DNA 결합 및 표적 부위 절단을 모두 가능하게 한다. 일부 구현예에서, 도메인은 모두 단일 폴리펩티드 내에 위치한다. 일부 구현예에서, 제1 도메인은 하나의 폴리펩티드에 있고, 제2 도메인은 제2 폴리펩티드에 있다. 예를 들어, 일부 구현예에서, 서열은 제1 폴리펩티드 및 제2 폴리펩티드 사이에서 분할될 수 있고, 예를 들어 여기서 제1 폴리펩티드는 역전사효소(RT) 도메인을 포함하고, 제2 폴리펩티드는 DNA-결합 도메인 및 엔도뉴클레아제 도메인, 예를 들어 닉카제 도메인을 포함한다. 추가의 예로서, 일부 구현예에서, 제1 폴리펩티드 및 제2 폴리펩티드는 각각 DNA 결합 도메인(예를 들어, 제1 DNA 결합 도메인 및 제2 DNA 결합 도메인)을 포함한다. 일부 구현예에서, 제1 및 제2 폴리펩티드는 스플릿(split)-인테인을 통해 번역 후 함께 합쳐져서 단일 유전자 변형 폴리펩티드를 형성할 수 있다.

[0156] 일부 양태에서, 본 명세서에 기재된 유전자 변형 폴리펩티드는 다음을 포함한다(예를 들어, 본 명세서에 기재된 시스템은 다음을 포함하는 유전자 변형 폴리펩티드를 포함함): 1) Cas 도메인(예를 들어, Cas 닉카제 도메인, 예를 들어 Cas9 닉카제 도메인); 2) 표 D의 역전사효소(RT) 도메인, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열(여기서, RT 도메인은 Cas 도메인의 C-말단임); RT 도메인과 Cas 도메인 사이에 배치된 링커(여기서, 링커는 표 D의 RT 도메인과 동일한 행의 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 가짐).

[0157] 일부 구현예에서, RT 도메인은 표 D의 RT 도메인과 100% 동일성을 갖는 서열을 갖고 링커는 표 D의 RT 도메인과 동일한 행의 링커 서열에 대하여 100% 동일성을 갖는 서열을 갖는다. 일부 구현예에서, Cas 도메인은 표 4의 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 98%, 또는 99% 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 서열 목록의 서열번호 1 내지 3332 중 임의의 것에 따른 아미노산 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 포함한다.

[0158] 일부 구현예에서, 유전자 변형 폴리펩티드는 Cas 도메인과 링커 사이에 GC 아미노산 서열, RT 도메인과 제2 NLS 사이에 AG 아미노산 서열, 및/또는 링커와 RT 도메인 사이에 GG 아미노산 서열을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 제1 NLS 및 Cas 도메인을 포함하는 서열번호 4000의 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 98%, 또는 99% 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 제2 NLS를 포함하는 서열번호 4001의 서열, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 98%, 또는 99% 동일성을 갖는 서열을 포함한다.

[0159] **예시적인 N-말단 NLS-Cas9 도메인**

[0160] MPAAKRVKLDGGDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMA
KVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYN
QLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN
LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNRED
LLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFD
KNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKII
KDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLT
FKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQ
LQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGG
LSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFV
YGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIL
PKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRM
LASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTL
TNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGDGG (서열번호 4000)

[0161] **예시적인 C-말단 서열**

[0162] AGKRTADGSEFEKRTADGSEFESPKKKAKVE (서열번호 4001)

[0163] *라이팅 도메인*

[0164] 본 발명의 특정 양태에서, 유전자 변형 시스템의 라이팅 도메인은 중합효소 기능을 이용하여 게놈 DNA의 변형을 구동한다. 중합효소 기능은, 예를 들어 핵산 주형으로부터 DNA를 합성하는 DNA 중합효소일 수 있다. 핵산 주형은, 예를 들어 DNA 또는 RNA일 수 있다. RNA 주형을 사용할 수 있는 DNA 중합효소, 예를 들어 RNA-의존성 DNA 중합효소, 예를 들어 역전사효소의 경우, 유전자 변형 폴리펩티드 구성요소에 상기 기재된 바와 같은 주형 RNA가 제공될 수 있다. 그러한 일 예는, 주형으로서 DNA 또는 RNA를 사용하는 것으로 나타난 진핵생물 DNA 중합효소인 DNA 중합효소 Θ(POLQ에 의해 인코딩됨, 본 명세서에서 "POLQ" 또는 "PolΘ"로도 지칭될 수 있는 폴리펩티드 생성물)이다. Chandramouly et al. 2021 (DOI: 10.1126/sciadv.abf1771). 따라서 POLQ(또는 POLQ의 구성요소)에 (선택적으로 링커를 통해) 융합된 Cas9 기능은 (gRNA 스페이서를 통해) 변형을 위해 원하는 게놈 부위를 표적화하고, (gRNA 스캐폴드를 통해) Cas9 기능을 동원하며, (주형 RNA를 통해) DNA 합성을 프라이밍하고 주형으로 사용하는 주형 RNA를 이용하여 유기체 또는 세포에 투여될 때 게놈 변형을 위한 드라이버로 사용될 수 있다.

[0165] 주형으로서 DNA를 사용하는 DNA 중합효소는 또한 Cas9 기능을 갖는 융합체에 통합되어 게놈 변형을 일으킬 수 있다. 이 경우, 주형 핵산은 공유 결합 또는 링커에 의해 서로 끝과 끝이 연결된, sgRNA와 DNA 주형의 융합체이다. pol세타는 DNA 주형으로부터 DNA를 합성할 수 있으므로, pol세타(또는 이의 구성요소)는 또한 이러한 방식으로 기능할 수 있다. 본 명세서에 기재된 바와 같이 주형 RNA를 지칭하는 구현예는 리보뉴클레오티드를 포함하는 주형 핵산, 또는 리보뉴클레오티드 및 데옥시리보뉴클레오티드를 포함하는 주형 핵산(예를 들어, 하나 이상의 DNA 영역에 커플링된 하나 이상의 RNA 영역을 포함하는 주형 RNA)을 포함할 수 있음이 이해된다. 일부 구현예에서, 본 명세서에 기재된 유전자 변형 폴리펩티드는 표 1에 따른 아미노산 서열, 또는 이에 대하여 적어도 70%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 중합효소 도메인을 포함한다. 일부 구현예에서, 본 명세서에 기재된 핵산은 표 1에 따른 아미노산 서열, 또는 이에 대하여 적어도 70%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 중합효소 도메인을 인코딩한다.

[0166] 본 명세서에 기재된 바와 같이 역전사효소 또는 역전사효소 도메인을 지칭하는 구현예는 표 1에 열거된 바와 같은 중합효소를 포함할 수 있음이 이해된다.

[0167] 표 1: 게놈 조작 폴리펩티드에 사용하기 위한 예시적인 중합효소.

표 1

[0168]

행	Pol	서열	길이
1	에타(η)	MATGQDRVVALVDMDCFFVQVEQRQNPHLRNKPCAVVQYKSWKGGGIIAVSYEARAFGVTRSMWADDAKKLCPDLLLAQVRES RGKANLTKYREASVEVMEIMSRFAVIERASIDEAYVDLTSAVQERLQKLQGQPISADLLPSTYIEGLPQGPTTAEETVQKEGM RKQGLFQWLDSLQIDNLTSPDLQLTVGAVIVEEMRAAIERETGFQCSAGISHNKVLAKLACGLNKPNRQTLVSHGSVPQLFSQ MPIRKIRSLGGKLGASVIEILGIEYMGELTQFTESQLQSHFGEKNGSWLYAMCRGIEHDPVKPRQLPKTIGCSKNFPGKTALA TREQVQWWLLQLAQELEERLTKDRNDNDRVATQLVVSIRVQGDKRLSSLRRCCALTRYDAHKMSHDAFTVIKNCNTSGIQTEW SPPLTMLFLCATKFSASAPSSSTDITSFLSSDPSSLPKVPVTSSEAKTQGSGPAVTATKKATTSLESFFQKAAERQKVKEASL SSLTAPTQAPMSNSPSKPSLPFQTSQSTGTEPFFKQKSLLLKQKQLNNSSVSSPQQNPWSNCKALPNSLPTEYPGCVPVCEGV SKLEESSKATPAEMDLAHNSQSMHASSASKSVLEVTQKATPNPSLLAAEDQVPCEKCGSLVPVWDMPEHMDYHFALELQKSFL QPHSSNPQVVSAVSHQGKRNPKSPLACTNKRPRPEGMQTLESFFKPLTH	713
2	이오타(ι)	MEKLGVEPEEEGGGDDDEEDAEAWAMELADVGAAASSQGVHDQVLPTPNASSRVIVHVDLDCFYAQVEMISNPELKDKPLGVQ QKYLVVTCNYEARKLGVKKLMNVRDAKEKCPQLVLVNGEDLTRYREMSYKVTELLEEFSPVVERLGFDENFVDLTEMVEKRLQ QLQSDELSAVTVSGHVYNNQSINLLDVLHIRLLVGSQIAAEMREAMYNQLGLTGCAGVASNKLLAKLVSGVFKPNQQTVLLPE SCQHLIHSLNHIKEIPGIGYKTAKCLEALGINSVRDLQTFSPKILEKELGISVAQRIQKLSFGEDNSPVILSGPPQSFSEEDS FKKCSSEVEAKNKIEELLASLLNRVCQDGRKPHTVRLIIRRYSSEKHYGRESRQCPIPSHVIQKLGTGNYDVMTPMVDILMKL FRNMVNVKMPFHLTLLSVCFCNLKALNTAKKGLIDYYLMPSLSTTSRSGKHSFKMKDTHMEDFPKDKETNRDFLPSGRIESTR TRESPLDTTNFSKEKDINEFPLCSLPEGVDQEVFKQLPVDIQEEILSGKSREKFQGKGSVSCPLHASRGVLSFFSKKQMQDIP INPRDHLSSSKQVSSVSPCEPGTSGFNSSSSSYMSSQKDYSYYLDNRLKDERISQGPKEPQGFHFTNSNPAVSAFHSFPNLQS EQLFSRNHTTDSHKQTVATDSHEGLTENREPDSVDEKITFPSDIDPQVFYELPEAVQKELLAEWKRAGSDFHIGHK	740
3	카파(κ)	MDSTKEKCDSYKDDLLLRMGLNDNKAGMEGLDKEKINKIIMEATKGSRFYGNELKKEKQVNQRIENMMQQKAQITSQQLRKAQ LQVDRFAMELEQSRNLSNTIVHIDMDAFYAAVEMRDNPELKDKPIAVGSMSMLSTSNYHARRFGVRAAMPGFIAKRLCPQLII VPPNFDKYRAVSKEVKEILADYDPNFMAMSLDEAYLNITKHLEERQNWPEDKRRYFIKMGSSVENDNPGKEVNKLSEHERSIS PLLFEESPSDVQPPGDPFQVNFEEQNNPQILQNSVVFGTSAQEVVKEIRFRIEQKTTLTASAGIAPNTMLAKVCSDKNKPNGQ YQILPNRQAVMDFIKDLPIRKVSGIGKVTEKMLKALGIITCTELYQQRALLSLLFSETSWHYFLHISLGLGSTHLTRDGERKS MSVERTFSEINKAEEQYSLCQELCSELAQDLQKERLKGRTVTIKLKNVNFEVKTRASTVSSVVSTAEEIFAIAKELLKTEIDA DFPHPLRLRLMGVRISSFPNEEDRKHQQRSIIGFLQAGNQALSATECTLEKTDKDKFVKPLEMSHKKSFFDKKRSERKWSHQD TFKCEAVNKQSFQTSQPFQVLKKKMNENLEISENSDDCQILTCPVCFRAQG시스LEALNKHVDECLDGPSISENFKMFSCSH VSATKVNKKENVPASSLCEKQDYEAHPKIKEISSVDCIALVDTIDNSSKAESIDALSNKHSKEECSSLPSKSFNIEHCHQNSS STVSLENEDVGSFRQEYRQPYLCEVKTGQALVCPVCNVEQKTSDLTLFNVHVDVCLNKSFIQELRKDKFNPVNQPKESSRSTG SSSGVQKAVTRTKRPGLMTKYSTSKKIKPNNPKHTLDIFFK	870
4	REV1	MRRGGWRKRAENDGWETWGGYMAAKVQKLEEQFRSDAAMQKDGTSSTIFSGVAIYVNGYTDPSAEELRKLMMLHGGQYHVYYS RSKTTHIIATNLPNAKIKELKGEKVIRPEWIVESIKAGRLLSYIPYQLYTKQSSVQKGLSFNPVCRPEDPLPGPSNIAKQLNN RVNHIVKKIETENEVKVNGMNSWNEEDENNDFSFVDLEQTSPGRKQNGIPHPRGSTAIFNGHTPSSNGALKTQDCLVPMVNSV ASRLSPAFSQEEDKAEKSSTDFRDCTLQQLQQSTRNTDALRNPHRTNSFSLSPLHSNTKINGAHHSTVQGPSSTKSTSSVSTF SKAAPSVPSKPSDCNFISNFYSHSRLHHISMWKCELTEFVNTLQRQSNGIFPGREKLKKMKTGRSALVVTDTGDMSVLNSPRH QSCIMHVDMDCFFVSVGIRNRPDLKGKPVAVTSNRGTGRAPLRPGANPQLEWQYYQNKILKGKAADIPDSSLWENPDSAQANG IDSVLSRAEIASCSYEARQLGIKNGMFFGHAKQLCPNLQAVPYDFHAYKEVAQTLYETLASYTHNIEAVSCDEALVDITEILA ETKLTPDEFANAVRMEIKDQTKCAASVGIGSNILLARMATRKAKPDGQYHLKPEEVDDFIRGQLVTNLPGVGHSMESKLASLG IKTCGDLQYMTMAKLQKEFGPKTGQMLYRFCRGLDDRPVRTEKERKSVSAEINYGIRFTQPKEAEAFLLSLSEEIQRRLEATG MKGKRLTLKIMVRKPGAPVETAKFGGHGICDNIARTVTLDQATDNAKIIGKAMLNMFHTMKLNISDMRGVGIHVNQLVPTNLN PSTCPSRPSVQSSHFPSGSYSVRDVFQVQKAKKSTEEEHKEVFRAAVDLEISSASRTCTFLPPFPAHLPTSPDTNKAESSGKW NGLHTPVSVQSRLNLSIEVPSPSQLDQSVLEALPPDLREQVEQVCAVQQAESHGDKKKEPVNGCNTGILPQPVGTVLLQIPEP QESNSDAGINLIALPAFSQVDPEVFAALPAELQRELKAAYDQRQRQGENSTHQQSASASVPKNPLLHLKAAVKEKKRNKKKKT IGSPKRIQSPLNNKLLNSPAKTLPGACGSPQKLIDGFLKHEGPPAEKPLEELSASTSGVPGLSSLQSDPAGCVRPPAPNLAGA VEFNDVKTLLREWITTISDPMEEDILQVVKYCTDLIEEKDLEKLDLVIKYMKRLMQQSVESVWNMAFDFILDNVQVVLQQTYG STLKVT	1251
5	베타(β)	MSKRKAPQETLNGGITDMLTELANFEKNVSQAIHKYNAYRKAASVIAKYPHKIKSGAEAKKLPGVGTKIAEKIDEFLATGKLR KLEKIRQDDTSSSINFLTRVSGIGPSAARKFVDEGIKTLEDLRKNEDKLNHHQRIGLKYFGDFEKRIPREEMLQMQDIVLNEV KKVDSEYIATVCGSFRRGAESSGDMDVLLTHPSFTSESTKQPKLLHQVVEQLQKVHFITDTLSKGETKFMGVCQLPSKNDEKE YPHRRIDIRLIPKDQYYCGVLYFTGSDIFNKNMRAHALEKGFTINEYTIRPLGVTGVAGEPLPVDSEKDIFDYIQWKYREPKD RSE	335
6	람다(λ)	MDPRGILKAFPKRQKIHADASSKVLAKIPRREEGEEAEEWLSSLRAHVVRTGIGRARAELFEKQIVQHGGQLCPAQGPGVTHI VVDEGMDYERALRLLRLPQLPPGAQLVKSAWLSLCLQERRLVDVAGFSIFIPSRYLDHPQPSKAEQDASIPPGTHEALLQTAL SPPPPPTRPVSPPQKAKEAPNTQAQPISDDEASDGEETQVSAADLEALISGHYPTSLEGDCEPSPAPAVLDKWVCAQPSSQKA TNHNLHITEKLEVLAKAYSVQGDKWRALGYAKAINALKSFHKPVTSYQEACSIPGIGKRMAEKIIEILESGHLRKLDHISESV PVLELFSNIWGAGTKTAQMWYQQGFRSLEDIRSQASLTTQQAIGLKHYSDFLERMPREEATEIEQTVQKAAQAFNSGLLCVAC GSYRRGKATCGDVDVLITHPDGRSHRGIFSRLLDSLRQEGFLTDDLVSQEENGQQQKYLGVCRLPGPGRRHRRLDIIVVPYSE FACALLYFTGSAHFNRSMRALAKTKGMSLSEHALSTAVVRNTHGCKVGPGRVLPTPTEKDVFRLLGLPYREPAERDW	575
7	뮤(μ)	MLPKRRRARVGSPSGDAASSTPPSTRFPGVAIYLVEPRMGRSRRAFLTGLARSKGFRVLDACSSEATHVVMEETSAEEAVSWQ ERRMAAAPPGCTPPALLDISWLTESLGAGQPVPVECRHRLEVAGPRKGPLSPAWMPAYACQRPTPLTHHNTGLSEALEILAEA AGFEGSEGRLLTFCRAASVLKALPSPVTTLSQLQGLPHFGEHSSRVVQELLEHGVCEEVERVRRSERYQTMKLFTQIFGVGVK TADRWYREGLRTLDDLREQPQKLTQQQKAGLQHHQDLSTPVLRSDVDALQQVVEEAVGQALPGATVTLTGGFRRGKLQGHDVD FLITHPKEGQEAGLLPRVMCRLQDQGLILYHQHQHSCCESPTRLAQQSHMDAFERSFCIFRLPQPPGAAVGGSTRPCPSWKAV RVDLVVAPVSQFPFALLGWTGSKLFQRELRRFSRKEKGLWLNSHGLFDPEQKTFFQAASEEDIFRHLGLEYLPPEQRNA	494

8	TdT	MDPPRASHLSPRKKRPRQTGALMASSPQDIKFQDLVVFILEKKMGTTRRAFLMELARRKGFRVENELSDSVTHIVAENNSGSD VLEWLQAQKVQVSSQPELLDVSWLIECIRAGKPVEMTGKHQLVVRRDYSDSTNPGPPKTPPIAVQKISQYACQRRTTLNNCNQ IFTDAFDILAENCEFRENEDSCVTFMRAASVLKSLPFTIISMKDTEGIPCLGSKVKGIIEEIIEDGESSEVKAVLNDERYQSF KLFTSVFGVGLKTSEKWFRMGFRTLSKVRSDKSLKFTRMQKAGFLYYEDLVSCVTRAEAEAVSVLVKEAVWAFLPDAFVTMTG GFRRGKKMGHDVDFLITSPGSTEDEEQLLQKVMNLWEKKGLLLYYDLVESTFEKLRLPSRKVDALDHFQKCFLIFKLPRQRVD SDQSSWQEGKTWKAIRVDLVLCPYERRAFALLGWTGSRQFERDLRRYATHERKMILDNHALYDKTKRIFLKAESEEEIFAHLG LDYIEPWERNA	509
9	세타 (Θ)	MNLLRRSGKRRRSESGSDSFSGSGGDSSASPQFLSGSVLSPPPGLGRCLKAAAAGECKPTVPDYERDKLLLANWGLPKAVLEK YHSFGVKKMFEWQAECLLLGQVLEGKNLVYSAPTSAGKTLVAELLILKRVLEMRKKALFILPFVSVAKEKKYYLQSLFQEVGI KVDGYMGSTSPSRHFSSLDIAVCTIERANGLINRLIEENKMDLLGMVVVDELHMLGDSHRGYLLELLLTKICYITRKSASCQA DLASSLSNAVQIVGMSATLPNLELVASWLNAELYHTDFRPVPLLESVKVGNSIYDSSMKLVREFEPMLQVKGDEDHVVSLCYE TICDNHSVLLFCPSKKWCEKLADIIAREFYNLHHQAEGLVKPSECPPVILEQKELLEVMDQLRRLPSGLDSVLQKTVPWGVAF HHAGLTFEERDIIEGAFRQGLIRVLAATSTLSSGVNLPARRVIIRTPIFGGRPLDILTYKQMVGRAGRKGVDTVGESILICKN SEKSKGIALLQGSLKPVRSCLQRREGEEVTGSMIRAILEIIVGGVASTSQDMHTYAACTFLAASMKEGKQGIQRNQESVQLGA IEACVMWLLENEFIQSTEASDGTEGKVYHPTHLGSATLSSSLSPADTLDIFADLQRAMKGFVLENDLHILYLVTPMFEDWTTI DWYRFFCLWEKLPTSMKRVAELVGVEEGFLARCVKGKVVARTERQHRQMAIHKRFFTSLVLLDLISEVPLREINQKYGCNRGQ IQSLQQSAAVYAGMITVFSNRLGWHNMELLLSQFQKRLTFGIQRELCDLVRVSLLNAQRARVLYASGFHTVADLARANIVEVE VILKNAVPFKSARKAVDEEEEAVEERRNMRTIWVTGRKGLTEREAAALIVEEARMILQQDLVEMGVQWNPCALLHSSTCSLTH SESEVKEHTFISQTKSSYKKLTSKNKSNTIFSDSYIKHSPNIVQDLNKSREHTSSFNCNFQNGNQEHQTCSIFRARKRASLDI NKEKPGASQNEGKTSDKKVVQTFSQKTKKAPLNFNSEKMSRSFRSWKRRKHLKRSRDSSPLKDSGACRIHLQGQTLSNPSLCE DPFTLDEKKTEFRNSGPFAKNVSLSGKEKDNKTSFPLQIKQNCSWNITLTNDNFVEHIVTGSQSKNVTCQATSVVSEKGRGVA VEAEKINEVLIQNGSKNQNVYMKHHDIHPINQYLRKQSHEQTSTITKQKNIIERQMPCEAVSSYINRDSNVTINCERIKLNTE ENKPSHFQALGDDISRTVIPSEVLPSAGAFSKSEGQHENFLNISRLQEKTGTYTTNKTKNNHVSDLGLVLCDFEDSFYLDTQS EKIIQQMATENAKLGAKDTNLAAGIMQKSLVQQNSMNSFQKECHIPFPAEQHPLGATKIDHLDLKTVGTMKQSSDSHGVDILT PESPIFHSPILLEENGLFLKKNEVSVTDSQLNSFLQGYQTQETVKPVILLIPQKRTPTGVEGECLPVPETSLNMSDSLLFDSF SDDYLVKEQLPDMQMKEPLPSEVTSNHFSDSLCLQEDLIKKSNVNENQDTHQQLTCSNDESIIFSEMDSVQMVEALDNVDIFP VQEKNHTVVSPRALELSDPVLDEHHQGDQDGGDQDERAEKSKLTGTRQNHSFIWSGASFDLSPGLQRILDKVSSPLENEKLKS MTINFSSLNRKNTELNEEQEVISNLETKQVQGISFSSSNNEVKSKIEMLENNANHDETSSLLPRKESNIVDDNGLIPPTPIPTS ASKLTFPGILETPVNPWKTNNVLQPGESYLFGSPSDIKNHDLSPGSRNGFKDNSPISDTSFSLQLSQDGLQLTPASSSSESLS IIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRSLTSSKTATIGSRFKQASSPQEIPIRDDGFPIKGCDDTLVVGLAVCWGG RDAYYFSLQKEQKHSEISASLVPPSLDPSLTLKDRMWYLQSCLRKESDKECSVVIYDFIQSYKILLLSCGISLEQSYEDPKVA CWLLDPDSQEPTLHSIVTSFLPHELPLLEGMETSQGIQSLGLNAGSEHSGRYRASVESILIFNSMNQLNSLLQKENLQDVFRK VEMPSQYCLALLELNGIGFSTAECESQKHIMQAKLDAIETQAYQLAGHSFSFTSSDDIAEVLFLELKLPPNREMKNQGSKKTL GSTRRGIDNGRKLRLGRQFSTSKDVLNKLKALHPLPGLILEWRRITNAITKVVFPLQREKCLNPFLGMERIYPVSQSHTATGR ITFTEPNIQNVPRDFEIKMPTLVGESPPSQAVGKGLLPMGRGKYKKGFSVNPRCQAQMEERAADRGMPFSISMRHAFVPFPGG SILAADYSQLELRILAHLSHDRRLIQVLNTGADVFRSIAAEWKMIEPESVGDDLRQQAKQICYGIIYGMGAKSLGEQMGIKEN DAACYIDSFKSRYTGINQFMTETVKNCKRDGFVQTILGRRRYLPGIKDNNPYRKAHAERQAINTIVQGSAADIVKIATVNIQK QLETFHSTFKSHGHREGMLQSDQTGLSRKRKLQGMFCPIRGGFFILQLHDELLYEVAEEDVVQVAQIVKNEMESAVKLSVKLK VKVKIGASWGELKDFDV	2590
10	뉴 (ν)	MENYEALVGFDLCNTPLSSVAQKIMSAMHSGDLVDSKTWGKSTETMEVINKSSVKYSVQLEDRKTQSPEKKDLKSLRSQTSRG SAKLSPQSFSVRLTDQLSADQKQKSISSLTLSSCLIPQYNQEASVLQKKGHKRKHFLMENINNENKGSINLKRKHITYNNLSE KTSKQMALEEDTDDAEGYLNSGNSGALKKHFCDIRHLDDWAKSQLIEMLKQAAALVITVMYTDGSTQLGADQTPVSSVRGIVV LVKRQAEGGHGCPDAPACGPVLEGFVSDDPCIYIQIEHSAIWDQEQEAHQQFARNVLFQTMKCKCPVICFNAKDFVRIVLQFF GNDGSWKHVADFIGLDPRIAAWLIDPSDATPSFEDLVEKYCEKSITVKVNSTYGNSSRNIVNQNVRENLKTLYRLTMDLCSKL KDYGLWQLFRTLELPLIPILAVMESHAIQVNKEEMEKTSALLGARLKELEQEAHFVAGERFLITSNNQLREILFGKLKLHLLS QRNSLPRTGLQKYPSTSEAVLNALRDLHPLPKIILEYRQVHKIKSTFVDGLLACMKKGSISSTWNQTGTVTGRLSAKHPNIQG ISKHPIQITTPKNFKGKEDKILTISPRAMFVSSKGHTFLAADFSQIELRILTHLSGDPELLKLFQESERDDVFSTLTSQWKDV PVEQVTHADREQTKKVVYAVVYGAGKERLAACLGVPIQEAAQFLESFLQKYKKIKDFARAAIAQCHQTGCVVSIMGRRRPLPR IHAHDQQLRAQAERQAVNFVVQGSAADLCKLAMIHVFTAVAASHTLTARLVAQIHDELLFEVEDPQIPECAALVRRTMESLEQ VQALELQLQVPLKVSLSAGRSWGHLVPLQEAWGPPPGPCRTESPSNSLAAPGSPASTQPPPLHFSPSFCL	900

[0169] **표 1a: 표 1에 열거된 중합효소의 특성**

[0170] [표 1a]

행	패밀리	Pol	숙주	유전자 기호	UniProt 수탁	UniProt 명칭	촉매 서브유닛 질량(kDa)	추가 서브유닛	생물학적 경로(들)	주형 핵산
1	Y	에타 (η)	호모 사피엔스(Homo sapiens)	POLH	Q9Y253	POLH_HUMAN	78	단량체	TLS	
2	Y	이오타 (ι)	호모 사피엔스	POLI	Q9UNA4	POLI_HUMAN	80	단량체	TLS	
3	Y	카파 (κ)	호모 사피엔스	POLK	Q9UBT6	POLK_HUMAN	76	단량체	TLS	
4	Y	REV1	호모 사피엔스	REV1	Q9UBZ9	REV1_HUMAN	138	단량체	TLS	
5	X	베타(β)	호모 사피엔스	POLB	P06746	DPOLB_HUMAN	39	단량체	BER	
6	X	람다 (λ)	호모 사피엔스	POLL	Q9UGP5	DPOLL_HUMAN	66	단량체	BER, NHEJ, TLS	
7	X	뮤(μ)	호모 사피엔스	POLM	Q9NP87	DPOLM_HUMAN	55	단량체	NHEJ, V(D)J	
8	X	TdT	호모 사피엔스	DNTT	P04053	TDT_HUMAN	56	단량체	V(D)J	주형-독립적
9	A	세타 (θ)	호모 사피엔스	POLQ	O75417	DPOLQ_HUMAN	290	단량체	TLS, MMEJ, RTDR	DNA 및/또는 RNA
10	A	뉴(ν)	호모 사피엔스	POLN	Q7Z5Q5	DPOLN_HUMAN	100	단량체	ICL 복구	

[0171]

[0172] 생물학적 경로에 있어서, TLS = 트랜스레시온(translesion) 합성; BER = 염기-절제 복구(base-excision repair); NHEJ = 비-상동성 말단 연결(non-homologous end joining); V(D)J = V(D)J 재조합 과정; MMEJ = 미세상동성-매개 말단 연결(microhomology-mediated end joining); RTDR = RNA-주형 DNA 복구.

[0173] 본 발명의 특정 양태에서, 유전자 변형 시스템의 라이팅 도메인은 역전사효소 활성을 가지며, 역전사효소 도메인(RT 도메인)으로도 지칭된다. 일부 구현예에서, RT 도메인은 RT 촉매부 및 RNA-결합 영역(예를 들어, 주형 RNA에 결합하는 영역)을 포함한다.

[0174] 일부 구현예에서, 역전사효소를 인코딩하는 핵산은, 예를 들어 인간 세포에 대해 개선된, 변경된 코돈 사용을 갖기 위해 그의 천연 서열에서 변경된다. 일부 구현예에서 역전사효소 도메인은 레트로바이러스 유래의 이종성 역전사효소이다. 일부 구현예에서, 유전자 변형 폴리펩티드를 포함하는 RT 도메인은 원래의 아미노산 서열로부터 돌연변이되었으며, 예를 들어 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 또는 100개 치환을 갖는다. 일부 구현예에서, RT 도메인은 레트로바이러스의 RT, 예를 들어 HIV-1 RT, 몰로니 뮤린 백혈병 바이러스(MMLV) RT, 조류 골수아구증 바이러스(AMV) RT, 또는 라우스 육종 바이러스(RSV) RT로부터 유래된다.

[0175] 일부 구현예에서, 레트로바이러스 역전사효소(RT) 도메인은, 예를 들어 내인성 RT 도메인에 비해 표적-프라이밍된 역전사(TPRT) 개시의 향상된 엄격성을 나타낸다. 일부 구현예에서, RT 도메인은, 제1 가닥 닉의 바로 상류의 표적 부위에서 3 nt, 예를 들어 RNA 주형을 프라이밍하는 게놈 DNA가 RNA 주형에서 3 nt의 상동성에 대해 적어도 66% 또는 100% 상보성을 갖는 경우, TPRT를 개시한다. 일부 구현예에서, RT 도메인은, 주형 RNA 상동성과 표적 DNA 프라이밍 역전사 사이에 5 nt 미만의 미스매치(예를 들어, 1, 2, 3, 4, 또는 5 nt 미만의 미스매치)가 존재하는 경우 TPRT를 개시한다. 일부 구현예에서, RT 도메인은 TPRT 반응을 프라이밍하는 데 있어 미스매치에 대한 엄격성이 증가하도록 변형되며, 예를 들어 여기서 RT 도메인은 야생형(예를 들어, 비변형된) RT 도메인에 비해 프라이밍 영역에서 임의의 미스매치를 용인하지 않거나, 더 적은 미스매치를 용인한다. 일부 구현예에서, RT 도메인은 HIV-1 RT 도메인을 포함한다. 구현예에서, HIV-1 RT 도메인은 대안적인 RT 도메인에 비해 3개의 뉴클레오티드 미스매치가 있는 경우에도 더 낮은 수준의 합성을 개시한다(예를 들어, 문헌[Jamburuthugoda and Eickbush J Mol Biol 407(5):661-672 (2011)]에 기재된 바와 같으며; 그 전체가 본 명세서에 참고로 포함됨). 일부 구현예에서, RT 도메인은 이량체(예를 들어, 이종이량체 또는 동종이량체)를 형성한다. 일부 구현예에서, RT 도메인은 단량체성이다. 일부 구현예에서, RT 도메인은 단량체 또는 이량체(예를 들어, 이종이량체 또는 동종이량체)로서 자연적으로 기능한다. 일부 구현예에서, RT 도메인은 단량체로서 자연적으로 기능하며, 예를 들어 바이러스로부터 유래되고, 여기서 이는 단량체로서 기능한다. 구현예에서, RT 도메인은 뮤린 백혈병 바이러스(MLV; 때때로 MoMLV로 지칭됨)(예를 들어, P03355), 돼지 내인성 레트로바이러스(PERV)(예를 들어, UniProt Q4VFZ2), 마우스 유방 종양 바이러스(MMTV)(예를 들어, UniProt P03365), Mason-Pfizer 원숭이 바이러스

(MPMV)(예를 들어, UniProt P07572), 소 백혈병 바이러스(BLV)(예를 들어, UniProt P03361), 인간 T-세포 백혈병 바이러스-1(HTLV-1)(예를 들어, UniProt P03362), 인간 포미(foamy) 바이러스(HFV)(예를 들어, UniProt P14350), 원숭이 포미 바이러스(SFV)(예를 들어, UniProt P23074), 또는 소 포미/세포융합 바이러스(BFV/BSV)(예를 들어, UniProt O41894), 또는 이의 기능성 단편 또는 변이체(예를 들어, 이와 적어도 70%, 80%, 90%, 95%, 또는 99% 동일성을 갖는 아미노산 서열)로부터의 RT 도메인으로부터 선택된다. 일부 구현예에서, RT 도메인은 그의 천연 기능에서 이량체성이다. 일부 구현예에서, RT 도메인은 바이러스로부터 유래되며, 여기서 이는 이량체로서 기능한다. 구현예에서, RT 도메인은 조류 육종/백혈병, 바이러스(ASLV)(예를 들어, UniProt A0A142BKH1), 라우스 육종 바이러스(RSV)(예를 들어, UniProt P03354), 조류 골수아구증 바이러스(AMV)(예를 들어, UniProt Q83133), 인간 면역결핍 바이러스 I형(HIV-1)(예를 들어, UniProt P03369), 인간 면역결핍 바이러스 II형(HIV-2)(예를 들어, UniProt P15833), 원숭이 면역결핍 바이러스(SIV)(예를 들어, UniProt P05896), 소 면역결핍 바이러스(BIV)(예를 들어, UniProt P19560), 말 전염성 빈혈 바이러스(EIAV)(예를 들어, UniProt P03371), 또는 고양이 면역결핍 바이러스(FIV)(예를 들어, UniProt P16088)(Herschhorn and Hizi *Cell Mol Life Sci* 67(16):2717-2747 (2010)), 또는 이의 기능성 단편 또는 변이체(예를 들어, 이와 적어도 70%, 80%, 90%, 95%, 또는 99% 동일성을 갖는 아미노산 서열)로부터의 RT 도메인으로부터 선택된다. 천연 이종이량체성 RT 도메인은 또한 일부 구현예에서 동종이량체로서 기능할 수 있다. 일부 구현예에서, 이량체성 RT 도메인은 융합 단백질, 예를 들어 동종이량체성 융합 단백질 또는 이종이량체성 융합 단백질로서 발현된다. 일부 구현예에서, 시스템의 RT 기능은 (예를 들어, 본 명세서에 기재된 바와 같은) 다수의 RT 도메인에 의해 수행된다. 추가의 구현예들에서, 다수의 RT 도메인은 융합 또는 분리되어, 예를 들어 동일한 폴리펩티드 상에 또는 상이한 폴리펩티드 상에 있을 수 있다.

[0176] 일부 구현예에서, 본 명세서에 기재된 유전자 변형 시스템은 인테그라제(integrase) 도메인을 포함하고, 예를 들어 여기서 인테그라제 도메인은 RT 도메인의 일부일 수 있다. 일부 구현예에서, (예를 들어, 본 명세서에 기재된 바와 같은) RT 도메인은 인테그라제 도메인을 포함한다. 일부 구현예에서, (예를 들어, 본 명세서에 기재된 바와 같은) RT 도메인은 인테그라제 도메인을 결여하거나, 또는 돌연변이에 의해 비활성화되거나 결실된 인테그라제 도메인을 포함한다. 일부 구현예에서, 본 명세서에 기재된 유전자 변형 시스템은 RNase H 도메인을 포함하고, 예를 들어 여기서 RNase H 도메인은 RT 도메인의 일부일 수 있다. 일부 구현예에서, RNase H 도메인은 RT 도메인의 일부가 아니고 가요성 링커를 통해 공유결합적으로 연결된다. 일부 구현예에서, (예를 들어, 본 명세서에 기재된 바와 같은) RT 도메인은 RNase H 도메인, 예를 들어 내인성 RNase H 도메인 또는 이종 RNase H 도메인을 포함한다. 일부 구현예에서, (예를 들어, 본 명세서에 기재된 바와 같은) RT 도메인은 RNase H 도메인이 결여되어 있다. 일부 구현예에서, (예를 들어, 본 명세서에 기재된 바와 같은) RT 도메인은 이종 RNase H 도메인에 대해 추가, 결실, 돌연변이 또는 교환된 RNase H 도메인을 포함한다. 일부 구현예에서, 폴리펩티드는 비활성화된 내인성 RNase H 도메인을 포함한다. 일부 구현예에서, 폴리펩티드의 다른 도메인 중 하나로부터의 내인성 RNase H 도메인은 폴리펩티드에 포함되지 않도록 유전적으로 제거되며, 예를 들어 내인성 RNase H 도메인은 포함하는 도메인으로부터 부분적으로 또는 완전히 절단된다. 일부 구현예에서, RNase H 도메인의 돌연변이는 일부 구현예에서, RNase H 도메인의 돌연변이는, 문헌[Kotewicz et al. *Nucleic Acids Res* 16(1):265-277 (1988)(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 방법에 의해 결정되는 바와 같은, 예를 들어 돌연변이가 없는 다른 유사한 도메인과 비교하여 적어도 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 또는 90%만큼 더 낮은 RNase 활성을 나타내는 폴리펩티드를 산출한다. 일부 구현예에서, RNase H 활성은 파괴된다.

[0177] 일부 구현예에서, RT 도메인은, 돌연변이가 없는 다른 유사한 도메인에 비해 정확도(fidelity)가 증가하도록 돌연변이된다. 예를 들어, 일부 구현예에서, RT 도메인(예를 들어, 역전사효소에서)에서 YADD 또는 YMDD 모티프는 YVDD로 대체된다. 구현예에서, YADD 또는 YMDD 또는 YVDD의 대체는 레트로바이러스 역전사효소 활성에서 더 높은 정확도를 생성한다(예를 들어, 문헌[Jamburuthugoda and Eickbush J Mol Biol 2011]에 기재된 바와 같으며; 이는 그 전체가 본 명세서에 참고로 포함됨).

[0178] 일부 구현예에서, 본 명세서에 기재된 유전자 변형 폴리펩티드는 표 2에 따른 아미노산 서열, 또는 이에 대하여 적어도 70%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 RT 도메인을 포함한다. 일부 구현예에서, 본 명세서에 기재된 핵산은 표 2에 따른 아미노산 서열, 또는 이에 대하여 적어도 70%, 80%, 85%, 90%, 95%, 97%, 98%, 또는 99% 동일성을 갖는 서열을 갖는 RT 도메인을 인코딩한다.

[0179] **표 2: 레트로바이러스 유래의 예시적인 역전사효소 도메인**

표 2

[0180]

RT 명칭	RT 아미노산 서열
AVIRE_P03360	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPW NTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRL PQGFKNSPTLFDEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSN SRTQAILQIPVPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRGGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSG AAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPR VRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAE LIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGLLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVA REVAIRPLSTQATIS
AVIRE_P03360_3mut	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPW NTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRL PQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSN SRTQAILQIPVPKTKRQVREFLGTIGYCRLWIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSG AAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPR VRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAE LIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVA REVAIRPLSTQATIS
AVIRE_P03360_3mutA	TAPLEEEYRLFLEAPIQNVTLLEQWKREIPKVWAEINPPGLASTQAPIHVQLLSTALPVRVRQYPITLEAKRSLRETIRKFRAAGILRPVHSPW NTPLLPVRKSGTSEYRMVQDLREVNKRVETIHPTVPNPYTLLSLLPPDRIWYSVLDLKDAFFCIPLAPESQLIFAFEWADAEEGESGQLTWTRL PQGFKNSPTLFNEALNRDLQGFRLDHPSVSLLQYVDDLLIAADTQAACLSATRDLLMTLAELGYRVSGKKAQLCQEEVTYLGFKIHKGSRSLSN SRTQAILQIPVPKTKRQVREFLGKIGYCRLFIPGFAELAQPLYAATRPGNDPLVWGEKEEEAFQSLKLALTQPPALALPSLDKPFQLFVEETSG AAKGVLTQALGPWKRPVAYLSKRLDPVAAGWPRCLRAIAAAALLTREASKLTFGQDIEITSSHNLESLLRSPPDKWLTNARITQYQVLLLDPPR VRFKQTAALNPATLLPETDDTLPIHHCLDTLDSLTSTRPDLTDQPLAQAEATLFTDGSSYIRDGKRYAGAAVVTLDSVIWAEPLPIGTSAQKAE LIALTKALEWSKDKSVNIYTDSRYAFATLHVHGMIYRERGWLTAGGKAIKNAPEILALLTAVWLPKRVAVMHCKGHQKDDAPTSTGNRRADEVA REVAIRPLSTQATIS
BAEVM_P10272	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNT PLLPVKKPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQ GFKNSPTLFDEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGR IETVARIPPPRNPREVREFLGTAGFCRLWIPGFAELAAPLYALTKESTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAK GVLTQKLGPWKRPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQF GPPVTLNPATLLPVPENQPSPHDCRQVLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIA LTKALELSKGKKANIYTDSRYAFATAHTHGSIYERRGLLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQA AMAEVLTLATEPDNTSHIT
BAEVM_P10272_3mut	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNT PLLPVKKPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQ GFKNSPTLFNEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGR IETVARIPPPRNPREVREFLGTAGFCRLWIPGFAELAAPLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAK GVLTQKLGPWKRPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQF GPPVTLNPATLLPVPENQPSPHDCRQVLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIA LTKALELSKGKKANIYTDSRYAFATAHTHGSIYERRGWLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQA AMAEVLTLATEPDNTSHIT
BAEVM_P10272_3mutA	TVSLQDEHRLFDIPVTTSLPDVWLQDFPQAWAETGGLGRAKCQAPIIIDLKPTAVPVSIKQYPMSLEAHMGIRQHIIKFLELGVLRPCRSPWNT PLLPVKKPGTQDYRPVQDLREINKRTVDIHPTVPNPYNLLSTLKPDYSWYTVLDLKDAFFCLPLAPQSQELFAFEWKDPERGISGQLTWTRLPQ GFKNSPTLFNEALHRDLTDFRTQHPEVTLLQYVDDLLLAAPTKKACTQGTRHLLQELGEKGYRASAKKAQICQTKVTYLGYILSEGKRWLTPGR IETVARIPPPRNPREVREFLGKAGFCRLFIPGFAELAAPLYALTKPSTPFTWQTEHQLAFEALKKALLSAPALGLPDTSKPFTLFLDERQGIAK GVLTQKLGPWKRPVAYLSKKLDPVAAGWPPCLRIMAATAMLVKDSAKLTLGQPLTVITPHTLEAIVRQPPDRWITNARLTHYQALLLDTDRVQF GPPVTLNPATLLPVPENQPSPHDCRQVLAETHGTREDLKDQELPDADHTWYTDGSSYLDSGTRRAGAAVVDGHNTIWAQSLPPGTSAQKAELIA LTKALELSKGKKANIYTDSRYAFATAHTHGSIYERRGWLTSEGKEIKNKAEIIALLKALFLPQEVAIIHCPGHQKGQDPVAVGNRQADRVARQA AMAEVLTLATEPDNTSHIT
BLVAU_P25059	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRVTNALTKPIPALSPGPPDL TAIPTHLPHIICLDLKDAFFQIPVEDRFRSYFAFTLPTPGGLQPHRRFAWRVLPQGFINSPALFERALQEPLRQVSAAFSQSLLVSYMDDILYV SPTEEQRLQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQLQTVLGDLQWVSRGTPTTRRPLQLLYS SLKGIDDPRAIIHLSPEQQQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQ AQALSSYAKTILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQFSPEPIPAALCLFSDGAARRGAYC LWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPLNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIFVGHVRSHSSASH PIASLNNYVDQL
BLVAU_P25059_2mut	GVLDAPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRVTNALTKPIPALSPGPPDL TAIPTHLPHIICLDLKDAFFQIPVEDRFRSYFAFTLPTPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYV SPTEEQRLQCYQTMAAHLRDLGFQVASEKTRQTPSPVPFLGQMVHERMVTYQSLPTLQISSPISLHQLQTVLGDLQWVSRGTPTTRRPLQLLYS SLKPIDDPRAIIHLSPEQQQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQ AQALSSYAKTILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLVTRAEVFLTPQFSPEPIPAALCLFSDGAARRGAYC LWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPLNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIFVGHVRSHSSASH PIASLNNYVDQL

BLVJ_P03361	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDL TAIPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFERALQEPLRQVSAAFSQSLLVSYMDDILYA SPTEEQRSQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYS SLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQ TQALSSYAKPILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYC LWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPVNIWVDSKYLYSLLRTLVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASH PIASLNNYVDQL
BLVJ_P03361_2mut	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDL TAIPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFNRALQEPLRQVSAAFSQSLLVSYMDDILYA SPTEEQRSQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYS SLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQ TQALSSYAKPILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYC LWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPVNIWVDSKYLYSLLRTWVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASH PIASLNNYVDQL
BLVJ_P03361_2mutB	GVLDTPPSHIGLEHLPPPPEVPQFPLNLERLQALQDLVHRSLEAGYISPWDGPGNNPVFPVRKPNGAWRFVHDLRATNALTKPIPALSPGPPDL TAPPTHPPHIICLDLKDAFFQIPVEDRFRFYLSFTLPSPGGLQPHRRFAWRVLPQGFINSPALFQRALQEPLRQVSAAFSQSLLVSYMDDILYA SPTEEQRSQCYQALAARLRDLGFQVASEKTSQTPSPVPFLGQMVHEQIVTYQSLPTLQISSPISLHQLQAVLGDLQWVSRGTPTTRRPLQLLYS SLKRHHDPRAIIQLSPEQLQGIAELRQALSHNARSRYNEQEPLLAYVHLTRAGSTLVLFQKGAQFPLAYFQTPLTDNQASPWGLLLLLGCQYLQ TQALSSYAKPILKYYHNLPKTSLDNWIQSSEDPRVQELLQLWPQISSQGIQPPGPWKTLITRAEVFLTPQFSPDPIPAALCLFSDGATGRGAYC LWKDHLLDFQAVPAPESAQKGELAGLLAGLAAAPPEPVNIWVDSKYLYSLLRTWVLGAWLQPDPVPSYALLYKSLLRHPAIVVGHVRSHSSASH PIASLNNYVDQL
FFV_093209	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKP LELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVL IQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLP QGFLNSPGLFTGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLE NITAPTTLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSTKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRY YNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDP QMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLG GNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_093209_2mut	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKP LELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVL IQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLP QGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLE NITAPTTLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRY YNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDP QMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLG GNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_093209_2mutA	MDLLKPLTVERKGVKIKGYWNSQADITCVPKDLLQGEEPVRQQNVTTIHGTQEGDVYYVNLKIDGRRINTEVIGTTLDYAIITPGDVPWILKKP LELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVINDLLKQGVL IQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGKQYCWTVLP QGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLTDTFKEKLE NITAPTTLKQLQSILGKLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNASYTTGYIRY YNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDPRIRFFYDP QMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFALKKCLPLG GNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQASFKVH
FFV_093209-Pro	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVIN DLLKQGVLIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGK QYCWTVLPQGFLNSPGLFTGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLT DTFKEKLENITAPTTLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSTKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNAS YTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDP RIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFA LKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQAS FKVH
FFV_093209-Pro_2mut	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVIN DLLKQGVLIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGK QYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLT DTFKEKLENITAPTTLKQLQSILGLLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNAS YTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDP RIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFA LKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQAS FKVH

FFV_093209-Pro_2mutA	VPWILKKPLELTIKLDLEEQQGTLLNNSILSKKGKEELKQLFEKYSALWQSWENQVGHRRIRPHKIATGTVKPTPQKQYHINPKAKPDIQIVIN DLLKQGVLIQKESTMNTPVYPVPKPNGRWRMVLDYRAVNKVTPLIAVQNQHSYGILGSLFKGRYKTTIDLSNGFWAHPIVPEDYWITAFTWQGK QYCWTVLPQGFLNSPGLFNGDVVDLLQGIPNVEVYVDDVYISHDSEKEHLEYLDILFNRLKEAGYIISLKKSNIANSIVDFLGFQITNEGRGLT DTFKEKLENITAPTTLKQLQSILGKLNFARNFIPDFTELIAPLYALIPKSPKNYVPWQIEHSTTLETLITKLNGAEYLQGRKGDKTLIMKVNAS YTTGYIRYYNEGEKKPISYVSIVFSKTELKFTELEKLLTTVHKGLLKALDLSMGQNIHVYSPIVSMQNIQKTPQTAKKALASRWLSWLSYLEDP RIRFFYDPQMPALKDLPAVDTGKDNKKHPSNFQHIFYTDGSAITSPTKEGHLNAGMGIVYFINKDGNLQKQQEWSISLGNHTAQFAEIAAFEFA LKKCLPLGGNILVVTDSNYVAKAYNEELDVWASNGFVNNRKKPLKHISKWKSVADLKRLRPDVVVTHEPGHQKLDSSPHAYGNNLADQLATQAS FKVH
FLV_P10273	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWN TPLLPVKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLP QGFKNSPTLFDEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKA RKEAILSIPVPKNSRQVREFLGTAGYCRLWIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFA KGVLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVH FGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIA LTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRRGLLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKA ATETHSSLTVLP
FLV_P10273_3mut	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWN TPLLPVKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLP QGFKNSPTLFNEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKA RKEAILSIPVPKNSRQVREFLGTAGYCRLWIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFA KGVLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVH FGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIA LTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRRGWLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKA ATETHSSLTVLP
FLV_P10273_3mutA	TLQLEEEYRLFEPESTQKQEMDIWLKNFPQAWAETGGMGTAHCQAPVLIQLKATATPISIRQYPMPHEAYQGIKPHIRRMLDQGILKPCQSPWN TPLLPVKKPGTEDYRPVQDLREVNKRVEDIHPTVPNPYNLLSTLPPSHPWYTVLDLKDAFFCLRLHSESQLLFAFEWRDPEIGLSGQLTWTRLP QGFKNSPTLFNEALHSDLADFRVRYPALVLLQYVDDLLLAAATRTECLEGTKALLETLGNKGYRASAKKAQICLQEVTYLGYSLKDGQRWLTKA RKEAILSIPVPKNSRQVREFLGKAGYCRLFIPGFAELAAPLYPLTRPGTLFQWGTEQQLAFEDIKKALLSSPALGLPDITKPFELFIDENSGFA KGVLVQKLGPWKRPVAYLSKKLDTVASGWPPCLRMVAAIAILVKDAGKLTLGQPLTILTSHPVEALVRQPPNKWLSNARMTHYQAMLLDAERVH FGPTVSLNPATLLPLPSGGNHHDCLQILAETHGTRPDLTDQPLPDADLTWYTDGSSFIRNGEREAGAAVTTESEVIWAAPLPPGTSAQRAELIA LTQALKMAEGKKLTVYTDSRYAFATTHVHGEIYRRRGWLTSEGKEIKNKNEILALLEALFLPKRLSIIHCPGHQKGDSPQAKGNRLADDTAKKA ATETHSSLTVLP
FOAMV_P14350	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLT QQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWT RLPQGFLNSPALFTADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKT KLLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIASAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGY VRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKAL KIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350_2mut	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLT QQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWT RLPQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKT KLLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGY VRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKAL KIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350_2mutA	MNPLQLLQPLPAEIKGTKLLAHWNSGATITCIPESFLEDEQPIKKTLIKTIHGEKQQNVYYVTFKVKGRKVEAEVIASPYEYILLSPTDVPWLT QQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGKQYCWT RLPQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLTDTFKT KLLNITPPKDLKQLQSILGKLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTSPSAGY VRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFACKKAL KIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGSYVVN
FOAMV_P14350-Pro	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVID DLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGK QYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLT DTFKTKLLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIASAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTS PSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDP RIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFA CKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGS YVVN

FOAMV_P14350-Pro_2mut	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVID DLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGK QYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLT DTFKTKLLNITPPKDLKQLQSILGLLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTS PSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDP RIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFA CKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGS YVVN
FOAMV_P14350-Pro_2mutA	VPWLTQQPLQLTILVPLQEYQEKILSKTALPEDQKQQLKTLFVKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVID DLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPESYWLTAFTWQGK QYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQVYVDDIYLSHDDPKEHVQQLEKVFQILLQAGYVVSLKKSEIGQKTVEFLGFNITKEGRGLT DTFKTKLLNITPPKDLKQLQSILGKLNFARNFIPNFAELVQPLYNLIAPAKGKYIEWSEENTKQLNMVIEALNTASNLEERLPEQRLVIKVNTS PSAGYVRYYNETGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDP RIQFHYDKTLPELKHIPDVYTSSQSPVKHPSQYEGVFYTDGSAIKSPDPTKSNNAGMGIVHATYKPEYQVLNQWSIPLGNHTAQMAEIAAVEFA CKKALKIPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKKPLKHISKWKSIAECLSMKPDITIQHEKGISLQIPVFILKGNALADKLATQGS YVVN
GALV_P21414	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNT PLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQ GFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPAR KATVMKIPVPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVAR GVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFA PPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALT QALRLAEGKNINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAAL STRVLAGTTKP
GALV_P21414_3mut	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNT PLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQ GFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPAR KATVMKIPVPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVAR GVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFA PPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALT QALRLAEGKNINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAAL STRVLAGTTKP
GALV_P21414_3mutA	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQKFLDLGVLVPCRSPWNT PLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSYTWYSVLDLKDAFFCLRLHPNSQPLFAFEWKDPEKGNTGQLTWTRLPQ GFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYEDCKKGTQKLLQELSKLGYRVSAKKAQLCQREVTYLGYLLKEGKRWLTPAR KATVMKIPVPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTKPSIPFIWTEEHQQAFDHIKKALLSAPALALPDLTKPFTLYIDERAGVAR GVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFA PPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLEDQPLPGVPTWYTDGSSFITEGKRRAGAPIVDGKRTVWASSLPEGTSAQKAELVALT QALRLAEGKNINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPRRVAIIHCPGHQRGSNPVATGNRRADEAAKQAAL STRVLAGTTKP
HTL1A_P03362	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTL AHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDL LLLSEATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQRHTD PRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSY GLLCQTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYI LWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHT NLPDPISRLNALTDALLITPVLQL
HTL1A_P03362_2mut	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTL AHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDL LLLSEATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTD PRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSY GLLCQTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYI LWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHT NLPDPISRLNALTDALLITPVLQL
HTL1A_P03362_2mutB	AVLGLEHLPRPPQISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSPPTTL AHLQTIDLRDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHEDL LLLSEATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTD PRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKEQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSY GLLCQTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSRAAYI LWDKQILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHT NLPDPISRLNALTDALLITPVLQL
HTL1C_P14078	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTL AHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWRVLPQGFKNSPTLFEMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADL QLLSEATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPKVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQRHTD PRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSY GLLCQTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSQAAYI LWDKHILSQRSFPLPPPHKSAQRAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHT NLPDPISRLNALTDALLITPVLQL

HTL1C_P14078_2mut	AVLGLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTIDLSSSSPGPPDLSSLPTTL AHLQTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWRVLPQGFKNSPTLFQMQLAHILQPIRQAFPQCTILQYMDDILLASPSHADL QLLSEATMASLISHGLPVSENKTQQTPGTIKFLGQIISPNHLTYDAVPKVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQPHTD PRDQIYLNPSQVQSLVQLRQALSQNCRSRLVQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSY GLLCQTIHHNISTQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTTAPLAPVKALMPVFTLSPVIINTAPCLFSDGSTSQAAYI LWDKHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWRCLNIFLDSKYLYHYLRTLALGTFQGRSSQAPFQALLPRLLSRKVVYLHHVRSHT NLPDPISRLNALTDALLITPVLQL
HTL1L_P0C211	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSLPTTLAHL QTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFEMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQL SEATMASLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRD QIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLL CQTIHHNISIQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWD KHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCLNIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLP DPISKLNALTDALLITPIL
HTL1L_P0C211_2mut	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSLPTTLAHL QTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQL SEATMASLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRD QIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLL CQTIHHNISIQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWD KHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCLNIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLP DPISKLNALTDALLITPIL
HTL1L_P0C211_2mutB	GLEHLPRPPEISQFPLNPERLQALQHLVRKALEAGHIEPYTGPGNNPVFPVKKANGTWRFIHDLRATNSLTVDLSSSSPGPPDLSSPPTTLAHL QTIDLKDAFFQIPLPKQFQPYFAFTVPQQCNYGPGTRYAWKVLPQGFKNSPTLFQMQLASILQPIRQAFPQCVILQYMDDILLASPSPEDLQQL SEATMASLISHGLPVSQDKTQQTPGTIKFLGQIISPNHITYDAVPTVPIRSRWALPELQALLGEIQWVSKGTPTLRQPLHSLYCALQGHTDPRD QIYLNPSQVQSLMQLQQALSQNCRSRLAQTLPLLGAIMLTLTGTTTVVFQSKQQWPLVWLHAPLPHTSQCPWGQLLASAVLLLDKYTLQSYGLL CQTIHHNISIQTFNQFIQTSDHPSVPILLHHSHRFKNLGAQTGELWNTFLKTAAPLAPVKALTPVFTLSPIIINTAPCLFSDGSTSQAAYILWD KHILSQRSFPLPPPHKSAQQAELLGLLHGLSSARSWHCLNIFLDSKYLYHYLRTLAWGTFQGKSSQAPFQALLPRLLAHKVIYLHHVRSHTNLP DPISKLNALTDALLITPIL
HTL32_Q0R5R2	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSLPQGLPHL RTIDLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFEQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAAL TDKVTNALTKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRD TIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQV CKSFHHNISNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWD RQVIHQQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFLDSKFLIGHLRRMALGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQD PISRLNEATDALMLAPLLPL
HTL32_Q0R5R2_2mut	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSLPQGLPHL RTIDLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFQQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAAL TDKVTNALTKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRD TIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQV CKSFHHNISNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWD RQVIHQQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFLDSKFLIGHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQD PISRLNEATDALMLAPLLPL
HTL32_Q0R5R2_2mutB	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSVTRDLASPSPGPPDLTSPPQGLPHL RTIDLTDAFFQIPLPTIFQPYFAFTLPQPNNYGPGTRYSWRVLPQGFKNSPTLFQQQLSHILTPVRKTFPNSLIIQYMDDILLASPAPGELAAL TDKVTNALTKEGLPLSPEKTQATPGPIHFLGQVISQDCITYETLPSINVKSTWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRD TIKLTSIQVQALRTIQKALTLNCRSRLVNQLPILALIMLRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAVIILDKYSLQHYGQV CKSFHHNISNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQMPQIFQNIDVLRPPFTISPVVINHAPCLFSDGSASKAAFIIWD RQVIHQQVLSLPSTCSAQAGELFGLLAGLQKSQPWVALNIFLDSKFLIGHLRRMAWGAFPGPSTQCELHTQLLPLLQGKTVYVHHVRSHTLLQD PISRLNEATDALMLAPLLPL
HTL3P_Q4U0X6	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSLPQDLPHL RTIDLTDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFEQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTAL TDKVTNALTKEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRD TIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQI CKSFHHNISNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWD KQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLDSKFLIGHLRRMALGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQD PISRLNEATDALMLAPLLPL
HTL3P_Q4U0X6_2mut	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSLPQDLPHL RTIDLTDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTAL TDKVTNALTKEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRD TIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQI CKSFHHNISNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWD KQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLDSKFLIGHLRRMALGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQD PISRLNEATDALMLAPLLPL

HTL3P_Q4U0X6_2mutB	GLEHLPPPPEVSQFPLNPERLQALTDLVSRALEAKHIEPYQGPGNNPIFPVKKPNGKWRFIHDLRATNSLTRDLASPSPGPPDLTSPPQDLPHLRTIDLTDAFFQIPLPAVFQPYFAFTLPQPNNHGPGTRYSWRVLPQGFKNSPTLFQQQLSHILAPVRKAFPNSLIIQYMDDILLASPALRELTALTDKVTNALTKEGLPMSLEKTQATPGSIHFLGQVISPDCITYETLPSIHVKSIWSLAELQSMLGELQWVSKGTPVLRSSLHQLYLALRGHRDPRDTIELTSTQVQALKTIQKALALNCRSRLVSQLPILALIILRPTGTTAVLFQTKQKWPLVWLHTPHPATSLRPWGQLLANAIITLDKYSLQHYGQICKSFHHNISNQALTYYLHTSDQSSVAILLQHSHRFHNLGAQPSGPWRSLLQVPQIFQNIDVLRPPFIISPVVIDHAPCLFSDGATSKAAFILWDKQVIHQQVLPLPSTCSAQAGELFGLLAGLQKSKPWPALNIFLDSKFLIGHLRRMAWGAFLGPSTQCDLHARLFPLLQGKTVYVHHVRSHTLLQDPISRLNEATDALMLAPLLPL	
HTLV2_P03363_2mut	HLPPPPQVDQFPLNLPERLQALNDLVSKALEAGHIEPYSGPGNNPVFPVKKPNGKWRFIHDLRATNAITTTLTSPSPGPPDLTSLPTALPHLQTIDLTDAFFQIPLPKQYQPYFAFTIPQPCNYGPGTRYAWTVLPQGFKNSPTLFQQQLAAVLNPMRKMFPTSTIVQYMDDILLASPTNEELQQLSQLTLQALTTHGLPISQEKTQQTPGQIRFLGQVISPNHITYESTPTIPIKSQWTLTELQVILGEIQWVSKGTPILRKHLQSLYSALHPYRDPRACITLTPQQLHALHAIQQALQHNCRGRLNPALPLLGLISLSTSGTTSVIFQPKQNWPLAWLHTPHPPTSLCPWGHLLACTILTLDKYTLQHYGQLCQSFHHNMSKQALCDFLRNSPHPSVGILIHHMGRFHNLGSQPSGPWKTLLHLPTLLQEPRLLRPIFTLSPVVLDTAPCLFSDGSPQKAAYVLWDQTILQQDITPLPSHETHSAQKGELLALICGLRAAKPWPSLNIFLDSKYLIKYLHSLAIGAFLGTSAHQTLQAALPPLLQGKTIYLHHVRSHTNLPDPISTFNEYTDSLILAPLVPL	
JSRV_P31623	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQLRLGHIEPSTSAWNSPIFVIKKKSGKWRLLQDLRKVNETMMHMGALQPGLPTPSAIPDKSYIIVIDLKDCFYTIPLAPQDCKRFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFPQLYLVHYMDDILLAHTDEHLLYQAFSILKQHLSLNGLVIADEKIQTHFPYNYLGFSLYPRVYNTQLVKLQTDHLKTLNDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASPRTLSLEGRTALQSIEEAIRQQQITYCDYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYYELVAKIIAKGRHEAIQYFGMEPPFICVPYALEQQDWLFQFSDNWSIAFANYPGQITHHYPSDKLLQFASSHAFIFPKIVRRQPIPEATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQALLTVPTSFNLFTDSSYVVGALQMIETVPIIGTTSPEVLNLFTLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVLTKQVFFQS	
JSRV_P31623_2mutB	PLGTSDSPVTHADPIDWKSEEPVWVDQWPLTQEKLSAAQQLVQEQLRLGHIEPSTSAWNSPIFVIKKKSGKWRLLQDLRKVNETMMHMGALQPGLPTPSPIPDKSYIIVIDLKDCFYTIPLAPQDCKRFAFSLPSVNFKEPMQRYQWRVLPQGMTNSPTLCQKFVATAIAPVRQRFPQLYLVHYMDDILLAHTDEHLLYQAFSILKQHLSLNGLVIADEKIQTHFPYNYLGFSLYPRVYNTQLVKLQTDHLKTLNDFQKLLGDINWIRPYLKLPTYTLQPLFDILKGDSDPASPRTLSLEGRTALQSIEEAIRQQQITYCDYQRSWGLYILPTPRAPTGVLYQDKPLRWIYLSATPTKHLLPYYELVAKIIAKGRHEAIQYFGMEPPFICVPYALEQQDWLFQFSDNWSIAFANYPGQITHHYPSDKLLQFASSHAFIFPKIVRRQPIPEATLIFTDGSSNGTAALIINHQTYYAQTSFSSAQVVELFAVHQALLTVPTSFNLFTDSSYVVGALQMIETVPIIGTTSPEVLNLFTLIQQVLHCRQHPCFFGHIRAHSTLPGALVQGNHTADVLTKQVFFQS	
KORV_Q9TTC1	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN	
KORV_Q9TTC1_3mut	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN	
KORV_Q9TTC1_3mutA	TLGDQGSRGSDPLPEPRVTLTVEGIPTEFLVNTGAEHSVLTKPMGKMGSKRTVVAGATGSKVYPWTTKRLLKIGQKQVTHSFLVIPECPAPLLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN	
KORV_Q9TTC1-Pro	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQYPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFDEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLGYRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTREKVPFTWTEAHQEAFGRIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHNLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRRQAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN	

KORV_Q9TTC1-Pro_3mut	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQ YPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFC LKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLG YRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFG RIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHN LESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRR QAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPK RVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN
KORV_Q9TTC1-Pro_3mutA	LLGRDLLTKLKAQIQFSTEGPQVTWEDRPAMCLVLNLEEEYRLHEKPVPPSIDPSWLQLFPMVWAEKAGMGLANQVPPVVVELKSDASPVAVRQ YPMSKEAREGIRPHIQRFLDLGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFC LKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLASFRALNPQVVMLQYVDDLLVAAPTYRDCKEGTRRLLQELSKLG YRVSAKKAQLCREEVTYLGYLLKGGKRWLTPARKATVMKIPTPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTRPKVPFTWTEAHQEAFG RIKEALLSAPALALPDLTKPFALYVDEKEGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAIAAVALLLKDADKLTLGQNVLVIAPHN LESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAILNPATLLPVESDDTPIHICSEILAEETGTRPDLRDQPLPGVPAWYTDGSSFIMDGRR QAGAAIVDNKRTVWASNLPEGTSAQKAELIALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPK RVAIIHCPGHQRGTDPVATGNRKADEAAKQAAQSTRILTETTKN
MLVAV_P03356	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRL PQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRV QFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAARE AAIKTPPDTSTLL
MLVAV_P03356_3mut	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRV QFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAARE AAIKTPPDTSTLL
MLVAV_P03356_3mutA	TLNLEDEYRLYETSAEPEVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHRWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLLTLGNLGYRASAKKAQLCQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLRKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRV QFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAARE AAIKTPPDTSTLL
MLVBM_Q7SVK7	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRL PQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTE ARKETVMGQPVPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRV QFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAARE AAIKTPPDTSTLL
MLVBM_Q7SVK7_3mut	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTE ARKETVMGQPVPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRV QFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAARE AAIKTPPDTSTLL

MLVBM_Q7SVK7_3mut	TLGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTE ARKETVMGQPVPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRV QFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAARE AAIKTPPDTSTLL
MLVBM_Q7SVK7_3mutA_WS	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWN TPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEA RKETVMGQPVPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYA KGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQ FGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIA LTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREA AIKTPPDTSTLLI
MLVBM_Q7SVK7_3mutA_WS	LGIEDEYRLHETSTEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIQQYPMSHEARLGIKPHIQRLLDQGILVPCQSPWN TPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGMGISGQLTWTRLP QGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDILLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLREGQRWLTEA RKETVMGQPVPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFSWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYA KGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQ FGPVVALNPATLLPLPEEGAPHDCLEILAETHGTRPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWAGALPAGTSAQRAELIA LTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAAREA AIKTPPDTSTLLI
MLVCB_P08361	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFDEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPIPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARE VATRETPETSTLL
MLVCB_P08361_3mut	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPIPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARE VATRETPETSTLL
MLVCB_P08361_3mutA	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLAGFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPIPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAFQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHDCLDILAEAHGTRSDLMDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARE VATRETPETSTLL
MLVF5_P26810	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRL PQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGLCRLWIPGFAEMAAPLYPLTKTGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPIVALNPATLLPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAARE VATRETPETSTLL
MLVF5_P26810_3mut	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGLCRLWIPGFAEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPIVALNPATLLPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAARE VATRETPETSTLL

MLVF5_P26810_3mutA	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAFRQAPLIISLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWKDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGLCRLFIPGFAEMAAPLYPLTKPGTLFKWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDVGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPIVALNPATLLPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRRAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAAGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNHAEARGNRMADQAARE VATRETPETSTLL
MLVFF_P26809_3mut	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPIVALNPATLLPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNRAEARGNRMADQAARE VATRETPETSTLL
MLVFF_P26809_3mutA	TLNIEDEYRLHETSKGPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQSLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGDLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFEWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPIVALNPATLLPLPEEGLQHDCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVVWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGNRAEARGNRMADQAARE VATRETPETSTLL
MLVMS_P03355	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARK AAITETPDTSTLL
MLVMS_P03355	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARK AAITETPDTSTLL
MLVMS_P03355_3mut	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARK AAITETPDTSTLL
MLVMS_P03355_3mutA_WS	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARK AAITETPDTSTLL

MLVMS_P03355_3mutA_WS	TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARK AAITETPDTSTLL
MLVMS_P03355_PLV919	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARK AAITETPDTSTLLIENSSPSGGSKRTADGSEFE
MLVMS_P03355_PLV919	TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRL PQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTE ARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRV QFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELI ALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARK AAITETPDTSTLLIENSSPSGGSKRTADGSEFE
MLVRD_P11227	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQGLREVNKRVEDIHPTVPNPYNLLSGLPTSHRWYTVLDLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRL PQGFKNSPTLFDEALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRASAKKAQICQKQVKYLGYLLREGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPRFAEMAAPLYPLTKTGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRV QFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYKRRGLLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAARE AAIKTPPDTSTLL
MLVRD_P11227_3mut	TLNIEDEYRLHEISTEPDVSPGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEAKLGIKPHIQRLLDQGILVPCQSPW NTPLLPVKKPGTNDYRPVQGLREVNKRVEDIHPTVPNPYNLLSGLPTSHRWYTVLDLKDAFFCLRLHPTSQPLFASEWRDPGMGISGQLTWTRL PQGFKNSPTLFNEALHRGLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLKTLGNLGYRASAKKAQICQKQVKYLGYLLREGQRWLTE ARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPRFAEMAAPLYPLTKPGTLFNWGPDQQKAYHEIKQALLTAPALGLPDLTKPFELFVDEKQGY AKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRV QFGPVVALNPATLLPLPEEGAPHDCLEILAETHGTEPDLTDQPIPDADHTWYTDGSSFLQEGQRKAGAAVTTETEVIWARALPAGTSAQRAELI ALTQALKMAEGKRLNVYTDSRYAFATAHIHGEIYKRRGWLTSEGREIKNKSEILALLKALFLPKRLSIIHCLGHQKGDSAEARGNRLADQAARE AAIKTPPDTSTLL
MMTVB_P03365	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGR DIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKA ILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLG NINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHIS PKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEK GIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQ EKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGR DIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKA ILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLG NINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHIS PKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEK GIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQ EKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365_2mut	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGR DIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKA ILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLG NINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHIS PKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEK GIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQ EKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT

MMTVB_P03365_2mut_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRD IMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAI LTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISP KVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKG IVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQE KFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTVB_P03365_2mut_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRD IMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAI LTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISP KVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKG IVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQE KFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTVB_P03365_2mutB	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGR DIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKA ILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLG NINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHIS PKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEK GIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQ EKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365_2mutB	WVQEISDSRPMLHIYLNGRRFLGLLNTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGR DIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRL LQDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKA ILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLG NINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHIS PKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEK GIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQ EKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365_2mutB_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRD IMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPPAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAI LTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISP KVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKG IVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQE KFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTVB_P03365_2mutB_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRD IMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPPAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAI LTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISP KVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKG IVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQE KFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTVB_P03365_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRD IMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAI LTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISP KVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKG IVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQE KFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA
MMTVB_P03365_WS	VQEISDSRPMLHIYLNGRRFLGLLDTGADKTCIAGRDWPANWPIHQTESSLQGLGMACGVARSSQPLRWQHEDKSGIIHPFVIPTLPFTLWGRD IMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKWRLL QDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVDKAI LTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKLLGN INWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPHISP KVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPLEKG IVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHKRQE KFYIGHIRGHTGLPGPLAQGNAYADSLTRILTA

MMTVB_P03365-Pro	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKW RLLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKL LGNINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPH ISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPL EKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHK RQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365-Pro	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKW RLLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKL LGNINWIRPFLKLTTGELKPLFEILNGDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPH ISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPL EKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHK RQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365-Pro_2mut	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKW RLLQDLRAVNATMHDMGALQPGLPSPVAVPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKL LGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPH ISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPL EKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHK RQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MMTVB_P03365-Pro_2mutB	GRDIMKDIKVRLMTDSPDDSQDLMIGAIESNLFADQISWKSDQPVWLNQWPLKQEKLQALQQLVTEQLQLGHLEESNSPWNTPVFVIKKKSGKW RLLQDLRAVNATMHDMGALQPGLPSPVAPPKGWEIIIIDLQDCFFNIKLHPEDCKRFAFSVPSPNFKRPYQRFQWKVLPQGMKNSPTLCQKFVD KAILTVRDKYQDSYIVHYMDDILLAHPSRSIVDEILTSMIQALNKHGLVVSTEKIQKYDNLKYLGTHIQGDSVSYQKLQIRTDKLRTLNDFQKL LGNINWIRPFLKLTTGELKPLFEILNPDSNPISTRKLTPEACKALQLMNERLSTARVKRLDLSQPWSLCILKTEYTPTACLWQDGVVEWIHLPH ISPKVITPYDIFCTQLIIKGRHRSKELFSKDPDYIVVPYTKVQFDLLLQEKEDWPISLLGFLGEVHFHLPKDPLLTFTLQTAIIFPHMTSTTPL EKGIVIFTDGSANGRSVTYIQGREPIIKENTQNTAQQAEIVAVITAFEEVSQPFNLYTDSKYVTGLFPEIETATLSPRTKIYTELKHLQRLIHK RQEKFYIGHIRGHTGLPGPLAQGNAYADSLTRILT
MPMV_P07572	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQ PGLPSPVAIPQGYLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMD DILIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKP LFDTLKGDSDPNSHRSLSKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILG RDHSKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFVFPQIISKTPLNNALLVFTDGSSTGMAAYTL TDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGP IAQGNQRADLATKIVASNINT
MPMV_P07572_2mutB	LTAAIDILAPQQCAEPITWKSDEPVWVDQWPLTNDKLAAAQQLVQEQLEAGHITESSSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQ PGLPSPVAPPQGYLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATAIHKVRHAWKQMYIIHYMD DILIAGKDGQQVLQCFDQLKQELTAAGLHIAPEKVQLQDPYTYLGFELNGPKITNQKAVIRKDKLQTLNDFQKLLGDINWLRPYLKLTTGDLKP LFDTLKPDSDPNSHRSLSKEALASLEKVETAIAEQFVTHINYSLPLIFLIFNTALTPTGLFWQDNPIMWIHLPASPKKVLLPYYDAIADLIILG RDHSKKYFGIEPSTIIQPYSKSQIDWLMQNTEMWPIACASFVGILDNHYPPNKLIQFCKLHTFVFPQIISKTPLNNALLVFTDGSSTGMAAYTL TDTTIKFQTNLNSAQLVELQALIAVLSAFPNQPLNIYTDSAYLAHSIPLLETVAQIKHISETAKLFLQCQQLIYNRSIPFYIGHVRAHSGLPGP IAQGNQRADLATKIVASNINT
PERV_Q4VFZ2	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWN TPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLP QGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEA RKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVA RGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTF APPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMA LTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGLLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQA AQGVNLL

PERV_Q4VFZ2	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFDEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKEKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_Q4VFZ2_3mut	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_Q4VFZ2_3mut	TLQLDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGTAGFCRLWIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLL
PERV_Q4VFZ2_3mutA_WS	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLP
PERV_Q4VFZ2_3mutA_WS	LDDEYRLYSPLVKPDQNIQFWLEQFPQAWAETAGMGLAKQVPPQVIQLKASATPVSVRQYPLSKEAQEGIRPHVQRLIQQGILVPVQSPWNTPLLPVRKPGTNDYRPVQDLREVNKRVQDIHPTVPNPYNLLCALPPQRSWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPGTGRTGQLTWTRLPQGFKNSPTIFNEALHRDLANFRIQHPQVTLLQYVDDLLLAGATKQDCLEGTKALLLELSDLGYRASAKKAQICRREVTYLGYSLRDGQRWLTEARKKTVVQIPAPTTAKQVREFLGKAGFCRLFIPGFATLAAPLYPLTKPKGEFSWAPEHQKAFDAIKKALLSAPALALPDVTKPFTLYVDERKGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPVCLKAIAAVAILVKDADKLTLGQNITVIAPHALENIVRQPPDRWMTNARMTHYQSLLLTERVTFAPPAALNPATLLPEETDEPVTHDCHQLLIEETGVRKDLTDIPLTGEVLTWFTDGSSYVVEGKRMAGAAVVDGTRTIWASSLPEGTSAQKAELMALTQALRLAEGKSINIYTDSRYAFATAHVHGAIYKQRGWLTSAGREIKNKEEILSLLEALHLPKRLAIIHCPGHQKAKDPISRGNQMADRVAKQAAQGVNLLP
SFV1_P23074	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVANANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074_2mut	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH
SFV1_P23074_2mutA	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPEAFLEDERPIQTMLIKTIHGEKQQDVYYLTFKVQGRKVEAEVLASPYDYILLNPSDVPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVIDDLLKQGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGKQYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLTDTFKQKLLNITPPKDLKQLQSILGKLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSSPSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDPRIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFACKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGSYVVH

SFV1_P23074-Pro	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVID DLLKQGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGK QYCWTRLPQGFLNSPALFTADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLT DTFKQKLLNITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVANANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSS PSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDP RIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFA CKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGS YVVH
SFV1_P23074-Pro_2mut	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVID DLLKQGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGK QYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLT DTFKQKLLNITPPKDLKQLQSILGLLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSS PSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDP RIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEYKIVHQWSIPLGDHTAQLAEIAAVEFA CKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGS YVVH
SFV1_P23074-Pro_2mutA	VPWLMKKPLQLTVLVPLHEYQERLLQQTALPKEQKELLQKLFLKYDALWQHWENQVGHRRIKPHNIATGTLAPRPQKQYPINPKAKPSIQIVID DLLKQGVLIQQNSTMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIYRGKYKTTLDLTNGFWAHPITPESYWLTAFTWQGK QYCWTRLPQGFLNSPALFNADVVDLLKEIPNVQAYVDDIYISHDDPQEHLEQLEKIFSILLNAGYVVSLKKSEIAQREVEFLGFNITKEGRGLT DTFKQKLLNITPPKDLKQLQSILGKLNFARNFIPNYSELVKPLYTIVAPANGKFISWTEDNSNQLQHIISVLNQADNLEERNPETRLIIKVNSS PSAGYIRYYNEGSKRPIMYVNYIFSKAEAKFTQTEKLLTTMHKGLIKAMDLAMGQEILVYSPIVSMTKIQRTPLPERKALPVRWITWMTYLEDP RIQFHYDKSLPELQQIPNVTEDVIAKTKHPSEFAMVFYTDGSAIKHPDVNKSHSAGMGIAQVQFIPEFKIVHQWSIPLGDHTAQLAEIAAVEFA CKKALKISGPVLIVTDSFYVAESANKELPYWKSNGFLNNKKKPLRHVSKWKSIAECLQLKPDIIIMHEKGHQQPMTTLHTEGNNLADKLATQGS YVVH
SFV3L_P27401	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLM KKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQ GVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWT RLPQGFLNSPALFTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQ KLLNITPPRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATANGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGY IRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFH YDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKAL KIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401_2mut	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLM KKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQ GVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWT RLPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQ KLLNITPPRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGY IRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFH YDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKAL KIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401_2mutA	MDPLQLLQPLEAEIKGTKLKAHWNSGATITCVPQAFLEEEVPIKNIWIKTIHGEKEQPVYYLTFKIQGRKVEAEVISSPYDYILVSPSDIPWLM KKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVINDLLKQ GVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQQYCWT RLPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLTETFKQ KLLNITPPRDLKQLQSILGKLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTSPSAGY IRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDPRIQFH YDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFACKKAL KIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGSYVVN
SFV3L_P27401-Pro	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVIN DLLKQGVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQ QYCWTRLPQGFLNSPALFTADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLT ETFKQKLLNITPPRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATANGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTS PSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDP RIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFA CKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGS YVVN
SFV3L_P27401-Pro_2mut	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVIN DLLKQGVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQ QYCWTRLPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLT ETFKQKLLNITPPRDLKQLQSILGLLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTS PSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDP RIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFA CKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGS YVVN

SFV3L_P27401-Pro_2mutA	IPWLMKKPLQLTTLVPLQEYEERLLKQTMLTGSYKEKLQSLFLKYDALWQHWENQVGHRRIKPHHIATGTVNPRPQKQYPINPKAKASIQTVIN DLLKQGVLIQQNSIMNTPVYPVPKPDGKWRMVLDYREVNKTIPLIAAQNQHSAGILSSIFRGKYKTTLDLSNGFWAHSITPESYWLTAFTWLGQ QYCWTRLPQGFLNSPALFNADVVDLLKEVPNVQVYVDDIYISHDDPREHLEQLEKVFSLLLNAGYVVSLKKSEIAQHEVEFLGFNITKEGRGLT ETFKQKLLNITPPRDLKQLQSILGKLNFARNFIPNFSELVKPLYNIIATAPGKYITWTTDNSQQLQNIISMLNSAENLEERNPEVRLIMKVNTS PSAGYIRFYNEFAKRPIMYLNYVYTKAEVKFTNTEKLLTTIHKGLIKALDLGMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMSYLEDP RIQFHYDKTLPELQQVPTVTDDIIAKIKHPSEFSMVFYTDGSAIKHPNVNKSHNAGMGIAQVQFKPEFTVINTWSIPLGDHTAQLAEVAAVEFA CKKALKIDGPVLIVTDSFYVAESVNKELPYWQSNGFFNNKKKPLKHVSKWKSIADCIQLKPDIIIIHEKGHQPTASTFHTEGNNLADKLATQGS YVVN
SFVCP_Q87040	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLT QQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWT RLPQGFLNSPALFTADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKT KLLNVTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGY VRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKAL KVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040_2mut	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLT QQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWT RLPQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKT KLLNVTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGY VRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKAL KVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040_2mutA	MNPLQLLQPLPAEVKGTKLLAHWNSGATITCIPESFLEDEQPIKQTLIKTIHGEKQQNVYYLTFKVKGRKVEAEVIASPYEYILLSPTDVPWLT QQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVIDDLLKQ GVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGKQYCWT RLPQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLTDTFKT KLLNVTPPKDLKQLQSILGKLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTSPSAGY VRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDPRIQFH YDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFACKKAL KVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGSYVVN
SFVCP_Q87040-Pro	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVID DLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGK QYCWTRLPQGFLNSPALFTADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLT DTFKTKLLNVTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSKGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTS PSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDP RIQFHYDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFA CKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGS YVVN
SFVCP_Q87040-Pro_2mut	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVID DLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGK QYCWTRLPQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLT DTFKTKLLNVTPPKDLKQLQSILGLLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTS PSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDP RIQFHYDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFA CKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGS YVVN
SFVCP_Q87040-Pro_2mutA	VPWLTQQPLQLTILVPLQEYQDRILNKTALPEEQKQQLKALFTKYDNLWQHWENQVGHRKIRPHNIATGDYPPRPQKQYPINPKAKPSIQIVID DLLKQGVLTPQNSTMNTPVYPVPKPDGRWRMVLDYREVNKTIPLTAAQNQHSAGILATIVRQKYKTTLDLANGFWAHPITPDSYWLTAFTWQGK QYCWTRLPQGFLNSPALFNADAVDLLKEVPNVQVYVDDIYLSHDNPHEHIQQLEKVFQILLQAGYVVSLKKSEIGQRTVEFLGFNITKEGRGLT DTFKTKLLNVTPPKDLKQLQSILGKLNFARNFIPNFAELVQTLYNLIASSPGKYIEWTEDNTKQLNKVIEALNTASNLEERLPDQRLVIKVNTS PSAGYVRYYNESGKKPIMYLNYVFSKAELKFSMLEKLLTTMHKALIKAMDLAMGQEILVYSPIVSMTKIQKTPLPERKALPIRWITWMTYLEDP RIQFHYDKTLPELKHIPDVYTSSIPPLKHPSQYEGVFCTDGSAIKSPDPTKSNNAGMGIVHAIYNPEYKILNQWSIPLGHHTAQMAEIAAVEFA CKKALKVPGPVLVITDSFYVAESANKELPYWKSNGFVNNKKEPLKHISKWKSIAECLSIKPDITIQHEKGHQPINTSIHTEGNALADKLATQGS YVVN
SMRVH_P03364	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQP GLPSPVAIPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDD ILLACDSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPL NNILKGDPNPLSVRALTPEAKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSL LAMLIIKGRYTGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSS NGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLR AHLNLPGPLAEGNALADAATQIFPIISD

SMRVH_P03364_2mut	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQP GLPSPVAIPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDD ILLACDSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPL NNILKPDPNPLSVRALTPEAKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSL LAMLIIKGRYTGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSS NGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLR AHLNLPGPLAEGNALADAATQIFPIISD
SMRVH_P03364_2mutB	PRSRAIDIPVPHADKISWKITDPVWVDQWPLTYEKTLAAIALVQEQLAAGHIEPTNSPWNTPIFIIKKKSGSWRLLQDLRAVNKVMVPMGALQP GLPSPVAPPLNYHKIVIDLKDCFFTIPLHPEDRPYFAFSVPQINFQSPMPRYQWKVLPQGMANSPTLCQKFVAAAIAPVRSQWPEAYILHYMDD ILLACDSAEAAKACYAHIISCLTSYGLKIAPDKVQVSEPFSYLGFELHHQQVFTPRVCLKTDHLKTLNDFQKLLGDIQWLRPYLKLPTSALVPL NNILKPDPNPLSVRALTPEAKQSLALINKAIQNQSVQQISYNLPLVLLLLPTPHTPTAVFWQPNGTDPTKNGSPLLWLHLPASPSKVLLTYPSL LAMLIIKGRYTGRQLFGRDPHSIIIPYTQDQLTWLLQTSDEWAIALSSFTGDIDNHYPSDPVIQFAKLHQFIFPKITKCAPIPQATLVFTDGSS NGIAAYVIDNQPISIKSPYLSAQLVELYAILQVFTVLAHQPFNLYTDSAYIAQSVPLLETVPFIKSSTNATPLFSKLQQLILNRQHPFFIGHLR AHLNLPGPLAEGNALADAATQIFPIISD
SRV2_P51517	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQ PGLPSPVAIPQGYFKIVIDLKDCFFTIPLQPVDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYMD DILIAGKLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRDKLQTLNDFQKLLGDINWLRPYLHLTTGDLKP LFDILKGDSNPNSPRSLSEAALASLQKVETAIAEQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILG RDNSKKYFGLEPSTIIQPYSKSQIHWLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAVVFPRIISKTPLDNALLVFTDGSSTGIAAYTF EKTTVRFKTSHTSAQLVELQALIAVLSAFPHRALNVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGP LSQGNHITDLATKVVATTLTT
SRV2_P51517_2mutB	LATAVDILAPQRYADPITWKSDEPVWVDQWPLTQEKLAAAQQLVQEQLQAGHIIESNSPWNTPIFVIKKKSGKWRLLQDLRAVNATMVLMGALQ PGLPSPVAPPQGYFKIVIDLKDCFFTIPLQPVDQKRFAFSLPSTNFKQPMKRYQWKVLPQGMANSPTLCQKYVAAAIEPVRKSWAQMYIIHYMD DILIAGKLGEQVLQCFAQLKQALTTTGLQIAPEKVQLQDPYTYLGFQINGPKITNQKAVIRRDKLQTLNDFQKLLGDINWLRPYLHLTTGDLKP LFDILKGDSNPNSPRSLSEAALASLQKVETAIAEQFVTQIDYTQPLTFLIFNTTLTPTGLFWQNNPVMWVHLPASPKKVLLPYYDAIADLIILG RDNSKKYFGLEPSTIIQPYSKSQIHWLMQNTETWPIACASYAGNIDNHYPPNKLIQFCKLHAVVFPRIISKTPLDNALLVFTDGSSTGIAAYTF EKTTVRFKTSHTSAQLVELQALIAVLSAFPHRALNVYTDSAYLAHSIPLLETVSHIKHISDTAKFFLQCQQLIYNRSIPFYLGHIRAHSGLPGP LSQGNHITDLATKVVATTLTT
WDSV_O92815	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAG RDEYRMIHDLRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFSQALYQ SLHKIKFKISSEICIYMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQI RAFLGLVGYCRHWIPEFSIHSKFLEKQLKKDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPI AFLSSKFDAIESGLPPCLKACASIHRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQS CENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSDGSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVN IYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WDSV_O92815_2mut	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAG RDEYRMIHDLRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQ SLHKIKFKISSEICIYMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQI RAFLGLVGYCRHWIPEFSIHSKFLEKQLKPDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPI AFLSSKFDAIESGLPPCLKACASIHRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQS CENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSDGSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVN IYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WDSV_O92815_2mutA	SCQTKNTLNIDEYLLQFPDQLWASLPTDIGRMLVPPITIKIKDNASLPSIRQYPLPKDKTEGLRPLISSLENQGILIKCHSPCNTPIFPIKKAG RDEYRMIHDLRAINNIVAPLTAVVASPTTVLSNLAPSLHWFTVIDLSNAFFSVPIHKDSQYLFAFTFEGHQYTWTVLPQGFIHSPTLFNQALYQ SLHKIKFKISSEICIYMDDVLIASKDRDTNLKDTAVMLQHLASEGHKVSKKKLQLCQQEVVYLGQLLTPEGRKILPDRKVTVSQFQQPTTIRQI RAFLGKVGYCRHFIPEFSIHSKFLEKQLKPDTAEPFQLDDQQVEAFNKLKHAITTAPVLVVPDPAKPFQLYTSHSEHASIAVLTQKHAGRTRPI AFLSSKFDAIESGLPPCLKACASIHRSLTQADSFILGAPLIIYTTHAICTLLQRDRSQLVTASRFSKWEADLLRPELTFVACSAVSPAHLYMQS CENNIPPHDCVLLTHTISRPRPDLSDLPIPDPDMTLFSDGSYTTGRGGAAVVMHRPVTDDFIIIHQQPGGASAQTAELLALAAACHLATDKTVN IYTDSRYAYGVVHDFGHLWMHRGFVTSAGTPIKNHKEIEYLLKQIMKPKQVSVIKIEAHTKGVSMEVRGNAAADEAAKNAVFLVQR
WMSV_P03359	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNT PLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQ GFKNSPTLFDEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPAR KATVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKESIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVAR GVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFA PPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALT QALRLAEGKDINIYTDSRYAFATAHIHGAIYKQRGLLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGNDPVATGNRRADEAAKQAAL STRVLAETTKP
WMSV_P03359_3mut	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNT PLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQ GFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPAR KATVMKIPPPTTPRQVREFLGTAGFCRLWIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVAR GVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFA PPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALT QALRLAEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGNDPVATGNRRADEAAKQAAL STRVLAETTKP

WMSV_P03359_3mutA	VLNLEEEYRLHEKPVPSSIDPSWLQLFPTVWAERAGMGLANQVPPVVVELRSGASPVAVRQYPMSKEAREGIRPHIQRFLDLGVLVPCQSPWNTPLLPVKKPGTNDYRPVQDLREINKRVQDIHPTVPNPYNLLSSLPPSHTWYSVLDLKDAFFCLKLHPNSQPLFAFEWRDPEKGNTGQLTWTRLPQGFKNSPTLFNEALHRDLAPFRALNPQVVLLQYVDDLLVAAPTYRDCKEGTQKLLQELSKLGYRVSAKKAQLCQKEVTYLGYLLKEGKRWLTPARKATVMKIPPPTTPRQVREFLGKAGFCRLFIPGFASLAAPLYPLTKPSIPFIWTEEHQKAFDRIKEALLSAPALALPDLTKPFTLYVDERAGVARGVLTQTLGPWRRPVAYLSKKLDPVASGWPTCLKAVAAVALLLKDADKLTLGQNVTVIASHSLESIVRQPPDRWMTNARMTHYQSLLLNERVSFAPPAVLNPATLLPVESEATPVHRCSEILAEETGTRRDLKDQPLPGVPAWYTDGSSFIAEGKRRAGAAIVDGKRTVWASSLPEGTSAQKAELVALTQALRLAEGKDINIYTDSRYAFATAHIHGAIYKQRGWLTSAGKDIKNKEEILALLEAIHLPKRVAIIHCPGHQKGNDPVATGNRRADEAAKQAALSTRVLAETTKP
XMRV6_A1Z651	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEAPHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHVHGEIYRRRGLLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARE AAMKAVLETSTLL
XMRV6_A1Z651_3mut	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEAPHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHVHGEIYRRRGWLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARE AAMKAVLETSTLL
XMRV6_A1Z651_3mutA	TLNIEDEYRLHETSKEPDVPLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALHRDLADFRIQHPDLILLQYVDDLLLAATSEQDCQRGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGKAGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQAMLLDTDRVQFGPVVALNPATLLPLPEKEAPHDCLEILAETHGTRPDLTDQPIPDADYTWYTDGSSFLQEGQRRAGAAVTTETEVIWARALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHVHGEIYRRRGWLTSEGREIKNKNEILALLKALFLPKRLSIIHCPGHQKGNSAEARGNRMADQAARE AAMKAVLETSTLL

[0181] 일부 구현예에서, 역전사효소 도메인은, 예를 들어 부위-특이적 돌연변이에 의해 변형된다. 일부 구현예에서, 역전사효소 도메인은 개선된 특성을 갖도록 조작되며, 예를 들어 SuperScript IV(SSIV) 역전사효소는 MMLV RT로부터 유래되었다. 일부 구현예에서, 역전사효소 도메인은 예를 들어 본 명세서에 참고로 포함된 WO2001068895에 기재된 바와 같이 더 낮은 오류율을 갖도록 조작될 수 있다. 일부 구현예에서, 역전사효소 도메인은 더 열안정성이 있도록 조작될 수 있다. 일부 구현예에서, 역전사효소 도메인은 더 진행성(processive)이 되도록 조작될 수 있다. 일부 구현예에서, 역전사효소 도메인은 억제제에 대한 내성을 갖도록 조작될 수 있다. 일부 구현예에서, 역전사효소 도메인은 더 신속하도록 조작될 수 있다. 일부 구현예에서, 역전사효소 도메인은 RNA 주형에서 변형된 뉴클레오티드를 더 양호하게 용인하도록 조작될 수 있다. 일부 구현예에서, 역전사효소 도메인은 변형된 DNA 뉴클레오티드를 삽입하도록 조작될 수 있다. 일부 구현예에서, 역전사효소 도메인은 주형 RNA에 결합하도록 조작된다. 일부 구현예에서, 하나 이상의 돌연변이는 뮤린 백혈병 바이러스 역전사효소의 RT 도메인 내 D200N, L603W, T330P, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, W313F, L435G, N454K, H594Q, L671P, E69K, 또는 D653N로부터 선택되거나, 또는 또 다른 RT 도메인의 상응하는 위치에서의 상응하는 돌연변이이다.

[0182] 일부 구현예에서, 유전자 변형 폴리펩티드는, 예를 들어 다음 서열을 포함하는, 레트로바이러스 역전사효소, 예를 들어 야생형 M-MLV RT로부터 유래된 RT 도메인을 포함한다:

[0183] *M-MLV(WT):*

TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALHRDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGTAGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPPCLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLTDQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIKNKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLI (서열번호 4012)

[0184] 일부 구현예에서, 유전자 변형 폴리펩티드는, 예를 들어 다음 서열을 포함하는, 레트로바이러스 역전사효소, 예

를 들어 M-MLV RT로부터 유래된 RT 도메인을 포함한다:

[0185] TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVK
KPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALH
RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGT
AGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP
CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT
DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIK
NKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLL (서열번호 4013)

[0186] 일부 구현예에서, 유전자 변형 폴리펩티드는 NP_057933의 아미노산 659 내지 1329의 서열을 포함하는 레트로바이러스 역전사효소로부터 유래된 RT 도메인을 포함한다. 구현예에서, 유전자 변형 폴리펩티드는, 예를 들어 하기 나타낸 바와 같이, NP_057933의 아미노산 659 내지 1329의 서열의 N-말단에 하나의 추가 아미노산을 추가로 포함한다:

[0187] TLNIEDEHRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLP**VK**
KPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFDEALH
RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGT
AGFCRLWIPGFAEMAAPLYPLTKTGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP
CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT
DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGLLTSEGKEIK
NKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAA (서열번호 4014)

[0188] **코어 RT**(굵은 글씨), 위에 주석이 달림

[0189] RNAseH(밑줄), 위에 주석이 달림

[0190] 구현예에서, 유전자 변형 폴리펩티드는 NP_057933의 아미노산 659 내지 1329의 서열의 C-말단에 하나의 추가 아미노산을 추가로 포함한다: 구현예에서, 유전자 변형 폴리펩티드는 RNaseH1 도메인(예를 들어, NP_057933의 아미노산 1178 내지 1318)을 포함한다.

[0191] 일부 구현예에서, 레트로바이러스 역전사효소 도메인, 예를 들어, M-MLV RT는 RT의 특징, 예를 들어 열안정성, 가공성, 및/또는 주형 결합을 개선시킬 수 있는 야생형 서열로부터 하나 이상의 돌연변이를 포함할 수 있다. 일부 구현예에서, M-MLV RT 도메인은 상기 M-MLV(WT) 서열에 대하여, 예를 들어 D200N, L603W, T330P, T306K, W313F, D524G, E562Q, D583N, P51L, S67R, E67K, T197A, H204R, E302K, F309N, L435G, N454K, H594Q, D653N, R110S, K103L, 예를 들어 D200N, L603W, 및 T330P와 같은 돌연변이의 조합(선택적으로 T306K 및 W313F를 추가로 포함함)으로부터 선택되는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, 본 명세서에 사용된 M-MLV RT는 돌연변이 D200N, L603W, T330P, T306K 및 W313F를 포함한다. 구현예에서, 돌연변이체 M-MLV RT는 다음 아미노산 서열을 포함한다:

[0192] *M-MLV(PE2):*
TLNIEDEYRLHETSKEPDVSLGSTWLSDFPQAWAETGGMGLAVRQAPLIIPLKATSTPVSIKQYPMSQEARLGIKPHIQRLLDQGILVPCQSPWNTPLLPVK
KPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLSGLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRDPEMGISGQLTWTRLPQGFKNSPTLFNEALH
RDLADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRALLQTLGNLGYRASAKKAQICQKQVKYLGYLLKEGQRWLTEARKETVMGQPTPKTPRQLREFLGK
AGFCRLFIPGFAEMAAPLYPLTKPGTLFNWGPDQQKAYQEIKQALLTAPALGLPDLTKPFELFVDEKQGYAKGVLTQKLGPWRRPVAYLSKKLDPVAAGWPP
CLRMVAAIAVLTKDAGKLTMGQPLVILAPHAVEALVKQPPDRWLSNARMTHYQALLLDTDRVQFGPVVALNPATLLPLPEEGLQHNCLDILAEAHGTRPDLT
DQPLPDADHTWYTDGSSLLQEGQRKAGAAVTTETEVIWAKALPAGTSAQRAELIALTQALKMAEGKKLNVYTDSRYAFATAHIHGEIYRRRGWLTSEGKEIK
NKDEILALLKALFLPKRLSIIHCPGHQKGHSAEARGNRMADQAARKAAITETPDTSTLLI (서열번호 4015)

[0193] 일부 구현예에서, 라이팅 도메인(예를 들어, RT 도메인)은, 예를 들어 RNA 서열에 특이적으로 결합하는 RNA-결합 도메인을 포함한다. 일부 구현예에서, 주형 RNA는 라이팅 도메인의 RNA-결합 도메인에 의해 특이적으로 결합된 RNA 서열을 포함한다.

[0194] 일부 구현예에서, 역전사 도메인은 특정 주형, 예를 들어 시스템의 주형 RNA만을 인식하고 역전사한다. 일부 구현예에서, 주형은 역전사 도메인에 의한 인식 및 역전사를 가능하게 하는 서열 또는 구조를 포함한다. 일부 구현예에서, 주형은 본 명세서에 기재된 게놈 조작 시스템의 폴리펩티드 구성요소의 RNA-결합 도메인과의 회합을 가능하게 하는 서열 또는 구조를 포함한다. 일부 실시형태에서, 게놈 조작 시스템은 바람직하게는 회합 서열이

결여된 주형에 비해 회합 서열을 포함하는 주형을 역전사한다.

[0195] 라이팅 도메인은 또한 DNA-의존성 DNA 중합효소 활성을 포함할 수 있으며, 예를 들어 주형 DNA 서열로부터 게놈 내로 DNA를 작성할 수 있는 효소 활성을 포함할 수 있다. 일부 구현예에서, DNA-의존성 DNA 중합은 표적 부위 편집의 제2 가닥 합성을 완료하는 데 이용된다. 일부 구현예에서, DNA-의존성 DNA 중합효소 활성은 폴리펩티드 내 DNA 중합효소 도메인에 의해 제공된다. 일부 구현예에서, DNA-의존성 DNA 중합효소 활성은 또한 DNA-의존성 DNA 중합, 예를 들어 제2 가닥 합성이 가능한 역전사효소 도메인에 의해 제공된다. 일부 구현예에서, DNA-의존성 DNA 중합효소 활성은 시스템의 제2 폴리펩티드에 의해 제공된다. 일부 구현예에서, DNA-의존성 DNA 중합효소 활성은 선택적으로 게놈 조작 시스템의 구성요소에 의해 표적 부위로 동원되는 내인성 숙주 세포 중합효소에 의해 제공된다.

[0196] 일부 구현예에서, 역전사효소 도메인은 참조 역전사효소 도메인에 비해 더 낮은 확률의 시험관 내 조기 종료율(P_{off})을 갖는다. 일부 구현예에서, 참조 역전사효소 도메인은 바이러스 역전사효소 도메인, 예를 들어 M-MLV 유래의 RT 도메인이다.

[0197] 일부 구현예에서, 역전사효소 도메인은, 예를 들어 1094 nt RNA 상에서 측정 시, 약 5 x 10^{-3}/nt, 5 x 10^{-4}/nt, 또는 5 x 10^{-6}/nt 미만의, 낮은 확률의 시험관 내 조기 종료율(P_{off})을 갖는다. 구현예에서, 시험관 내 조기 종료율은 문헌[Bibillo and Eickbush (2002) J Biol Chem 277(38):34836-34845(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이 결정된다.

[0198] 일부 구현예에서, 역전사효소 도메인은 적어도 약 30% 또는 50%의 세포 내 통합을 완료할 수 있다. 완전한 통합 백분율은 실질적으로 전장 통합 사건의 수(예를 들어, 예측된 통합된 서열의 적어도 98%를 포함하는 게놈 부위)를 세포 집단에서 총(실질적으로 전장 및 부분적 통합을 포함함) 통합 사건의 수로 나눔으로써 측정될 수 있다. 구현예에서, 세포 내 통합은, 예를 들어 문헌[Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903(이는 그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, (예를 들어, 통합 부위에 걸쳐) 긴-판독 암플리콘 시퀀싱을 이용하여 결정된다.

[0199] 구현예에서, 세포 내 통합의 정량화는 주형 RNA에 상응하는 DNA 서열을 적어도 약 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 함유하는 통합의 분획을 계수하는 것을 포함한다(예를 들어, 주형 RNA는 적어도 0.05, 0.1, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.5, 2, 3, 4, 또는 5 kb, 예를 들어 길이가 0.5 내지 0.6, 0.6 내지 0.7, 0.7 내지 0.8, 0.8 내지 0.9, 1.0 내지 1.2, 1.2 내지 1.4, 1.4 내지 1.6, 1.6 내지 1.8, 1.8 내지 2.0, 2 내지 3, 3 내지 4, 또는 4 내지 5 kb의 길이를 가짐).

[0200] 일부 구현예에서, 역전사효소 도메인은 시험관 내에서 dNTP를 중합할 수 있다. 구현예에서, 역전사효소 도메인은 시험관 내에서 0.1 내지 50 nt/sec(예를 들어, 0.1 내지 1, 1 내지 10, 또는 10 내지 50 nt/sec)의 비로 dNTP를 중합할 수 있다. 구현예에서, 역전사효소 도메인에 의한 dNTP의 중합은 단일 분자 분석에 의해, 예를 들어 문헌[Schwartz and Quake (2009) *PNAS* 106(48):20294-20299(그 전체가 참고로 포함됨)]에 기재된 바와 같이 측정된다.

[0201] 일부 구현예에서, 역전사효소 도메인은, 예를 들어 문헌[Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 1 x 10^{-3} 내지 1 x 10^{-4} 또는 1 x 10^{-4} 내지 1 x 10^{-5} 치환/nt의 시험관 내 오류율(예를 들어, 뉴클레오티드의 오혼입(misincorporation))을 갖는다. 일부 구현예에서, 역전사효소 도메인은, 예를 들어 문헌[Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 예를 들어 긴-판독 암플리콘 시퀀싱에 의해, 1 x 10^{-3} 내지 1 x 10^{-4} 또는 1 x 10^{-4} 내지 1 x 10^{-5} 치환/nt의 세포 내 오류율을 갖는다.

[0202] 일부 구현예에서, 역전사효소 도메인은 시험관 내 표적 RNA의 역전사를 수행할 수 있다. 일부 구현예에서, 역전사효소는 주형의 역전사를 개시하는 데 적어도 3개 뉴클레오티드의 프라이머를 필요로 한다. 일부 구현예에서, 표적 RNA의 역전사는, 예를 들어 문헌[Bibillo and Eickbush (2002) *J Biol Chem* 277(38):34836-34845(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이 표적 RNA로부터의 cDNA의 검출에 의해 결정된다(예를 들어, ssDNA 프라이머가 제공되는 경우, 예를 들어, 3' 말단에서 적어도 3, 4, 5, 6, 7, 8, 9, 또는 10 nt를 이용하여 표적에 어닐링함).

[0203] 일부 구현예에서, 예를 들어 그의 RNA 주형을 cDNA로 전환하는 경우, 역전사효소 도메인은, 예를 들어 단백질

결합 모티프(예를 들어, 3' UTR)가 결여된 RNA 주형과 비교하여, (예를 들어, cDNA 생산에 의해) 적어도 5 또는 10배 이상 효율적으로 역전사를 수행한다. 구현예에서, 역전사의 효율은 문헌[Yasukawa et al. (2017) *Biochem Biophys Res Commun* 492(2):147-153(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이 측정된다.

[0204] 일부 구현예에서, 예를 들어 세포(예를 들어, HEK293T 세포) 내에서 발현된 경우, 역전사효소 도메인은 임의의 내인성 세포 RNA보다 더 높은 빈도(예를 들어, 약 5 또는 10배 더 높은 빈도)로 특이적 RNA 주형에 특이적으로 결합한다. 구현예에서, 역전사효소 도메인과 주형 RNA간의 특이적 결합의 빈도는, 예를 들어 문헌[Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이 CLIP-seq에 의해 측정된다.

[0205] *주형 핵산 결합 도메인*

[0206] 유전자 변형 폴리펩티드는 통상적으로 주형 핵산(예를 들어, 주형 RNA)과 회합할 수 있는 영역을 함유한다. 일부 구현예에서, 주형 핵산 결합 도메인은 RNA 결합 도메인이다. 일부 구현예에서, RNA 결합 도메인은 특이적 시그니처, 예를 들어 구조적 모티프를 함유하는 RNA 분자와 회합할 수 있는 모듈 도메인이다. 다른 구현예에서, 주형 핵산 결합 도메인(예를 들어, RNA 결합 도메인)은 역전사 도메인 내에 함유되고, 예를 들어 역전사효소-유래된 구성요소는 RNA 선호도에 대한 알려진 시그니처를 갖는다.

[0207] 다른 구현예에서, 주형 핵산 결합 도메인(예를 들어, RNA 결합 도메인)은 표적 DNA 결합 도메인 내에 포함된다. 예를 들어, 일부 구현예에서, DNA 결합 도메인은 gRNA를 포함하는 주형 핵산(예를 들어, 주형 RNA)의 구조를 인식하는 CRISPR-연관 단백질이다. 일부 구현예에서, 유전자 변형 폴리펩티드는 DNA-결합 도메인이 표적 게놈 DNA 서열에 결합하도록 허용하는 gRNA 스캐폴드와 회합하는 CRISPR-연관 단백질을 포함하는 DNA-결합 도메인을 포함한다. 일부 구현예에서, gRNA 스캐폴드 및 gRNA 스페이서는 주형 핵산(예를 들어, 주형 RNA) 내에 포함되며, 따라서 DNA-결합 도메인은 또한 주형 핵산 결합 도메인이다. 일부 구현예에서, 폴리펩티드는 다수의 도메인에서 RNA 결합 기능을 가지며, 예를 들어 CRISPR-연관 DNA 결합 도메인에서 gRNA 구조 및 역전사효소 도메인 내 추가의 서열 또는 구조에 결합할 수 있다.

[0208] 일부 구현예에서, RNA 결합 도메인은 참조 RNA 결합 도메인보다 더 큰 친화도로 주형 RNA에 결합할 수 있다. 일부 구현예에서, 참조 RNA 결합 도메인은 S. 피오게네스(S. pyogenes)의 Cas9로부터의 RNA 결합 도메인이다. 일부 구현예에서, RNA 결합 도메인은 100 pM 내지 10 nM(예를 들어, 100 pM 내지 1 nM 또는 1 nM 내지 10 nM)의 친화도로 주형 RNA에 결합할 수 있다. 일부 구현예에서, 그의 주형 RNA에 대한 RNA 결합 도메인의 친화도는 시험관 내에서, 예를 들어 문헌[Asmari et al. Methods 146:107-119 (2018)(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 예를 들어 열영동에 의해 측정된다. 일부 구현예에서, 그의 주형 RNA에 대한 RNA 결합 도메인의 친화도는 세포 내에서 (예를 들어, FRET 또는 CLIP-Seq에 의해) 측정된다.

[0209] 일부 구현예에서, RNA 결합 도메인은 스크램블드(scrambled) RNA를 이용한 경우보다 적어도 약 5배 또는 10배 더 높은 빈도로 시험관 내에서 주형 RNA와 회합된다. 일부 구현예에서, RNA 결합 도메인과 주형 RNA 또는 스크램블드 RNA 간의 회합의 빈도는, 예를 들어 문헌[Lin and Miles (2019) *Nucleic Acids Res* 47(11):5490-5501(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이 CLIP-seq에 의해 측정된다. 일부 구현예에서, RNA 결합 도메인은 스크램블드 RNA를 이용한 경우보다 적어도 약 5배 또는 10배 더 높은 빈도로 세포(예를 들어, HEK293T 세포) 내에서 주형 RNA와 회합된다. 일부 구현예에서, RNA 결합 도메인과 주형 RNA 또는 스크램블드 RNA 간의 회합의 빈도는, 예를 들어 문헌[Lin and Miles (2019) 상동]에 기재된 바와 같이 CLIP-seq에 의해 측정된다.

[0210] *엔도뉴클레아제 도메인 및 DNA 결합 도메인*

[0211] 일부 구현예에서, 유전자 변형 폴리펩티드는 엔도뉴클레아제 도메인을 통한 DNA 표적 부위 절단 기능을 갖는다. 일부 구현예에서, 유전자 변형 폴리펩티드는, 예를 들어 표적 핵산에 결합하기 위한, DNA 결합 도메인을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드의 도메인(예를 들어, Cas 도메인)은 2개 이상의 더 작은 도메인, 예를 들어 DNA 결합 도메인 및 엔도뉴클레아제 도메인을 포함한다. DNA 결합 도메인(예를 들어, Cas 도메인)이 표적 핵산 서열에 결합한다고 할 때, 일부 구현예에서 결합은 gRNA에 의해 매개되는 것으로 이해된다.

[0212] 일부 구현예에서, 도메인은 2가지 기능을 갖는다. 예를 들어, 일부 구현예에서, 엔도뉴클레아제 도메인은 또한 DNA-결합 도메인이다. 일부 구현예에서, 엔도뉴클레아제 도메인은 또한 주형 핵산(예를 들어, 주형 RNA) 결합 도메인이다. 예를 들어, 일부 구현예에서, 폴리펩티드는 gRNA를 포함하는 주형 RNA에 결합하고, (예를 들어, gRNA의 일부에 대한 상보성을 갖는) 표적 DNA 서열에 결합하고, 표적 DNA 서열을 절단하는 CRISPR-연관 엔도뉴

클레아제 도메인을 포함한다. 일부 구현예에서, 이종성 공급원 유래의 엔도뉴클레아제 도메인 또는 엔도뉴클레아제/DNA-결합 도메인이 본 명세서에 기재된 유전자 변형 시스템에서 사용될 수 있거나 (예를 들어, 하나 이상의 잔기의 삽입, 결실, 또는 치환에 의해) 변형될 수 있다.

[0213] 일부 구현예에서, 엔도뉴클레아제 도메인 또는 엔도뉴클레아제/DNA 결합 도메인을 인코딩하는 핵산은, 예를 들어 인간 세포에 대해 개선된, 변경된 코돈 사용을 갖기 위해 그의 천연 서열에서 변경된다. 일부 구현예에서, 엔도뉴클레아제 요소는 이종성 엔도뉴클레아제 요소, 예컨대 Cas 엔도뉴클레아제(예를 들어, Cas9), II형 제한 엔도뉴클레아제(예를 들어, Fok1), 메가뉴클레아제(예를 들어, I-SceI), 또는 다른 엔도뉴클레아제 도메인이다.

[0214] 특정 양태에서, 본 명세서에 기재된 유전자 변형 폴리펩티드의 DNA-결합 도메인은 원하는 숙주 DNA 표적 서열에 대한 결합을 위해 선택, 설계 또는 구축된다. 특정 구현예에서, 폴리펩티드의 DNA-결합 도메인은 이종성 DNA-결합 요소이다. 일부 구현예에서, 이종성 DNA 결합 요소는 징크-핑거 요소 또는 TAL 이펙터 요소, 예를 징크-핑거 또는 TAL 폴리펩티드 또는 이의 기능성 단편이다. 일부 구현예에서, 이종성 DNA 결합 요소는 서열-가이드된 DNA 결합 요소, 예컨대 Cas9, Cpf1, 또는 엔도뉴클레아제 활성을 갖지 않도록 변경된 기타 CRISPR-관련 단백질이다. 일부 구현예에서, 이종성 DNA 결합 요소는 엔도뉴클레아제 활성을 보유한다. 일부 구현예에서, 이종성 DNA 결합 요소는 ssDNA를 절단하는 부분적인 엔도뉴클레아제 활성을 보유하며, 예를 들어 닉카제 활성을 갖는다. 특정 구현예에서, 이종성 DNA 결합 도메인은 Cas9, TAL 도메인, ZF 도메인, Myb 도메인, 이의 조합, 또는 이의 다수 중 임의의 하나 이상일 수 있다.

[0215] 일부 구현예에서, DNA-결합 도메인은, 예를 들어 부위-특이적 돌연변이, DNA-결합 요소(예를 들어, 징크 핑거의 수 및/또는 특이성)의 증가 또는 감소 등에 의해 변형되어, DNA-결합 특이성 및 친화도를 변경시킨다. 일부 구현예에서, DNA 결합 도메인을 인코딩하는 핵산 서열은, 예를 들어 인간 세포에 대해 개선된, 변경된 코돈 사용을 갖기 위해 그의 천연 서열에서 변경된다. 구현예에서, DNA 결합 도메인은 야생형 DNA 결합 도메인에 대한 하나 이상의 변형, 예를 들어 지향된 진화, 예를 들어 파지-보조된 지속적 진화(phage-assisted continuous evolution; PACE)를 통한 변형을 포함한다.

[0216] 일부 구현예에서, DNA 결합 도메인은 메가뉴클레아제 도메인(예를 들어, 본 명세서에, 예를 들어 엔도뉴클레아제 도메인 섹션에 기재된 바와 같음), 또는 이의 기능성 단편을 포함한다. 일부 구현예에서, 메가뉴클레아제 도메인은 엔도뉴클레아제 활성, 예를 들어 이중 가닥 절단 및/또는 닉카제 활성을 갖는다. 다른 구현예에서, 메가뉴클레아제 도메인은 감소된 활성을 갖고, 예를 들어 엔도뉴클레아제 활성이 결여되고, 예를 들어 메가뉴클레아제는 촉매적으로 비활성이다. 일부 구현예에서, 촉매적으로 비활성인 메가뉴클레아제는, 예를 들어 그 전체가 본 명세서에 참고로 포함된 문헌[Fonfara et al. Nucleic Acids Res 40(2):847-860 (2012)]에 기재된 바와 같이, DNA 결합 도메인으로 사용된다.

[0217] 일부 구현예에서, 유전자 변형 폴리펩티드는, 예를 들어 야생형 폴리펩티드에 비해 DNA-결합 도메인에 대한 변형을 포함한다. 일부 구현예에서, DNA-결합 도메인은 원래 DNA-결합 도메인의 아미노산 서열에 대한 추가, 결실, 대체, 또는 변형을 포함한다. 일부 구현예에서, DNA-결합 도메인은 관심 표적 핵산(예를 들어, DNA) 서열에 특이적으로 결합하는 이종 기능성 도메인을 포함하도록 변형된다. 일부 구현예에서, 기능성 도메인은 폴리펩티드의 이전 DNA-결합 도메인의 적어도 일부(예를 들어, 전체)를 대체한다. 일부 구현예에서, 기능성 도메인은 징크 핑거(예를 들어, 관심 표적 핵산(예를 들어, DNA)에 특이적으로 결합하는 징크 핑거)를 포함한다. 일부 구현예에서, 기능성 도메인은 Cas 도메인(예를 들어, 관심 표적 핵산(예를 들어, DNA)에 특이적으로 결합하는 Cas 도메인)을 포함한다. 일부 구현예에서, Cas 도메인은 Cas9 또는 이의 돌연변이체 또는 변이체(예를 들어, 본 명세서에 기재된 바와 같음)를 포함한다. 구현예에서, Cas 도메인은, 예를 들어 본 명세서에 기재된 바와 같이 가이드 RNA(gRNA)와 회합된다. 구현예에서, Cas 도메인은 gRNA에 의해 관심 표적 핵산(예를 들어, DNA) 서열로 지시된다. 구현예에서, Cas 도메인은 gRNA와 동일한 핵산(예를 들어, RNA) 분자에서 인코딩된다. 구현예에서, Cas 도메인은 gRNA와 상이한 핵산(예를 들어, RNA) 분자에서 인코딩된다.

[0218] 일부 구현예에서, DNA 결합 도메인은 참조 DNA 결합 도메인보다 더 큰 친화도로 표적 서열(예를 들어, dsDNA 표적 서열)에 결합할 수 있다. 일부 구현예에서, 참조 DNA 결합 도메인은 S. 피오게네스의 Cas9로부터의 DNA 결합 도메인이다. 일부 구현예에서, DNA 결합 도메인은 100 pM 내지 10 nM(예를 들어, 100 pM 내지 1 nM 또는 1 nM 내지 10 nM)의 친화도로 표적 서열(예를 들어, dsDNA 표적 서열)에 결합할 수 있다.

[0219] 일부 구현예에서, 그의 표적 서열(예를 들어, dsDNA 표적 서열)에 대한 DNA 결합 도메인의 친화도는 시험관 내에서, 예를 들어 문헌[Asmari et al. Methods 146:107-119 (2018)(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 예를 들어 열영동에 의해 측정된다.

[0220] 구현예에서, DNA 결합 도메인은 몰 과량, 예를 들어 약 100배 몰 과량의 스크램블드 서열 경쟁자 dsDNA 존재 하에서, 100 pM 내지 10 nM(예를 들어, 100 pM 내지 1 nM 또는 1 nM 내지 10 nM) 사이의 친화도로 표적 서열(예를 들어, dsDNA 표적 서열)에 결합할 수 있다.

[0221] 일부 구현예에서, DNA 결합 도메인은, 예를 들어 ChIP-seq(예를 들어, HEK293T 세포에서)에 의해, 예를 들어 문헌[He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이 측정 시, 표적 세포, 예를 들어 인간 표적 세포의 게놈 내의 임의의 다른 서열보다 더 빈번하게 그의 표적 서열(예를 들어, dsDNA 표적 서열)과 회합하는 것으로 발견된다. 일부 구현예에서, DNA 결합 도메인은, 예를 들어 ChIP-seq(예를 들어, HEK293T 세포에서)에 의해, 예를 들어 문헌[He and Pu (2010), 상동]에 기재된 바와 같이 측정 시, 표적 세포의 게놈 내의 임의의 다른 서열보다 적어도 약 5배 또는 10배 더 빈번하게 그의 표적 서열(예를 들어, dsDNA 표적 서열)과 회합하는 것으로 발견된다.

[0222] 일부 구현예에서, 엔도뉴클레아제 도메인은 닉카제 활성을 갖고, 표적 DNA의 하나의 가닥을 절단한다. 일부 구현예에서, 닉카제 활성은 표적 부위에서 이중 가닥 파손의 형성을 감소시킨다. 일부 구현예에서, 엔도뉴클레아제 도메인은 표적 DNA의 제1 가닥과 제2 가닥에 엇갈린(staggered) 닉 구조를 형성한다. 일부 구현예에서, 엇갈린 닉 구조는 표적 부위에 자유 3' 오버행(overhang)을 생성한다. 일부 구현예에서, 표적 부위의 자유 3' 오버행은, 예를 들어 주형 핵산의 3' 상동성 영역의 접근 및 어닐링을 향상시킴으로써 편집 효율성을 개선시킨다. 일부 구현예에서, 엇갈린 닉 구조는 표적 부위에서 이중 가닥 파손의 형성을 감소시킨다.

[0223] 일부 구현예에서, 엔도뉴클레아제 도메인은 표적 DNA의 가닥 둘 다를 절단하며, 예를 들어 절단 부위의 양쪽에 ssDNA 오버행이 없는 표적의 평활 말단(blunt-end) 절단을 초래한다. 본 명세서에 기재된 유전자 변형 시스템의 엔도뉴클레아제 도메인의 아미노산 서열은 본 명세서에 기재된 엔도뉴클레아제 도메인의 아미노산 서열에 대하여 적어도 약 50%, 적어도 약 60%, 적어도 약 70%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95%, 적어도 약 96%, 적어도 약 97%, 적어도 약 98%, 적어도 약 99% 동일할 수 있다.

[0224] 특정 구현예에서, 이종성 엔도뉴클레아제는 Fok1 또는 이의 기능성 단편이다. 특정 구현예에서, 이종성 엔도뉴클레아제는 홀리데이 접합 리졸바제(resolvase) 또는 이의 동족체, 예컨대 솔폴로부스 솔파타리쿠스-솔헤(Sulfolobus solfataricus--Ssol Hje)로부터의 홀리데이 접합 리졸바제이다(Govindaraju et al., *Nucleic Acids Research* 44:7, 2016). 특정 구현예에서, 이종성 엔도뉴클레아제는 Prp8과 같은 스플라이스좀 단백질의 큰 단편의 엔도뉴클레아제이다(Mahbub et al., *Mobile DNA* 8:16, 2017). 특정 구현예에서, 이종성 엔도뉴클레아제는 CRISPR-연관 단백질, 예를 들어 Cas9로부터 유래된다. 특정 구현예에서, 이종성 엔도뉴클레아제는 단지 ssDNA 절단 활성만, 예를 들어 닉카제 활성만 갖도록 조작되며, 예를 들어 Cas9 닉카제, 예를 들어 D10A, H840A, 또는 N863A 돌연변이를 갖는 SpCas9가 된다. 표 3은 예시적인 Cas 단백질 및 닉카제 활성과 연관된 돌연변이를 제공한다. 또 다른 구현예에서, 동종성 엔도뉴클레아제 도메인은, 예를 들어 부위-특이적 돌연변이에 의해 변형되어, DNA 엔도뉴클레아제 활성을 변경시킨다. 또 다른 구현예에서, 엔도뉴클레아제 도메인은, 예를 들어 DNA-서열 특이성을 부여하는 도메인을 제거하기 위한 절단 또는 DNA-서열 특이성을 부여하는 영역을 비활성화시키기 위한 돌연변이에 의해, DNA-서열 특이성을 감소시키도록 변형된다.

[0225] 일부 구현예에서, 엔도뉴클레아제 도메인은 닉카제 활성을 갖고, 이중 가닥 파손을 형성하지 않는다. 일부 구현예에서, 엔도뉴클레아제 도메인은 이중 가닥 파손보다 더 높은 빈도로 단일 가닥 파손을 형성하며, 예를 들어 적어도 90%, 95%, 96%, 97%, 98%, 또는 99%의 절단은 단일 가닥 파손이거나, 또는 10%, 5%, 4%, 3%, 2%, 또는 1% 미만의 절단은 이중 가닥 파손이다. 일부 구현예에서, 엔도뉴클레아제는 실질적으로 이중 가닥 파손을 형성하지 않는다. 일부 구현예에서, 엔도뉴클레아제는 검출가능한 수준의 이중 가닥 파손을 형성하지 않는다.

[0226] 일부 구현예에서, 엔도뉴클레아제 도메인은 제1 가닥의 표적 부위 DNA에 닉을 생성하는 닉카제 활성을 갖고; 예를 들어, 일부 구현예에서 엔도뉴클레아제 도메인은 라이팅 도메인에 의해 연장될 가닥 상에서 변경 부위 근처의 표적 부위의 게놈 DNA를 절단한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 제1 가닥의 표적 부위 DNA에 닉을 생성하고 제2 가닥의 표적 부위 DNA에 닉을 생성하지 않는 닉카제 활성을 갖는다. 예를 들어, 폴리펩티드가 닉카제 활성을 갖는 CRISPR-연관 엔도뉴클레아제 도메인을 포함하는 경우, 일부 구현예에서 상기 CRISPR-연관 엔도뉴클레아제 도메인은 PAM 부위를 함유하는 표적 부위 DNA 가닥에 닉을 생성한다(예를 들어, 그리고 PAM 부위를 함유하지 않는 표적 부위 DNA 가닥에 닉을 생성하지 않음). 추가의 예로서, 폴리펩티드가 닉카제 활성을 갖는 CRISPR-연관 엔도뉴클레아제 도메인을 포함하는 경우, 일부 구현예에서 상기 CRISPR-연관 엔도뉴클레아제 도메인은 PAM 부위를 함유하지 않는 표적 부위 DNA 가닥에 닉을 생성한다(예를 들어, 그리고 PAM 부위를 함유하는 표적 부위 DNA 가닥에 닉을 생성하지 않음).

[0227] 일부 다른 구현예에서, 엔도뉴클레아제 도메인은 제1 가닥 및 제2 가닥의 표적 부위 DNA에 닉을 생성하는 닉카제 활성을 갖는다. 이론에 구속되고자 함이 없이, 본 명세서에 기재된 폴리펩티드의 라이팅 도메인(예를 들어, RT 도메인)이 주형 핵산(예를 들어, 주형 RNA)의 이종 목적 서열로부터 중합(예를 들어, 역전사)된 후, 세포 DNA 복구 기구는 제1 DNA 가닥 상의 닉을 복구해야 한다. 표적 부위 DNA는 이제 제1 DNA 가닥에 대해 2개의 상이한 서열을 함유하며, 하나는 원래 게놈 DNA(예를 들어, 자유 5' 말단을 가짐)에 상응하고 두 번째는 이종 목적 서열(예를 들어, 자유 3' 말단을 가짐)로부터 중합된 것에 상응한다. 2개의 상이한 서열은 서로 평형을 이루는 것으로 여겨지며, 첫 번째 서열이 두 번째 가닥을 혼성화한 다음 다른 하나를 혼성화하고, 세포 DNA 복구 장치의 서열이 복구된 표적 부위로 통합하는 것은 확률적 공정일 수 있다. 이론에 구속되고자 함이 없이, 제2 가닥에 추가의 닉을 도입하는 것은 세포 DNA 복구 기구가 원래 게놈 서열보다 이종 목적 서열 기반 서열을 더 빈번하게 채용하도록 편향시킬 수 있는 것으로 생각된다(Anzalone et al. Nature 576:149-157 (2019)). 일부 구현예에서, 추가의 닉은 표적 부위 변형(예를 들어, 삽입, 결실, 또는 치환)의 적어도 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 또는 150개의 뉴클레오티드 5' 또는 3' 또는 제1 가닥의 닉에 위치한다.

[0228] 대안적으로 또는 추가적으로, 이론에 구속되고자 함이 없이, 제2 가닥에 대한 추가의 닉은 제2 가닥 합성을 촉진할 수 있는 것으로 여겨진다. 일부 구현예에서, 유전자 변형 시스템이 제1 가닥의 일부를 삽입하거나 치환하는 경우, 제2 가닥에서의 삽입/치환에 상응하는 새로운 서열의 합성이 필요하다.

[0229] 일부 구현예에서, 폴리펩티드는 엔도뉴클레아제 활성을 갖는 단일 도메인(예를 들어, 단일 엔도뉴클레아제 도메인)을 포함하고, 상기 도메인은 제1 가닥 및 제2 가닥 둘 모두에 닉을 생성한다. 예를 들어, 이러한 구현예에서 엔도뉴클레아제 도메인은 CRISPR-연관 엔도뉴클레아제 도메인일 수 있고, 주형 핵산(예를 들어, 주형 RNA)은 제1 가닥의 닉 생성을 지시하는 gRNA 스페이서 및 제2 가닥의 닉 생성을 지시하는 추가의 gRNA 스페이서를 포함한다. 일부 구현예에서, 폴리펩티드는 엔도뉴클레아제 활성을 갖는 복수의 도메인을 포함하고, 제1 엔도뉴클레아제 도메인은 제1 가닥에 닉을 생성하고 제2 엔도뉴클레아제 도메인은 제2 가닥에 닉을 생성한다(선택적으로, 제1 엔도뉴클레아제 도메인은 제2 가닥에 닉을 생성하지 않고(예를 들어, 닉을 생성할 수 없음), 제2 엔도뉴클레아제 도메인은 제1 가닥에 닉을 생성하지 않는다(예를 들어, 닉을 생성할 수 없음).

[0230] 일부 구현예에서, 엔도뉴클레아제 도메인은 제1 가닥 및 제2 가닥에 닉을 생성할 수 있다. 일부 구현예에서, 제1 및 제2 가닥 닉은 표적 부위 내 동일 위치에서 발생하지만 반대 가닥에서 발생한다. 일부 구현예에서, 제2 가닥 닉은 엇갈린 위치, 예를 들어 제1 닉으로부터 상류 또는 하류에서 발생한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 제2 가닥 닉이 제1 가닥 닉의 상류에 있는 경우 표적 부위 결실을 생성한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 제2 가닥 닉이 제1 가닥 닉의 하류에 있는 경우 표적 부위 복제를 생성한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 제1 및 제2 가닥 닉이 표적 부위의 동일 위치에서 발생하는 경우 복제 및/또는 결실을 생성하지 않는다. 일부 구현예에서, 엔도뉴클레아제 도메인은, 예를 들어 제1 또는 제2 가닥의 닉 생성을 촉진하는 단백질 적합(conformation) 또는 RNA-결합 상태에 따라 변경된 활성을 갖는다(예를 들어, 그 전체가 본 명세서에 참고로 포함된 문헌[Christensen et al. PNAS 2006]에 기재된 바와 같음).

[0231] 일부 구현예에서, 엔도뉴클레아제 도메인은 메가뉴클레아제, 또는 이의 기능성 단편을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 호밍(homing) 엔도뉴클레아제, 또는 이의 기능성 단편을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 LAGLIDADG, GIY-YIG, HNH, His-Cys Box, 또는 PD-(D/E) XK 패밀리로부터의 예를 들어 패밀리 명칭에 표시된 바와 같이 보존된 아미노산 모티프를 갖는 메가뉴클레아제, 또는 이의 기능성 단편 또는 변이체를 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 예를 들어 I-SmaMI(Uniprot F7WD42), I-SceI(Uniprot P03882), I-AniI(Uniprot P03880), I-DmoI(Uniprot P21505), I-CreI(Uniprot P05725), I-TevI(Uniprot P13299), I-OnuI(Uniprot Q4VWW5), 또는 I-BmoI(Uniprot Q9ANR6)로부터 선택되는 메가뉴클레아제, 또는 이의 단편을 포함한다. 일부 구현예에서, 메가뉴클레아제는 기능성 형태에서 자연적으로 단량체, 예를 들어 I-SceI, I-TevI이며, 또는 이량체, 예를 들어 I-CreI이다. 예를 들어, LAGLIDADG 모티프의 단일 카피를 갖는 LAGLIDADG 메가뉴클레아제는 일반적으로 동종이량체를 형성하는 한편, LAGLIDADG 모티프의 2개의 카피를 갖는 구성원은 일반적으로 단량체로 발견된다. 일부 구현예에서, 보통 이량체로 형성되는 메가뉴클레아제는 융합물로서 발현되며, 예를 들어 2개의 서브유닛은 단일 ORF로 발현되고, 선택적으로, 링커, 예를 들어 I-CreI 이량체 융합물에 의해 연결된다(Rodriguez-Fornes et al. Gene Therapy 2020; 그 전체가 본 명세서에 참고로 포함됨). 일부 구현예에서, 메가뉴클레아제 또는 이의 기능성 단편은 이중 가닥 DNA 분자, 예를 들어, I-SceI(K122I 및/또는 K223I)(Niu et al. J Mol Biol 2008), I-AniI(K227M)(McConnell Smith et al. PNAS 2009), I-DmoI(Q42A 및/또는 K120M)(Molina et al. J Biol Chem 2015) 중 하나의 가닥에 대한 닉카제 활

성을 선호하도록 변경된다. 일부 구현예에서, 메가뉴클레아제 또는 단일가닥 절단에 대한 이러한 선호도를 갖는 이의 기능성 단편은, 예를 들어 닉카제 활성을 갖는 엔도뉴클레아제 도메인으로 사용된다. 일부 구현예에서, 엔도뉴클레아제 도메인은 세이프 하버 부위, 예를 들어 I-CreI 표적화 SH6 부위를 자연적으로 표적화하거나 표적화하도록 조작된 메가뉴클레아제 또는 이의 기능성 단편을 포함한다(Rodriguez-Fornes et al., 상동). 일부 구현예에서, 엔도뉴클레아제 도메인은 서열 내성 촉매 도메인, 예를 들어 최소 모티프 CNNNG를 인식하는 I-TevI를 갖는 메가뉴클레아제 또는 이의 기능성 단편을 포함한다(Kleinstiver et al. PNAS 2012). 일부 구현예에서, 표적 서열 내성 촉매 도메인은, 예를 들어 활성을 지시하기 위해, 예를 들어 I-TevI를 (i) Tev-ZFE를 생성하기 위한 징크 핑거(Kleinstiver et al. PNAS 2012), (ii) MegaTevs를 생성하기 위한 기타 메가뉴클레아제(Wolfs et al. Nucleic Acids Res 2014), 및/또는 (iii) TevCas9를 생성하기 위한 Cas9(Wolfs et al. PNAS 2016)에 융합시킴으로써, DNA 결합 도메인에 융합된다.

[0232] 일부 구현예에서, 엔도뉴클레아제 도메인은 제한 효소, 예를 들어 IIS형 또는 IIP형 제한 효소를 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 IIS형 제한 효소, 예를 들어 FokI, 또는 이의 단편 또는 변이체를 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 IIP형 제한 효소, 예를 들어 PvuII, 또는 이의 단편 또는 변이체를 포함한다. 일부 구현예에서, 이량체성 제한 효소는, 그가 단일 사슬로서 기능하도록 하는 융합물, 예를 들어 FokI 이량체 융합물로서 발현된다(Minczuk et al. Nucleic Acids Res 36(12):3926-3938 (2008)).

[0233] 추가적인 엔도뉴클레아제 도메인의 사용은, 예를 들어 문헌[Guha and Edgell Int J Mol Sci 18(22):2565 (2017)]에 설명되어 있으며, 이는 그 전체가 본 명세서에 참고로 포함된다.

[0234] 일부 구현예에서, 유전자 변형 폴리펩티드는, 예를 들어 야생형 Cas 단백질에 비해 엔도뉴클레아제 도메인에 대한 변형을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 야생형 Cas 단백질의 아미노산 서열에 대한 추가, 결실, 대체, 또는 변형을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인은 관심 표적 핵산(예를 들어, DNA) 서열에 특이적으로 결합하고/하거나 이의 엔도뉴클레아제 절단을 유도하는 이종 기능성 도메인을 포함하도록 변형된다. 일부 구현예에서, 엔도뉴클레아제 도메인은 징크 핑거를 포함한다. 구현예에서, Cas 도메인을 포함하는 엔도뉴클레아제 도메인은, 예를 들어 본 명세서에 기재된 바와 같이 가이드 RNA(gRNA)와 회합된다. 일부 구현예에서, 엔도뉴클레아제 도메인은 특정 표적 핵산(예를 들어, DNA) 서열을 표적화하지 않는 기능성 도메인을 포함하도록 변형된다. 구현예에서, 엔도뉴클레아제 도메인은 Fok1 도메인을 포함한다.

[0235] 일부 구현예에서, 엔도뉴클레아제 도메인은 스크램블드 dsDNA를 이용한 경우보다 적어도 약 5배 또는 10배 더 높은 빈도로 시험관 내에서 표적 dsDNA와 회합된다. 일부 구현예에서, 엔도뉴클레아제 도메인은, 예를 들어 세포(예를 들어, HEK293T 세포)에서, 스크램블드 dsDNA를 이용한 경우보다 적어도 약 5배 또는 10배 더 높은 빈도로 시험관 내에서 표적 dsDNA와 회합된다. 일부 구현예에서, 엔도뉴클레아제 도메인과 표적 DNA 또는 스크램블드 DNA 간의 회합의 빈도는, 예를 들어 문헌[He and Pu (2010) *Curr. Protoc Mol Biol* Chapter 21(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이 ChIP-seq에 의해 측정된다.

[0236] 일부 구현예에서, 엔도뉴클레아제 도메인은, 예를 들어 비-표적 서열과 비교하여(예를 들어, 표적 세포의 게놈에서 임의의 다른 게놈 서열과 비교하여) 적어도 약 5배 또는 약 10배의 증가로 표적 서열에서 닉의 형성을 촉매할 수 있다. 일부 구현예에서, 닉 형성의 수준은, 예를 들어 문헌[Elacqua et al. (2019) *bioRxiv* doi.org/10.1101/867937(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, NickSeq을 사용하여 결정된다.

[0237] 일부 구현예에서, 엔도뉴클레아제 도메인은 시험관 내에서 DNA에 닉을 생성할 수 있다. 구현예에서, 닉은 노출된 염기를 생성한다. 구현예에서, 노출된 염기는, 예를 들어 문헌[Chaudhry and Weinfeld (1995) *Nucleic Acids Res* 23(19):3805-3809(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 뉴클레아제 민감도 분석을 사용하여 검출될 수 있다. 구현예에서, (예를 들어, 뉴클레아제 민감도 분석에 의해 검출된) 노출된 염기의 수준은 참조 엔도뉴클레아제 도메인과 비교하여 적어도 10%, 50% 또는 그 이상 증가된다. 일부 구현예에서, 참조 엔도뉴클레아제 도메인은 S. 피오게네스의 Cas9로부터의 엔도뉴클레아제 도메인이다.

[0238] 일부 구현예에서, 엔도뉴클레아제 도메인은 세포 내에서 DNA에 닉을 생성할 수 있다. 구현예에서, 엔도뉴클레아제 도메인은 HEK293T 세포 내에서 DNA에 닉을 생성할 수 있다. 구현예에서, Rad51 부재 하에서 복제를 진행하는 복구되지 않는 닉은 닉 부위에서 증가된 NHEJ 비를 초래하며, 이는, 예를 들어 문헌[Bothmer et al. (2017) *Nat Commun* 8:13905(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 예를 들어 Rad51 억제 분석을 사용하여 검출될 수 있다. 구현예에서, NHEJ 비는 0 내지 5% 넘게 증가된다. 구현예에서, NHEJ 비는, 예를 들어 Rad51 억제시, 20 내지 70%(예를 들어, 30% 내지 60% 또는 40 내지 50%)로 증가된다.

[0239] 일부 구현예에서, 엔도뉴클레아제 도메인은 절단 후 표적을 방출한다. 일부 구현예에서, 표적의 방출은, 예를 들어 문헌[Yourik at al. RNA 25(1):35-44 (2019)(그 전체가 본 명세서에 참고로 포함됨)]에 기재되고, 도 2에 나타낸 바와 같이, 효소에 의한 다수의 전환수(turnover)를 평가함으로써 간접적으로 표시된다. 일부 구현예에서, 엔도뉴클레아제 도메인의 k_{exp}는 그러한 방법에 의해 측정 시, 1 x 10^{-3} 내지 1 x 10^{-}5 분-1이다.

[0240] 일부 구현예에서, 엔도뉴클레아제 도메인은 시험관 내에서 약 1 x 10^{8} s^{-1} M^{-1} 초과의 촉매 효율(k_{cat}/K_m)을 갖는다. 구현예에서, 엔도뉴클레아제 도메인은 시험관 내에서 약 1 x 10^{5}, 1 x 10^{6}, 1 x 10^{7}, 또는 1 x 10^{8}, s^{-1} M^{-1} 초과의 촉매 효율을 갖는다. 구현예에서, 촉매 효율은 문헌[Chen et al. (2018) *Science* 360(6387):436-439(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이 결정된다. 일부 구현예에서, 엔도뉴클레아제 도메인은 세포 내에서 약 1 x 10^{8} s^{-1} M^{-1} 초과의 촉매 효율(k_{cat}/K_m)을 갖는다. 구현예에서, 엔도뉴클레아제 도메인은 세포 내에서 약 1 x 10^{5}, 1 x 10^{6}, 1 x 10^{7}, 또는 1 x 10^{8}, s^{-1} M^{-1} 초과의 촉매 효율을 갖는다.

[0241] *Cas 도메인을 포함하는 유전자 변형 폴리펩티드*

[0242] 일부 구현예에서, 본 명세서에 기재된 유전자 변형 폴리펩티드는 Cas 도메인을 포함한다. 일부 구현예에서, Cas 도메인은 gRNA 스페이서에 의해 지정된 표적 부위로 유전자 변형 폴리펩티드를 지시하여, 이에 의해 "시스"로 표적 핵산 서열을 변형시킬 수 있다. 일부 구현예에서, 유전자 변형 폴리펩티드는 Cas 도메인에 융합된다. 유전자 변형 폴리펩티드는 CRISPR/Cas 도메인(CRISPR-연관 단백질로도 지칭됨)을 포함한다. 일부 구현예에서, CRISPR/Cas 도메인은 클러스터링된 조절 간격의 짧은 회문 반복(CRISPR) 시스템에 관여하는 단백질, 예를 들어, Cas 단백질을 포함하고, 선택적으로 가이드 RNA, 예를 들어 단일 가이드 RNA(sgRNA)에 결합한다.

[0243] CRISPR 시스템은 원래 박테리아 및 고세균에서 발견된 적응 방어 시스템이다. CRISPR 시스템은 CRISPR-연관 또는 "Cas" 엔도뉴클레아제(예를 들어, Cas9 또는 Cpf1)로 지칭되는 RNA-유도 뉴클레아제를 사용하여 외래 DNA를 절단한다. 예를 들어, 통상적인 CRISPR-Cas 시스템에서, 엔도뉴클레아제는 단일- 또는 이중-가닥 DNA 서열을 표적화하는 서열-특이적, 비-코딩 "가이드 RNA"에 의해 표적 뉴클레오티드 서열(예를 들어 서열-편집할 게놈 내의 부위)로 지시된다. CRISPR 시스템의 3가지 클래스(I 내지 III)가 확인되었다. 클래스 II CRISPR 시스템은 (다수의 Cas 단백질이 아닌) 단일 Cas 엔도뉴클레아제를 사용한다. 하나의 클래스 II CRISPR 시스템은 Cas9와 같은 II형 Cas 엔도뉴클레아제, CRISPR RNA("crRNA"), 및 트랜스-활성화 crRNA("tracrRNA")를 포함한다. crRNA는 통상적으로 표적 DNA 서열에 상응하는 약 20-뉴클레오티드 RNA 서열인 "스페이서" 서열("프로토스페이서")을 함유한다. 야생형 시스템에서, 그리고 일부 조작된 시스템에서, crRNA는 또한 tracrRNA에 결합하여 RNase III에 의해 절단되는 부분적으로 이중 가닥인 구조를 형성하여 crRNA/tracrRNA 하이브리드 분자를 생성하는 영역을 함유한다. 그 다음 crRNA/tracrRNA 하이브리드는 Cas 엔도뉴클레아제가 표적 DNA 서열을 인식하고 절단하도록 지시한다. 표적 DNA 서열은 일반적으로 주어진 Cas 엔도뉴클레아제에 특이적이고 crRNA의 스페이서와 일치하는 표적 부위에서 절단 활성에 필요한 "프로토스페이서 인접 모티프"(protospacer adjacent motif; "PAM")에 인접한다. 다양한 원핵생물 종으로부터 확인된 CRISPR 엔도뉴클레아제는, 예를 들어 표 3의 예시적인 Cas 효소에 대해 열거된 바와 같이 고유한 PAM 서열 요건을 가지며; PAM 서열의 예로는 5'-NGG(스트렙토코커스 피오게네스(Streptococcus pyogenes)), 5'-NNAGAA(스트렙토코커스 써모필루스(Streptococcus thermophilus) CRISPR1), 5'-NGGNG(스트렙토코커스 써모필루스 CRISPR3), 및 5'-NNNGATT(네이세리아 메닝지디티스(Neisseria meningiditis))가 있다. 일부 엔도뉴클레아제, 예를 들어 Cas9 엔도뉴클레아제는 G-풍부 PAM 부위, 예를 들어 5'-NGG와 회합되며, PAM 부위로부터(5'로부터)의 위치 3인 뉴클레오티드 상류에서 표적 DNA의 평활 말단 절단을 수행한다. 또 다른 클래스 II CRISPR 시스템은, Cas9보다 작은 V 형 엔도뉴클레아제 Cpf1를 포함하며, 예에는 AsCpf1((애시드아미노코커스(Acidaminococcus) 종 유래) 및 LbCpf1((라크노스피라세(Lachnospiraceae) 종 유래)가 포함된다. Cpf1-연관 CRISPR 어레이는 tracrRNA의 필요 없이 성숙한 crRNA로 처리되며; 즉, Cpf1 시스템은 일부 구현예에서 표적 DNA 서열을 절단하기 위해 Cpf1 뉴클레아제 및 crRNA만을 포함한다. Cpf1 엔도뉴클레아제는 통상적으로 T-풍부 PAM 부위, 예를 들어 5'-TTN과 회합된다. Cpf1은 5'-CTA PAM 모티프도 또한 인식할 수 있다. Cpf1은 통상적으로 4- 또는 5-뉴클레오티드 5' 오버행(overhang)을 갖는 오프셋 또는 엇갈린 이중 가닥 파손을 도입함으로써, 예를 들어 코딩 가닥 상에서 PAM 부위로부터(이의 3'로부터) 18개 뉴클레오티드 하류, 및 상보적 가닥 상에서 PAM 부위로부터 23개 뉴클레오티드 하류에 위치한 5-뉴클레오티드 오프셋 또는 엇갈린 절단물을 갖는 표적 DNA를 절단하며; 그러한 오프셋 절단으로 인한 5-뉴클레오티드 오버행은 평활 말단 절단된 DNA에서의 삽입에 의한 것보다는, 상동성 재조합에 의한 DNA 삽입에 의해 더 정확한 유전체 편집을 가능하게 한

다. 예를 들어 문헌[Zetsche et al. (2015) Cell, 163:759 - 771]을 참조한다.

[0244] 다양한 CRISPR 연관(Cas) 유전자 또는 단백질이 본 개시내용에 의해 제공되는 기술에 사용될 수 있으며, Cas 단백질의 선택은 본 방법의 특정 조건에 따라 달라질 것이다. Cas 단백질의 특정 예는 Cas1, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9, Cas10, Cpf1, C2C1, 또는 C2C3을 포함하는 클래스 II 시스템을 포함한다. 일부 구현예에서, Cas 단백질, 예를 들어 Cas9 단백질은 다양한 원핵생물 종들 중 임의의 것으로부터 유래될 수 있다. 일부 구현예에서, 특정 Cas 단백질, 예를 들어 특정 Cas9 단백질은 특정 프로토스페이서-인접 모티프(PAM) 서열을 인식하도록 선택된다. 일부 구현예에서, DNA-결합 도메인 또는 엔도뉴클레아제 도메인은 서열 표적화 폴리펩티드, 예컨대 Cas 단백질, 예를 들어 Cas9를 포함한다. 특정 구현예에서, Cas 단백질, 예를 들어 Cas9 단백질은 박테리아 또는 고세균으로부터 수득되거나, 또는 알려진 방법을 이용하여 합성될 수 있다. 특정 구현예에서, Cas 단백질은 그람 양성 박테리아 또는 그람 음성 박테리아에서 유래할 수 있다. 특정 구현예에서, Cas 단백질은 스트렙토코커스(예를 들어, S. 피오게네스, 또는 S. 써모필루스), 프란시셀라(Francisella)(예를 들어, F. 노비시다(novicida)), 스태필로코커스(예를 들어, S. 아우레우스(aureus)), 애시드아미노코커스(예를 들어, 애시드아미노코커스 BV3L6), 나이세리아(예를 들어, N. 메닝지티디스), 크립토코커스, 코리네박테리움, 헤모필루스(Haemophilus), 유박테리움(Eubacterium), 파스퇴렐라(Pasteurella), 프레보텔라(Prevotella), 베일로넬라(Veillonella), 또는 마리노박터(Marinobacter)에서 유래할 수 있다.

[0245] 일부 구현예에서, 유전자 변형 폴리펩티드는 표 3 또는 4에 열거된 바와 같은 Cas 도메인, 또는 이에 대하여 적어도 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 동일성을 갖는 서열을 포함할 수 있다.

[0246] **표 3. CRISPR/Cas 단백질, 종, 및 돌연변이**

표 3

[0247]

명칭	*효소*	*종*	*AA의 수*	*PAM*	*PAM 인식을 변경시키기 위한 돌연변이*	*촉매적으로 사멸시키기 위한 돌연변이*
FnCas9	Cas9	프란시셀라 노비시다	1629	5'-NGG-3'	Wt	D11A/H969A/N995A
FnCas9 RHA	Cas9	프란시셀라 노비시다	1629	5'-YG-3'	E1369R/E1449H/R1556A	D11A/H969A/N995A
SaCas9	Cas9	스태필로코커스 아우레우스	1053	5'-NNGRRT-3'	Wt	D10A/H557A
SaCas9 KKH	Cas9	스태필로코커스 아우레우스	1053	5'-NNNRRT-3'	E782K/N968K/R1015H	D10A/H557A
SpCas9	Cas9	스트렙토코커스 피오게네스	1368	5'-NGG-3'	Wt	D10A/D839A/H840A/N863A
SpCas9 VQR	Cas9	스트렙토코커스 피오게네스	1368	5'-NGA-3'	D1135V/R1335Q/T1337R	D10A/D839A/H840A/N863A
AsCpf1 RR	Cpf1	애시드아미노코커스 종 *BV3L6*	1307	5'-TYCV-3'	S542R/K607R	E993A
AsCpf1 RVR	Cpf1	애시드아미노코커스 종 *BV3L6*	1307	5'-TATV-3'	S542R/K548V/N552R	E993A
FnCpf1	Cpf1	프란시셀라 노비시다	1300	5'-NTTN-3'	Wt	D917A/E1006A/D1255A
NmCas9	Cas9	네이세리아 메닝지티디스	1082	5'-NNNGATT-3'	Wt	D16A/D587A/H588A/N611A

[0248] **표 4 CRISPR/Cas 단백질의 아미노산 서열, 종, 및 돌연변이**

표 4

[0249]

변이체	모 숙주(들)	단백질 서열	닉카제 (HNH)	닉카제 (HNH)	닉카제 (RuvC)
Nme2Cas9	네이세리아 메닝지티디스	MAAFKPNPINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAE VPKTGDSLAMARRLARSVRRLTRRRAHRLLRARRLLKREGVLQAADFDEN GLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGET ADKELGALLKGVANNAHALQTGDFRTPAELALNKFEKESGHIRNQRGDYS HTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDA VQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDT ERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEM KAYHAISRALEKEGLKDKKSPLNLSSELQDEIGTAFSLFKTDEDITGRLK DRVQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYG DHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPAR IHIETAREVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKS KDILKLRLYEQQHGKCLYSGKEINLVRLNEKGYVEIDHALPFSRTWDDSF NNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQ RILLQKFDEDGFKECNLNDTRYVNRFLCQFVADHILLTGKGKRRVFASNG QITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEM NAFDGKTIDKETGKVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEA DTPEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGAHKDTLRSAK RFVKHNEKISVKRVWLTEIKLADLENMVNYKNGREIELYEALKARLEAYG GNAKQAFDPKDNPFYKKGGQLVKAVRVEKTQESGVLLNKKNAYTIADNGD MVRVDVFCKVDKKGKNQYFIVPIYAWQVAENILPDIDCKGYRIDDSYTFC FSLHKYDLIAFQKDEKSKVEFAYYINCDSSNGRFYLAWHDKGSKEQQFRI STQNLVLIQKYQVNELGKEIRPCRLKKRPPVR	N611A	H588A	D16A
PpnCas9	파스퇴렐라 뉴모트로피카	MQNNPLNYILGLDLGIASIGWAVVEIDEESSPIRLIDVGVRTFERAEVAK TGESLALSRRLARSSRRLIKRRAERLKKAKRLLKAEKILHSIDEKLPINV WQLRVKGLKEKLERQEWAAVLLHLSKHRGYLSQRKNEGKSDNKELGALLS GIASNHQMLQSSEYRTPAEIAVKKFQVEEGHIRNQRGSYTHTFSRLDLLA EMELLFQRQAELGNSYTSTTLLENLTALLMWQKPALAGDAILKMLGKCTF EPSEYKAAKNSYSAERFVWLTKLNNLRILENGTERALNDNERFALLEQPY EKSKLTYAQVRAMLALSDNAIFKGVRYLGEDKKTVESKTTLIEMKFYHQI RKTLGSAELKKEWNELKGNSDLLDEIGTAFSLYKTDDDICRYLEGKLPER VLNALLENLNFDKFIQLSLKALHQILPLMLQGQRYDEAVSAIYGDHYGKK STETTRLLPTIPADEIRNPVVLRTLTQARKVINAVVRLYGSPARIHIETA REVGKSYQDRKKLEKQQEDNRKQRESAVKKFKEMFPHFVGEPKGKDILKM RLYELQQAKCLYSGKSLELHRLLEKGYVEVDHALPFSRTWDDSFNNKVLV LANENQNKGNLTPYEWLDGKNNSERWQHFVVRVQTSGFSYAKKQRILNHK LDEKGFIERNLNDTRYVARFLCNFIADNMLLVGKGKRNVFASNGQITALL RHRWGLQKVREQNDRHHALDAVVVACSTVAMQQKITRFVRYNEGNVFSGE RIDRETGEIIPLHFPSPWAFFKENVEIRIFSENPKLELENRLPDYPQYNH EWVQPLFVSRMPTRKMTGQGHMETVKSAKRLNEGLSVLKVPLTQLKLSDL ERMVNRDREIALYESLKARLEQFGNDPAKAFAEPFYKKGGALVKAVRLEQ TQKSGVLVRDGNGVADNASMVRVDVFTKGGKYFLVPIYTWQVAKGILPNR AATQGKDENDWDIMDEMATFQFSLCQNDLIKLVTKKKTIFGYFNGLNRAT SNINIKEHDLDKSKGKLGIYLEVGVKLAISLEKYQVDELGKNIRPCRPTK RQHVR	N605A	H582A	D13A

SauCas9	스태필로코커스 아우레우스	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
SauCas9-KKH	스태필로코커스 아우레우스	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
SauriCas9	스태필로코커스 아우리쿨라리스	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNRRSKRGARRLKRRRIHRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNISVSMGDEEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQYIDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYSADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKSGKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSIKVINAVINRFGLPEDIIIELAREKNSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIEKIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQSENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEERDINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTHKALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRQLINDTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLMTILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFRFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYEAEKQKKKIKESDLFVGSFYYNDLIMYEDELFRVIGVNSDINNLVELNMVDITYKDFCEVNNVTGEKRIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL	N588A	H565A	D15A

SauriCas9-KKH	스태필로코커스 아우리쿨라리스	MQENQQKQNYILGLDIGITSVGYGLIDSKTREVIDAGVRLFPEADSENNSNRRSKRGARRLKRRRIHRLNRVKDLLADYQMIDLNNVPKSTDPYTIRVKGLREPLTKEEFAIALLHIAKRRGLHNISVSMGDEEQDNELSTKQQLQKNAQQLQDKYVCELQLERLTNINKVRGEKNRFKTEDFVKEVKQLCETQRQYHNIDDQFIQQYIDLVSTRREYFEGPGNGSPYGWDGDLLKWYEKLMGRCTYFPEELRSVKYAYSADLFNALNDLNNLVVTRDDNPKLEYYEKYHIIENVFKQKKNPTLKQIAKEIGVQDYDIRGYRITKSGKPQFTSFKLYHDLKNIFEQAKYLEDVEMLDEIAKILTIYQDEISIKKALDQLPELLTESEKSQIAQLTGYTGTHRLSLKCIHIVIDELWESPENQMEIFTRLNLKPKKVEMSEIDSIPTTLVDEFILSPVVKRAFIQSIKVINAVINRFGLPEDIIIELARENKSKDRRKFINKLQKQNEATRKKIEQLLAKYGNTNAKYMIEKIKLHDMQEGKCLYSLEAIPLEDLLSNPTHYEVDHIIPRSVSFDNSLNNKVLVKQSENSKKGNRTPYQYLSSNESKISYNQFKQHILNLSKAKDRISKKKRDMLLEERDINKFEVQKEFINRNLVDTRYATRELSNLLKTYFSTHDYAVKVKTINGGFTNHLRKVWDFKKHRNHGYKHHAEDALVIANADFLFKTHKALRRTDKILEQPGLEVNDTTVKVDTEEKYQELFETPKQVKNIKQFRDFKYSHRVDKKPNRKLINDTLYSTREIDGETYVVQTLKDLYAKDNEKVKKLFTERPQKILMYQHDPKTFEKLMTILNQYAEAKNPLAAYYEDKGEYVTKYAKKGNGPAIHKIKYIDKKLGSYLDVSNKYPETQNKLVKLSLKSFRFDIYKCEQGYKMVSIGYLDVLKKDNYYYIPKDKYEAEKQKKKIKESDLFVGSFYKNDLIMYEDELFRVIGVNSDINNLVELNMVDITYKDFCEVNNVTGEKHIKKTIGKRVVLIEKYTTDILGNLYKTPLPKKPQLIFKRGEL	N588A	H565A	D15A
ScaCas9-Sc++	스트렙토코커스 카니스	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGKLATEEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A	H849A	D10A

SpyCas9	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-NG	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A

SpyCas9-SpRY	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAERTRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTRLGAPRAFKYFDTTIDPKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
St1Cas9	스트렙토코커스 써모필루스	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTMPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A	H599A	D9A

BlatCas9	브레비바실러스 라테로스포루스	MAYTMGIDVGIASCGWAIVDLERQRIIDIGVRTFEKAENPKNGEALAVPR REARSSRRRLRRKKHRIERLKHMFVRNGLAVDIQHLEQTLRSQNEIDVWQ LRVDGLDRMLTQKEWLRVLIHLAQRRGFQSNRKTDGSSEDGQVLVNVTEN DRLMEEKDYRTVAEMMVKDEKFSDHKRNKNGNYHGVVSRSSLLVEIHTLF ETQRQHHNSLASKDFELEYVNIWSAQRPVATKDQIEKMIGTCTFLPKEKR APKASWHFQYFMLLQTINHIRITNVQGTRSLNKEEIEQVVNMALTKSKVS YHDTRKILDLSEEYQFVGLDYGKEDEKKKVESKETIIKLDDYHKLNKIFN EVELAKGETWEADDYDTVAYALTFFKDDEDIRDYLQNKYKDSKNRLVKNL ANKEYTNELIGKVSTLSFRKVGHLSLKALRKIIPFLEQGMTYDKACQAAG FDFQGISKKKRSVVLPVIDQISNPVVNRALTQTRKVINALIKKYGSPETI HIETARELSKTFDERKNITKDYKENRDKNEHAKKHLSELGIINPTGLDIV KYKLWCEQQGRCMYSNQPISFERLKESGYTEVDHIIPYSRSMNDSYNNRV LVMTRENREKGNQTPFEYMGNDTQRWYEFEQRVTTNPQIKKEKRQNLLLK GFTNRRELEMLERNLNDTRYITKYLSHFISTNLEFSPSDKKKKVVNTSGR ITSHLRSRWGLEKNRGQNDLHHAMDAIVIAVTSDSFIQQVTNYYKRKERR ELNGDDKFPLPWKFFREEVIARLSPNPKEQIEALPNHFYSEDELADLQPI FVSRMPKRSITGEAHQAQFRRVVGKTKEGKNITAKKTALVDISYDKNGDF NMYGRETDPATYEAIKERYLEFGGNVKKAFSTDLHKPKKDGTKGPLIKSV RIMENKTLVHPVNKGKGVVYNSSIVRTDVFQRKEKYYLLPVYVTDVTKGK LPNKVIVAKKGYHDWIEVDDSFTFLFSLYPNDLIFIRQNPKKKISLKKRI ESHSISDSKEVQEIHAYYKGVDSSTAAIEFIIHDGSYYAKGVGVQNLDCF EKYQVDILGNYFKVKGEKRLELETSDSNHKGKDVNSIKSTSR	N607A	H584A	D8A
cCas9-v16	스태필로코커스 아우레우스	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK RGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKL SEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYV AELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDT YIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYA YNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIA KEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQ IAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAI NLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVV KRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQ TNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP FNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKIS YETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTR YATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKH HAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEY KEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTL IVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDE KNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNS RNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEA KKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNSDKNNLIEVNMIDIT YREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQII KKG	N580A	H557A	D10A
cCas9-v17	스태필로코커스 아우레우스	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSK RGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKL SEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYV AELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDT YIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYA YNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIA KEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQ IAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAI NLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVV KRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQ TNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNP FNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKIS YETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTR YATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKH HAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEY KEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTL IVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDE KNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNS RNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEA KKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNSTRNIVELNMIDIT YREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQII KKG	N580A	H557A	D10A

cCas9-v21	스태필로코커스 아우레우스	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNSDDRNIIELNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
cCas9-v42	스태필로코커스 아우레우스	MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGARRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLVITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLNSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIELAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKGNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRKLINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGDEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRNKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKKISNQAEFIASFYKNDLIKINGELYRVIGVNNNRLNKIELNMIDITYREYLENMNDKRPPHIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG	N580A	H557A	D10A
CdiCas9	코리네박테리움 디프테리아	MKYHVGIDVGTFSVGLAAIEVDDAGMPIKTLSLVSHIHDSGLDPDEIKSAVTRLASSGIARRTRRLYRRKRRRLQQLDKFIQRQGWPVIELEDYSDPLYPWKVRAELAASYIADEKERGEKLSVALRHIARHRGWRNPYAKVSSLYLPDGPSDAFKAIREEIKRASGQPVPETATVGQMVTLCELGTLKLRGEGGVLSARLQQSDYAREIQEICRMQEIGQELYRKIIDVVFAAESPKGSASSRVGKDPLQPGKNRALKASDAFQRYRIAALIGNLRVRVDGEKRILSVEEKNLVFDHLVNLTPKKEPEWVTIAEILGIDRGQLIGTATMTDDGERAGARPPTHDTNRSIVNSRIAPLVDWWKTASALEQHAMVKALSNAEVDDFDSPEGAKVQAFFADLDDDVHAKLDSLHLPVGRAAYSEDTLVRLTRRMLSDGVDLYTARLQEFGIEPSWTPPTPRIGEPVGNPAVDRVLKTVSRWLESATKTWGAPERVIIEHVREGFVTEKRAREMDGDMRRRAARNAKLFQEMQEKLNVQGKPSRADLWRYQSVQRQNCQCAYCGSPITFSNSEMDHIVPRAGQGSTNTRENLVAVCHRCNQSKGNTPFAIWAKNTSIEGVSVKEAVERTRHWVTDTGMRSTDFKKFTKAVVERFQRATMDEEIDARSMESVAWMANELRSRVAQHFASHGTTVRVYRGSLTAEARRASGISGKLKFFDGVGKSRLDRRHHAIDAAVIAFTSDYVAETLAVRSNLKQSQAHRQEAPQWREFTGKDAEHRAAWRVWCQKMEKLSALLTEDLRDDRVVVMSNVRLRLGNGSAHKETIGKLSKVKLSSQLSVSDIDKASSEALWCALTREPGFDPKEGLPANPERHIRVNGTHVYAGDNIGLFPVSAGSIALRGGYAELGSSFHHARVYKITSGKKPAFAMLRVYTIDLLPYRNQDLFSVELKPQTMSMRQAEKKLRDALATGNAEYLGWLVVDDELVVDTSKIATDQVKAVEAELGTIRRWRVDGFFSPSKLRLRPLQMSKEGIKKESAPELSKIIDRPGWLPAVNKLFSDGNVTVVRRDSLGRVRLESTAHLPVTWKVQ	N597A	H573A	D8A

CjeCas9	캄필로박터 제주니	MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRL ARSARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLIS PYELRFRALNELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIK QNEEKLANYQSVGEYLYKEYFQKFKENSKEFTNVRNKKESYERCIAQSFL KDELKLIFKKQREFGFSFSKKFEEEVLSVAFYKRALKDFSHLVGNCSFFT DEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGILYTKDDLNALLNEVLK NGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKALGEHNLSQD DLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKA LKLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNP VVLRAIKEYRKVLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNE NYKAKKDAELECEKLGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQ DEKMLEIDHIYPYSRSFDDSYMNKVLVFTKQNQEKLNQTPFEAFGNDSAK WQKIEVLAKNLPTKKQKRILDKNYKDKEQKNFKDRNLNDTRYIARLVLNY TKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGMLTSALRHTWGFSAKD RNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAKKISELDYK NKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQ SYGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIY TMDFALKVLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKD MQEPEFVYYNAFTSSTVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAK SIGIQNLKVFEKYIVSALGEVTKAEFRQREDFKK	N582A	H559A	D8A
GeoCas9	제오바실러스 스테아로써모필루스	MRYKIGLDIGITSVGWAVMNLDIPRIEDLGVRIFDRAENPQTGESLALPR RLARSARRRLRRRKHRLERIRRLVIREGILTKEELDKLFEEKHEIDVWQL RVEALDRKLNNDELARVLLHLAKRRGFKSNRKSERSNKENSTMLKHIEEN RAILSSYRTVGEMIVKDPKFALHKRNKGENYTNTIARDDLEREIRLIFSK QREFGNMSCTEEFENEYITIWASQRPVASKDDIEKKVGFCTFEPKEKRAP KATYTFQSFIAWEHINKLRLISPSGARGLTDEERRLLYEQAFQKNKITYH DIRTLLHLPDDTYFKGIVYDRGESRKQNENIRFLELDAYHQIRKAVDKVY GKGKSSSFLPIDFDTFGYALTLFKDDADIHSYLRNEYEQNGKRMPNLANK VYDNELIEELLNLSFTKFGHLSLKALRSILPYMEQGEVYSSACERAGYTF TGPKKKQKTMLLPNIPPIANPVVMRALTQARKVVNAIIKKYGSPVSIHIE LARDLSQTFDERRKTKKEQDENRKKNETAIRQLMEYGLTLNPTGHDIVKF KLWSEQNGRCAYSLQPIEIERLLEPGYVEVDHVIPYSRSLDDSYTNKVLV LTRENREKGNRIPAEYLGVGTERWQQFETFVLTNKQFSKKKRDRLLRLHY DENEETEFKNRNLNDTRYISRFFANFIREHLKFAESDDKQKVYTVNGRVT AHLRSRWEFNKNREESDLHHAVDAVIVACTTPSDIAKVTAFYQRREQNKE LAKKTEPHFPQPWPHFADELRARLSKHPKESIKALNLGNYDDQKLESLQP VFVSRMPKRSVTGAAHQETLRRYVGIDERSGKIQTVVKTKLSEIKLDASG HFPMYGKESDPRTYEAIRQRLLEHNNDPKKAFQEPLYKPKKNGEPGPVIR TVKIIDTKNQVIPLNDGKTVAYNSNIVRVDVFEKDGKYYCVPVYTMDIMK GILPNKAIEPNKPYSEWKEMTEDYTFRFSLYPNDLIRIELPREKTVKTAA GEEINVKDVFVYYKTIDSANGGLELISHDHRFSLRGVGSRTLKRFEKYQV DVLGNIYKVRGEKRVGLASSAHSKPGKTIRPLQSTRD	N605A	H582A	D8A

iSpyMacCas9	스트렙토코커스 종	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRKLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLKREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEIQTVGQNGGLFDDNPKSPLEVTPSKLVPLKKELNPKKYGGYQKPTTAYPVLLITDTKQLIPISVMNKKQFEQNPVKFLRDRGYQQVGKNDFIKLPKYTLVDIGDGIKRLWASSKEIHKGNQLVVSKKSQILLYHAHHLDSDLSNDYLQNHNQQFDVLFNEIISFSKKCKLGKEHIQKIENVYSNKKNSASIEELAESFIKLLGFTQLGATSPFNFLGVKLNQKQYKGKKDYILPCTEGTLIRQSITGLYETRVDLSKIGEDSGGSGGSKRTADGSEFES	N863A	H840A	D10A
NmeCas9	네이세리아 메닝지티디스	MAAFKPNSINYILGLDIGIASVGWAMVEIDEEENPIRLIDLGVRVFERAEVPKTGDSLAMARRLARSVRRLTRRRAHRLLRTRRLLKREGVLQAANFDENGLIKSLPNTPWQLRAAALDRKLTPLEWSAVLLHLIKHRGYLSQRKNEGETADKELGALLKGVAGNAHALQTGDFRTPAELALNKFEKESGHIRNQRSDYSHTFSRKDLQAELILLFEKQKEFGNPHVSGGLKEGIETLLMTQRPALSGDAVQKMLGHCTFEPAEPKAAKNTYTAERFIWLTKLNNLRILEQGSERPLTDTERATLMDEPYRKSKLTYAQARKLLGLEDTAFFKGLRYGKDNAEASTLMEMKAYHAISRALEKEGLKDKKSPLNLSPELQDEIGTAFSLFKTDEDITGRLKDRIQPEILEALLKHISFDKFVQISLKALRRIVPLMEQGKRYDEACAEIYGDHYGKKNTEEKIYLPPIPADEIRNPVVLRALSQARKVINGVVRRYGSPARIHIETARDVGKSFKDRKEIEKRQEENRKDREKAAAKFREYFPNFVGEPKSKDILKLRLYEQQHGKCLYSGKEINLGRLNEKGYVEIDHALPFSRTWDDSFNNKVLVLGSENQNKGNQTPYEYFNGKDNSREWQEFKARVETSRFPRSKKQRILLQKFDEDGFKERNLNDTRYVNRFLCQFVADRMRLTGKGKKRVFASNGQITNLLRGFWGLRKVRAENDRHHALDAVVVACSTVAMQQKITRFVRYKEMNAFDGKTIDKETGEVLHQKTHFPQPWEFFAQEVMIRVFGKPDGKPEFEEADTLEKLRTLLAEKLSSRPEAVHEYVTPLFVSRAPNRKMSGQGHMETVKSAKRLDEGVSVLRVPLTQLKLKDLEKMVNREREPKLYEALKARLEAHKDDPAKAFAEPFYKYDKAGNRTQQVKAVRVEQVQKTGVWVRNHNGIADNATMVRVDVFEKGDKYYLVPIYSWQVAKGILPDRAVVQGKDEEDWQLIDDSFNFKFSLHPNDLVEVITKKARMFGYFASCHRGTGNININIRIHDLDHKIGKNGILEGIGVKTALSFQKYQIDELGKEIRPCRLKKRPPVR	N611A	H588A	D16A

ScaCas9	스트렙토코커스 카니스	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVDGIKHRKRTTKLATQEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNRNFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASATELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A	H849A	D10A
ScaCas9-HiFi-Sc++	스트렙토코커스 카니스	MEKKYSIGLDIGTNSVGWAVITDDYKVPSKKFKVLGNTNRKSIKKNLMGALLFDSGETAEATRLKRTARRRYTRRKNRIRYLQEIFANEMAKLDDSFFQRLEESFLVEEDKKNERHPIFGNLADEVAYHRNYPTIYHLRKKLADSPEKADLRLIYLALAHIIKFRGHFLIEGKLNAENSDVAKLFYQLIQTYNQLFEESPLDEIEVDAKGILSARLSKSKRLEKLIAVFPNEKKNGLFGNIIALALGLTPNFKSNFDLTEDAKLQLSKDTYDDDLDELLGQIGDQYADLFSAAKNLSDAILLSDILRSNSEVTKAPLSASMVKRYDEHHQDLALLKTLVRQQFPEKYAEIFKDDTKNGYAGYVGADKKLRKRSGKLATEEEFYKFIKPILEKMDGAEELLAKLNRDDLLRKQRTFDNGSIPHQIHLKELHAILRRQEEFYPFLKENREKIEKILTFRIPYYVGPLARGNSRFAWLTRKSEEAITPWNFEEVVDKGASAQSFIERMTNFDEQLPNKKVLPKHSLLYEYFTVYNELTKVKYVTERMRKPEFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEIIGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRHYTGWGRLSRKMINGIRDKQSGKTILDFLKSDGFSNANFMQLIHDDSLTFKEEIEKAQVSGQGDSLHEQIADLAGSPAIKKGILQTVKIVDELVKVMGHKPENIVIEMARENQTTTKGLQQSRERKKRIEEGIKELESQILKENPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFIKDDSIDNKVLTRSVENRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSEADKAGFIKRQLVETRQITKHVARILDSRMNTKRDKNDKPIREVKVITLKSKLVSDFRKDFQLYKVRDINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKRFFYSNIMNFFKTEVKLANGEIRKRPLIETNGETGEVVWNKEKDFATVRKVLAMPQVNIVKKTEVQTGGFSKESILSKRESAKLIPRKKGWDTRKYGGFGSPTVAYSILVVAKVEKGKAKKLKSVKVLVGITIMEKGSYEKDPIGFLEAKGYKDIKKELIFKLPKYSLFELENGRRRMLASAKELQKANELVLPQHLVRLLYYTQNISATTGSNNLGYIEQHREEFKEIFEKIIDFSEKYILKNKVNSNLKSSFDEQFAVSDSILLSNSFVSLLKYTSFGASGGFTFLDLDVKQGRLRYQTVTEVLDATLIYQSITGLYETRTDLSQLGGD	N872A	H849A	D10A

SpyCas9-3var-NRRH	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTAAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLHKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGVPAAFKYFDTTIDKKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-3var-NRTH	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIGFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASASVLHKGNELALPSKYVNFLYLASHYEKLKGSSEDNKQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGASAAFKYFDTTIGRKLYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A

SpyCas9-3var-NRCH	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMVKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQGDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRLRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGGHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKGNSDKLIARKKDWDPKKYGGFNSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTINRKQYNTTKEVLDATLIRQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-HF1	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A

SpyCas9-QQR1	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADAQLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTFKQKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-SpG	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFLWPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAKQLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A

SpyCas9-VQR	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKQYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-VRER	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKEYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A

SpyCas9-xCas	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGVLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A
SpyCas9-xCas-NG	스트렙토코커스 피오게네스	MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDTKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKLYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGIIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEKVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGDQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFIQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESIRPKRNSDKLIARKKDWDPKKYGGFVSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASARFLQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPRAFKYFDTTIDRKVYRSTKEVLDATLIHQSITGLYETRIDLSQLGGD	N863A	H840A	D10A

St1Cas9-CNRZ1066	스트렙토코커스 써모필루스	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKATGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPKQKHYVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A	H599A	D9A
St1Cas9-LMG1831	스트렙토코커스 써모필루스	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEEQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKKDETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQMNEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLLGNPIDITPENSKNKVVLQSLKPWRTDVYFNKATGKYEILGLKYADLQFEKGTGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRTLPKQKHYVELKPYDKQKFEGGEALIKVLGNVANGGQCIKGLAKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A	H599A	D9A
St1Cas9-MTH17CL396	스트렙토코커스 써모필루스	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHYVELKPYNRQKFEGSEYLIKSLGTVAKGGQCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A	H599A	D9A

St1Cas9-TH1477	스트렙토코커스 써모필루스	MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTNRQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEELFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERYQTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINRYLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAKASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAKLLSCDVADIKGYRIDKSGKAEIHTFEEAYRKMKTLETLDIEQMDRETLDKLAYVLTLNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMMELIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIKIVNAAIKEYGDFDNIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKAELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSITFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNKKKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVRGQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQLLDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATIYATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQTFEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYYDSKLGNHIDITPKDSNNKVVLQSLKPWRTDVYFNKNTGKYEILGLKYSDMQFEKGTGKYSISKEQYENIKVREGVDENSEFKFTLYKNDLLLLKDSENGEQILLRFTSRNDTSKHYVELKPYNRQKFEGSEYLIKSLGTVVKGGRCIKGLGKSNISIYKVRTDVLGNQHIIKNEGDKPKLDF	N622A	H599A	D9A
sRGN3.1	스태필로코커스 종	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGSRRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIALLHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRITKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSKKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFEVQKEFINRNLVDTRYATRELTNYLKAYFSANNMNVKVKTINGSFTDYLRKVWKFKKERNHGYKHHAEDALIIANADFLFKENKKLKAVNSVLEKPEIETKQLDIQVDSEDNYSEMFIIPKQVQDIKDFRNFKYSHRVDKKPNRQLINDTLYSTRKKDNSTYIVQTIKDIYAKDNTTLKKQFDKSPEKFLMYQHDPRTFEKLEVIMKQYANEKNPLAKYHEETGEYLTKYSKKNNGPIVKSLKYIGNKLGSHLDVTHQFKSSTKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELEKKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	N585A	H562A	D10A
sRGN3.3	스태필로코커스 종	MNQKFILGLDIGITSVGYGLIDYETKNIIDAGVRLFPEANVENNEGRRSKRGSRRLKRRRIHRLERVKLLLTEYDLINKEQIPTSNNPYQIRVKGLSEILSKDELAIALLHLAKRRGIHNVDVAADKEETASDSLSTKDQINKNAKFLESRYVCELQKERLENEGHVRGVENRFLTKDIVREAKKIIDTQMQYYPEIDETFKEKYISLVETRREYFEGPGQGSPFGWNGDLKKWYEMLMGHCTYFPQELRSVKYAYSADLFNALNDLNNLIIQRDNSEKLEYHEKYHIIENVFKQKKKPTLKQIAKEIGVNPEDIKGYRITKSGTPEFTSFKLFHDLKKVVKDHAILDDIDLLNQIAEILTIYQDKDSIVAELGQLEYLMSEADKQSISELTGYTGTHSLSLKCMNMIIDELWHSSMNQMEVFTYLNMRPKKYELKGYQRIPTDMIDDAILSPVVKRTFIQSINVINKVIEKYGIPEDIIIELARENNSDDRKKFINNLQKKNEATRKRINEIIGQTGNQNAKRIVEKIRLHDQQEGKCLYSLESIPLEDLLNNPNHYEVDHIIPRSVSFDNSYHNKVLVKQSENSKKSNLTPYQYFNSGKSKLSYNQFKQHILNLSKSQDRISKKKKEYLLEERDINKFEVQKEFINRNLVDTRYATRELTSYLKAYFSANNMDVKVKTINGSFTNHLRKVWRFDKYRNHGYKHHAEDALIIANADFLFKENKKLQNTNKILEKPTIENNTKKVTVEKEEDYNNVFETPKLVEDIKQYRDYKFSHRVDKKPNRQLINDTLYSTRMKDEHDYIVQTITDIYGKDNTNLKKQFNKNPEKFLMYQNDPKTFEKLSIIMKQYSDEKNPLAKYYEETGEYLTKYSKKNNGPIVKKIKLLGNKVGNHLDVTNKYENSTKKLVKLSIKNYRFDVYLTEKGYKFVTIAYLNVFKKDNYYYIPKDKYQELEKKKKIKDTDQFIASFYKNDLIKLNGDLYKIIGVNSDDRNIIELDYYDIKYKDYCEINNIKGEPRIKKTIGKKTESIEKFTTDVLGNLYLHSTEKAPQLIFKRGL	N585A	H562A	D10A

[0250] 일부 구현예에서, Cas 단백질은 Cas 단백질이 결합 및/또는 기능하기 위한 표적 DNA 서열에 존재하거나 이에 인접하기 위해 프로토스페이서 인접 모티프(PAM)를 필요로 한다. 일부 구현예에서, PAM은 5'에서 3'로 NGG, YG, NNGRRT, NNNRRT, NGA, TYCV, TATV, NTTN, 또는 NNNGATT이거나, 이를 포함하며, 여기서 N은 임의의 뉴클레오티드를 나타내고, Y는 C 또는 T를 나타내고, R은 A 또는 G를 나타내고, V는 A 또는 C 또는 G를 나타낸다. 일부 구현예에서, Cas 단백질은 표 3 또는 4에 열거된 단백질이다. 일부 구현예에서, Cas 단백질은 그의 PAM을 변경하는 하나 이상의 돌연변이를 포함한다. 일부 구현예에서, Cas 단백질은 E1369R, E1449H, 및 R1556A 돌연변이 또는 상기 위치에 상응하는 아미노산에 대한 유사한 치환을 포함한다. 일부 구현예에서, Cas 단백질은 E782K, N968K, 및 R1015H 돌연변이 또는 상기 위치에 상응하는 아미노산에 대한 유사한 치환을 포함한다. 일부 구현예에서, Cas 단백질은 D1135V, R1335Q, 및 T1337R 돌연변이 또는 상기 위치에 상응하는 아미노산에 대한 유사한 치환을 포함한다. 일부 구현예에서, Cas 단백질은 S542R 및 K607R 돌연변이 또는 상기 위치에 상응하는 아미노산에 대한 유사한 치환을 포함한다. 일부 구현예에서, Cas 단백질은 S542R, K548V, 및 N552R 돌연변이 또는 상기 위치에 상응하는 아미노산에 대한 유사한 치환을 포함한다. 변경된 PAM 서열을 인식하기 위한 Cas 효소의 조작에서의 예시적인 발전은, 그 전체가 본 명세서에 참고로 포함된 문헌[Collias et al Nature Communications 12:555 (2021)]에 검토되어 있다.

[0251] 일부 구현예에서, Cas 단백질은 촉매적으로 활성이며, 표적 DNA 부위의 하나 또는 2개의 가닥을 절단한다. 일부 구현예에서, 표적 DNA 부위의 절단 후 변경, 예를 들어 세포 복구 기구에 의한, 예를 들어 삽입 또는 결실이 형성된다.

[0252] 일부 구현예에서, Cas 단백질은 뉴클레아제, 예를 들어, 뉴클레아제-결핍 Cas9를 탈활성화 또는 부분적으로 탈활성화하도록 변형된다. 야생형 Cas9는 gRNA에 의해 표적화된 특이적인 DNA 서열에서 이중 가닥 파손(DSB)을 생성하는 한편, 변형된 기능기들을 갖는 다수의 CRISPR 엔도뉴클레아제가 이용가능하며, 예를 들어: 부분적으로 탈활성화된 Cas9의 "닉카제" 형태는 단일 가닥 파손만을 생성하고; 촉매적으로 비활성인 Cas9("dCas9")는 표적 DNA를 절단하지 않는다. 일부 구현예에서, DNA 서열에 결합하는 dCas9는 입체 장애에 의해 그 부위에서 전사를 간섭할 수 있다. 일부 구현예에서, 앵커 서열에 결합하는 dCas9는 기하학적 복합체(예를 들어 ASMC) 형성 및/또는 유지를 간섭(예를 들어, 감소 또는 방지)할 수 있다. 일부 구현예에서, DNA-결합 도메인은 촉매적으로 비활성인 Cas9, 예를 들어 dCas9를 포함한다. 많은 촉매적으로 비활성인 Cas9 단백질이 당업계에 알려져 있다. 일부 구현예에서, dCas9는 Cas 단백질의 각각의 엔도뉴클레아제 도메인에서 돌연변이, 예를 들어 D10A 및 H840A 또는 N863A 돌연변이를 포함한다. 일부 구현예에서, 촉매적으로 비활성이거나 부분적으로 비활성인 CRISPR/Cas 도메인은 하나 이상의 돌연변이, 예를 들어 표 3에 열거된 하나 이상의 돌연변이를 포함하는 Cas 단백질을 포함한다. 일부 구현예에서, 표 3의 제공된 열에 기재된 Cas 단백질은 표 3의 동일한 열에 열거된 1, 2, 3, 또는 모든 돌연변이를 포함한다. 일부 구현예에서, 예를 들어 표 3에 기재되지 않은, Cas 단백질은 표 3의 열에 열거된 1, 2, 3, 또는 모든 돌연변이, 또는 Cas 단백질에서 상응하는 부위에서 상응하는 돌연변이를 포함한다.

[0253] 일부 구현예에서, 촉매적으로 비활성인, 예를 들어 dCas9 또는 부분적으로 탈활성화된 Cas9 단백질은 D11 돌연변이(예를 들어, D11A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사한 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9 또는 부분적으로 탈활성화된 Cas9 단백질은 H969 돌연변이(예를 들어, H969A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9 또는 부분적으로 탈활성화된 Cas9 단백질은 N995 돌연변이(예를 들어, N995A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9는 위치 D11, H969, 및 N995 중 1, 2 또는 3개에서의 돌연변이(예를 들어, D11A, H969A, 및 N995A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다.

[0254] 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9 또는 부분적으로 탈활성화된 Cas9 단백질은 D10 돌연변이(예를 들어, D10A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9 또는 부분적으로 탈활성화된 Cas9 단백질은 H557 돌연변이(예를 들어, H557A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9는 D10 돌연변이(예를 들어, D10A 돌연변이) 및 H557 돌연변이(예를 들어, H557A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다.

[0255] 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 D839 돌연변이(예를 들어, D839A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함

한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 H840 돌연변이(예를 들어, H840A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 N863 돌연변이(예를 들어, N863A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9는 D10 돌연변이(예를 들어, D10A), D839 돌연변이(예를 들어, D839A), H840 돌연변이(예를 들어, H840A), 및 N863 돌연변이(예를 들어, N863A) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다.

[0256] 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 E993 돌연변이(예를 들어, E993A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다.

[0257] 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 D917 돌연변이(예를 들어, D917A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 E1006 돌연변이(예를 들어, E1006A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 D1255 돌연변이(예를 들어, D1255A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9는 D917 돌연변이(예를 들어, D917A), E1006 돌연변이(예를 들어, E1006A), 및 D1255 돌연변이(예를 들어, D1255A) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다.

[0258] 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 D16 돌연변이(예를 들어, D16A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 D587 돌연변이(예를 들어, D587A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 부분적으로 탈활성화된 Cas 도메인은 닉카제 활성을 갖는다. 일부 구현예에서, 부분적으로 탈활성화된 Cas9 도메인은 Cas9 닉카제 도메인이다. 일부 구현예에서, 촉매적으로 비활성인 Cas 도메인 또는 사멸 Cas 도메인은 검출가능한 이중 가닥 파손 형성을 생성하지 않는다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 H588 돌연변이(예를 들어, H588A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9, 또는 부분적으로 탈활성화된 Cas9 단백질은 N611 돌연변이(예를 들어, N611A 돌연변이) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다. 일부 구현예에서, 촉매적으로 비활성인 Cas9 단백질, 예를 들어 dCas9는 D16 돌연변이(예를 들어, D16A), D587 돌연변이(예를 들어, D587A), H588 돌연변이(예를 들어, H588A), 및 N611 돌연변이(예를 들어, N611A) 또는 상기 위치에 상응하는 아미노산에 대한 유사 치환을 포함한다.

[0259] 일부 구현예에서, DNA-결합 도메인 또는 엔도뉴클레아제 도메인은 gRNA(예를 들어, 주형 핵산, 예를 들어 gRNA를 포함하는 주형 RNA)를 포함하거나 그에 연결된(예를 들어, 공유결합적으로) Cas 분자를 포함할 수 있다.

[0260] 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 스트렙토코커스 피오게네스 Cas9(SpCas9) 또는 이의 기능성 단편 또는 변이체를 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 변형된 SpCas9를 포함한다. 구현예에서, 변형된 SpCas9는 프로토스페이서-인접 모티프(PAM) 특이성을 변경시키는 변형을 포함한다. 구현예에서, PAM은 핵산 서열 5'-NGT-3'에 대한 특이성을 갖는다. 구현예에서, 변형된 SpCas9는 하나 이상의 아미노산 치환을, 예를 들어 위치 L1111, D1135, G1218, E1219, A1322, 또는 R1335 중 하나 이상에서, 예를 들어 L1111R, D1135V, G1218R, E1219F, A1322R, R1335V로부터 선택되는 위치에서 포함한다. 구현예에서, 변형된 SpCas9는 아미노산 치환 T1337R 및 예를 들어 L1111, D1135L, S1136R, G1218S, E1219V, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, R1335Q, T1337, T1337L, T1337Q, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337H, T1337Q, 및 T1337M으로부터 선택되는 하나 이상의 추가의 아미노산 치환, 또는 이에 상응하는 아미노산 치환을 포함한다. 구현예에서, 변형된 SpCas9는 (i) D1135L, S1136R, G1218S, E1219V, A1322R, R1335Q, 및 T1337로부터 선택되는 하나 이상의 아미노산 치환; 및 (ii) L1111R, G1218R, E1219F, D1332A, D1332S, D1332T, D1332V, D1332L, D1332K, D1332R, T1337L, T1337I, T1337V, T1337F, T1337S, T1337N, T1337K, T1337R, T1337H, T1337Q, 및 T1337M으로부터 선택되는 하나 이상의

치환 또는 이에 상응하는 아미노산 치환을 포함한다.

[0261] 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 Cas 도메인, 예를 들어 Cas9 도메인을 포함한다. 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 뉴클레아제 활성 Cas 도메인, Cas 닉카제(nCas) 도메인, 또는 뉴클레아제-비활성 Cas(dCas) 도메인을 포함한다. 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 뉴클레아제-활성 Cas9 도메인, Cas9 닉카제(nCas9) 도메인, 또는 뉴클레아제-비활성 Cas9(dCas9) 도메인을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 Cas9(예를 들어, dCas9 및 nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 또는 Cas12i의 Cas9 도메인을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 Cas9(예를 들어, dCas9 및 nCas9), Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 또는 Cas12i를 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 S. 피오게네스 또는 S. 써모필루스 Cas9, 또는 이의 기능성 단편을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은, 예를 들어 본 명세서에 참고로 포함된 문헌[Chylinski, Rhun, and Charpentier (2013) RNA Biology 10:5, 726-737]에 기재된 바와 같은, Cas9 서열을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 HNH 뉴클레아제 서브도메인 및/또는 Cas, 예를 들어 본 명세서에 기재된 바와 같은 Cas9의 RuvC1 서브도메인, 또는 이의 변이체를 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, 또는 Cas12i를 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 Cas 폴리펩티드(예를 들어, 효소), 또는 이의 기능성 단편을 포함한다. 구현예에서, Cas 폴리펩티드(예를 들어, 효소)는 Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas5d, Cas5t, Cas5h, Cas5a, Cas6, Cas7, Cas8, Cas8a, Cas8b, Cas8c, Cas9(예를 들어, Csn1 또는 Csx12), Cas10, Cas10d, Cas12a/Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12d/CasY, Cas12e/CasX, Cas12g, Cas12h, Cas12i, Csy1 , Csy2, Csy3, Csy4, Cse1, Cse2, Cse3, Cse4, Cse5e, Csc1, Csc2, Csa5, Csn1, Csn2, Csm1, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx1S, Csx11, Csf1, Csf2, Cs0, Csf4, Csd1, Csd2, Cst1, Cst2, Csh1, Csh2, Csa1, Csa2, Csa3, Csa4, Csa5, II형 Cas 이펙터 단백질, V형 Cas 이펙터 단백질, VI형 Cas 이펙터 단백질, CARF, DinG, Cpf1, Cas12b/C2c1, Cas12c/C2c3, Cas12b/C2c1, Cas12c/C2c3, SpCas9(K855A), eSpCas9(1.1), SpCas9-HF1, 초정확한 Cas9 변이체(HypaCas9), 이의 동족체, 변형된 또는 이의 조작된 형태들, 및/또는 이의 기능성 단편들로부터 선택된다. 구현예에서, Cas9는 예를 들어 H840A, D10A, P475A, W476A, N477A, D1125A, W1126A, 및 D1127A로부터 선택되는 하나 이상의 치환을 포함한다. 구현예에서, Cas9는 D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, 및/또는 A987로부터 선택되는 위치에서 하나 이상의 돌연변이, 예를 들어 D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, 및/또는 D986A로부터 선택되는 하나 이상의 치환을 포함한다. 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 코리네박테리움 울세란스(Corynebacterium ulcerans), 코리네박테리움 디프테리아(Corynebacterium diphtheria), 스피로플라스마 시르피디콜라(Spiroplasma syrphidicola), 프레보텔라 인터미디아(Prevotella intermedia), 스피로플라스마 타이와넨스(Spiroplasma taiwanense), 스트렙토코커스 이니에(Streptococcus iniae), 벨리엘라 발티카(Belliella baltica), 사이크로플렉서스 토르퀴스(Psychroflexus torquis), 스트렙토코커스 써모필루스, 리수테리아 이노쿠아(Listeria innocua), 캄필로박터 제주니(Campylobacter jejuni), 나이세리아 메닝지티디스, 스트렙토코커스 피오게네스, 또는 스태필로코커스 아우레우스(Staphylococcus aureus), 또는 이들의 단편 또는 변이체로부터의 Cas(예를 들어, Cas9) 서열을 포함한다.

[0262] 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은, 예를 들어 D917A, E1006A, D1255A, D917A/E1006A, D917A/D1255A, E1006A/D1255A, 및 D917A/E1006A/D1255A로부터 선택된, 예를 들어 D917, E1006A, D1255 위치 또는 이들의 임의의 조합에서, 예를 들어 하나 이상의 치환을 포함하는 Cpf1 도메인을 포함한다.

[0263] 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA 결합 도메인은 spCas9, spCas9-VRQR(서열번호 4019), spCas9- VRER(서열번호 4020), xCas9(sp), saCas9, saCas9-KKH, spCas9-MQKSER(서열번호 4021), spCas9-LRKIQK(서열번호 4022), 또는 spCas9- LRVSQL(서열번호 4023)를 포함한다.

[0264] 일부 구현예에서, 유전자 변형 폴리펩티드는 Cas9 닉카제, 예를 들어 Cas9 H840A를 포함하는 엔도뉴클레아제 도메인을 갖는다. 구현예에서, Cas9 H840A는 다음 아미노산 서열을 갖는다:

[0265] *Cas9 닉카제(H840A):*

[0266] DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEE

SFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASG
VDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILR
VNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGS
IPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKH
SLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENE
DILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYYLQ
NGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKR
QLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKM
IAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARK
KDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNE
LALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFKYF
DTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0267] 일부 구현예에서, 유전자 변형 폴리펩티드는 D10A 및/또는 H840A 돌연변이를 포함하는 dCas9 서열, 예를 들어 다음 서열을 포함한다:

[0268] SMDKKYSIGLAIGTNSVGWAVITDDYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRL
EESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINA
SGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDI
LRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDN
GSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLP
KHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEE
NEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQV
SGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKLYLYY
LQNGRDMYVDQELDINRLSDYDVDAIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDKAGFI
KRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVR
KMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIA
RKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKG
NELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD (서열번호 4007)

[0269] *TAL 이펙터 및 징크 핑거 뉴클레아제*

[0270] 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA-결합 도메인은 TAL 이펙터 분자를 포함한다. TAL 이펙터 분자, 예를 들어 DNA 서열에 특이적으로 결합하는 TAL 이펙터 분자는 복수의 TAL 이펙터 도메인들 또는 이의 단편들을 통상적으로 포함하며, 선택적으로 자연에 존재하는 TAL 이펙터들의 하나 이상의 추가의 부분들을 포함한다(예를 들어, TAL 이펙터 도메인들의 N- 및/또는 C-말단). 많은 TAL 이펙터가 당업자에게 알려져 있으며, 예를 들어 Thermo Fisher Scientific으로부터, 상업적으로 구매가능하다.

[0271] 자연에 존재하는 TALE는 숙주 식물에서의 유전자 발현을 조절하고, 박테리아 콜로니화 및 생존을 용이하게 하는 식물 병원체 잔토모나스(*Xanthomonas*)를 포함하는 박테리아 병원체의 다양한 종에 의해 분비된 천연 이펙터 단백질이다. TAL 이펙터의 특이적 결합은 통상적으로 33 또는 234개의 아미노산의 나란히 정렬된 거의 동일한 반복부들의 중심 반복 도메인을 기준으로 한다(반복부-가변 이(di)-잔기, RVD 도메인).

[0272] TAL 이펙터 과의 원들은 그들의 반복부들의 수와 순서에서 주로 상이하다. 반복부들의 수는 통상적으로 1.5 내지 33.5개의 반복부 범위이며, C-말단 반복부는 일반적으로 길이가 더 짧고(예를 들어, 약 20개 아미노산), 일반적으로 "절반-반복부"로서 지칭된다. TAL 이펙터의 각각의 반복부는 일반적으로, 상이한 염기쌍 특이성을 나타내는 상이한 반복부 유형과 하나의 반복-대-하나의 염기쌍 상관관계를 특징으로 한다(하나의 반복부는 표적 유전자 서열 상에서 일 염기쌍을 인식함). 일반적으로, 반복부의 수가 더 적을수록, 단백질-DNA 상호작용은 더 약해진다. 다수의 6.5개 반복부는 리포터 유전자의 전사를 활성화하기에 충분한 것으로 보인다(Scholze et al., 2010).

[0273] 반복부 대 반복부 변이는 아미노산 위치 12 및 13에서 주로 발생하며, 따라서 이는 "초가변"으로 지칭되고, 표적 DNA 프로모터 서열과의 상호작용의 특이성에 책임이 있고, 이는 예시적인 반복부 가변 2잔기(RVD) 및 이들의

핵산 염기 표적에 대한 상동성 열거한 표 5에 나타낸 바와 같다.

[0274] **표 5 - RVD 및 핵산 염기 특이성**

표 5

[0275]

표적	가능한 RVD 아미노산 조합												
A	NI	NN	CI	HI	KI								
G	NN	GN	SN	VN	LN	DN	QN	EN	HN	RH	NK	AN	FN
C	HD	RD	KD	ND	AD								
T	NG	HG	VG	IG	EG	MG	YG	AA	EP	VA	QG	KG	RG

[0276] 이에 따라, TAL 이펙터의 반복부를 표적 특이적 DAN 서열로 변형하는 것이 가능하다. 추가의 연구는, RVD NK가 G를 표적화할 수 있음을 보여준다. TAL 이펙터의 표적 부위는 제1 반복부에 의해 표적화되는 5' 염기에 측접하는 T를 포함하는 경향도 있지만, 이러한 인식의 정확한 기작은 알려지지 않았다. 113개가 넘는 TAL 이펙터 서열이 오늘날까지 알려졌다. 잔토모나스로부터의 TAL 이펙터의 비제한적 예는 Hax2, Hax3, Hax4, AvrXa7, AvrXa10 및 AvrBs3를 포함한다.

[0277] 이에 따라, 본 명세서에 기재된 TAL 이펙터 분자의 TAL 이펙터 도메인은 임의의 박테리아 종(예를 들어, 잔토모나스 종, 예컨대 아프리카 균주인, 잔토모나스 오리제(*oryzae*) pv. 오리제(Yu et al. 2011), 잔토모나스 캄페스트리스 pv. 라파니(*campestris* pv. *raphani*) 균주 756C 및 잔토모나스오리제 pv. 오리지콜라(*oryzicola*) 균주 BLS256 (Bogdanove et al. 2011)으로부터의 TAL 이펙터로부터 유도될 수 있다. 일부 구현예에서, TAL 이펙터 도메인은 RVD 도메인뿐만 아니라 자연에 존재하는 TAL 이펙터로부터의 측접 서열(들)(RVD 도메인의 N-말단 및/또는 C-말단 측부 상에서의 서열)도 포함한다. 이는 자연에 존재하는 TAL 이펙터의 RVD보다 더 많거나 더 적은 반복부들을 포함할 수 있다. TAL 이펙터 분자는 상기 코드를 기준으로 한 소정의 DNA 서열 및 본 기술 분야에 알려진 다른 것들을 표적화하도록 설계될 수 있다. TAL 이펙터 도메인의 수(예를 들어, 반복부(단량체 또는 모듈)) 및 그의 특이적 서열은 원하는 DNA 표적 서열을 기준으로 선택될 수 있다. 예를 들어, TAL 이펙터 도메인, 예를 들어 반복부들은 특이적 표적 서열에 맞도록 제거 또는 부가될 수 있다. 구현예에서, 본 발명의 TAL 이펙터 분자는 6.5 내지 33.5 TAL 이펙터 도메인, 예를 들어 반복부를 포함한다. 구현예에서, 본 발명의 TAL 이펙터 분자는 8 내지 33.5 TAL 이펙터 도메인, 예를 들어 반복부, 예를 들어 10 내지 25 TAL 이펙터 도메인, 예를 들어 반복부, 예를 들어 10 내지 14 TAL 이펙터 도메인, 예를 들어 반복부를 포함한다.

[0278] 일부 구현예에서, TAL 이펙터 분자는 DNA 표적 서열에 대한 완전한 매치에 상응하는 TAL 이펙터 도메인을 포함한다. 일부 구현예에서, DNA 표적 서열 상에서 반복부와 표적 염기쌍 사이의 미스매치는 TAL 이펙터 분자를 포함하는 폴리펩티드의 기능을 가능하게 하는 한 허용된다. 일반적으로, TALE 결합은 미스매치의 수와 역의 상관관계를 갖는다. 일부 구현예에서, 본 발명의 폴리펩티드의 TAL 이펙터 분자는 표적 DNA 서열과, 7개의 미스매치, 6개의 미스매치, 5개의 미스매치, 4개의 미스매치, 3개의 미스매치, 2개의 미스매치, 또는 1개의 미스매치 이하를 포함하고, 선택적으로 미스매치가 없다. 이론에 구속되고자 함이 없이, 일반적으로 TAL 이펙터 분자 내 TAL 이펙터 도메인의 수가 더 적을수록, 더 적은 수의 미스매치가 용인될 것이고, TAL 이펙터 분자를 포함하는 폴리펩티드의 기능을 여전히 허용할 것이다. 결합 친화도는 매칭되는 반복부-DNA 조합의 합에 따라 달라지는 것으로 생각된다. 예를 들어, 25 이상의 TAL 이펙터 도메인을 갖는 TAL 이펙터 분자는 7 이하의 미스매치를 용인할 수 있다.

[0279] TAL 이펙터 도메인에 더하여, 본 발명의 TAL 이펙터 분자는 자연에 존재하는 TAL 이펙터로부터 유도된 추가의 서열들을 포함할 수 있다. TAL 이펙터 분자의 TAL 이펙터 도메인의 각각의 측면에 포함된 C-말단 및/또는 N-말단 서열(들)의 길이는 다양할 수 있으며, 예를 들어 문헌[Zhang et al. (2011)]의 연구를 기준으로 당업자에 의해 선택될 수 있다. 상기 문헌[Zhang et al.]은 Hax3 유도된 TAL-이펙터를 기초로 한 단백질에서 많은 C-말단 및 N-말단 절단 돌연변이체를 특징화하였으며, 표적 서열에 대한 최적 결합 및 이에 따라 전사 활성화에 기여하는, 주요 요소를 확인하였다. 일반적으로, 전사 활성은 N-말단의 길이와 역의 상관관계에 있는 것으로 밝혀졌다. C-말단에 대하여, Hax 3 서열의 최초 68개 아미노산 내 DNA 결합 잔기에 중요한 요소가 확인되었다. 이에 따라 일부 구현예에서, 자연에 존재하는 TAL 이펙터의 TAL 이펙터 도메인의 C-말단부 상에서 최초 68개 아미노산은 TAL 이펙터 분자 내에 포함된다. 따라서, 구현예에서 TAL 이펙터 분자는 1) 자연에 존재하는 TAL 이펙터로부터 유래된 하나 이상의 TAL 이펙터 도메인; 2) TAL 이펙터의 N-말단부에 대한 자연에 존재하는 TAL 이펙

터로부터 적어도 70, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260, 270, 280개 이상의 아미노산; 및/또는 3) TAL 이펙터 도메인의 C-말단부 상에서 자연에 존재하는 TAL 이펙터로부터의 적어도 68, 80, 90, 100, 110, 120, 130, 140, 150, 170, 180, 190, 200, 220, 230, 240, 250, 260 이상의 아미노산을 포함한다.

[0280] 일부 구현예에서, 엔도뉴클레아제 도메인 또는 DNA-결합 도메인은 Zn 핑거 분자이거나, 또는 이를 포함한다. Zn 핑거 분자는 Zn 핑거 단백질, 예를 들어, 자연에 존재하는 Zn 핑거 단백질 또는 조작된 Zn 핑거 단백질, 또는 이의 단편을 포함한다. 많은 Zn 핑거 단백질이 당업자에게 알려져 있으며, 예를 들어 Sigma-Aldrich로부터, 상업적으로 구매가능하다.

[0281] 일부 구현예에서, Zn 핑거 분자는, 선택된 표적 DNA 서열에 결합하도록 조작된, 자연에 존재하지 않는 Zn 핑거 단백질을 포함한다. 예를 들어, 그 전체가 본 명세서에 참고로 포함된, 문헌[Beerli, et al. (2002) Nature Biotechnol. 20:135-141; Pabo, et al. (2001) Ann. Rev. Biochem. 70:313-340; Isalan, et al. (2001) Nature Biotechnol. 19:656-660; Segal, et al. (2001) Curr. Opin. Biotechnol. 12:632-637; Choo, et al. (2000) Curr. Opin. Struct. Biol. 10:411-416]; 미국 특허 6,453,242; 6,534,261; 6,599,692; 6,503,717; 6,689,558; 7,030,215; 6,794,136; 7,067,317; 7,262,054; 7,070,934; 7,361,635; 7,253,273; 및 미국 특허 공개 2005/0064474; 2007/0218528; 2005/0267061을 참조한다.

[0282] 조작된 Zn 핑거 단백질은, 자연에 존재하는 Zn 핑거 단백질에 비해, 새로운 결합 특이성을 가질 수 있다. 조작 방법은 합리적 설계(rational design) 및 다양한 유형의 선택을 포함하지만 이에 제한되지 않는다. 합리적 설계는, 예를 들어 삼중자(또는 사중자) 뉴클레오티드 서열 및 개별적인 Zn 핑거 아미노산 서열을 포함하는 데이터베이스의 이용을 포함하며, 여기서 각각의 삼중자 또는 사중자 뉴클레오티드 서열은, 특정 삼중자 또는 사중자 서열에 결합하는 징크 핑거의 아미노산 서열 하나 이상과 회합된다. 예를 들어, 그 전체가 본 명세서에 참고로 포함된, 미국 특허 6,453,242 및 6,534,261호를 참조한다.

[0283] 예시적인 선택 방법은, 파지 디스플레이 및 2-하이브리드 시스템을 포함하고, 미국 특허 5,789,538; 5,925,523; 6,007,988; 6,013,453; 6,410,248; 6,140,466; 6,200,759; 및 6,242,568; 국제 특허 공개 WO 98/37186; WO 98/53057; WO 00/27878; 및 WO 01/88197 및 GB 2,338,237에 개시되어 있다. 추가적으로, 징크 핑거 단백질에 대한 결합 특이성의 향상은, 예를 들어 국제 특허 공개 WO 02/077227에 기재되어 있다.

[0284] 추가적으로, 이들 및 기타 참고에 개시된 것과 같이, 징크 핑거 도메인 및/또는 다중-핑거의 징크 핑거 단백질은 임의의 적합한 링커 서열을 함께 이용하여 링크될 수 있으며, 이는 예를 들어 길이 5개 이상의 아미노산의 링커를 포함한다. 또한, 6개 이상의 아미노산 길이의 예시적인 링커 서열에 대해, 미국 특허 6,479,626; 6,903,185; 및 7,153,949를 참조한다. 본 명세서에 기재된 단백질은 단백질의 개별적인 징크 핑거들 간에 적합한 링커의 임의의 조합을 포함할 수 있다. 추가적으로, 징크 핑거 결합 도메인에 대한 결합 특이성의 향상은, 예를 들어 공동 소유된 국제 특허 공보 WO 02/077227에 기재되어 있다.

[0285] Zn 핑거 단백질 및 융합 단백질의 설계 및 구축 방법(및 이를 인코딩하는 폴리뉴클레오티드)는 당업자에게 알려져 있으며, 미국 특허 6,140,0815; 789,538; 6,453,242; 6,534,261; 5,925,523; 6,007,988; 6,013,453; 및 6,200,759; 국제 특허 공개 WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197; WO 02/099084; WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536; 및 WO 03/016496에 상세하게 기재되어 있다.

[0286] 추가적으로, 이들 및 기타 참고에 개시된 바와 같이, Zn 핑거 단백질 및/또는 다중-핑거 Zn 핑거 단백질은, 예를 들어 길이 5개 이상의 아미노산의 링커를 포함하는, 임의의 적합한 링커 서열을 이용하여, 예를 들어 융합 단백질로서, 함께 연결될 수 있다. 또한, 6개 이상의 아미노산 길이의 예시적인 링커 서열에 대해, 미국 특허 6,479,626; 6,903,185; 및 7,153,949를 참조한다. 본 명세서에 기재된 Zn 핑거 분자는, 개별적인 징크 핑거 단백질 및/또는 Zn 핑거 분자의 다중-핑거 Zn 핑거 단백질간에 적합한 링커의 임의의 조합을 포함할 수 있다.

[0287] 특정 구현예에서, DNA-결합 도메인 또는 엔도뉴클레아제 도메인은 표적 DNA 서열에 (서열-특이적 방식으로) 결합하는 조작된 징크 핑거 단백질을 포함하는 Zn 핑거 분자를 포함한다. 일부 구현예에서, Zn 핑거 분자는 하나의 Zn 핑거 단백질 또는 이의 단편을 포함한다. 다른 구현예에서, Zn 핑거 분자는 복수의 Zn 핑거 단백질들(또는 이의 단편), 예를 들어, 2, 3, 4, 5, 6개 이상의 Zn 핑거 단백질(및 선택적으로 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 또는 2 이하의 Zn 핑거 단백질)을 포함한다. 일부 구현예에서, Zn 핑거 분자는 적어도 3개의 Zn 핑거 단백질을 포함한다. 일부 구현예에서, Zn 핑거 분자는 4, 5, 또는 6개의 핑거를 포함한다. 일부 구현예에서, Zn

핑거 분자는 8, 9, 10, 11 또는 12개의 핑거를 포함한다. 일부 구현예에서, 3개의 Zn 핑거 단백질을 포함하는 Zn 핑거 분자는 9 또는 10개의 뉴클레오티드를 포함하는 표적 DNA 서열을 인식한다. 일부 구현예에서, 4개의 Zn 핑거 단백질을 포함하는 Zn 핑거 분자는 12 내지 14개의 뉴클레오티드를 포함하는 표적 DNA 서열을 인식한다. 일부 구현예에서, 6개의 Zn 핑거 단백질을 포함하는 Zn 핑거 분자는 18 내지 21개의 뉴클레오티드를 포함하는 표적 DNA 서열을 인식한다.

[0288] 일부 구현예에서, Zn 핑거 분자는 양손(two-handed) Zn 핑거 단백질을 포함한다. 양손 징크 핑거 단백질은, 여기서 징크 핑거 단백질의 2개의 덩어리는 개재하는 아미노산들에 의해 분리되어, 2개의 징크 핑거 도메인이 2개의 불연속적인 표적 DNA 서열에 결합되도록 하는 단백질들이다. 징크 핑거 결합 단백질의 양손 유형의 예는 SIP1이며, 여기서 4개의 징크 핑거 단백질의 덩어리는 단백질의 아미노 말단에 위치되고, 3개의 징크 핑거 단백질의 덩어리는 카르복실 말단에 위치한다(문헌[Remade, et al. (1999) EMBO Journal 18(18):5073-5084] 참조). 이들 단백질의 징크 핑거의 각각의 덩어리는 이들 단백질에서 징크 핑거의 각각의 클러스터가 독특한 표적 서열에 결합할 수 있고, 2개의 표적 서열 사이의 간격은 많은 뉴클레오티드를 포함할 수 있다.

[0289] *링커*

[0290] 일부 구현예에서, 유전자 변형 폴리펩티드는 링커, 예를 들어 펩티드 링커, 예를 들어 표 6에 기재된 바와 같은 링커를 포함할 수 있다. 일부 구현예에서, 유전자 변형 폴리펩티드는 N-말단에서 C-말단 방향으로, Cas 도메인(예를 들어, 표 3의 Cas 도메인), 표 6의 링커(또는 이에 대하여 적어도 70%, 80%, 85%, 90%, 95%, 또는 99% 동일성을 갖는 서열), 및 RT 도메인(예를 들어, 표 2의 RT 도메인)을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 엔도뉴클레아제와 RT 도메인 사이에 가요성 링커, 예를 들어 아미노산 서열 SGGSSGGSSGSETPGTSESATPESSGGSSGGSS를 포함하는 링커를 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드의 RT 도메인은 엔도뉴클레아제 도메인의 C-말단에 위치할 수 있다. 일부 구현예에서, 유전자 변형 폴리펩티드의 RT 도메인은 엔도뉴클레아제 도메인의 N-말단에 위치할 수 있다.

[0291] **표 6 예시적인 링커 서열**

표 6

[0292]

아미노산 서열	서열번호
GGS	4101
GGSGGS	4102
GGSGGSGGS	4103
GGSGGSGGSGGS	4104
GGSGGSGGSGGSGGS	4105
GGSGGSGGSGGSGGSGGS	4106
GGGGS	4107
GGGGSGGGGS	4108
GGGGSGGGGSGGGGS	4109
GGGGSGGGGSGGGGSGGGGS	4110
GGGGSGGGGSGGGGSGGGGSGGGGS	4111
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	4112
GGG	4113
GGGG	4114
GGGGG	4115
GGGGGG	4116
GGGGGGG	4117
GGGGGGGG	4118
GSS	4119
GSSGSS	4120
GSSGSSGSS	4121
GSSGSSGSSGSS	4122
GSSGSSGSSGSSGSS	4123
GSSGSSGSSGSSGSSGSS	4124
EAAAK	4125
EAAAKEAAAK	4126
EAAAKEAAAKEAAAK	4127
EAAAKEAAAKEAAAKEAAAK	4128

EAAAKEAAAKEAAAKEAAAKEAAAK	4129
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	4130
PAP	4131
PAPAP	4132
PAPAPAP	4133
PAPAPAPAP	4134
PAPAPAPAPAP	4135
PAPAPAPAPAPAP	4136
GGSGGG	4137
GGGGGS	4138
GGSGSS	4139
GSSGGS	4140
GGSEAAAK	4141
EAAAKGGS	4142
GGSPAP	4143
PAPGGS	4144
GGGGSS	4145
GSSGGG	4146
GGGEAAAK	4147
EAAAKGGG	4148
GGGPAP	4149
PAPGGG	4150
GSSEAAAK	4151
EAAAKGSS	4152
GSSPAP	4153
PAPGSS	4154
EAAAKPAP	4155
PAPEAAAK	4156
GGSGGGGSS	4157
GGSGSSGGG	4158
GGGGGSGSS	4159
GGGGSSGGS	4160
GSSGGSGGG	4161
GSSGGGGGS	4162
GGSGGGEAAAK	4163
GGSEAAAKGGG	4164
GGGGGSEAAAK	4165
GGGEAAAKGGS	4166
EAAAKGGSGGG	4167
EAAAKGGGGGS	4168
GGSGGGPAP	4169
GGSPAPGGG	4170
GGGGGSPAP	4171
GGGPAPGGS	4172
PAPGGSGGG	4173
PAPGGGGGS	4174
GGSGSSEAAAK	4175
GGSEAAAKGSS	4176
GSSGGSEAAAK	4177
GSSEAAAKGGS	4178
EAAAKGGSGSS	4179
EAAAKGSSGGS	4180
GGSGSSPAP	4181
GGSPAPGSS	4182
GSSGGSPAP	4183
GSSPAPGGS	4184
PAPGGSGSS	4185
PAPGSSGGS	4186

GGSEAAAKPAP	4187
GGSPAPEAAAK	4188
EAAAKGGSPAP	4189
EAAAKPAPGGS	4190
PAPGGSEAAAK	4191
PAPEAAAKGGS	4192
GGGGSSEAAAK	4193
GGGEAAAKGSS	4194
GSSGGGEAAAK	4195
GSSEAAAKGGG	4196
EAAAKGGGGSS	4197
EAAAKGSSGGG	4198
GGGGSSPAP	4199
GGGPAPGSS	4200
GSSGGGPAP	4201
GSSPAPGGG	4202
PAPGGGGSS	4203
PAPGSSGGG	4204
GGGEAAAKPAP	4205
GGGPAPEAAAK	4206
EAAAKGGGPAP	4207
EAAAKPAPGGG	4208
PAPGGGEAAAK	4209
PAPEAAAKGGG	4210
GSSEAAAKPAP	4211
GSSPAPEAAAK	4212
EAAAKGSSPAP	4213
EAAAKPAPGSS	4214
PAPGSSEAAAK	4215
PAPEAAAKGSS	4216
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	4217
GGGGSEAAAKGGGGS	4218
EAAAKGGGGSEAAAK	4219
SGSETPGTSESATPES	4220
GSAGSAAGSGEF	4221
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	4222

[0293] 일부 구현예에서, 유전자 변형 폴리펩티드의 링커는 $(SGGS)_n$(서열번호 4025), $(GGGS)_n$(서열번호 4026), $(GGGGS)_n$(서열번호 4027), $(G)_n$, $(EAAAK)_n$(서열번호 4028), $(GGS)_n$, 또는 $(XP)_n$으로부터 선택되는 모티프를 포함한다.

[0294] *풀링된 스크리닝에 의한 유전자 변형 폴리펩티드 선택*

[0295] 후보 유전자 변형 폴리펩티드를 스크리닝하여 후보 유전자 편집 능력을 평가할 수 있다. 예를 들어, 인간 게놈에서 코딩 서열의 표적화된 편집을 위해 설계된 RNA 유전자 변형 시스템이 사용될 수 있다. 특정 구현예에서, 그러한 유전자 변형 시스템은 풀링된 스크리닝 접근법과 함께 사용될 수 있다.

[0296] 예를 들어, 유전자 변형 폴리펩티드 후보 및 주형 가이드 RNA(tgRNA)의 라이브러리를 포유동물 세포 내로 도입하여 풀링된 스크리닝 접근법에 의해 후보의 유전자 편집 능력을 시험할 수 있다. 특정 구현예에서, 유전자 변형 폴리펩티드 후보의 라이브러리를 포유동물 세포 내로 도입한 후 tgRNA를 세포 내로 도입한다.

[0297] 스크리닝에 사용될 수 있는 포유동물 세포의 대표적인 비제한적 예는 HEK293T 세포, U2OS 세포, HeLa 세포, HepG2 세포, Huh7 세포, K562 세포, 또는 iPS 세포를 포함한다.

[0298] 유전자 변형 폴리펩티드 후보는 1) Cas-뉴클레아제, 예를 들어 야생형 Cas 뉴클레아제, 예를 들어 야생형 Cas9 뉴클레아제, 돌연변이 Cas 뉴클레아제, 예를 들어 Cas 닉카제, 예를 들어 Cas9 닉카제, 예컨대 Cas9 N863A 닉카제, 또는 **표 3 또는 표 4**로부터 선택되는 Cas 뉴클레아제, 2) 다양한 정도의 길이, 가요성, 소수성, 및/또는 2

차 구조를 나타낼 수 있는 펩티드 링커, 예를 들어 **표 D 또는 표 6**의 서열; 및 3) 역전사효소(RT), 예를 들어 **표 D 또는 표 2**의 RT 도메인을 포함할 수 있다. 유전자 변형 폴리펩티드 후보 라이브러리는 Cas 뉴클레아제, 펩티드 링커 또는 RT 도메인 구성요소 중 1개, 2개 또는 3개 전부와 관련하여 서로 상이한 복수의 상이한 유전자 변형 폴리펩티드 후보, 또는 그러한 유전자 변형 폴리펩티드 후보를 인코딩하는 복수의 핵산 발현 벡터를 포함한다.

[0299] 유전자 변형 폴리펩티드 후보의 스크리닝을 위해, 유전자 변형 폴리펩티드 구성요소 및 tgRNA 구성요소를 포함하는 2-구성요소 시스템이 사용될 수 있다. 유전자 변형 구성요소는, 예를 들어 유전자 변형 폴리펩티드 후보를 인코딩하고, 예를 들어 유전자 변형 폴리펩티드 후보를 인코딩하는 인간 코돈-최적화 핵산, 예를 들어 상기 기재된 바와 같은 Cas-링커-RT 융합체를 포함하는 발현 벡터, 예를 들어 발현 플라스미드 또는 렌티바이러스 벡터를 포함할 수 있다. 특정 구현예에서, (i) 포유동물 세포에서의 발현을 위한 프로모터, 예를 들어 CMV 프로모터; (ii) 유전자 변형 라이브러리 후보, 예를 들어 **표 3 또는 표 4**의 Cas 뉴클레아제, **표 6**의 펩티드 링커 및 **표 2**의 RT를 포함하는 Cas-링커-RT 융합체, 예를 들어 **표 D**의 Cas-링커-RT 융합체; (iii) 자가-절단 폴리펩티드, 예를 들어 T2A 펩티드; (iv) 포유동물 세포에서 선택을 가능하게 하는 마커, 예를 들어 퓨로마이신 내성 유전자; 및 (v) 종결 신호, 예를 들어 폴리 A 테일을 포함하는 렌티바이러스 카세트가 이용된다.

[0300] tgRNA 구성요소는 tgRNA를 생산하고, tgRNA의 발현을 구동하는 U6 프로모터를 이용하며(여기서, tgRNA는 Cas에 의해 인식되는 비-코딩 RNA 서열임), tgRNA를 관심 게놈 유전자좌에 국재화시키며, 또한 RT 도메인에 의해 원하는 편집물을 게놈으로 역전사하는 주형을 제공하는, tgRNA 또는 발현 벡터, 예를 들어 발현 플라스미드를 포함할 수 있다.

[0301] 유전자 변형 폴리펩티드 라이브러리 후보를 발현하는 세포의 풀을 준비하기 위해, 포유동물 세포, 예를 들어 HEK293T 또는 U2OS 세포는 유전자 변형 후보 폴리펩티드 라이브러리의 풀링된 유전자 변형 폴리펩티드 후보 발현 벡터 제제, 예를 들어 렌티바이러스 제제를 이용하여 형질도입될 수 있다. 특정 구현예에서, 렌티바이러스 플라스미드가 이용되고, HEK293 Lenti-X 세포는 렌티바이러스 플라스미드 형질감염 전에 15 cm 플레이트(약 $12x10^6$개 세포)에 시딩된다. 그러한 구현예에서, 렌티바이러스 플라스미드 형질감염은 Lentiviral Packaging Mix(Biosettia)를 사용하여 수행될 수 있으며 유전자 변형 후보 라이브러리에 대한 플라스미드 DNA의 형질감염은 제조업체의 프로토콜에 따라 Lipofectamine 2000 및 Opti-MEM 배지를 사용하여 다음 날 수행된다. 그러한 구현예에서, 세포외 DNA는 다음 날 전체 배지 교체에 의해 제거될 수 있고 바이러스-함유 배지는 48시간 후에 수확될 수 있다. 렌티바이러스 배지는 Lenti-X Concentrator(TaKaRa Biosciences)를 사용하여 농축될 수 있으며 5 mL의 렌티바이러스 분취액이 제조되고 -80° C에서 보관될 수 있다. 렌티바이러스 적정은 선택 후, 예를 들어 퓨로마이신 선택 후 콜로니 형성 단위를 계수함으로써 수행된다.

[0302] 표적 DNA의 유전자 편집을 모니터링하기 위해, 표적 DNA를 운반하는 포유동물 세포, 예를 들어 HEK293T 또는 U2OS 세포가 이용될 수 있다. 표적 DNA의 유전자 편집을 모니터링하기 위한 다른 구현예에서, 표적 DNA 게놈 랜딩 패드(landing pad)를 운반하는 포유동물 세포, 예를 들어 HEK293T 또는 U2OS 세포가 이용될 수 있다. 특정 구현예에서, 표적 DNA 게놈 랜딩 패드는 관심 질환 또는 장애의 치료를 위해 편집될 유전자를 포함할 수 있다. 다른 특정 구현예에서, 표적 DNA는 유전자 편집이 발생했는지 여부를 결정하기 위해 모니터링될 수 있는 검출가능한 특성을 나타내는 단백질을 발현하는 유전자 서열이다. 예를 들어, 특정 구현예에서, 청색 형광 단백질(BFP)- 또는 녹색 형광 단백질(GFP)-발현 게놈 랜딩 패드가 이용된다. 특정 구현예에서, 표적 DNA, 예를 들어 표적 DNA 게놈 랜딩 패드를 포함하는 포유동물 세포, 예를 들어 HEK293T 또는 U2OS 세포는 유전자 변형 라이브러리 후보당 500x 내지 3000x 세포로 배양 플레이트에 시딩되고 0.2 내지 0.3의 감염다중도(MOI)로 형질도입되어 세포당 다중 감염을 최소화한다. 감염된 세포의 선택을 가능하게 하기 위해 감염 후 48시간 후에 퓨로마이신(2.5 ug/mL)이 첨가될 수 있다. 그러한 구현예에서, 세포는 적어도 7일 동안 퓨로마이신 선택 하에 유지된 다음, tgRNA 도입, 예를 들어 tgRNA 전기천공을 위해 규모가 확대될 수 있다.

[0303] 유전자 편집이 발생하는지 여부를 확인하기 위해, 편집된 표적 DNA를 함유하는 포유동물 세포를 유전자 변형 폴리펩티드 라이브러리 후보로 감염시킨 다음 표적 DNA 편집에 사용하도록 설계된 tgRNA로 형질감염시킬 수 있다. 이어서, 세포를 분석하여 설계된 결과에 따라 표적 유전자좌의 편집이 발생했는지, 또는 편집이 발생하지 않았거나 불완전한 편집이 발생했는지 여부를, 예를 들어 세포 분류 및 서열 분석을 사용하여 결정할 수 있다.

[0304] 특정 구현예에서, 게놈 편집이 발생하는지 여부를 확인하기 위해, BFP- 또는 GFP-발현 포유동물 세포, 예를 들어 HEK293T 또는 U2OS 세포를 유전자 변형 라이브러리 후보로 감염시킨 다음, 예를 들어 라이브러리 후보당 250x 내지 1000x 초과의 커버리지를 보장하는 세포 수에서 BFP에서 GFP로 또는 GFP에서 BFP로 전환시키도록 설

계된 20 ng의 tgRNA 플라스미드를 이용하여 웰당 250,000개 세포를 전기천공시킴으로써, tgRNA 플라스미드 또는 RNA로 형질감염시키거나 전기천공시킬 수 있다. 그러한 구현예에서, 이 분석에서 다양한 작제물의 게놈 편집 능력은 전기천공 후 4 내지10일 후에 색-전환 형광 단백질(FP)의 발현을 위해 형광 활성화 세포 분류(FACS)에 의해 세포를 분류함으로써 평가될 수 있다. 세포는 편집되지 않은 세포(원래의 형광 단백질 신호를 나타냄), 편집된 세포(전환된 형광 단백질 신호를 나타냄) 및 불완전한 편집(형광 단백질 신호를 나타내지 않음) 세포의 별개의 집단으로 분류 및 수확된다. 분석 동안에 후보 농축을 결정하기 위해 분류되지 않은 세포 샘플은 또한 입력 집단으로 수확될 수 있다.

[0305] 어떤 유전자 변형 라이브러리 후보가 분석에서 게놈 편집 능력을 나타내는지를 결정하기 위해, 분류된 세포 집단에서 게놈 DNA(gDNA)를 수확하고, 각각의 집단에서 유전자 변형 라이브러리 후보를 시퀀싱함으로써 분석한다. 간략하게, 유전자 변형 후보는 유전자 변형 폴리펩티드 발현 벡터, 예를 들어 렌티바이러스 카세트에 특이적인 프라이머를 사용하여 게놈으로부터 증폭될 수 있고, 게놈 DNA를 희석하기 위해 PCR의 두 번째 라운드에서 증폭된 다음 시퀀싱되며, 예를 들어 차세대 시퀀싱 플랫폼에 의해 시퀀싱될 수 있다. 시퀀싱 판독물의 품질 관리 후, 적어도 약 1500개 뉴클레오티드 및 일반적으로 약 3200개 이하의 뉴클레오티드의 판독물이 유전자 변형 폴리펩티드 라이브러리 서열에 맵핑되고 라이브러리 서열과 최소 약 80% 일치하는 판독물은 이러한 폴링된 스크린의 목적을 위해 주어진 후보에 대해 성공적으로 정렬된 것으로 간주된다. 분석에서 유전자 편집, 예를 들어 BFP에서 GFP로 또는 GFP에서 BFP로의 편집을 수행할 수 있는 후보를 확인하기 위해, 편집된 집단에 있는 각각의 라이브러리 후보의 판독 수를 초기의 분류되지 않은 집단에서 판독 수와 비교한다.

[0306] 폴링된 스크리닝의 목적을 위해, 게놈 편집 능력이 있는 유전자 변형 후보는 분류되지 않은 (입력) 세포에 비해 편집된(전환된 FP) 집단의 농축을 기준으로 확인된다. 일부 구현예에서, 입력에 대한 적어도 1.0, 1.5, 2.0, 2.5, 3.0, 4.0, 5.0, 6.0, 7.0, 8.0, 9.0, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90배, 또는 적어도 100배의 농축, 예를 들어 적어도 2배의 농축은 잠재적으로 유용한 유전자 편집 활성을 나타낸다. 일부 구현예에서, 농축은 농축 비율의 로그 베이스 2를 위함으로써 로그 값으로 전환된다. 일부 구현예에서, 적어도 0, 1, 2, 3, 4, 5, 5.5, 6.0, 6.2, 6.3, 6.4, 6.5, 또는 적어도 6.6의 log2 농축 점수, 예를 들어 적어도 1.0의 log2 농축 점수는 잠재적으로 유용한 유전자 편집 활성을 나타낸다. 특정 구현예에서, 유전자 변형 후보에 대해 관찰된 농축 값은 참조, 예를 들어 요소 ID 번호 17380을 이용하여 유사한 조건 하에 관찰된 농축 값과 비교될 수 있다.

[0307] 일부 구현예에서, 다중 tgRNA가 유전자 변형 후보 라이브러리를 스크리닝하는 데 사용될 수 있다. 특정 구현예에서, 복수의 tgRNA가, 예를 들어 특정 표적 유전자, 예를 들어 질환 치료를 위한 유전자 표적의 유전자 편집을 위해, 주형/Cas-링커-RT 융합 쌍을 최적화하는 데 이용될 수 있다. 특정 구현예에서, 유전자 변형 후보를 스크리닝하기 위한 폴링된 접근법은 배열된 형식으로 다수의 상이한 tgRNA를 사용하여 수행될 수 있다.

[0308] 일부 구현예에서, 다양한 유형의 편집, 예를 들어 다양한 길이의 삽입, 치환, 및/또는 결실이 유전자 변형 후보 라이브러리를 스크리닝하는 데 사용될 수 있다.

[0309] 일부 구현예에서, 다중 표적 서열, 예를 들어 다양한 형광 단백질이 유전자 변형 후보 라이브러리를 스크리닝하는 데 사용될 수 있다. 일부 구현예에서, 다중 표적 서열, 예를 들어 다양한 형광 단백질이 유전자 변형 후보 라이브러리를 스크리닝하는 데 사용될 수 있다. 일부 구현예에서, 다중 세포 유형, 예를 들어 HEK293T 또는 U2OS가 유전자 변형 후보 라이브러리를 스크리닝하는 데 사용될 수 있다. 당업자는 주어진 후보가 변경된 편집 능력을 나타낼 수 있거나 심지어 tgRNA 서열(예를 들어, 뉴클레오티드 변형, PBS 길이, RT 주형 길이), 표적 서열, 표적 위치, 편집 유형, 유전자 변형 폴리펩티드의 제1 가닥 닉에 대한 돌연변이의 위치, 또는 세포 유형을 포함하여, 다양한 조건에 걸쳐 임의의 관찰가능하거나 유용한 활성의 이득 또는 손실을 나타낼 수 있음을 이해할 것이다. 따라서, 일부 구현예에서, 유전자 변형 라이브러리 후보는, 예를 들어 적어도 2가지의 세포 유형에서 적어도 2개의 별개의 tgRNA를 이용하여 다수의 파라미터에 걸쳐 스크리닝되고, 유전자 편집 활성은 임의의 단일 조건에서의 강화에 의해 확인된다. 다른 구현예에서, 다양한 tgRNA 및 세포 유형에 걸쳐 더 강력한 활성을 갖는 후보는 적어도 2가지 조건, 예를 들어 스크리닝된 모든 조건에서의 강화에 의해 확인된다. 명확성을 위해, 임의의 주어진 조건 하에서 강화가 거의 나타나지 않거나 전혀 나타나지 않은 것으로 확인된 후보는 모든 조건에 걸쳐 비활성인 것으로 가정되지 않으며 다양한 파라미터로 스크리닝되거나, 예를 들어 도메인(예를 들어, RT 도메인), 링커, 또는 다른 신호(예를 들어, NLS)를 교환, 셔플링, 또는 변화시킴으로써, 폴리펩티드 수준에서 재구성될 수 있다.

[0310] *예시적인 Cas9-링커-RT 융합체의 서열*

[0311] 일부 구현예에서, 유전자 변형 폴리펩티드는 링커 서열 및 RT 서열을 포함한다. 일부 구현예에서, 유전자 변형

폴리펩티드는 표 D에 열거된 바와 같은 링커 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 아미노산 서열을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 표 D에 열거된 바와 같은 RT 도메인의 아미노산 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 아미노산 서열을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 표 D에 열거된 바와 같은 링커 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 아미노산 서열; 및 표 D에 열거된 바와 같은 RT 도메인의 아미노산, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 아미노산 서열을 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 (i) 표 D의 행에 열거된 바와 같은 링커 서열, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 아미노산 서열; 및 (ii) 표 D의 동일한 행에 열거된 바와 같은 RT 도메인의 아미노산, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 아미노산 서열을 포함한다.

[0312] *유전자 변형 시스템을 위한 국재화 서열*

[0313] 특정 구현예에서, 유전자 편집기 시스템 RNA는 세포내 국재화 서열, 예를 들어 핵 국재화 서열(NLS)을 추가로 포함한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 서열번호 4000 및/또는 서열번호 4001에 포함된 바와 같은 NLS, 또는 이에 대하여 적어도 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99% 동일성을 갖는 아미노산 서열을 갖는 NLS를 포함한다.

[0314] 핵 국재화 서열은 RNA의 핵 내로의 유입을 촉진하는 RNA 서열일 수 있다. 특정 구현예에서, 핵 위치 신호는 주형 RNA 상에 위치한다. 특정 구현예에서, 유전자 변형 폴리펩티드는 제1 RNA 상에서 인코딩되며, 주형 RNA는 제2의 별개의 RNA며, 핵 위치 신호는 주형 RNA 상에 위치되고, 유전자 변형 폴리펩티드를 인코딩하는 RNA 상에서는 위치하지 않는다. 이론에 구속되고자 하지는 않지만, 일부 구현예에서 유전자 변형 폴리펩티드를 인코딩하는 RNA는 일차적으로 세포질로 표적화되어 그의 번역을 촉진하는 한편, 주형 RNA는 일차적으로 핵으로 표적화되어 그의 게놈 내로의 삽입을 촉진한다. 일부 구현예에서, 핵 위치 신호는 3' 말단, 5' 말단에, 또는 주형 RNA의 내부 영역에 존재한다. 일부 구현예에서, 핵 위치 신호는 이종성 서열의 3'(예를 들어, 직접적으로 이종성 서열의 3')거나, 또는 이종성 서열의 5'(예를 들어, 직접적으로 이종성 서열의 5')이다. 일부 구현예에서, 핵 위치 신호는 주형 RNA의 5' UTR 외부 또는 3' UTR에 위치한다. 일부 구현예에서, 핵 위치 신호는 5' UTR과 3' UTR 사이에 위치되고, 여기서 선택적으로 핵 위치 신호는 이식유전자를 이용하여 전사되지 않는다(예를 들어, 핵 위치 신호는 항-센스 배향이거나, 전사 말단 신호 또는 폴리아데닐화 신호의 하류임). 일부 구현예에서, 핵 국재화 서열은 인트론의 내부에 위치한다. 일부 구현예에서, 복수의 동일 또는 상이한 핵 위치 신호는 RNA 내, 예를 들어 주형 RNA 내에 존재한다. 일부 구현예에서, 핵 위치 신호는 길이가 5, 10, 25, 50, 75, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900 또는 1000 bp 미만이다. 다양한 RNA 핵 국재화 서열이 사용될 수 있다. 예를 들어, 문헌[Lubelsky and Ulitsky, *Nature* 555 (107-111), 2018]은 RNA 국재화를 핵 내로 구동하는 RNA 서열을 기재한다. 일부 구현예에서, 핵 위치 신호는 SINE-유도된 핵 RNA 국재화(SIRLOIN) 신호이다. 일부 구현예에서, 핵 위치 신호는 핵-풍부화된 단백질에 결합한다. 일부 구현예에서, 핵 위치 신호는 HNRNPK 단백질에 결합한다. 일부 구현예에서, 핵 위치 신호는 피리미딘에 풍부하며, 예를 들어, C/T 풍부, C/U 풍부, C 풍부, T 풍부, 또는 U 풍부 영역이 있다. 일부 구현예에서, 핵 위치 신호는 긴 비-코딩 RNA로부터 유래된다. 일부 구현예에서, 핵 위치 신호는 MALATI 긴 비-코딩 RNA로부터 유래되거나, MALAT1의 600 뉴클레오티드 M 영역이다(문헌[Miyagawa et al., *RNA* 18, (738-751), 2012]에 기재됨). 일부 구현예에서, 핵 위치 신호는 BORG 긴 비-코딩 RNA로부터 유래되거나, AGCCC 모티프다(문헌[Zhang et al., *Molecular and Cellular Biology* 34, 2318-2329 (2014)]에 기재됨). 일부 구현예에서, 핵 국재화 서열은 문헌[Shukla et al., *The EMBO Journal* e98452 (2018)]에 기재되어 있다. 일부 구현예에서, 핵 위치 신호는 레트로바이러스로부터 유래된다.

[0315] 일부 구현예에서, 본 명세서에 기재된 폴리펩티드는 하나 이상(예를 들어, 2, 3, 4, 5)의 핵 표적화 서열, 예를 들어 핵 국재화 서열(NLS)를 포함한다. 일부 구현예에서, NLS는 이분절 NLS이다. 일부 구현예에서, NLS는 NLS를 포함하는 단백질의 세포 핵 내로의 유입을 촉진한다. 일부 구현예에서, NLS는 본 명세서에 기재된 바와 같은 유전자 변형 폴리펩티드의 N-말단에 융합된다. 일부 구현예에서, NLS는 유전자 변형 폴리펩티드의 C-말단에 융합된다. 일부 구현예에서, NLS는 Cas 도메인의 N-말단 또는 C-말단에 융합된다. 일부 구현예에서, 링커 서열은 NLS와 유전자 변형 폴리펩티드의 이웃 도메인 사이에 배치된다.

[0316] 일부 구현예에서, NLS는 아미노산 서열 MDSLLMNRRKFLYQFKNVRWAKGRRETYLC(서열번호 4009), PKKRKVEGADKRTADGSEFESPKKKRKV(서열번호 4010), RKSGKIAAIWKRPRKPKKKRKV(서열번호 4011), KRTADGSEFESPKKKRKV(서열번호 4012), KKTELQTTNAENKTKKL(서열번호 4013), 또는 KRGINDRNFWRGENGRKTR(서열번호

4014), KRPAATKKAGQAKKKK(서열번호 4015), 또는 이의 기능성 단편 또는 변이체를 포함한다. 예시적인 NLS 서열은 PCT/EP2000/011690에도 기재되어 있으며, 이의 내용은 예시적인 핵 국재화 서열의 개시에 대하여 본 명세서에 참고로 포함된다. 일부 구현예에서, NLS는 표 7에 개시된 바와 같은 아미노산 서열을 포함한다. 이 표의 NLS는, 핵으로의 세포내 국재화를 개선하기 위해, 폴리펩티드에서 하나 이상의 위치에서 폴리펩티드 내 하나 이상의 복제, 예를 들어 N-말단 도메인 내, 펩티드 도메인 사이, C-말단 도메인 내, 또는 위치의 조합에서 NLS의 1, 2, 3 이상의 복제를 사용하여 이용될 수 있다. 다수의 독특한 서열들이 단일 폴리펩티드 내에서 사용될 수 있다. 서열은 예를 들어 염기성 아미노산의 1 또는 2개의 스트레치(stretch)를 갖는, 자연히 단일분절 또는 이분절일 수 있거나, 또는 키메라 이분절 서열로 사용될 수 있다. 서열 참조는, 세포내 국재화 예측 알고리즘을 이용하여 캐낸 서열에 대해 SeqNLS로서 나타낸 경우를 제외하고는, UniProt 수탁번호에 상응한다(문헌[Lin et al BMC Bioinformat 13:157 (2012)], 이는 그 전체가 본 명세서에 참고로 포함된다).

[0317] **표 7 유전자 변형 시스템에서 사용하기 위한 예시적인 핵 국재화 신호**

표 7

[0318]

서열	서열 참조	서열번호
AHFKISGEKRPSTDPGKKAKNPKKKKKKDP	Q76IQ7	4223
AHRAKKMSKTHA	P21827	4224
ASPEYVNLPINGNG	SeqNLS	4225
CTKRPRW	O88622, Q86W56, Q9QYM2, O02776	4226
DKAKRVSRNKSEKKRR	O15516, Q5RAK8, Q91YB2, Q91YB0, Q8QGQ6, O08785, Q9WVS9, Q6YGZ4	4227
EELRLKEELLKGIYA	Q9QY16, Q9UHL0, Q2TBP1, Q9QY15	4228
EEQLRRRKNSRLNNTG	G5EFF5	4229
EVLKVIRTGKRKKKAWKRMVTKVC	SeqNLS	4230
HHHHHHHHHHHHHQPH	Q63934, G3V7L5, Q12837	4231
HKKKHPDASVNFSEFSK	P10103, Q4R844, P12682, B0CM99, A9RA84, Q6YKA4, P09429, P63159, Q08IE6, P63158, Q9YH06, B1MTB0	4232
HKRTKK	Q2R2D5	4233
IINGRKLKLKKSRRRSSQTSNNSFTSRRS	SeqNLS	4234
KAEQERRK	Q8LH59	4235
KEKRKRREELFIEQKKRK	SeqNLS	4236
KKGKDEWFSRGKKP	P30999	4237
KKGPSVQKRKKT	Q6ZN17	4238
KKKTVINDLLHYKKEK	SeqNLS, P32354	4239
KKNGGKGKNKPSAKIKK	SeqNLS	4240
KKPKWDDFKKKKK	Q15397, Q8BKS9, Q562C7	4241
KKRKKD	SeqNLS, Q91Z62, Q1A730, Q969P5, Q2KHT6, Q9CPU7	4242
KKRRKRRRK	SeqNLS	4243
KKRRRRARK	Q9UMS6, D4A702, Q91YE8	4244
KKSKRGR	Q9UBS0	4245
KKSRKRGS	B4FG96	4246
KKSTALSRELGKIMRRR	SeqNLS, P32354	4247
KKSYQDPEIIAHSRPRK	Q9U7C9	4248
KKTGKNRKLKSKRVKTR	Q9Z301, O54943, Q8K3T2	4249
KKVSIAGQSGKLWRWKR	Q6YUL8	4250
KKYENVVIKRSPRKRGRPRK	SeqNLS	4251
KNKKRK	SeqNLS	4252
KPKKKR	SeqNLS	4253
KRAMKDDSHGNSTSPKRRK	Q0E671	4254
KRANSNLVAAYEKAKKK	P23508	4255
KRASEDTTSGSPPKKSSAGPKR	Q9BZZ5, Q5R644	4256
KRFKRRWMVRKMKTKK	SeqNLS	4257
KRGLNSSFETSPKKVK	Q8IV63	4258
KRGNSSIGPNDLSKRKQRKK	SeqNLS	4259
KRIHSVSLSQSQIDPSKKVKRAK	SeqNLS	4260

KRKGKLKNKGSKRKK	O15381	4261
KRRRRRRREKRKR	Q96GM8	4262
KRSNDRTYSPEEEKQRRA	Q91ZF2	4263
KRTVATNGDASGAHRAKKMSK	SeqNLS	4264
KRVYNKGEDEQEHLPKGKKR	SeqNLS	4265
KSGKAPRRRAVSMDNSNK	Q9WVH4, O43524	4266
KVNFLDMSLDDIIIYKELE	Q9P127	4267
KVQHRIAKKTTRRRR	Q9DXE6	4268
LSPSLSPL	Q9Y261, P32182, P35583	4269
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC	Q9GZX7	4270
MPQNEYIELHRKRYGYRLDYHEKKRKKESRE AHERSKKAKKMIGLKAKLYHK	SeqNLS	4271
MVQLRPRASR	SeqNLS	4272
NNKLLAKRRKGGASPKDDPMDDIK	Q965G5	4273
NYKRPMDGTYGPPAKRHEGE	O14497, A2BH40	4274
PDTKRAKLDSSETTMVKKK	SeqNLS	4275
PEKRTKI	SeqNLS	4276
PGGRGKKK	Q719N1, Q9UBP0, A2VDN5	4277
PGKMDKGEHRQERRDRPY	Q01844, Q61545	4278
PKKGDKYDKTD	Q45FA5	4279
PKKKSRK	O35914, Q01954	4280
PKKNKPE	Q22663	4281
PKKRAKV	P04295, P89438	4282
PKPKKLKVE	P55263, P55262, P55264, Q64640	4283
PKRGRGR	Q9FYS5, Q43386	4284
PKRRLVDDA	P0C797	4285
PKRRRTY	SeqNLS	4286
PLFKRR	A8X6H4, Q9TXJ0	4287
PLRKAKR	Q86WB0, Q5R8V9	4288
PPAKRKCIF	Q6AZ28, O75928, Q8C5D8	4289
PPARRRRL	Q8NAG6	4290
PPKKKRKV	Q3L6L5, P03070, P14999, P03071	4291
PPNKRMKVKH	Q8BN78	4292
PPRIYPQLPSAPT	P0C799	4293
PQRSPFPKSSVKR	SeqNLS	4294
PRPRKVPR	P0C799	4295
PRRRVQRKR	SeqNLS, Q5R448, Q5TAQ9	4296
PRRVRLK	Q58DJ0, P56477, Q13568	4297
PSRKRPR	Q62315, Q5F363, Q92833	4298
PSSKKRKV	SeqNLS	4299
PTKKRVK	P07664	4300
QRPGPYDRP	SeqNLS	4301
RGKGGKGLGKGGAKRHRK	SeqNLS	4302
RKAGKGGGGHKTTKKRSAKDEKVP	B4FG96	4303
RKIKLKRAK	A1L3G9	4304
RKIKRKRAK	B9X187	4305
RKKEAPGPREELRSRGR	O35126, P54258, Q5IS70, P54259	4306
RKKRKGK	SeqNLS, Q29243, Q62165, Q28685, O18738, Q9TSZ6, Q14118	4307
RKKRRQRRR	P04326, P69697, P69698, P05907, P20879, P04613, P19553, P0C1J9, P20893, P12506, P04612, Q73370, P0C1K0, P05906, P35965, P04609, P04610, P04614, P04608, P05905	4308
RKKSIPLSIKNLKRKHKRKKNKITR	Q9C0C9	4309
RKLVKPKNTKMKTKLRTNPY	Q14190	4310
RKRLILSDKGQLDWKK	SeqNLS, Q91Z62, Q1A730, Q2KHT6, Q9CPU7	4311
RKRLKSK	Q13309	4312

RKRRVRDNM	Q8QPH4, Q809M7, A8C8X1, Q2VNC5, Q38SQ0, O89749, Q6DNQ9, Q809L9, Q0A429, Q20NV3, P16509, P16505, Q6DNQ5, P16506, Q6XT06, P26118, Q2ICQ2, Q2RCG8, Q0A2D0, Q0A2H9, Q9IQ46, Q809M3, Q6J847, Q6J856, B4URE4, A4GCM7, Q0A440, P26120, P16511,	4313
RKRSPKDKKEKDLDGAGKRRKT	Q7RTP6	4314
RKRTPRVDGQTGENDMNKRRRK	O94851	4315
RLPVRRRRRR	P04499, P12541, P03269, P48313, P03270	4316
RLRFRKPKSK	P69469	4317
RQQRKR	Q14980	4318
RRDLNSSFETSPKKVK	Q8K3G5	4319
RRDRAKLR	Q9SLB8	4320
RRGDGRRR	Q80WE1, Q5R9B4, Q06787, P35922	4321
RRGRKRKAEKQ	Q812D1, Q5XXA9, Q99JF8, Q8MJG1, Q66T72, O75475	4322
RRKKRR	Q0VD86, Q58DS6, Q5R6G2, Q9ERI5, Q6AYK2, Q6NYC1	4323
RRKRSKSEDMDSVESKRRR	Q7TT18	4324
RRKRSR	Q99PU7, D3ZHS6, Q92560, A2VDM8	4325
RRPKGKTLQKRKPK	Q6ZN17	4326
RRRGFERFGPDNMGRKRK	Q63014, Q9DBR0	4327
RRRGKNKVAAQNCRK	SeqNLS	4328
RRRKRR	Q5FVH8, Q6MZT1, Q08DH5, Q8BQP9	4329
RRRQKQKGGASRRR	SeqNLS	4330
RRRREGPRARRRR	P08313, P10231	4331
RRTIRLKLVYDKCDRSCKIQKKNRNKCQYCR FHKCLSVGMSHNAIRFGRMPRSEKAKLKAE	SeqNLS	4332
RRVPQRKEVSRCRKCRK	Q5RJN4, Q32L09, Q8CAK3, Q9NUL5	4333
RVGGRRQAVECIEDLLNEPGQPLDLSCKRPR P	P03255	4334
RVVKLRIAP	P52639, Q8JMN0	4335
RVVRRR	P70278	4336
SKRKTKISRKTR	Q5RAY1, O00443	4337
SYVKTVPNRTRTYIKL	P21935	4338
TGKNEAKKRKIA	P52739, Q8K3J5, Q5RAU9	4339
TLSPASSPSSVSCPVIPASTDESPGSALNI	SeqNLS	4340
VSKKQRTGKKIH	P52739, Q8K3J5, Q5RAU9	4341
SPKKKRKVE		4342
KRTAD GSEFE SPKKKRKVE		4343
PAAKRVKLD		4344
PKKKRKV		4345
MDSLLMNRRKFLYQFKNVRWAKGRRETYLC		4346
SPKKKRKVEAS		4347
MAPKKKRKVGIHRGVP		4348

[0319] 일부 구현예에서, NLS는 이분절 NLS이다. 이분절 NLS는 스페이서 서열에 의해 분리된 2개의 염기성 아미노산 클러스터를 통상적으로 포함한다(이는, 예를 들어 약 10개의 아미노산 길이). 단일분절 NLS는 통상적으로 스페이서가 결여되어 있다. 이분절 NLS의 예는 뉴클레오플라스민(nucleoplasmin) NLS으로, 이는 서열 KR[PAATKKAGQA]KKKK(서열번호 4015)를 갖고, 스페이서는 괄호표시되었다. 또 다른 예시적인 이분절 NLS는 서열 PKKKRKVEGADKRTADGSEFESPKKKRKV(서열번호 4016)를 갖는다. 예시적인 NLS는 국제 특허 WO2020051561에 기재되어 있으며, 이는 핵 국재화 서열에 대한 그의 개시내용을 포함하여 그 전체가 본 명세서에 참고로 포함된다.

[0320] 특정 구현예에서, 유전자 편집기 시스템 폴리펩티드(예를 들어, 본 명세서에 기재된 바와 같은 유전자 변형 폴리펩티드)는 세포내 국재화 서열, 예를 들어 핵 국재화 서열 및/또는 핵소체 국재화 서열을 추가로 포함한다. 핵 국재화 서열 및/또는 핵소체 국재화 서열은 단백질의 핵 및/또는 핵소체 내로의 유입을 촉진하는 아미노산 서열일 수 있으며, 이는 이종성 서열의 게놈 내로의 통합을 촉진할 수 있다. 특정 구현예에서, 유전자 편집기 시스템 폴리펩티드(예를 들어, 본 명세서에 기재된 바와 같은 유전자 변형 폴리펩티드)는 핵소체 국재화 서열을

추가로 포함한다. 특정 구현예에서, 유전자 변형 폴리펩티드는 제1 RNA 상에서 인코딩되고, 주형 RNA는 제2의 별개의 RNA이고, 핵소체 국재화 신호는 유전자 변형 폴리펩티드를 인코딩하는 RNA 상에서 인코딩되고, 주형 RNA 상에서는 인코딩되지 않는다. 일부 구현예에서, 핵소체 국재화 신호는 폴리펩티드의 N-말단, C-말단에, 또는 내부 영역에 위치한다. 일부 구현예에서, 복수의 동일하거나 상이한 핵소체 국재화 신호들이 사용된다. 일부 구현예에서, 핵 위치 신호는 길이가 5, 10, 25, 50, 75, 또는 100개 미만의 아미노산이다. 다양한 폴리펩티드 핵소체 국재화 신호가 사용될 수 있다. 예를 들어, 문헌[Yang et al., *Journal of Biomedical Science* 22, 33 (2015)]은 핵소체 국재화 신호로서 또한 기능하는 핵 위치 신호를 기재한다. 일부 구현예에서, 핵소체 국재화 신호는 핵 위치 신호일 수도 있다. 일부 구현예에서, 핵소체 국재화 신호는 핵 위치 신호와 중복될 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 뻗어 있는 염기성 잔기들을 포함할 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 아르기닌 및 리신 잔기가 풍부할 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 핵소체에서 풍부화된 단백질로부터 유래될 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 리보솜성 RNA 유전자좌에서 풍부화된 단백질로부터 유래될 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 rRNA에 결합하는 단백질로부터 유래될 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 MSP58로부터 유래될 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 단일분절 모티프일 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 이분절 모티프일 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 다수의 단일분절 또는 이분절 모티프로 이루어질 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 단일분절 또는 이분절 모티프의 혼합물로 이루어질 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 이중 이분절 모티프일 수 있다. 일부 구현예에서, 핵소체 국재화 모티프는 KRASSQALGTIPKRRSSSRFIKRKK(서열번호 4017)일 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 핵 인자-κB-유도 키나제로부터 유래될 수 있다. 일부 구현예에서, 핵소체 국재화 신호는 RKKRKKK 모티프(서열번호 4018)일 수 있다(문헌[Birbach et al., Journal of Cell Science, 117 (3615-3624), 2004]에 기재됨).

[0321] *유전자 변형 폴리펩티드 및 시스템의 진화된 변이체*

[0322] 일부 구현예에서, 본 발명은 본 명세서에 기재된 바와 같은 유전자 변형 폴리펩티드의 진화된 변이체를 제공한다. 진화된 변이체는 일부 구현예에서 참조 유전자 변형 폴리펩티드, 또는 그 안에 포함된 단편 또는 도메인 중 하나를 돌연변이시킴으로써 생산될 수 있다. 일부 구현예에서, 하나 이상의 도메인(예를 들어, 역전사효소 도메인)이 진화된다. 하나 이상의 그러한 진화된 변이체 도메인은, 일부 구현예에서 단독으로 또는 다른 도메인과 함께 진화될 수 있다. 진화된 변이체 도메인 또는 도메인들은, 일부 구현예에서 진화되지 않은 동족 구성요소(들) 또는 동족 구성요소(들)의 진화된 변이체와 조합될 수 있으며, 예를 들어 이는 병렬적 또는 연속적 방식으로 진화될 수 있다.

[0323] 일부 구현예에서, 참조 유전자 변형 폴리펩티드, 또는 이의 단편 또는 도메인을 돌연변이시키는 과정은 참조 유전자 변형 폴리펩티드 또는 이의 단편 또는 도메인을 돌연변이시키는 것을 포함한다. 구현예에서, 돌연변이유발은, 예를 들어 본 명세서에 기재된 바와 같이 연속적인 진화 방법(예를 들어, PACE) 또는 비연속적인 진화 방법(예를 들어, PANCE)을 포함한다. 일부 구현예에서, 진화된 유전자 변형 폴리펩티드 또는 이의 단편 또는 도메인은, 참조 유전자 변형 폴리펩티드의 아미노산 서열, 또는 이의 단편 또는 도메인에 비해 그의 아미노산 서열 내로 도입된 하나 이상의 아미노산 변이를 포함한다. 구현예에서, 아미노산 서열 변이는, 예를 들어 코딩 서열에서 임의의 특정 위치에서의 코돈의 변화, 하나 이상의 아미노산의 결실(예를 들어, 절단된 단백질), 하나 이상의 아미노산의 삽입, 또는 상기의 임의의 조합을 생성하는 유전자 변형 폴리펩티드를 인코딩하는 뉴클레오티드 서열에서의 변화 결과로서, 참조 유전자 변형 폴리펩티드의 아미노산 서열 내에 하나 이상의 돌연변이된 잔기(예를 들어, 보존적 치환, 비보존적 치환, 또는 이의 조합)를 포함할 수 있다. 진화된 변이체 유전자 변형 폴리펩티드는 유전자 변형 폴리펩티드의 하나 이상의 구성요소 또는 도메인에서의 변이체(예를 들어, 역전사효소 도메인 내로 도입된 변이체)를 포함할 수 있다.

[0324] 일부 양태에서, 본 개시내용은 유전자 변형 폴리펩티드, 시스템, 키트, 및 유전자 변형 폴리펩티드의 진화된 변이체를 사용 또는 포함하는 방법을 제공하며, 예를 들어 유전자 변형 폴리펩티드의 진화된 변이체 또는 PACE 또는 PANCE에 의해 생산되거나 생산가능한 유전자 변형 폴리펩티드를 이용한다. 구현예에서, 진화되지 않은 참조 유전자 변형 폴리펩티드는 본 명세서에 개시된 바와 같은 유전자 변형 폴리펩티드이다.

[0325] 본 명세서에 사용된 바와 같은 용어 "파지-보조된 지속적 진화(PACE)"는, 바이러스 벡터로서 파지를 이용하는 연속적 진화를 일반적으로 지칭한다. PACE 기술의 예는, 예를 들어 2009년 9월 8일 출원되고, 2010년 3월 11일 WO 2010/028347로 공개된, 국제 PCT 출원 PCT/US 2009/056194; 2011년 12월 22일 출원되고, 2012년 6월 28일 WO 2012/088381로 공개된, 국제 PCT 출원 PCT/US2011/066747; 2015년 5월 5일 발행된 미국 특허 9,023,594; 2017년 9월 26일 발행된 미국 특허 9,771,574; 2016년 7월 19일 발행된 미국 특허 9,394,537; 2015년 1월 20일

출원되고, 2015년 9월 11일 WO 2015/134121로 공개된, 국제 PCT 출원 PCT/US2015/012022; 2019년 1월 15일 발행된 미국 특허 10,179,911; 및 2016년 4월 15일 출원되고, 2016년 10월 20일 공개된 WO 2016/168631로 공개된 국제 PCT 출원 PCT/US2016/027795에 기재되어 있으며, 이들 각각의 전체 내용은 본 명세서에 참고로 포함된다.

[0326] 본 명세서에 사용된 바와 같은 용어 "파지-보조된 비연속적 진화(PANCE)"는 바이러스 벡터로서 파지를 이용하는 비연속적 진화를 일반적으로 지칭한다. PANCE 기술의 예는, 예를 들어 그 전체가 본 명세서에 참고로 포함된 문헌[Suzuki T. et al, Crystal structures reveal an elusive functional domain of pyrrolysyl-tRNA synthetase, Nat Chem Biol. 13(12): 1261-1266 (2017)]에 기재되어 있다. 간략하게, PANCE는, 신선한 숙주 세포(예를 들어, 이. 콜라이(E. coli) 세포)에 걸쳐, 진화될 관심 유전자를 함유하는 선택 파지(SP)의 연속 플라스크 수송을 사용한 신속한 생체 내 유도 진화(directed evolution)를 위한 기법이다. SP 내에 함유된 유전자가 연속적으로 진화하는 동안, 숙주 세포 내부의 유전자는 일정하게 유지될 수 있다. 파지 성장 후, 감염된 세포의 분취액은 숙주 이. 콜라이를 함유하는 연속된 플라스크를 형질주입시키는 데 사용될 수 있다. 이 과정은 원하는 표현형이 진화될 때까지, 예를 들어 많은 수송이 필요한 대로 반복 및/또는 계속될 수 있다.

[0327] PACE 및 PANCE를 유전자 변형 폴리펩티드에 적용하는 방법은 특히 상기 참고문헌들을 참조하여 당업자에 의해 쉽게 이해될 수 있다. 예를 들어 파지 입자를 사용하여, 예를 들어 숙주 세포의 집단에서 게놈-변형 단백질 또는 시스템의 연속 진화를 지시하기 위한 추가의 예시적인 방법은 유전자 변형 폴리펩티드의 진화된 변이체, 또는 이의 단편 또는 서브도메인을 생성하도록 적용될 수 있다. 그러한 방법의 비제한적인 예는 2009년 9월 8일 출원되고, 2010년 3월 11일 WO 2010/028347로 공개된, 국제 PCT 출원 PCT/US 2009/056194; 2011년 12월 22일 출원되고, 2012년 6월 28일 WO 2012/088381로 공개된, 국제 PCT 출원 PCT/US2011/066747; 2015년 5월 5일 발행된 미국 특허 9,023,594; 2017년 9월 26일 발행된 미국 특허 9,771,574; 2016년 7월 19일 발행된 미국 특허 9,394,537; 2015년 1월 20일 출원되고, 2015년 9월 11일 WO 2015/134121로 공개된, 국제 PCT 출원 PCT/US2015/012022; 2019년 1월 15일 발행된 미국 특허 10,179,911; 2019년 6월 14일 출원되고, 2019년 1월 31일 국제 특허 공개 WO 2019/023680으로 공개된 국제 출원 PCT/US2019/37216, 2016년 4월 15일 출원되고, 2016년 10월 20일 공개된 WO 2016/168631로 공개된 국제 PCT 출원 PCT/US2016/027795, 및 2019년 8월 23일 출원된 국제 특허 공개 PCT/US2019/47996에 기재되어 있으며, 이들 각각은 그 전체가 본 명세서에 참고로 포함된다.

[0328] 제한되지 않는 일부 예시적인 구현예에서, 진화된 변이체 유전자 변형 폴리펩티드, 이의 단편 또는 도메인의 진화 방법은: (a) 숙주 세포 집단을 관심 유전자(시작 유전자 변형 폴리펩티드, 또는 이의 단편 또는 도메인)를 포함하는 바이러스 벡터의 집단과 접촉시키는 단계를 포함하며, 여기서 (1) 숙주 세포는 바이러스 벡터에 의한 감염이 잘 되고; (2) 숙주 세포는 바이러스 입자의 생성에 필요한 바이러스 유전자를 발현하고; (3) 감염성 바이러스 입자의 생산에 필요한 적어도 하나의 바이러스 유전자의 발현이 관심 유전자의 기능에 의존하고; 및/또는 (4) 바이러스 벡터가 숙주 세포에서 단백질의 발현을 가능하게 하며, 이는 숙주 세포에 의해 바이러스 입자로 복제 및 패키징될 수 있다. 일부 구현예에서, 이 방법은 (b) 돌연변이율을 상승시키는 돌연변이를 가진 숙주 세포, 또는 이의 조합을 사용하여, (예를 들어, 돌연변이 플라스미드 또는 일부 게놈 변형 중 어느 하나, 예를 들어 프루핑-손상된 DNA 중합효소, SOS 유전자, 예컨대 UmuC, UmuD', 및/또는 RecA를 운반함으로써, 플라스미드-결합된 경우, 돌연변이는 유도성 프로모터의 제어 하에 있을 수 있음) 숙주 세포를 돌연변이 유발원과 접촉시키는 단계를 포함한다. 일부 구현예에서, 이 방법은 (c) 바이러스 입자의 바이러스 복제 및 생산을 가능하게 하는 조건 하에서 숙주 세포의 집단을 인큐베이션하는 단계를 포함하고, 여기서 숙주 세포는 숙주 세포 집단으로부터 제거되고, 신선하고 감염되지 않은 숙주 세포는 숙주 세포 집단 내로 도입되어, 이에 따라 숙주 세포의 집단을 보충하고 숙주 세포의 흐름을 생성한다. 일부 구현예에서, 세포는 관심 유전자가 돌연변이의 획득을 가능하게 하는 조건 하에서 인큐베이션된다. 일부 구현예에서, 이 방법은 (d) 바이러스 벡터의 돌연변이된 형태를 단리하고, 숙주 세포의 집단으로부터 진화된 유전자 생성물(예를 들어, 진화된 변이체 유전자 변형 폴리펩티드, 또는 이의 단편 또는 도메인)을 인코딩하는 단계를 추가로 포함한다.

[0329] 당업자는 상기 기재된 프레임워크 내에서 이용가능한 다양한 특징들을 이해할 것이다. 예를 들어 일부 구현예에서, 바이러스 벡터 또는 파지는 사상 파지(filamentous phage), 예를 들어 M13 파지, 예를 들어 M13 선택 파지이다. 특정 구현예에서, 감염성 바이러스 입자들의 생산에 필요한 유전자는 M13 유전자 III(gIII)이다. 구현예에서, 파지는 기능성 gIII이 결여될 수 있지만, 다르게는 gI, gII, gIV, gV, gVI, gVII, gVIII, gIX, 및 gX를 포함한다. 일부 구현예에서, 감염성 VSV 입자의 생성은 외피막(envelope) 단백질 VSV-G를 수반한다. 다양한 구현예는 레트로바이러스 벡터, 예를 들어 뮤린 백혈병 바이러스 벡터, 또는 렌티바이러스 벡터를 사용할 수 있다. 구현예에서, 레트로바이러스 벡터는, 예를 들어 바이러스의 천연 외피막 단백질에 대한 대체물로서 VSV-G 외피막 단백질로 효율적으로 패키징될 수 있다.

[0330] 일부 구현예에서, 숙주 세포는 적합한 횟수의 바이러스 생애 주기, 예를 들어 적어도 10, 적어도 20, 적어도 30, 적어도 40, 적어도 50, 적어도 100, 적어도 200, 적어도 300, 적어도 400, 적어도, 500, 적어도 600, 적어도 700, 적어도 800, 적어도 900, 적어도 1000, 적어도 1250, 적어도 1500, 적어도 1750, 적어도 2000, 적어도 2500, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 7500, 적어도 10000회, 또는 그 이상의 연속된 바이러스 생애 주기에 따라 인큐베이션되며, 이는 M13 파지의 예시적이며 비제한적인 예에서 바이러스 생애 주기 당 10 내지 20분이다. 유사하게, 숙주 세포가 숙주 세포의 집단으로 남아 있는 시간을 조정하기 위해 조건이 조절될 수 있으며, 예를 들어 약 10, 약 11, 약 12, 약 13, 약 14, 약 15, 약 16, 약 17, 약 18, 약 19, 약 20, 약 21, 약 22, 약 23, 약 24, 약 25, 약 30, 약 35, 약 40, 약 45, 약 50, 약 55, 약 60, 약 70, 약 80, 약 90, 약 100, 약 120, 약 150, 또는 약 180분이다. 숙주 세포 집단은 부분적으로는 숙주 세포의 밀도, 또는 일부 구현예에서 유입물 내 숙주 세포 밀도에 의해 제어될 수 있으며, 예를 들어 10^3개 세포/ml, 약 10^4개 세포/ml, 약 10^5개 세포/ml, 약 $5 \cdot 10^5$개 세포/ml, 약 10^6개 세포/ml, 약 $5 \cdot 10^6$개 세포/ml, 약 10^7개 세포/ml, 약 $5 \cdot 10^7$개 세포/ml, 약 10^8개 세포/ml, 약 $5 \cdot 10^8$개 세포/ml, 약 10^9개 세포/ml, 약 $5 \cdot 10^9$개 세포/ml, 약 10^{10}개 세포/ml, 또는 약 $5 \cdot 10^{10}$개 세포/ml이다.

[0331] *인테인*

[0332] 일부 구현예에서, 하기에 더 상세히 기재되는 바와 같이, N-말단 부분의 C-말단 부분으로의 연결을 위해 인테인-N(intN) 도메인은 본 명세서에 기재된 유전자 변형 폴리펩티드의 제1 도메인의 N-말단 부분에 융합될 수 있고, 인테인-C(intC) 도메인은 본 명세서에 기재된 유전자 변형 폴리펩티드의 제2 도메인의 C-말단 부분에 융합될 수 있으며, 이에 의해 제1 및 제2 도메인을 연결할 수 있다. 일부 구현예에서, 제1 및 제2 도메인은 DNA 결합 도메인, RNA 결합 도메인, RT 도메인, 및 엔도뉴클레아제 도메인으로부터 각각 독립적으로 선택된다.

[0333] 인테인은 자기-스플라이싱 단백질 인트론(예를 들어, 펩티드)으로서 발생할 수 있으며, 예를 들어 측접하는 N-말단과 C-말단 엑스테인(extein)(예를 들어, 연결될 단편)을 결찰한다. 인테인은 일부 경우에서, 그 자신을 절제하고 남아있는 단편(엑스테인)을 단백질 스플라이싱으로 알려진 과정에서 펩티드 결합을 이용하여 연결할 수 있는 단백질 단편을 포함한다. 인테인은 또한 "단백질 인트론"으로도 지칭된다. 그 자신을 절제하고 단백질의 나머지 부분을 연결하는 인테인의 과정은 본 명세서에서 "단백질 스플라이싱" 또는 "인테인-매개 단백질 스플라이싱"으로 명명된다.

[0334] 일부 구현예에서, 전구체 단백질의 인테인(인테인-매개 단백질 스플라이싱 전에 단백질을 함유하는 인테인)은 2개의 유전자로부터 유래한다. 그러한 인테인은 본 명세서에서 스플릿 인테인(예를 들어, 스플릿 인테인-N 및 스플릿 인테인-C)으로 지칭된다. 따라서, 인테인-기반 접근법은 제1 폴리펩티드 서열과 제2 폴리펩티드 서열을 함께 연결하는 데 사용될 수 있다. 예를 들어, 시아노박테리아에서, DNA 중합효소 III의 촉매 서브유닛인 DnaE는 2개의 별도의 유전자 dnaE-n 및 dnaE-c에 의해 인코딩된다. dnaE-n 유전자에 의해 인코딩되는 인테인-N-도메인은 제1 폴리펩티드 서열의 일부로서 위치할 때 제1 폴리펩티드 서열을 제2 폴리펩티드 서열과 연결할 수 있으며, 여기서 제2 폴리펩티드 서열은 dnaE-c 유전자에 의해 인코딩되는 것과 같은 인테인-C 도메인을 포함한다. 따라서, 일부 구현예에서, 단백질은 제1 및 제2 폴리펩티드 서열을 인코딩하는 핵산을 제공함으로써 제조될 수 있으며(예를 들어, 여기서 제1 핵산 분자는 제1 폴리펩티드 서열을 인코딩하고 제2 핵산 분자는 제2 폴리펩티드 서열을 인코딩함), 핵산은 제1 및 제2 폴리펩티드 서열의 생산을 가능하게 하고 인테인-기반 메커니즘을 통해 제1 폴리펩티드 서열의 제2 폴리펩티드 서열로의 연결을 가능하게 하는 조건 하에서 세포 내로 도입된다.

[0335] 이종성 단백질 단편을 연결하기 위한 인테인의 사용은, 예를 들어 문헌[Wood et al., J. Biol. Chem.289(21); 14512-9 (2014) (그 전체가 본 명세서에 참고로 포함됨)]에 기재되어 있다. 예를 들어, 별도의 단백질 단편에 융합될 때, 인테인 IntN 및 IntC는 서로 인식하고/하거나, 그들 자신을 스플라이스로 제거하고/하거나 그들이 융합된 단백질 단편의 측접하는 N-말단 및 C-말단 엑스테인을 동시에 결찰하여, 이에 의해 2개의 단백질 단편으로부터 전장 단백질을 재구성할 수 있다.

[0336] 일부 구현예에서, dnaE 인테인, Cfa-N(예를 들어, 스플릿 인테인-N) 및 Cfa-C(예를 들어, 스플릿 인테인-C) 인테인 쌍에 기반한 합성 인테인이 사용된다. 그러한 인테인의 예는, 예를 들어 문헌[Stevens et al., J Am Chem Soc. 2016 Feb. 24; 138(7):2162-5(그 전체가 본 명세서에 참고로 포함됨)]에 기재되어 있다. 본 개시내용에 따라 사용될 수 이는 인테인 쌍의 비제한적인 예는 다음을 포함한다: Cfa DnaE 인테인, Ssp GyrB 인테인, Ssp DnaX 인테인, Ter DnaE3 인테인, Ter ThyX 인테인, Rma DnaB 인테인 및 Cne Prp8 인테인(예를 들어, 본 명세서

에 참고로 포함된 미국 특허 8,394,604에 기재된 바와 같음).

[0337] 스플릿 Cas9를 포함하는 일부 구현예에서, 인테인-N 및 인테인-C는 스플릿 Cas9의 N-말단 부분과 스플릿 Cas9의 C-말단 부분의 연결을 위해, 스플릿 Cas9의 N-말단 부분과 스플릿 Cas9의 C-말단 부분에 각각 융합될 수 있다. 예를 들어, 일부 구현예에서, 인테인-N은 스플릿 Cas9의 N-말단 부분의 C-말단에 융합되어, 즉 N-스플릿 Cas9의 N-말단 부분]-[인테인-N]~C의 구조를 형성한다. 일부 구현예에서, 인테인-C는 스플릿 Cas9의 C-말단 부분의 N-말단에 융합되어, 즉 N-[인테인-C]~[스플릿 Cas9의 C-말단 부분]-C의 구조를 형성한다. 인테인이 융합되는 단백질(예를 들어, 스플릿 Cas9)의 연결을 위한 인테인-매개 단백질 스플라이싱의 메커니즘은 문헌[Shah et al., Chem Sci. 2014; 5(1):446-461(본 명세서에 참고로 포함됨)]에 기재되어 있다. 인테인을 설계 및 사용하는 방법은 당업계에 알려져 있고, 예를 들어 WO2020051561, WO2014004336, WO2017132580, US20150344549, 및 US20180127780에 기재되어 있으며, 이들 각각은 그 전체가 본 명세서에 참고로 포함된다.

[0338] 일부 구현예에서, 분할은 둘 이상의 단편으로의 분열을 지칭한다. 일부 구현예에서, 스플릿 Cas9 단백질 또는 스플릿 Cas9는, 2개의 별도의 뉴클레오티드 서열에 의해 인코딩된 N-말단 단편 및 C-말단 단편으로서 제공되는 Cas9 단백질을 포함한다. Cas9 단백질의 N-말단 부분 및 C-말단 부분에 상응하는 폴리펩티드는 스플라이싱되어 재구성된 Cas9 단백질을 형성할 수 있다. 구현예에서, Cas9 단백질은 단백질의 정돈되지 않은(disorderd) 영역 내에서, 예를 들어 문헌[Nishimasu et al., Cell, Volume 156, Issue 5, pp. 935-949, 2014]에 기재된 바와 같이, 또는 문헌[Jiang et al. (2016) Science 351: 867-871 및 PDB file]에 기재된 바와 같이(이들 각각은 그 전체가 본 명세서에 참고로 포함됨), 2개의 단편으로 분열된다. 정돈되지 않은 영역은, 제한없이 X-선 결정분석, NMR 분광법, 전자 현미경(예를 들어, cryoEM) 및/또는 인실리코(*in silico*) 단백질 모델링을 포함하여 당업계에서 알려진 하나 이상의 단백질 구조 결정 기법에 의해 결정될 수 있다. 일부 구현예에서, 단백질은 임의의 C, T, A 또는 S에서, 예를 들어 아미노산 A292 내지 G364, F445 내지 K483, 또는 E565 내지 T637 간의 spCas9 영역 내에서, 또는 임의의 기타 Cas9, Cas9 변이체(예를 들어, nCas9, dCas9) 또는 다른 napDNAbp에서 상응하는 위치에서, 2개의 단편으로 분열된다. 일부 구현예에서, 단백질은 SpCas9 T310, T313, A456, S469, 또는 C574에서 2개의 단편으로 분열된다. 일부 구현예에서, 단백질을 2개의 단편으로 나누는 과정은 단백질 분할(spitting)로 지칭된다.

[0339] 일부 구현예에서, 단백질 단편은 길이가 약 2 내지 1000개 아미노산(예를 들어, 2 내지 10, 10 내지 50, 50 내지 100, 100 내지 200, 200 내지 300, 300 내지 400, 400 내지 500, 500 내지 600, 600 내지 700, 700 내지 800, 800 내지 900, 또는 900 내지 1000개 아미노산)의 범위이다. 일부 구현예에서, 단백질 단편은 길이가 약 5 내지 500개 아미노산(예를 들어, 5 내지 10, 10 내지 50, 50 내지 100, 100 내지 200, 200 내지 300, 300 내지 400, 또는 400 내지 500개 아미노산)의 범위이다. 일부 구현예에서, 단백질 단편은 길이가 약 20 내지 200개 아미노산(예를 들어, 20 내지 30, 30 내지 40, 40 내지 50, 50 내지 100, 또는 100 내지 200개 아미노산)의 범위이다.

[0340] 일부 구현예에서, 유전자 변형 폴리펩티드의 일부 또는 단편은 인테인에 융합된다. 뉴클레아제는 인테인의 N-말단 또는 C-말단에 융합될 수 있다. 일부 구현예에서, 융합 단백질의 일부 또는 단편은 인테인에 융합되고 AAV 캡시드 단백질에 융합된다. 인테인, 뉴클레아제 및 캡시드 단백질은 임의의 배열로(예를 들어, 뉴클레아제-인테인-캡시드, 인테인-뉴클레아제-캡시드, 캡시드-인테인-뉴클레아제 등) 함께 융합될 수 있다. 일부 구현예에서, 인테인의 N-말단은 융합 단백질의 C-말단에 융합되고, 인테인의 C-말단은 AAV 캡시드 단백질의 N-말단에 융합된다

[0341] 일부 구현예에서, 엔도뉴클레아제 도메인(예를 들어, 닉카제 Cas9 도메인)은 인테인-N에 융합되고, RT 도메인을 포함하는 폴리펩티드는 인테인-C에 융합된다.

[0342] 인테인-N 도메인 및 적합한 인테인 C-도메인의 예시적인 뉴클레오티드 및 아미노산 서열이 하기에 제공되어 있다:

[0343] DnaE 인테인-N DNA:

[0344] TGCCTGTCATACGAAACCGAGATACTGACAGTAGAATATGGCCTTCTGCCAATCGGGAAGATTGTGGAGAAACGGATAGAATGCACAGTTTACTCTGTCGAT
AACAATGGTAACATTTATACTCAGCCAGTTGCCCAGTGGCACGACCGGGGAGAGCAGGAAGTATTCGAATACTGTCTGGAGGATGGAAGTCTCATTAGGGCC
ACTAAGGACCACAAATTTATGACAGTCGATGGCCAGATGCTGCCTATAGACGAAATCTTTGAGCGAGAGTTGGACCTCATGCGAGTTGACAACCTTCCTAAT
(서열번호 4029)

[0345] DnaE 인테인-N 단백질:

[0346] CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN (서열번호 4030)

[0347] DnaE 인테인-C DNA:

[0348] ATGATCAAGATAGCTACAAGGAAGTATCTTGGCAAACAAAACGTTTATGATATTGGAGTCGAAAGAGATCACAACTTTGCTCTGAAGAACGGATTCATAGCT TCTAAT (서열번호 4031)

[0349] DnaE 인테인-C 단백질:

[0350] MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN (서열번호 4032)

[0351] Cfa-N DNA:

[0352] TGCCTGTCTTATGATACCGAGATACTTACCGTTGAATATGGCTTCTTGCCTATTGGAAAGATTGTCGAAGAGAGAATTGAATGCACAGTATATACTGTAGAC AAGAATGGTTTCGTTTACACACAGCCCATTGCTCAATGGCACAATCGCGGCGAACAAGAAGTATTTGAGTACTGTCTCGAGGATGGAAGCATCATACGAGCA ACTAAAGATCATAAATTCATGACCACTGACGGGCAGATGTTGCCAATAGATGAGATATTCGAGCGGGGCTTGGATCTCAAACAAGTGGATGGATTG CCA (서열번호 4033)

[0353] Cfa-N 단백질:

[0354] CLSYDTEILTVEYGFLPIGKIVEERIECTVYTVDKNGFVYTQPIAQWHNRGEQEVFEYCLEDGSIIRATKDHKFMTTDGQMLPIDEIFERGLDLKQVDGLP (서열번호 4034)

[0355] Cfa-C DNA:

[0356] ATGAAGAGGACTGCCGATGGATCAGAGTTTGAATCTCCCAAGAAGAAGAGGAAAGTAAAGATAATATCTCGAAAAAGTCTTGGTACCCAAAATGTCTATGAT ATTGGAGTGGAGAAAGATCACAACTTCCTTCTCAAGAACGGTCTCGTAGCCAGCAAC (서열번호 4035)

[0357] Cfa-C 단백질:

[0358] MKRTADGSEFESPKKKRKVKIISRKSLGTQNVYDIGVEKDHNFLLKNGLVASN (서열번호 4036)

[0359] *추가 도메인*

[0360] 유전자 변형 폴리펩티드는 표적 DNA 서열 및 주형 핵산(예를 들어, 주형 RNA)에 결합하고, 표적 부위에 닉을 생성하고, 주형을 DNA 내로 작성하여(예를 들어, 역전사), 표적 부위의 변형을 생성할 수 있다. 일부 구현예에서, 추가의 도메인이 과정의 효율을 향상시키기 위해 폴리펩티드에 첨가될 수 있다. 일부 구현예에서, 유전자 변형 폴리펩티드는 역전사된 DNA를 표적 부위의 DNA에 연결시키기 위한 추가의 DNA 결찰 도메인을 함유할 수 있다. 일부 구현예에서, 폴리펩티드는 이종성 RNA-결합 도메인을 포함할 수 있다. 일부 구현예에서, 폴리펩티드는 5'에서 3' 엑소뉴클레아제 활성을 갖는 도메인을 포함할 수 있다(예를 들어, 여기서 5'에서 3' 엑소뉴클레아제 활성은, 예를 들어 원래의 게놈 서열에 대한 변경을 위해, 표적 부위의 변경의 복구를 증가시킴). 일부 구현예에서, 폴리펩티드는 3'에서 5' 엑소뉴클레아제 활성, 예를 들어 교정 활성을 갖는 도메인을 포함할 수 있다. 일부 구현예에서, 라이팅 도메인, 예를 들어 RT 도메인은 3'에서 5' 엑소뉴클레아제 활성, 예를 들어 교정 활성을 갖는다.

[0361] ***주형 핵산***

[0362] 본 명세서에 기재된 유전자 변형 시스템은 주형 핵산 서열을 이용하여 숙주 표적 DNA 부위를 변형시킬 수 있다. 일부 구현예에서, 본 명세서에 기재된 유전자 변형 시스템은 표적-프라이밍된 역전사(TPRT)에 의해 RNA 서열 주형을 숙주 표적 DNA 부위로 전사한다. RNA 서열 주형의 숙주 게놈 내로의 역전사를 통해 DNA 서열(들)을 변형시킴으로써, 유전자 변형 시스템은(예를 들어, CRISPR 시스템과 달리) 숙주 세포 내로 도입되는 외인성 DNA 서열에 대한 필요 없이 목적 서열을 표적 게놈 내로 삽입할 뿐만 아니라, 외인성 DNA 삽입 단계를 제거할 수 있다. 유전자 변형 시스템은 표적 게놈으로부터의 서열을 결실 또는 목적 서열을 이용하여 치환을 도입할 수도 있다. 따라서, 유전자 변형 시스템은 목적 서열, 예를 들어 이종성 유전자 코딩 및/또는 기능 정보를 포함하는 서열을 함유하는 맞춤화된 RNA 서열 주형의 이용을 위한 플랫폼을 제공한다.

[0363] 일부 구현예에서, 주형 핵산은 유전자 변형 폴리펩티드에 결합하는 하나 이상의 서열(예를 들어, 2개의 서열)을 포함한다.

[0364] 일부 구현예에서, 주형 핵산은 동일한 가닥에 리보뉴클레오티드와 데옥시리보뉴클레오티드 잔기를 둘 다 갖는

하이브리드를 포함한다.

[0365] 일부 구현예에서, 본 명세서에 기재된 시스템 또는 방법은 단일 주형 핵산(예를 들어, 주형 RNA)을 포함한다. 일부 구현예에서, 본 명세서에 기재된 시스템 또는 방법은 복수의 주형 핵산(예를 들어, 주형 RNA)을 포함한다. 예를 들어, 본 명세서에 기재된 시스템은 (예를 들어, 5'에서 3'으로) 유전자 변형 폴리펩티드(예를 들어, DNA-결합 도메인 및/또는 엔도뉴클레아제 도메인, 예를 들어 gRNA)에 결합하는 서열 및 표적 부위(예를 들어, 표적 게놈 내 부위의 제2 가닥)에 결합하는 서열을 포함하는 제1 RNA, 및 선택적으로 유전자 변형 폴리펩티드에 결합하는(예를 들어, RT 도메인에 특이적으로 결합하는) 서열, 이종 목적 서열, 및 PBS 서열을 포함하는 제2 RNA(예를 들어, 주형 RNA)를 포함한다. 일부 구현예에서, 시스템이 복수의 핵산을 포함하는 경우, 각각의 핵산은 접합 도메인을 포함한다. 일부 구현예에서, 접합 도메인은, 예를 들어 상보적 서열의 혼성화에 의해 핵산 분자의 회합을 가능하게 한다. 예를 들어, 일부 구현예에서, 제1 RNA는 제1 접합 도메인을 포함하고 제2 RNA는 제2 접합 도메인을 포함하며, 제1 및 제2 접합 도메인은, 예를 들어 엄격한 조건 하에서 서로 혼성화할 수 있다. 일부 구현예에서, 혼성화를 위한 엄격한 조건은 4× 염화나트륨/나트륨 시트레이트(SSC) 중, 약 65C에서의 혼성화에 이어 약 65C에서 1xSSC 내 세척을 포함한다.

[0366] 일부 구현예에서, 주형 핵산은 RNA를 포함한다. 일부 구현예에서, 주형 핵산은 DNA(예를 들어, 단일 가닥 또는 이중 가닥 DNA)를 포함한다. 일부 구현예에서, 주형 핵산은 동일한 가닥에 리보뉴클레오티드와 데옥시리보뉴클레오티드 잔기를 둘 다 갖는 하이브리드를 포함한다.

[0367] 일부 구현예에서, 주형 핵산은 표적 핵산에 대해 상동성을 갖는 하나 이상(예를 들어, 2개)의 상동성 도메인을 포함한다. 일부 구현예에서, 상동성 도메인은 길이가 약 10 내지 20, 20 내지 50, 또는 50 내지 100개 뉴클레오티드이다.

[0368] 일부 구현예에서, 주형 RNA는, 예를 들어 유전자 변형 폴리펩티드를 관심 표적 부위로 지시하기 위한 gRNA 서열을 포함할 수 있다. 일부 구현예에서, 주형 RNA는 (예를 들어, 5'에서 3'으로) (i) 선택적으로 표적 부위에 결합하는 gRNA 스페이서(예를 들어, 표적 게놈 내 부위의 제2 가닥), (ii) 선택적으로 본 명세서에 기재된 폴리펩티드에 결합하는 gRNA 스캐폴드(예를 들어, 유전자 변형 폴리펩티드 또는 Cas 폴리펩티드), (iii) 돌연변이 영역을 포함하는 이종 목적 서열(선택적으로 이종 목적 서열은 5'에서 3'으로 제1 상동성 영역, 돌연변이 영역, 및 제2 상동성 영역을 포함함), 및 (iv) 3' 표적 상동성 도메인을 포함하는 프라이머 결합 부위(PBS) 서열을 포함한다.

[0369] 본 명세서에 기재된 게놈 편집 시스템의 주형 핵산(예를 들어, 주형 RNA) 구성요소는 통상적으로 시스템의 유전자 변형 폴리펩티드에 결합할 수 있다. 일부 구현예에서 주형 핵산(예를 들어, 주형 RNA)은 유전자 변형 폴리펩티드에 결합할 수 있는 3' 영역을 갖는다. 결합 영역, 예를 들어 3' 영역은 시스템의 유전자 변형 폴리펩티드에 결합할 수 있는, 예를 들어 적어도 1, 2 또는 3개의 헤어핀 루프를 갖는, 구조화된 RNA 영역일 수 있다. 결합 영역은 주형 핵산(예를 들어, 주형 RNA)을 임의의 폴리펩티드 모듈과 회합시킬 수 있다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 결합 영역은 폴리펩티드에서 RNA-결합 도메인과 회합할 수 있다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 결합 영역은 유전자 변형 폴리펩티드의 역전사 도메인과 회합할 수 있다(예를 들어, RT 도메인에 특이적으로 결합할 수 있음). 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 폴리펩티드의 DNA 결합 도메인, 예를 들어 Cas9-유래 DNA 결합 도메인과 회합하는 gRNA와 회합할 수 있다. 일부 구현예에서, 결합 영역은 또한 DNA 표적 인식, 예를 들어 표적 DNA 서열과 혼성화하고 폴리펩티드, 예를 들어 Cas9 도메인에 결합하는 gRNA를 제공할 수 있다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 폴리펩티드의 다중 구성요소, 예를 들어 DNA 결합 도메인 및 역전사 도메인과 회합할 수 있다.

[0370] 일부 구현예에서 주형 RNA는 3' 말단에서 폴리-A 테일을 갖는다. 일부 구현예에서 주형 RNA는 3' 말단에서 폴리-A 테일을 갖지 않는다.

[0371] 일부 구현예에서, 주형 RNA는 표적 세포의 게놈 DNA에서(예를 들어, 대상체에서, 예를 들어 생체 외 또는 생체 내에서, 예를 들어 표적 조직 또는 기관에서) 주어진 돌연변이를 교정하도록 맞춤화될 수 있다. 예를 들어, 돌연변이는 야생형 서열에 비해 질환-연관 돌연변이일 수 있다. 이론에 구속되고자 함이 없이, 임의의 주어진 표적 부위 및 편집물은 다양한 편집 효율 및 충실도를 초래할 유전자 변형 시스템에 사용하기 위한 다수의 가능한 주형 RNA 분자를 가질 것이다. 이러한 스크리닝 부담을 부분적으로 감소시키기 위해, 경험적 파라미터의 세트는 주형 RNA 또는 이의 일부의 최적의 초기 인실리코 설계를 보장하는 데 도움이 된다. 비제한적인 예시적 예로서, 선택된 돌연변이에 대해 다음 설계 파라미터가 이용될 수 있다. 일부 구현예에서, 설계는 표적 영역의 역할을 하는 돌연변이의 양쪽에서 서열에 측접하는 대략 500 bp(예를 들어, 최대 50, 100, 150, 200, 250, 300, 350,

400, 450, 500, 550, 600, 650, or 700 bp, 및 선택적으로 적어도 20, 30, 40, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 또는 650 bp)를 획득함으로써 개시된다. 일부 구현예에서, 주형 핵산은 gRNA를 포함한다. 일부 구현예에서, gRNA는 표적 부위에 결합하는 서열(예를 들어, CRISPR 스페이서)을 포함한다. 일부 구현예에서, 주형 핵산을 표적 영역으로 표적화하는 데 사용하기 위한 표적 부위에 결합하는 서열(예를 들어, CRISPR 스페이서)은 사용되는 특정 유전자 변형 폴리펩티드(예를 들어, CRISPR/Cas 도메인을 포함하는, 예를 들어 엔도뉴클레아제 도메인 또는 라이팅 도메인)를 고려함으로써 선택된다(예를 들어, Cas9의 경우, 20개 뉴클레오티드 gRNA 결합 영역의 바로 3'에 있는 NGG의 프로토스페이서-인접 모티프(PAM)). 일부 구현예에서, CRISPR 스페이서는 먼저 PAM이 유전자 변형 시스템 유도 편집에 의해 붕괴될 것인지 여부에 의해 순위를 매김으로써 선택된다. 일부 구현예에서, PAM의 붕괴는 편집 효율성을 증가시킬 것이다. 일부 구현예에서, PAM은 유전자 변형 동안 표적 부위에서 침묵 돌연변이(예를 들어, 존재하는 경우, 표적 핵산 서열에 의해 인코딩된 아미노산 잔기를 변경하지 않는 돌연변이)를 (예를 들어, 게놈 DNA에서 표적 부위에 또 다른 변형의 일부 또는 첨가로서) 도입함에 의해서도 붕괴될 수 있다. 일부 구현예에서, CRISPR 스페이서는 원하는 편집 위치에 대한 그들의 상응하는 게놈 부위 인접에 의해 서열의 순위를 매김으로써 선택된다. 일부 구현예에서, gRNA는 gRNA 스캐폴드를 포함한다. 일부 구현예에서, 사용된 gRNA 스캐폴드는 표준 스캐폴드(예를 들어, Cas9의 경우, 5'-GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC-3')일 수 있거나, 하나 이상의 뉴클레오티드 치환을 함유할 수 있다. 일부 구현예에서, 이종 목적 서열은 유전자 변형에 의해 표적 부위 내로 작성될 수 있는 임의의 삽입, 치환 또는 결실을 제외하고, 적어도 90% 동일성, 예를 들어, 적어도 90%, 95%, 98%, 99%, 또는 100% 동일성을 갖거나, 또는 제1 가닥 닉의 3'(예를 들어, 제1 가닥 닉의 바로 3' 또는 제1 가닥 닉의 3'에서 최대 1, 2, 3, 4, 또는 5개의 뉴클레오티드)에서 표적 부위에 대해 동일성이 없는 1, 2, 3, 4, 또는 5개 이하의 위치를 포함한다. 일부 구현예에서, 3' 표적 상동성 도메인은 적어도 90% 동일성, 예를 들어, 적어도 90%, 95%, 98%, 99%, 또는 100% 동일성을 함유하거나, 또는 제1 가닥 닉의 5'(예를 들어, 제1 가닥 닉의 바로 5' 또는 제1 가닥 닉의 3'에서 최대 1, 2, 3, 4, 또는 5개의 뉴클레오티드)에서 표적 부위에 대해 동일성이 없는 1, 2, 3, 4, 또는 5개 이하의 위치를 포함한다.

[0372] 일부 구현예에서, 주형 핵산은 주형 RNA이다. 일부 구현예에서, 주형 RNA는 하나 이상의 변형된 뉴클레오티드를 포함한다. 예를 들어, 일부 구현예에서, 주형 RNA는 하나 이상의 데옥시리보뉴클레오티드를 포함한다. 일부 구현예에서, 주형 RNA의 영역은 DNA 뉴클레오티드에 의해 대체되어, 예를 들어 분자의 안정성을 향상시킨다. 예를 들어, 주형의 3' 말단은 DNA 뉴클레오티드를 포함할 수 있는 한편, 주형의 나머지는 역전사될 수 있는 RNA 뉴클레오티드를 포함한다. 예를 들어 일부 구현예에서, 이종 목적 서열은 주로 또는 전체적으로 RNA 뉴클레오티드(예를 들어, 적어도 90%, 95%, 98%, 또는 99% RNA 뉴클레오티드)로 구성된다. 일부 구현예에서, PBS 서열은 주로 또는 전체적으로 DNA 뉴클레오티드(예를 들어, 적어도 90%, 95%, 98%, 또는 99% DNA 뉴클레오티드)로 구성된다. 다른 구현예에서, 게놈 내로 작성하기 위한 이종 목적 서열은 DNA 뉴클레오티드를 포함할 수 있다. 일부 구현예에서, 주형 내 DNA 뉴클레오티드는 DNA-의존성 DNA 중합효소 활성이 가능한 도메인에 의해 게놈 내로 카피된다. 일부 구현예에서, DNA-의존성 DNA 중합효소 활성은 폴리펩티드 내 DNA 중합효소 도메인에 의해 제공된다. 일부 구현예에서, DNA-의존성 DNA 중합효소 활성은 또한 DNA-의존성 DNA 중합, 예를 들어 제2 가닥 합성이 가능한 역전사효소 도메인에 의해 제공된다. 일부 구현예에서, 주형 분자는 DNA 뉴클레오티드만으로 구성된다. 일부 구현예에서, 주형 핵산은 동일한 가닥에 리보뉴클레오티드와 데옥시리보뉴클레오티드 잔기를 둘 다 갖는 하이브리드를 포함한다.

[0373] 일부 구현예에서, 본 명세서에 기재된 시스템은 본 명세서에 기재된 주형 RNA의 서열을 함께 포함하는 2개의 핵산을 포함한다. 일부 구현예에서, 2개의 핵산은 서로 비공유적으로 회합되어 있으며, 예를 들어 (예를 들어, 염기쌍 형성을 통해) 서로 직접적으로 회합되어 있거나, 하나 이상의 추가 분자를 포함하는 복합체의 일부로서 간접적으로 회합되어 있다.

[0374] 본 명세서에 기재된 주형 RNA는 5'에서 3'으로 (1) gRNA 스페이서; (2) gRNA 스캐폴드; (3) 이종 목적 서열 (4) 프라이머 결합 부위(PBS) 서열을 포함할 수 있다. 이들 구성요소 각각은 이제 보다 상세히 기재된다.

[0375] *gRNA 스페이서 및 gRNA 스캐폴드*

[0376] 본 명세서에 기재된 주형 RNA는 유전자 변형 시스템을 표적 핵산으로 지시하는 gRNA 스페이서, 및 주형 RNA와 유전자 변형 폴리펩티드의 Cas 도메인의 회합을 촉진하는 gRNA 스캐폴드를 포함할 수 있다. 본 명세서에 기재된 시스템은 또한 주형 핵산의 일부가 아닌 gRNA를 포함할 수 있다. 예를 들어, gRNA 스페이서 및 gRNA 스캐폴드를 포함하지만, 이종 목적 서열 또는 PBS 서열을 포함하지 않는 gRNA는, 예를 들어 본 명세서에서 "제2 가닥 닉 생

성"라는 표제의 섹션에 기재된 바와 같이, 예를 들어 제2 가닥 닉 생성을 유도하는 데 사용될 수 있다.

[0377] 일부 구현예에서, gRNA는 CRISPR-연관 단백질 결합에 참여하는 스캐폴드 서열, 및 게놈 표적을 위한 사용자-정의된 약 20개의 뉴클레오티드 표적화 서열로 구성되는 짧은 합성 RNA이다. 완전한 gRNA의 구조는 문헌[Nishimasu et al. Cell 156, P935-949 (2014)]에 의해 설명되었다. gRNA(단일-가이드 RNA의 경우 sgRNA로도 지칭됨)는 인공 테트라루프에 의해 연결된 crRNA- 및 tracrRNA-유래 서열로 이루어진다. crRNA 서열은 가이드(20 nt) 및 반복(12 nt) 영역으로 나뉠 수 있는 반면, tracrRNA 서열은 항-반복(14 nt) 및 3개의 tracrRNA 스템 루프로 나뉠 수 있다(Nishimasu et al. Cell 156, P935-949 (2014)). 실제로, 가이드 RNA 서열은 17 내지 24개의 뉴클레오티드(예를 들어, 19, 20, 또는 21개의 뉴클레오티드) 길이를 갖고 표적화된 핵산 서열에 상보적이 되도록 일반적으로 설계된다. 효과적인 가이드 RNA의 설계에 사용하기 위한 맞춤형 gRNA 생성기 및 알고리즘이 상업적으로 구매 가능하다. 일부 구현예에서, gRNA는 천연 CRISPR 시스템 유래의 2개의 RNA 구성요소, 예를 들어 crRNA 및 tracrRNA를 포함한다. 당업계에 잘 알려진 바와 같이, gRNA는 또한 (뉴클레아제에 결합하기 위한) tracrRNA 및 (편집/결합을 위해 뉴클레아제를 표적화된 서열로 가이드하는) 적어도 하나의 crRNA 둘 다로부터의 서열을 함유하는 키메라 단일 가이드 RNA(sgRNA)를 포함할 수 있다. 화학적으로 변형된 sgRNA가 또한 CRISPR-연관 단백질과 함께 사용하기에 효과적인 것으로 입증되었으며; 예를 들어, 문헌[Hendel et al. (2015) Nature Biotechnol., 985 - 991]을 참조한다. 일부 구현예에서, gRNA 스페이서는 표적 유전자와 연관된 DNA 서열에 상보적인 핵산 서열을 포함한다.

[0378] 일부 구현예에서, gRNA를 포함하는 주형 핵산, 예를 들어 주형 RNA의 영역은 표적 DNA에 결합된 gRNA의 덜 감긴(underwound) 리본-유사 구조를 채용한다(예를 들어, 문헌[Mulepati et al. Science 19 Sep 2014:Vol. 345, Issue 6203, pp. 1479-1484]에 기재된 바와 같음) 이론에 구속되고자 함이 없이, 이러한 비-정규 구조는 RNA-DNA 하이브리드 중 매 6 번째 뉴클레오티드의 회전에 의해 용이하게 되는 것으로 생각된다. 따라서, 일부 구현예에서, gRNA를 포함하는 주형 핵산, 예를 들어 주형 RNA의 영역은 어느 정도의 간격, 예를 들어 매 6 번째 염기에서 표적 부위와의 증가된 미스매칭을 용인할 수 있다. 일부 구현예에서, 표적 부위에 대해 상동성을 포함하는 gRNA를 포함하는 주형 핵산, 예를 들어 주형 RNA의 영역은 규칙적인 간격, 예를 들어 매 6 번째 염기에서 동요 위치를 가질 수 있으며, 이는 표적 부위와의 염기쌍을 필요로 하지 않는다.

[0379] 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은, 예를 들어 유전자 변형 폴리펩티드의 Cas9도메인(표 3)에 적절한 길이의 gRNA 스페이서 서열을 포함하는, 예를 들어 5' 말단에서, 표적 부위에 대하여 적어도 80%, 85%, 90%, 95%, 99%, 또는 100% 상동성의 적어도 15, 16, 17, 18, 19, 20, 21, 22, 23, 또는 24개 염기를 갖는다.

[0380] 일부 구현예에서, 향상된 활성을 갖는 Cas9 유도체가 유전자 변형 폴리펩티드에 사용될 수 있다. 일부 구현예에서, Cas9 유도체는 HNH 엔도뉴클레아제 도메인의 활성을 개선시키는 돌연변이, 예를 들어 SpyCas9 R221K, N394K, 또는 R-루프 형성을 개선시키는 돌연변이, 예를 들어 SpyCas9 L1245V를 포함할 수 있거나, 그러한 돌연변이의 조합, 예를 들어 SpyCas9 R221K/N394K, SpyCas9 N394K/L1245V, SpyCas9 R221K/L1245V, 또는 SpyCas9 R221K/N394K/L1245V를 포함할 수 있다(예를 들어, 문헌[Spencer and Zhang Sci Rep 7:16836 (2017)] 참조, Cas 유도체 및 돌연변이를 포함함, 이는 본 명세서에 참고로 포함됨). 일부 구현예에서, Cas9 유도체는 본 명세서에 기재된 돌연변이 중 하나 이상의 유형, 예를 들어 PAM-변형 돌연변이, 단백질 안정화 돌연변이, 활성 향상 돌연변이, 및/또는 모 효소에 비해 하나 또는 2개 엔도뉴클레아제 도메인을 부분적으로 또는 완전히 비활성화시키는 돌연변이(예를 들어, 표적 DNA의 하나 또는 두 가닥 모두에 대한 엔도뉴클레아제 활성을 없애는 하나 이상의 돌연변이, 예를 들어 닉카제 또는 촉매적 사멸 효소)를 포함할 수 있다. 일부 구현예에서, 본 명세서에 기재된 시스템에 사용되는 Cas9 효소는 촉매 효율을 개선시키는 돌연변이(예를 들어, SpyCas9 R221K, N394K, 및/또는 L1245V)에 추가적으로, 효소에 대한 닉카제 활성을 부여하는 돌연변이(예를 들어, SpyCas9 N863A 또는 H840A)를 포함할 수 있다. 일부 구현예에서, 본 명세서에 기재된 시스템에 사용되는 Cas9 효소는 촉매 효율을 개선시키는 R221K 및 N394K 돌연변이에 추가적으로, 닉카제 활성을 부여하는 N863A 돌연변이를 추가로 포함하는 SpyCas9 효소 또는 유도체이다.

[0381] 표 8은 유전자 변형을 위해 표 3에 열거된 Cas 변이체에 적용하기 위하여 gRNA 및/또는 주형 RNA를 설계하기 위한 구성요소를 정의하는 파라미터를 제공한다. 절단 부위는 검증되거나 예측된 프로토스페이서 인접 모티프(PAM) 요건, (PAM 부위의 가장 상류 염기에 대해) 절단 부위의 검증되거나 예측된 위치를 나타낸다. 주어진 효소에 대한 gRNA는 crRNA, 테트라루프, 및 tracrRNA 서열을 연쇄화(concatenating)시키고, 표적 부위에서 프로토스페이서와 일치하는 스페이서(최소) 및 스페이서(최대) 내 길이의 5' 스페이서를 추가로 첨가함으로써 조립될 수 있다. 또한, 표적에서 ssDNA의 예측된 위치는, 표적 프라이밍된 역전사를 개시하기 위해, 닉의 바로 5'에서 서열에 어닐링할 수 있는 주형 RNA의 PBS 서열을 설계하는데 중요하다. 일부 구현예에서, 본 명세서에 기재

된 gRNA 스캐폴드는 5'에서 3' 방향으로, 표 8의 crRNA, 표 8의 동일한 행의 테트라루프, 및 표 8의 동일한 행의 tracrRNA를 포함하는 핵산 서열, 또는 이에 대하여 적어도 70%, 80%, 85%, 90%, 95%, 또는 99% 동일성을 갖는 서열을 포함한다. 일부 구현예에서, 스캐폴드를 포함하는 gRNA 또는 주형 RNA는 표 8의 동일한 행에 나타낸 스페이서(최소) 및 스페이서(최대) 내의 길이를 갖는 gRNA 스페이서를 추가로 포함한다. 일부 구현예에서, 표 8에 따른 서열을 갖는 gRNA 또는 주형 RNA는 유전자 변형 폴리펩티드를 추가로 포함하는 시스템에 포함되며, 여기서 유전자 변형 폴리펩티드는 표 8의 동일한 행에 기재된 Cas 도메인을 포함한다.

[0382] **표 8 유전자 변형 시스템에서 표 3에 열거된 Cas 변이체에 적용하기 위하여 gRNA 및/또는 주형 RNA를 설계하기 위한 구성요소를 정의하는 파라미터**

표 8

[0383]

변이체	PAM(들)	절단	단계	스페이서(최소)	스페이서(최대)	crRNA	테트라루프	tracrRNA
Nme2Cas9	NNNNCC	-3	1	22	24	GTTGTAGCTCCCTTTCTCATTTCG	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
PpnCas9	NNNNRTT		1	21	24	GTTGTAGCTCCCTTTTTCATTTCGC	GAAA	GCGAAATGAAAAACGTTGTTACAATAAGAGATGAATTTCTCGCAAAGCTCTGCCTCTTGAAATTTCGGTTTCAAGAGGCATC
SauCas9	NNGRR;NNGRRT	-3	1	21	23	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauCas9-KKH	NNNRR;NNNRRT	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauriCas9	NNGG	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
SauriCas9-KKH	NNRG	-3	1	21	21	GTTTTAGTACTCTG	GAAA	CAGAATCTACTAAAACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
ScaCas9-Sc++	NNG	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9	NGG	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-NG	NG (NGG=NGA=NGT>NGC)	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-SpRY	NRN>NYN	-3	1	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
St1Cas9	NNAGAAW>NNAGGAW=NNGGAAW	-3	1	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
BlatCas9	NNNNCNAA>NNNNCNDD>NNNNC	-3	1	19	23	GCTATAGTTCCTTACT	GAAA	GGTAAGTTGCTATAGTAAGGGCAACAGACCCGAGGCGTTGGGGATCGCCTAGCCCGTGTTTACGGGCTCTCCCCATATTCAAAATAATGACAGACGAGCACCTTGGAGCATTTATCTCCGAGGTGCT
cCas9-v16	NNVACT;NNVATGM;NNVATT;NNVGCT;NNVGTG;NNVGTT	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
cCas9-v17	NNVRRN	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA

cCas9-v21	NNVACT;NNVATGM;NNVATT;NNVGCT;NNVGTG;NNVGTT	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
cCas9-v42	NNVRRN	-3	2	21	21	GTCTTAGTACTCTG	GAAA	CAGAATCTACTAAGACAAGGCAAAATGCCGTGTTTATCTCGTCAACTTGTTGGCGAGA
CdiCas9	NNRHHHY;NNRAAAY		2	22	22	ACTGGGGTTCAG	GAAA	CTGAACCTCAGTAAGCATTGGCTCGTTTCCAATGTTGATTGCTCCGCCGGTGCTCCTTATTTTTAAGGGCGCCGGC
CjeCas9	NNNNRYAC	-3	2	21	23	GTTTTAGTCCCT	GAAA	AGGGACTAAAATAAAGAGTTTGCGGGACTCTGCGGGGTTACAATCCCCTAAAACCGC
GeoCas9	NNNNCRAA		2	21	23	GTCATAGTTCCCCTGA	GAAA	TCAGGGTTACTATGATAAGGGCTTTCTGCCTAAGGCAGACTGACCCGCGGCGTTGGGGATCGCCTGTCGCCCGCTTTTGGCGGGCATTCCCCATCCTT
iSpyMacCas9	NAAN	-3	2	19	21	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
NmeCas9	NNNNGAYT;NNNNGYTT;NNNNGAYA;NNNNGTCT	-3	2	20	24	GTTGTAGCTCCCTTTCTCATTTCG	GAAA	CGAAATGAGAACCGTTGCTACAATAAGGCCGTCTGAAAAGATGTGCCGCAACGCTCTGCCCCTTAAAGCTTCTGCTTTAAGGGGCATCGTTTA
ScaCas9	NNG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
ScaCas9-HiFi-Sc++	NNG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-3var-NRRH	NRRH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-3var-NRTH	NRTH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-3var-NRCH	NRCH	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-HF1	NGG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-QQR1	NAAG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-SpG	NGN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VQR	NGAN	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-VRER	NGCG	-3	2	20	20	GTTTTAGAGCTA	GAAA	TAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas	NG;GAA;GAT	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
SpyCas9-xCas-NG	NG	-3	2	20	20	GTTTAAGAGCTATGCTG	GAAA	CAGCATAGCAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC
St1Cas9-CNRZ1066	NNACAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT

St1Cas9-LMG1831	NNGCAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-MTH17CL396	NNAAAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
St1Cas9-TH1477	NNGAAA	-3	2	20	20	GTCTTTGTACTCTG	GTAC	CAGAAGCTACAAAGATAAGGCTTCATGCCGAAATCAACACCCTGTCATTTTATGGCAGGGTGTTTT
sRGN3.1	NNGG		1	21	23	GTTTTAGTACTCTG	GAAA	CAGAATCTACTGAAACAAGACAATATGTCGTGTTTATCCCATCAATTTATTGGTGGGATTTT
sRGN3.3	NNGG		1	21	23	GTTTTAGTACTCTG	GAAA	CAGAATCTACTGAAACAAGACAATATGTCGTGTTTATCCCATCAATTTATTGGTGGGATTTT

[0384] 추가로, 말단 U 및 T는 선택적으로 tracrRNA 서열에 추가되거나 제거될 수 있고 RNA로서 제공되는 경우 변형되거나 변형되지 않을 수 있다. 예에 의해 구속되고자 함이 없이, 표 8에 예시된 것에 대해 대안적인 gRNA 스캐폴드 서열의 형태는 또한 표 4에 예시된 다양한 Cas9 효소 또는 이의 유도체, 예를 들어 뉴클레오티드 첨가, 치환, 또는 결실이 있는 대체 gRNA 스캐폴드 서열, 예를 들어 스템-루프 구조가 추가되거나 제거된 서열과 함께 기능할 수 있다. 본 명세서에서 gRNA 스캐폴드 서열은 주어진 시스템, Cas-RT 융합 폴리펩티드, 적응증, 표적 돌연변이, 주형 RNA, 또는 전달 비히클에 대해 유사하게 최적화될 수 있는 유전자 변형 시스템의 구성요소를 나타내는 것으로 고려된다.

[0385] 본 명세서에서, RNA 서열(예를 들어, 주형 RNA 서열)이 티민(T)을 포함하는 특정 서열(예를 들어, 표 8의 서열 또는 이의 일부)을 포함한다고 말할 때, 물론 RNA 서열은 T 대신에 우라실(U)을 포함할 수 있다(그리고 흔히 포함한다)고 이해된다. 예를 들어, RNA 서열은 표 8의 서열에서 T로 나타낸 모든 위치에서 U를 포함할 수 있다. 더 구체적으로, 본 개시내용은 표 8의 모든 gRNA 스캐폴드 서열에 따른 RNA 서열을 제공하며, 여기서 RNA 서열은 표 8의 서열에서 각각의 T 대신에 U를 갖는다.

[0386] *이종 목적 서열*

[0387] 본 명세서에 기재된 주형 RNA는 유전자 변형 폴리펩티드가 역전사를 위한 주형으로서 사용하여, 원하는 서열을 표적 핵산에 작성할 수 있는 이종 목적 서열을 포함할 수 있다. 일부 구현예에서, 이종 목적 서열은 5'에서 3'으로 사후 편집 상동성 영역, 돌연변이 영역, 및 사전 편집 상동성 영역을 포함한다. 이론에 구속되고자 함이 없이, 주형 RNA에 대해 역전사를 수행하는 RT는 먼저 사전 편집 상동성 영역, 그 돌연변이 영역, 그 다음 사후 편집 상동성 영역을 역전사하여, 이에 의해 양쪽에 상동성 영역이 있는 원하는 돌연변이를 포함하는 DNA 가닥을 형성한다.

[0388] 일부 구현예에서, 이종 목적 서열은 길이가 적어도 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 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, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 120, 140, 160, 180, 200, 500, 또는 1,000개의 뉴클레오티드(nt), 또는 길이가 적어도 1, 1.5, 2, 2.5, 3, 3.5, 4, 4.5, 5, 5.5, 6, 6.5, 7, 7.5, 8, 8.5, 9, 9.5, 또는 10 킬로베이스이다. 일부 구현예에서, 이종 목적 서열은 길이가 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 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, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 120, 140, 160, 180, 200, 500, 1,000, 또는 2000개 이하의 뉴클레오티드(nt), 또는 길이가 단지 20, 15, 10, 9, 8, 7, 6, 5, 4, 또는 3 킬로베이스 이하이다. 일부 구현예에서, 이종 목적 서열은 길이가 30 내지 1000, 40 내지 1000, 50 내지 1000, 60 내지 1000, 70 내지 1000, 74 내지 1000, 75 내지 1000, 76 내지 1000, 77 내지 1000, 78 내지 1000, 79 내지 1000, 80 내지 1000, 85 내지 1000, 90 내지 1000, 100 내지 1000, 120 내지 1000, 140 내지 1000, 160 내지 1000, 180 내지 1000, 200 내지 1000, 500 내지 1000, 30 내지 500, 40 내지 500, 50 내지 500, 60 내지 500, 70 내지 500, 74 내지 500, 75 내지 500, 76 내지 500, 77 내지 500, 78 내지 500, 79 내지 500, 80 내지 500, 85 내지 500, 90 내지 500, 100 내지 500, 120 내지 500, 140 내지 500, 160 내지 500, 180 내지 500, 200 내지 500, 30 내지 200, 40 내지 200, 50 내지 200, 60 내지 200, 70 내지 200, 74 내지 200, 75 내지 200, 76 내지 200, 77 내지 200, 78 내지 200, 79 내지 200, 80 내지 200, 85 내지 200,

90 내지 200, 100 내지 200, 120 내지 200, 140 내지 200, 160 내지 200, 180 내지 200, 30 내지 100, 40 내지 100, 50 내지 100, 60 내지 100, 70 내지 100, 74 내지 100, 75 내지 100, 76 내지 100, 77 내지 100, 78 내지 100, 79 내지 100, 80 내지 100, 85 내지 100, 또는 90 내지 100개의 뉴클레오티드(nt), 또는 길이가 1 내지 20, 1 내지 15, 1 내지 10, 1 내지 9, 1 내지 8, 1 내지 7, 1 내지 6, 1 내지 5, 1 내지 4, 1 내지 3, 1 내지 2, 2 내지 20, 2 내지 15, 2 내지 10, 2 내지 9, 2 내지 8, 2 내지 7, 2 내지 6, 2 내지 5, 2 내지 4, 2 내지 3, 3 내지 20, 3 내지 15, 3 내지 10, 3 내지 9, 3 내지 8, 3 내지 7, 3 내지 6, 3 내지 5, 3 내지 4, 4 내지 20, 4 내지 15, 4 내지 10, 4 내지 9, 4 내지 8, 4 내지 7, 4 내지 6, 4 내지 5, 5 내지 20, 5 내지 15, 5 내지 10, 5 내지 9, 5 내지 8, 5 내지 7, 5 내지 6, 6 내지 20, 6 내지 15, 6 내지 10, 6 내지 9, 6 내지 8, 6 내지 7, 7 내지 20, 7 내지 15, 7 내지 10, 7 내지 9, 7 내지 8, 8 내지 20, 8 내지 15, 8 내지 10, 8 내지 9, 9 내지 20, 9 내지 15, 9 내지 10, 10 내지 15, 10 내지 20, 또는 15 내지 20 킬로베이스이다. 일부 구현예에서, 이종 목적 서열은 길이가 10 내지 100, 10 내지 90, 10 내지 80, 10 내지 70, 10 내지 60, 10 내지 50, 10 내지 40, 10 내지 30, 또는 10 내지 20개의 nt, 예를 들어 길이가 10 내지 80, 10 내지 50, 또는 10 내지 20개의 nt, 예를 들어 길이 약 10 내지 20개의 nt이다. 일부 구현예에서, 이종 목적 서열은 길이가 8 내지 30, 9 내지 25, 10 내지 20, 11 내지 16, 또는 12 내지 15개의 뉴클레오티드이고, 예를 들어 길이가 11 내지 16개의 nt이다. 이론에 구속되고자 함이 없이, 일부 구현예에서, 더 큰 삽입 크기, 더 큰 편집 영역(예를 들어, 표적 영역에서 제1 편집/치환과 제2 편집/치환 사이의 거리), 및 더 많은 수의 원하는 편집(예를 들어, 이종 목적 서열과 표적 게놈의 불일치)은 더 긴 최적의 이종 목적 서열을 생성할 수 있다.

[0389] 특정 구현예에서, 주형 핵산은, 예를 들어 이종성 코딩 영역을 게놈에 도입; 엑손 구조/대안 스플라이싱에 작용 또는 유발, 예를 들어 하나 이상의 엑손의 엑손 스키핑(exon skipping)을 야기; 내인성 유전자의 붕괴를 유발, 예를 들어 유전자 녹아웃을 형성; 내인성 유전자의 전사 활성화를 유발; 내인성 DNA의 후성유전학적 조절을 유발; 하나 이상의 작동가능하게 연결된 유전자의 상향-조절을 유발, 예를 들어 유전자 활성화 또는 과발현을 야기; 하나 이상의 작동가능하게 연결된 유전자의 하향-조절을 유발, 예를 들어 유전자 녹다운을 형성 등에 의해, 숙주 게놈 기능을 변경 또는 특정하는 서열을 함유하도록 확인, 설계, 조작, 및 구축될 수 있는 맞춤화된 RNA 서열 주형을 포함한다. 특정 구현예에서, 맞춤화된 RNA 서열 주형은 엑손 및/또는 이식유전자를 코딩하는 서열을 함유하고, 전사 인자 활성제, 억제자, 인핸서 등 및 이의 조합에 대한 결합 부위를 제공하도록 조작될 수 있다. 일부 구현예에서, 맞춤화된 주형은 표적 부위에 작동가능하게 연결된 내인성 RNA 전사체 또는 내인성 단백질에 발현될 핵산 또는 펩티드 태그를 인코딩하도록 조작될 수 있다. 다른 구현예에서, 코딩 서열은 스플라이스 공여 부위, 스플라이스 수여 부위, 또는 폴리-A 테일을 이용하여 추가로 맞춤화될 수 있다.

[0390] 시스템의 주형 핵산(예를 들어, 주형 RNA)은 통상적으로 표적 DNA 내로 원하는 서열을 작성하기 위한 목적 서열(예를 들어, 이종 목적 서열)을 포함한다. 목적 서열은 코딩 또는 비-코딩일 수 있다. 주형 핵산(예를 들어, 주형 RNA)은 표적 DNA 유전자좌에서 삽입, 돌연변이, 또는 결실을 초래하도록 설계될 수 있다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)는 표적 DNA에서 삽입을 유발하도록 설계될 수 있다. 예를 들어, 주형 핵산(예를 들어, 주형 RNA)은 이종성 서열을 함유할 수 있으며, 여기서 역전사는 표적 DNA 내로 이종성 서열의 삽입을 초래할 것이다. 다른 구현예에서, RNA 주형은 표적 DNA 내로 결실을 도입하도록 설계될 수 있다. 예를 들어, 주형 핵산(예를 들어, 주형 RNA)은 원하는 결실의 상류 및 하류에서 표적 DNA와 일치할 수 있고, 여기서 역전사는 개재 서열 없이, 예를 들어 개재 서열의 결실을 유발하면서, 주형 핵산(예를 들어, 주형 RNA)로부터 상류 및 하류 서열의 카피를 초래할 것이다. 다른 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 표적 DNA 내로 편집물을 도입하도록 설계될 수 있다. 예를 들어, 주형 RNA는 하나 이상의 뉴클레오티드를 제외하고 표적 DNA 서열과 일치할 수 있으며, 여기서 역전사는 이들 편집물의 표적 DNA 내로의 카피를 초래하여, 돌연변이, 예를 들어 전이 또는 전환 돌연변이를 생성할 것이다.

[0391] 일부 구현예에서, 목적 서열의 표적 부위 내로의 작성은 뉴클레오티드의 치환을 초래하며, 예를 들어 여기서 목적 서열의 전체 길이는 하나 이상의 미스매치 염기가 있는 표적 부위의 일치 길이에 상응한다. 일부 구현예에서, 이종 목적 서열은 서열 변경이 조합, 예를 들어 동시 추가 및 결실, 추가 및 치환, 또는 결실 및 치환이 발생할 수 있도록 설계될 수 있다.

[0392] 일부 구현예에서, 이종 목적 서열은 오픈 리딩 프레임 또는 오픈 리딩 프레임의 단편을 함유할 수 있다. 일부 구현예에서, 이종 목적 서열은 코작 서열을 갖는다. 일부 구현예에서, 이종 목적 서열은 내부 리보솜 도입 부위를 갖는다. 일부 구현예에서, 이종 목적 서열은 T2A 또는 P2A 부위와 같은 자기-절단 펩티드를 갖는다. 일부 구현예에서, 이종 목적 서열은 출발 코돈을 갖는다. 일부 구현예에서, 주형 RNA는 스플라이스 수여 부위를 갖는다. 일부 구현예에서, 주형 RNA는 스플라이스 공여 부위를 갖는다. 예시적인 스플라이스 수여 및 스플라이

스 공여 부위는, 그 전체가 본 명세서에 참고로 포함된 WO2016044416에 기재되어 있다. 예시적인 스플라이스 수여 부위 서열은 기술 분야의 당업자에게 알려져 있다. 일부 구현예에서, 주형 RNA는 정지 코돈의 하류에서 마이크로RNA 결합 부위를 갖는다. 일부 구현예에서, 주형 RNA는 오픈 판독 프레임의 정지 코돈의 하류에서 폴리A 테일을 갖는다. 일부 구현예에서, 주형 RNA는 하나 이상의 엑손을 포함한다. 일부 구현예에서, 주형 RNA는 하나 이상의 인트론을 포함한다. 일부 구현예에서, 주형 RNA는 진핵성 전사 종결자를 포함한다. 일부 구현예에서, 주형 RNA는 향상된 번역 요소 또는 번역 향상 요소를 포함한다. 일부 구현예에서, RNA는 인간 T-세포 백혈병 바이러스(HTLV-1) R 영역을 포함한다. 일부 구현예에서, RNA는 B형 간염 바이러스(HPRE) 또는 우드척(Woodchuck) 간염 바이러스(WPRE)의 것과 같은 핵 수출을 향상하는 번역후 조절 요소를 포함한다.

[0393] 일부 구현예에서, 이종 목적 서열은 비-코딩 서열을 함유할 수 있다. 예를 들어, 주형 핵산(예를 들어, 주형 RNA)은 조절 요소, 예를 들어 프로모터 또는 인핸서 서열 또는 miRNA 결합 부위를 포함할 수 있다. 일부 구현예에서, 표적 부위에서 목적 서열의 통합은 내인성 유전자의 상향조절을 초래할 것이다. 일부 구현예에서, 표적 부위에서 목적 서열의 통합은 내인성 유전자의 하향조절을 초래할 것이다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 조직 특이적 프로모터 또는 인핸서를 포함하며, 이들 각각은 단방향 또는 양방향일 수 있다. 일부 구현예에서 프로모터는 RNA 중합효소 I 프로모터, RNA 중합효소 II 프로모터, 또는 RNA 중합효소 III 프로모터이다. 일부 구현예에서 프로모터는 TATA 요소를 포함한다. 일부 구현예에서 프로모터는 B 인식 요소를 포함한다. 일부 구현예에서 프로모터는 전사 인자에 대한 하나 이상의 결합 부위를 갖는다.

[0394] 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 후성유전학적 변형을 조정하는 부위를 포함한다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 염색질 절연체를 포함한다. 예를 들어, 주형 핵산(예를 들어, 주형 RNA)은 CTCF 부위 또는 DNA 메틸화를 위해 표적화된 부위를 포함한다.

[0395] 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 이펙터 서열에 작동가능하게 연결된 적어도 하나의 조절 영역으로 구성된 유전자 발현 단위를 포함한다. 이펙터 서열은, RNA로 전사되는 서열(예를 들어, 코딩 서열 또는 비-코딩 서열, 예컨대 마이크로RNA를 인코딩하는 서열)일 수 있다.

[0396] 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 이종 목적 서열은 내인성 인트론에서 표적 게놈 내로 삽입된다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 이종 목적 서열은 표적 게놈 내로 삽입되고, 이에 의해 새로운 엑손으로 작용한다. 일부 구현예에서, 이종 목적 서열의 표적 게놈 내로의 삽입은 천연 엑손의 대체 또는 천연 엑손의 스키핑을 초래한다.

[0397] 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 이종 목적 서열은 게놈 세이프 하버 부위, 예컨대 AAVS1, CCR5, ROSA26, 또는 알부민 유전자좌에서 표적 게놈 내로 삽입된다. 일부 구현예에서, 유전자 변형은 T-세포 수용체 α 상수(*TRAC*) 유전자좌 내로 CAR를 통합하는 데 사용된다(Eyquem et al Nature 543, 113-117 (2017)). 일부 구현예에서, 유전자 변형 시스템은 T-세포 수용체 β 상수(*TRBC*) 유전자좌 내로 CAR를 통합하는 데 사용된다. 많은 다른 세이프 하버부는 컴퓨터 계산 접근에 의해 확인되었으며(Pellenz et al Hum Gen Ther 30, 814-828 (2019)), 유전자 변형 시스템-매개 통합에 사용될 수 있다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 이종 목적 서열은 유전자간 또는 유전자내 영역에서 게놈에 첨가된다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 이종 목적 서열은 내인성 활성 유전자의 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, 또는 100 kb 내 5' 또는 3'에서 게놈에 첨가된다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 이종 목적 서열은 내인성 프로모터 또는 인핸서의 0.1 kb, 0.25 kb, 0.5 kb, 0.75, kb, 1 kb, 2 kb, 3 kb, 4 kb, 5 kb, 7.5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 50, 75 kb, 또는 100 kb 내 5' 또는 3'에서 게놈에 첨가된다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)의 이종 목적 서열은, 예를 들어 50 내지 50,000 염기쌍(예를 들어, 50 내지 40,000 bp, 500 내지 30,000 bp, 500 내지 20,000 bp, 100 내지 15,000 bp, 500 내지 10,000 bp, 50 내지 10,000 bp, 50 내지 5,000 bp일 수 있다.

[0398] 주형 핵산(예를 들어, 주형 RNA)은 표적 DNA 유전자좌에서 삽입, 돌연변이, 또는 결실을 초래하도록 설계될 수 있다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)는 표적 DNA에서 삽입을 유발하도록 설계될 수 있다. 예를 들어, 주형 핵산(예를 들어, 주형 RNA)은 이종 목적 서열을 함유할 수 있으며, 여기서 역전사는 표적 DNA 내로 이종 목적 서열의 삽입을 초래할 것이다. 다른 구현예에서, RNA 주형은 표적 DNA 내로 결실을 작성하도록 설계될 수 있다. 예를 들어, 주형 핵산(예를 들어, 주형 RNA)은 원하는 결실의 상류 및 하류에서 표적 DNA와 일치할 수 있고, 여기서 역전사는 개재 서열 없이, 예를 들어 개재 서열의 결실을 유발하면서, 주형 핵산(예를 들어, 주형 RNA)로부터 상류 및 하류 서열의 카피를 초래할 것이다. 다른 구현예에서, 주형 핵산(예를 들어, 주형

RNA)은 표적 DNA 내로 편집물을 작성하도록 설계될 수 있다. 예를 들어, 주형 RNA는 하나 이상의 뉴클레오티드를 제외하고 표적 DNA 서열과 일치할 수 있으며, 여기서 역전사는 이들 편집물의 표적 DNA 내로의 카피를 초래하여, 돌연변이, 예를 들어 전이 또는 전환 돌연변이를 생성할 것이다.

[0399] 일부 구현예에서, 사전 편집 상동성 도메인은 표적 핵산 분자에 포함된 핵산 서열과 100% 서열 동일성을 갖는 핵산 서열을 포함한다.

[0400] 일부 구현예에서, 사후 편집 상동성 도메인은 표적 핵산 분자에 포함된 핵산 서열과 100% 서열 동일성을 갖는 핵산 서열을 포함한다.

[0401] *PBS 서열*

[0402] 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 PBS 서열을 포함한다. 일부 구현예에서, PBS 서열은 이종 목적 서열의 3'에 배치되고, 본 명세서에 기재된 시스템에 의해 변형될 부위에 인접한 서열에 상보적이거나, 시스템/유전자 변형 폴리펩티드에 의해 변형될 부위에 인접한 서열에 상보적인 서열에 대하여 1, 2, 3, 4, 또는 5개 이하의 미스매치를 포함한다. 일부 구현예에서, PBS 서열은 표적 핵산 분자 내 닉 부위의 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개 뉴클레오티드 내에 결합한다. 일부 구현예에서, 표적 핵산 분자에 대한 PBS 서열의 결합은, 예를 들어 표적-프라이밍된 역전사(TPRT)에 대한 프라이머로서 작용하는 3' 상동성 도메인을 이용하여, TPRT의 개시를 허용한다. 일부 구현예에서, PBS 서열은 길이가 3 내지 5, 5 내지 10, 10 내지 30, 10 내지 25, 10 내지 20, 10 내지 19, 10 내지 18, 10 내지 17, 10 내지 16, 10 내지 15, 10 내지 14, 10 내지 13, 10 내지 12, 10 내지 11, 11 내지 30, 11 내지 25, 11 내지 20, 11 내지 19, 11 내지 18, 11 내지 17, 11 내지 16, 11 내지 15, 11 내지 14, 11 내지 13, 11 내지 12, 12 내지 30, 12 내지 25, 12 내지 20, 12 내지 19, 12 내지 18, 12 내지 17, 12 내지 16, 12 내지 15, 12 내지 14, 12 내지 13, 13 내지 30, 13 내지 25, 13 내지 20, 13 내지 19, 13 내지 18, 13 내지 17, 13 내지 16, 13 내지 15, 13 내지 14, 14 내지 30, 14 내지 25, 14 내지 20, 14 내지 19, 14 내지 18, 14 내지 17, 14 내지 16, 14 내지 15, 15 내지 30, 15 내지 25, 15 내지 20, 15 내지 19, 15 내지 18, 15 내지 17, 15 내지 16, 16 내지 30, 16 내지 25, 16 내지 20, 16 내지 19, 16 내지 18, 16 내지 17, 17 내지 30, 17 내지 25, 17 내지 20, 17 내지 19, 17 내지 18, 18 내지 30, 18 내지 25, 18 내지 20, 18 내지 19, 19 내지 30, 19 내지 25, 19 내지 20, 20 내지 30, 20 내지 25, 또는 25 내지 30개의 뉴클레오티드, 예를 들어 길이가 10 내지 17, 12 내지 16, 또는 12 내지 14개의 뉴클레오티드이다. 일부 구현예에서, PBS 서열은 길이가 5 내지 20, 8 내지 16, 8 내지 14, 8 내지 13, 9 내지 13, 9 내지 12, 또는 10 내지 12개 뉴클레오티드, 예를 들어 길이가 9 내지 12개 뉴클레오티드이다.

[0403] 주형 핵산(예를 들어, 주형 RNA)은 표적 DNA에 대해 어느 정도 상동성을 가질 수 있다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA) PBS 서열 도메인은, 표적 DNA가 주형 핵산(예를 들어, 주형 RNA)의 역전사를 프라이밍하도록 위치하도록, 표적 DNA에 대한 어닐링 영역의 역할을 할 수 있다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 RNA의 3' 말단에서 표적 DNA에 대해 정확히 상동성인 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 175, 200개 또는 그 이상의 염기를 갖는다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)은 주형 핵산(예를 들어, 주형 RNA)의 5' 말단에서, 표적 DNA에 대해 적어도 50%, 60%, 70%, 80%, 85%, 90%, 95%, 97%, 98%, 99% 또는 100% 상동성인 적어도 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 175, 200개 또는 그 이상의 염기를 갖는다.

[0404] ***유도성 활성을 갖는 gRNA***

[0405] 일부 구현예에서, 본 명세서에 기재된 gRNA(예를 들어, 제2 가닥 닉 생성에 사용된 주형 RNA 또는 gRNA의 일부인 gRNA)는 유도성 활성을 갖는다. 유도성 활성은 (gRNA에 추가하여) 차단 도메인을 추가로 포함하는 주형 핵산, 예를 들어 주형 RNA에 의해 달성될 수 있으며, 여기서 차단 도메인의 일부 또는 전체 서열은 gRNA의 일부 또는 전체에 대해 적어도 부분적으로 상보적이다. 따라서 차단 도메인은 gRNA의 일부 또는 전체에 혼성화 또는 실질적으로 혼성화할 수 있다. 일부 구현예에서, gRNA가 차단 도메인이 gRNA에 혼성화되거나 또는 실질적으로 혼성화되는 제1 입체형태, 및 차단 도메인이 gRNA에 혼성화되지 않거나 또는 실질적으로 혼성화되지 않는 제2 입체형태를 채용할 수 있도록, 차단 도메인 및 유도성 활성 gRNA는 주형 핵산, 예를 들어 주형 RNA 상에 배치된다. 일부 구현예에서, 제1 입체형태에서, gRNA는 유전자 변형 폴리펩티드(예를 들어, 주형 핵산 결합 도메인, DNA 결합 도메인, 또는 엔도뉴클레아제 도메인(예를 들어, CRISPR/Cas 단백질))에 결합할 수 없거나 또는 차단 도메인이 결여된 다른 유사한 주형 RNA와 비교하여 실질적으로 감소된 친화도로 결합한다. 일부 구현예에서, 제2 입체형태에서, gRNA는 유전자 변형 폴리펩티드(예를 들어, 주형 핵산 결합 도메인, DNA 결합 도메인, 또는 엔

도뉴클레아제 도메인(예를 들어, CRISPR/Cas 단백질))에 결합할 수 있다. 일부 구현예에서, gRNA가 제1 또는 제2 입체형태인지 여부는 DNA 결합 또는 (예를 들어, CRISPR/Cas 단백질 중 유전자 변형 폴리펩티드가 포함하는) 유전자 변형 폴리펩티드의 엔도뉴클레아제 활성이 활성인지 여부에 영향을 미칠 수 있다.

[0406] 일부 구현예에서, 제2 닉을 조정하는 gRNA는 유도성 활성을 갖는다. 일부 구현예에서, 제2 닉을 조정하는 gRNA는 주형이 역전사 후에 유도된다. 일부 구현예에서, gRNA의 차단 도메인으로의 혼성화는 오프너(opener) 분자를 사용하여 붕괴될 수 있다. 일부 구현예에서, 오프너 분자는 gRNA 또는 차단 도메인의 일부 또는 전체에 결합하는 작용제를 포함하고, gRNA의 차단 도메인에 대한 혼성화를 억제한다. 일부 구현예에서, 오프너 분자는, 예를 들어 gRNA, 차단 도메인, 또는 둘 다에 부분적으로 또는 전체적으로 상보적인 서열을 포함하는 핵산을 포함한다. 적절한 오프너 분자를 선택 또는 설계함으로써, 오프너 분자의 제공은, CRISPR/Cas 단백질과 회합하고 CRISPR/Cas 단백질의 회합된 기능(예를 들어, DNA 결합 및/또는 엔도뉴클레아제 활성)을 제공하도록, gRNA의 입체형태에서의 변화를 촉진할 수 있다. 이론에 구속되고자 함이 없이, 선택된 시간 및/또는 위치에서 오프너 분자의 제공은 gRNA, CRISPR/Cas 단백질, 또는 이를 포함하는 유전자 변형 시스템의 활성의 공간적 및 시간적 제어를 가능하게 할 수 있다. 일부 구현예에서, 오프너 분자는 유전자 변형 폴리펩티드 및 또는 주형 핵산을 포함하는 세포에 대해 외인성이다. 일부 구현예에서, 오프너 분자는 (예를 들어, 유전자 변형 폴리펩티드 및 또는 gRNA를 포함하는 주형 핵산 및 차단 도메인을 포함하는 세포에 대해 내인성인) 내인성 작용제를 포함한다. 예를 들어, 유도성 gRNA, 차단 도메인, 및 오프너 분자는, 오프너 분자가 표적 세포 또는 조직에서 발현된 내인성 작용제이도록 선택될 수 있고, 예를 들어 이에 의해 표적 세포 또는 조직에서 유전자 변형 시스템의 활성을 보장한다. 추가의 예로서, 유도성 gRNA, 차단 도메인, 및 오프너 분자는, 오프너 분자가 하나 이상의 비-표적 세포 또는 조직에서 부재하거나 실질적으로 발현되지 않도록 선택될 수 있고, 예를 들어 이에 의해 유전자 변형 시스템의 활성이 하나 이상의 비-표적 세포 또는 조직에서 발생하지 않거나 실질적으로 발생하지 않거나, 또는 표적 세포 또는 조직과 비교하여 감소된 수준으로 발생하는 것을 보장한다. 예시적인 차단 도메인, 오프너 분자, 및 이의 용도는 PCT 출원 공개 WO2020044039A1에 기재되어 있으며, 이는 그 전체가 본 명세서에 참고로 포함된다. 일부 구현예에서, 주형 핵산, 예를 들어 주형 RNA는, 예를 들어 역전사효소 또는 RNA 결합 도메인, 및 gRNA에 의해, 유전자 변형 폴리펩티드의 하나 이상의 구성요소에 의한 결합을 위한 하나 이상의 서열 또는 구조를 포함할 수 있다. 일부 구현예에서, gRNA는 유전자 변형 폴리펩티드의 주형 핵산 결합 도메인(예를 들어, RNA 결합 도메인)과의 상호작용을 용이하게 한다. 일부 구현예에서, gRNA는, 예를 들어 표적 세포 게놈에서 유전자 변형 폴리펩티드를 매칭 표적에 지시한다.

[0407] ***유전자 변형 시스템의 환형 RNA 및 리보자임***

[0408] 표적 세포 내에서 제형화, 전달, 또는 유전자 변형 반응 동안 환형 및/또는 선형 RNA 상태를 이용하는 것은 유용할 수 있는 것으로 고려된다. 따라서 본 명세서에 기재된 임의의 양태의 일부 구현예에서, 유전자 변형 시스템은 하나 이상의 환형 RNA(circRNA)를 포함한다. 본 명세서에 기재된 임의의 양태의 일부 구현예에서, 유전자 변형 시스템은 하나 이상의 선형 RNA를 포함한다. 일부 구현예에서, 본 명세서에 기재된 바와 같은 핵산(예를 들어, 주형 핵산, 유전자 변형 폴리펩티드를 인코딩하는 핵산 분자, 또는 둘 다)은 circRNA이다. 일부 구현예에서, 환형 RNA 분자는 유전자 변형 폴리펩티드를 인코딩한다. 일부 구현예에서, 유전자 변형 폴리펩티드를 인코딩하는 circRNA 분자는 숙주 세포로 전달된다. 일부 구현예에서, 환형 RNA 분자는, 예를 들어 본 명세서에 기재된 바와 같은 재조합효소를 인코딩한다. 일부 구현예에서, 재조합효소를 인코딩하는 circRNA 분자는 숙주 세포로 전달된다. 일부 구현예에서, 유전자 변형 폴리펩티드를 인코딩하는 circRNA 분자는 (예를 들어, 숙주 세포 내에서, 예를 들어 숙주 세포의 핵에서) 번역 전에 선형화된다.

[0409] 환형 RNA(circRNA)는 세포 내에서 자연 발생하는 것으로 발견되었으며, 인간 세포에서 비-코딩 및 단백질 코딩 역할을 둘 다 포함하여, 다양한 기능을 갖는 것으로 발견되어 왔다. circRNA는 자기-스플라이싱 인트론을 RNA 분자(또는 RNA 분자를 인코딩하는 DNA)로 혼입시킴으로써 조작되어 RNA의 환형화를 초래할 수 있으며, 그 조작된 cirRNA는 향상된 단백질 생산 및 안정성을 가질 수 있는 것으로 나타난다(Wesselhoeft et al. *Nature Communications* 2018). 일부 구현예에서, 유전자 변형 폴리펩티드는 circRNA로서 인코딩된다. 일부 구현예에서, 주형 핵산은 DNA, 예컨대 dsDNA 또는 ssDNA이다. 특정 구현예에서, circDNA는 주형 RNA를 포함한다.

[0410] 일부 구현예에서, circRNA는 하나 이상의 리보자임 서열을 포함한다. 일부 구현예에서, 리보자임 서열은 예를 들어 숙주 세포에서 자가절단에 의해 활성화되며, 예를 들어 이에 의해 circRNA의 선형화를 초래한다. 일부 구현예에서, 리보자임은 마그네슘 농도가, 예를 들어 숙주 세포 내에서, 절단에 충분한 수준으로 도달하는 경우 활성화된다. 일부 구현예에서, circRNA는 숙주 세포로의 전달 전에 낮은 마그네슘 환경으로 유지된다. 일부 구현예에서, 리보자임은 단백질-반응성 리보자임이다. 일부 구현예에서, 리보자임은 핵산-반응성 리보자임이다.

일부 구현예에서, circRNA는 절단 부위를 포함한다. 일부 구현예에서, circRNA는 제2 절단 부위를 포함한다.

[0411] 일부 구현예에서, circRNA는 표적 세포의 핵에서 선형화된다. 일부 구현예에서, 세포의 핵에서 circRNA의 선형화는, 예를 들어 절단 사건을 활성화시키기 위해, 세포의 핵에 존재하는 구성요소를 포함한다. 일부 구현예에서, 핵 요소, 예를 들어 핵 단백질, 예를 들어 게놈-상호작용 단백질, 예를 들어 후성유전학적 변형자, 예를 들어 EZH2에 반응성인 리보자임, 예를 들어 B2 또는 ALU 요소로부터의 리보자임이, 예를 들어 유전자 변형 시스템의 circRNA 내로 포함된다. 일부 구현예에서, circRNA의 핵 국재화는 리보자임의 자가촉매 활성 및 circRNA의 선형화의 증가를 초래한다.

[0412] 일부 구현예에서, 리보자임은 유전자 변형 시스템의 다른 구성요소 중 하나 이상에 대해 이종성이다. 일부 구현예에서, 유도성 리보자임(예를 들어, 본 명세서에 기재된 바와 같은 circRNA)은, 예를 들어 단백질 리간드-반응성 압타머 설계를 이용함으로써, 합성적으로 형성된다. 담배 둥근무늬 바이러스 해머헤드 리보자임의 위성 RNA를 MS2 외피 단백질 압타머와 함께 이용하기 위한 시스템이 문헌[Kennedy et al. Nucleic Acids Res 42(19):12306-12321 (2014)(그 전체가 본 명세서에 참고로 포함됨)]에 기재되었으며, MS2 외피 단백질의 존재하에 리보자임 활성의 활성화를 초래한다. 구현예에서, 그러한 시스템은 세포질 또는 핵에 국재화된 단백질 리간드에 반응한다. 일부 구현예에서, 단백질 리간드는 MS2가 아니다. 리간드를 표적화하기 위해 RNA 압타머를 생성하기 위한 방법은, 예를 들어 지수적 농축에 의한 리간드의 계통적 진화(systematic evolution of ligands by exponential enrichment; SELEX)를 기반으로 기재되었으며(Tuerk and Gold, Science 249(4968):505-510 (1990); Ellington and Szostak, Nature 346(6287):818-822 (1990); 이들 각각의 방법은 본 명세서에 참고로 포함됨], 일부 경우에서, 인실리코 설계의 도움을 받는다(Bell et al. PNAS 117(15):8486-8493, 이 방법은 본 명세서에 참고로 포함됨). 따라서 일부 구현예에서, 표적 리간드에 대한 압타머가 생성되고, 예를 들어 단백질 리간드 존재 하에서, 예를 들어 리보자임-매개 절단 및 circRNA 선형화를 촉발하기 위해 합성 리보자임 시스템 내로 혼입된다. 일부 구현예에서, circRNA 선형화는, 예를 들어 세포질 내 리간드와 회합하는 압타머를 사용하여, 세포질 내에서 촉발된다. 일부 구현예에서, circRNA 선형화는, 예를 들어 핵 내 리간드와 회합하는 압타머를 사용하여, 핵 내에서 촉발된다. 구현예에서, 핵 내 리간드는 후성유전학적 변형자 또는 전사 인자를 포함한다. 일부 구현예에서, 선형화를 촉발하는 리간드는 표적-외 세포보다 표적상 세포에서 더 높은 수준으로 존재한다.

[0413] 핵산-반응성 리보자임 시스템이 circRNA 선형화를 위해 이용될 수 있다는 것이 추가로 고려된다. 예를 들어, 리보자임 활성화를 촉발하기 위해 정의된 표적 핵산 분자를 감지하는 바이오센서는, 예를 들어 문헌[Penchovsky, Biotechnology Advances 32(5):1015-1027 (2014)(본 명세서에 참고로 포함됨)]에 기재되어 있다. 이들 방법에 의해, 리보자임은 비활성 상태로 자연적으로 접히고, 정의된 표적 핵산 분자(예를 들어, RNA 분자) 존재 하에서만 활성화된다. 일부 구현예에서, 유전자 변형 시스템의 circRNA는 정의된 표적 핵산, 예를 들어 RNA, 예를 들어 mRNA, miRNA, 가이드 RNA, gRNA, sgRNA, ncRNA, lncRNA, tRNA, snRNA, 또는 mtRNA 존재 하에서 활성화되는 핵산-반응성 리보자임을 포함한다. 일부 구현예에서, 선형화를 촉발하는 핵산은 표적-외 세포보다 표적상 세포에서 더 높은 수준으로 존재한다.

[0414] 본 명세서의 임의의 양태의 일부 구현예에서, 유전자 변형 시스템은 관심 표적 조직 또는 표적 세포에 대한 유도성 특이성을 갖는 하나 이상의 리보자임, 예를 들어 관심 표적 조직 또는 표적 세포에서 더 높은 수준으로 존재하는 리간드 또는 핵산에 의해 활성화되는 리보자임을 포함한다. 일부 구현예에서, 유전자 변형 시스템은 세포내 구획, 예를 들어, 핵, 핵소체, 세포질, 또는 미토콘드리아에 대해 유도성 특이성을 갖는 리보자임을 포함한다. 일부 구현예에서, 표적 세포내 구획에서 더 높은 수준으로 존재하는 리간드 또는 핵산에 의해 활성화되는 리보자임. 일부 구현예에서, 유전자 변형 시스템의 RNA 구성요소는, 예를 들어 선형화에 의해 활성화되는 circRNA로서 제공된다. 일부 구현예에서, 유전자 변형 폴리펩티드를 인코딩하는 circRNA의 선형화는 분자를 번역에 대해 활성화시킨다. 일부 구현예에서, 유전자 변형 시스템의 circRNA 구성요소를 활성화시키는 신호는, 예를 들어 그 시스템이 이들 세포에서 특이적으로 활성화되도록, 표적상 세포 또는 조직에서 더 높은 수준으로 존재한다.

[0415] 일부 구현예에서, 유전자 변형 시스템의 RNA 구성요소는 선형화에 의해 비활성화되는 circRNA로서 제공된다. 일부 구현예에서, 유전자 변형 폴리펩티드를 인코딩하는 circRNA는 절단 및 분해에 의해 비활성화된다. 일부 구현예에서, 유전자 변형 폴리펩티드를 인코딩하는 circRNA는 폴리펩티드의 코딩 서열로부터 번역 신호를 분리하는 절단에 의해 비활성화된다. 일부 구현예에서, 유전자 변형 시스템의 circRNA 구성요소를 비활성화시키는 신호는, 그 시스템이 이들 세포에서 특이적으로 비활성화되도록, 표적-외 세포 또는 조직에서 더 높은 수준으로

존재한다.

[0416] ***표적 핵산 부위***

[0417] 일부 구현예에서, 유전자 변형 후, 편집된 서열 주위의 표적 부위는, 예를 들어 문헌[Karst et al. (2020) *bioRxiv* doi.org/10.1101/645903(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 예를 들어 표적 부위의 긴-판독 암플리콘 시퀀싱에 의해 결정 시, 예를 들어 편집 사건의 약 50% 또는 10% 미만으로 제한된 수의 삽입 또는 결실을 함유한다. 일부 구현예에서, 표적 부위는, 예를 들어 문헌[Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020)(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 예를 들어 표적 부위의 긴-판독 암플리콘 시퀀싱에 의해 결정 시, 다수 연속 편집 사건, 예를 들어 헤드-투-테일(head-to-tail) 또는 헤드-투-헤드 중복을 나타내지 않는다. 일부 구현예에서, 표적 부위는 주형 RNA에 상응하는 통합된 서열을 함유한다. 일부 구현예에서, 표적 부위는, 예를 들어 문헌[Karst et al. *bioRxiv* doi.org/10.1101/645903 (2020) (그 전체가 본 명세서에 참고로 포함됨).]에 기재된 바와 같이, 예를 들어 표적 부위의 긴-판독 암플리콘 시퀀싱에 의해 결정 시, 약 1% 또는 10% 초과의 사건에서 내인성 RNA로부터 생성되는 삽입물을 함유하지 않는다. 일부 구현예에서, 표적 부위는 주형 RNA에 상응하는 통합된 서열을 함유한다.

[0418] 본 발명의 특정 양태에서, 유전자 변형 시스템에 의해 통합된 숙주 DNA-결합 부위는 유전자 내, 인트론 내, 엑손 내, 임의의 유전자의 코딩 영역 외부인 ORF, 유전자의 조절 영역 내, 또는 유전자의 조절 영역 외부일 수 있다. 다른 양태에서, 폴리펩티드는 하나 이상의 숙주 DNA 서열에 결합할 수 있다.

[0419] 일부 구현예에서, 유전자 변형 시스템은 다수의 대립유전자에서 표적 유전자좌를 편집하는 데 사용된다. 일부 구현예에서, 유전자 변형 시스템은 특이적 대립유전자를 편집하도록 설계된다. 예를 들어, 유전자 변형 폴리펩티드는 하나의 대립유전자에만 존재하는 특이적 서열로 지시될 수 있으며, 예를 들어 표적 대립유전자, 예를 들어 gRNA 또는 어닐링 도메인에 대해서는 상동성을 갖지만, 제2의 동족 대립유전자에 대해서는 상동성이 없는 주형 RNA를 포함한다. 일부 구현예에서, 유전자 변형 시스템은 일배체형-특이적 대립유전자를 변경시킬 수 있다. 일부 구현예에서, 특이적 대립유전자를 표적화하는 유전자 변형 시스템은, 예를 들어 표적 대립유전자에 대해 적어도 2, 4, 6, 8, 또는 10배 선호를 갖는 대립 유전자를 우선적으로 표적화한다.

[0420] ***제2 가닥 닉 생성***

[0421] 일부 구현예에서, 본 명세서에 기재된 유전자 변형 시스템은 (예를 들어, 유전자 변형 폴리펩티드에서) 제1 가닥에 닉을 생성하는 닉카제 활성, 및 (예를 들어, 유전자 변형 폴리펩티드와 별개의 폴리펩티드에서) 표적 DNA의 제2 가닥에 닉을 생성하는 닉카제 활성을 포함한다. 본 명세서에 논의된 바와 같이, 이론에 구속되고자 함이 없이, 표적 부위 DNA의 제1 가닥의 닉 생성은 주형 RNA의 서열, 예를 들어 이종 목적 서열을 역전사하기 위해 RT 도메인에 의해 사용될 수 있는 3' OH를 제공하는 것으로 생각된다. 이론에 구속되고자 함이 없이, 제2 가닥에 추가의 닉을 도입하는 것은 세포 DNA 복구 기구가 원래 게놈 서열보다 이종성 목적 서열 기반 서열을 더 빈번하게 채용하도록 편향시킬 수 있는 것으로 생각된다. 일부 구현예에서, 제2 가닥에 대한 추가의 닉은 제1 가닥에 대한 닉과 동일한 엔도뉴클레아제 도메인(예를 들어, 닉카제 도메인)에 의해 만들어진다. 일부 구현예에서, 동일한 유전자 변형 폴리펩티드는 제1 가닥에 대한 닉과 제2 가닥에 대한 닉 둘 다를 수행한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 CRISPR/Cas 도메인을 포함하고, 제2 가닥에 대한 추가의 닉은, 예를 들어 CRISPR/Cas 도메인이 제2 가닥에 닉을 생성하도록 지시하는 제2 gRNA를 포함하는, 추가의 핵산에 의해 지시된다. 다른 구현예에서, 추가의 제2 가닥 닉은 제1 가닥에 대한 닉 외에 상이한 엔도뉴클레아제 도메인(예를 들어, 닉카제 도메인)에 의해 만들어진다. 일부 구현예에서, 상기 상이한 엔도뉴클레아제 도메인은 유전자 변형 폴리펩티드와 별도의, 추가의 폴리펩티드(예를 들어, 본 발명의 시스템은 추가의 폴리펩티드를 추가로 포함함)에 위치한다. 일부 구현예에서, 추가의 폴리펩티드는 본 명세서에 기재된 엔도뉴클레아제 도메인(예를 들어, 닉카제 도메인)을 포함한다. 일부 구현예에서, 추가의 폴리펩티드는, 예를 들어 본 명세서에 기재된 바와 같은, DNA 결합 도메인을 포함한다.

[0422] 본 명세서에서, 제2 가닥 닉이 일어나는 위치는, 제1 가닥 닉에 상대적으로, 원하는 유전자 변형 DNA 변형의 수득, 원하지 않는 이중 가닥 파손(DSB)의 발생, 원하지 않는 삽입의 발생, 또는 원하지 않는 결실의 발생 중 하나 이상에 영향을 미칠 수 있는 정도인 것으로 고려된다. 이론에 구속되고자 함이 없이, 제2 가닥 닉 생성은 두 가지 일반적 배향, 즉 내측 및 외측 닉으로 발생할 수 있다.

[0423] 일부 구현예에서, 내측 닉 배향에서, RT 도메인은 (예를 들어, 주형 RNA(예를 들어, 이종 목적 서열)를 사용하여) 제2 가닥 닉에서 멀어지며 중합한다. 일부 구현예에서, 내측 닉 배향에서, 제1 가닥에 대한 닉의 위치 및

제2 가닥에 대한 닉의 위치는(예를 들어, 두 닉 모두가 CRISPR/Cas 도메인을 포함하는 폴리펩티드(예를 들어, 유전자 변형 폴리펩티드)에 의해 만들어진다는 시나리오에서) 제1 PAM 부위 및 제2 PAM 부위 사이에 위치한다. 일부 구현예에서, 내측 닉 배향에서, 제1 가닥에 대한 닉의 위치 및 제2 가닥에 대한 닉의 위치는 폴리펩티드와 추가의 폴리펩티드가 표적 DNA에 결합하는 부위 사이에 있다. 일부 구현예에서, 내측 닉 배향에서, 제2 가닥에 대한 닉의 위치는 제1 가닥에 대한 닉의 위치에 대하여 폴리펩티드 및 추가의 폴리펩티드의 결합 부위의 동일한 쪽에 위치한다. 일부 구현예에서, 내측 닉 배향에서, 제1 가닥에 대한 닉의 위치 및 제2 가닥에 대한 닉의 위치는 PAM 부위와 표적 부위로부터 떨어진 부위 사이에 위치한다.

[0424] 내측 닉 배향을 제공하는 유전자 변형 시스템의 예는 CRISPR/Cas 도메인을 포함하는 유전자 변형 폴리펩티드, 제1 가닥 상의 표적 부위 DNA의 닉 생성을 지시하는 gRNA를 포함하는 주형 RNA, 및 제1 닉의 위치로부터 멀리 떨어진 부위에서 닉 생성을 지시하는 추가의 gRNA를 포함하는 추가의 핵산을 포함하고, 여기서 제1 닉의 위치 및 제2 닉의 위치는 두 gRNA가 유전자 변형 폴리펩티드에 지시하는 부위 중의 PAM 부위 사이에 있다. 추가의 예로서, 내측 닉 배향을 제공하는 또 다른 유전자 변형 시스템은, 징크 핑거 분자 및 제1 닉카제 도메인을 포함하는 유전자 변형 폴리펩티드(여기서, 징크 핑거 분자는, 제1 닉카제 도메인이 표적 부위의 제1 가닥에 닉을 생성하도록 지시하는 방식으로 표적 DNA에 결합함); CRISPR/Cas 도메인을 포함하는 추가의 폴리펩티드, 및 그 추가의 폴리펩티드가 제2 가닥 상에서 표적 부위 DNA로부터 떨어진 부위에 닉을 생성하도록 지시하는 gRNA를 포함하는 추가의 핵산을 포함하며, 여기서 제1 닉의 위치 및 제2 닉의 위치는 PAM 부위와 징크 핑거 분자가 결합하는 부위 사이에 있다. 추가의 예로서, 내측 닉 배향을 제공하는 또 다른 유전자 변형 시스템은, 징크 핑거 분자 및 제1 닉카제 도메인을 포함하는 유전자 변형 폴리펩티드(여기서, 징크 핑거 분자는, 제1 닉카제 도메인이 표적 부위의 제1 가닥에 닉을 생성하도록 지시하는 방식으로 표적 DNA에 결합함); TAL 이펙터 분자 및 제2 닉카제 도메인을 포함하는 추가의 폴리펩티드를 포함하고, 여기서 TAL 이펙터 분자는 추가의 폴리펩티드가 제2 가닥에 닉을 생성하도록 지시하는 방식으로 표적 부위로부터 떨어진 부위에 결합하고, 여기서 제1 닉의 위치와 제2 닉의 위치는 TAL 이펙터 분자가 결합하는 부위와 징크 핑거 분자가 결합하는 부위 사이에 있다.

[0425] 일부 구현예에서, 외측 닉 배향에서, RT 도메인은 (예를 들어, 주형 RNA(예를 들어, 이종 목적 서열)를 사용하여) 제2 가닥 닉을 향해 중합한다. 일부 구현예에서, 내측 닉 배향에서 제1 및 제2 닉 모두가 CRISPR/Cas 도메인을 포함하는 폴리펩티드(예를 들어, 유전자 변형 폴리펩티드)에 의해 만들어지는 경우, 제1 PAM 부위와 제2 PAM 부위는 제1 가닥에 대한 닉의 위치와 제2 가닥에 대한 닉의 위치 사이에 위치한다. 일부 구현예에서, 내측 닉 배향에서, 폴리펩티드(예를 들어, 유전자 변형 폴리펩티드)와 추가의 폴리펩티드는 제1 가닥에 대한 닉의 위치와 제2 가닥에 대한 닉의 위치 사이의 표적 DNA 상의 부위에 결합한다. 일부 구현예에서, 내측 닉 배향에서, 제2 가닥에 대한 닉의 위치는 제1 가닥에 대한 닉의 위치에 대하여 폴리펩티드 및 추가의 폴리펩티드의 결합 부위의 반대 쪽에 위치한다. 일부 구현예에서, 내측 배향에서, PAM 부위와 표적 부위로부터 떨어진 부위는 제1 가닥에 대한 닉의 위치와 제2 가닥에 대한 닉의 위치 사이에 위치한다.

[0426] 외측 닉 배향을 제공하는 유전자 변형 시스템의 예는 CRISPR/Cas 도메인을 포함하는 유전자 변형 폴리펩티드, 제1 가닥 상의 표적 부위 DNA의 닉 생성을 지시하는 gRNA를 포함하는 주형 RNA, 및 제1 닉의 위치로부터 멀리 떨어진 부위에서 닉 생성을 지시하는 추가의 gRNA를 포함하는 추가의 핵산을 포함하고, 여기서 제1 닉의 위치 및 제2 닉의 위치는 두 gRNA가 유전자 변형 폴리펩티드에 지시하는 부위 중의 PAM 부위의 외측에 있다(즉, PAM 부위는 제1 닉의 위치와 제2 닉의 위치 사이에 있음). 추가의 예로서, 외측 닉 배향을 제공하는 또 다른 유전자 변형 시스템은, 징크 핑거 분자 및 제1 닉카제 도메인을 포함하는 유전자 변형 폴리펩티드(여기서, 징크 핑거 분자는, 제1 닉카제 도메인이 표적 부위의 제1 가닥에 닉을 생성하도록 지시하는 방식으로 표적 DNA에 결합함); CRISPR/Cas 도메인을 포함하는 추가의 폴리펩티드, 및 그 추가의 폴리펩티드가 제2 가닥 상에서 표적 부위 DNA로부터 떨어진 부위를 닉 생성하도록 지시하는 gRNA를 포함하는 추가의 핵산을 포함하며, 여기서 제1 닉의 위치 및 제2 닉의 위치는 PAM 부위와 징크 핑거 분자가 결합하는 부위의 외측에 있다(즉, PAM 부위와 징크 핑거 분자가 결합하는 부위는 제1 닉의 위치와 제2 닉의 위치 사이에 있음). 추가의 예로서, 외측 닉 배향을 제공하는 또 다른 유전자 변형 시스템은, 징크 핑거 분자 및 제1 닉카제 도메인을 포함하는 유전자 변형 폴리펩티드(여기서, 징크 핑거 분자는, 제1 닉카제 도메인이 표적 부위의 제1 가닥에 닉을 생성하도록 지시하는 방식으로 표적 DNA에 결합함); TAL 이펙터 분자 및 제2 닉카제 도메인을 포함하는 추가의 폴리펩티드를 포함하고, 여기서 TAL 이펙터 분자는 추가의 폴리펩티드가 제2 가닥에 닉을 생성하도록 지시하는 방식으로 표적 부위로부터 떨어진 부위에 결합하고, 여기서 제1 닉의 위치와 제2 닉의 위치는 TAL 이펙터 분자가 결합하는 부위와 징크 핑거 분자가 결합하는 부위의 외측에 있다(즉, TAL 이펙터 분자가 결합하는 부위와 징크 핑거 분자가 결합하는 부위는 제1 닉의 위치와 제2 닉의 위치 사이에 있음).

[0427] 이론에 구속되고자 함이 없이, 제2 가닥 닉이 제공되는 유전자 변형 시스템의 경우, 일부 구현예에서 외측 닉 배향이 바람직한 것으로 생각된다. 본 명세서에 기재된 바와 같이, 내측 닉은 외측 닉 배향보다 더 많은 수의 이중 가닥 파손(DSB)를 생성할 수 있다. DSB는 세포의 핵에서 DSB 복구 경로에 의해 인식될 수 있으며, 이는 원하지 않는 삽입 및 결실을 생성한다. 외측 닉 배향은 DSB 형성의 감소된 위험, 및 상응하는 더 적은 양의 원하지 않는 삽입 및 결실을 제공할 수 있다. 일부 구현예에서, 원하지 않는 삽입 및 결실은 이종 목적 서열에 의해 인코딩되지 않는 삽입 및 결실, 예를 들어 이종 목적 서열에 의해 인코딩된 변형에 관계없는 이중 가닥 파손 복구 경로에 의해 생성되는 삽입 및 결실이다. 일부 구현예에서, 원하는 유전자 변형은 이종 목적 서열에 의해 인코딩되는(예를 들어, 표적 부위 내로 이종 목적 서열을 유전자 변형 작성함으로써 달성되는) 표적 DNA에 대한 변화(예를 들어, 치환, 삽입, 또는 결실)를 포함한다. 일부 구현예에서, 제1 가닥 닉 및 제2 가닥 닉은 외측 배향이다.

[0428] 추가적으로, 제1 가닥 닉과 제2 가닥 닉 사이의 거리는, 원하는 유전자 변형 시스템 DNA 변형의 수득, 원하지 않는 이중 가닥 파손(DSB)의 발생, 원하지 않는 삽입의 발생, 또는 원하지 않는 결실의 발생 중 하나 이상에 영향을 미칠 수 있는 정도이다. 이론에 구속되고자 함이 없이, 제2 가닥 닉 이득인, 이종 목적 서열의 표적 DNA로의 혼입을 향한 DNA 복구의 편향은, 제1 가닥과 제2 가닥 닉 사이의 거리가 감소됨에 따라 증가하는 것으로 생각된다. 그러나, 또한 DSB 형성의 위험은 제1 가닥 닉과 제2 가닥 닉 사이의 거리가 감소됨에 따라 증가하는 것으로 생각된다. 따라서, 원하지 않는 삽입 및/또는 결실은 제1 가닥 닉과 제2 가닥 닉 사이의 거리가 감소함에 따라 증가할 수 있는 것으로 생각된다. 일부 구현예에서, 제1 가닥 닉과 제2 가닥 닉 사이의 거리는 이종 목적 서열의 표적 DNA 내로의 혼입을 향한 DNA 복구의 편향의 이득과 DSB 형성 및 원하지 않는 결실 및/또는 삽입의 위험 간에 균형을 맞추도록 선택된다. 일부 구현예에서, 제1 가닥 닉과 제2 가닥 닉이 적어도 역치 거리로 떨어져 있는 시스템은, 제1 닉과 제2 닉이 역치 거리 보다 적게 떨어져 있는 다른 유사한 내측 닉 배향 시스템에 상대적으로, 원하는 유전자 변형 시스템 변형 결과의 증가된 수준, 원하지 않는 결실의 감소된 수준, 및/또는 원하지 않는 삽입의 감소된 수준을 갖는다. 일부 구현예에서, 역치 거리(들)가 하기 제공된다.

[0429] 일부 구현예에서, 제1 닉과 제2 닉은 적어도 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 또는 200개 뉴클레오티드 떨어져 있다. 일부 구현예에서, 제1 닉과 제2 닉은 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 또는 250개 이하 뉴클레오티드 떨어져 있다. 일부 구현예에서, 제1 닉과 제2 닉은 20 내지 200, 30 내지 200, 40 내지 200, 50 내지 200, 60 내지 200, 70 내지 200, 80 내지 200, 90 내지 200, 100 내지 200, 110 내지 200, 120 내지 200, 130 내지 200, 140 내지 200, 150 내지 200, 160 내지 200, 170 내지 200, 180 내지 200, 190 내지 200, 20 내지 190, 30 내지 190, 40 내지 190, 50 내지 190, 60 내지 190, 70 내지 190, 80 내지 190, 90 내지 190, 100 내지 190, 110 내지 190, 120 내지 190, 130 내지 190, 140 내지 190, 150 내지 190, 160 내지 190, 170 내지 190, 180 내지 190, 20 내지 180, 30 내지 180, 40 내지 180, 50 내지 180, 60 내지 180, 70 내지 180, 80 내지 180, 90 내지 180, 100 내지 180, 110 내지 180, 120 내지 180, 130 내지 180, 140 내지 180, 150 내지 180, 160 내지 180, 170 내지 180, 20 내지 170, 30 내지 170, 40 내지 170, 50 내지 170, 60 내지 170, 70 내지 170, 80 내지 170, 90 내지 170, 100 내지 170, 110 내지 170, 120 내지 170, 130 내지 170, 140 내지 170, 150 내지 170, 160 내지 170, 20 내지 160, 30 내지 160, 40 내지 160, 50 내지 160, 60 내지 160, 70 내지 160, 80 내지 160, 90 내지 160, 100 내지 160, 110 내지 160, 120 내지 160, 130 내지 160, 140 내지 160, 150 내지 160, 20 내지 150, 30 내지 150, 40 내지 150, 50 내지 150, 60 내지 150, 70 내지 150, 80 내지 150, 90 내지 150, 100 내지 150, 110 내지 150, 120 내지 150, 130 내지 150, 140 내지 150, 20 내지 140, 30 내지 140, 40 내지 140, 50 내지 140, 60 내지 140, 70 내지 140, 80 내지 140, 90 내지 140, 100 내지 140, 110 내지 140, 120 내지 140, 130 내지 140, 20 내지 130, 30 내지 130, 40 내지 130, 50 내지 130, 60 내지 130, 70 내지 130, 80 내지 130, 90 내지 130, 100 내지 130, 110 내지 130, 120 내지 130, 20 내지 120, 30 내지 120, 40 내지 120, 50 내지 120, 60 내지 120, 70 내지 120, 80 내지 120, 90 내지 120, 100 내지 120, 110 내지 120, 20 내지 110, 30 내지 110, 40 내지 110, 50 내지 110, 60 내지 110, 70 내지 110, 80 내지 110, 90 내지 110, 100 내지 110, 20 내지 100, 30 내지 100, 40 내지 100, 50 내지 100, 60 내지 100, 70 내지 100, 80 내지 100, 90 내지 100, 20 내지 90, 30 내지 90, 40 내지 90, 50 내지 90, 60 내지 90, 70 내지 90, 80 내지 90, 20 내지 80, 30 내지 80, 40 내지 80, 50 내지 80, 60 내지 80, 70 내지 80, 20 내지 70, 30 내지 70, 40 내지 70, 50 내지 70, 60 내지 70, 20 내지 60, 30 내지 60, 40 내지 60, 50 내지 60, 20 내지 50, 30 내지 50, 40 내지 50, 20 내지 40, 30 내지 40, 또는 20 내지 30개 뉴클레오티드 떨어져 있다. 일부 구현예에서, 제1 가닥 닉 및 제2 가닥 닉은 40 내지 100개 뉴클레오티드 떨어져 있다.

[0430] 이론에 구속되고자 함이 없이, 제2 가닥 닉이 제공되고, 내측 닉 배향이 선택된 유전자 변형 시스템의 경우, 제1 가닥 닉과 제2 가닥 닉 사이의 거리를 증가시키는 것이 바람직할 수 있는 것으로 생각된다. 본 명세서에 기재된 바와 같이, 내측 닉 배향은 외측 닉 배향보다 많은 수의 DSB를 생성할 수 있고, 외측 닉 배향보다 더 많은 양의 원하지 않는 삽입 및 결실을 생성할 수 있지만, 닉 사이의 거리를 증가시키는 것은, DSB, 원하지 않는 결실, 및/또는 원하지 않는 삽입의 증가를 완화할 수 있다. 일부 구현예에서, 제1 닉과 제2 닉이 적어도 역치 거리로 떨어져 있는 내측 닉 배향은, 제1 닉과 제2 닉이 역치 거리 보다 적게 떨어져 있는 다른 유사한 내측 닉 배향 시스템에 상대적으로, 원하는 유전자 변형 시스템 변형 결과의 증가된 수준, 원하지 않는 결실의 감소된 수준, 및/또는 원하지 않는 삽입의 감소된 수준을 갖는다. 일부 구현예에서 역치 거리가 하기 제공된다.

[0431] 일부 구현예에서, 제1 가닥 닉 및 제2 가닥 닉은 내측 배향이다. 일부 구현예에서, 제1 가닥 닉 및 제2 가닥 닉은 내측 배향이고 제1 가닥 닉 및 제2 가닥 닉은 적어도 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 220, 240, 260, 280, 300, 350, 400, 450, 또는 500개 뉴클레오티드, 적어도 100개 뉴클레오티드 떨어져 있다(및 선택적으로 500, 400, 300, 200, 190, 180, 170, 160, 150, 140, 130, 또는 120개 뉴클레오티드 떨어져 있음). 일부 구현예에서, 제1 가닥 닉 및 제2 가닥 닉은 내측 배향이고 제1 가닥 닉 및 제2 가닥 닉은 100 내지 200, 110 내지 200, 120 내지 200, 130 내지 200, 140 내지 200, 150 내지 200, 160 내지 200, 170 내지 200, 180 내지 200, 190 내지 200, 100 내지 190, 110 내지 190, 120 내지 190, 130 내지 190, 140 내지 190, 150 내지 190, 160 내지 190, 170 내지 190, 180 내지 190, 100 내지 180, 110 내지 180, 120 내지 180, 130 내지 180, 140 내지 180, 150 내지 180, 160 내지 180, 170 내지 180, 100 내지 170, 110 내지 170, 120 내지 170, 130 내지 170, 140 내지 170, 150 내지 170, 160 내지 170, 100 내지 160, 110 내지 160, 120 내지 160, 130 내지 160, 140 내지 160, 150 내지 160, 100 내지 150, 110 내지 150, 120 내지 150, 130 내지 150, 140 내지 150, 100 내지 140, 110 내지 140, 120 내지 140, 130 내지 140, 100 내지 130, 110 내지 130, 120 내지 130, 100 내지 120, 110 내지 120, 또는 100 내지 110개 뉴클레오티드 떨어져 있다.

[0432] ***화학적으로 변형된 핵산 및 핵산 말단 특징부***

[0433] 본 명세서에 기재된 핵산(예를 들어, 주형 핵산, 예를 들어 주형 RNA; 또는 유전자 변형 폴리펩티드를 인코딩하는 핵산(예를 들어, mRNA); 또는 gRNA)은 비변형 또는 변형된 핵염기를 포함할 수 있다. 자연 발생 RNA는 4개의 염기성 리보뉴클레오티드, 즉 ATP, CTP, UTP 및 GTP로부터 합성되지만, 번역후 변형된 뉴클레오티드를 함유할 수 있다. 또한, 대략 100개의 다양한 뉴클레오시드 변형이 RNA에서 확인되었다(Rozenski, J, Crain, P, and McCloskey, J. (1999). The RNA Modification Database: 1999 update. Nucl Acids Res 27: 196-197). RNA는 또한 자연에서 발생하지 않는 완전히 합성인 뉴클레오티드를 포함할 수 있다.

[0434] 일부 구현예에서, 화학적 변형은 WO/2016/183482, 미국 특허 공개 20090286852, 국제 출원 WO/2012/019168, WO/2012/045075, WO/2012/135805, WO/2012/158736, WO/2013/039857, WO/2013/039861, WO/2013/052523, WO/2013/090648, WO/2013/096709, WO/2013/101690, WO/2013/106496, WO/2013/130161, WO/2013/151669, WO/2013/151736, WO/2013/151672, WO/2013/151664, WO/2013/151665, WO/2013/151668, WO/2013/151671, WO/2013/151667, WO/2013/151670, WO/2013/151666, WO/2013/151663, WO/2014/028429, WO/2014/081507, WO/2014/093924, WO/2014/093574, WO/2014/113089, WO/2014/144711, WO/2014/144767, WO/2014/144039, WO/2014/152540, WO/2014/152030, WO/2014/152031, WO/2014/152027, WO/2014/152211, WO/2014/158795, WO/2014/159813, WO/2014/164253, WO/2015/006747, WO/2015/034928, WO/2015/034925, WO/2015/038892, WO/2015/048744, WO/2015/051214, WO/2015/051173, WO/2015/051169, WO/2015/058069, WO/2015/085318, WO/2015/089511, WO/2015/105926, WO/2015/164674, WO/2015/196130, WO/2015/196128, WO/2015/196118, WO/2016/011226, WO/2016/011222, WO/2016/011306, WO/2016/014846, WO/2016/022914, WO/2016/036902, WO/2016/077125, 또는 WO/2016/077123에 제공된 것이며, 이들 각각은 그 전체가 본 명세서에 참고로 포함된다. 화학적으로 변형된 뉴클레오티드의 폴리뉴클레오티드로의 혼입은, 뉴클레오티드에서 변형의 위치에 따라, 핵염기, 백본 또는 이들 둘 다 내로 포함되는 변형을 초래할 수 있음이 이해된다. 일부 구현예에서, 백본 변형은 EP 2813570에 제공된 것이며, 이는 그 전체가 본 명세서에 참고로 포함된다. 일부 구현예에서, 변형된 캡은 미국 특허 공개 20050287539에 제공된 것이며, 이는 그 전체가 본 명세서에 참고로 포함된다.

[0435] 일부 구현예에서, 화학적으로 변형된 핵산(예를 들어 RNA, 예를 들어 mRNA)은 ARCA: 항-역전 캡 유사체(m27.3'-OGP3G), GP3G(비메틸화 캡 유사체), m7GP3G(모노메틸화 캡 유사체), m32.2.7GP3G(트리메틸화 캡 유사체), m5CTP(5'-메틸-시티딘 트리포스페이트), m6ATP(N6-메틸-아데노신-5'-트리포스페이트), s2UTP(2-티오-우리딘 트리포스페이트), 및 Ψ(슈도우리딘 트리포스페이트) 중 하나 이상을 포함한다.

[0436] 일부 구현예에서, 화학적으로 변형된 핵산은 5' 캡, 예를 들어 7-메틸구아노신 캡(예를 들어, O-Me-m7G 캡); 초메틸화 캡 유사체; NAD+-유래 캡 유사체(예를 들어, 문헌[Kiledjian, Trends in Cell Biology 28, 454-464 (2018)]에 기재된 바와 같음); 또는 변형된, 예를 들어 비오티닐화된 캡 유사체(예를 들어, 문헌[Bednarek et al., Phil Trans R Soc B 373, 20180167 (2018)]에 기재된 바와 같음)를 포함한다.

[0437] 일부 구현예에서, 화학적으로 변형된 핵산은 다음 중 하나 이상으로부터 선택되는 3' 특징부를 포함한다: 폴리A 테일; 쌍을 형성하지 않은 5개 뉴클레오티드에 의해 측접된 16-뉴클레오티드의 긴 스템-루프 구조(예를 들어, 문헌[Mannironi et al., Nucleic Acid Research 17, 9113-9126 (1989)]에 기재된 바와 같음); 삼중-나선형 구조(예를 들어, 문헌[Brown et al., PNAS 109, 19202-19207 (2012)]에 기재된 바와 같음); tRNA, Y RNA, 또는 볼트 RNA 구조(예를 들어, 문헌[Labno et al., Biochemica et Biophysica Acta 1863, 3125-3147 (2016)]에 기재된 바와 같음); 하나 이상의 데옥시리보뉴클레오티드 트리포스페이트(dNTP), 2'O-메틸화 NTP, 또는 포스포로티오에이트-NTP의 혼입; 단일 뉴클레오티드 화학적 변형(예를 들어, 3' 말단 리보스의 반응성 알데히드로의 산화에 이어 알데히드-반응성 변형된 뉴클레오티드의 접합); 또는 또 다른 핵산 분자에 대한 화학적 결찰.

[0438] 일부 구현예에서, 핵산(예를 들어, 주형 핵산)은, 예를 들어 디하이드로우리딘, 이노신, 7-메틸구아노신, 5-메틸시티딘(5mC), 5' 포스페이트 리보티미딘, 2'-O-메틸 리보티미딘, 2'-O-에틸 리보티미딘, 2'-플루오로 리보티미딘, C-5 프로피닐-데옥시시티딘(pdC), C-5 프로피닐-데옥시우리딘(pdU), C-5 프로피닐-시티딘(pC), C-5 프로피닐-우리딘(pU), 5-메틸 시티딘, 5-메틸 우리딘, 5-메틸 데옥시시티딘, 5-메틸 데옥시우리딘 메톡시, 2,6-디아미노퓨린, 5'-디메톡시트리틸-N4-에틸-2'-데옥시시티딘, C-5 프로피닐-f-시티딘(pfC), C-5 프로피닐-f-우리딘(pfU), 5-메틸 f-시티딘, 5-메틸 f-우리딘, C-5 프로피닐-m-시티딘(pmC), C-5 프로피닐-f-우리딘(pmU), 5-메틸 m-시티딘, 5-메틸 m-우리딘, LNA(잠금 핵산), MGB(마이너 그루브 바인더) 슈도우리딘(Ψ), 1-N-메틸슈도우리딘(1-Me-Ψ), 또는 5-메톡시우리딘(5-MO-U)으로부터 선택되는, 하나 이상의 변형된 뉴클레오티드를 포함한다.

[0439] 일부 구현예에서, 핵산은 백본 변형, 예를 들어 백본에서 당 또는 포스페이트 기에 대한 변형을 포함한다. 일부 구현예에서, 핵산은 핵염기 변형을 포함한다.

[0440] 일부 구현예에서, 핵산은 표 9의 하나 이상의 화학적으로 변형된 뉴클레오티드, 표 10의 하나 이상의 화학적 백본 변형, 표 11의 하나 이상의 화학적으로 변형된 캡을 포함한다. 예를 들어 일부 구현예에서, 핵산은 둘 이상(예를 들어, 3, 4, 5, 6, 7, 8, 9, 또는 10개 또는 이상)의 다양한 유형의 화학적 변형을 포함한다. 예로서, 핵산은, 예를 들어 본 명세서에 기재된 바와 같은, 예를 들어 표 9의 둘 이상(예를 들어, 3, 4, 5, 6, 7, 8, 9, 또는 10개 또는 그 이상)의 다양한 유형의 변형된 핵염기를 포함할 수 있다. 대안적으로 또는 조합하여, 핵산은, 예를 들어 본 명세서에 기재된 바와 같은, 예를 들어 표 10의 둘 이상(예를 들어, 3, 4, 5, 6, 7, 8, 9, 또는 10개 또는 그 이상)의 다양한 유형의 백본 변형을 포함할 수 있다. 대안적으로 또는 조합하여, 핵산은, 예를 들어 본 명세서에 기재된 바와 같은, 예를 들어 표 11의 하나 이상의 변형된 캡을 포함할 수 있다. 예를 들어 일부 구현예에서, 핵산은 하나 이상의 유형의 변형된 핵염기 및 하나 이상의 유형의 백본 변형; 하나 이상의 유형의 변형된 핵염기 및 하나 이상의 변형된 캡; 하나 이상의 유형의 변형된 캡 및 하나 이상의 유형의 백본 변형; 또는 하나 이상의 유형의 변형된 핵염기, 하나 이상의 유형의 백본 변형, 및 하나 이상의 유형의 변형된 캡을 포함한다.

[0441] 일부 구현예에서, 핵산은 하나 이상(예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000개, 또는 그 이상)의 변형된 핵염기를 포함한다. 일부 구현예에서, 핵산의 모든 핵염기는 변형된다. 일부 구현예에서, 핵산은 백본의 하나 이상(예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 900, 1000개, 또는 그 이상)의 위치에서 변형된다. 일부 구현예에서, 핵산의 모든 백본 위치는 변형된다.

[0442] **표 9. 변형된 뉴클레오티드**

표 9

[0443]

5-아자-우리딘 2-티오-5-아자-미딘 2-티오우리딘 4-티오-슈도우리딘 2-티오-슈도우리딘 5-하이드록시우리딘 3-메틸우리딘 5-카르복시메틸-우리딘 1-카르복시메틸-슈도우리딘 5-프로피닐-우리딘 1-프로피닐-슈도우리딘 5-타우리노메틸우리딘 1-타우리노메틸-슈도우리딘 5-타우리노메틸-2-티오-우리딘 1-타우리노메틸-4-티오-우리딘 5-메틸-우리딘 1-메틸-슈도우리딘	N2-메틸-6-티오-구아노신 N2,N2-디메틸-6-티오-구아노신 피리딘-4-온 리보뉴클레오시드 2-티오-5-아자-우리딘 2-티오미딘 4-티오-슈도미딘 2-티오-슈도와이딘 3-메틸미딘 1-프로피닐-슈도미딘 1-메틸-1-데아자-슈도미딘 2-티오-1-메틸-1-데아자-슈도우리딘 4-메톡시-슈도미딘 5'-O-(1-티오포스페이트)-아데노신 5'-O-(1-티오포스페이트)-시티딘 5'-O-(1-티오포스페이트)-구아노신 5'-O-(1-티오포스페이트)-우리딘 5'-O-(1-티오포스페이트)-슈도우리딘
4-티오-1-메틸-슈도우리딘 2-티오-1-메틸-슈도우리딘 1-메틸-1-데아자-슈도우리딘 2-티오-1-메틸-1-데아자-슈도미딘 디하이드로우리딘 디하이드로슈도우리딘 2-티오-디하이드로미딘 2-티오-디하이드로슈도우리딘 2-메톡시우리딘 2-메톡시-4-티오-우리딘 4-메톡시-슈도우리딘 4-메톡시-2-티오-슈도우리딘 5-아자-시티딘 슈도이소시티딘 3-메틸-시티딘 N4-아세틸시티딘 5-포르밀시티딘	2'-O-메틸-아데노신 2'-O-메틸-시티딘 2'-O-메틸-구아노신 2'-O-메틸-우리딘 2'-O-메틸-슈도우리딘 2'-O-메틸-이노신 2-메틸아데노신 2-메틸티오-N6-메틸아데노신 2-메틸티오-N6 이소펜테닐아데노신 2-메틸티오-N6-(시스-하이드록시이소펜테닐)아데노신 N6-메틸-N6-트레오닐카르바모일아데노신 N6-하이드록시노르발릴카르바모일아데노신 2-메틸티오-N6-하이드록시노르발릴 카르바모일아데노신 2'-O-리보실아데노신(포스페이트) 1,2'-O-디메틸이노신 5,2'-O-디메틸시티딘 N4-아세틸-2'-O-메틸시티딘

N4-메틸시티딘	리시딘
5-하이드록시메틸시티딘	7-메틸구아노신
1-메틸-슈도이소시티딘	N2,2'-O-디메틸구아노신
피롤로-시티딘	N2,N2,2'-O-트리메틸구아노신
피롤로-슈도이소시티딘	2'-O-리보실구아노신(포스페이트)
2-티오-시티딘	와이부토신
2-티오-5-메틸-시티딘	퍼옥시와이부토신
4-티오-슈도이소시티딘	하이드록시와이부토신
4-티오-1-메틸-슈도이소시티딘	과소변형된 하이드록시와이부토신
4-티오-1-메틸-1-데아자-슈도이소시티딘	메틸와이오신
1-메틸-1-데아자-슈도이소시티딘	케오신
제불라린	애폭시케오신
5-아자-제불라린	갈락토실-케오신
5-메틸-제불라린	만노실-케오신
5-아자-2-티오-제불라린	7-시아노-7-데아자구아노신
2-티오-제불라린	7-아미노메틸-7-데아자구아노신
2-메톡시-시티딘	아르케오신(archaeosine)
	5,2'-O-디메틸우리딘
	4-티오우리딘
	5-메틸-2-티오우리딘
	2-티오-2'-O-메틸우리딘
	3-(3-아미노-3-카르복시프로필)우리딘
	5-메톡시우리딘
	우리딘 5-옥시아세트산
	우리딘 5-옥시아세트산 메틸 에스테르
	5-(카르복시하이드록시메틸)우리딘)
	5-(카르복시하이드록시메틸)우리딘 메틸 에스테르
	5-메톡시카르보닐메틸우리딘
	5-메톡시카르보닐메틸-2'-O-메틸우리딘
	5-메톡시카르보닐메틸-2-티오우리딘
	5-아미노메틸-2-티오우리딘
	5-메틸아미노메틸우리딘
	5-메틸아미노메틸-2-티오우리딘
	5-메틸아미노메틸-2-셀레노우리딘
	5-카르바모일메틸우리딘
	5-카르바모일메틸-2'-O-메틸우리딘
	5-카르복시메틸아미노메틸우리딘
	5-카르복시메틸아미노메틸-2'-O-메틸우리딘
	5-카르복시메틸아미노메틸-2-티오우리딘
	N4,2'-O-디메틸시티딘
	5-카르복시메틸우리딘
	N6,2'-O-디메틸아데노신
	N,N6,O-2'-트리메틸아데노신
	N2,7-디메틸구아노신
	N2,N2,7-트리메틸구아노신
	3,2'-O-디메틸우리딘
	5-메틸디하이드로우리딘
	5-포르밀-2'-O-메틸시티딘
	1,2'-O-디메틸구아노신
	4-데메틸와이오신
	이소와이오신
	N6-아세틸아데노신

2-메톡시-5-메틸-시티딘 4-메톡시-슈도이소시티딘 4-메톡시-1-메틸-슈도이소시티딘 2-아미노퓨린 2,6-디아미노퓨린 7-데아자-아데닌 7-데아자-8-아자-아데닌 7-데아자-2-아미노퓨린 7-데아자-8-아자-2-아미노퓨린 7-데아자-2,6- 디아미노퓨린 7-데아자-8-아자-2,6-디아미노퓨린 1-메틸아데노신 N6-이소펜테닐아데노신 N6-(시스-하이드록시이소펜테닐)아데노신 2-메틸티오-N6-(시스-하이드록시이소펜테닐) 아데노신 N6-글리시닐카르바모일아데노신 N6-트레오닐카르바모일아데노신 2-메틸티오-N6-트레오닐 카르바모일아데노신 N6,N6-디메틸아데노신 7-메틸아데닌 2-메틸티오-아데닌 2-메톡시-아데닌 이노신 1-메틸-이노신 와이오신 와이부토신 7-데아자-구아노신 7-데아자-8-아자-구아노신 6-티오-구아노신 6-티오-7-데아자-구아노신 6-티오-7-데아자-8-아자-구아노신 7-메틸-구아노신 6-티오-7-메틸-구아노신 7-메틸이노신 6-메톡시-구아노신 1-메틸구아노신 N2-메틸구아노신 N2,N2-디메틸구아노신 8-옥소-구아노신 7-메틸-8-옥소-구아노신 1-메틸-6-티오-구아노신	

[0444] **표 10. 백본 변형**

표 10

[0445]

2'-O-메틸 백본
펩티드 핵산(PNA) 백본
포스포로티오에이트 백본
모르폴리노 백본
카르바메이트 백본
실록산 백본
설파이드 백본
설폭사이드 백본
설폰 백본
포름아세틸 백본
티오포름아세틸 백본
메틸렌포름아세틸 백본
리보아세틸 백본
알켄 함유 백본
설파메이트 백본
설포네이트 백본
설폰아미드 백본
메틸렌이미노 백본
메틸렌하이드라지노 백본
아미드 백본

[0446] **표 11. 변형된 캡**

표 11

[0447]

m7GpppA
m7GpppC
m2,7GpppG
m2,2,7GpppG
m7Gpppm7G
m7,2'OmeGpppG
m72'dGpppG
m7,3'OmeGpppG
m7,3'dGpppG
GppppG
m7GppppG
m7GppppA
m7GppppC
m2,7GppppG
m2,2,7GppppG
m7Gppppm7G
m7,2'OmeGppppG
m72'dGppppG
m7,3'OmeGppppG
m7,3'dGppppG

[0448] 유전자 변형 시스템의 주형을 포함하는 뉴클레오티드는 천연 또는 변형된 염기, 또는 이의 조합일 수 있다. 예를 들어, 주형은 슈도우리딘, 디하이드로우리딘, 이노신, 7-메틸구아노신, 또는 다른 변형된 염기를 함유할 수 있다. 일부 구현예에서, 주형은 잠금 핵산 뉴클레오티드를 함유할 수 있다. 일부 구현예에서, 주형에서 사용된 변형된 염기는 주형의 역전사를 억제하지 않는다. 일부 구현예에서, 주형에서 사용된 변형된 염기는 역전사, 예를 들어 특이성 또는 정확도를 개선시킬 수 있다.

[0449] 일부 구현예에서, 시스템의 RNA 구성요소(예를 들어, 주형 RNA 또는 gRNA)는 하나 이상의 뉴클레오티드 변형을 포함한다. 일부 구현예에서, gRNA의 변형 패턴은, 미변형된 또는 말단-변형된 가이드와 비교하여 생체 내 활성에 현저한 영향을 줄 수 있다(예를 들어, 문헌[Finn et al. *Cell Rep* 22(9):2227-2235 (2018)]의 도 1D에 나타낸 바와 같음; 그 전체가 본 명세서에 참고로 포함됨). 이론에 구속되고자 함이 없이, 이 과정은 적어도 부분적으로는 변형에 의해 부여된 RNA의 안정화로 인한 것일 수 있다. 그러한 변형의 비제한적인 예는 2'-O-메틸(2'-O-Me), 2'-O-(2-메톡시에틸)(2'-O-MOE), 2'- 플루오로(2'-F), 뉴클레오티드들 사이의 포스포로티오에이트(PS) 결합, G-C 치환, 및 뉴클레오티드들 사이의 역전된 무염기 연결 및 이들의 균등물을 포함할 수 있다.

[0450] 일부 구현예에서, 주형 RNA(예를 들어, 표적 부위에 결합하는 이의 부분에서) 또는 가이드 RNA는 5' 말단 영역을 포함한다. 일부 구현예에서, 주형 RNA 또는 가이드 RNA는 5' 말단 영역을 포함하지 않는다. 일부 구현예에서, 5' 말단 영역은, 예를 들어 문헌[Briner AE et al, Molecular Cell 56: 333-339 (2014)(그 전체가 본 명세서에 참고로 포함됨; 본 명세서에서, 예를 들어 모든 가이드 RNA에 대해 적용가능함)]에 sgRNA에 대하여 기재된 바와 같은, gRNA 스페이서 영역을 포함한다. 일부 구현예에서, 5' 말단 영역은 5' 말단 변형을 포함한다. 일부 구현예에서, 스페이서 영역을 갖는 또는 갖지 않는 5' 말단 영역은 crRNA, trRNA, sgRNA 및/또는 dgRNA와 회합될 수 있다. gRNA 스페이서 영역은, 일부 경우에서 가이드 영역, 가이드 도메인, 또는 표적화 도메인을 포함할 수 있다.

[0451] 일부 구현예에서, 주형 RNA(예를 들어, 표적 부위에 결합하는 이의 일부에서) 또는 본 명세서에 기재된 가이드 RNA는, 그 전체가 본 명세서에 참고로 포함된 WO2018107028A1의 표 4에 나타낸 임의의 서열을 포함한다. 일부 구현예에서, 서열이 가이드 및/또는 스페이서 영역을 나타내는 경우, 조성물은 이 영역을 포함하거나 포함하지 않을 수 있다. 일부 구현예에서, 가이드 RNA는, 예를 들어 서열번호에 의해 확인되는 바와 같이, WO2018107028A1의 표 4에 나타낸 임의의 서열의 하나 이상의 변형을 포함한다. 구현예에서, 뉴클레오티드는 동일하거나 상이할 수 있고/있거나, 나타낸 변형 패턴은 WO2018107028A1의 표 4에 나타낸 바와 같은 가이드 서열의 변형 패턴과 동일하거나 유사할 수 있다. 일부 구현예에서, 변형 패턴은 gRNA 또는 gRNA의 영역(예를 들어 5' 말단 영역, 하부 스템 영역, 벌지 영역, 상부 스템 영역, 연계 영역, 헤어핀 1 영역, 헤어핀 2 영역, 3' 말단 영역)의 변형의 상대적 위치 및 동일성을 포함한다. 일부 구현예에서, 변형 패턴은 WO2018107028A1의 표 4의 서열 열에 나타낸 서열들 중 어느 하나의 및/또는 서열의 하나 이상의 영역에 걸친 변형의 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100%를 함유한다. 일부 구현예에서, 변형 패턴은 WO2018107028A1의 표 4의 서열 열에 나타낸 서열들 중 어느 하나의 변형 패턴에 대하여 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일하다. 일부 구현예에서, 변형 패턴은 예를 들어 5' 말단 영역, 하부 스템 영역, 벌지 영역, 상부 스템 영역, 연계 영역, 헤어핀 1 영역, 헤어핀 2 영역, 및/또는 3' 말단 영역에서, WO2018107028A1의 표 4에 나타낸 서열의 하나 이상의 영역에 대해 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일하다. 일부 구현예에서, 변형 패턴은 5' 말단 영역에 걸쳐 서열의 변형 패턴에 대해 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99%, 또는 100% 동일하다. 일부 구현예에서, 변형 패턴은 하부 스템에 걸쳐 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 동일하다. 일부 구현예에서, 변형 패턴은 벌지에 걸쳐 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 동일하다. 일부 구현예에서, 변형 패턴은 상부 스템에 걸쳐 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 동일하다. 일부 구현예에서, 변형 패턴은 연계에 걸쳐 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 동일하다. 일부 구현예에서, 변형 패턴은 헤어핀 1에 걸쳐 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 동일하다. 일부 구현예에서, 변형 패턴은 헤어핀 2에 걸쳐 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 동일하다. 일부 구현예에서, 변형 패턴은 3' 말단에 걸쳐 적어도 50%, 55%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% 동일하다. 일부 구현예에서, 변형 패턴은 WO2018107028A1의 표 4에 서열, 또는 그러한 서열의 영역(예를 들어, 5' 말단, 하부 스템, 벌지, 상부 스템, 연계, 헤어핀 1 영역, 헤어핀 2 영역, 3' 말단 영역)의 변형 패턴과, 예를 들어 0, 1, 2, 3, 4, 5, 6개, 또는 그 이상의 뉴클레오티드에서 상이하다. 일부 구현예에서, gRNA는 WO2018107028A1의 표 4의 서열의 변형과, 예를 들어 0, 1, 2, 3, 4, 5, 6개, 또는 그 이상의 뉴클레오티드에서 상이하다. 일부 구현예에서, gRNA는 WO2018107028A1의 표 4의 서열의 영역(예를 들어, 5' 말단 영역, 하부 스템, 벌지, 상부 스템, 연계, 헤어핀 1 영역, 헤어핀 2 영역, 3' 말단)의 변형과, 예를 들어 0, 1, 2, 3, 4, 5, 6개, 또는 그 이상의 뉴클레오티드에서 상이한 변형을 포함한다.

[0452] 일부 구현예에서, 주형 RNA(예를 들어, 표적 부위에 결합하는 이의 부분에서) 또는 gRNA는 2'-O-메틸(2'-O-Me)

변형된 뉴클레오티드를 포함한다. 일부 구현예에서, gRNA는 2'-O-(2-메톡시 에틸)(2'-O-moe) 변형된 뉴클레오티드를 포함한다. 일부 구현예에서, gRNA는 2'-플루오로(2'-F) 변형된 뉴클레오티드를 포함한다. 일부 구현예에서, gRNA는 뉴클레오티드들 사이에 포스포로티오에이트(PS) 결합을 포함한다. 일부 구현예에서, gRNA는 5' 말단 변형, 3' 말단 변형, 또는 5' 및 3' 말단 변형을 포함한다. 일부 구현예에서, 5' 말단 변형은 뉴클레오티드들 사이에 포스포로티오에이트(PS) 결합을 포함한다. 일부 구현예에서, 5' 말단 변형은 2'-O-메틸(2'-O-Me), 2'-O-(2-메톡시 에틸)(2'-O-MOE), 및/또는 2'-플루오로(2'-F) 변형된 뉴클레오티드를 포함한다. 일부 구현예에서, 5' 말단 변형은 적어도 하나의 포스포로티오에이트(PS) 결합, 및 2'-O-메틸(2'-O- Me), 2'-O-(2-메톡시 에틸)(2'-O-MOE), 및/또는 2'-플루오로(2'-F) 변형 뉴클레오티드 중 하나 이상을 포함한다. 말단 변형은 포스포로티오에이트(PS), 2'-O-메틸(2'-O-Me), 2'-O-(2- 메톡시에틸)(2'-O-MOE), 및/또는 2'-플루오로(2'-F) 변형을 포함할 수 있다. 균등한 말단 변형은 본 명세서에 기재된 구현예들에 의해서도 포함된다. 일부 구현예에서, 주형 RNA 또는 gRNA는 주형 RNA 또는 gRNA의 하나 이상의 영역의 변형과 조합된 말단 변형을 포함한다. RNA, 예를 들어 gRNA 및 이의 제형을 보호하기 위한 추가의 예시적인 변형 및 방법은 WO2018126176A1에 기재된 바와 같으며, 이는 그 전체가 본 명세서에 참고로 포함된다.

[0453] 일부 구현예에서, 구조-가이드된 계통적 접근법들이 예를 들어 문헌[Mir et al. *Nat Commun* 9:2641 (2018)(그 전체가 본 명세서에 참고로 포함됨)]에 기재된 바와 같이, 주형 RNA 또는 가이드 RNA에 변형(예를 들어, 2'-OMe-RNA, 2'-F-RNA, 및 PS 변형)을 도입하는 데 사용된다. 일부 구현예에서, 2'-F-RNA의 도입은, 예를 들어 C3'-엔도 당 퍼커링(puckering)을 최소로 간섭하면서, RNA:RNA 또는 RNA:DNA 듀플렉스(duplex)의 열 및 뉴클레아제 안정성을 증가시킨다. 일부 구현예에서, 2'-F는 2'-OH가 RNA:DNA 듀플렉스 안정성에 중요한 위치에서, 2'-OMe보다 더 양호하게 용인될 수 있다. 일부 구현예에서, crRNA는, 예를 들어 그 전체가 본 명세서에 참고로 포함된 문헌[Mir et al. *Nat Commun* 9:2641 (2018)]의 보충 표 1에 기재된 바와 같이, Cas9 활성을 감소시키지 않는 하나 이상의 변형, 예를 들어 C10, C20, 또는 C21(완전히 변형됨)을 포함한다. 일부 구현예에서, tracrRNA는 Cas9 활성을 감소시키지 않는 하나 이상의 변형, 예를 들어 문헌[Mir et al. *Nat Commun* 9:2641 (2018)의 보충 표 1]의 T2, T6, T7, 또는 T8(완전히 변형됨)을 포함한다. 일부 구현예에서, crRNA는 (예를 들어, 본 명세서에 기재된 바와 같은) 하나 이상의 변형, 예를 들어 C20 및 T2를 포함하는 tracrRNA와 쌍을 형성할 수 있는 하나 이상의 변형을 포함한다. 일부 구현예에서, gRNA는 예를 들어 crRNA 및 tracrRNA의 키메라를 포함한다(예를 들어, 문헌[Jinek et al. *Science* 337(6096):816-821 (2012)]). 구현예에서, crRNA 및 tracrRNA로부터의 변형은, 예를 들어 향상된 안정성을 갖는 변형된 gRNA를 생산하도록 단일-가이드 키메라 상에서 맵핑된다.

[0454] 일부 구현예에서, gRNA 분자는 자연 발생 구조적 구성요소, 예를 들어 헤어핀의 부가 또는 공제에 의해 변형될 수 있다. 일부 구현예에서, gRNA는, 예를 들어 그 전체가 본 명세서에 참고로 포함된 WO2018106727에 기재된 바와 같은, 결실된 하나 이상의 3' 헤어핀 요소를 갖는 gRNA를 포함할 수 있다. 일부 구현예에서, gRNA는 부가된 헤어핀 구조, 예를 들어 스페이서 영역에서 부가된 헤어핀 구조를 함유할 수 있으며, 이는 문헌[Kocak et al. *Nat Biotechnol* 37(6):657-666 (2019)]의 교시에서 CRISPR-Cas 시스템의 특이성을 증가시키는 것으로 나타났다. 단축된 gRNA 및 생체 활성을 개선시키는 특이적 변형의 예를 포함하는 추가의 변형들은, 그 전체가 본 명세서에 참고로 포함된 US20190316121에서 확인할 수 있다.

[0455] 일부 구현예에서, 구조-가이드된 계통적 접근법(예를 들어, 문헌[Mir et al. *Nat Commun* 9:2641 (2018)]에 기재된 바와 같음; 그 전체가 본 명세서에 참고로 포함됨)이 주형 RNA에 대한 변형을 찾기 위해 이용된다. 구현예에서, 변형은 주형 RNA의 가이드 영역의 포함 또는 배제로 확인된다. 일부 구현예에서, 주형 RNA에 결합된 폴리펩티드의 구조는, 이후에 예를 들어 RNA와 폴리펩티드의 회합을 붕괴할 위험이 더 적은, 변형에 대해 선택될 수 있는 RNA의 비-단백질 접촉된 뉴클레오티드를 결정하는 데 사용된다. 주형 RNA에서 제2 구조는 또한, 예를 들어 변형, 예를 들어 헤어핀, 스템 및/또는 벌지를 선택하기 위한 2차 구조를 결정하기 위해, 소프트웨어 도구, 예를 들어 rrna.urmc.rochester.edu/RNAstructureWeb에서 입수가능한 RNAstructure 도구에 의해 인실리코로 예측될 수도 있다(Bellaousov et al. *Nucleic Acids Res* 41:W471-W474 (2013); 그 전체가 본 명세서에 참고로 포함됨).

[0456] ***조성물 및 시스템의 생산***

[0457] 당업자에 의해 이해되는 바와 같이, 핵산 작제물 및 단백질 또는 폴리펩티드를 설계 및 구축하는 방법(예컨대, 본 명세서에 기재된 시스템, 작제물 및 폴리펩티드)은 당업계에서 일상적이다. 일반적으로, 재조합 방법이 사용될 수 있다. 일반적으로 문헌[Smales & James (Eds.), *Therapeutic Proteins: Methods and Protocols* (Methods in Molecular Biology), Humana Press (2005)]; 및 [Crommelin, Sindelar & Meibohm (Eds.), *Pharmaceutical Biotechnology: Fundamentals and Applications*, Springer (2013)]을 참조한다. 핵산 조성물의 설계, 제조, 평

가, 정제 및 조작 방법은 문헌[Green and Sambrook (Eds.), *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012)]에 기재되어 있다.

[0458] 본 개시내용은 부분적으로는 핵산, 예를 들어 본 명세서에 기재된 유전자 변형 폴리펩티드를 인코딩하는, 벡터, 본 명세서에 기재된 주형 핵산, 또는 둘 다를 제공한다. 일부 구현예에서, 벡터는 선택적 마커, 예를 들어 항생제 내성 마커를 포함한다. 일부 구현예에서, 항생제 내성 마커는 카나마이신 내성 마커이다. 일부 구현예에서, 항생제 내성 마커는 베타-락탐 항생제에 대한 내성을 부여하지 않는다. 일부 구현예에서, 벡터는 암피실린 내성 마커를 포함하지 않는다. 일부 구현예에서, 벡터는 카나마이신 내성 마커를 포함하고, 암피실린 내성 마커는 포함하지 않는다. 일부 구현예에서, 유전자 변형 폴리펩티드를 인코딩하는 벡터는 (예를 들어, 표적 세포, 조직, 기관 또는 대상체에의 투여시) 표적 세포 게놈 내로 통합된다. 일부 구현예에서, 유전자 변형 폴리펩티드를 인코딩하는 벡터는 (예를 들어, 표적 세포, 조직, 기관 또는 대상체에의 투여시) 표적 세포 게놈 내로 통합되지 않는다. 일부 구현예에서, 주형 핵산(예를 들어, 주형 RNA)을 인코딩하는 벡터는 (예를 들어, 표적 세포, 조직, 기관 또는 대상체에의 투여시) 표적 세포 게놈 내로 통합되지 않는다. 일부 구현예에서, 벡터가 표적 세포 게놈에서 표적 부위 내로 통합되는 경우, 선택적 마커는 게놈 내로 통합되지 않는다. 일부 구현예에서, 벡터가 표적 세포 게놈에서 표적 부위 내로 내로 통합되는 경우, 벡터 유지에 연관된 유전자 또는 서열(예를 들어, 플라스미드 유지 유전자)은 게놈 내로 통합되지 않는다. 일부 구현예에서, 벡터가 표적 세포 게놈에서 표적 부위 내로 통합되는 경우, 수송 조절 서열(예를 들어, AAV로부터의, 예를 들어 역 말단 반복부)은 게놈 내로 통합되지 않는다. 일부 구현예에서, 표적 세포, 조직, 기관, 또는 대상체에 대한 (예를 들어, 본 명세서에 기재된 유전자 변형 폴리펩티드, 본 명세서에 기재된 주형 핵산, 또는 둘 다를 인코딩하는) 벡터의 투여는 상기 표적 세포, 조직, 기관 또는 대상체의 게놈(들)에서 하나 이상의 표적 부위 내로 벡터의 부분의 통합을 초래한다. 일부 구현예에서, 통합 물질을 포함하는 표적 부위의 99, 95, 90, 80, 70, 60, 50, 40, 30, 20, 10, 5, 4, 3, 2, 또는 1% 미만(예를 들어, 표적 부위 없음)은 선택 마커(예를 들어, 항생제 내성 유전자), 수송 조절 서열(예를 들어, AAV 유래의, 예를 들어 역 말단 반복부), 또는 벡터 유래의 둘 다를 포함한다.

[0459] 재조합 단백질은 또한 곤충 세포, 효모, 박테리아 또는 기타 세포를 사용하여 적절한 프로모터 제어 하에 생산될 수도 있지만, 본 명세서에 기재된 치료적 약학 단백질 또는 폴리펩티드를 생산하기 위한 예시적인 방법은 포유동물 세포에서의 발현을 포함한다. 포유동물 발현 벡터는 복제 기점, 적합한 프로모터, 및 기타 5' 또는 3' 측접하는 비-전사 서열과 같은 비-전사 요소들, 및 필요한 리보솜 결합 부위, 폴리아데닐화 부위, 스플라이스 공여체 및 수용체 부위, 및 종료 서열과 같은 5' 또는 3' 비번역 서열을 포함할 수 있다. SV40 바이러스 게놈, 예를 들어 SV40 기원으로부터 유래된 DNA 서열, 초기 프로모터, 스플라이스 및 폴리아데닐화 부위는 이종 DNA 서열의 발현에 필요한 기타 유전자 요소를 제공하는 데 사용될 수 있다. 박테리아, 진균, 효모, 및 포유동물 세포 숙주와의 사용을 위한 적절한 클로닝과 발현 벡터가 문헌[Green & Sambrook, *Molecular Cloning: A Laboratory Manual* (Fourth Edition), Cold Spring Harbor Laboratory Press (2012)]에 기재되어 있다.

[0460] 다양한 포유동물 세포 배양 시스템이 재조합 단백질 발현 및 제조를 위해 사용될 수 있다. 포유동물 발현 시스템의 예는 CHO, COS, HEK293, HeLA, 및 BHK 세포주를 포함한다. 단백질 치료제의 생산을 위한 숙주 세포 배양의 과정은 문헌[Zhou and Kantardjieff (Eds.), *Mammalian Cell Cultures for Biologics Manufacturing (Advances in Biochemical Engineering/Biotechnology)*, Springer (2014)]에 기재되어 있다. 본 명세서에 기재된 조성물은 재조합 단백질을 인코딩하는, 벡터, 예컨대 바이러스 벡터, 예를 들어 렌티바이러스 벡터를 포함할 수 있다. 일부 구현예에서, 벡터, 예를 들어 바이러스 벡터는 재조합 단백질을 인코딩하는 핵산을 포함할 수 있다.

[0461] 단백질 치료제의 정제는 문헌[Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); and in Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010)]에 기재되어 있다.

[0462] 본 개시내용은 또한 유전자 변형 폴리펩티드 및/또는 게놈 표적 부위에 대한 특이성을 갖는 주형 핵산 분자(예를 들어, 주형 RNA)의 생산을 위한 조성물 및 방법을 제공한다. 일 양태에서, 방법은 상류 상동성 절편, 이종 목적 서열 절편, 유전자 변형 폴리펩티드 결합 모티프, 및 gRNA 절편을 포함하여 RNA 절편의 생산을 포함한다.

[0463] ***적용***

[0464] ***치료 적용***

[0465] 본 명세서에 기재된 바와 같은 유전자 변형 시스템은 세포(예를 들어, 동물 세포, 식물 세포, 또는 진균 세포)

를 변형시키는 데 사용될 수 있다. 본 명세서에 기재된 바와 같은 유전자 변형 시스템은 포유동물 세포(예를 들어, 인간 세포)를 변형시키는 데 사용될 수 있다. 일부 구현예에서, 본 명세서에 기재된 바와 같은 유전자 변형 시스템은 가축 동물(예를 들어, 소, 말, 양, 염소, 돼지 라마, 알파카, 낙타, 야크, 닭, 오리, 거위, 또는 타조) 유래의 세포를 변형시키는 데 사용될 수 있다. 일부 구현예에서, 본 명세서에 기재된 바와 같은 유전자 변형 시스템은, 예를 들어 동물 세포, 예를 들어 포유동물 세포(예를 들어, 인간 세포), 식물 세포, 또는 진균 세포를 변형시키는 데, 실험실 도구 또는 연구 도구로 사용될 수 있거나, 실험실 방법 또는 연구 방법에서 사용될 수 있다.

[0466] 코딩 유전자를 RNA 서열 주형 내로 통합시킴으로써, 유전자 변형 시스템은, 예를 들어 기능 손실 돌연변이를 갖는 개인에서 치료용 이식유전자의 발현을 제공함으로써, 기능 수득 돌연변이를 정상의 이식유전자로 대체함으로써, 기능 수득 돌연변이 발현을 제거하기 위한 조절 서열을 제공함으로써, 및/또는 작동가능하게 연결된 유전자, 이식유전자 및 이의 시스템의 발현을 제어함으로써 치료적 필요를 다룰 수 있다. 특정 구현예에서, RNA 서열 주형은 숙주 세포의 치료적 필요에 특이적인 프로모터 영역, 예를 들어 조직 특이적 프로모터 또는 인핸서를 인코딩한다. 또 다른 구현예에서, 프로모터는 코딩 서열에 작동가능하게 연결될 수 있다.

[0467] 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 표적 유전자의 발현(예를 들어, 전사 또는 번역)을 증가시키거나 감소시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 프로모터 또는 인핸서, 예를 들어 전사 인자에 결합하는 서열에서 서열을 변경, 추가, 또는 결실시킴으로써 표적 유전자의 발현(예를 들어, 전사 또는 번역)을 증가시키거나 감소시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 표적 유전자의 번역을 변경시키고(예를 들어, 아미노산 서열을 변경함), 시작 또는 정지 코돈을 삽입 또는 결실시키고, 유전자의 번역 프레임을 변경 또는 고정시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은, 예를 들어 스플라이스 어셉터 수여 또는 공여 부위를 삽입, 결실, 또는 변경시킴으로써, 표적 유전자의 스플라이싱을 변경시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 전사체 또는 단백질 반감기를 변경시킨다. 일부 구현예에서, 삽입, 결실, 치환, 또는 이의 조합은 표적 유전자, 예를 들어 표적 유전자에 의해 인코딩된 단백질의 활성을 변경, 증가, 감소시킨다.

[0468] *보상 편집*

[0469] 일부 구현예에서, 본 명세서에 제공된 시스템 또는 방법은 보상 편집을 도입하는 데 사용될 수 있다. 일부 구현예에서, 보상 편집은 질환 또는 장애와 연관된 유전자의 위치에 존재하며, 이는 질환-유발 돌연변이의 위치와 상이하다. 일부 구현예에서, 보상 돌연변이는 원인성 돌연변이를 함유하는 유전자 내에는 없다. 일부 구현예에서, 보상 편집은 질환-유발 돌연변이를 무효화할 수 있거나 이를 보상할 수 있다. 일부 구현예에서, 보상 편집은 질환-유발 돌연변이의 돌연변이체 효과를 억제 또는 역전시키기 위해 본 명세서에 제공된 시스템 또는 방법에 의해 도입될 수 있다.

[0470] *조절 편집*

[0471] 일부 구현예에서, 본 명세서에 제공된 시스템 또는 방법은 조절 편집을 도입하는 데 사용될 수 있다. 일부 구현예에서, 조절 편집은 유전자의 조절 서열, 예를 들어 유전자 프로모터, 유전자 인핸서, 유전자 억제자, 또는 유전자 스플라이싱을 조절하는 서열에 도입된다. 일부 구현예에서, 조절 편집은 표적 유전자의 발현 수준을 증가 또는 감소시킨다. 일부 구현예에서, 표적 유전자는 질환-유발 돌연변이를 함유하는 유전자와 동일하다. 일부 구현예에서, 표적 유전자는 질환-유발 돌연변이를 함유하는 유전자와 상이하다.

[0472] *반복부 확장 질환*

[0473] 일부 구현예에서, 본 명세서에 제공된 시스템 또는 방법은 반복부 확장 질환을 치료하는 데 사용될 수 있다. 일부 구현예에서, 예를 들어 유전자 변형 폴리퀩티드를 포함하는, 본 명세서에 제공된 시스템 또는 방법은 맞춤화된 RNA 주형에 따라 유전자좌에서 반복부의 수를 재설정함으로써 반복부 확장 질환을 치료하는 데 사용될 수 있다.

[0474] *치료 적응증*

[0475] 일부 구현예에서, 본 명세서에 제공된 시스템 또는 방법은 하기 표 12 내지 15 중 임의의 것의 적응증을 치료하는 데 사용될 수 있다. 예를 들어, 일부 구현예에서 유전자 변형 시스템은 세포 내 게놈 DNA의 표적 부위를 변형시키며, 여기서 표적 부위는 표 12 내지 15 중 임의의 것의 유전자에, 예를 들어 표 12 내지 15 중 임의의 것에 열거된 상응하는 적응증을 갖는 대상체에 있다. 일부 구현예에서, 세포는 간 세포이고, 표적 부위는 표 12의 유전자에, 예를 들어 표 12에 열거된 상응하는 적응증을 갖는 대상체에 있다. 일부 구현예에서, 세포는 조혈 줄

기 세포(HSC)이고, 표적 부위는 표 13의 유전자에, 예를 들어 표 13에 열거된 상응하는 적응증을 갖는 대상체에 있다. 일부 구현예에서, 세포는 중추신경계(CNS) 세포이고, 표적 부위는 표 14의 유전자에, 예를 들어 표 14에 열거된 상응하는 적응증을 갖는 대상체에 있다. 일부 구현예에서, 세포는 눈의 세포이고, 표적 부위는 표 15의 유전자에, 예를 들어 표 15에 열거된 상응하는 적응증을 갖는 대상체에 있다. 일부 구현예에서, 표적 부위는 유전자의 코딩 영역에 있다. 일부 구현예에서, 표적 부위는 프로모터에 있다. 일부 구현예에서, 표적 부위는 표 12 내지 15 중 임의의 것의 유전자의 5' UTR 또는 3' UTR에 있다. 일부 구현예에서, 표적 부위는 유전자의 인트론 또는 엑손에 있다. 일부 구현예에서, 유전자 변형 시스템은 유전자의 돌연변이를 수정한다. 일부 구현예에서, 유전자 변형 폴리펩티드는 (예를 들어, 질환-유발 돌연변이를 통해) 유전자로부터 결실되었던 서열을 삽입한다. 일부 구현예에서, 유전자 변형 시스템은 유전자에서 (예를 들어, 질환-유발 돌연변이를 통해) 복제되었던 서열을 결실시킨다. 일부 구현예에서, 유전자 변형 시스템은 돌연변이(예를 들어, 질환-유발 돌연변이)를 상응하는 야생형 서열로 대체한다. 일부 구현예에서, 돌연변이는 치환, 삽입, 결실 또는 역전이다.

[0476] **표 12: 예를 들어, 간에서의 적응증 및 유전자 표적**

표 12

[0477]

질환	**침범된 유전자**
급성 간헐성 포르피린증	HMBS
알파-1-항트립신 결핍(AAT)	SERPINA1
아르기나제 결핍	ARG1
아르기노숙시네이트 리아제 결핍	ASL
카바모일 포스페이트 합성효소 I 결핍	CPS1
시트린 결핍	SLC25A13
시트룰린혈증 I형	ASS1
크리글러-나자르(Crigler-Najjar) 증후군(고빌리루빈혈증)	UGT1A1
파브리 병	GLA
가족성 과콜레스테롤혈증 4(동형접합성 가족성 콜레스테롤혈증)	LDLRAP1
글루타르산뇨증 I	GCDH
글루타르산뇨증 II(다중 아실-CoA 탈수소효소 결핍)	GA IIA: ETFA GA IIB: ETFB GA IIC: ETFDH
글리코겐 축적병 IV형	GBE1
혈우병 A	F8
혈우병 B	F9
유전성 혈색소증	HFE
호모시스틴뇨증	CBS
메이플시럽뇨 질환(MSUD)	Ia형: BCKDHA Ib형: BCKDHB II형: DBT
메틸말론산 혈증(메틸말로닐-CoA 뮤타제 결핍)	MMUT
MPS 1S(샤이에(Scheie) 증후군)	IDUA
MPS2	IDS
MPS3(산 필리포 신드롬)	IIIa형: SGSH IIIb형: NAGLU IIIc형: HGSNAT IIId형: GNS
MPS4	IVA형: GALNS IVB형: GLB1
MPS6	ARSB
MPS7	GUSB
오르니틴 트랜스카바밀라제 결핍	OTC
페닐케톤뇨증(페닐알라닌 히드록실라제 결핍)	PAH
다낭성 간 질환	PCLD1: PRKCSH PCLD2: SEC63 PLCD3: ALG8 PCLD4: LRP5
폼페병	GAA

원발성 수산과잉뇨 1(수산증)	AGXT
진행성 가족성 간내 수산과잉뇨 1형	ATP8B1
진행성 가족성 간내 수산과잉뇨 2형	ABCB11
진행성 가족성 간내 수산과잉뇨 3형	ABCB4
프로피온산혈증	PCCB; PCCA
피루베이트 카르복실라제 결핍	PC
티로신혈증 I형	FAH
윌슨 병	ATP7B

[0478] **표 13: HSC에 대한 적응증 및 유전자 표적**

표 13

[0479]

질환	침범된 유전자
부신백질이영양증(CALD)	ABCD1
알파-만노사이드 축적증	MAN2B1
블랙판-다이아몬드 빈혈	
선천성 무거대핵 혈소판감소증	MPL
선천성 이상각화증	TERC
판코니 빈혈	FANC
고셔병	GBA
공세포 백색질 장애(크라베(Krabbe) 병)	GALC
혈구탐식성 림프조직구 증식증	PRF1; STX11; STXBP2; UNC13D
악성 영아골다공증- 상염색체 열성 골다공증	많은 유전자들이 연루됨
이염성 백질이영양증	PSAP
MPS 1S(샤이에(Scheie) 증후군)	IDUA
MPS2	IDS
MPS7	GUSB
뮤코지질증 II	GNPTAB
니만-피크(Niemann-Pick) 병 A 및 B	SMPD1
니만-피크 병 C	NPC1
폼페병	GAA
피루베이트 키나제 결핍(PKD)	PKLR
겸상 세포 질환(SCD)	HBB
테이-삭스	HEXA
지중해빈혈	HBB

[0480] **표 14: CNS에 대한 적응증 및 유전자 표적**

표 14

[0481]

질환	침범된 유전자
알파-만노사이드 축적증	MAN2B1
운동실조-모세혈관확장증 증후군	ATM
CADASIL	NOTCH3
카나반 병	ASPA
카르바모일-포스페이트 합성효소 1 결핍	CPS1
CLN1 병	PPT1
CLN2 병	TPP1
CLN3 병(소아 신경세포 세로이드 라이포푸스신증, 바텐(Batten)병)	CLN3
코핀-로우리(Coffin-Lowry) 증후군	RPS6KA3
선천성 근무력증 증후군 5	COLQ
코넬리아디란지 증후군(NIPBL)	NIPBL
코넬리아디란지 증후군(SMC1A)	SMC1A
드라베 증후군(SCN1A)	SCN1A
글리신 뇌병증(GLDC)	GLDC
GM1 강글리오시드축적증	GLB1

헌팅턴 병	HTT
실비우스의 수도관 협착을 동반한 수두증	L1CAM
라이(Leigh) 증후군	SURF1
이염성 백질이영양증(ARSA)	ARSA
MPS 2형	IDS
MPS 3형	3a형: SGSH 3b형: NAGLU
뮤코지질증 IV	MCOLN1
신경섬유종증 1형	NF1
신경섬유종증 2형	NF2
판토텐네이트 키나아제 연관 신경퇴화	PANK2
피리독신-의존성 뇌전증	ALDH7A1
레트 증후군(MECP2)	MECP2
샌드호프(sandhoff) 병	HEXB
의미 치매(전두측두엽 치매)	MAPT
거대 축삭신경병증을 동반한 척수소뇌 운동실조증(안구운동 행위상실증을 동반한 운동실조증)	SETX
테이-삭스병	HEXA
X-연관 부신백질이영양증	ABCD1

[0482] **표 15: 눈에 대한 적응증 및 유전자 표적**

표 15

[0483]

질환	**침범된 유전자**
색맹	CNGB3
선천성 흑내장(LCA1)	GUCY2D
선천성 흑내장(LCA10)	CEP290
선천성 흑내장(LCA2)	RPE65
선천성 흑내장(LCA8)	CRB1
맥락막 결손	CHM
원뿔-막대 세포 이상증(ABCA4)	ABCA4
원뿔-막대 세포 이상증(GUCY2D)	GUCY2D
시스틴증, 눈 비신장병	CTNS
도이네(Doyne) 벌집형 망막 이상증(DHRD)	EFEMP1
가족성 눈연수막 아밀로이드증	TTR
각막염-어린선-난청(KID)	GJB2
격자 각막 이상증 I형	TGFBI
반점 각막 이상증(MCD)	CHST6
미즈만(Meesmann) 각막 이상증	KRT12; KRT3
시신경 위축	OPA1
망막세포변성증(AR)	USH2A
망막세포변성증(AD)	RHO
소스비 펀두스(Sorsby Fundus) 이상증	TIMP3
스타르가르트 병	ABCA4

[0484] *식물에의 적용*

[0485] 일부 구현예에서, 본 명세서에 제공된 시스템 또는 방법은, 예를 들어 식물의 적합성을 증가시키기 위해 식물 또는 식물 부분(예를 들어, 잎, 뿌리, 꽃, 과실 또는 종자)을 변형하는 데 사용될 수 있다.

[0486] *식물로의 전달*

[0487] 본 명세서에 기재된 유전자 변형 시스템을 식물로 전달하는 방법이 본 명세서에서 제공된다. 식물 또는 이의 부분을 유전자 변형 시스템과 접촉시킴으로써 유전자 변형 시스템을 식물에 전달하는 방법이 포함된다. 본 방법은, 예를 들어 식물의 적합성을 증가시키기 위해 식물을 변형시키는 데 유용하다.

[0488] 더 구체적으로, 일부 구현예에서, 본 명세서에 기재된 핵산(예를 들어, 유전자 변형 시스템을 인코딩하는 핵

산)은 벡터 내에 인코딩될 수 있으며, 예를 들어 식물 프로모터, 예를 들어 식물 벡터(예를 들어, pHUC411) 내 옥수수 유비퀴틴 프로모터(ZmUBI)에 인접하여 삽입될 수 있다. 일부 구현예에서, 본 명세서에 기재된 핵산은 아그로박테리아(agrobacteria)를 통해 식물(예를 들어, 자포니카(japonica) 벼) 또는 식물의 부분(예를 들어, 식물의 캘러스(callus)) 내로 도입된다. 일부 구현예에서, 본 명세서에 기재된 시스템 및 방법은 식물 유전자(예를 들어, 하이그로마이신 포스포전이효소(HPT))를 비대립유전자(null allele)로 대체함으로써(예를 들어, 개시 코돈에서 염기 치환을 함유함) 식물에서 사용될 수 있다. 식물 게놈의 변형을 위한 시스템 및 방법은 문헌[Xu et. al. Development of plant prime-editing systems for precise genome editing, 2020, Plant Communications]에 기재되어 있다.

[0489] 일 양태에서, 식물의 적합성을 증가시키는 방법이 본 명세서에서 제공되며, 이 방법은 처리되지 않은 식물(예를 들어, 유전자 변형 시스템을 전달받지 않은 식물)에 비해 식물의 적합성을 증가시키기 위해 본 명세서에 기재된 유전자 변형 시스템을 (예를 들어, 효과적인 양 및 기간으로) 식물에 전달하는 단계를 포함한다.

[0490] 유전자 변형 시스템의 전달 결과 식물의 적합성의 증가는 다양한 방식으로, 예를 들어 이에 의해 식물의 더 양호한 생산, 예를 들어 개선된 산출량, 식물의 활력 또는 식물로부터 수확된 생성물의 품질의 개선, 농업 또는 원예에 바람직한 것으로 여겨지는 수확전 또는 수확 후 특성에서의 개선(예를 들어, 맛, 외관, 유통기한), 또는 다른 인간에게 유익한 특성들의 개선(예를 들어, 알레르겐 생성 감소)으로 표명될 수 있다. 식물의 개선된 산출량은, 동일 조건 하에서 생산되었지만 본 조성물의 적용이 없는 식물의 동일 생성물의 산출량에 비해, 또는 통상의 식물 변형제의 적용과 비교하여 측정가능한 양으로, 식물의 생성물의 산출량에서의 증가(예를 들어, 식물 바이오매스, 낟알, 종자 또는 과실 산출량, 단백질 함량, 탄수화물 또는 오일 함량 또는 잎 면적에 의해 측정되는 바와 같음)와 관련된다. 예를 들어, 산출량은 적어도 약 0.5%, 약 1%, 약 2%, 약 3%, 약 4%, 약 5%, 약 10%, 약 20%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 약 100%, 또는 100% 초과만큼 증가될 수 있다. 일부 경우에서, 방법은 처리되지 않은 식물에 비해, 산출량을 약 2x배, 5x배, 10x배, 25x배, 50x배, 75x배, 100x배, 또는 100x배 초과만큼 증가시키는 데 효과적이다. 산출량은 식물의 중량 또는 부피에 의한 양 또는 동일 기준에서의 식물의 생산물 측면에서 표현될 수 있다. 기준은 시간, 성장 면적, 생산된 식물의 중량, 또는 사용된 원료의 양의 측면에서 표현될 수 있다. 예를 들어, 그러한 방법은 종자, 과실, 낟알(kernel), 꼬투리(boll), 괴경, 뿌리 및 잎을 포함하지만 이에 제한되지 않는 식물 조직의 산출량을 증가시킬 수 있다.

[0491] 유전자 변형 시스템의 전달 결과로서 식물의 적합성의 증가는 또한, 동일 조건 하에서 그러나 본 조성물의 투여가 없거나 또는 통상의 식물 변형제를 적용하여 생산된 식물의 동일 인자에 비해 측정가능한 또는 현저한 양만큼의, 기타 수단들, 예컨대 활력 평가, 스탠드(stand)(단위 면적 당 식물의 수), 식물 높이, 줄기 둘레, 줄기 길이, 잎 수, 잎 크기, 식물 캐노피(canopy), 시각적 외관(예컨대 더 푸른 잎 색상), 뿌리 평가, 출아(emergence), 단백질 함량, 증가된 분열(tillering), 더 큰 잎, 더 많은 잎, 더 적은 죽은 기부 잎(basal leaves), 더 강한 틸러(tillers), 더 적게 요구되는 비료, 적게 요구되는 종자, 더 생산적인 틸러, 더 이른 개화, 이른 낟알 또는 종자 성숙, 적은 식물 버스(verse)(쓰러짐), 증가된 싹 생장, 조기 발아, 또는 이들 인자들의 임의의 조합의 증가 또는 개선에 의해 측정될 수도 있다.

[0492] 따라서, 본 명세서에 제공된 임의의 유전자 변형 시스템의 유효량을 식물에 전달하는 단계를 포함하는, 식물을 변형시키는 방법이 본 명세서에서 제공되며, 여기서 이 방법은 식물을 변형시키며, 이에 의해 식물에서 유익한 특성을, 미처리된 식물에 비해 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼) 도입 또는 증가시킨다. 특히, 이 방법은 식물의 적합성을 미처리된 식물에 비해 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼) 증가시킬 수 있다.

[0493] 일부 경우에서, 식물 적합성의 증가는, 질환 저항성, 내건성, 내열성, 저온내성, 내염성, 내금속성, 제초제 내성, 내화학성, 용수 효율, 질소 이용률, 질소 스트레스에 대한 내성, 질소 고정, 병해충내성, 초식동물 내성, 병원체 내성, 산출량, 물이 제한된 조건 하에서의 산출량, 활력, 성장, 광합성능, 영양, 단백질 함량, 탄수화물 함량, 오일 함량, 바이오매스, 싹 길이, 뿌리 길이, 뿌리 구조, 종자 종량, 또는 수확가능한 생성물의 양에서 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼의) 증가이다.

[0494] 일부 경우에, 적합성의 증가는 발달, 성장, 산출량, 비생물적 스트레스요인에 대한 내성, 또는 생물적 스트레스요인에 대한 내성에서 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼의) 증가이다. 비생물적 스트레스는, 예를 들어 건조 스트레스, 염 스트레스, 열 스트레스, 한냉

스트레스, 및 저영양 스트레스를 포함하는, 식물 또는 식물의 부분이 처하게 되는 환경적 스트레스 조건을 지칭한다. 생물적 스트레스는 식물 또는 식물 부분이 처하게 되는 환경적 스트레스 조건을 지칭하며, 예를 들어 선충 스트레스, 곤충 초식동물 스트레스, 진균성 병원체 스트레스, 세균성 병원체 스트레스, 또는 바이러스 병원체 스트레스를 포함한다. 스트레스는 일시적, 예를 들어 몇시간, 며칠, 몇달 또는 영구, 예를 들어 식물의 생애 동안일 수 있다.

[0495] 일부 경우에, 식물 적합성의 증가는 식물로부터 수확된 생성물의 품질에서 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼의) 증가이다. 예를 들어, 식물 적합성의 증가는 식물로부터 수확된 생성물의 상업적으로 선호되는 특징들(예를 들어, 맛 또는 외관)의 개선일 수 있다. 다른 경우들에서, 식물 적합성의 증가는 식물로부터 수확된 생성물의 유통 기한에서의 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼의) 증가이다.

[0496] 대안적으로, 적합성의 증가는, 알레르기원 생성에서의 감소와 같이, 인간 또는 동물의 건강에 유리한 특성의 변경일 수 있다. 예를 들어, 적합성의 증가는 동물(예를 들어, 인간)에서 면역 반응을 자극하는 알레르기원(예를 들어, 꽃가루) 생성의 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼의) 감소일 수 있다.

[0497] 식물의 변형(예를 들어, 적합성의 증가)은 하나 이상의 식물 부분의 변형으로부터 야기될 수 있다. 예를 들어, 식물은 식물의 잎, 종자, 꽃가루, 뿌리, 과실, 싹, 꽃, 세포, 원형질체 또는 조직(예를 들어, 분열 조직)을 접촉시킴으로써 변형될 수 있다. 그와 같이, 또 다른 양태에서 식물의 적합성을 증가시키는 방법이 본 명세서에서 제공되며, 이 방법은 식물의 꽃가루를 본 명세서의 임의의 식물-변형 조성물 유효량과 접촉시키는 단계를 포함하며, 여기서 이 방법은 미처리된 식물에 비해 식물의 적합성을 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼) 증가시킨다.

[0498] 또 다른 양태에서, 식물의 적합성을 증가시키는 방법이 본 명세서에서 제공되며, 이 방법은 식물의 종자를 본 명세서에 개시된 임의의 유전자 변형 시스템의 유효량과 접촉시키는 단계를 포함하며, 여기서 이 방법은 미처리된 식물에 비해 식물의 적합성을 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼) 증가시킨다.

[0499] 또 다른 양태에서, 식물의 원형질체를 본 명세서에 기재된 임의의 유전자 변형 시스템의 유효량과 접촉시키는 단계를 포함하는 방법이 본 명세서에서 제공되며, 여기서 이 방법은 미처리된 식물에 비해 식물의 적합성을 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼) 증가시킨다.

[0500] 추가의 양태에서, 식물의 적합성을 증가시키는 방법이 본 명세서에서 제공되며, 이 방법은 식물의 식물 세포를 본 명세서에 기재된 임의의 유전자 변형 시스템의 유효량과 접촉시키는 단계를 포함하며, 여기서 이 방법은 미처리된 식물에 비해 식물의 적합성을 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼) 증가시킨다.

[0501] 또 다른 양태에서 식물의 적합성을 증가시키는 방법이 본 명세서에서 제공되며, 이 방법은 식물의 분열 조직을 본 명세서의 임의의 식물-변형 조성물 유효량과 접촉시키는 단계를 포함하며, 여기서 이 방법은 미처리된 식물에 비해 식물의 적합성을 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼) 증가시킨다.

[0502] 또 다른 양태에서 식물의 적합성을 증가시키는 방법이 본 명세서에서 제공되며, 이 방법은 식물의 배아를 본 명세서의 임의의 식물-변형 조성물 유효량과 접촉시키는 단계를 포함하며, 여기서 이 방법은 미처리된 식물에 비해 식물의 적합성을 (예를 들어, 약 1%, 2%, 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, 또는 100% 초과만큼) 증가시킨다.

[0503] *식물 투여 방법*

[0504] 본 명세서에 기재된 식물은 본 명세서에 기재된 임의의 유전자 변형 시스템 조성물에, 그 조성물이 식물에 전달 또는 투여되는 것을 가능하게 하는 임의의 적합한 방식으로 노출될 수 있다. 유전자 변형 시스템은 단독으로 또는 기타 활성 물질(예를 들어, 수정화제) 또는 비활성 물질과 조합되어 전달될 수 있으며, 식물-변형 조성물의 유효 농도를 전달하도록 제형화된, 농축 액체, 겔, 용액, 현탁액, 분무액, 분말, 펠렛, 조개탄, 벽돌 등의 형태로, 예를 들어 식물 전체에 걸쳐, 분무, 주사(예를 들어, 미세주입), 붓거나, 침지에 의해 적용될 수 있다. 본 명세서에 기재된 조성물의 적용을 위한 양 및 위치는 일반적으로, 식물의 서식지, 식물-변형 조성물에 의해 표

적화될 수 있는 식물의 생애주기 중 단계, 적용이 이루어지는 부위, 및 식물-변형 조성물의 물리적 및 기능적 특성에 의해 결정된다.

[0505] 일부 경우에, 조성물은, 예를 들어 백팩 분무, 공중 분무, 작물 분무/더스팅(dusting) 등에 의해 식물, 예를 들어 작물 상에 직접 분무된다. 유전자 변형 시스템이 식물에 전달되는 경우에, 유전자 변형 시스템을 받는 식물은 임의의 식물 성장 단계에 있을 수 있다. 예를 들어, 제형화된 식물-변형 조성물은 식물 성장의 초기 단계에서 종자-코팅 또는 뿌리 처리로서 또는 작물 주기의 후기 단계에서 전체 식물 처리로서 적용될 수 있다. 일부 경우에, 식물-변형 조성물은 식물에 대해 국소제로서 적용될 수 있다.

[0506] 나아가, 유전자 변형 시스템은 (예를 들어, 식물이 성장하는 토양에서, 또는 식물에 물을 주는 데 사용되는 물 중에서), 식물의 조직을 통해 흡수 및 분포되는 전신투여제로서 적용될 수 있다. 일부 경우에, 식물 또는 식용 유기체는 유전자 변형 시스템을 발현하도록 유전적으로 형질전환될 수 있다.

[0507] 지연된 또는 연속 방출은 또한 유전자 변형 시스템 또는 조성물을 용해가능하거나 생체부식성 코팅 층, 예컨대 젤라틴을 포함하는 식물-변형 조성물(들)로 코팅하고, 이 코팅은 사용 환경에서 용해 또는 부식하여, 식물-변형 조성물 위치를 이용가능하도록 만들거나, 또는 용해가능하거나 부식성 기질 내에 작용제를 분산시킴으로써 달성될 수 있다. 그러한 연속 방출 및/또는 분배 수단 장치는 본 명세서에 기재된 하나 이상의 식물-변형 조성물의 유효 농도를 일관되게 유지하기 위해 유리하게 이용될 수 있다.

[0508] 일부 경우에, 유전자 변형 시스템은 식물의 부분, 예를 들어 잎, 종자, 꽃가루, 뿌리, 과실, 싹, 또는 꽃 또는 이의 조직, 세포 또는 원형질체에 전달된다. 일부 경우에, 유전자 변형 시스템은 식물의 세포로 전달된다. 일부 경우에, 유전자 전달 시스템은 식물의 원형질체에 전달된다. 일부 경우에, 유전자 전달 시스템은 식물의 조직에 전달된다. 일부 경우에, 유전자 전달 시스템은 식물의 조직에 전달된다. 예를 들어, 조성물은 식물의 분열 조직(예를 들어, 정단 분열조직, 측부 분열조직, 또는 절간 분열조직(intercalary meristem))에 전달될 수 있다. 일부 경우에, 조성물은 식물의 영구 조직(예를 들어 단순 조직(예를 들어, 유조직(parenchyma), 후각세포(collenchyma), 또는 후막조직(sclerenchyma)) 또는 복합 영구 조직(예를 들어, 목질부 또는 체관부))에 전달된다. 일부 경우에, 유전자 변형 시스템은 식물 배아에 전달된다.

[0509] *식물*

[0510] 본 명세서에 기재된 유전자 변형 시스템이 다양한 식물로 전달되거나, 이를 이용하여 다양한 식물이 처리될 수 있다. 본 방법에 따라 유전자 변형 시스템으로 전달될 수 있는(즉, "처리된") 식물은 전체 식물 또는 이의 일부를 포함하며, 이는 싹 영양기관/구조체/구조(예를 들어, 잎, 스템 및 괴경), 뿌리, 꽃 및 꽃 기관/구조(예를 들어, 포엽, 꽃받침, 꽃잎, 수술, 암술, 꽃밥 및 씨방), 종자(배아, 배유(endosperm), 자엽(cotyledon), 및 종자 외피를 포함) 및 과실(성숙한 난소), 식물 조직(예를 들어, 유관속 조직, 기본 조직 등) 및 세포(예를 들어, 공변 세포(guard cells), 난세포 등), 및 그의 자손을 포함하지만 이에 제한되지 않는다. 식물 부분은 싹, 뿌리, 스템, 종자, 턱잎(stipules), 잎, 꽃잎, 꽃, 씨방, 포엽, 가지, 잎자루, 절간(internodes), 나무껍질, 솜털(pubescence), 틸러, 근경, 양치잎(frond), 블레이드(blade), 꽃가루, 수술 등과 같은 식물의 부분을 추가로 지칭할 수 있다.

[0511] 본 명세서에 개시된 방법에서 처리될 수 있는 식물의 부류는 고등 식물과 하등 식물의 부류를 포함하며, 이는 속씨 식물(단자엽 및 쌍자엽 식물), 겉씨 식물, 양치류, 쇠뜨기, 고생솔잎난류(psilophytes), 라이코파이트(lycophytes), 이끼류, 및 조류(예를 들어, 다세포 또는 단세포 조류)를 포함한다. 본 발명에 따라 처리될 수 있는 식물은 임의의 관다발 식물, 예를 들어 단자엽식물 또는 쌍자엽 식물 또는 겉씨 식물을 더 포함하며, 이는 이에 제한되지는 않지만, 알팔파, 사과, 아기장대, 바나나, 보리, 카놀라, 피마자, 국화, 클로버, 코코아, 커피, 목화, 목화씨, 옥수수, 크렘베(crambe), 크랜베리, 오이, 덴드로븀(dendrobium), 디오스코레아(dioscorea), 유칼립투스, 페스큐(fescue), 아마, 글라디올러스, 백합과, 아마인, 밀레(millet), 머스크멜론, 겨자, 귀리, 오일 야자, 유채, 파파야, 땅콩, 파인애플, 관상식물, 강낭콩 속, 감자, 유채(rapeseed), 벼, 호밀, 라이그래스(ryegrass), 잇꽃, 참깨, 수수, 대두, 사탕무, 사탕수수, 해바라기, 딸기, 담배, 토마토, 잔디(turfgrass), 밀과 채소 작물 예컨대 상추, 셀러리, 브로콜리, 콜리플라워, 조롱박(cucurbit); 과실 및 견과 나무, 예컨대 사과, 배, 복숭아, 오렌지, 자몽, 레몬, 라임, 아몬드, 피칸, 호두, 헤이즐; 덩굴류, 예컨대 포도(예를 들어 포도밭), 키위, 호프; 과실 관목 및 브램블(bramble), 예컨대 라스베리, 블랙베리, 구스베리; 삼림수 예컨대 물푸레나무, 소나무, 전나무, 단풍나무, 오크, 밤나무, 포플러(popular); 이와 함께, 알팔파, 카놀라, 피마자, 옥수수, 목화, 크렘베, 아마, 아마인, 겨자, 오일야자, 유채, 땅콩, 감자, 벼, 잇꽃, 참깨, 대두, 사탕무, 해바라기, 담배 및 밀을 포함한다. 본 발명의 방법에 따라 처리될 수 있는 식물은 임의의 농작물,

예를 들어 사료 작물, 유지 작물, 곡물, 유실 작물, 채소 작물, 섬유 작물, 향신 작물, 견과 작물, 잔디 작물, 설탕 작물, 음료 작물 및 수풀을 포함한다. 특정 경우에, 본 방법에서 처리되는 농작물은 대두 식물이다. 다른 특정 경우에, 농작물은 밀이다. 특정 경우에, 농작물은 옥수수이다. 특정 경우에, 농작물은 목화이다. 특정 경우에, 농작물은 알팔파이다. 특정 경우에, 농작물은 사탕무이다. 특정 예에서, 농작물은 벼다. 특정 경우에, 농작물은 감자이다. 특정 경우에, 농작물은 토마토이다.

[0512] 특정 경우에, 식물은 작물이다. 그러한 작물 식물의 예는, 단자엽 및 쌍자엽 식물을 포함하며, 이에 제한되지는 않지만, 다음으로부터 선택되는 꼴(fodder) 또는 사료용 콩류, 관상 식물, 농작물, 나무, 또는 관목류: 단풍나무류(에이서(Acer) spp.), 알리움(Allium spp.), 아마란스(Amaranthus spp.), 파인애플(아나나스 코모서스(Ananas comosus)), 샐러리(아피움 그라베오렌스(Apium graveolens)), 땅콩류(아라키스(Arachis) spp)), 아스파라거스 오피시날리스(Asparagus officinalis), 비트(베타 불가리스(Beta vulgaris)), 브라시카류(Brassica spp.)(예를 들어, 브라시카 나푸스(Brassica napus)), 브라시카 라파(Brassica rapa ssp.)(카놀라, 유채, 순무 유채), 차(카멜리아 시넨시스(Camellia sinensis)), 칸나 인디카(Canna indica), 칸나비스 살리바(Cannabis saliva), 고추류(캅시쿰(Capsicum) spp.)), 밤류(카스타네아(Castanea spp.)), 치커리(치코리움 엔디비아(Cichorium endivia)), 수박(시트룰러스 라나터스(Citrullus lanatus), 감귤류(시트러스(Citrus) spp.), 코코스류(Cocos spp.), 커피류(Coffea spp.), 고수(코리앤드럼 사티붐(Coriandrum sativum)), 개암나무속(코릴러스(Corylus) spp.), 산사나무(크라태구스(Crataegus spp.)), 호박(Cucurbita(커큐비타) spp.), 커큐미스류(Cucumis spp.), 당근(다우쿠스 카로타(Daucus carota), 너도밤나무속(파구스(Fagus) spp.), 무화과(피쿠스 카리카(Ficus carica)), 딸기속(프라가리아(Fragaria) spp.), 은행나무(징코 빌로바(Ginkgo biloba)), 야생콩(글리신(Glycine) spp.)(예를 들어, 대두(글리신 맥스(Glycine max)), 소야 히스피다(Soja hispida) 또는 소야 맥스(Soja max)), 목화(고시피움 히르수툼(Gossypium hirsutum)), 해바라기 속(힐란터스(Helianthus) spp.(예를 들어, 해바라기(힐란터스 아누스(Helianthus annuus)), 무궁화속(히비스커스(Hibiscus) spp.), 보리속(호르데움(Hordeum) spp.)(예를 들어, 보리(호르데움 불가레(Hordeum vulgare)), 고구마(이포모에아 바타타스(Ipomoea batatas)), 호두(저글란(Juglans) spp.), 상추(락투카 사티바(Lactuca sativa), 아마(리눔 우시타티시뭄(Linum usitatissimum)), 리치(리치 치넨시스(Litchi chinensis)), 로터스류(Lotus spp.), 각수세미(루파 아쿠탕굴라(Luffa acutangula)), 루피너스류(Lupinus spp.), 토마토(라이코페르시콘(Lycopersicon) spp.)(예를 들어, 라이코페르시콘 에스쿨렌턴(Lycopersicon esculenturn), 라니코페르시콘 라이코페르시쿰(Lycopersicon lycopersicum), 라이코페르시콘 파이리포르메(Lycopersicon pyriforme)), 사과나무속(말루스(Malus) spp.), 자주개자리(메디카고 사티바(Medicago sativa)), 박하(멘타(Mentha) spp.), 참억새(미스칸투스 시넨시스(Miscanthus sinensis)), 검정 오디(머루스 니그라(Morus nigra)), 바나나(무사(Musa spp.)), 담배(니코타(Nicotiana) spp.), 올리브속(올레아(Olea) spp.), 쌀(오리자(Oryza) spp.)(예를 들어, 벼(오리자 사티바(Oryza sativa)), 벼속(오리자 라티폴리아(Oryza latifolia)), 기장(파니쿰 밀리아세움(Panicum miliaceum)), 마니쿰 버가툼(Panicum virgatum)), 괴물시계초(파시플로라 에듈리스(Passiflora edulis)), 파슬리(페트로셀리눔 크리스품(Petroselinum crispum)), 파세올러스류(Phaseolus spp.), 소나무(파이너스(Pinus) spp.), 피스타치오(피스타치아 베라(Pistacia vera)), 완두콩류(피섬(Pisum) spp.), 포아류(Poa spp.), 사시나무류(포퓰러스(Populus) spp.), 벚나무류(프루너스(Prunus) spp.), 배(파이러스 커뮤니스(Pyrus communis)), 참나무류(쿠어커스(Quercus) spp.), 무(라파누스 사티부스(Raphanus sativus), 대황(레움 라바바룸(Rheum rhabarbarum)), 까치밥나무류(리베스(Ribes) spp.), 피마자(리시너스 커뮤니스(Ricinus communis)), 산딸기속(루버스(Rubus) spp.), 사탕수수류(사카룸(Saccharum) spp.), 버들류(살릭스(Salix) sp.), 삼부커스류(Sambucus spp.), 호밀(세칼레 세레알레(Secale cereale)), 참깨류(세사뭄(Sesamum) spp.), 시나피스류(Sinapis spp.), 솔라넘류(Solanum spp.)(예를 들어, 감자(솔라넘 투베로숨(Solanum tuberosum), 땅고추(솔라넘 인테그리폴리움(Solanum integrifolium) 또는 토마토(솔라넘 리코페르시쿰(Solanum lycopersicum)), 수수(소검 바이컬러 (Sorghum bicolor)), 시리아수수새(소검 할레펜스(Sorghum halepense), 스피나치아류(Spinacia spp.), 타마린(타마린두스 인디카(Tamarindus indica)), 카카오(테오브로마 카카오(Theobroma cacao)), 토끼풀류(트리폴륨(Trifolium) spp.), 트리티코세칼레 림파위(Triticosecale rimpaui), 밀(트리티쿰(Triticum) spp.)(예를 들어, 일반 밀(트리티쿰 아에스티붐(Triticum aestivum)), 듀럼 밀(트리티쿰 듀럼(Triticum durum)), 트리티쿰 투르기둠(Triticum turgidum), 트리티쿰 하이버 (Triticum hybernum), 트리티쿰 마카(Triticum macha), 트리티쿰 사티붐(Triticum sativum) 또는 트리티쿰 불가레(Triticum vulgare)), 산앵도나무속(바키니움(Vaccinium) spp.), 비키아류(Vicia spp.), 녹두류(비그나(Vigna) spp.), 제비꽃(비올라 오도라타(Viola odorata), 포도(비티스(Vitis) spp.), 및 옥수수(Zea mays)를 포함하지만 이에 제한되지 않는다. 특정 구현예에서, 농작물은 벼, 유채, 카놀라, 대두, 옥수수(메이즈), 목화, 사탕수수, 알팔파, 수수, 또는 밀이다.

[0513] 본 발명에서의 사용을 위한 식물 또는 식물 부분은 식물 발달의 임의의 단계의 식물을 포함한다. 특정 경우에, 전달은 발아, 묘목 생장, 영양 생장, 및 생식 생장 단계 동안 일어날 수 있다. 특정 경우에, 식물로의 전달은 영양 및 생식 생장 동안 발생한다. 일부 경우에, 조성물은 식물의 꽃가루로 전달된다. 일부 경우에, 조성물은 식물의 종자로 전달된다. 일부 경우에, 조성물은 식물의 원형질체로 전달된다. 일부 경우에, 조성물은 식물의 조직으로 전달된다. 일부 경우에, 유전자 전달 시스템은 식물의 조직에 전달된다. 예를 들어, 조성물은 식물의 분열 조직(예를 들어, 정단 분열조직, 측부 분열조직, 또는 절간 분열조직)에 전달될 수 있다. 일부 경우에, 조성물은 식물의 영구 조직(예를 들어 단순 조직(예를 들어, 유조직, 후각세포, 또는 후막조직) 또는 복합 영구 조직(예를 들어, 목질부 또는 체관부))에 전달된다. 일부 경우에, 조성물은 식물 배아로 전달된다. 일부 경우에, 조성물은 식물 세포로 전달된다. 영양 생장 및 생식 생장의 단계는 또한 본 명세서에서 "성체" 또는 "성숙" 식물로서 지칭되기도 한다.

[0514] 유전자 변형 시스템이 식물 부분에 전달되는 경우, 식물 부분은 식물-변형제에 의해 변형될 수 있다. 대안적으로, 유전자 변형 시스템은 식물-변형제에 의해 순차적으로 변형된 식물의 다른 부분들에 (예를 들어, 식물의 순환 시스템에 의해) 분포될 수 있다.

[0515] ***투여 및 전달***

[0516] 본 명세서에 기재된 조성물 및 시스템은 시험과 내 또는 생체 내에서 사용될 수 있다. 일부 구현예에서, 시스템 또는 시스템의 구성요소는, 예를 들어 시험관 내 또는 생체 내에서, 세포(예를 들어, 포유동물 세포, 예를 들어 인간 세포)에 전달된다. 일부 구현예에서, 세포는 진핵 세포, 예를 들어 다세포 유기체, 예를 들어 식물 또는 동물, 예를 들어 포유동물(예를 들어, 인간, 돼지, 소), 조류(예를 들어, 가금류, 예컨대 닭, 칠면조, 또는 오리), 또는 어류의 세포이다. 일부 구현예에서, 세포는 비-인간 동물 세포(예를 들어, 실험실 동물, 가축 동물, 또는 반려 동물)이다. 일부 구현예에서, 세포는 줄기 세포(예를 들어, 조혈 줄기 세포), 섬유아세포, 또는 T 세포이다. 일부 구현예에서, 세포는 면역 세포, 예를 들어 T 세포(예를 들어, Treg, CD4, CD8, γδ 또는 기억 T 세포), B 세포(예를 들어, 기억 B 세포 또는 형질 세포), 또는 NK 세포이다. 일부 구현예에서, 세포는 비-분열 세포, 예를 들어 비-분열 섬유아세포 또는 비-분열 T 세포이다. 일부 구현예에서, 세포는 HSC이고 p53은 상향조절되지 않거나, 예를 들어 PCT/US2019/048607의 실시예 30에 기재된 방법에 따라 결정 시 10%, 5%, 2%, 또는 1% 미만으로 상향조절된다. 당업자는 유전자 변형 시스템의 구성요소가 폴리펩티드, 핵산(예를 들어, DNA, RNA), 및 이들의 조합의 형태로 전달될 수 있음을 이해할 것이다.

[0517] 일 구현예에서 시스템 및/또는 시스템의 구성요소는 핵산으로 전달된다. 예를 들어, 유전자 변형 폴리펩티드는 폴리펩티드를 인코딩하는 DNA 또는 RNA의 형태로 전달될 수 있고, 주형 RNA는 RNA 또는 RNA로 전사될 그의 상보적인 DNA의 형태로 전달될 수 있다. 일부 구현예에서, 시스템 또는 시스템의 구성요소는 1, 2, 3, 4개, 또는 그 이상의 별개의 핵산 분자로 전달된다. 일부 구현예에서 시스템 또는 시스템의 구성요소는 DNA와 RNA의 조합으로서 전달된다. 일부 구현예에서 시스템 또는 시스템의 구성요소는 DNA와 단백질의 조합으로서 전달된다. 일부 구현예에서 시스템 또는 시스템의 구성요소는 RNA와 단백질의 조합으로서 전달된다. 일부 구현예에서 유전자 변형 폴리펩티드는 단백질로서 전달된다.

[0518] 일부 구현예에서 시스템 또는 시스템의 구성요소는 벡터를 사용하여 세포, 예를 들어 포유동물 세포 또는 인간 세포로 전달된다. 벡터는, 예를 들어 플라스미드 또는 바이러스일 수 있다. 일부 구현예에서, 전달은 생체 내, 시험관 내, 생체 외 또는 제 위치에서 이루어진다. 일부 구현예에서 바이러스는 아데노 관련 바이러스(AAV), 렌티바이러스, 또는 아데노바이러스이다. 일부 구현예에서, 시스템 또는 시스템의 구성성분은 바이러스-유사 입자 또는 바이로솜을 이용하여 세포에 전달된다. 일부 구현예에서 전달은 하나 초과의 바이러스, 바이러스-유사 입자 또는 바이로솜을 사용한다.

[0519] 일 구현예에서, 본 명세서에 기재된 조성물 및 시스템은 리포좀 또는 다른 유사한 소포 내에 제형화될 수 있다. 리포좀은 내부 수성 구획을 둘러싸고 있는 단층 또는 다층 지질 이중층 및 상대적으로 불침투성인 외부 친유성 인지질 이중층으로 구성된 구형의 소포 구조이다. 리포좀은 음이온성, 중성 또는 양이온성일 수 있다. 리포좀은 생체적합성이고, 비독성이며, 친수성 및 친유성 약물 분자 둘 모두를 전달하고, 혈장 효소에 의한 분해로부터 그들의 카고를 보호하고, 생물학적 막 및 혈액-뇌 장벽(BBB)을 가로질러 그들의 로드를 수송할 수 있다(예를 들어, 검토를 위해 문헌[Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조).

[0520] 소포는 몇몇 상이한 유형의 지질로부터 제조될 수 있지만; 약물 담체로서 리포좀을 생성하는 데 인지질이 가장 흔히 사용된다. 다층 소포 지질을 제조하기 위한 방법이 당업계에 알려져 있다(예를 들어, 미국 특허 6,693,086

을 참조하며, 다층 소포 지질 제제에 관한 이의 교시내용은 본 명세서에 참고로 포함됨). 지질 필름이 수용액과 혼합될 때 소포 형성이 자발적일 수 있지만, 그것은 또한 균질기, 초음파처리기, 또는 압출 장치를 사용함으로써 전단 형태로 힘을 인가함으로써 촉진될 수 있다(예를 들어, 검토를 위해 문헌[Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679] 참조). 압출된 지질은 압출된 지질 제제에 관한 교시가 본 명세서에 참고로 포함되는 문헌[Templeton et al., Nature Biotech, 15:647-652, 1997]에 기재된 바와 같이 감소하는 크기의 필터를 통해 압출시킴으로써 제조될 수 있다.

[0521] 다양한 나노입자, 예컨대 리포좀, 지질 나노입자, 양이온성 지질 나노입자, 이온화성 지질 나노입자, 중합체성 나노입자, 금 나노입자, 덴드리머(dendrimer), 사이클로덱스트린 나노입자, 미셀 또는 상기의 조합이 전달을 위해 사용될 수 있다.

[0522] 지질 나노입자는 본 명세서에 기재된 약제학적 조성물을 위한 생체적합성 및 생분해성 전달 시스템을 제공하는 담체의 예이다. 나노구조화된 지질 담체(NLC)는 고체 지질 나노입자(SLN)의 특성을 보유하고, 약물 안정성 및 로딩 능력을 개선시키고, 약물 누출을 방지하는 변형된 고체 지질 나노입자(SLN)이다. 중합체 나노입자(PNP)는 약물 전달의 중요한 구성성분이다. 이들 나노입자는 특정 표적에 대한 약물 전달을 효과적으로 유도하고, 약물 안정성 및 약물의 제어 방출을 개선시킬 수 있다. 리포좀과 중합체를 조합한 담체의 유형인 지질-중합체 나노입자(PLN)가 또한 사용될 수 있다. 이들 나노입자는 PNP 및 리포좀의 상보적 이점을 가진다. PLN은 코어-셸 구조로 구성되며; 중합체 코어는 안정한 구조를 제공하고, 인지질 셸은 양호한 생체 적합성을 제공한다. 이와 같이, 2가지 구성성분은 약물 캡슐화 효율을 증가시키고, 표면 변형을 용이하게 하고, 수용성 약물의 누출을 방지한다. 검토를 위하여, 예를 들어, 문헌[Li et al. 2017, Nanomaterials 7, 122; doi:10.3390/nano7060122]을 참조한다.

[0523] 엑소좀도 또한, 본 명세서에 기재된 조성물 및 시스템을 위한 약물 운반 비히클로서 사용될 수 있다. 검토를 위해, 문헌[Ha et al. July 2016. Acta Pharmaceutica Sinica B. Volume 6, Issue 4, Pages 287-296; https://doi.org/10.1016/j.apsb.2016.02.001]을 참고한다.

[0524] 푸소좀은 표적 세포와 반응 및 융합하고, 따라서 다양한 분자에 대한 전달 비히클로 사용될 수 있다. 이들은 일반적으로 내강 또는 동공을 둘러싼 양친매성 지질의 이중층 및 양친매성 지질 이중층과 상호작용하는 푸소젠(fusogen)으로 이루어진다. 푸소젠 구성요소는 융합 및 페이로드 전달에 대한 표적 세포 특이성을 부여하여, 프로그램가능한 세포 특이성을 갖는 전달 비히클의 형성을 가능하게 하기 위해 조작가능한 것으로 나타났다(예를 들어, 특허출원 WO2020014209 참조, 푸소좀 설계, 제조, 및 용법에 관한 이의 교시 내용은 본 명세서에 참고로 포함됨).

[0525] 일부 구현예에서, 유전자 변형 시스템의 단백질 구성요소(들)는 주형 핵산(예를 들어, 주형 RNA)과 미리 회합될 수 있다. 예를 들어, 일부 구현예에서, 유전자 변형 폴리펩티드는 먼저 주형 핵산(예를 들어, 주형 RNA)과 조합되어, 리보핵단백질(RNP) 복합체를 형성할 수 있다. 일부 구현예에서, RNP는, 예를 들어 형질감염, 뉴클레오펙션, 바이러스, 소포, LNP, 엑소좀, 푸소좀을 통해 세포로 전달될 수 있다.

[0526] 유전자 변형 시스템은 세포, 조직 및 다세포 유기체로 도입될 수 있다. 일부 구현예에서, 시스템 또는 시스템의 구성요소는 기계적 수단 또는 물리적 수단을 통해 세포로 전달된다.

[0527] 단백질 치료제의 제형은 문헌[Meyer (Ed.), *Therapeutic Protein Drug Products: Practical Approaches to formulation in the Laboratory, Manufacturing, and the Clinic*, Woodhead Publishing Series (2012)]에 기재되어 있다.

[0528] *조직 특이적 활성/투여*

[0529] 일부 구현예에서, 본 명세서에 기재된 시스템은 하나 이상의 특징부(예를 들어, 프로모터 또는 마이크로RNA 결합 부위)를 사용하여 표적외 세포 또는 조직에서 활성을 제한할 수 있다.

[0530] 일부 구현예에서, 본 명세서에 기재된 핵산(예를 들어, 주형 RNA 또는 주형 RNA를 인코딩하는 DNA)은 프로모터 서열, 예를 들어 조직 특이적 프로모터 서열을 포함한다. 일부 구현예에서, 조직-특이적 프로모터는 유전자 변형 시스템의 표적-세포 특이성을 증가시키는 데 사용된다. 예를 들어, 프로모터는 표적 세포 유형에서 활성이지만 비-표적 세포 유형에서 활성이 아닌(또는 더 낮은 수준으로 활성인) 것에 기초하여 선택될 수 있다. 따라서, 프로모터가 비-표적 세포의 게놈 내로 통합되더라도, 통합된 유전자의 발현을 구동하지 않을 것이다(또는 단지 낮은 수준의 발현만 구동할 것임). 주형 RNA에 조직-특이적 프로모터 서열을 갖는 시스템은 또한, 예를 들어 본 명세서에 기재된 바와 같은, 예를 들어 주형 RNA 또는 유전자 변형 단백질을 인코딩하는 핵산에서, 마이크로RNA

결합 부위와 조합하여 사용될 수 있다. 주형 RNA에 조질-특이적 프로모터 서열을 갖는 시스템은 또한 조직-특이적 프로모터에 의해 구동되는, 유전자 변형 폴리펩티드를 인코딩하는 DNA와 조합하여 사용되어, 예를 들어 비-표적 세포보다 표적 세포에서 유전자 변형 단백질의 더 높은 수준을 달성할 수 있다. 일부 구현예에서, 예를 들어 간 적응증의 경우, 조직-특이적 프로모터는 본 명세서에 참고로 포함된 WO2020014209의 표 3으로부터 선택된다.

[0531] 일부 구현예에서, 본 명세서에 기재된 핵산(예를 들어, 주형 RNA 또는 주형 RNA를 인코딩하는 DNA)은 마이크로RNA 결합 부위를 포함한다. 일부 구현예에서, 마이크로RNA 결합 부위는 유전자 변형 시스템의 표적-세포 특이성을 증가시키는 데 사용된다. 예를 들어, 마이크로RNA 결합 부위는 비-표적 세포 유형에 존재하지만, 표적 세포 유형에는 존재하지 않는(또는 비-표적 세포에 비해 감소된 수준으로 존재하는) miRNA에 의해 인식된다는 것에 기초하여 선택될 수 있다. 따라서, 주형 RNA가 비-표적 세포에 존재하는 경우, miRNA에 의해 결합될 수 있고, 주형 RNA가 표적 세포에 존재하는 경우, 이는 miRNA에 의해 결합되지 않을 수 있다(또는 비-표적 세포에 대해 감소된 수준으로 결합한다). 이론에 구속되고자 하지는 않지만, miRNA의 주형 RNA에 대한 결합은 이의 활성을 간섭할 수 있고, 예를 들어 이종성 목적 서열의 게놈 내로의 삽입을 간섭할 수 있다. 따라서, 시스템은 비-표적 세포의 게놈을 편집하는 것보다 더 효율적으로 표적 세포의 게놈을 편집할 수 있을 것이며, 예를 들어 이종 목적 서열은 비-표적 세포의 게놈 내 보다 더 효율적으로 표적 세포의 게놈 내로의 삽입될 수 있거나, 삽입 또는 결실은 비-표적 세포보다 표적 세포에서 더 효율적으로 생산된다. 주형 RNA(또는 이를 인코딩하는 DNA)에서 마이크로RNA 결합 부위를 갖는 시스템은 유전자 변형 폴리펩티드를 인코딩하는 핵산과 조합되어 사용될 수도 있으며, 여기서 유전자 변형 폴리펩티드의 발현은, 예를 들어 본 명세서에 기재된 바와 같은 제2의 마이크로RNA 결합 부위에 의해 제어된다. 일부 구현예에서, 예를 들어 간 적응증의 경우, miRNA는 본 명세서에 참고로 포함된 WO2020014209의 표 4로부터 선택된다.

[0532] 일부 구현예에서, 주형 RNA는 마이크로RNA 서열, siRNA 서열, 가이드 RNA 서열, 또는 피위(piwi) RNA 서열을 포함한다.

[0533] *프로모터*

[0534] 일부 구현예에서, 하나 이상의 프로모터 또는 인핸서 요소는, 이종 목적 서열의 발현을 제어하는 유전자 변형 단백질을 인코딩하는 핵산 또는 주형 핵산에 작동가능하게 연결된다. 특정 구현예에서, 하나 이상의 프로모터 또는 인핸서 요소는 세포 유형 또는 조직 특이적 요소를 포함한다. 일부 구현예에서, 프로모터 또는 인핸서는 이종 목적 서열의 발현을 자연적으로 제어하는 프로모터 또는 인핸서와 동일하거나 이로부터 유래된다. 예를 들어, 오르니틴 트랜스카르보밀라제(transcarbomylase) 프로모터 및 인핸서는, 오르니틴 트랜스카르보밀라제 결핍을 교정하기 위해 본 발명에 의해 제공된 시스템 또는 방법에서 오르니틴 트랜스르카보밀라제 유전자의 발현을 제어하는 데 사용될 수 있다. 일부 구현예에서, 프로모터는 표 16 또는 17 또는 이의 기능성 단편 또는 변이체의 프로모터이다.

[0535] 상업적으로 구매가능한 예시적인 조직 특이적 프로모터는, 예를 들어 인터넷 주소(예를 들어, www.invivogen.com/tissue-specific-promoters)에서 찾을 수 있다. 일부 구현예에서, 프로모터는 천연 프로모터 또는 최소 프로모터로, 예를 들어 이는 소정의 유전자의 5' 영역으로부터 단일 단편으로 이루어진다. 일부 구현예에서, 천연 프로모터는 코어 프로모터 및 그의 천연 5' UTR을 포함한다. 일부 구현예에서, 5' UTR은 인트론을 포함한다. 다른 구현예에서, 이들은 복합 프로모터를 포함하며, 이는 상이한 기원의 프로모터 요소들을 조합하거나 원위 인핸서를 동일한 기원의 최소 프로모터와 조립함으로써 생성되었다.

[0536] 예시적인 세포 또는 조직 특이적 프로모터가 하기 표에 제공되며, 이를 인코딩하는 예시적인 핵산 서열은 당업계에 알려져 있고, 다양한 공급원, 예컨대 RefSeq를 포함하여 NCBI 데이터베이스, 및 진핵성 프로모터(//epd.epfl.ch//index.php) 데이터베이스를 이용하여 쉽게 접근될 수 있다.

[0537] **표 16. 예시적인 세포 또는 조직-특이적 프로모터**

표 16

[0538]

프로모터	**표적 세포**
B29 프로모터	B 세포
CD14 프로모터	단구 세포
CD43 프로모터	백혈구 및 혈소판
CD45 프로모터	조혈세포

CD68 프로모터	대식세포
데스민(Desmin) 프로모터	근육 세포
엘라스타제-1 프로모터	췌장 선방 세포
엔도글린(Endoglin) 프로모터	내피세포
피브로넥틴 프로모터	분화 세포, 치유 조직
Flt-1 프로모터	내피세포
GFAP 프로모터	성상세포
GPIIB 프로모터	거핵세포
ICAM-2 프로모터	내피세포
INF-베타 프로모터	조혈세포
Mb 프로모터	근육 세포
Nphs1 프로모터	족세포
OG-2 프로모터	조골세포, 치아모세포(Odonblast)
SP-B 프로모터	폐
Syn1 프로모터	뉴런
WASP 프로모터	조혈세포
SV40/bAlb 프로모터	간
SV40/bAlb 프로모터	간
SV40/Cd3 프로모터	백혈구 및 혈소판
SV40/CD45 프로모터	조혈세포
NSE/RU5' 프로모터	성숙 뉴런

[0539] **표 17. 추가의 예시적인 세포 또는 조직-특이적 프로모터**

표 17

[0540]

프로모터	**유전자 설명**	**유전자 특이성**
APOA2	아포지질단백질 A-II	간세포(간 전구세포 유래)
SERPINA1 (hAAT)	세르핀 펩티다제 억제제, 클레이드 A (알파-1 항프로테나제, 항트립신), 구성원 1 (알파 1 항-트립신으로도 명명됨)	간세포 (최종적인 내배엽 단계로부터)
CYP3A	시토크롬 P450, 패밀리 3, 서브패밀리 A, 폴리펩티드	성숙한 간세포
MIR122	마이크로RNA 122	간세포 (초기 배아 간 세포 유래) 및 내배엽
췌장 특이적 프로모터		
프로모터	**유전자 설명**	**유전자 특이성**
INS	인슐린	췌장 베타 세포 (최종 내배엽 단계로부터)
IRS2	인슐린 수용체 기질 2	췌장 베타 세포
Pdx1	췌장 및 십이지장 호메오박스 1	췌장 (최종 내배엽 단계로부터)
Alx3	아리스탈레스-유사(Aristaless-like) 호메오박스 3	췌장 베타 세포 (최종 내배엽 단계로부터)
Ppy	췌장 폴리펩티드	PP 췌장 세포 (감마 세포)
심장 특이적 프로모터		
프로모터	**유전자 설명**	**유전자 특이성**
Myh6(aMHC)	미오신, 중쇄 6, 심근, 알파	심장 근육세포의 후기 분화 마커(심방 특이성)
MYL2(MLC-2v)	미오신, 경쇄 2, 조절, 심장, 느림	심장 근육세포의 후기 분화 마커(심실 특이성)
ITNN13 (cTnI)	트로포닌 I 형 3(심장)	심근세포 (미성숙 상태로부터)

ITNN13 (cTnI)	트로포닌 I 형 3(심장)	심근세포 (미성숙 상태로부터)
NPPA(ANF)	나트륨뇨 펩티드 전구체 A(심방 나트륨뇨 인자로도 명명됨)	성인 세포에서 심방 특이성
Slc8a1(Ncx1)	용질 담체 패밀리 8(나트륨/칼슘 교환기), 구성원 1	조기 발생 단계로부터의 심근세포
CNS 특이적 프로모터		
SYN1(hSyn)	시냅신 I	뉴런
GFAP	아교세포섬유 산성 단백질	성상세포
INA	인터넥신 뉴런성 중간체 필라멘트 단백질, 알파(a-인터넥신)	신경전구세포(neuroprogenitor)
NES	네스틴	신경전구세포 및 외배엽
MOBP	미엘린-연관 희소돌기아교세포 염기성 단백질	희소돌기아교세포
MBP	미엘린 염기성 단백질	희소돌기아교세포
TH	티로신 하이드록실라제	도파민작용성 뉴런
FOXA2(HNF3 베타)	포크헤드(Forkhead) 박스 A2	도파민작용성 뉴런(내배엽의 마커로서도 사용됨)
피부 특이적 프로모터		
프로모터	**유전자 설명**	**유전자 특이성**
FLG	필라그린	과립층 유래의 각질형성세포
K14	케라틴 14	과립 및 기저 층으로부터의 각질형성세포
TGM3	트랜스글루타미나제 3	과립층 유래의 각질형성세포
면역 세포 특이적 프로모터		
프로모터	**유전자 설명**	**유전자 특이성**
ITGAM (CD11B)	인테그린, 알파 M(보체 구성요소 3 수용체 3 서브유닛)	단핵구, 대식세포, 과립구, 자연 살해 세포
요생식기 세포 특이적 프로모터		
프로모터	**유전자 설명**	**유전자 특이성**
Pbsn	프로바신(Probasin)	전립선 상피
Upk2	유로플라킨 2	방광
Sbp	스페르민 결합 단백질	전립선
Fer1l4	Fer-1-유사 4	방광
내피세포 특이적 프로모터		
프로모터	**유전자 설명**	**유전자 특이성**
ENG	엔도글린	내피세포
다능성 및 배아 세포 특이적 프로모터		
프로모터	**유전자 설명**	**유전자 특이성**
Oct4(POU5F1)	POU 클래스 5 호메오박스 1	다능성 세포 (생식 세포, ES 세포, iPS 세포)
NANOG	나노그(Nanog) 호메오박스	다능성 세포 (ES 세포, iPS 세포)
합성 Oct4	Oct-4 코어 인핸서 요소에 기반한 합성 프로모터	다능성 세포(ES 세포, iPS 세포)
T 단미증	단미증	중배엽
NES	네스틴	신경전구세포 및 외배엽
SOX17	SRY(성 결정 영역 Y)-박스 17	내배엽
FOXA2(HNFJ 베타)	포크헤드 박스 A2	내배엽(도파민작용성 뉴런의 마커로서도 사용됨)
MIR122	마이크로RNA 122	내배엽 및 간세포 (초기단계 배아 간 세포로부터~

[0541] 이용되는 숙주/벡터 시스템에 따라, 구성 및 유도성 프로모터, 전사 인핸서 요소, 전사 종결자 등을 포함하는 많은 임의의 적합한 전사 및 번역 제어 요소가 발현 벡터 상에서 사용될 수 있다(예를 들어, 문헌[Bitter et al. (1987) *Methods in Enzymology,* 153:516-544] 참조; 그 전체가 참조로서 여기 포함됨).

[0542] 일부 구현예에서, 유전자 변형 단백질 또는 주형 핵산을 인코딩하는 핵산은 제어 요소, 예를 들어 프로모터와 같은 전사 제어 요소에 작동가능하게 연결된다. 전사 제어 요소는 일부 구현예에서 진핵 세포, 예를 들어 포유

동물 세포; 또는 원핵 세포(예를 들어, 박테리아 또는 고세포) 중 어느 하나에서 기능성일 수 있다. 일부 구현예에서, 폴리펩티드를 인코딩하는 뉴클레오티드 서열은 다중 제어 요소들에 작동가능하게 연결되며, 예를 들어 이는 원핵 및 진핵 세포 모두에서 폴리펩티드를 인코딩하는 뉴클레오티드 서열의 발현을 가능하게 한다.

[0543] 예시적인 목적을 위해, 공간적으로 제한된 프로모터들의 예는 뉴런-특이적 프로모터, 지방세포-특이적 프로모터, 심근세포-특이적 프로모터, 평활근-특이적 프로모터, 광수용체-특이적 프로모터 등을 포함하지만 이에 제한되지 않는다. 뉴런-특이적인 공간적으로 제한된 프로모터는, 뉴런-특이적 에놀라제(NSE) 프로모터(예를 들어, EMBL HSENO2, X51956 참조); 방향족 아미노산 데카르복실라제(AADC) 프로모터, 신경미세섬유 프로모터(예를 들어, GenBank HUMNFL, L04147, 참조); 시냅신 프로모터(예를 들어, GenBank HUMSYNIB, M55301 참조); thy-1 프로모터(예를 들어, 문헌[Chen et al. (1987) Cell 51:7-19; 및 Llewellyn, et al. (2010) Nat. Med. 16(10):1161-1166] 참조); 세로토닌 수용체 프로모터(예를 들어, GenBank S62283, 참조); 티로신 하이드록실라제 프로모터(TH)(예를 들어, 문헌[Oh et al. (2009) Gene Ther 16:437; Sasaoka et al. (1992) Mol. Brain Res. 16:274; Boundy et al. (1998) J. Neurosci. 18:9989; 및 Kaneda et al. (1991) Neuron 6:583-594)] 참조); GnRH 프로모터(예를 들어, 문헌[Radovick et al. (1991) Proc. Natl. Acad. Sci. USA 88:3402-3406] 참조); L7 프로모터(예를 들어, 문헌[Oberdick et al. (1990) Science 248:223-226] 참조); DNMT 프로모터(예를 들어, 문헌[Bartge et al. (1988) Proc. Natl. Acad. Sci. USA 85:3648-3652] 참조); 엔케팔린(enkephalin) 프로모터(예를 들어, 문헌[Comb et al. (1988) EMBO J. 17:3793-3805] 참조); 미엘린 염기성 단백질(MBP) 프로모터; Ca2+-칼모듈린-의존성 단백질 키나제 II-알파(CamKII α) 프로모터(예를 들어, 문헌[Mayford et al. (1996) Proc. Natl. Acad. Sci. USA 93:13250; 및 Casanova et al. (2001) Genesis 31:37] 참조); CMV 인핸서/혈소판-유래 성장 인자- β 프로모터(예를 들어, 문헌[Liu et al. (2004) Gene Therapy 11:52-60] 참조); 등을 포함하지만 이에 제한되지 않는다.

[0544] 지방세포-특이적인 공간적으로 제한된 프로모터는 aP2 유전자 프로모터/인핸서, 예를 들어 인간 aP2 유전자의 -5.4 kb 내지 +21 bp의 영역(예를 들어, 문헌[Tozzo et al. (1997) Endocrinol. 138:1604; Ross et al. (1990) Proc. Natl. Acad. Sci. USA 87:9590; 및 Pavjani et al. (2005) Nat. Med. 11:797)] 참조); 포도당 수송체-4(GLUT4) 프로모터(예를 들어, 문헌[Knight et al. (2003) Proc. Natl. Acad. Sci. USA 100:14725)] 참조); 지방산 자리옮김효소(FAT/CD36) 프로모터(예를 들어, 문헌[Kuriki et al. (2002) Biol. Pharm. Bull. 25:1476; 및 Sato et al. (2002) J. Biol. Chem. 277:15703)] 참조); 스테아로일-CoA 불포화효소-1(SCD1) 프로모터(문헌[Tabor et al. (1999) J. Biol. Chem. 274:20603] 참조); 렙틴 프로모터(예를 들어, 문헌[Mason et al. (1998) Endocrinol. 139:1013; 및 Chen et al. (1999) Biochem. Biophys. Res. Comm. 262:187)] 참조); 아디포넥틴 프로모터(예를 들어, 문헌[Kita et al. (2005) Biochem. Biophys. Res. Comm. 331:484; 및 Chakrabarti (2010) Endocrinol. 151:2408] 참조); 아디프신 프로모터(예를 들어, 문헌[Platt et al. (1989) Proc. Natl. Acad. Sci. USA 86:7490] 참조); 레시스틴 프로모터(예를 들어, 문헌[Seo et al. (2003) Molec. Endocrinol. 17:1522] 참조); 등을 포함하지만 이에 제한되지 않는다.

[0545] 심근세포-특이적인 공간적으로 제한된 프로모터는 다음 유전자들로부터 유래된 제어 서열: 미오신 경쇄-2, α-미오신 중쇄, AE3, 심장 트로포닌 C, 심장 액틴 등을 포함하지만 이에 제한되지 않는다 (Franz et al. (1997) Cardiovasc. Res. 35:560-566; Robbins et al. (1995) Ann. N.Y. Acad. Sci. 752:492-505; Linn et al. (1995) Circ. Res. 76:584-591; Parmacek et al. (1994) Mol. Cell. Biol. 14:1870-1885; Hunter et al. (1993) Hypertension 22:608-617; 및 Sartorelli et al. (1992) Proc. Natl. Acad. Sci. USA 89:4047-4051).

[0546] 평활근-특이적인 공간적으로 제한된 프로모터는, 이에 제한되지는 않지만, SM22α 프로모터(예를 들어, 문헌[Akyurek et al. (2000) Mol. Med. 6:983]; 및 미국 특허 7,169,874 참조); 스무드린(smoothelin) 프로모터(예를 들어, WO 2001/018048 참조); α-평활근 액틴 프로모터; 등을 포함한다. 예를 들어, 그 안에 2개의 CArG 요소가 놓여있는, SM22α 프로모터의 0.4 kb 영역은 혈관 평활근 세포-특이적 발현을 매개하는 것으로 나타났다(예를 들어, 문헌[Kim, et al. (1997) Mol. Cell. Biol. 17, 2266-2278; Li, et al., (1996) J. Cell Biol. 132, 849-859; 및 Moessler, et al. (1996) Development 122, 2415-2425] 참조).

[0547] 광수용체-특이적인 공간적으로 제한된 프로모터는 로돕신 프로모터; 로돕신 키나제 프로모터(Young et al. (2003) Ophthalmol. Vis. Sci. 44:4076); 베타 포스포디에스테라제 유전자 프로모터(Nicoud et al. (2007) J. Gene Med. 9:1015); 망막세포 변성증 유전자 프로모터(Nicoud et al. (2007) 상동); 광수용체간(interphotoreceptor) 레티노이드-결합 단백질(IRBP) 유전자 인핸서(Nicoud et al. (2007) 상동); IRBP 유전자 프로모터(Yokoyama et al. (1992) Exp Eye Res. 55:225) 등을 포함하지만 이에 제한되지 않는다.

[0548] 일부 구현예에서, 유전자 변형 시스템, 예를 들어 유전자 변형 폴리펩티드를 인코딩하는 DNA, 주형 RNA를 인코딩하는 DNA, 또는 이종 목적 서열을 인코딩하는 DNA 또는 RNA는 하나 이상의 요소가 조직-특이적 프로모터, 예를 들어 T-세포에서 활성인 프로모터에 작동가능하게 연결되도록 설계된다. 추가 구현예에서, T-세포 활성 프로모터는 다른 세포 유형, 예를 들어 B-세포, NK 세포에서 비활성이다. 일부 구현예에서, T-세포 활성 프로모터는 T-세포 수용체의 구성요소, 예를 들어 TRAC, TRBC, TRGC, TRDC를 인코딩하는 유전자에 대한 프로모터로부터 유래된다. 일부 구현예에서, T-세포 활성 프로모터는 T-세포-특이적 분화 단백질 클러스터의 구성요소, 예를 들어 CD3, 예를 들어 CD3D, CD3E, CD3G, CD3Z를 인코딩하는 유전자에 대한 프로모터로부터 유래된다. 일부 구현예에서, 유전자 변형 시스템의 T-세포-특이적 프로모터는 세포 유형에 걸쳐 공적으로 입수가능한 유전자 발현 데이터를 비교하고, T-세포에서 향상된 발현을 갖는 유전자로부터 프로모터를 선택함으로써 발견된다. 일부 구현예에서, 프로모터는 원하는 발현 폭에 따라, 예를 들어 T-세포에서만 활성인 프로모터, NK 세포에서만 활성인 프로모터, T-세포 및 NK 세포 둘 다에서 활성인 프로모터를 선택할 수 있다.

[0549] 당업계에 알려진 세포-특이적 프로모터는 유전자 변형 단백질의 발현을, 예를 들어 본 명세서에 기재된 바와 같이, 지시하는 데 사용될 수 있다. 비제한적인 예시적인 포유동물 세포-특이적 프로모터는 세포-특이적 방식으로 Cre 재조합효소를 발현하는 마우스에서 특징화 및 사용된다. 특정 비제한적인 예시적인 포유동물 세포-특이적 프로모터는, 본 명세서에 참고로 포함된 US9845481의 표 1에 열거되었다.

[0550] 일부 구현예에서, 본 명세서에 기재된 벡터는 발현 카세트를 포함한다. 통상적으로, 발현 카세트는 프로모터 서열에 작동가능하게 연결된 본 발명의 핵산 분자를 포함한다. 예를 들어, 그 코딩 서열의 발현에 영향을 미치는 경우, 프로모터는 코딩 서열과 작동가능하게 연결되었다(예를 들어, 코딩 서열이 프로모터의 전사 제어 하에 존재한다). 서열 인코딩은 센스 또는 안티센스 배향으로 조절 서열에 작동가능하게 연결될 수 있다. 특정 구현예에서, 프로모터는 이종성 프로모터이다. 특정 구현예에서, 발현 카세트는 추가의 요소, 예를 들어 인트론, 인핸서, 폴리아데닐화 부위, 우드척 반응 요소(WRE) 및/또는 인코딩 서열의 발현 수준에 영향을 미치는 것으로 알려진 기타 요소를 포함할 수 있다. 프로모터는 코딩 서열 또는 기능성 RNA의 발현을 통상적으로 제어한다. 특정 구현예에서, 프로모터 서열은 근위 및 더 원위의 상류 요소를 포함하며, 인핸서 요소를 추가로 포함할 수 있다. 인핸서는 통상적으로 프로모터 활성을 자극할 수 있고, 프로모터의 선천적인 요소 또는 프로모터의 수준 또는 조직-특이성을 향상하기 위해 삽입된 이종성 요소일 수 있다. 특정 구현예에서, 프로모터는 그 전체가 자연의 유전자로부터 유래된다. 특정 구현예에서, 프로모터는 상이한 자연 발생 프로모터로부터 유래된 상이한 요소로 구성된다. 특정 구현예에서, 프로모터는 합성 뉴클레오티드 서열을 포함한다. 상이한 프로모터는 상이한 조직 또는 세포 유형에서 또는 발달의 상이한 단계에서 또는 상이한 환경 조건이나 약물 또는 전사 보조-인자의 존재 또는 부재에 반응하여 유전자의 발현을 지시할 것임은 당업자는 이해할 것이다. 편재성, 세포 유형-특이적, 조직-특이적, 발달 단계-특이적, 및 조건적 프로모터, 예를 들어 약물-반응성 프로모터(예를 들어, 테트라사이클린-반응성 프로모터)는 당업자에게 잘 알려져 있다. 프로모터의 예는 포스포글리세레이트 키나제(PKG) 프로모터, CAG(CMV 인핸서, 닭 베타 액틴 프로모터(CBA) 및 토끼 베타 글로빈 인트론의 복합체), NSE(뉴런 특이적 엔올라제), 시냅신 또는 NeuN 프로모터, SV40 초기 프로모터, 마우스 유방 종양 바이러스 LTR 프로모터; 아데노바이러스 주요 후기 프로모터(Ad MLP); 단순포진 바이러스(HSV) 프로모터, 거대세포바이러스(CMV) 프로모터, 예컨대 CMV 급 초기 프로모터 영역(CMVIE), SFFV 프로모터, 라우스 육종 바이러스(RSV) 프로모터, 합성 프로모터, 하이브리드 프로모터 등을 포함하지만 이에 제한되지 않는다. 기타 프로모터는 인간 기원 또는 쥐를 포함한 기타 종일 수 있다. 통상의 프로모터는 예를 들어 인간 거대세포바이러스(CMV) 급 초기 유전자 프로모터, SV40 초기 프로모터, 라우스 육종 바이러스 긴 말단 반복부, [베타]- 액틴, 쥐 인슐린 프로모터, 포스포글리세레이트 키나제 인간 알파-1 항트립신(hAAT) 프로모터, 트랜스타이레틴(transthyretin) 프로모터, TBG 프로모터 및 기타 간-특이적 프로모터, 데스민 프로모터 및 유사한 근육-특이적 프로모터, EF1 -알파 프로모터, 다수-조직 특이성을 갖는 하이브리드 프로모터, 시냅신과 같은 뉴런에 특이적인 프로모터 및 글리세르알데히드-3-포스페이트 탈수소화효소 프로모터를 포함하고, 이들 모두는 당업자에게 잘 알려지고 쉽게 입수가능하며, 관심 서열의 코딩 서열의 높은 수준의 발현을 수득하는 데 사용될 수 있다. 추가적으로, 쥐 메탈로티오네인(metallothionein) 유전자와 같은 비-바이러스 유전자로부터 유래된 서열은 본 발명에서의 이용을 또한 찾을 것이다. 그러한 프로모터 서열은 예를 들어 스트라타진(Stratagene)(San Diego, CA 소재)으로부터 상업적으로 구매가능하다. 추가의 예시적인 프로모터 서열은 예를 들어 WO2018213786A1(그 전체가 본 명세서에 참고로 포함됨)에 기재되어 있다.

[0551] 일부 구현예에서, 아포지질단백질 E 인핸서(ApoE) 또는 이의 기능성 단편은 예를 들어 간에서의 발현을 구동하는 데 사용된다. 일부 구현예에서, ApoE 인핸서의 2개의 복제물 또는 이의 기능성 단편이 사용된다. 일부 구현

예에서, ApoE 인핸서 또는 이의 기능성 단편은 프로모터, 예를 들어 인간 알파-1 항트립신(hAAT) 프로모터와 조합되어 사용된다.

[0552] 일부 구현예에서, 조절 서열은 조직-특이적 유전자 발현능을 부여한다. 일부 경우들에서, 조직-특이적 조절 서열은 조직 특이적 방식으로 전사를 유도하는 조직-특이적 전사 인자에 결합한다. 다양한 조직-특이적 조절 서열(예를 들어, 프로모터, 인핸서 등)은 당업계에 알려져 있다. 예시적인 조직-특이적 조절 서열은 다음의 조직-특이적 프로모터: 간-특이적 티록신 결합 글로불린(TBG) 프로모터, 인슐린 프로모터, 글루카곤 프로모터, 소마토스타틴 프로모터, 췌장 폴리펩티드(PPY) 프로모터, 시냅신-1(Syn) 프로모터, 크레아틴 키나제(MCK) 프로모터, 포유동물 데스민(DES) 프로모터, α-미오신 중쇄(a-MHC) 프로모터, 또는 심장 트로포닌 T(cTnT) 프로모터를 포함하지만 이에 제한되지 않는다. 기타 예시적인 프로모터는 베타-액틴 프로모터, B형 간염 바이러스 코어 프로모터(Sandig et al., Gene Ther., 3:1002-9 (1996)); 알파-페토단백질(AFP) 프로모터(Arbuthnot et al., Hum. Gene Ther., 7:1503-14 (1996)), 뼈 오스테오칼신 프로모터(Stein et al., Mol. Biol. Rep., 24:185-96 (1997)); 뼈 시알로단백질 프로모터(Chen et al., J. Bone Miner. Res., 11:654-64 (1996)), CD2 프로모터(Hansal et al., J. Immunol., 161:1063-8 (1998)); 면역글로불린 중쇄 프로모터; T 세포 수용체 α-사슬 프로모터, 뉴런성 예컨대 뉴런-특이적 엔올라제(NSE) 프로모터(Andersen et al., Cell. Mol. Neurobiol., 13:503-15 (1993)), 신경미세섬유 경쇄 유전자 프로모터(Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991)), 및 뉴런-특이적 vgf 유전자 프로모터(Piccioli et al., Neuron, 15:373-84 (1995)), 및 기타를 포함한다. 추가의 예시적인 프로모터 서열은 예를 들어 미국 특허 10300146(그 전체가 본 명세서에 참고로 포함됨)에 기재되어 있다. 일부 구현예에서, 조직-특이적 조절 요소, 예를 들어 조직-특이적 프로모터는, 예를 들어 RNA-seq 또는 단백질 발현 데이터, 또는 이의 조합에 의해 측정됨에 따라, 소정의 조직에서 고도로 발현되는 유전자에 작동가능하게 연결되는 것으로 알려진 것으로부터 선택된다. 발현에 의해 조직 특이성을 분석하기 위한 방법은, 그 전체가 본 명세서에 참고로 포함된, 문헌[Fagerberg et al. *Mol Cell Proteomics* 13(2):397-406 (2014)]에 교시되어 있다.

[0553] 일부 구현예에서, 본 명세서에 기재된 벡터는 멀티시스트로닉(multicistronic) 발현 작제물이다. 멀티시스트로닉 발현 작제물은, 예를 들어 제1 프로모터와 제1 인코딩 핵산 서열을 포함하는 제1 발현 카세트, 및 예를 들어 제2 프로모터와 제2 인코딩 핵산 서열을 포함하는 제2 발현 카세트를 보유하는 작제물을 포함한다. 그러한 멀티시스트로닉 발현 작제물은, 일부 경우에서 헤어핀 RNA와 같이 비번역된 유전자 생성물을, 폴리펩티드, 예를 들어 유전자 변형 폴리펩티드 및 유전자 변형 주형과 함께 전달하는데 특히 유용하다. 일부 구현예에서, 멀티시스트로닉 발현 작제물은, 예를 들어 프로모터 간섭 또는 인접한 비혼화성 핵산 요소 존재로 인해, 하나 이상의 포함된 이식 유전자의 감소된 발현 수준을 나타낼 수 있다. 멀티시스트로닉 발현 작제물이 바이러스 벡터의 일부인 경우, 자기-상보적 핵산 서열의 존재는 일부 경우에서 바이러스 복제 또는 패키징에 필요한 구조의 형성을 간섭할 수 있다.

[0554] 일부 구현예에서, 서열은 헤어핀을 갖는 RNA를 인코딩한다. 일부 구현예에서, 헤어핀 RNA는 가이드 RNA, 주형 RNA, shRNA, 또는 마이크로RNA이다. 일부 구현예에서, 제1 프로모터는 RNA 중합효소 I 프로모터이다. 일부 구현예에서, 제1 프로모터는 RNA 중합효소 II 프로모터이다. 일부 구현예에서, 제2 프로모터는 RNA 중합효소 III 프로모터이다. 일부 구현예에서, 제2 프로모터는 U6 또는 H1 프로모터이다.

[0555] 이론에 구속되고자 함이 없이, 멀티시스트로닉 발현 작제물은 단지 하나의 시스트론을 함유하는 발현 시스템과 비교하여 최적의 발현 수준을 달성할 수 없다. 둘 이상의 프로모터 요소를 포함하는 멀티시스트로닉 발현 작제물로 달성된 보다 적은 발현 수준에 대해 제안된 이유들 중 하나는 프로모터 간섭의 현상이다(예를 들어, 문헌[Curtin J A, Dane A P, Swanson A, Alexander I E, Ginn S L. *Bidirectional promoter interference between two widely used internal heterologous promoters in a late-generation lentiviral construct*. Gene Ther. 2008 March; 15(5):384-90; 및 Martin-Duque P, Jezzard S, Kaftansis L, Vassaux G. *Direct comparison of the insulating properties of two genetic elements in an adenoviral vector containing two different expression cassettes*. Hum Gene Ther. 2004 October; 15(10):995-1002] 참조; 프로모터 간섭 현상의 개시를 위해, 두 참고 모두 본 명세서에 참고로 포함됨). 일부 구현예에서, 프로모터 간섭의 문제는, 예를 들어 내부 리보솜 도입 부위에 의해 분리된 다중 인코딩 핵산 서열의 전사를 구동하는 단지 하나의 프로모터를 포함하는 멀티시스트로닉 발현 작제물을 생산함으로써, 또는 그 자신의 프로모터를 포함하는 시스트론을 전사 절연체 요소로 구분함으로써 극복될 수 있다. 일부 구현예에서, 다중 시스트론의 단일-프로모터 구동 발현은 시스트론의 균일하지 않은 발현 수준을 생성할 수 있다. 일부 구현예에서, 프로모터는 효율적으로 분리될 수 없고, 분리 요소는 일부 유전자 전이(gene transfer) 벡터, 예를 들어 일부 레트로바이러스 벡터들과 양립성이 아닐 수 있다.

[0556] *마이크로RNA*

[0557] 마이크로RNA(miRNA) 및 기타 간섭 핵산은 표적 RNA 전사체 절단/분해 또는 표적 메신저 RNA(mRNA)의 번역 저해를 통해 유전자 발현을 일반적으로 조절한다. miRNA는 일부 경우에서, 통상적으로 최종 19 내지 25의 비번역된 RNA 생성물로서 자연적으로 발현될 수 있다. miRNA는 표적 mRNA의 3' 비번역된 영역(UTR)과 서열-특이적 상호작용을 통한 그의 활성을 일반적으로 나타낸다. 이들 내인성적으로 발현된 miRNA는 miRNA 듀플렉스로, 나아가 성숙한 단일 가닥 miRNA 분자로 순차적으로 가공되는 헤어핀 전구체를 형성할 수 있다. 이러한 성숙한 miRNA는 일반적으로 다중단백질 복합체, miRISC를 가이드하며, 이는 성숙한 miRNA에 대한 그들의 상보성에 근거하여 표적 mRNA의 표적 3' UTR 영역을 확인한다. 유용한 이식유전자 생성물은 연결된 폴리펩티드의 발현을 조절하는, 예를 들어 miRNA 또는 miRNA 결합 부위를 포함할 수 있다. miRNA 유전자의 비제한적인 열거; 이들 유전자 및 이들의 동족체는 예를 들어 참조로서 본 명세서에 포함된 US10300146, 22:25-25:48에 열거된 바와 같은 방법에서, 이식유전자로서 또는 소간섭 핵산에 대한 표적으로서 유용하다(예를 들어, miRNA 스폰지, 안티센스 올리고뉴클레오티드). 일부 구현예에서, 하나 이상의 상기 miRNA에 대한 하나 이상의 결합 부위는 이식유전자, 예를 들어 rAAV 벡터에 의해 전달된 이식유전자 내에 포함되어, 예를 들어 이식유전자를 보유하는 동물의 하나 이상의 조직에서 그 이식유전자의 발현을 억제한다. 일부 구현예에서, 결합 부위는 조직 특이적 방식으로 이식유전자의 발현을 제어하도록 선택될 수 있다. 예를 들어, 간-특이적 miR-122에 대한 결합 부위는 간에서 그 이식유전자의 발현을 저해하기 위해 이식유전자 내로 포함될 수 있다. 추가의 예시적인 miRNA 서열은, 예를 들어 미국 특허 10,300,146(그 전체가 본 명세서에 참고로 포함됨)에 기재되어 있다.

[0558] miR 억제제 또는 miRNA 억제제는, 일반적으로 miRNA 발현 및/또는 가공을 차단하는 작용제이다. 그러한 작용물의 예는 마이크로RNA 안타고니스트, 마이크로RNA 특이적 안티센스, 마이크로RNA 스폰지 및 마이크로RNA 올리고뉴클레오티드(이중 가닥, 헤어핀, 짧은 올리고뉴클레오티드)를 포함하지만 이에 제한되지 않으며 이는 드로샤(Drosha) 복합체와의 miRNA 상호작용을 억제한다. 마이크로RNA 억제제, 예를 들어 miRNA 스폰지는 이식유전자로부터의 세포 내에서 발현될 수 있다(예를 들어, 문헌[Ebert, M. S. Nature Methods, Epub Aug. 12, 2007]에 기재된 바와 같음; 이는 그 전체가 본 명세서에 참고로 포함됨). 일부 구현예에서, 마이크로RNA 스폰지, 또는 기타 miR 억제제가 AAV와 사용된다. 마이크로RNA 스폰지는 상보적인 칠량체 시드(seed) 서열을 통해 miRNA를 전반적으로 특이하게 억제한다. 일부 구현예에서, miRNA의 전체 과는 단일 스폰지 서열을 사용하여 침묵화될 수 있다. miRNA 기능을 세포에서 침묵화하기 위한 기타 방법(miRNA 표적의 탈억제)는 당업자에게 명백할 것이다.

[0559] 일부 구현예에서, 유전자 변형 시스템, 주형 RNA, 또는 본 명세서에 기재된 폴리펩티드는 표적 조직, 예를 들어 제1 조직에 투여되거나 표적 조직에서 활성이다(예를 들어, 표적 조직에서 더 활성임). 일부 구현예에서, 유전자 변형 시스템, 주형 RNA, 또는 폴리펩티드는 비-표적 조직에 투여되지 않거나 비-표적 조직에서 덜 활성이다(예를 들어, 비-표적 조직에서 활성이 아님). 일부 구현예에서, 본 명세서에 기재된 유전자 변형 시스템, 주형 RNA, 또는 폴리펩티드는 표적 조직, 예를 들어 제1 조직 내에서 DNA를 변형시키는 데(예를 들어, 비-표적 조직에서 DNA를 변형시키지 않는 데) 유용하다.

[0560] 일부 구현예에서, 유전자 변형 시스템은 (a) 본 명세서에 기재된 폴리펩티드 또는 이를 인코딩하는 핵산, (b) 본 명세서에 기재된 주형 핵산(예를 들어, 주형 RNA), 및 (c) 표적 조직에 특이적인 하나 이상의 제1 조직-특이적 발현-제어 서열을 포함하고, 여기서 표적 조직에 특이적인 하나 이상의 제1 조직-특이적 발현-제어 서열은 (a), (b), 또는 (a) 및 (b)과 작동적으로 연관되며, 여기서 (a)에 회합된 경우, (a)는 폴리펩티드를 인코딩하는 핵산을 포함한다.

[0561] 일부 구현예에서, (b)에서 핵산은 RNA를 포함한다.

[0562] 일부 구현예에서, (b)에서 핵산은 DNA를 포함한다.

[0563] 일부 구현예에서, (b)에서 핵산은: (i) 단일 가닥이거나, 단일 가닥 절편을 포함하며, 예를 들어 단일 가닥 DNA이거나, 단일 가닥 절편 및 하나 이상의 이중 가닥 절편을 포함하고; (ii) 역 말단 반복부를 갖거나; 또는 (iii) (i)과 (ii) 둘 다이다.

[0564] 일부 구현예에서, (b)에서 핵산은 이중 가닥이거나 또는 이중 가닥 절편을 포함한다.

[0565] 일부 구현예에서, (a)는 폴리펩티드를 인코딩하는 핵산을 포함한다.

[0566] 일부 구현예에서, (a)에서 핵산은 RNA를 포함한다.

[0567] 일부 구현예에서, (a)에서 핵산은 DNA를 포함한다.

[0568] 일부 구현예에서, (a)에서 핵산은: (i) 단일 가닥이거나, 단일 가닥 절편을 포함하며, 예를 들어 단일 가닥 DNA이거나, 단일 가닥 절편 및 하나 이상의 이중 가닥 절편을 포함; (ii) 역 말단 반복부를 가짐; 또는 (iii) (i)과 (ii) 둘 다이다.

[0569] 일부 구현예에서, (a)에서 핵산은 이중 가닥이거나 또는 이중 가닥 절편을 포함한다.

[0570] 일부 구현예에서, (a), (b), 또는 (a) 및 (b)에서 핵산은 선형이다.

[0571] 일부 구현예에서, (a), (b), 또는 (a) 및 (b)에서 핵산은 환형, 예를 들어, 플라스미드 또는 미니서클이다.

[0572] 일부 구현예에서, 이종성 목적 서열은 제1 프로모터와 작동적으로 연관된다.

[0573] 일부 구현예에서, 하나 이상의 제1 조직-특이적 발현-제어 서열은 조직 특이적 프로모터를 포함한다.

[0574] 일부 구현예에서, 조직-특이적 프로모터는: (i) 이종성 목적서열, (ii) 레트로바이러스 RT를 인코딩하는 핵산, 또는 (iii) (i) 및 (ii)와 작동적으로 연관된 제1 프로모터를 포함한다.

[0575] 일부 구현예에서, 하나 이상의 제1 조직-특이적 발현-제어 서열은: (i) 이종성 목적서열, (ii) 레트로바이러스 RT 도메인을 인코딩하는 핵산, 또는 (iii) (i) 및 (ii)와 작동적으로 연관된 조직-특이적 마이크로RNA 인식 서열을 포함한다.

[0576] 일부 구현예에서, 시스템은 조직-특이적 프로모터를 포함하고, 시스템은 하나 이상의 조직-특이적 마이크로RNA 인식 서열을 추가로 포함하고, 여기서: (i) 조직-특이적 프로모터는: (I) 이종성 목적 서열, (II) 레트로바이러스 RT 도메인을 인코딩하는 핵산, 또는 (III) (I) 및 (II)와 작동적으로 연관되며; 및/또는 (ii) 하나 이상의 조직-특이적 마이크로RNA 인식 서열은: (I) 이종성 목적 서열, (II) 레트로바이러스 RT를 인코딩하는 핵산, 또는 (III) (I) 및 (II)와 작동적으로 연관된다.

[0577] 일부 구현예에서, (a)는 폴리펩티드를 인코딩하는 핵산을 포함하며, 핵산은 폴리펩티드를 인코딩하는 핵산과 작동적으로 연관된 프로모터를 포함한다.

[0578] 일부 구현예에서, 폴리펩티드를 인코딩하는 핵산은, 폴리펩티드 코딩 서열과 작동적으로 연관된 표적 조직에 대해 특이적인 하나 이상의 제2의 조직-특이적 발현-제어 서열을 포함한다.

[0579] 일부 구현예에서, 하나 이상의 제2 조직-특이적 발현-제어 서열은 조직 특이적 프로모터를 포함한다.

[0580] 일부 구현예에서, 조직-특이적 프로모터는 폴리펩티드를 인코딩하는 핵산과 작동적으로 연관된 프로모터이다.

[0581] 일부 구현예에서, 하나 이상의 제2 조직-특이적 발현-제어 서열은 조직-특이적 마이크로RNA 인식 서열을 포함한다.

[0582] 일부 구현예에서, 폴리펩티드를 인코딩하는 핵산과 작동적으로 연관된 프로모터는 조직-특이적 프로모터로, 시스템은 하나 이상의 조직-특이적 마이크로RNA 인식 서열을 추가로 포함한다.

[0583] 일부 구현예에서, 본 발명에 의해 제공된 시스템의 핵산 구성요소인, (예를 들어, 폴리펩티드를 인코딩하거나, 이종성 목적 서열을 포함하는) 서열은 단백질 발현 수준을 변형시키는 비번역 영역에 의해 측접된다. 다양한 5' 및 3' UTR은 단백질 발현에 영향을 미칠 수 있다. 예를 들어, 일부 구현예에서, 코딩 서열은 RNA 안정성 또는 단백질 번역을 변형시키는 5' UTR이 선행할 수 있다. 일부 구현예에서, 서열은 RNA 안정성 또는 단백질 번역을 변형시키는 3' UTR이 뒤따를 수 있다. 일부 구현예에서, 서열은 5' UTR이 선행하고 3' UTR이 뒤따를 수 있으며 이는 RNA 안정성 또는 번역을 변형시킨다. 일부 구현예에서, 보체 인자 3의 5' 및/또는 3' UTR(C3)(CACTCCTCCCCATCCTCTCCCTCTGTCCCTCTGTCCCTCTGACCCTGCACTGTCCCAGCACC) 또는 오로소뮤코이드(orosomucoid) 1(ORM1)(CAGGACACAGCCTTGGATCAGGACAGAGACTTGGGGGCCATCCTGCCCCTCCAACCCGACATGTGTACCTCAGCTTTTTCCCTCACTTGCATCAATAAAGCTTCTGTGTTTGGAACAGCTAA)로부터 선택될 수 있다(Asrani et al. RNA Biology 2018). 특정 구현예에서, 5' UTR은 C3으로부터의 5'UTR이고, 3' UTR은 ORM1로부터의 3' UTR이다. 특정 구현예에서, 단백질 발현을 위한 5' UTR 및 3' UTR, 예를 들어 유전자 변형 폴리펩티드 또는 이종성 목적 서열을 위한 mRNA(또는 그 RNA를 인코딩하는 DNA)는 최적화된 발현 서열을 포함한다. 일부 구현예에서, 5' UTR은 GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC 및/또는 UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCAGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA)을 포함하는 3' UTR을 포함하며, 이는 예를 들어, 문헌[Richner et al. *Cell* 168(6): P1114-1125

(2017)]에 기재된 바와 같으며, 이의 서열은 본 명세서에 참고로 포함된다.

[0584] 일부 구현예에서, 5' 및/또는 3' UTR은 단백질 발현을 향상시키기 위해 선택될 수 있다. 일부 구현예에서, 5' 및/또는 3' UTR은 과잉생산 저해가 최소화되도록 단백질 발현을 변형시키기 위해 선택될 수 있다. 일부 구현예에서, UTR은 코딩 서열 주위, 예를 들어 코딩 서열 외부에 존재하며, 다른 구현예는 코딩 서열에 인접하여 존재한다. 일부 구현예에서, 추가의 조절 요소들(예를 들어, miRNA 결합 부위, cis-조절 부위)이 UTR에 포함된다.

[0585] 일부 구현예에서, 유전자 변형 시스템의 오픈 판독 프레임, 예를 들어 유전자 변형 폴리펩티드를 인코딩하는 mRNA(또는 mRNA를 인코딩하는 DNA)의 ORF 또는 이종성 목적 서열의 mRNA(또는 mRNA를 인코딩하는 DNA)의 하나 이상의 ORF는, 그의 발현을 향상시키는 5' 및/또는 3' 비번역 영역(UTR)에 의해 측접된다. 일부 구현예에서, 시스템의 mRNA 구성요소(또는 DNA 구성요소로부터 생산된 전사체)의 5' UTR은 서열 5'-GGGAAAUAAGAGAGAAAAGAAGAGUAAGAAGAAAUAUAAGAGCCACC-3'을 포함한다. 일부 구현예에서, 시스템의 mRNA 구성요소(또는 DNA 구성요소로부터 생산된 전사체)의 3' UTR은 서열 5'-UGAUAAUAGGCUGGAGCCUCGGUGGCCAUGCUUCUUGCCCCUUGGGCCUCCCCCCAGCCCCUCCUCCCCUUCCUGCACCCGUACCCCCGUGGUCUUUGAAUAAAGUCUGA-3'을 포함한다. 5' UTR 및 3' UTR의 이러한 조합은 문헌[Richner et al. *Cell* 168(6): P1114-1125 (2017)]에 의해, 작동가능하게 연결된 ORF의 바람직한 발현을 생성하는 것으로 나타났으며, 여기 포함된 교시 및 서열은 여기에 참조로서 포함된다. 일부 구현예에서, 본 명세서에 기재된 시스템은 전사체를 인코딩하는 DNA를 포함하고, 여기서 DNA는, 상응하는 5' UTR 및 3' UTR 서열을 포함하고, 상기 열거된 서열에서 U는 T로 치환된다. 일부 구현예에서, 시스템의 RNA 구성요소를 생산하기 위해 사용된 DNA 벡터는 시험관 내 전사를 개시하기 위해 5' UTR의 프로모터 상류, 예를 들어 T7, T3, 또는 SP6 프로모터를 더 포함한다. 상기 5' UTR은 GGG로 시작하며, 이는 T7 RNA 중합효소를 이용하여 전사를 최적화하기 위한 적합한 출발이다. 전사 수준을 조절하고, 대안적인 5' UTR을 적합화하기 위해 전사 개시 부위를 변경하기 위해, 문헌[Davidson et al. *Pac Symp Biocomput* 433-443 (2010)]의 교시는 T7 프로모터 변이체 및 그의 발견 방법을 설명하며, 이는 이들 두 특성 모두를 충족한다.

[0586] *바이러스 벡터 및 이의 구성요소*

[0587] 바이러스는, 본 명세서에 기재된 바와 같은 관련 효소 또는 도메인, 예를 들어 본 명세서에 사용된 중합효소 또는 중합효소 기능, 예를 들어 DNA-의존성 DNA 중합효소, RNA-의존성 RNA 중합효소, RNA-의존성 DNA 중합효소, DNA-의존성 RNA 중합효소, 역전사효소의 공급원에 더하여, 본 명세서에 기재된 시스템을 위한 전달 비히클의 유용한 공급원이다. 일부 효소들, 예를 들어 역전사효소는 다중 활성을 가질 수 있으며, 예를 들어 RNA-의존성 DNA 중합 및 DNA-의존성 DNA 중합 둘 모두, 예를 들어 제1 및 제2 가닥 합성이 가능하다. 일부 구현예에서, 유전자 변형 전달 시스템 또는 이의 구성요소들의 공급원으로 사용되는 바이러스는 문헌[Baltimore *Bacteriol Rev* 35(3):235-241 (1971)]에 의해 기재된 바와 같은 그룹으로부터 선택될 수 있다.

[0588] 일부 구현예에서, 바이러스는 그룹 I 바이러스로부터 선택되고, 예를 들어 DNA 바이러스이며, dsDNA를 비리온 내로 패키징한다. 일부 구현예에서, 그룹 I 바이러스는 예를 들어, 아데노바이러스, 헤르페스바이러스, 폭스바이러스(Poxviruse)로부터 선택된다.

[0589] 일부 구현예에서, 바이러스는 그룹 II 바이러스로부터 선택되고, 예를 들어 DNA 바이러스이며 ssDNA를 비리온 내로 패키징한다. 일부 구현예에서, 그룹 II 바이러스는 예를 들어 파보바이러스(Parvovirus)로부터 선택된다. 일부 구현예에서, 파보바이러스는 디펜도파보바이러스(dependoparvovirus), 예를 들어 아데노-연관 바이러스(AAV)이다.

[0590] 일부 구현예에서, 바이러스는 그룹 III 바이러스로부터 선택되고, 예를 들어 RNA 바이러스이며 ssDNA를 비리온 내로 패키징한다. 일부 구현예에서, 그룹 III 바이러스는, 예를 들어 레오바이러스(Reoviruse)로부터 선택된다. 일부 구현예에서, 그러한 비리온에 함유된 dsRNA 중 하나 또는 두 가닥 모두는 숙주 세포 내로의 형질도입시 mRNA로서 직접 제공될 수 있는 코딩 분자이며, 예를 들어 숙주 세포 내로 형질도입시 임의의 개입하는 핵산 복제 또는 중합 단계를 필요로 함이 없이 단백질로 직접 번역될 수 있다.

[0591] 일부 구현예에서, 바이러스는 그룹 IV 바이러스로부터 선택되고, 예를 들어 RNA 바이러스이며, ssRNA(+)를 비리온 내로 패키징한다. 일부 구현예에서, 그룹 IV 바이러스는, 예를 들어 코로나바이러스, 피코르나바이러스(Picornavirus), 토가바이러스(Togavirus)로부터 선택된다. 일부 구현예에서, 그러한 비리온 내에 함유된 ssRNA(+)는 숙주 세포 내로의 형질도입시 mRNA로서 직접 제공될 수 있는 코딩 분자이며, 예를 들어 숙주 세포 내로 형질도입시 임의의 개입하는 핵산 복제 또는 중합 단계를 필요로 함이 없이 단백질로 직접 번역될 수

있다.

[0592] 일부 구현예에서, 바이러스는 그룹 V 바이러스로부터 선택되고, 예를 들어 RNA 바이러스이며, ssRNA(-)를 비리온 내로 패키징한다. 일부 구현예에서, 그룹 V 바이러스는, 예를 들어 오르토믹소바이러스(Orthomyxoviruses), 랍도바이러스(Rhabdoviruses)로부터 선택된다. 일부 구현예에서, ssRNA(-) 게놈을 갖는 RNA 바이러스는 바이러스 게놈과 함께 숙주 세포로 형질도입되는 비리온 내부의 효소, 예를 들어 RNA-의존성 RNA 중합효소 또한 운반하며, 이는 ssRNA(-)를 숙주에 의해 직접 번역될 수 있는 ssRNA(+)로 카피할 수 있다.

[0593] 일부 구현예에서, 바이러스는 그룹 VI 바이러스로부터 선택되고, 예를 들어 레트로바이러스이며, ssRNA(+)를 비리온 내로 패키징한다. 일부 구현예에서, 그룹 VI 바이러스는, 예를 들어 레트로바이러스로부터 선택된다. 일부 구현예에서, 레트로바이러스는 렌티바이러스, 예를 들어 HIV-1, HIV-2, SIV, BIV이다. 일부 구현예에서, 레트로바이러스는 스푸마바이러스(spumavirus), 예를 들어 포미 바이러스, 예를 들어, HFV, SFV, BFV이다. 일부 구현예에서, 그러한 비리온 내에 함유된 ssRNA(+)는 숙주 세포 내로의 형질도입시 mRNA로서 직접 제공될 수 있는 코딩 분자이며, 예를 들어 숙주 세포 내로 형질도입시 임의의 개입하는 핵산 복제 또는 중합 단계를 필요로 함이 없이 단백질로 직접 번역될 수 있다. 일부 구현예에서, ssRNA(+)는 먼저 역전사되고, 복제되어 dsDNA 게놈 중간체를 생성하고, 이로부터 mRNA가 숙주 세포 내에서 전사될 수 있다. 일부 구현예에서, ssRNA(+) 게놈을 갖는 RNA 바이러스는 바이러스 게놈을 이용하여 숙주 세포로 형질도입되는 비리온 내 효소, 예를 들어 RNA-의존성 DNA 중합효소도 운반하며, 이는 mRNA로 전사될 수 있고 숙주에 의해 번역될 수 있는 dsDNA 내로 ssRNA(+)를 카피할 수 있다. 일부 구현예에서, 그룹 VI 레트로바이러스로부터의 역전사효소는 유전자 변형 폴리펩티드의 역전사효소 도메인으로서 포함된다.

[0594] 일부 구현예에서, 바이러스는 그룹 VII 바이러스로부터 선택되고, 예를 들어 레트로바이러스이며, dsRNA를 비리온 내로 패키징한다. 일부 구현예에서, 그룹 VII 바이러스는, 예를 들어 헤파드나바이러스(Hepadnavirus)로부터 선택된다. 일부 구현예에서, 그러한 비리온에 함유된 dsRNA 중 하나 또는 두 가닥 모두는 숙주 세포 내로의 형질도입시 mRNA로서 직접 제공될 수 있는 코딩 분자이며, 예를 들어 숙주 세포 내로 형질도입시 임의의 개입하는 핵산 복제 또는 중합 단계를 필요로 함이 없이 단백질로 직접 번역될 수 있다. 일부 구현예에서, 그러한 비리온 내에 함유된 dsRNA의 하나 또는 두 가닥 모두는 먼저 역전사되고, 복제되어 dsDNA 게놈 중간체를 생성하며, 이로부터 mRNA가 숙주 세포 내에서 전사될 수 있다. 일부 구현예에서, dsRNA 게놈을 갖는 RNA 바이러스는 바이러스 게놈을 이용하여 숙주 세포로 형질도입되는 비리온 내 효소, 예를 들어 RNA-의존성 DNA 중합효소도 운반하며, 이는 mRNA로 전사될 수 있고 숙주에 의해 번역될 수 있는 dsDNA 내로 dsRNA를 카피할 수 있다. 일부 구현예에서, 그룹 VII 레트로바이러스로부터의 역전사효소는 유전자 변형 폴리펩티드의 역전사효소 도메인으로서 포함된다.

[0595] 일부 구현예에서, 본 발명에서 핵산을 전달하기 위해 사용된 비리온은 유전자 변형의 과정에 관여하는 효소들 또한 운반할 수 있다. 예를 들어, 레트로바이러스 비리온은 핵산과 함께 숙주 세포 내로 전달된 역전사효소 도메인을 함유할 수 있다. 일부 구현예에서, RNA 주형은 비리온 내에서 유전자 변형 폴리펩티드와, 둘 모두가 바이러스 입자로부터 핵산의 형질도입시 표적 세포에 함께 전달되도록 회합될 수 있다. 일부 구현예에서, 비리온 내 핵산은 DNA를 포함할 수 있으며, 예를 들어 선형 ssDNA, 선형 dsDNA, 환형 ssDNA, 환형 dsDNA, 미니서클 DNA, dbDNA, ceDNA를 포함할 수 있다. 일부 구현예에서, 비리온 내 핵산은 RNA, 예를 들어 RNA, 예를 들어, 선형 ssRNA, 선형 dsRNA, 환형 ssRNA, 환형 dsRNA를 포함할 수 있다. 일부 구현예에서, 바이러스 게놈은 숙주 세포 내로의 형질도입시 환형화될 수 있으며, 예를 들어 ssRNA 분자는 공유 결합되어 환형 ssRNA를 형성할 수 있고, 선형 dsRNA 분자는 공유 결합되어 환형 dsDNA 또는 하나 이상의 환형 ssRNA를 형성할 수 있다. 일부 구현예에서, 바이러스 게놈은 숙주 세포에서 회전 환 복제에 의해 복제될 수 있다. 일부 구현예에서, 바이러스 게놈은 단일 핵산 분자를 포함할 수 있고, 예를 들어 비-절편화된 게놈을 포함한다. 일부 구현예에서, 바이러스 게놈은 둘 이상의 핵산 분자를 포함할 수 있으며, 예를 들어 절편화된 게놈을 포함한다. 일부 구현예에서, 비리온 내 핵산은 하나 이상의 단백질과 회합될 수 있다. 일부 구현예에서, 비리온 내 하나 이상의 단백질은 형질도입시 숙주 세포로 전달될 수 있다. 일부 구현예에서, 천연 바이러스는, 비리온 패키징 신호를 표적 핵산에의 첨가함으로써 핵산 전달을 위해 조정될 수 있으며, 여기서 숙주 세포는 패키징 신호를 함유한 표적 핵산을 패키징하기 위해 사용된다.

[0596] 일부 구현예에서, 전달 비히클로 사용된 비리온은 공생하는 인간 바이러스를 포함할 수 있다. 일부 구현예에서, 전달 비히클로 사용된 비리온은 아넬로바이러스를 포함할 수 있으며, 이의 사용은 WO2018232017A1에 설명되고, 이는 그 전체가 본 명세서에 참고로 포함된다.

[0597] *AAV 투여*

[0598] 일부 구현예에서, 아데노-연관 바이러스(AAV)는 본 명세서에 기재된 시스템, 주형 핵산, 및/또는 폴리펩티드와 함께 사용된다. 일부 구현예에서, AAV는 본 명세서에 기재된 시스템, 주형 핵산, 및/또는 폴리펩티드를 전달, 투여 또는 패키징하기 위해 사용된다. 일부 구현예에서, AAV는 재조합 AAV(rAAV)이다.

[0599] 일부 구현예에서, 시스템은 (a) 본 명세서에 기재된 폴리펩티드 또는 이를 인코딩하는 핵산, (b) 본 명세서에 기재된 주형 핵산(예를 들어, 주형 RNA), 및 (c) 표적 조직에 특이적인 하나 이상의 제1 조직-특이적 발현-제어 서열을 포함하고, 여기서 표적 조직에 특이적인 하나 이상의 제1 조직-특이적 발현-제어 서열은 (a), (b), 또는 (a) 및 (b)과 작동적으로 연관되며, 여기서 (a)에 회합된 경우, (a)는 폴리펩티드를 인코딩하는 핵산을 포함한다.

[0600] 일부 구현예에서, 본 명세서에 기재된 시스템은 제1의 재조합 아데노-연관 바이러스(rAAV) 캡시드 단백질을 더 포함하고; 여기서(a) 또는 (b) 중 적어도 하나는 제1의 rAAV 캡시드 단백질과 회합되고, 여기서(a) 또는 (b) 중 적어도 하나는 AAV 역 말단 반복부(ITR)에 의해 측접된다.

[0601] 일부 구현예에서, (a) 및 (b)는 제1 rAAV 캡시드 단백질과 회합된다.

[0602] 일부 구현예에서, (a) 및 (b)는 단일 핵산 상에 존재한다.

[0603] 일부 구현예에서, 시스템은 제2의 rAAV 캡시드 단백질을 더 포함하고, 여기서(a) 또는 (b) 중 적어도 하나는 제2의 rAAV 캡시드 단백질과 회합되고, 여기서 제2의 rAAV 캡시드 단백질과 회합된 (a) 또는 (b) 중 적어도 하나는 제1의 rAAV 캡시드 단백질과 회합된 (a) 또는 (b) 중 적어도 하나와 상이하다.

[0604] 일부 구현예에서, 제1 또는 제2의 rAAV 캡시드 단백질과 회합된 (a) 또는 (b) 중 적어도 하나는 제1 또는 제2 rAAV 캡시드 단백질의 내부에 분산되며, 제1 또는 제2 rAAV 캡시드 단백질은 AAV 캡시드 입자의 형태로 존재한다.

[0605] 일부 구현예에서, 시스템은 나노입자를 더 포함하며, 여기서 나노입자는 (a) 또는 (b) 중 적어도 하나와 회합된다.

[0606] 일부 구현예에서, (a) 및 (b)는 각각: a) 제1 rAAV 캡시드 단백질 및 제2 rAAV 캡시드 단백질; b) 나노입자 및 제1 rAAV 캡시드 단백질; c) 제1 rAAV 캡시드 단백질; d) 제1 아데노바이러스 캡시드 단백질; e) 제1 나노입자 및 제2 나노입자; 또는 f) 제1 나노입자와 회합된다.

[0607] 바이러스 벡터는, 예를 들어 본 발명에 의해 제공된 방법에서의 이용을 위해, 본 발명에 의해 제공된 시스템 전체 또는 부분을 전달하는데 유용하다. 상이한 바이러스로부터 유래된 시스템은 폴리펩티드, 또는 핵산의 전달을 위해 사용되어왔으며; 예를 들어 인테그라제-결핍 렌티바이러스, 아데노바이러스, 아데노-연관 바이러스(AAV), 단순포진 바이러스, 및 배큘로바이러스(문헌[Hodge et al. Hum Gene Ther 2017; Narayanavari et al. Crit Rev Biochem Mol Biol 2017; Boehme et al. Curr Gene Ther 2015]에서 검토됨).

[0608] 아데노바이러스는 잘 정의된 생물학, 유전적 안정성, 높은 형질도입 효율, 및 거대규모 생산의 용이성을 제공하는 유전자 전달 비히클로 사용되어온 흔한 바이러스다(예를 들어, 문헌[Lee et al. Genes & Diseases 2017]에 의해 검토됨, 참고). 이들은 선형 dsDNA 게놈을 갖고, 조직 및 세포 향성(tropism)에서 상이한 다양한 혈청형으로 나타난다. 수령자 세포에서 감염성 바이러스의 복제를 방지하기 위해, 패키징을 위해 사용된 아데노바이러스 게놈은 일부 또는 모든 내인성 바이러스 단백질이 결실되며, 이는 바이러스 생산 세포에 트랜스로 제공된다. 이는 게놈을 헬퍼-의존성으로 만들며, 이는 소위 헬퍼 기능에 의해 제공되는 상실 구성요소들의 존재 하에서만 이들이 바이러스 입자들 내로 복제 및 패키징될 수 있음을 의미한다. 모든 바이러스 ORF가 제거된 헬퍼-의존성 아데노바이러스 시스템은 약 37 kb에 달하는 외래 DNA 패키징에 적합할 수 있다(Parks et al. J Virol 1997). 일부 구현예에서, 아데노바이러스 벡터는 유전자 변형 시스템의 폴리펩티드 또는 주형 구성요소에 상응하는 DNA를 전달하기 위해 사용되거나, 또는 이들 모두는 별개의 또는 동일한 아데노바이러스 벡터 상에 함유된다. 일부 구현예에서, 아데노바이러스는 자기-패키징을 할 수 없는 헬퍼-의존성 아데노바이러스(HD-AdV)이다. 일부 구현예에서, 아데노바이러스는 아데노바이러스 입자 내로의 패키징에 필요한 서열 구성요소를 유지하는 한편, 결실된 내인성 ORF의 모든 또는 실질적인 부분들을 가졌던 고용량 아데노바이러스(HC-AdV)이다. 이 유형의 벡터에 대해, 게놈 패키징에 필요한 아데노바이러스 서열만이 비코딩 서열이다: 양 말단 모두에서 역 말단 반복부(ITR) 및 5'-말단에서 패키징 신호(Jager et al. Nat Protoc 2009). 일부 구현예에서, 아데노바이러스 게놈은 최적 생산 및 안정성을 위한 최소 게놈 크기를 충족시키기 위한 스터퍼 DNA를 또한 포함한다(예를 들어, 문헌[Hausl et

al. Mol Ther 2010] 참조). 일부 구현예에서, 아데노바이러스는 유전자 변형 시스템을 간으로 전달하는 데 사용된다.

[0609] 일부 구현예에서, 아데노바이러스는 유전자 변형 시스템을 HSCs, 예를 들어, HDAd5/35++로 전달하기 위해 사용된다. HDAd5/35++는, 벡터를 간으로부터 탈표적화하는 변형된 혈청형 35를 갖는 아데노바이러스이다(Wang et al. Blood Adv 2019). 일부 구현예에서, 유전자 변형 시스템을 HSC로 전달하는 아데노바이러스는 원시적 HSC, 예를 CD46 상에서 특이적으로 발현되는 수용체를 이용한다.

[0610] 아데노-연관 바이러스(AAV)는 파보비리데(parvoviridae) 과에 속하며, 더 구체적으로는 디펜도파보바이러스 속을 구성한다. AAV 게놈은 선형 단일 가닥 DNA 분자로 구성되며, 이는 대략 4.7 킬로베이스(kb)를 함유하고, 비-구조적 Rep(복제) 및 구조적 Cap(캡시드) 단백질을 인코딩하는 2개의 주요한 오픈 판독 프레임(ORF)로 이루어진다. 캡 유전자 내에서 제2의 ORF는 조립-활성화 단백질(AAP)을 인코딩하는 것으로 확인되었다. AAV 코딩 영역에 측접하는 DNA는, 대략 145개 뉴클레오티드 길이인 2개의 시스-작용 역 말단 반복부로, DNA 복제의 프라이머로서 기능하는 에너지적으로 안정한 헤어핀 구조로 접힐 수 있는 단속된 회문 서열을 갖는다. DNA 복제에서 이들의 역할에 더하여, ITR 서열은 바이러스 DNA의 세포 게놈 내로의 통합, 숙주 게놈 또는 플라스미드로부터 구제, 및 바이러스 핵산의 성숙한 비리온 내로의 캡슐화에 관여되는 것으로 나타났다(Muzyczka, (1992) Curr. Top. Micro. Immunol. 158:97-129). 일부 구현예에서, 하나 이상의 유전자 변형 핵산 구성요소는, 바이러스 패키징을 위해, AAV로부터 유래된 ITR에 의해 측접된다. 예를 들어, WO2019113310을 참조한다.

[0611] 일부 구현예에서, 유전자 변형 시스템의 하나 이상의 구성요소는 적어도 하나의 AAV 벡터를 통해 운반된다. 일부 구현예에서, 적어도 하나의 AAV 벡터는 특정 세포, 조직, 유기체에 대한 향성에 대해 선택된다. 일부 구현예에서, AAV 벡터는 슈도타입(pseudotype)화, 예를 들어 AAV2/8되며, 여기서 캡시드 단백질이 AAV8로부터의 것으로 대체된 것을 제외한 AAV2는 작제물의 설계를 설명한다. 기재된 임의의 벡터는 슈도타입 유도체일 수 있는 것으로 이해되며, 여기서 AAV 게놈을 패키징하기 위해 사용된 캡시드 단백질은 상이한 AAV 혈청형의 것으로부터 유래된다. 벡터 선택에 제한을 두고자 하지 않지만, 예시적인 AAV 혈청형의 목록은 표 18에서 찾을 수 있다. 일부 구현예에서, 유전자 변형을 위해 사용되는 AAV는, 문헌에서 예증되어 온 바와 같이, 새로운 세포 또는 조직 향성에 대해 진전될 수 있다(예를 들어, 문헌[Davidsson et al. Proc Natl Acad Sci USA 2019]).

[0612] 일부 구현예에서, AAV 전달 벡터는 2개의 AAV 역 말단 반복부(ITR) 및 관심 뉴클레오티드 서열(예를 들어, 유전자 변형 폴리펩티드를 코딩하는 서열 또는 DNA 주형, 또는 이들 모두)을 갖는 벡터로, 상기 ITR 각각은 단속된(또는 비인접) 회문 서열, 즉 3개의 절편: 5'→판독시에는 동일하지만, 서로에 대하여 반대로 위치할 때 혼성화되는 제1 절편 및 마지막 절편, 및 동일한 절편들을 구분하는 상이한 절편으로 구성된 서열을 갖는다. 예를 들어, WO2012123430을 참조한다.

[0613] 통상적으로, 캡시드가 있는 AAV 비리온은 rAAV 또는 scAAV 게놈, Rep 단백질, 및 Cap 단백질을 인코딩하는 플라스미드 또는 플라스미드들을 도입함으로써 생산된다(Grimm et al, 1998). 이들 헬퍼 플라스미드들의 인 트랜스(in trans) 도입시, AAV 게놈은 숙주 게놈으로부터 "구제"되고(즉, 방출 및 순차적으로 회수됨), 감염성 AAV를 생산하기 위해 추가로 단백질막으로 둘러싸인다(encapsidated). 일부 구현예에서, 하나 이상의 유전자 변형 핵산은 ITR-측접된 핵산을 헬퍼 기능과 함께 패키징 세포 내로 도입함으로써 AAV 입자 내로 패키징된다.

[0614] 일부 구현예에서, AAV 게놈은 소위 자기-상보적인 게놈으로도 명명되며(scAAV로도 지칭됨), ITR들 사이에 위치한 서열은, 원하는 핵산 서열의 역전 보체에 추가하여, 원하는 핵산 서열(예를 들어, 유전자 변형 폴리펩티드 또는 주형, 또는 둘 모두를 인코딩하는 DNA) 둘 모두를 함유하여, 이들 두 구성요소들이 접혀 자기-하이브리드 형성될 수 있도록 한다. 일부 구현예에서, 자기-상보적인 모듈은 DNA가 둘로 되접히는 것을 가능하게 하는 개재 서열에 의해 분리되어, 예를 들어 스템-루프를 형성한다. scAAV는 dsDNA를 형성하기 위해 ITR 프라이밍 및 제2 가닥 합성에 우선 의존하기보다는, 핵으로의 도입시에 전사에 대한 준비가 된 장점을 갖는다. 일부 구현예에서, 유전자 변형 구성요소들 중 하나 이상은 scAAV로서 설계되며, 여기서 AAV ITR 사이의 서열은 자기-혼성화하여 dsDNA를 생성할 수 있는 2개의 역 상보화 모듈을 함유한다.

[0615] 일부 구현예에서, 세포로 전달된(예를 들어, 폴리펩티드를 인코딩하는, 또는 주형, 또는 둘 모두) 핵산은 폐쇄된-말단, 선형 듀플렉스 DNA(CELiD DNA 또는 ceDNA)이다. 일부 구현예에서, ceDNA는 AAV 게놈의 복제 형태로부터 유래된다(Li et al. PLoS One 2013). 일부 구현예에서, 핵산(예를 들어, 폴리펩티드를 인코딩하는, 또는 주형 DNA, 또는 이들 모두)은 ITR, 예를 들어 AAV ITR에 의해 측접되며, 여기서 적어도 하나의 ITR은 말단 분해 부위 및 반복 단백질 결합 부위를 포함한다(때때로 복제 단백질 결합 부위로서도 지칭됨). 일부 구현예에서, ITR은 아데노-연관 바이러스, 예를 들어, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10,

AAV11, AAV12, 또는 이의 조합으로부터 유래된다. 일부 구현예에서, ITR은 대칭적이다. 일부 구현예에서, ITR은 비대칭적이다. 일부 구현예에서, 적어도 하나의 Rep 단백질은 작제물의 복제를 가능하게 하기 위해 제공된다. 일부 구현예에서, 적어도 하나의 Rep 단백질은 아데노-연관 바이러스, 예를 들어, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, 또는 이의 조합으로부터 유래된다. 일부 구현예에서, ceDNA는 생산 세포에 (i) ITR, 예를 들어 AAV ITR에 의해 측접된 DNA, 및 (ii) ITR-의존성 복제에 필요한 구성 요소들, 예를 들어 AAV 단백질 Rep78 및 Rep52(또는 단백질을 인코딩하는 핵산)을 갖는 제공함으로써 생성된다. 일부 구현예에서, ceDNA는 어떤 캡시드 단백질도 없으며, 예를 들어 감염성 AAV 입자 내로 패키징되지 않는다. 일부 구현예에서, ceDNA는 LNP로 제형화된다(예를 들어, WO2019051289A1 참조).

[0616] 일부 구현예에서, ceDNA 벡터는 2개의 자가 상보적인 서열로 이루어지며, 예를 들어 비대칭 또는 대칭 또는 실질적으로 대칭인 여기 정의된 것과 같은 ITR, 상기 발현 카세트에 측접되고, 여기서 ceDNA 벡터는 캡시드 단백질과 회합되지 않는다. 일부 구현예에서, ceDNA 벡터는 AAV 게놈 내에서 발견되는 2개의 자가 상보적인 서열을 포함하며, 여기서 적어도 하나의 ITR은 작동적 Rep-결합 요소(RBE)(종종 여기서 "RBS"로 지칭되기도 함) 및 AAV의 말단 분해 부위(trs) 또는 RBE의 기능성 변이체를 포함한다. 예를 들어, WO2019113310을 참조한다.

[0617] 일부 구현예에서, AAV 게놈은 4개의 복제 단백질 및 3개의 캡시드 단백질을 인코딩하는 2개의 유전자를 각각 포함한다. 일부 구현예에서, 이 유전자들은 145-bp 역 말단 반복부(ITR)에 의해 어느 하나의 측면 상에서 측접된다. 일부 구현예에서, 비리온은 1:1:10 비로 생산된 3개 이하의 캡시드 단백질(Vp1, Vp2, 및/또는 Vp3)을 포함한다. 일부 구현예에서, 캡시드 단백질은 동일한 오픈 판독 프레임 및/또는 대체 스플라이싱(Vp1) 및 대안적인 번역 개시 부위(각각 Vp2 및 Vp3)로부터 생산된다. Vp3은 비리온에서 가장 풍부한 서브유닛이며, 바이러스의 향성을 규정하는 세포 표면에서 수용체 인식에 참여한다. 일부 구현예에서, Vp1은 포스포리파제 도메인, 예를 들어 Vp1의 N-말단에서 바이러스 감염성에의 기능을 포함한다.

[0618] 일부 구현예에서, 바이러스 벡터의 패키징 용량은 벡터 내로 패키징될 수 있는 유전자 변형 시스템의 크기를 제한한다. 예를 들어, AAV의 패키징 용량은 약 4.5 kb(예를 들어, 약 3.0, 3.5, 4.0, 4.5, 5.0, 5.5, 또는 6.0 kb)일 수 있으며, 예를 들어 하나 또는 2개의 역 말단 반복부(ITR), 예를 들어, 145 염기 ITR을 포함한다.

[0619] 일부 구현예에서, 재조합 AAV(rAAV)는 시스-작용 145-bp ITRs 측접하는 벡터 이식유전자 카세트를 포함하며, 이는 예를 들어 외래 DNA의 패키징을 위해 4.5 kb까지 제공한다. 감염에 이어, rAAV는 일부 경우에서, 본 발명의 융합 단백질을 발현하고, 환형 헤드-투-테일 콘카티머(concatemer)에서 에피좀적으로 존재함으로써 숙주 게놈 내로의 통합 없이 유지된다. rAAV는 예를 들어 시험관 내 및 생체 내에서 사용될 수 있다. 일부 구현예에서, AAV-매개 유전자 전달은 유전자의 코딩 서열의 길이가 야생형 AAV 게놈보다 동일하거나 더 클 것을 요구한다.

[0620] 이 크기를 초과하는 유전자의 AAV 전달 및/또는 거대한 생리학적 조절 요소의 사용은, 예를 들어 둘 이상의 단편으로 전달되도록 단백질(들)을 나눔으로써 달성될 수 있다. 일부 구현예에서, N-말단 단편은 인테인-N 서열에 융합된다. 일부 구현예에서, C-말단 단편은 인테인-C 서열에 융합된다. 구현예에서, 단편은 둘 이상의 AAV 벡터 내로 패키징된다.

[0621] 일부 구현예에서, 이중 AAV 벡터는 거대 이식유전자 발현 카세트를 2개의 별도의 절반(5 및 3 말단, 또는 헤드 및 테일)으로 분할시켜 제조되며, 예를 들어 여기서 카세트의 각각의 절반은 단일 AAV 벡터(<5 kb) 내로 패키징된다. 전장 이식유전자 발현 카세트의 재조립은, 이중 AAV 벡터 둘 모두에 의해 이후 일부 구현예에서 동일한 세포의 공동 감염시에 달성될 수 있다. 일부 구현예에서, 공동 감염은 다음 중 하나 이상에 따른다: (1) 5 및 3 게놈 사이의 상동성 재조합(HR)(이중 AAV 중복 벡터); (2) ITR-매개된 테일-투-헤드 콘카티머화(이중 AAV 트랜스-스플라이싱 벡터); 및/또는 (3) 이들 2개의 메커니즘의 조합(이중 AAV 하이브리드 벡터). 일부 구현예에서, 생체 내에서 이중 AAV 벡터의 사용은 전장 단백질의 발현을 생성한다. 일부 구현예에서, 이중 AAV 벡터 플랫폼의 이용은 크기가 약 4.0, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 또는 5.0 kb보다 더 큰 이식 유전자에 대한 효율적이고 독자생존가능한 유전자 전이 전략을 나타낸다. 일부 구현예에서, AAV 벡터는 예를 들어 핵산 및 펩티드의 시험관 내 생산에 있어서 표적 핵산을 이용하여 세포를 형질도입하는 데 사용될 수도 있다. 일부 구현예에서, AAV 벡터는 생체 내 및 생체 외 유전자 치료 절차에 사용될 수 있다(예를 들어, 문헌[West et al., Virology 160:38-47 (1987); 미국 특허 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest.94:1351 (1994)] 참조; 이들 각각은 그 전체가 본 명세서에 참고로 포함됨). 재조합 AAV 벡터의 구축은 문헌[미국 특허 5,173,414; Tratschin et al., Mol. Cell. Biol.5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol.4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); 및 Samulski et al., J. Virol.63:03822-3828 (1989)](그 전체가 본 명세서에 참고로 포함됨)을

포함한 많은 간행물에 기재되어 있다.

[0622] 일부 구현예에서, 본 명세서에 기재된 유전자 변형 폴리뉍티드(예를 들어, 하나 이상의 가이드 핵산 존재 또는 없음)는, AAV, 렌티바이러스, 아데노바이러스 또는 기타 플라스미드 또는 바이러스 벡터 유형을 이용하여, 특히 예를 들어 미국 특허 8,454,972(아데노바이러스용 제형, 용량), 미국 특허 8,404,658(AAV용 제형, 용량) 및 미국 특허 5,846,946(DNA 플라스미드용 제형, 용량), 및 렌티바이러스, AAV 및 아데노바이러스에 대한 임상 시험 및 간행물들로부터의 제형 및 용량을 사용하여 전달될 수 있다. 예를 들어, AAV의 경우, 투여 경로, 제형 및 용량은 미국 특허 8,454,972 및 AAV에 관한 임상 시험에서 기재된 것과 같을 수 있다. 아데노바이러스의 경우, 투여 경로, 제형 및 용량은 미국 특허 8,404,658 및 아데노바이러스에 관한 임상 시험에서 기재된 것과 같을 수 있다. 플라스미드 전달의 경우, 투여 경로, 제형 및 용량은 미국 특허 5,846,946 및 플라스미드를 포함하는 임상 연구에 기재된 것과 같을 수 있다. 용량은 평균 70 kg의 개인(예를 들어, 성인 남성)을 기준으로 하거나 외삽될 수 있으며, 환자, 대상, 상이한 체중의 포유동물 및 종에 대해 조절될 수 있다. 투여 빈도는 의약 또는 수의학 전문가(의사, 수의사)의 영역에 있으며, 환자 또는 대상의 연령, 성별, 일반 건강 상태 및 다루어지는 특정 상태 또는 증상을 포함한 일반적인 인자들에 따라 달라진다. 일부 구현예에서, 바이러스 벡터는 관심 조직 내로 주사될 수 있다. 세포 유형 특이적 유전자 변형에 대해, 유전자 변형 폴리뉍티드 및 선택적인 가이드 핵산의 발현은, 일부 구현예에서 세포-유형 특이적 프로모터에 의해 구동될 수 있다.

[0623] 일부 구현예에서, AAV는, 예를 들어 면역 반응을 활성화할 수 있는 세포 입자들의 초원심분리를 필요로 하지 않는 정제 방법으로 인해 낮은 독성을 가능하게 한다. 일부 구현예에서, AAV는 예를 들어 숙주 게놈 내로 실질적으로 통합되지 않기 때문에, 삽입성 돌연변이 생성 유발의 낮은 가능성을 가능하게 한다.

[0624] 일부 구현예에서, AAV는 약 4.4, 4.5, 4.6, 4.7, 또는 4.75 kb의 패키징 한계를 갖는다. 일부 구현예에서, 유전자 변형 폴리뉍티드-인코딩 서열, 프로모터, 및 전사 종결자는 단일 바이러스 벡터 내로 맞추어질 수 있다. SpCas9(4.1 kb)는 일부 경우에 AAV 내로 패키징하기 어려울 수 있다. 따라서, 일부 구현예에서, 다른 유전자 변형 폴리뉍티드 코딩 서열 또는 염기 편집기보다 길이가 더 짧은 유전자 변형 폴리뉍티드 코딩 서열이 사용된다. 일부 구현예에서, 유전자 변형 폴리뉍티드 인코딩 서열은 약 4.5 kb, 4.4 kb, 4.3 kb, 4.2 kb, 4.1 kb, 4 kb, 3.9 kb, 3.8 kb, 3.7 kb, 3.6 kb, 3.5 kb, 3.4 kb, 3.3 kb, 3.2 kb, 3.1 kb, 3 kb, 2.9 kb, 2.8 kb, 2.7 kb, 2.6 kb, 2.5 kb, 2 kb, 또는 1.5 kb 미만이다.

[0625] AAV는 AAV1, AAV2, AAV5 또는 이들의 임의의 조합일 수 있다. 일부 구현예에서, AAV의 유형은 표적화되는 세포에 대하여 선택된다; 예를 들어, AAV 혈청형 1, 2, 5 또는 하이브리드 캡시드 AAV1, AAV2, AAV5 또는 이들의 임의의 조합은 뇌 또는 뉴런 세포를 표적화하기 위해 선택될 수 있거나; 또는 AAV4는 심장 조직을 표적화하기 위해 선택될 수 있다. 일부 구현예에서, AAV8은 간으로의 전달을 위해 선택된다. 이들 세포와 관련하여 예시적인 AAV 혈청형은 예를 들어 문헌[Grimm, D. et al, J. Virol.82: 5887-5911 (2008)]에 기재되어 있다(그 전체가 본 명세서에 참고로 포함됨). 일부 구현예에서, AAV는 혈청형, 하위유형, 및 자연 발생 AAV 및 재조합 AAV를 지칭한다. AAV는 바이러스 그 자체 또는 그의 유도체를 지칭하는 데 사용될 수 있다. 일부 구현예에서, AAV는 AAV1, AAV2, AAV3, AAV3B, AAV4, AAV5, AAV6, AAV6.2, AAV7, AAVrh.64R1, AAVhu.37, AAVrh.8, AAVrh.32.33, AAV8, AAV9, AAV-DJ, AAV2/8, AAVrh10, AAVLK03, AV10, AAV11, AAV 12, rh10, 및 이들의 하이브리드, 조류 AAV, 소 AAV, 개 AAV, 말 AAV, 영장류 AAV, 비영장류 AAV, 및 양 AAV를 포함한다. AAV의 다양한 혈청형의 게놈 서열 및 천연 말단 반복부(TR)의 서열, Rep 단백질, 및 캡시드 서브유닛이 당업계에 알려져 있다. 그러한 서열은 문헌 또는 GeneBank와 같은 공공 데이터베이스에서 찾을 수 있다. 추가의 예시적인 AAV 혈청형이 표 18에 열거되어 있다.

[0626] **표 18. 예시적인 AAV 혈청형.**

표 18

[0627]

표적 조직	비히클	참조
간	AAV (AAV8[1], AAVrh.8[1], AAVhu.37[1], AAV2/8, AAV2/rh10[2], AAV9, AAV2, NP40[3], NP59[2,3], AAV3B[5], AAV-DJ[4], AAV-LK01[4], AAV-LK02[4], AAV-LK03[4], AAV-LK19[4], AAV5[7] 아데노바이러스 (Ad5, HC-AdV[6])	1. Wang *et al.*, *Mol. Ther.* **18**, 118-25 (2010) 2. Ginn *et al.*, *JHEP Reports*, 100065 (2019) 3. Paulk *et al.*, *Mol. Ther.* **26**, 289-303 (2018). 4. L. Lisowski *et al.*, *Nature*. **506**, 382-6 (2014). 5. L. Wang *et al.*, *Mol. Ther.* **23**, 1877-87 (2015). 6. Hausl *Mol Ther* (2010) 7. Davidoff *et al.*, *Mol. Ther.* 11, 875-88 (2005)
폐	AAV (AAV4, AAV5, AAV6[1], AAV9, H22[2]) 아데노바이러스 (Ad5, Ad3, Ad21, Ad14)[3]	1. Duncan *et al.*, *Mol Ther Methods Clin Dev* (2018) 2. Cooney *et al.*, Am J Respir Cell Mol Biol (2019) 3. Li *et al.*, *Mol Ther Methods Clin Dev* (2019)
피부	AAV (AAV6[1], AAV-LK19[2])	1. Petek et al., *Mol. Ther.* (2010)2. L. Lisowski *et al.*, *Nature*. **506**, 382-6 (2014).
HSCs	아데노바이러스 (HDAd5/35[++])	Wang et al. *Blood Adv* (2019)

[0628] 일부 구현예에서, 약제학적 조성물(예를 들어, 본 명세서에 기재된 바와 같은 AAV를 포함하는)은 10% 미만의 중공 캡시드(empty capsid), 8% 미만의 중공 캡시드, 7% 미만의 중공 캡시드, 5% 미만의 중공 캡시드, 3% 미만의 중공 캡시드, 또는 1 % 미만의 중공 캡시드를 갖는다. 일부 구현예에서, 약제학적 조성물은 약 5% 미만의 중공 캡시드를 갖는다. 일부 구현예에서, 중공 캡시드의 수는 검출 한계 아래이다. 일부 구현예에서, 예를 들어 중공 캡시드는 예를 들어 적거나 실질적으로 치료 이익이 없으면서 불리한 반응(예를 들어, 면역 반응, 염증 반응, 간 반응 및/또는 심장 반응)을 생성할 수 있기 때문에, 약제학적 조성물은 적은 양의 중공 캡시드를 갖는 것이 유리하다.

[0629] 일부 구현예에서, 약제학적 조성물 중 잔류 숙주 세포 단백질(rHCP)은, 1 x 10^{13} vg/ml 당 100 ng/ml 이하의 rHCP, 예를 들어, 1 x 10^{13} vg/ml 당 40 ng/ml 이하의 rHCP 또는 1 x 10^{13} vg/ml 당 1 내지 50 ng/ml rHCP이다. 일부 구현예에서, 약제학적 조성물은 1 x 10^{13} vg 당 10 ng 미만의 rHCP 또는 1 x 10^{13} vg 당 5 ng 미만의 rHCP, 1 x 10^{13} vg 당 4 ng 미만의 rHCP, 또는 1 x 10^{13} vg 당 3 ng 미만의 rHCP, 또는 이들 사이의 임의의 농도를 포함한다. 일부 구현예에서, 약제학적 조성물 중 잔류 숙주 세포 DNA(hcDNA)는 1 x 10^{13} vg/ml 당 5 x 10^{6} pg/ml 이하의 hcDNA, 1 x 10^{13} vg/ml 당 1.2 x 10^{6} pg/ml 이하의 hcDNA, 또는 1 x 10^{13} vg/ml당 1x 10^{5} pg/ml hcDNA이다. 일부 구현예에서, 상기 약제학적 조성물 중 잔류 숙주 세포는 DNA는 1 x 10^{13} vg 당 5.0 x 10^{5} pg 미만, 1 x 10^{13} vg 당 2.0 x 10^{5} pg 미만, 1 x 10^{13} vg 당 1.1 x 10^{5} pg 미만, 1 x 10^{13} vg 당 1.0 x 10^{5} pg 미만의 hcDNA, 1 x 10^{13} vg 당 0.9 x 10^{5} pg 미만의 hcDNA, 1 x 10^{13} vg 당 0.8 x 10^{5} pg 미만의 hcDNA, 또는 이들 사이의 임의의 농도이다.

[0630] 일부 구현예에서, 약제학적 조성물 중 잔류 플라스미드 DNA는 1.0 x 10^{13} vg/ml 당 1.7 x 10^{5} pg/ml 이하, 또는 1 x 1.0 x 10^{13} vg/ml 당 1 x 10^{5} pg/ml, 또는 1.0 x 10^{13} vg/ml 당 1.7 x 10^{6} pg/ml이다. 일부 구현예에서, 약제학적 조성물 중 잔류 DNA 플라스미드는 1.0 x 10^{13} vg을 기준으로 10.0 x 10^{5} pg 미만, 1.0 x 10^{13} vg을 기준으로 8.0 x 10^{5} pg 미만 또는 1.0 x 10^{13} vg을 기준으로 6.8 x 10^{5} pg 미만이다. 구현예에서,

약제학적 조성물은 1.0 x 10^{13} vg 당 0.5 ng 미만, 1.0 x 10^{13} vg 당 0.3 ng 미만, 1.0 x 10^{13} vg 당 0.22 ng 미만 또는 1.0 x 10^{13} vg 당 0.2 ng 미만 또는 임의의 중간 농도의 소 혈청 알부민(BSA)을 포함한다. 구현예에서, 약제학적 조성물 중 벤조나제는 1.0 x 10^{13} vg을 기준으로 0.2 ng 미만, 1.0 x 10^{13} vg을 기준으로 0.1 ng 미만, 1.0 x 10^{13} vg을 기준으로 0.09 ng 미만, 1.0 x 10^{13} vg을 기준으로 0.08 ng 미만 또는 임의의 중간 농도를 포함한다. 구현예에서, 약제학적 조성물 중 폴록사머 188은 약 10 내지 150 ppm, 약 15 내지 100 ppm 또는 약 20 내지 80 ppm이다. 구현예에서, 약제학적 조성물 중 세슘은 50 pg/g(ppm) 미만, 30 pg/g(ppm) 미만 또는 20 pg/g(ppm) 미만 또는 임의의 중간 농도이다.

[0631] 구현예에서, 약학 조성물은 예를 들어 SDS-PAGE에 의해 결정 시, 10% 미만, 8% 미만, 7% 미만, 6% 미만, 5% 미만, 4% 미만, 3% 미만, 2% 미만 또는 이들 사이의 임의의 백분율의 총 불순물을 포함한다. 구현예에서, 예를 들어 SDS-PAGE에 의해 결정 시, 총 순도는 90% 초과, 92% 초과, 93% 초과, 94% 초과, 95% 초과, 96% 초과, 97% 초과, 98% 초과 또는 이들 사이의 임의의 농도이다. 구현예에서, 예를 들어 SDS-PAGE에 의해 측정될 때, 하나가 아닌 명명되지 않은 관련 불순물은 5% 초과, 4% 초과, 3% 초과, 2% 초과, 또는 이들 사이의 임의의 백분율이다. 구현예에서, 약학 조성물은 총 캡시드에 대하여 충전된 캡시드의 백분율로(예를 들어 분석용 초원심분리에 의해 측정 시 피크 1 + 피크 2) 85% 초과, 86% 초과, 87% 초과, 88% 초과, 89% 초과, 90% 초과, 91% 초과, 91.9% 초과, 92% 초과, 93% 초과, 또는 이들 사이의 임의의 백분율이다. 약학 조성물의 구현예에서, 분석용 초원심분리에 의해 피크 1로 측정된 충전된 캡시드의 백분율은 20 내지 80%, 25 내지 75%, 30 내지 75%, 35 내지 75%, 또는 37.4 내지 70.3%이다. 약학 조성물의 구현예들에서, 분석용 초원심분리에 의해 피크 2로 측정된 충전된 캡시드의 백분율은 20 내지 80%, 20 내지 70%, 22 내지 65%, 24 내지 62%, 또는 24.9 내지 60.1%다.

[0632] 일 구현예에서, 약학 조성물은 1.0 내지 5.0 x 10^{13} vg/mL, 1.2 내지 3.0 x 10^{13} vg/mL 또는 1.7 내지 2.3 x 10^{13} vg/ml의 유전체 역가를 포함한다. 일 구현예에서, 약학 조성물은 5 CFU/mL 미만, 4 CFU/mL 미만, 3 CFU/mL 미만, 2 CFU/mL 미만 또는 1 CFU/mL 미만 또는 임의의 중간 생략산의 생물학적 부하를 나타낸다. 구현예에서, USP, 예를 들어 USP <85>(그 전체가 참고로서 포함됨)에 따른 내독소의 양은 1.0 EU/mL 미만, 0.8 EU/mL 미만, 또는 0.75 EU/mL 미만이다. 구현예에서, USP, 예를 들어 USP <785>(그 전체가 참고로서 포함됨)에 따른 약학 조성물의 삼투압농도는 350 내지 450 mOsm/kg, 370 내지 440 mOsm/kg 또는 390 내지 430 mOsm/kg이다. 구현예에서, 약학 조성물은 용기 당 25 ㎛ 초과인 입자 1200개 미만, 용기 당 25 ㎛ 초과인 입자 1000개 미만, 용기 당 25 ㎛ 초과인 입자 500개 미만, 또는 임의의 중간체 값을 함유한다. 구현예에서, 약학 조성물은 용기 당 10 ㎛ 초과인 입자 10,000개 미만, 용기 당 10 ㎛ 초과인 입자 8000개 미만, 용기 당 10 pm 초과인 입자 600개 미만을 함유한다.

[0633] 일 구현예에서, 약학 조성물은 0.5 내지 5.0 x 10^{13} vg/mL, 1.0 내지 4.0 x 10^{13} vg/mL, 1.5 내지 3.0 x 10^{13} vg/ml 또는 1.7 내지 2.3 x 10^{13} vg/ml의 유전체 역가를 갖는다. 일 구현예에서, 본 명세서에 기재된 약학 조성물은 다음 중 하나 이상을 포함한다: 1.0 x 10^{13} vg 당 약 0.09 ng 미만의 벤조나제, 약 30 pg/g(ppm) 미만의 세슘, 약 20 내지 80 ppm의 폴록사머 188, 1.0 x 10^{13} vg당 약 0.22 ng 미만의 BSA, 1.0 x 10^{13} vg 당 약 6.8 x 10^{5} pg 미만의 잔류 DNA 플라스미드, 1.0 x 10^{13} vg 당 약 1.1 x 10^{5} pg 미만의 잔류 hcDNA, 1.0 x 10^{13} vg 당 약 4 ng 미만의 rHCP, pH 7.7 내지 8.3, 약 390 내지 430 mOsm/kg, 용기 당 > 25 ㎛ 크기의 입자 약 600개 미만, 용기 당 > 10 ㎛ 크기의 입자 약 6000개 미만, 약 1.7 x 10^{13} 내지 2.3 x 10^{13} vg/mL의 유전체 역가, 1.0 x 10^{13} vg 당 약 3.9 x 10^{8} 내지 8.4 x 10^{10} IU의 감염 역가, 1.0 x 10^{13} vg당 약 100 내지 300 pg의 총 단백질, A7SMA 마우스에서 7.5 x 10^{13} vg/kg 용량의 바이러스 벡터를 이용하여 >24일의 평균 생존, 시험관 내 세포 기반 분석을 기준으로 약 70 내지 130%의 상대 및/또는 약 5% 의 중공 캡시드. 다양한 구현예에서, 본 명세서에 기재된 약학 조성물은 여기 논의된 임의의 바이러스 입자를 포함하고, 참조 표준의 ±20% 사이, ±15% 사이, ±10% 사이 또는 ±5% 내의 역가를 유지한다. 일부 구현예에서, 역가는 적합한 시험관 내 세포 분석 또는 생체 내 동물 모델을 사용하여 측정된다.

[0634] AAV 입자의 제조, 특성화 및 투약하는 추가의 방법은, 그 전체가 본 명세서에 참고로 포함된 WO2019094253에 교시되어 있다.

[0635] 본 발명과 합치되는 이용될 수 있는 추가의 rAAV 작제물은 문헌[Wang et al 2019]에 기재되어 있는 것들을 포함

하며, 이는 그의 표 1을 포함하여, 그 전체가 본 명세서에 참고로 포함된 //doi.org/10.1038/s41573-019-0012-9에서 입수가능하다.

[0636] *지질 나노입자*

[0637] 본 명세서에 제공된 방법 및 시스템은, 임의의 적합한 담체 또는 전달 양식을 이용할 수 있으며, 이는 특정 구현예에서 지질 나노입자(LNP)를 포함한다. 지질 나노입자는, 일부 구현예에서 하나 이상의 이온성 지질, 예컨대 하나 이상의 이온성 지질, 예컨대 비-양이온성 지질(예를 들어, 중성 또는 음이온, 또는 쯔비터이온성 지질); 하나 이상의 접합된 지질(예컨대 PEG-접합된 지질 또는 그 전체가 본 명세서에 참고로 포함된 WO2019217941의 표 5에 기재된 중합체에 접합된 지질); 하나 이상의 스테롤(예를 들어, 콜레스테롤); 및, 선택적으로, 하나 이상의 표적화 분자(예를 들어, 접합된 수용체, 수용체 리간드, 항체); 또는 상기의 조합을 포함한다.

[0638] 나노입자 형성에서 사용될 수 있는 지질(예를 들어, 지질 나노입자)은, 예를 들어, 잠고문헌으로 포함된 WO2019217941의 표 4에 기재된 것들을 포함하며 - 예를 들어 지질 함유 나노입자는 WO2019217941의 표 4의 하나 이상의 지질을 포함한다. 지질 나노입자는 중합체, 예컨대 참고로서 포함된 WO2019217941의 표 5에 기재된 중합체와 같은 추가의 요소들을 포함할 수 있다.

[0639] 일부 구현예에서, 존재하는 경우, 접합된 지질은 PEG-디아실글리세롤(DAG)(예컨대, 1-(모노메톡시-폴리에틸렌글리콜)-2,3-디미리스토일글리세롤(PEG-DMG)), PEG-디알킬옥시프로필(DAA), PEG-인지질, PEG-세라마이드(Cer), 페길화 포스파티딜에탄올아민(PEG-PE), PEG 숙시네이트 디아실글리세롤(PEGS-DAG)(예컨대 4-O-(2',3'-디(테트라데카노일옥시)프로필-1-O-(w-메톡시(폴리에톡시)에틸) 부탄디오에이트(PEG-S-DMG)), PEG 디알콕시프로필카르밤, N- (카르보닐-메톡시폴리에틸렌글리콜 2000)-1,2-디스테아로일-sn-글리세로-3-포스포에탄올아민 나트륨염, WO2019051289(참고로 포함됨)의 표 2에 기재된 것들, 및 상기의 조합들 중 하나 이상을 포함할 수 있다.

[0640] 일부 구현예에서, 지질 나노입자 내에 포함될 수 있는 스테롤은 하나 이상의 콜레스테롤 또는 콜레스테롤 유도체, 예컨대 참고로서 포함된 WO2009/127060 또는 US2010/0130588에 있는 것들을 포함한다. 추가의 예시적인 스테롤은 파이토스테롤을 포함하며, 본 명세서에 참고로 포함된 문헌[Eygeris et al (2020), dx.doi.org/10.1021/acs.nanolett.0c01386]에 기재된 것들을 포함한다.

[0641] 일부 구현예에서, 지질 입자는 이온화 지질, 비-양이온성 지질, 입자들의 집합을 저해하는 접합된 지질, 및 스테롤을 포함한다. 이들 구성요소들의 양은 독립적으로, 그리고 원하는 특성을 달성하기 위해 변화될 수 있다. 예를 들어, 일부 구현예에서, 지질 나노입자는 이온화 지질을 총 지질의 약 20 몰% 내지 약 90 몰%(다른 구현예에서 이는 지질 나노입자에 존재하는 총 지질의 20 내지 70%(몰), 30 내지 60%(몰) 또는 40 내지 50%(몰); 약 50 몰% 내지 약 90 몰%), 비-양이온성 지질을 총 지질의 약 5 몰% 내지 약 30 몰%, 접합된 지질을 총 지질의 약 0.5 몰% 내지 약 20 몰%의 양으로, 및 스테롤을 총 지질의 약 20 몰% 내지 약 50 몰%의 양으로 포함한다. 총 지질 대 핵산(예를 들어, 유전자 변형 폴리펩티드를 인코딩하는 핵산 또는 주형 핵산)의 비는 원하는 바에 따라 변화될 수 있다. 예를 들어, 총 지질 대 핵산(질량 또는 중량) 비는 약 10:1 내지 약 30:1이다.

[0642] 일부 구현예에서, 이온화 지질은 양이온성 지질, 이온화성 양이온성 지질, 예를 들어 pH에 따라 양의 하전된 형태 또는 중성 형태로 존재할 수 있는 양이온성 지질, 또는 쉽게 양자화될 수 있는 아민-함유 지질일 수 있다. 일부 구현예에서, 양이온성 지질은, 예를 들어 생리학적 조건 하에서 양으로 하전될 수 있는 지질이다. 예시적인 양이온성 지질은 양의 하전을 갖는 하나 이상의 아민 기(들)을 포함한다. 일부 구현예에서, 지질 입자는 제형 내에서 하나 이상의 중성 지질, 이온화성 아민-함유 지질, 생물분해성 알킨 지질, 스테로이드, 다중불포화 지질을 포함하는 인지질, 구조 지질(예를 들어 스테롤), PEG, 콜레스테롤 및 중합체 접합된 지질과 함께 양이온성 지질을 포함한다. 일부 구현예에서, 양이온성 지질은 이온화 양이온성 지질일 수 있다. 본 명세서에 개시된 바와 같은 예시적인 양이온성 지질은 6.0에 걸쳐 효과적인 pKa를 가질 수 있다. 구현예에서, 지질 나노입자는, 제1 양이온성 지질보다는, (예를 들어, 제1의 유효 pKa보다 더 큰) 상이한 유효한 pKa를 갖는 제2의 양이온성 지질을 포함할 수 있다. 지질 나노입자는 40 내지 60 몰%의 양이온성 지질, 중성 지질, 스테로이드, 중합체 접합된 지질, 및 치료제, 예를 들어 본 명세서에 기재된 핵산(예를 들어, RNA)(예를 들어, 주형 핵산 또는 유전자 변형 폴리펩티드를 인코딩하는 핵산)을, 지질 나노입자 내에 캡슐화되거나 또는 이와 회합되어 포함할 수 있다. 일부 구현예에서, 핵산은 양이온성 지질과 공동 제형화된다. 핵산은 LNP, 예를 들어 양이온성 지질을 포함하는 LNP의 표현에 흡착될 수 있다. 일부 구현예에서, 핵산은 LNP, 예를 들어 양이온성 지질을 포함하는 LNP 내에 캡슐화될 수 있다. 일부 구현예에서, 지질 나노입자는 예를 들어 표적화제로 코팅된 표적화 모이어티를 포함할 수 있다. 구현예에서, LNP 제형은 생물분해성이다. 일부 구현예에서, 본 명세서에 기재된 하나 이상의 지질을 포함하는 지질 나노입자, 예를 들어 화학식 i, 화학식 ii, 화학식 ii, 화학식 vii 및/또는 화학식 ix는 적어도 1%,

적어도 5%, 적어도 10%, 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 92%, 적어도 95%, 적어도 97%, 적어도 98% 또는 100%의 RNA 분자, 예를 들어 주형 RNA 및/또는 유전자 변형 폴리펩티드를 인코딩하는 mRNA를 캡슐화한다.

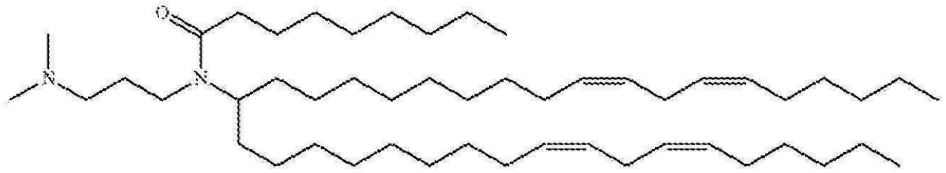
[0643] 일부 구현예에서, 지질 대 핵산 비(질량/질량 비; w/w 비)는 약 1:1 내지 약 25:1, 약 10:1 내지 약 14:1, 약 3:1 내지 약 15: 1, 약 4:1 내지 약 10:1, 약 5:1 내지 약 9:1, 또는 약 6:1 내지 약 9:1의 범위일 수 있다. 지질 및 핵산의 양은 원하는 N/P 비, 예를 들어 3, 4, 5, 6, 7, 8, 9, 10 이상의 N/P 비를 제공하도록 조절될 수 있다. 일반적으로, 지질 나노입자 제형의 전체 지질 함량은 약 5 mg/ml 내지 약 30 mg/mL의 범위일 수 있다.

[0644] 지질 입자 제형에서 사용될 수 있는 예시적인 이온화 지질은, 제한 없이 본 명세서에 참조로 포함된, WO2019051289의 표 1에 열거된 것들을 포함한다. 추가의 예시적인 지질은, 제한 없이 다음 화학식 중 하나 이상을 포함한다: US2016/0311759의 X; US20150376115 또는 US2016/0376224의 I; US20160151284의 I, II 또는 III; US20170210967의 I, IA, II, 또는 IIA; US20150140070의 I-c; US2013/0178541의 A; US2013/0303587 또는 US2013/0123338의 I; US2015/0141678의 I; US2015/0239926의 II, III, IV, 또는 V; US2017/0119904의 I; WO2017/117528의 I 또는 II; US2012/0149894의 A; US2015/0057373의 A; WO2013/116126의 A; US2013/0090372의 A; US2013/0274523의 A; US2013/0274504의 A; US2013/0053572의 A; WO2013/016058의 A; WO2012/162210의 A; US2008/042973의 I; US2012/01287670의 I, II, III, 또는 IV; US2014/0200257의 I 또는 II; US2015/0203446의 I, II, 또는 III; US2015/0005363의 I 또는 III; US2014/0308304의 I, IA, IB, IC, ID, II, IIA, IIB, IIC, IID, 또는 III-XXIV; US2013/0338210의; WO2009/132131의 I, II, III, 또는 IV; US2012/01011478의 A; US2012/0027796의 I 또는 XXXV; US2012/0058144의 XIV 또는 XVII; US2013/0323269의; US2011/0117125의 I; US2011/0256175의 I, II, 또는 III; US2012/0202871의 I, II, III, IV, V, VI, VII, VIII, IX, X, XI, XII; US2011/0076335의 I, II, III, IV, V, VI, VII, VIII, X, XII, XIII, XIV, XV, 또는 XVI; US2006/008378의 I 또는 II; US2013/0123338의 I; US2015/0064242의 I 또는 X-A-Y-Z; US2013/0022649의 XVI, XVII, 또는 XVIII; US2013/0116307의 I, II, 또는 III; US2013/0116307의 I, II, 또는 III; US2010/0062967의 I 또는 II; US2013/0189351의 I-X; US2014/0039032의 I; US2018/0028664의 V; US2016/0317458의 I; US2013/0195920의 I; US10,221,127의 5, 6, 또는 10; WO2018/081480의 III-3; WO2020/081938의 I-5 또는 I-8; US9,867,888의 18 또는 25; US2019/0136231의 A; WO2020/219876의 II; US2012/0027803의 1; US2019/0240349의 OF-02; US10,086,013의 23; 문헌 [Miao et al (2020)]의 cKK-E12/A6; WO2010/053572의 C12-200; 문헌 [Dahlman et al (2017)]의 7C1; 문헌 [Whitehead et al]의 304-013 또는 503-013; US9,708,628의 TS-P4C2; WO2020/106946의 I; WO2020/106946의 I.

[0645] 일부 구현예에서, 이온화 지질은 MC3(6Z,9Z,28Z,3 1Z)-헵타트리아콘타-6,9,28,3 1-테트라엔-19-일-4-(디메틸아미노) 부타노에이트(DLin-MC3-DMA 또는 MC3), 예를 들어 이는 WO2019051289A9(그 전체가 본 명세서에 참고로 포함됨)의 실시예 9에 기재된 바와 같다. 일부 구현예에서, 이온화 지질은 예를 들어 WO2019051289A9(그 전체가 본 명세서에 참고로 포함됨)의 실시예 10에 기재된 바와 같은 지질 ATX-002이다. 일부 구현예에서, 이온화 지질은 (13Z,16Z)-A,A-디메틸-3- 노닐도코사-13, 16-디엔-1-아민(화합물 32)로, 이는 예를 들어 WO2019051289A9(그 전체가 본 명세서에 참고로 포함됨)의 실시예 11에 기재된 바와 같다. 일부 구현예에서, 이온화 지질은 화합물 6 또는 화합물 22로, 이는 예를 들어 WO2019051289A9(그 전체가 본 명세서에 참고로 포함됨)의 실시예12에 기재된 바와 같다. 일부 구현예에서, 이온화 지질은 헵타데칸-9-일 8-((2-히드록시에틸)(6-옥소-6-(운데실옥시)헥실)아미노)옥타노에이트(SM-102); 예를 들어, US9,867,888(그 전체가 본 명세서에 참고로 포함됨)의 실시예 1에 기재된 바와 같다. 일부 구현예에서, 이온화 지질은 9Z,12Z)-3-((4,4-비스(옥틸옥시)부타노일)옥시)-2-((((3-(디에틸아미노)프로폭시)카르보닐)옥시)메틸)프로필 옥타데카-9,12-디에노에이트(LP01)이며, 예를 들어 WO2015/095340(그 전체가 본 명세서에 참고로 포함됨)의 실시예 13에서 합성되는 바와 같다. 일부 구현예에서, 이온화 지질은 디((Z)-비-2-엔-1-일) 9-((4-디메틸아미노)부타노일)옥시)헵타데칸디오에이트(L319)며, 예를 들어 US2012/0027803(그 전체가 본 명세서에 참고로 포함됨)의 실시예 7, 8 또는 9에서 합성되는 바와 같다. 일부 구현예에서, 이온화 지질은 1,1'-((2-(4-(2-((2-(비스(2-히드록시도데실)아미노)에틸)(2-히드록시도데실)아미노)에틸)피페라진-1-일)에틸)아잔디일)비스(도데칸-2-올)(C12-200)며, 예를 들어 WO2010/053572(그 전체가 본 명세서에 참고로 포함됨)의 실시예 14 및 16에서 합성되는 바와 같다. 일부 구현예에서, 이온화 지질은; 이미다졸 콜레스테롤 에스테르(ICE) 지질(3S, 10R, 13R, 17R)-10, 13-디메틸-17- ((R)-6-메틸헵탄-2-일)-2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17-테트라데카하이드로-1H-시클로펜타[a]페난트렌-3-일 3-(1H-이미다졸-4-일)프로파노에이트, 예를 들어 WO2020/106946(그 전체가 본 명세서에 참고로 포함됨)의 구조식 (I)이다.

[0646] 본 명세서에 기재된 조성물의 전달을 위해 지질 나노입자를 형성하기 위해(예를 들어, 다른 지질 성분과 조합되어) 사용될 수 있는 지질 화합물의 일부 비제한적인 예, 예를 들어 본 명세서에 기재된 핵산(예를 들어, RNA)(예를 들어, 주형 핵산 또는 유전자 변형 폴리펩티드를 인코딩하는 핵산)은 다음을 포함한다:

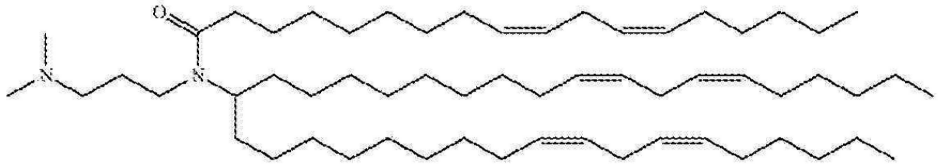
[0647] [화학식 i]



[0648] .

[0649] 일부 구현예에서, 화학식 i를 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

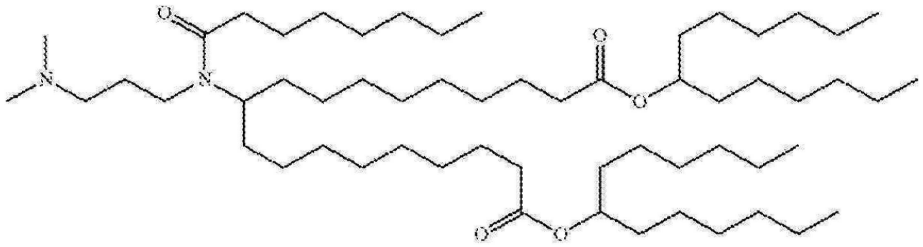
[0650] [화학식 ii]



[0651] .

[0652] 일부 구현예에서, 화학식 ii를 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

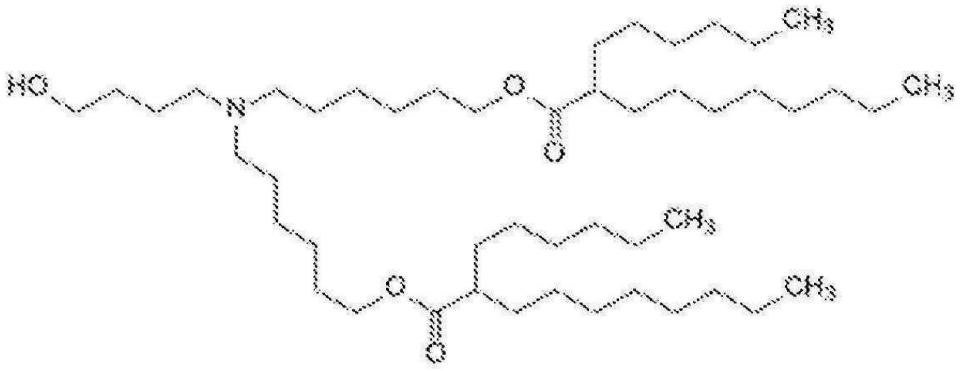
[0653] [화학식 iii]



[0654] .

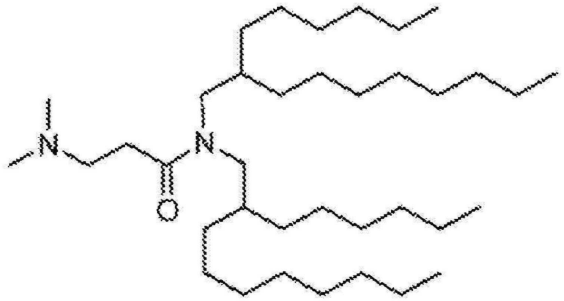
[0655] 일부 구현예에서, 화학식 iii을 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

[0656] [화학식 iv]



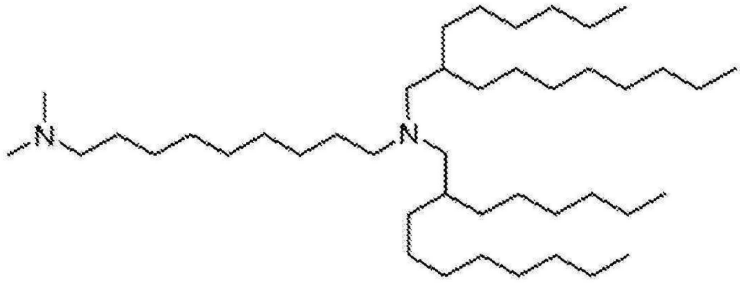
[0657]

[0658] [화학식 v]

[0659] 

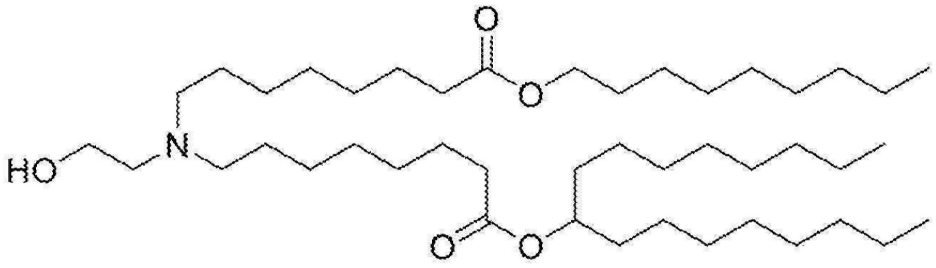
[0660] 일부 구현예에서, 화학식 v를 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

[0661] [화학식 vi]

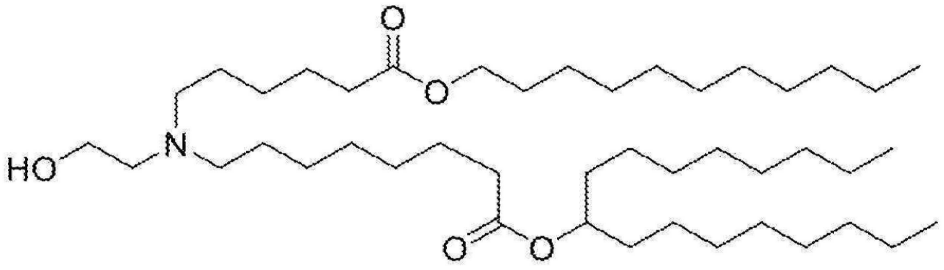
[0662] 

[0663] 일부 구현예에서, 화학식 vi을 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

[0664] [화학식 vii]

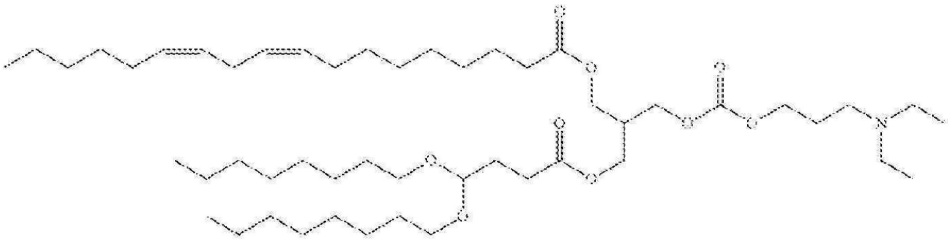
[0665] 

[0666] [화학식 viii]

[0667] 

[0668] 일부 구현예에서, 화학식 viii을 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

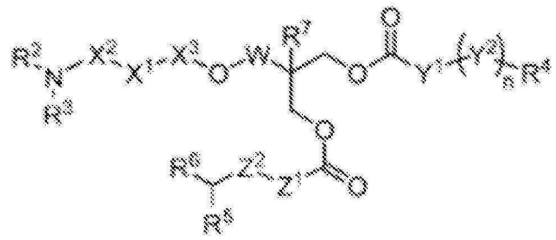
[0669] [화학식 ix]



[0670]

[0671] 일부 구현예에서, 화학식 ix을 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

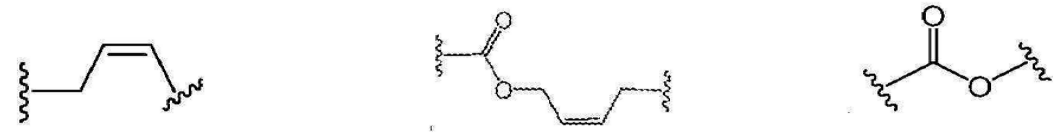
[0672] [화학식 x]



[0673]

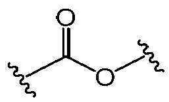
[0674] 여기서

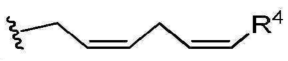
[0675] X^1는 O, NR^1, 또는 직접 결합이고, X^2는 C2-5 알킬렌이고, X^3은 C(=O) 또는 직접 결합이고, R^1은 H 또는 Me이고, R^3은 Ci-3 알킬이고, R^2는 Ci-3 알킬이거나, R^2는 이것이 부착된 질소 원자 및 X^2의 1 내지 3개의 탄소 원자와 함께 4-, 5-, 또는 6-원 고리를 형성하거나, 또는 X^1은 NR^1이고, R^1 및 R^2는 이들이 부착된 질소 원자와 함께 5- 또는 6-원 고리를 형성하거나, 또는 R^2는 R^3 및 이들이 부착된 질소 원자와 함께 5-, 6-, 또는 7-원 고리를 형성하고, Y^1은 C2-12 알킬렌이고, Y^2는 다음에서 선택되고,

[0676] 

[0677] (어느 하나의 배향), (어느 하나의 배향), (어느 하나의 배향),

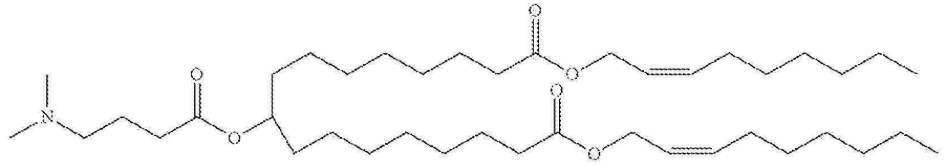
[0678] n은 0 내지 3이고, R^4는 Ci-15 알킬이고, Z^1은 Ci-6 알킬렌 또는 직접 결합이고,

[0679] Z는  (어느 하나의 배향) 또는 부재하며, 단, Z^1이 직접 결합인 경우, Z^2는 부재하고;

[0680] R^5는 C5-9 알킬 또는 C6-10 알콕시이고, R^6은 C5-9 알킬 또는 C6-10 알콕시이고, W는 메틸렌 또는 직접 결합이며, R^7은 H 또는 이의 염, 단, R^3 및 R^2는 C2 알킬이고, X^1은 O이고, X^2는 선형 C3 알킬렌이고, X^3은 C(=O)이고, Y^1은 선형 Ce 알킬렌이고, $(Y^2)_n$-R^4는  이고, R^4는 선형 C5 알킬이고, Z^1은 C2 알킬렌이고, Z^2는 부재하고, W는 메틸렌이고, R^7은 H이며, R^5 및 R^6은 Cx 알콕시가 아니다.

[0681] 일부 구현예에서, 화학식 xii을 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

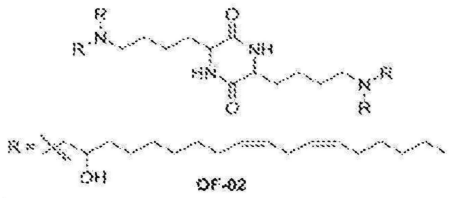
[0682] [화학식 xi]



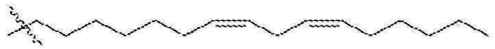
[0683]

[0684] 일부 구현예에서, 화학식 xi를 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

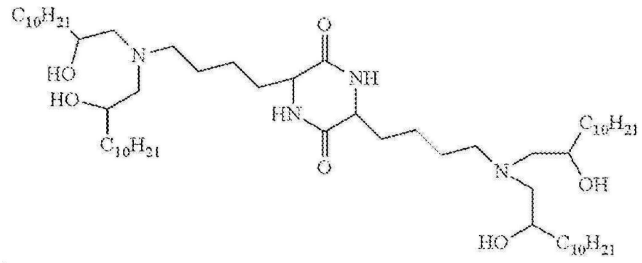
[0685] [화학식 xii]



[0686]

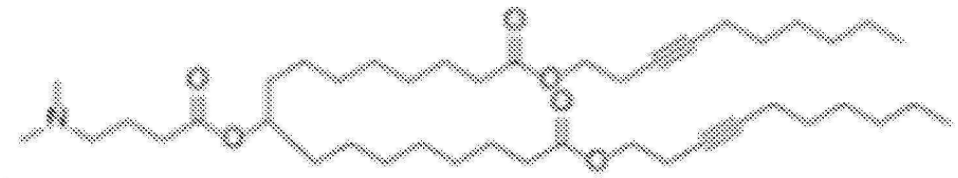
[0687] (여기서, R= )

[0688] [화학식 xiii]



[0689]

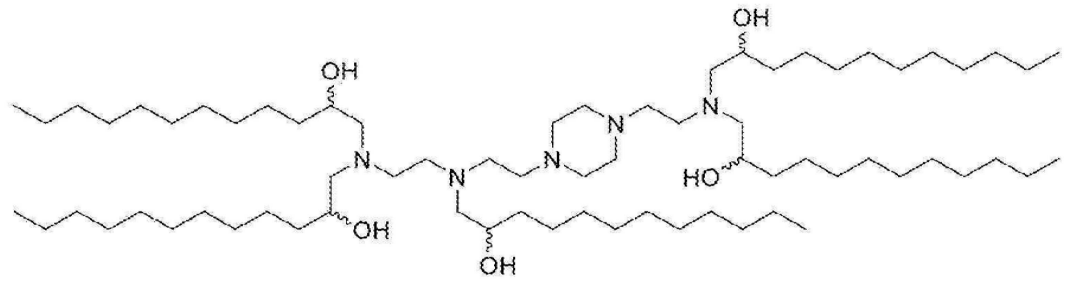
[0690] [화학식 xiv]



[0691] .

[0692] 일부 구현예에서 LNP는 화학식 xiii의 화합물 및 화학식 xiv의 화합물을 포함한다.

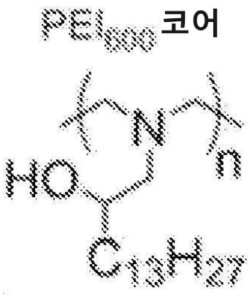
[0693] [화학식 xv]



[0694]

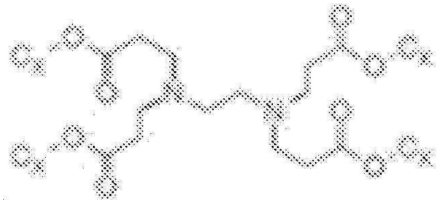
[0695] 일부 구현예에서, 화학식 xv를 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 간 및/또는 간세포에 전달하는 데 사용된다.

[0696] [화학식 xvi]

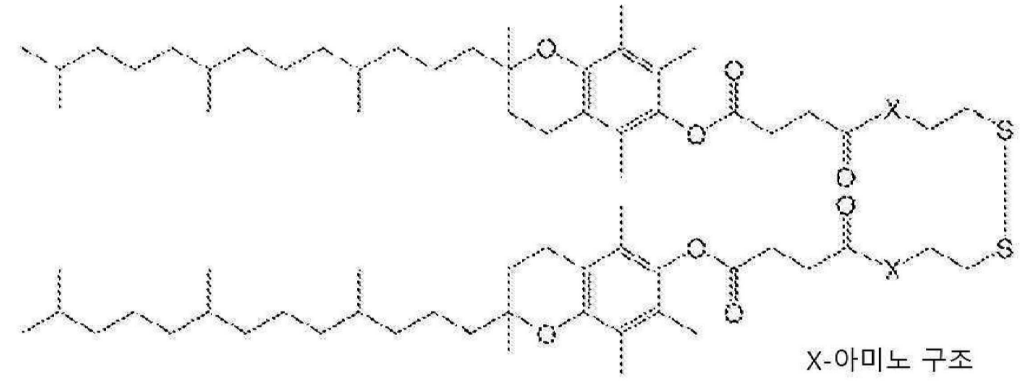
[0697] 

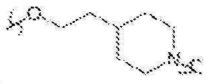
[0698] 일부 구현예에서, 화학식 xvi의 제형을 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 폐 내피 세포로 전달하는 데 사용된다.

[0699] [화학식 xvii]

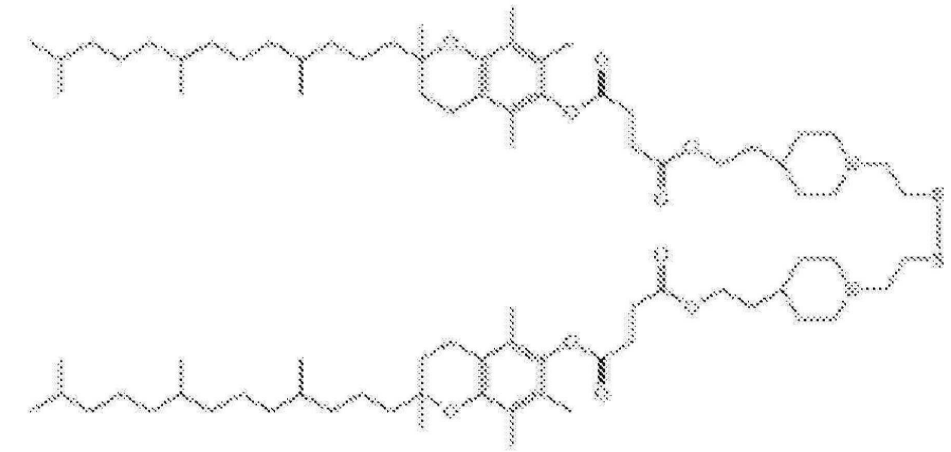
[0700] 

[0701] [화학식 xviii(a)]

[0702] 

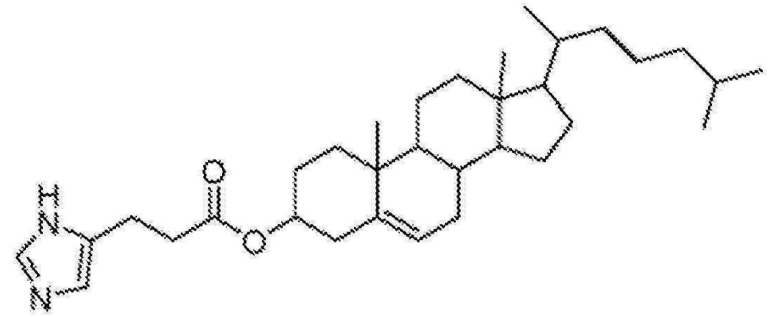
[0703] 여기서, X= 

[0704] [화학식 xviii(b)]



[0705]

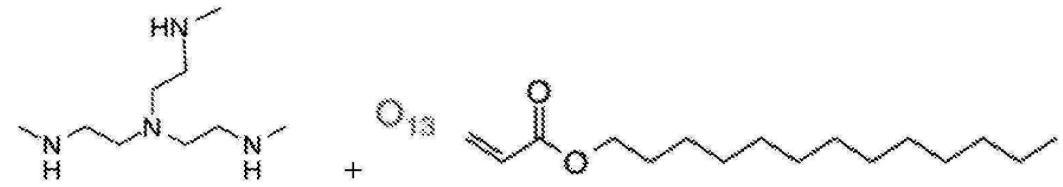
[0706] [화학식 xix]



[0707]

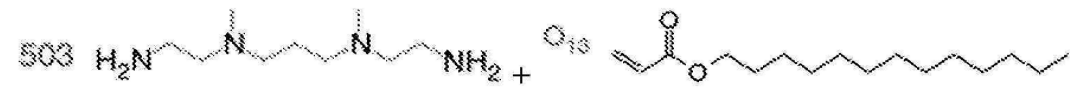
[0708] 일부 구현예에서, 본 명세서에 기재된 조성물의 전달을 위해 지질 나노입자를 형성하는 데 사용되는 지질 화합물, 예를 들어 본 명세서에 기재된 핵산(예를 들어, RNA)(예를 들어, 주형 핵산 또는 유전자 변형 폴리펩티드를 인코딩하는 핵산)은 다음 반응들 중 어느 하나에 의해 제조된다:

[0709] [화학식 xx(a)]



[0710]

[0711] [화학식 xx(b)]

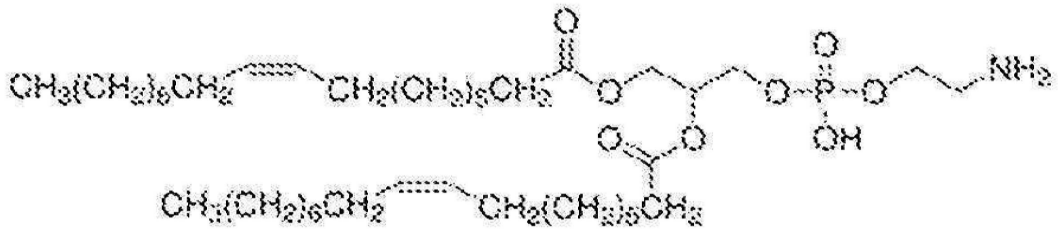


[0712]

[0713] 예시적인 비-양이온성 지질은 디스테아로일-sn-글리세로-포스포에탄올아민, 디스테아로릴포스파티딜콜린(DSPC), 디올레일포스파티딜콜린(DOPC), 디팔미토일포스파티딜콜린(DPPC), 디올레일포스파티딜글리세롤(DOPG), 디팔미토일포스파티딜글리세롤(DPPG), 디올레일포스파티딜에탄올아민(DOPE), 1,2-디올레일-sn-글리세로-3-포스포에탄올아민(DOPE), 팔미토일올레일포스파티딜콜린(POPC), 팔미토일올레일포스파티딜에탄올아민(POPE), 디올레일포스파티딜에탄올아민 4-(N-말레이미도메틸)-시클로헥산-1 - 카르복실레이트(DOPE-mal), 디팔미토일 포스파티딜 에탄올아민(DPPE), 디미리스토일포스포에탄올아민(DMPE), 디스테아로일-포스파티딜-에탄올아민(DSPE), 모노메틸-포스파티딜에탄올아민(예컨대 16-O-모노메틸 PE), 디메틸-포스파티딜에탄올아민(예컨대 16-O-디메틸 PE), 18-1-트랜스 PE, 1-스테아로일-2-올레일-포스파티딜에탄올아민(SOPE), 수소화 콩 포스파티딜콜린(HSPC), 달걀 포스파티딜콜린(EPC), 디올레일포스파티딜세린(DOPS), 스핑고미엘린(SM), 디미리스토일 포스파티딜콜린(DMPC), 디미리스

토일포스파티딜글리세롤(DMPG), 디스테아로일포스파티딜글리세롤(DSPG), 디에루코일포스파티딜콜린(DEPC), 팔미토일올레일포스파티딜글리세롤(POPG), 디엘라이도일-포스파티딜에탄올아민(DEPE), 레시틴, 포스파티딜에탄올아민, 리소레시틴, 리소포스파티딜에탄올아민, 포스파티딜세린, 포스파티딜이노시톨, 스핑고미엘린, 달갈 스핑고미엘린(ESM), 세팔린, 카르디올리핀, 포스파티드산, 세레브로사이드, 디세틸포스페이트, 리소포스파티딜콜린, 디리올레일포스파티딜콜린, 또는 이들의 혼합물을 포함하지만 이에 제한되지 않는다. 기타 디아실포스파티딜콜린 및 디아실포스파티딜에탄올아민 인지질이 또한 사용될 수 있는 것으로 이해된다. 이들 지질에서 아실 기는 바람직하게는 C10-C24의 탄소 사슬을 갖는 지방산, 예를 들어 라우로일, 미리스토일, 팔미토일, 스테아로일 또는 올레일로부터 유도된 아실 기이다. 추가의 예시적인 지질은 특정 구현예에서, 본 명세서에 참고로 포함된 문헌[Kim et al. (2020) dx.doi.org/10.1021/acs.nanolett.0c01386]을 제한없이 포함한다. 그러한 지질은 일부 구현예에서, mRNA와 함께 간 형질주입을 개선하는 것으로 밝혀진 식물 지질(예를 들어, DGTS)을 포함한다. 일부 구현예에서, 비-양이온성 지질은 다음 구조를 가질 수 있다:

[0714] [화학식 xxi]



[0715]

[0716] 지질 나노입자에서의 이용에 적합한 비-양이온성 지질의 기타 예는, 예를 들어, 스테아릴아민, 도데실아민(dodeeylamine), 헥사데실아민, 아세틸 팔미테이트, 글리세롤 리시놀레이트, 헥사데실 스테아레이트, 이소프로필 미리스테이트, 양쪽성 아크릴성 중합체, 트리에탄올아민-라우릴 설페이트, 알킬-아릴 설페이트 폴리에틸옥실레이트화 지방산 아미드, 디옥타데실 디메틸 암모늄 브로마이드, 세라마이드, 스핑고미엘린 등과 같은 비-인지질을 제한없이 포함한다. 기타 비-양이온성 지질은, 그 내용 전체가 본 명세서에 참고로 포함된, WO2017/099823 또는 미국 특허 공개 US2018/0028664에 기재되어 있다.

[0717] 일부 구현예에서, 비-양이온성 지질은 올레산 또는, 그 전체가 본 명세서에 참고로 포함된, US2018/0028664의 화학식 I, II 또는 IV의 화합물이다. 비-양이온성 지질은, 예를 들어 지질 나노입자에 존재하는 총 지질 중 0 내지 30%(몰)를 포함할 수 있다. 일부 구현예에서, 비-양이온성 지질 함량은 지질 나노입자에 존재하는 총 지질의 5 내지 20%(몰) 또는 10 내지 15%(몰)이다. 일 구현예에서, 이온화 가능한 지질 대 중성 지질의 몰 비율은 약 2:1 내지 약 8:1(예를 들어, 약 2:1, 3:1, 4:1, 5:1, 6:1, 7: 1 또는 8:1)의 범위이다.

[0718] 일부 구현예에서, 지질 나노입자는 임의의 인지질을 포함하지 않는다.

[0719] 일부 구현예에서, 지질 나노입자 막 온전성을 제공하기 위해, 스테롤과 같은 구성요소를 추가로 포함할 수 있다. 지질 나노입자에서 사용될 수 있는 일 예시적인 스테롤은 콜레스테롤 및 이의 유도체이다. 콜레스테롤 유도체의 비제한적인 예는, 극성 유사체들, 예컨대 5a-콜레스탄올(choiestanol), 53-코프로스탄올, 콜레스테릴-(2'-하이드록시)-에틸 에테르, 콜레스테릴-(4'-하이드록시)-부틸 에테르, 및 6-케토콜레스탄올; 비-극성 유사체들, 예컨대 5a-콜레스탄, 콜레스테논, 5a-콜레스타논, 5p-콜레스타논, 및 콜레스테릴 데카노에이트; 및 이들의 혼합물을 포함한다. 일부 구현예에서, 콜레스테롤 유도체는 극성 유사체, 예를 들어 콜레스테릴-(4 '-하이드록시)-부틸 에테르이다. 예시적인 콜레스테롤 유도체는 PCT 공개 WO2009/127060 및 미국 특허 공개 US2010/0130588에 기재되어 있으며, 이들 각각은 그 전체가 본 명세서에 참고로 포함된다.

[0720] 일부 구현예에서, 막 온전성을 제공하는 구성요소, 예컨대 스테롤은 지질 나노입자에 존재하는 총 지질의 0 내지 50%(몰)(예를 들어, 0 내지 10%, 10 내지 20%, 20 내지 30%, 30 내지 40%, 또는 40 내지 50%)를 포함할 수 있다. 일부 구현예에서, 그러한 구성요소는 지질 나노입자에 존재하는 총 지질 함량의 20 내지 50%(몰) 30 내지 40%(몰)이다.

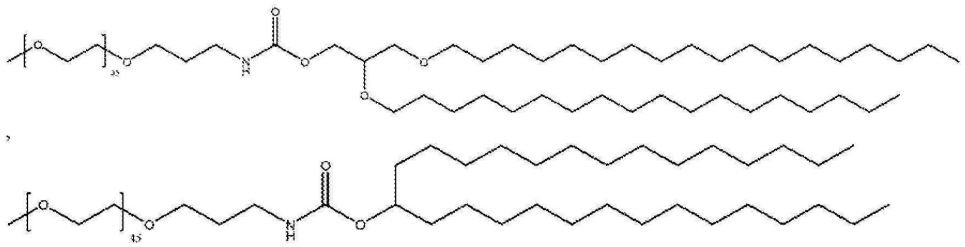
[0721] 일부 구현예에서, 지질 나노입자는 폴리에틸렌 글리콜(PEG) 또는 접합된 지질(conjugated lipid) 분자를 포함할 수 있다. 일반적으로, 이들은 지질 나노입자의 집합을 저해하는 데 사용되고/되거나 입체 안정화를 제공한다. 예시적인 접합된 지질은 이에 제한되지는 않지만, PEG-지질 접합물, 폴리옥사졸린(POZ)-지질 접합물, 폴리아미드-지질 접합물(예컨대 ATTA-지질 접합물), 양이온성-중합체 지질(CPL) 접합물, 및 이들의 혼합물을 포함한다.

일부 구현예에서, 접합된 지질 분자는 PEG-지질 접합물, 예를 들어(메톡시 폴리에틸렌글리콜)-접합된 지질이다.

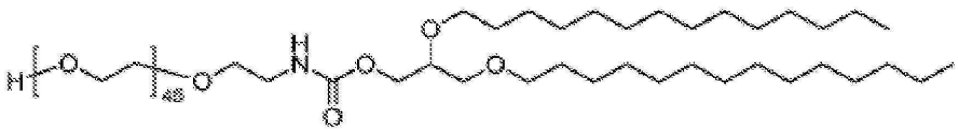
[0722] 예시적인 PEG-지질 접합물은 PEG-디아실글리세롤(DAG)(예컨대, 1-(모노메톡시-폴리에틸렌글리콜)-2,3-디미리스토일글리세롤(PEG-DMG)), PEG-디알킬옥시프로필(DAA), PEG-인지질, PEG-세라마이드(Cer), 페길화 포스파티딜에탄올아민(PEG-PE), 1,2-디미리스토일-sn-글리세롤, 메톡시폴리에틸렌 글리콜(DMG-PEG-2K), PEG 석시네이트 디아실글리세롤(PEGS-DAG)(예컨대 4-O-(2',3'-디(테트라데카노일옥시)프로필-1-O-(w-메톡시(폴리에톡시)에틸) 부탄디오에이트(PEG-S-DMG)), PEG 디알콕시프로필카르밤, N-(카르보닐-메톡시폴리에틸렌 글리콜 2000)-1,2-디스테아릴-sn-글리세로-3-포스포에탄올아민 나트륨 염, 또는 이들의 혼합물을 포함하지만 이에 제한되지 않는다. 추가의 예시적인 PEG-지질 접합물은, 예를 들어 US5,885,613, US6,287,591, US2003/0077829, US2003/0077829, US2005/0175682, US2008/0020058, US2011/0117125, US2010/0130588, US2016/0376224, US2017/0119904, 및 US/099823에 기재되어 있으며, 이들 모두의 내용은 그 전체가 본 명세서에 참고로 포함된다.

[0723] 일부 구현예에서, PEG-지질은, 그 내용 전체가 본 명세서에 참고로 포함된 US2018/0028664의 화학식 III, III-a-I, III-a-2, III-b-1, III-b-2, 또는 V의 화합물이다. 일부 구현예에서, PEG-지질은, 이들의 내용 전체가 본 명세서에 참고로 포함된, US20150376115 또는 US2016/0376224의 화학식 II이다. 일부 구현예에서, PEG-DAA 접합물은, 예를 들어 PEG-디라우릴옥시프로필, PEG-디미리스틸옥시프로필, PEG-디팔미틸옥시프로필, 또는 PEG-디스테아릴옥시프로필일 수 있다. PEG-지질은 PEG-DMG, PEG-디라우릴글리세롤, PEG-디팔미토일글리세롤, PEG-디스테릴글리세롤, PEG-디라우릴글리카미드, PEG-디미리스틸글리카미드, PEG-디팔미토일글리카미드, PEG-디스테릴글리카미드, PEG-콜레스테롤(1-[8'-(콜레스트-5-엔-3[베타]-옥시)카르복스아미드-3',6'-디옥시옥타닐] 카르바모일-[오메가]-메틸-폴리(에틸렌 글리콜), PEG-DMB(3,4-디테트라데옥시벤질-[오메가]-메틸-폴리(에틸렌 글리콜) 에테르), 및 1,2-디미리스토일-sn-글리세로-3-포스포에탄올아민-N-[메톡시(폴리에틸렌 글리콜)-2000] 중 하나 이상일 수 있다. 일부 구현예에서, PEG-지질은 PEG-DMG 1,2-디미리스토일-sn-글리세로-3-포스포에탄올아민-N-[메톡시(폴리에틸렌글리콜))-2000]을 포함한다. 일부 구현예에서, PEG-지질은 다음으로부터 선택되는 구조를 포함한다:

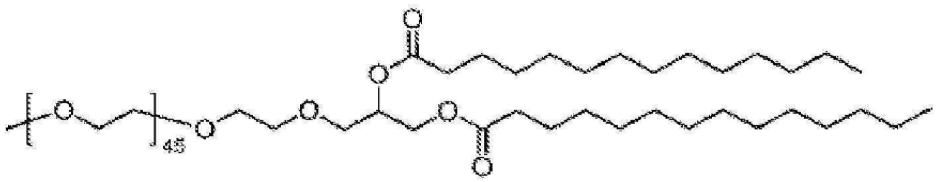
[0724] [화학식 xxii]

[0725] 

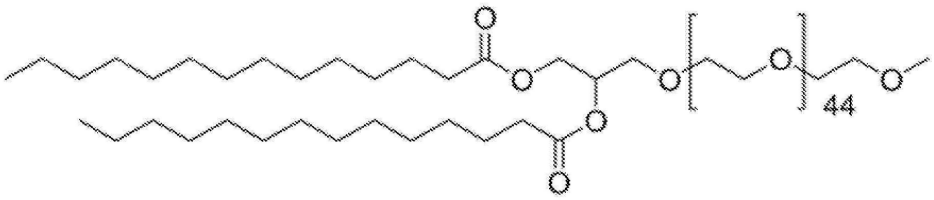
[0726] [화학식 xxiii]

[0727] 

[0728] [화학식 xxiv]

[0729]  , 및

[0730] [화학식 xxv]

[0731] 

[0732] 일부 구현예에서, PEG 외의 분자와 접합된 지질은 PEG-지질 대신 사용될 수도 있다. 예를 들어, 폴리옥사졸린(POZ)-지질 접합물, 폴리아미드-지질 접합물(예컨대 ATTA-지질 접합물), 및 양이온성-중합체 지질(GPL) 접합물은 PEG-지질 대신 또는 이에 추가하여 사용될 수 있다.

[0733] 예시적인 접합된 지질, 즉 PEG-지질, (POZ)-지질 접합물, ATTA-지질 접합물 및 양이온성 중합체-지질은 WO2019051289A9 및 WO2020106946A1의 표 2에 열거된 PCT 및 LIS에 기재되었으며, 이들 내용 모두는 그 전체가 본 명세서에 참고로 포함된다.

[0734] 일부 구현예에서, LNP는 화학식 xix의 화합물, 화학식 xxi의 화합물 및 화학식 xxv의 화합물을 포함한다. 일부 구현예에서, 화학식 xix, 화학식 xxi 및 화학식 xxv의 제형을 포함하는 LNP는 본 명세서에 기재된 유전자 변형 조성물을 폐 또는 허파 세포로 전달하는 데 사용된다.

[0735] 일부 구현예에서, 모든 지질 나노입자는 화학식 i, 화학식 ii, 화학식 iii, 화학식 vii, 및 화학식 ix로부터 선택되는 하나 이상의 양이온성 지질을 포함할 수 있다. 일부 구현예에서, LNP는 하나 이상의 중성 지질, 예를 들어 DSPC, DPPC, DMPC, DOPC, POPC, DOPE, SM, 스테로이드, 예를 들어 콜레스테롤, 및/또는 하나 이상의 중합체 접합 지질, 예를 들어 페길화 지질, 예를 들어 PEG-DAG, PEG-PE, PEG-S-DAG, PEG-cer 또는 PEG 디알콕시프로필 카르바메이트를 추가로 포함할 수 있다.

[0736] 일부 구현예에서, PEG 또는 접합된 지질은 지질 나노입자에 존재하는 총 지질의 0 내지 20%(몰)을 포함할 수 있다. 일부 구현예에서, PEG 또는 접합된 지질 함량은 지질 나노입자에 존재하는 총 지질의 0.5 내지 10% 또는 2 내지 5%(몰)이다. 이온성 지질, 비양이온성 지질, 스테롤 및 PEG/접합된 지질의 몰 비는 필요에 따라 변화될 수 있다. 예를 들어, 지질 입자는 조성물의 몰 기준 또는 총 중량 기준으로 30 내지 70%의 이온화 지질, 조성물의 몰 기준 또는 총 중량 기준으로 0 내지 60%의 콜레스테롤, 조성물의 몰 기준 또는 총 중량 기준으로 0 내지 30%의 비-양이온성 지질, 및 조성물의 몰 기준 또는 총 중량 기준으로 1 내지 10%의 접합된 지질을 포함할 수 있다. 바람직하게는, 조성물은 조성물의 몰 기준 또는 총 중량 기준으로 30 내지 40%의 이온화 지질, 조성물의 몰 기준 또는 총 중량 기준으로 40 내지 50%의 콜레스테롤, 및 조성물의 몰 기준 또는 총 중량 기준으로 10 내지 20%의 비-양이온성 지질을 포함한다. 일부 기타 구현예들에서, 조성물은 조성물의 몰 기준 또는 총 중량 기준으로 50 내지 75% 이온화 지질, 조성물의 몰 기준 또는 총 중량 기준으로 20 내지 40%의 콜레스테롤, 및 조성물의 몰 기준 또는 총 중량 기준으로 5 내지 10%의 비-양이온성 지질 및 조성물의 몰 기준 또는 총 중량 기준으로 1 내지 10%의 접합된 지질이다. 조성물은 조성물의 몰 기준 또는 총 중량 기준으로 60 내지 70%의 이온화 지질, 조성물의 몰 기준 또는 총 중량 기준으로 25 내지 35%의 콜레스테롤, 및 조성물의 몰 기준 또는 총 중량 기준으로 5 내지 10%의 비-양이온성 지질을 포함할 수 있다. 조성물은 조성물의 몰 기준 또는 총 중량 기준으로 90%의 이온화 지질 및 조성물의 몰 기준 또는 총 중량 기준으로 2 내지 15%의 비-양이온성 지질도 함유할 수 있다. 제형은, 예를 들어 조성물의 몰 기준 또는 총 중량 기준으로 8 내지 30%의 이온화 지질, 조성물의 몰 기준 또는 총 중량 기준으로 5 내지 30%의 비-양이온성 지질, 및 조성물의 몰 기준 또는 총 중량 기준으로 0 내지 20%의 콜레스테롤; 조성물의 몰 기준 또는 총 중량 기준으로 4 내지 25%의 이온화 지질, 조성물의 몰 기준 또는 총 중량 기준으로 4 내지 25%의 비-양이온성 지질, 조성물의 몰 기준 또는 총 중량 기준으로 2 내지 25%의 콜레스테롤, 조성물의 몰 기준 또는 총 중량 기준으로 10 내지 35%의 접합 지질, 및 조성물의 몰 기준 또는 총 중량 기준으로 5%의 콜레스테롤; 또는 조성물의 몰 기준 또는 총 중량 기준으로 2 내지 30%의 이온화 지질, 조성물의 몰 기준 또는 총 중량 기준으로 2 내지 30%의 비-양이온성 지질, 조성물의 몰 기준 또는 총 중량 기준으로 1 내지 15%의 콜레스테롤, 조성물의 몰 기준 또는 총 중량 기준으로 2 내지 35%의 접합 지질, 및 조성물의 몰 기준 또는 총 중량 기준으로 1 내지 20%의 콜레스테롤; 또는 심지어는 조성물의 몰 기준 또는 총 중량 기준으로 90% 이하의 이온화 지질 및 조성물의 몰 기준 또는 총 중량 기준으로 2 내지 10%의 비-양이온성 지질, 또는 조성물의 몰 기준 또는 총 중량 기준으로 심지어는 100%의 양이온성 지질을 포함하는 지질 나노입자 제형일 수도 있다. 일부 구현예에서, 지질 입자 제제는 50: 10:38.5: 1.5 몰 비의 이온화 지질, 인지질, 콜레스테롤 및 PEG-일화 지질을 포함한다. 일부 다른 구현예에서, 지질 입자 제형은 60:38.5: 1.5 몰 비의 이온화 지질, 콜레스테

롤 및 PEG화 지질을 포함한다.

[0737] 일부 구현예에서, 지질 입자는 이온화 지질, 비-양이온성 지질(예를 들어, 인지질), 스테롤(예를 들어, 콜레스테롤) 및 PEG화 지질을 포함하며, 여기서 지질의 몰 비는 이온화 지질의 경우 40~60을 목표로, 20~70 몰 백분율의 범위이며, 비-양이온성 지질의 몰 백분율은 0 내지 15를 목표로, 0 내지 30의 범위이며, 스테롤의 몰 백분율은 30 내지 50을 목표로, 20 내지 70의 범위이며, PEG화 지질의 몰 백분율은 2 내지 5를 목표로, 1 내지 6의 범위다.

[0738] 일부 구현예에서, 지질 입자는 50: 10:38.5: 1.5 몰 비의 이온화 지질 / 비-양이온성 지질/ 스테롤/ 접합된 지질을 포함한다.

[0739] 일 양태에서, 본 개시내용은 인지질, 레시틴, 포스파티딜콜린 및 포스파티딜에탄올아민을 포함하는 지질 나노입자 제형을 제공한다.

[0740] 일부 구현예에서, 하나 이상의 추가 화합물이 포함될 수도 있다. 그러한 화합물은 별도로 투여될 수 있거나, 또는 추가의 화합물은 본 발명의 지질 나노입자에 포함될 수 있다. 즉, 지질 나노입자는, 핵산에 더하여 기타 화합물 또는 첫 번째와 다른, 적어도 제2의 핵산을 함유할 수 있다. 제한없이, 기타 추가의 화합물들은 작은 또는 큰 유기 또는 무기 분자, 단당류, 이당류, 삼당류, 올리고당, 다당류, 펩티드, 단백질, 펩티드 유사체 및 이의 유도체, 펩티도미메틱스(peptidomimetics), 핵산, 핵산 유사체 및 이의 유도체, 생물학적 재료로 제조된 추출물, 또는 이의 임의의 조합으로 이루어진 군으로부터 선택될 수 있다.

[0741] 일부 구현예에서, 지질 나노입자(또는 지질 나노입자를 포함하는 제형)는 반응성 불순물(예를 들어, 알데히드 또는 케톤)을 결여하거나, 또는 예측된 수준 미만의 반응성 불순물(예를 들어, 알데히드 또는 케톤)을 포함한다. 이론에 구속되고자 하지는 않지만, 일부 구현예에서 지질 시약이 지질 나노입자 제형을 만드는 데 사용되며, 지질 시약은 오염성 반응성 불순물(예를 들어, 알데히드 또는 케톤)을 포함할 수 있다. 지질 시약은 예측된 수준 미만의 반응성 불순물(예를 들어, 알데히드 또는 케톤)을 갖는 것에 기초한 제조를 위해 선택될 수 있다. 이론에 구속되고자 함이 없이, 일부 구현예에서 알데히드는 RNA의 변형 및 손상, 예를 들어 염기들 간의 가교결합 및/또는 지질의 RNA로의 공유결합 접합(예를 들어, 지질-RNA 부가체를 형성)을 일으킬 수 있다. 이는 일부 경우에서, 역전사효소 반응의 실패 및/또는 부적절한 염기들의, 예를 들어 병소(들)의 부위(들)에서의 혼입, 예를 들어 새롭게 합성된 표적 DNA에서의 돌연변이를 야기할 수 있다.

[0742] 일부 구현예에서, 지질 나노입자 제형은 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 총 반응성 불순물(예를 들어, 알데히드) 함량을 포함하는 지질 시약을 사용하여 생산된다. 일부 구현예에서, 지질 나노입자 제형은 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 임의의 단일 반응성 불순물(예를 들어, 알데히드) 종을 포함하는 지질 시약을 사용하여 생산된다. 일부 구현예에서, 지질 나노입자 제형은, (i) 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 총 반응성 불순물(예를 들어, 알데히드) 함량; 및 (ii) 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 임의의 단일 반응성 불순물(예를 들어, 알데히드) 종을 포함하는 지질 시약을 사용하여 생산된다. 일부 구현예에서, 지질 나노입자 제형은 복수의 지질 시약을 사용하여 생산되며, 그 다수 중 각각의 지질 시약은 이 문단에서 설명된 하나 이상의 기준을 독립적으로 충족한다. 일부 구현예에서, 복수의 각각의 지질 시약은 동일한 기준, 예를 들어 이 문단의 기준을 충족한다.

[0743] 일부 구현예에서, 지질 나노입자 제형은 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 총 반응성 불순물(예를 들어, 알데히드) 함량을 포함한다. 일부 구현예에서, 지질 나노입자 제형은 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 임의의 단일 반응성 불순물(예를 들어, 알데히드) 종을 포함한다. 일부 구현예에서, 지질 나노입자 제형은, (i) 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 총 반응성 불순물(예를 들어, 알데히드) 함량; 및 (ii) 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 임의의 단일 반응성 불순물(예를 들어, 알데히드) 종을 포함한다.

[0744] 일부 구현예에서, 본 명세서에 기재된 바와 같은 지질 나노입자 또는 그의 제형에 사용된 지질 시약 중 하나 이상, 또는 선택적으로 모두는 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 총 반응성 불순물(예를 들어, 알데히드) 함량을 포함한다. 일부 구현예에서, 본 명세서에 기재된 바와 같은 지질 나노입자 또는 그의 제형에 사용된 지질 시약 중 하나 이상, 또는 선택적으로 모두는 5%, 4%,

3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 임의의 단일 반응성 불순물(예를 들어, 알데히드) 종을 포함한다. 일부 구현예에서, 본 명세서에 기재된 바와 같은 지질 나노입자 또는 그의 제형에 사용된 지질 시약 중 하나 이상, 또는 선택적으로 모두는, (i) 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 총 반응성 불순물(예를 들어, 알데히드) 함량; 및 (ii) 5%, 4%, 3%, 2%, 1%, 0.9%, 0.8%, 0.7%, 0.6%, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만의 임의의 단일 반응성 불순물(예를 들어, 알데히드) 종을 포함한다.

[0745] 일부 구현예에서, 총 알데히드 함량 및/또는 임의의 단일 반응성 불순물(예를 들어, 알데히드) 종의 양은, 예를 들어 PCT/US21/20948의 실시예 40에서 기재된 방법에 따라, 예를 들어 직렬 질량분석법(MS/MS)와 커플링된, 액체 크로마토그래피에 의해 결정된다. 일부 구현예에서, 반응성 불순물(예를 들어, 알데히드) 함량 및/또는 반응성 불순물(예를 들어, 알데히드) 종의 양은, 예를 들어 지질 시약에서 반응성 불순물(예를 들어 알데히드)의 존재와 회합된 핵산 분자(예를 들어 본 명세서에 기재된 바와 같은, 예를 들어 RNA 분자)의 하나 이상의 화학적 변형을 검출함으로써 결정된다. 일부 구현예에서, 반응성 불순물(예를 들어, 알데히드) 함량 및/또는 반응성 불순물(예를 들어, 알데히드) 종의 양은, 예를 들어 PCT/US21/20948의 실시예 41에 기재된 방법에 따라, 예를 들어 지질 시약에서, 반응성 불순물(예를 들어, 알데히드)의 존재와 회합된 뉴클레오티드 또는 뉴클레오시드(예를 들어, 리보뉴클레오티드 또는 리보뉴클레오시드, 예를 들어 본 명세서에 기재된 바와 같이, 예를 들어 주형 핵산에 포함된 또는 그로부터 단리된)의 하나 이상의 화학적 변형을 검출함으로써 결정된다. 구현예에서, 핵산 분자, 뉴클레오티드, 또는 뉴클레오시드의 화학적 변형은, 예를 들어 PCT/US21/20948의 실시예 41에 기재된 방법에 따라, 예를 들어 LC-MS/MS 분석을 사용하여, 하나 이상의 변형된 뉴클레오티드 또는 뉴클레오시드의 존재를 결정함으로써 검출된다.

[0746] 일부 구현예에서, 본 명세서에 기재된(예를 들어, 주형 핵산 또는 유전자 변형 폴리펩티드를 인코딩하는 핵산) 핵산(예를 들어, RNA은 알데히드 변형을 포함하지 않거나, 또는 알데히드 변형의 예비선택된 양 미만을 포함한다. 일부 구현예에서, 평균적으로, 핵산은 1000 뉴클레오티드 당 50, 20, 10, 5, 2, 또는 1 미만의 알데히드 변형을 갖고, 예를 들어 여기서 2개의 뉴클레오티드의 단일 가교결합은 단일 알데히드 변형이다. 일부 구현예에서, 알데히드 변형은 RNA 부가체(예를 들어, 지질-RNA 부가체)이다. 일부 구현예에서, 알데히드-변형된 뉴클레오티드는 염기들 간에 가교결합한다. 일부 구현예에서, 본 명세서에 기재된 핵산(예를 들어, RNA)은 뉴클레오티드 간의 50, 20, 10, 5, 2, 또는 1 미만의 가교결합을 포함한다.

[0747] 일부 구현예에서, LNP는 표적화 도메인의 첨가에 의해 특이적 조직으로 향한다. 예를 들어, 생물학적 리간드는 LNP의 표면 상에 디스플레이되어 동족 수용체를 디스플레이는 세포와의 상호작용을 향상할 수 있어서, 이에 따라 세포가 수용체를 발현하는 조직과의 회합 및 조직으로의 카고 전달을 구동한다. 일부 구현예에서, 생물학적 리간드는 간으로의 전달을 구동하는 리간드, 예를 들어 GalNAc를 디스플레이하는 LNP일 수 있으며, 이는 아시알로당단백질 수용체(ASGPR)를 디스플레이하는 간세포로 핵산 카고의 전달을 생성한다. 문헌[Akinc et al. *Mol Ther* 18(7):1357-1364 (2010)]의 연구는, 관찰가능한 LNP 카고 효과에 대한 ASGPR에 따라 달라지는 LNP를 산출하기 위한 PEG-지질(GalNAc-PEG-DSG)로의 3가 GalNAc 리간드의 접합을 교시한다(예를 들어, 이의 도 6 참조). 예를 들어 폴레이트, 트랜스페린 또는 항체를 포함하는, 기타 리간드를 디스플레이하는 LNP 제형은 그 전체가 본 명세서에 참고로 포함된 WO2017223135에서 그 안에 사용된 참조, 즉 문헌[Kolhatkar et al., Curr Drug Discov Technol. 2011 8:197-206; Musacchio and Torchilin, Front Biosci. 2011 16:1388-1412; Yu et al., Mol Membr Biol. 2010 27:286-298; Patil et al., Crit Rev Ther Drug Carrier Syst. 2008 25:1-61 ; Benoit et al., Biomacromolecules. 2011 12:2708-2714; Zhao et al., Expert Opin Drug Deliv. 2008 5:309-319; Akinc et al., Mol Ther. 2010 18:1357-1364; Srinivasan et al., Methods Mol Biol. 2012 820:105-116; Ben-Arie et al., Methods Mol Biol. 2012 757:497-507; Peer 2010 J Control Release. 20:63-68; Peer et al., Proc Natl Acad Sci U S A. 2007 104:4095-4100; Kim et al., Methods Mol Biol. 2011 721:339-353; Subramanya et al., Mol Ther. 2010 18:2028-2037; Song et al., Nat Biotechnol. 2005 23:709-717; Peer et al., Science. 2008 319:627-630; 및 Peer and Lieberman, Gene Ther. 2011 18:1127-1133]에 추가하여 논의된다.

[0748] 일부 구현예에서, LNP는, 전통적인 성분들, 예컨대 이온화 양이온성 지질, 양친매성 인지질, 콜레스테롤 및 폴리(에틸렌 글리콜)(PEG) 지질을 포함하는 제제에 선택적 기관 표적화(SORT) 분자의 첨가에 의해 조직-특이적 활성에 대해 선택된다. 문헌[Cheng et al. Nat Nanotechnol 15(4):313-320 (2020)]의 교시는 보충 "SORT" 구성성분의 첨가가 생체 내 RNA 전달 프로파일을 정확하게 변경하고 SORT 분자의 백분율 및 생물물리학적 특성의 함수로서 조직 특이적(예를 들어, 폐, 간, 비장) 유전자 전달 및 편집을 매개함을 보여준다.

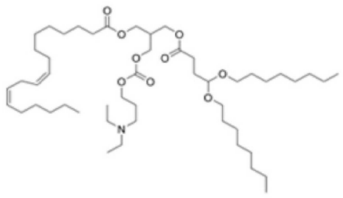
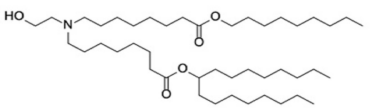
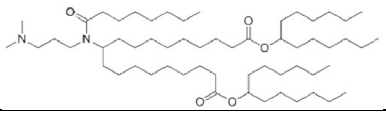
[0749] 일부 구현예에서, LNP는 생물분해성, 이온화 지질을 포함한다. 일부 구현예에서, LNP는 3-((4,4-비스(옥틸옥시)부타노일)옥시)-2-((((3-(디에틸아미노)프로폭시)카르보닐)옥시)메틸)프로필(9Z,12Z)-옥타데카-9,12-디에노에이트) 로도 지칭되는 (9Z,12Z)-3-((4,4-비스(옥틸옥시)부타노일)옥시)-2-((((3-(디에틸아미노)프로폭시)카르보닐)옥시)메틸)프로필 옥타데카-9,12-디에노에이트, 또는 또 다른 이온화 지질을 포함한다. 예를 들어, WO2019/067992, WO/2017/173054, WO2015/095340, 및 WO2014/136086뿐만 아니라, 그 안에 제공된 참고문헌의 지질을 참조한다. 일부 구현예에서, LNP 지질의 맥락에서 양이온성 및 이온화라는 용어는 상호 교환가능하며, 예를 들어 여기서 이온화 지질은 pH에 따라 달라지는 양이온성이다.

[0750] 일부 구현예에서, 본 명세서에 기재된 LNP는 표 19에 기재된 지질을 포함한다.

[0751] **표 19 예시적인 지질**

표 19

[0752]

지질 ID	화학명	분자량	구조
LIPIDV003	(9Z,12Z)-3-((4,4- 비스(옥틸옥시)부타노일)옥시)-2-((((3-(디에틸아미노)프로폭시)카르보닐)옥시)메틸)프로필 옥타데카-9, 12-디에노에이트	852.29	
LIPIDV004	헵타데칸-9-일 8-((2-하이드록시에틸)(8-(노닐옥시)-8-옥소옥틸)아미노)옥타노에이트	710.18	
LIPIDV005		919.56	

[0753] 일부 구현예에서, 유전자 변형 시스템의 다수의 구성요소는 단일 LNP 제형으로서 제조될 수 있으며, 예를 들어 LNP 제형은 유전자 변형 폴리펩티드 및 RNA 주형을 인코딩하는 mRNA를 포함한다. 핵산 구성요소의 비는 치료제의 특성을 최대화하기 위해 변화될 수 있다. 일부 구현예에서, RNA 주형 대 유전자 변형 폴리펩티드를 인코딩하는 mRNA의 비는 몰 비로, 약 1:1 내지 100:1, 예를 들어, 약 1:1 내지 20:1, 약 20:1 내지 40:1, 약 40:1 내지 60:1, 약 60:1 내지 80:1, 또는 약 80:1 내지 100:1이다. 다른 구현예에서, 다수의 핵산의 시스템은 별도의 제형, 예를 들어 주형 RNA를 포함하는 하나의 LNP 제형 및 유전자 변형 폴리펩티드를 인코딩하는 mRNAf를 포함하는 제2의 LNP 제형에 의해 제조될 수 있다. 일부 구현예에서, 시스템은 LNP로 제조된 2개가 넘는 핵산 구성요소를 포함할 수 있다. 일부 구현예에서, 시스템은 단백질, 예를 들어 유전자 변형 폴리펩티드, 및 적어도 하나의 LNP 제형으로 제조된 주형 RNA를 포함할 수 있다.

[0754] 일부 구현예에서, LNP 제형의 평균 LNP 직경은, 예를 들어 동적 광산란(DLS)에 의해 측정하여 10 범위의 nm 및 100 범위의 nm 사이일 수 있다. 일부 구현예에서, LNP 제형의 평균 LNP 직경은 약 40 nm 내지 약 150 nm, 예컨대 약 40 nm, 45 nm, 50 nm, 55 nm, 60 nm, 65 nm, 70 nm, 75 nm, 80 nm, 85 nm, 90 nm, 95 nm, 100 nm, 105 nm, 110 nm, 115 nm, 120 nm, 125 nm, 130 nm, 135 nm, 140 nm, 145 nm, 또는 150 nm일 수 있다. 일부 구현예에서, LNP 제형의 평균 LNP 직경은 약 50 nm 내지 약 100 nm, 약 50 nm 내지 약 90 nm, 약 50 nm 내지 약 80 nm, 약 50 nm 내지 약 70 nm, 약 50 nm 내지 약 60 nm, 약 60 nm 내지 약 100 nm, 약 60 nm 내지 약 90 nm, 약 60 nm 내지 약 80 nm, 약 60 nm 내지 약 70 nm, 약 70 nm 내지 약 100 nm, 약 70 nm 내지 약 90 nm, 약 70 nm 내지 약 80 nm, 약 80 nm 내지 약 100 nm, 약 80 nm 내지 약 90 nm, 또는 약 90 nm 내지 약 100 nm일 수 있다. 일부 구현예에서, LNP 제형의 평균 LNP 직경은 약 70 nm 내지 약 100 nm일 수 있다. 특정 구현예에서, LNP 제형의 평균 LNP 직경은 약 80 nm일 수 있다. 일부 구현예에서, LNP 제형의 평균 LNP 직경은 약 100 nm일 수 있다. 일부 구현예에서, LNP 제형의 평균 LNP 직경은 약 1 mm 내지 약 500 mm, 약 5 mm 내지 약 200 mm, 약 10 mm 내지 약 100 mm, 약 20 mm 내지 약 80 mm, 약 25 mm 내지 약 60 mm, 약 30 mm 내지 약 55 mm, 약 35 mm 내지 약 50 mm, 또는 약 38 mm 내지 약 42 mm의 범위이다.

[0755] LNP는 일부 경우에서 상대적으로 균질할 수 있다. 다분산성 지수는 LNP의 균질성, 예를 들어 지질 나노입자의

입자 크기 분포를 나타내는 데 사용될 수 있다. 작은(예를 들어, 0.3 미만) 다분산성 지수는 일반적으로 좁은 입자 크기 분포를 나타낸다. LNP는 약 0 내지 약 0.25, 예컨대 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.10, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.20, 0.21, 0.22, 0.23, 0.24, 또는 0.25의 다분산성 지수를 가질 수 있다. 일부 구현예에서, LNP의 다분산성 지수는 약 0.10 내지 약 0.20일 수 있다.

[0756] LNP의 제타 전위는 조성물의 동전위를 나타내는 데 사용될 수 있다. 일부 구현예에서, 제타 전위는 LNP의 표면 하전을 기재할 수 있다. 상대적으로 낮은 양 또는 음의 하전을 갖는 지질 나노입자가 일반적으로 바람직한데, 이는 고하전 종이 많을수록 신체에서 세포, 조직 및 기타 요소와 바람직하지 않게 상호작용할 수 있다. 일부 구현예에서, LNP의 제타 전위는 약 -10 mV 내지 약 +20 mV, 약 -10 mV 내지 약 +15 mV, 약 -10 mV 내지 약 +10 mV, 약 -10 mV 내지 약 +5 mV, 약 -10 mV 내지 약 0 mV, 약 -10 mV 내지 약 -5 mV, 약 -5 mV 내지 약 +20 mV, 약 -5 mV 내지 약 +15 mV, 약 -5 mV 내지 약 +10 mV, 약 -5 mV 내지 약 +5 mV, 약 -5 mV 내지 약 0 mV, 약 0 mV 내지 약 +20 mV, 약 0 mV 내지 약 +15 mV, 약 0 mV 내지 약 +10 mV, 약 0 mV 내지 약 +5 mV, 약 +5 mV 내지 약 +20 mV, 약 +5 mV 내지 약 +15 mV, 또는 약 +5 mV 내지 약 +10 mV일 수 있다.

[0757] 단백질 및/또는 핵산, 예를 들어 유전자 변형 폴리펩티드 또는 폴리펩티드를 인코딩하는 mRNA의 캡슐화의 효율은, 제공된 최초 양에 비해, 캡슐화되거나 다르게는, 제조 후 LNP와 회합된, 단백질 및/또는 핵산의 양을 기재한다. 캡슐화 효율은 높은 것(예를 들어, 100%에 가까운 것)이 바람직하다. 캡슐화 효율은, 예를 들어 하나 이상의 유기 용매 또는 세제로 지질 나노입자를 분해하기 전과 후에 지질 나노입자를 함유하는 용액 중의 단백질 또는 핵산의 양을 비교함으로써, 측정될 수 있다. 음이온 교환 수지는 용액 중 자유 단백질 또는 핵산(예를 들어, RNA)의 양을 측정하는 데 사용될 수 있다. 형광은 용액 중 자유 단백질 및/또는 핵산(예를 들어, RNA)의 양을 측정하는 데 사용될 수 있다. 본 명세서에 기재된 지질 나노입자의 경우, 단백질 및/또는 핵산의 캡슐화 효율은 적어도 50%, 예를 들어 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, 또는 100%일 수 있다. 일부 구현예에서, 캡슐화 효율은 적어도 80%일 수 있다. 일부 구현예에서, 캡슐화 효율은 적어도 90%일 수 있다. 일부 구현예에서, 캡슐화 효율은 적어도 95%일 수 있다.

[0758] LNP는 하나 이상의 코팅을 선택적으로 포함할 수 있다. 일부 구현예에서, LNP는 코팅을 갖는 캡슐, 필름 또는 정제로 제형화될 수 있다. 본 명세서에 기재된 조성물을 포함하는 캡슐, 필름, 또는 정제는 임의의 유용한 크기, 인장 강도, 경도 또는 밀도를 가질 수 있다.

[0759] LNP의 추가의 예시적인 지질, 제형, 방법, 및 특성화는, WO2020061457에 의해 교시되며, 이는 그 전체가 본 명세서에 참고로 포함된다.

[0760] 일부 구현예에서, 시험관 내 또는 생체 외 세포 리포펙션은 Lipofectamine MessengerMax(Thermo Fisher) 또는 TransIT-mRNA 형질주입 시약(Mirus Bio)을 사용하여 수행된다. 특정 구현예에서, LNP는 GenVoy_ILM 이온화 지질 혼합물(Precision Nanosystems)을 사용하여 제형화된다. 특정 구현예에서, LNP는 2,2-디리놀레일-4-디메틸아미노에틸-[1,3]-디옥솔란(DLin-KC2-MA) 또는 디리놀레일메틸-4-디메틸아미노부티레이트(DLin-MC3-DMA 또는 MC3)을 사용하여 제형화되며, 상기 제형 및 이의 생체 내 사용은, 그 전체가 본 명세서에 참고로 포함된, 문헌 [Jayaraman et al. Angew Chem Int Ed Engl 51(34):8529-8533 (2012)]에 교시된다.

[0761] CRISPR-Cas 시스템, 예를 들어 Cas9-gRNA RNP, gRNA, Cas9 mRNA의 전달에 최적화된 LNP 제형은, 둘 다 참고로 포함된, WO2019067992 및 WO2019067910에 기재되어 있다.

[0762] 핵산의 전달에 유용한 추가적인 특이적 LNP 제형은, 둘 다가 참고로 포함된 US8158601 및 US8168775에 기재되어 있으며, 이는 ONPATTRO라는 명칭 하에 판매되는, 파티시란(patisiran)에서 사용된 제형을 포함한다.

[0763] 유전자 변형 LNP의 예시적인 투약량은 약 0.1, 0.25, 0.3, 0.5, 1, 2, 3, 4, 5, 6, 8, 10, 또는 100 mg/kg(RNA)을 포함할 수 있다. 시스템의 하나 이상의 구성요소를 인코딩하는 핵산을 포함하는 AAV의 예시적인 투여량은 약 10^{11}, 10^{12}, 10^{13}, 및 10^{14} vg/kg의 MOI를 포함할 수 있다.

[0764] ***키트, 제조 물품, 및 약제학적 조성물***

[0765] 일 양태에서 본 개시내용은, 예를 들어 본 명세서에 기재된 바와 같은, 유전자 변형 폴리펩티드 또는 유전자 변형 시스템을 포함하는 키트를 제공한다. 일부 구현예에서, 키트는 유전자 변형 폴리펩티드(폴리펩티드를 인코딩하는 핵산) 및 주형 RNA(또는 상기 주형 RNA를 인코딩하는 DNA)를 포함한다. 일부 구현예에서, 키트는 시스템을 세포 내로 도입하기 위한 시약, 예를 들어 형질감염 시약, LNP 등을 추가로 포함한다. 일부 구현예에서, 키트는

본 명세서에 기재된 임의의 방법에 적합하다. 일부 구현예에서, 키트는, 예를 들어 제조 물품 내에 배치된, 하나 이상의 요소, 조성물(예를 들어, 약제학적 조성물), 유전자 변형 폴리펩티드, 및/또는 유전자 변형 시스템, 또는 이의 기능성 단편 또는 구성요소를 포함한다. 일부 구현예에서, 키트는 이의 사용을 위한 지침을 포함한다.

[0766] 일 양태에서, 본 개시내용은 제조 물품을 제공하며, 예를 들어 그 안에 본 명세서에 기재된 바와 같은 키트, 또는 이의 구성요소가 배치된다.

[0767] 일 양태에서 본 개시내용은, 예를 들어 본 명세서에 기재된 바와 같은, 유전자 변형 폴리펩티드 또는 유전자 변형 시스템을 포함하는 약제학적 조성물을 제공한다. 일부 구현예에서, 약제학적 조성물은 약제학적으로 허용 가능한 담체 또는 부형제를 추가로 포함한다. 일부 구현예에서, 약제학적 조성물은 주형 RNA 및/또는 폴리펩티드를 인코딩하는 RNA를 포함한다. 구현예에서, 약제학적 조성물은 다음 특성 중 하나 이상(예를 들어, 1, 2, 3, 또는 4개)을 갖는다:

[0768] (a) 예를 들어 몰 기준으로, 주형 RNA 및/또는 폴리펩티드를 인코딩하는 RNA에 대하여, 1% 미만(예를 들어, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만)의 DNA 주형;

[0769] (b) 예를 들어 몰 기준으로, 주형 RNA 및/또는 폴리펩티드를 인코딩하는 RNA에 대하여, 1% 미만(예를 들어, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만)의 캡핑되지 않은 RNA;

[0770] (c) 예를 들어 몰 기준으로, 주형 RNA 및/또는 폴리펩티드를 인코딩하는 RNA에 대하여, 1% 미만(예를 들어, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만)의 부분 길이 RNA;

[0771] (d) 반응되지 않은 캡 디뉴클레오티드가 실질적으로 결여됨.

[0772] ***화학, 제조, 및 제어(CMC)***

[0773] 단백질 치료제의 정제는, 예를 들어 문헌[Franks, Protein Biotechnology: *Isolation, Characterization, and Stabilization*, Humana Press (2013); 및 Cutler, *Protein Purification Protocols (Methods in Molecular Biology)*, Humana Press (2010)]에 기재되어 있다.

[0774] 일부 구현예에서, 유전자 변형 시스템, 폴리펩티드, 및/또는 주형 핵산(예를 들어, 주형 RNA)은 특정 품질 표준에 따른다. 일부 구현예에서, 본 명세서에 기재된 방법에 의해 생산된 유전자 변형 시스템, 폴리펩티드, 및/또는 주형 핵산(예를 들어, 주형 RNA)은 특정 품질 표준에 따른다. 따라서, 본 개시내용은, 일부 양태에서 특정 품질 표준에 따르는 유전자 변형 시스템, 폴리펩티드, 및/또는 주형 핵산(예를 들어, 주형 RNA)의 제조 방법에 관한 것이며, 예를 들어 여기서 상기 품질 표준이 분석된다. 본 개시내용은, 일부 양태에서 유전자 변형 시스템, 폴리펩티드, 및/또는 주형 핵산(예를 들어, 주형 RNA)에서 상기 품질 표준의 분석 방법에도 관한 것이다. 일부 구현예에서, 품질 표준은 다음 중 하나 이상(예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 또는 12개)을 포함하지만 이에 제한되지 않는다:

[0775] (i) 주형 RNA의 길이, 예를 들어 주형 RNA가 상기 참조 길이 초과 또는 참조 길이 범위 내의 길이를 갖는지 여부, 예를 들어 존재하는 주형 RNA의 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%는 100, 125, 150, 175, 또는 200개 뉴클레오티드 길이 초과인지 여부;

[0776] (ii) 주형 RNA 상에서 폴리A 테일의 존재, 부재, 및/또는 길이, 예를 들어 존재하는 주형 RNA의 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%가 폴리A 테일(예를 들어, 길이가 적어도 5, 10, 20, 30, 50, 70, 100개 뉴클레오티드인 폴리A 테일)을 함유하는지 여부;

[0777] (iii) 주형 RNA 상의 5' 캡의 존재, 부재, 및/또는 유형, 예를 들어 존재하는 주형 RNA의 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%가 5' 캡을 함유하는지 여부, 예를 들어 그 캡이 7-메틸구아노신 캡, 예를 들어 O-Me-m7G 캡인지 여부;

[0778] (iv) 주형 RNA에서 (예를 들어, 슈도우리딘, 디하이드로우리딘, 이노신, 7-메틸구아노신, 1-N-메틸슈도우리딘(1-Me-Ψ), 5-메톡시우리딘(5-MO-U), 5-메틸시티딘(5mC), 또는 잠금 뉴클레오티드로부터 선택되는) 하나 이상의 변형된 뉴클레오티드의 존재, 부재, 및/또는 유형, 예를 들어 존재하는 주형 RNA의 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%가 하나 이상의 변형된 뉴클레오티드를 함유하는지 여부;

[0779] (v) (예를 들어, 시간 경과에 따라 및/또는 미리 선택된 조건 하에서) 주형 RNA의 안정성, 예를 들어 안정성 시험 후 주형 RNA의 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%가 온전한 상태(예를 들어, 100, 125,

150, 175, 또는 200개 초과의 뉴클레오티드 길이)로 남아 있는지 여부;

[0780] (vi) DNA 변형 시스템에서 주형 RNA의 역가, 예를 들어 주형 RNA를 포함하는 시스템이 역가에 대해 분석된 후 표적 부위의 적어도 1%가 변형되는지 여부;

[0781] (vii) 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드의 길이, 예를 들어 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드가 참조 길이 초과 또는 참조 길이 범위 내의 길이를 갖는지 여부, 예를 들어 존재하는 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드의 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%가 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, 또는 2000개 아미노산 길이를 초과하는지(및 선택적으로, 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, 또는 600개 아미노산 길이 이하) 여부;

[0782] (viii) 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드 상의 번역후 변형의 존재, 부재 및/또는 유형, 예를 들어 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드의 적어도 80, 85, 90, 95, 96, 97, 98, 또는 99%가 포스포릴화, 메틸화, 아세틸화, 미리스토일화, 팔미토일화, 이소프레닐화, 글리피아티온 또는 리포일화 또는 이들의 임의의 조합을 함유하는지 여부;

[0783] (ix) 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드에서 하나 이상의 인공, 합성 또는 비-천연 아미노산(예를 들어, 오르니틴, β-알라닌, GABA, δ아미노레불린산, PABA, D-아미노산(예를 들어, D-알라닌 또는 D-글루타메이트), 아미노이소부티르산, 데하이드로알라닌, 시스타티온, 란티오닌, 드젠콜산, 디아미노피멜산, 호모알라닌, 노르발린, 노르류신, 호모노르류신, 호모세린, O-메틸-호모세린 및 O-에틸-호모세린, 에티오닌, 셀레노시스테인, 셀레노호모시스테인, 셀레노메티오닌, 셀레노에티오닌, 텔루로시스테인, 또는 텔루로메티오닌으로부터 선택됨)의 존재, 부재, 및/또는 유형, 예를 들어 존재하는 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드의 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%가 하나 이상의 인공, 합성, 또는 비-정규 아미노산을 함유하는지 여부;

[0784] (x) (예를 들어, 시간 경과에 따라 및/또는 미리 선택된 조건 하에서) 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드의 안정성, 예를 들어 안정성 시험 후 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드의 적어도 80%, 85%, 90%, 95%, 96%, 97%, 98%, 또는 99%가 온전한 상태(예를 들어, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1600, 1700, 1800, 1900, 또는 2000개 초과의 아미노산 길이(및 선택적으로, 2500, 2000, 1500, 1400, 1300, 1200, 1100, 1000, 900, 800, 700, 또는 600개 이하의 아미노산 길이))로 남아 있는지 여부;

[0785] (xi) DNA 변형 시스템에서 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드의 역가, 예를 들어 폴리펩티드, 제1 폴리펩티드, 또는 제2 폴리펩티드를 포함하는 시스템이 역가에 대해 분석된 후 표적 부위의 적어도 1%가 변형되는지 여부; 또는

[0786] (xii) 하나 이상의 발열원, 바이러스, 진균류, 박테리아 병원체, 또는 숙주 세포 단백질의 존재, 부재, 및/또는 수준, 예를 들어 시스템에 발열원, 바이러스, 진균류, 박테리아 병원체, 또는 숙주 세포 단백질 오염이 없거나 실질적으로 없는지 여부.

[0787] 일부 구현예에서, 본 명세서에 기재된 시스템 또는 약제학적 조성물에는 내독소가 없다.

[0788] 일부 구현예에서, 하나 이상의 발열원, 바이러스, 진균류, 박테리아 병원체, 및/또는 숙주 세포 단백질의 존재, 부재, 및/또는 수준이 결정된다. 구현예에서, 시스템에 발열원, 바이러스, 진균류, 박테리아 병원체, 및/또는 숙주 세포 단백질 오염이 없거나 실질적으로 없는지 여부가 결정된다.

[0789] 일부 구현예에서, 본 명세서에 기재된 바와 같은 약제학적 조성물 또는 시스템은 다음 특성 중 하나 이상(예를 들어, 1, 2, 3, 또는 4개)을 갖는다:

[0790] (a) 예를 들어 몰 기준으로, 주형 RNA 및/또는 폴리펩티드를 인코딩하는 RNA에 대하여, 1% 미만(예를 들어, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만)의 DNA 주형;

[0791] (b) 예를 들어 몰 기준으로, 주형 RNA 및/또는 폴리펩티드를 인코딩하는 RNA에 대하여, 1% 미만(예를 들어, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만)의 캡핑되지 않은 RNA;

[0792] (c) 예를 들어 몰 기준으로, 주형 RNA 및/또는 폴리펩티드를 인코딩하는 RNA에 대하여, 1% 미만(예를 들어, 0.5%, 0.4%, 0.3%, 0.2%, 또는 0.1% 미만)의 부분 길이 RNA;

[0793] (d) 반응되지 않은 캡 디뉴클레오티드가 실질적으로 결여됨.

[0794] **실시예**

[0795] 본 발명은 하기 실시예에 의해 추가로 설명된다. 실시예는 예시적인 목적으로만 제공되며 어떤 식으로든 본 발명의 범주 또는 내용을 제한하는 것으로 해석되어서는 안 된다.

[0796] **실시예 1: 인간 세포에서 정확한 편집을 가능하게 하기 위해 다양한 Pol 도메인을 갖는 Cas-Pol 융합체를 포함하는 유전자 변형 시스템**

[0797] 이 실시예는 인간 세포의 게놈 DNA에 돌연변이를 프로그래밍 가능한 방식으로 설치하는 Cas-Pol 융합체의 능력을 설명한다. 더 구체적으로, Cas-Pol 융합체의 중합효소 도메인, 예를 들어 본 출원에 기재된 중합효소 도메인, 예를 들어 인간 DNA Polθ는 신규 중합효소 조합을 이용하는 Cas-Pol 융합체의 게놈 편집 능력을 결정하기 위해 다양하다. 주형 핵산은 Cas-Pol 융합체의 표적 편집 활성을 결정하기 위해 인간 세포에서 Cas-Pol 발현 플라스미드와 함께 공동 전달된다.

[0798] 게놈 조작 폴리펩티드에 대한 도메인 라이브러리를 생성하기 위해, Cas 이펙터 단백질을 선택하며; 표 3 및 표 4를 참조한다. 표적 서열의 PAM 요건, 닉카제 활성을 부여하기 위한 예측된 돌연변이(예를 들어, SpCas9의 경우 D10A, H840A, 또는 N863A)를 포함하는 특징부로서 본 명세서에 기재된 게놈 조작 폴리펩티드에 사용하기 위해 추가의 Cas9 도메인을 추가로 선택하며, 단일-가이드 구성을 포함한 gRNA 특징부, 예를 들어 동족 Cas 효소에 대하여 폴리펩티드 결합을 부여하기 위한 특정 스페이서 매개변수 및 gRNA 스캐폴드 서열을 결정할 수 있다(표 8). 융합 단백질을 최적화하기 위해 길이, 가요성 및 구성의 다양성에 대한 검색에 기반하여 Cas 및 Pol 도메인을 연결하는 링커 서열을 수집한다(표 6). 중합효소 도메인은 RNA 및/또는 DNA-의존성 DNA 중합 활성이 가능한 야생형 중합효소, 개선된 특성(예를 들어, 열안정성, 가공성, 정확도)을 갖는 유도된 중합효소, 비활성화되거나 재활성화된 기능성 도메인(예를 들어, 5'-3' 또는 3'-5' 엑소뉴클레아제 활성, 교정 활성, 헬리카제 활성, 또는 RNase 활성을 부여하는 비활성화되거나 재활성화된 도메인)을 갖는 유도된 중합효소, 합성적으로 변화된 RNA-의존성 및/또는 DNA-의존성 DNA 중합효소 활성을 갖는 중합효소(예를 들어, 그 전체가 본 명세서에 참고로 포함된 문헌[Ellefson et al Science 352(6293):1590-1593 (2016)]에 기재된 바와 같은, 써모코커스 코다카렌시스(Thermococcus kodakarensis)(KOD)의 PolB로부터 유래된 RTX 중합효소)를 포함하여, 본 출원에 기재된 바와 같은 문헌 및 중합효소 단백질 특징을 사용하여 다양한 공급원으로부터 발견될 수 있다.

[0799] 구체적으로, 게놈을 성공적으로 편집하기 위한 유전자 변형 폴리펩티드의 맥락에서 신규 Pol 도메인의 사용을 평가하기 위해, Cas9(N863A) 닉카제에 대한 융합을 위해 예시적인 Pol 도메인의 서브세트를 선택한다. 간략하게, 단량체성 인간 중합효소의 단백질 서열은 UniProt(The UniProt Consortium Nucleic Acids Res 49(D1):D480-D489 (2021))을 사용하여 결정되고 기능성 도메인은 InterPro(Blum et al Nucleic Acids Res 49(D1):D344-D354 (2021)) 및 InterProScan(Jones et al Bioinformatics 30(9):1236-1240 (2014))(여기서 표 X)을 사용하여 추가로 예측되고 참조된다. 그러한 예에 의해 제한되고자 하지 않지만, 단일 서브유닛을 포함하는 중합효소 패밀리 Y, X, 및 A 유래의 단백질이 기재되었다(예를 들어, 그 전체가 본 명세서에 참고로 포함된 문헌[Hoitsma et al Cell Mol Life Sci 77(1):35-59 (2020)] 참조).

[0800] 게놈 조작 시스템 Cas-Pol 융합체를 사용하여 정확한 편집을 생성하기 위해, 주형 핵산을 구축하여 Pol 도메인에 의해 게놈 표적 부위로 편집물 중합의 주형으로 사용한다. 주형 핵산은 (i) Cas-Pol을 표적 영역으로 안내하기 위한 gRNA 스페이서 서열, 예를 들어 HEK3 유전자좌의 20개 뉴클레오티드 서열에 상보적인 서열; (ii) 표적-프라이밍된 중합을 위해 닉이 생성된 DNA의 단일 가닥과 상보적인 염기쌍을 형성할 수 있는 프라이머-결합 서열(PBS); (iii) 의도된 최종 표적 서열을 추가로 포함하는 중합을 위한 주형을 제공하는 이종 목적 서열; 및 (iv) Cas9-Pol 폴리펩티드 융합체의 Cas9 도메인과 회합하는 gRNA 스캐폴드 서열을 포함하도록 설계된다. 여기에 이용된 작제물은 구체적으로 5'에서 3' 방향으로 (i), (iv), (iii), (ii)를 따른다. 일부 구현예에서, (iii)은 RNA 및/또는 DNA 뉴클레오티드를 포함할 수 있다. 일부 구현예에서, (ii)는 RNA 및/또는 DNA 뉴클레오티드를 포함할 수 있다. 예에 의해 제한되고자 하지 않지만, (i) 및 (iv)는 이들 실험에서 RNA 뉴클레오티드를 포함한다. 주형 조성물은 표 21(주형 P1, P2, P3)에 기재되어 있으며, 여기서 (ii) 및 (iii)은 각각 RNA 또는 DNA 뉴클레오티드로서 포함될 수 있다. 주형 분자는 선택적으로 5' 캡 및 3' 폴리A 테일을 추가로 포함한다.

[0801] **표 20: 실시예 1에서 사용된 주형 핵산 및 제2-닉 gRNA**

표 20

[0802]

명칭	설명	스페이서 (i)	스캐폴드 (iv)	Pol 주형 (iii)	PBS (ii)	전체 주형 분자
주형 P1	HEK3_8PBS_10RT(CTTat1)	GGCCCAGACTGAGCACGTGA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	TCTGCCATCAAAG	CGTGCTCA	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCCATCAAAGCGTGCTCA
주형 P2	HEK3_13PBS_10RT(CTTat1)	GGCCCAGACTGAGCACGTGA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	TCTGCCATCAAAG	CGTGCTCAGTCTG	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCCATCAAAGCGTGCTCAGTCTG
주형 P3	HEK3_17PBS_10RT(CTTat1)	GGCCCAGACTGAGCACGTGA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	TCTGCCATCAAAG	CGTGCTCAGTCTGGGCC	GGCCCAGACTGAGCACGTGAGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTCTGCCATCAAAGCGTGCTCAGTCTGGGCC
주형 P4	HBB_13PBS_10RT(TtoAat4)	GCATGGTGCACCTGACTCCTG	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	AGACTTCTCCACAG	GAGTCAGGTGCAC	GCATGGTGCACCTGACTCCTGGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCAGACTTCTCCACAGGAGTCAGGTGCAC
2gRNA P5	HEK3_+90	GTCAACCAGTATCCCGGTGC	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	NA	NA	NA
2gRNA P6	HBB_+72	GCCTTGATACCAACCTGCCCA	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	NA	NA	NA

[0803] 약 800 ng의 Cas9-Pol 융합체(예를 들어, pol 세타 융합체) 발현 플라스미드, 200 ng의 화학적으로 합성된 주형 핵산 분자, 및 선택적으로 83 ng의 추가 제2-닉 gRNA(주형 P1, P2, P3의 경우 2gRNA P5 또는 주형 P4의 경우 2gRNA P6)를 이용하여 250,000개 세포/웰을 전기천공함으로써 U2OS 또는 HEK293T 세포를 형질감염시킨다(표 21). Cas-Pol 융합체의 게놈 편집 능력을 평가하기 위해, 게놈 DNA(gDNA)를 형질감염 후 3일째에 수집한다. 표적 유전자좌(주형 P1, P2, P3의 경우 HEK3 또는 주형 P4의 경우 HBB)에서 의도된(설계된 바와 같이 정확하고 흉터 없는 편집) 및 의도하지 않은(표적 서열에 대한 임의의 의도하지 않은 변경) 편집의 빈도를 암플리콘 시퀀싱으로 분석한다. 본 명세서에 사용된 바와 같이, 표적 부위의 암플리콘 시퀀싱은 표적 부위의 PCR 증폭에서 부위-특이적 프라이머의 사용, Illumina MiSeq에서의 암플리콘 시퀀싱, 및 CRISPResso2 파이프라인을 사용한 편집 사건의 검출 및 특성화를 포함한다(Clement et al Nat Biotechnol 37(3):224-226 (2019)). 일부 구현예에서, 활성 Cas-Pol 융합체는 검출가능한 수준의 편집을 초래하며, 예를 들어 시퀀싱 판독물의 적어도 0.1%가 표적 부위 편집을 입증한다. 일부 구현예에서, 바람직한 Cas-Pol 융합체는 의도하지 않은 편집에 대한 의도된 편집의 더 높은 빈도, 예를 들어 의도하지 않은 편집에 대한 의도된 편집의 적어도 2배 더 높은 빈도를 나타낸다.

[0804] **실시예 2: 링커 선택을 통한 Cas-Pol 융합체의 발현 개선**

[0805] 이 실시예는 포유동물 세포에서 단백질 발현을 개선하기 위한 Cas-Pol 융합체의 최적화를 설명한다. 상기 실시예 1에 기재된 바와 같이 새로운 기능적 도메인의 치환에 의한 신규 Cas-Pol 융합체의 구축은 게놈 조작 폴리펩티드의 낮거나 중간 정도의 발현을 초래할 수 있다. 따라서, 융합체의 변형된 구성이 상이한 도메인의 맥락에서 유리할 수 있다는 것이 여기서 고려된다. 예에 의해 제한되고자 하지 않지만, 새로운 융합체의 발현 및 안정성을 개선하기 위한 그러한 접근법 중 하나는 링커 라이브러리의 사용을 통한다. 여기서 Cas-Pol 융합체의 Cas와 Pol 도메인 사이의 펩티드 링커 서열은 링커 서열 라이브러리를 사용하여 다양해진다. 더 구체적으로, 하기 표 21의 링커를 사용하여 Cas9 융합 작제물의 새로운 변이체를 생성하고 개선된 Cas-Pol 단백질 발현을 스크리닝하기 위해 인간 세포에 전달한다.

[0806] 다양한 정도의 길이, 가요성, 소수성, 및 2차 구조를 갖는 22개의 펩티드 링커 세트(표 21)를 먼저 사용하여 원래 링커를 대체하여 Cas-Pol 융합 단백질의 변이체를 생성하였다(상기 참조된 실시예 30 참조). 200 ng의 단일-가이드 RNA 플라스미드와 함께 약 800 ng의 각 Cas9-Pol 융합 플라스미드를 사용하여 250,000개 세포/웰을 전기천공함으로써 HEK293T 세포를 형질감염시킨다. Cas9-Pol 융합체의 발현 수준을 평가하기 위해, 형질감염 후 2일째에 세포 용해물을 수집하고 Cas9에 대한 1차 항체를 사용하여 웨스턴 블롯으로 분석한다.

[0807] **표 21. Cas-Pol 융합체를 포함하는 게놈 조작 폴리펩티드에서 Cas와 Pol 도메인 사이의 링커로 사용하기 위한 펩티드 서열**

표 21

[0808]

#	링커 서열	비고
1	GGS	짧음
2	GGGGS	가요성, 짧음
3	(GGGGS)2	가요성
4	(GGGGS)3	가요성, 김
5	(GGGGS)4	가요성, 매우 김
6	(G)6	가요성
7	(G)8	가요성
8	GSAGSAAGSGEF	가요성
9	(GSSGSS)	중간
10	(GSSGSS)2	중간, 가요성
11	(GSSGSS)3	중간
12	SGSETPGTSESATPES	XTEN
13	(EAAAK)	강성 나선, 짧음
14	(EAAAK)2	강성 나선, 중간
15	(EAAAK)3	강성 나선, 김
16	PAP	강성, 짧음
17	PAPAP	강성, 짧음
18	PAPAPAPAP	강성, 중간
19	A(EAAAK)4ALEA(EAAAK)4A	강성, 나선형으로 매우 김
20	GGGGS(EAAAK)GGGGS	가요성 - 나선 - 가요성
21	(EAAAK)GGGGS(EAAAK)	나선 - 가요성 - 나선
22	SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	가요성 - XTEN - 가요성

[0809] **실시예 3: HEK293T & U2OS 세포에서 풀링된 스크리닝에 의한 유전자 변형 폴리펩티드 선택**

[0810] 이 실시예는 인간 게놈 내 코딩 서열의 표적화된 편집을 위한 RNA 유전자 변형 시스템의 사용을 설명한다. 더 구체적으로, 이 실시예는, 예를 들어 풀링된 스크리닝 접근법에 의해 인간 세포에서 활성을 편집하기 위해 새로운 유전자 변형 폴리펩티드를 평가하는 수단으로서, 유전자 변형 후보의 라이브러리로 HEK293T 및 U2OS 세포를 감염시킨 후 세포에서 시험관 내 유전자 변형을 위한 주형 가이드 RNA(tgRNA)를 형질감염시키는 것을 설명한다.

[0811] 본 명세서에서 분석된 유전자 변형 폴리펩티드 라이브러리 후보는 각각 다음을 포함한다: 1) 하나의 엔도뉴클레아제 활성 부위를 비활성화시키는 N863A 돌연변이를 함유하는 S. 피오게네스(Spy) Cas9 닉카제; 2) **표 6**에 기재된 122개의 펩티드 링커 중 하나; 및 3) 레트로바이러스 기원의 **표 2**의 역전사효소(RT) 도메인. 이용된 특정 레

트로바이러스 RT 도메인이 단량체로 기능할 것으로 예상되는 경우 이를 선택하였다. 각각의 선택된 RT 도메인에 대해, 야생형 서열뿐만 아니라, 주요 야생형 서열에 점 돌연변이가 설치된 버전을 시험하였다. 특히, 야생형이거나 다양한 돌연변이를 함유하는 143개의 RT 도메인을 시험하였다. 총 17,446개의 Cas-링커-RT 유전자 변형 폴리펩티드를 시험하였다.

[0812] 본 명세서에 기재된 시스템은 다음을 포함하는 2-구성요소 시스템이다: 1) 렌티바이러스 카세트 내에 인간 코돈-최적화 유전자 변형 폴리펩티드 라이브러리 후보를 인코딩하는 발현 플라스미드, 및 2) Cas에 의해 인식되고 관심 게놈 유전자좌로 국재화하며, 또한 U6 프로모터에 의해 구동되는 RT 도메인에 의해 원하는 편집물을 게놈으로 역전사하는 주형을 제공하는 비-코딩 tgRNA 서열을 발현하는 tgRNA 발현 플라스미드. 렌티바이러스 카세트는 (i) 포유동물 세포에서의 발현을 위한 CMV 프로모터; (ii) 나타낸 바와 같은 유전자 변형 폴리펩티드 라이브러리 후보; (iii) 자가-절단 T2A 폴리펩티드; (iv) 포유동물 세포에서 선택을 가능하게 하는 퓨로마이신 내성 유전자; 및 (v) 폴리A 테일 종결 신호를 포함한다.

[0813] 유전자 변형 폴리펩티드 라이브러리 후보를 발현하는 세포의 풀을 준비하기 위해, HEK293T 또는 U2OS 세포를 유전자 변형 후보 플라스미드 라이브러리의 풀링된 렌티바이러스 제제를 이용하여 형질도입시켰다. HEK293 Lenti-X 세포를 렌티바이러스 플라스미드 형질감염 전에 15 cm 플레이트(약 $12x10^6$개 세포)에 시딩하였다. Lentiviral Packaging Mix(Biosettia, 27 ug) 및 유전자 변형 후보 라이브러리에 대한 플라스미드 DNA(27 ug)를 사용한 렌티바이러스 플라스미드 형질감염을 제조업체의 프로토콜에 따라 Lipofectamine 2000 및 Opti-MEM 배지를 사용하여 다음 날 수행하였다. 다음 날 전체 배지 교체에 의해 세포외 DNA를 제거하였고 바이러스-함유 배지를 48시간 후에 수확하였다. 렌티바이러스 배지를 Lenti-X Concentrator(TaKaRa Biosciences)를 사용하여 농축시켰고 5 mL의 렌티베아러스 분취액을 제조하였으며 -80° C에서 보관하였다. 렌티바이러스 적정은 퓨로마이신 선택 후 콜로니 형성 단위를 계수함으로써 수행하였다. BFP-발현 게놈 랜딩 패드를 보유하는 HEK293T 또는 U2OS 세포를 배양 플레이트에 $6x10^7$개 세포로 시딩하였고 0.3의 감염다중도(MOI)로 형질도입하여 세포당 다중 감염을 최소화하였다. 감염된 세포의 선택을 가능하게 하기 위해 감염 후 48시간 후에 퓨로마이신(2.5 ug/mL)을 첨가하였다. 세포를 적어도 7일 동안 퓨로마이신 선택 하에 유지시킨 다음, tgRNA 전기천공을 위해 규모를 확대시켰다.

[0814] 분석에서 유전자 변형 라이브러리 후보의 게놈-편집 능력을 결정하기 위해, 그 다음 감염된 BFP-발현 HEK293T 또는 U2OS 세포를 라이브러리 후보당 >1000x 커버리지에 대해 충분한 세포 수에서 BFP를 GFP로 전환시키도록 설계된, 200 ng의 tgRNA(g4 또는 g10) 플라스미드를 이용하여 250,000개 세포/웰을 전기천공함으로써 형질감염시켰다.

[0815] g4 tgRNA(5'에서 3')는 다음과 같다: 20개 뉴클레오티드 스페이서 영역(GCCGAAGCACTGCACGCCGT), 스캐폴드 영역(GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC), BFP를 GFP로 변형시키는 단일 염기 치환을 인코딩하는 주형 영역(**굵은 글씨**) 및 기능성 유전자 변형 반응의 완료 시 유전자 변형 폴리펩티드의 재결합을 방지하는 SpyCas9 PAM에 동의어 점 돌연변이(NGG에서 NCG)를 도입한 PAM 비활성화(밑줄)(ACCCTGAC**G**T**T**ACG), 및 13 뉴클레오티드 PBS(GCGTGCAGTGCTT).

[0816] 유사하게, g10 tgRNA(5'에서 3')는 다음과 같다: 20개 뉴클레오티드 스페이서 영역(AGAAGTCGTGCTGCTTCATG), 스캐폴드 영역(GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC), BFP를 GFP로 변형시키는 단일 염기 치환을 인코딩하는 주형 영역(**굵은 글씨**) 및 기능성 유전자 변형 반응의 완료 시 유전자 변형 폴리펩티드의 재결합을 방지하는 SpyCas9 PAM에 동의어 점 돌연변이(NGG에서 NGA)를 도입한 PAM 비활성화(밑줄)(ACCCTGACC<u>T</u>ACGGCGTGCAGTGCTTCGGCCGCTACCCCGA**T**CACAT), 및 13 뉴클레오티드 PBS(GAAGCAGCACGAC).

[0817] 분석에서 다양한 작제물의 게놈 편집 능력을 분석하기 위해, 전기천공 후 6 내지 7일 후에 GFP 발현을 위해 형광 활성화 세포 분류(FACS)에 의해 세포를 분류하였다. 세포를 편집되지 않은(BFP+) 세포, 편집된 (GFP+) 세포 및 불완전한 편집(BFP-, GFP-) 세포의 별개의 집단으로 분류하고 수확하였다. 분석 동안에 농축을 결정하기 위해 분류되지 않은 세포 샘플을 또한 입력 집단으로 수확하였다.

[0818] 어떤 유전자 변형 라이브러리 후보가 분석에서 게놈 편집 능력을 갖는지를 결정하기 위해, 분류된 세포 집단과 분류되지 않은 세포 집단에서 게놈 DNA(gDNA)를 수확하고, 각각의 집단에서 유전자 변형 라이브러리 후보를 시퀀싱함으로써 분석하였다. 간략하게, 유전자 변형 서열을 렌티바이러스 카세트에 특이적인 프라이머를 사용하여 게놈으로부터 증폭시켰고, 게놈 DNA를 희석하기 위해 PCR의 두 번째 라운드에서 증폭시킨 다음, 제조업체의 프로토콜에 따라 Oxford Nanopore Sequencing Technology를 사용하여 시퀀싱하였다.

[0819] 시퀀싱 판독물의 품질 관리 후, 적어도 1500개 및 3200개 이하의 뉴클레오티드의 판독물을 유전자 변형 폴리펩

티드 라이브러리 서열에 맵핑하였고 라이브러리 서열과 최소 약 80% 일치하는 판독물을 주어진 후보에 대해 성공적으로 정렬된 것으로 간주하였다. 분석에서 유전자 편집을 수행할 수 있는 유전자 변형 후보를 확인하기 위해, 편집된 집단에 있는 각각의 라이브러리 후보의 판독 수를 초기의 분류되지 않은 집단에서 판독 수와 비교하였다. 이러한 풀링된 스크린의 목적을 위해, 게놈-편집 능력이 있는 유전자 변형 후보를 분류되지 않은(입력) 세포에 비해 전환된(GFP+) 집단이 농축된 후보로 선택하였으며, 여기서 농축은 참조(요소 ID 번호: 17380)의 농축 수준 이상인 것으로 결정되었다.

[0820] 다수의 유전자 변형 폴리펩티드 후보는 GFP+ 세포 집단이 농축된 것으로 결정되었다. 예를 들어, 시험한 17,446개 후보 중, 3,300개 초과가 표 D에 나타낸 유사한 실험 조건(g4 tgRNA를 사용한 HEK293T; g10 tgRNA를 사용한 HEK293T 세포; 또는 g4 tgRNA를 사용한 U2OS 세포) 하에서 참조와 적어도 동등한 GFP+ 분류 집단(분류되지 않은 집단에 비해)의 농축을 나타내었다. 17,446개의 후보를 또한 g10 tgRNA를 사용한 U2OS 세포에서 시험하였지만, 풀링된 스크린은 실험 조건 하에서 분류되지 않은(입력) 세포에 비해 전환된(GFP+) 집단이 농축된 후보를 산출하지 않았으며; 이러한 결과를 설명하기 위해 추가 조사가 필요하다.

[0821] **표 D. 스크리닝된 링커와 RT 서열의 조합. 이 표의 각각의 RT의 아미노산 서열은 표 6에 제공되어 있다.**

[0822] [표 D]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
PAPEAAAK	MLVFF_P26809_3mutA
EAAAKPAPGGG	MLVFF_P26809_3mutA
GSSGSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mut
PAPGGGEAAAK	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355_PLV919
GSSEAAAK	MLVFF_P26809_3mutA
EAAAKPAPGGS	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	XMRV6_A1Z651_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	MLVFF_P26809_3mutA
PAPEAAAKGSS	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	AVIRE_P03360_3mutA
PAPAPAPAPAP	MLVCB_P08361_3mutA
PAPAPAPAPAP	MLVFF_P26809_3mutA
EAAAKGGSPAP	PERV_Q4VFZ2_3mutA_WS
PAP	MLVMS_P03355_PLV919
PAPGGGGSS	WMSV_P03359_3mutA
SGSETPGTSESATPES	MLVFF_P26809_3mutA
PAPEAAAKGSS	XMRV6_A1Z651_3mutA
EAAAKGGSGGG	MLVMS_P03355_PLV919
GGGGSGGGGS	MLVFF_P26809_3mutA
GGGPAPGSS	MLVAV_P03356_3mutA
GGSGGSGGSGGSGGSGGS	XMRV6_A1Z651_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVCB_P08361_3mutA
GSSPAP	AVIRE_P03360_3mutA

[0823]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGSSPAP	MLVFF_P26809_3mutA
GSSGGGEAAAK	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	MLVMS_P03355_3mutA_WS
PAPAPAPAP	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	XMRV6_A1Z651_3mutA
EAAAKGGSPAP	MLVMS_P03355_3mutA_WS
PAPGGSEAAAK	AVIRE_P03360_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	AVIRE_P03360_3mutA
EAAAKGGGGSEAAAK	MLVCB_P08361_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	WMSV_P03359_3mutA
GSS	MLVMS_P03355_PLV919
GSSGSSGSSGSS	MLVMS_P03355_PLV919
GSSPAPEAAAK	XMRV6_A1Z651_3mutA
GGSPAPEAAAK	MLVFF_P26809_3mutA
GGGEAAAKGGS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGGGGG	PERV_Q4VFZ2_3mut
GGGPAP	MLVCB_P08361_3mutA
PAPAPAPAPAP	MLVCB_P08361_3mutA
GGSGGSGGSGGSGGSGGS	MLVCB_P08361_3mutA
PAP	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAP	MLVMS_P03355_PLV919
EAAAKPAPGSS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
EAAAKGGS	MLVMS_P03355_3mutA_WS
GGGGSEAAAKGGGGS	MLVFF_P26809_3mutA
EAAAKPAPGSS	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
EAAAKGGGGGS	MLVMS_P03355_PLV919
GGSPAP	XMRV6_A1Z651_3mutA
EAAAKGGGPAP	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mutA
PAP	MLVCB_P08361_3mutA
EAAAK	XMRV6_A1Z651_3mutA
GGSGSSPAP	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_PLV919
GSSEAAAKGGG	MLVAV_P03356_3mutA
GGGEAAAKGGS	XMRV6_A1Z651_3mutA
EAAAKGGGGSEAAAK	MLVAV_P03356_3mutA

[0824]

링커 아미노산 서열	RT 도메인 명칭
GGGGSGGGGSGGGGS	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	AVIRE_P03360_3mutA
SGSETPGTSESATPES	AVIRE_P03360_3mutA
GGGEAAAKPAP	MLVFF_P26809_3mutA
EAAAKGSSGGG	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mut
GGSGGSGGSGGS	XMRV6_A1Z651_3mutA
GGSEAAAKPAP	MLVFF_P26809_3mutA
EAAAKGSSGGG	XMRV6_A1Z651_3mutA
GGGGS	MLVFF_P26809_3mutA
GGGEAAAKGSS	MLVMS_P03355_PLV919
PAPAPAPAPAP	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGSGGGGS	MLVCB_P08361_3mutA
GGGEAAAKGSS	MLVCB_P08361_3mutA
PAPGGSGSS	MLVFF_P26809_3mutA
GSAGSAAGSGEF	MLVCB_P08361_3mutA
PAPGGSEAAAK	MLVMS_P03355_3mutA_WS
GGSGSS	XMRV6_A1Z651_3mutA
PAPGGGGSS	MLVMS_P03355_PLV919
GSSGSSGSS	XMRV6_A1Z651_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355_3mutA_WS
EAAAK	MLVMS_P03355_PLV919
GSSGSSGSSGSS	MLVFF_P26809_3mutA
PAPGGGGSS	MLVCB_P08361_3mutA
GGGEAAAKGGS	MLVCB_P08361_3mutA
PAPGGGEAAAK	MLVMS_P03355_PLV919
GGGGGSPAP	XMRV6_A1Z651_3mutA
EAAAKGGS	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	XMRV6_A1Z651_3mut
PAPEAAAK	MLVAV_P03356_3mutA
GGSGGSGGSGGS	MLVMS_P03355_3mutA_WS
GGGPAPGGS	MLVMS_P03355_PLV919
GSSGSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGS	MLVCB_P08361_3mutA
GSSGSS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	FLV_P10273_3mutA
GSS	MLVFF_P26809_3mutA
EAAAKEAAAK	MLVMS_P03355_3mutA_WS
PAPEAAAKGGG	MLVAV_P03356_3mutA

[0825]

링커 아미노산 서열	RT 도메인 명칭
GGSGSSEAAAK	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2
GSSEAAAKPAP	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mutA
EAAAKGGG	MLVFF_P26809_3mutA
GSSPAPGGG	MLVCB_P08361_3mutA
GGGPAPGSS	MLVMS_P03355_PLV919
GGGGGS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	WMSV_P03359_3mutA
EAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
PAPAPAPAP	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	MLVCB_P08361_3mutA
PAPEAAAKGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mutA
EAAAKGGGGSEAAAK	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	MLVMS_P03355_3mutA_WS
EAAAKGGGPAP	XMRV6_A1Z651_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FLV_P10273_3mutA
GGSEAAAKGGG	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	KORV_Q9TTC1-Pro_3mutA
GGGPAPGGS	MLVCB_P08361_3mutA
PAPAPAPAPAP	XMRV6_A1Z651_3mutA
GGSGSSGGG	XMRV6_A1Z651_3mutA
GGSGSSGGG	MLVCB_P08361_3mutA
GGGEAAAKGGS	MLVMS_P03355_3mutA_WS
EAAAK	MLVCB_P08361_3mutA
GGSPAPGSS	MLVMS_P03355_3mutA_WS
GGGGSSEAAAK	PERV_Q4VFZ2_3mut
PAPAPAPAP	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAK	MLVAV_P03356_3mutA
GGGGGSGSS	MLVCB_P08361_3mutA
EAAAKGSSPAP	MLVMS_P03355_3mutA_WS
PAPAPAPAPAP	MLVMS_P03355_3mutA_WS
GSSGGGGGS	MLVMS_P03355_3mutA_WS
PAPGSSGGG	MLVMS_P03355_PLV919
GGSGGGPAP	MLVCB_P08361_3mutA

[0826]

[0827]

링커 아미노산 서열	RT 도메인 명칭
GGGGGGG	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	MLVCB_P08361_3mutA
GGGPAPGGS	MLVFF_P26809_3mutA
EAAAKGGSGGG	PERV_Q4VFZ2_3mut
EAAAKGGGSS	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_3mut
GGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAPAP	MLVMS_P03355_PLV919
GGGEAAAKGGS	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355_3mut
GSAGSAAGSGEF	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	MLVFF_P26809_3mutA
EAAAKGGSGSS	MLVFF_P26809_3mutA
PAPGGG	MLVFF_P26809_3mutA
GGGPAPGSS	XMRV6_A1Z651_3mutA
PAPEAAAKGGS	AVIRE_P03360_3mutA
PAPGGGEAAAK	MLVFF_P26809_3mut
GGGGSSEAAAK	MLVCB_P08361_3mutA
EAAAK	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mutA
GSSGGGEAAAK	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVFF_P26809_3mutA
GSSPAPGGS	XMRV6_A1Z651_3mutA
GGSGGGPAP	MLVMS_P03355_PLV919
EAAAK	AVIRE_P03360_3mutA
GSS	XMRV6_A1Z651_3mutA
GGSGGSGGS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	AVIRE_P03360_3mut
PAPEAAAKGGG	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	BAEVM_P10272_3mutA
GGSGSSGGG	MLVMS_P03355_3mutA_WS
GGGGGGG	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	WMSV_P03359_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVFF_P26809_3mut
GGGEAAAKGGS	AVIRE_P03360_3mutA
GGSPAPGGG	AVIRE_P03360_3mutA
GSAGSAAGSGEF	MLVAV_P03356_3mutA
EAAAK	MLVAV_P03356_3mutA
EAAAKPAPGSS	WMSV_P03359_3mutA

[0828]

링커 아미노산 서열	RT 도메인 명칭
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	MLVCB_P08361_3mutA
PAPAPAPAPAP	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	MLVMS_P03355_3mut
GGSGGSGGSGGS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mutA
GGG	AVIRE_P03360_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2_3mut
GGSGGSGGSGGS	MLVMS_P03355_3mutA_WS
GGGEAAAK	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GSSGGGPAP	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	MLVFF_P26809_3mutA
EAAAKEAAAK	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVCB_P08361_3mut
GGGGGG	MLVMS_P03355_3mutA_WS
GGSGSSGGG	MLVFF_P26809_3mutA
GSSGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAP	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401_2mut
EAAAKGGSGGG	BAEVM_P10272_3mutA
GGGGSSPAP	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKPAP	MLVMS_P03355_PLV919
GGSGGGPAP	BAEVM_P10272_3mutA
PAPGSSGGS	MLVMS_P03355_PLV919
GGSGGGPAP	MLVMS_P03355_3mutA_WS
EAAAKGGSPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGGG	MLVMS_P03355_3mutA_WS
PAPGSSGGG	MLVFF_P26809_3mutA
GSSEAAAKGGS	MLVFF_P26809_3mutA
PAPGSSEAAAK	MLVFF_P26809_3mutA
EAAAKGSSPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPGSSEAAAK	MLVMS_P03355_PLV919
EAAAKGSSGGG	MLVMS_P03355_3mutA_WS
EAAAKGGGGS	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
PAPAPAPAPAP	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGS	MLVCB_P08361_3mutA

링커 아미노산 서열	RT 도메인 명칭
PAPGGSEAAAK	MLVCB_P08361_3mutA
PAPGSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGSS	AVIRE_P03360_3mutA
GGSPAPGSS	WMSV_P03359_3mutA
PAPGGSGGG	MLVMS_P03355_PLV919
EAAAKGGSGSS	MLVMS_P03355_3mutA_WS
GGSGGG	MLVFF_P26809_3mutA
GGSEAAAKGSS	KORV_Q9TTC1_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVCB_P08361_3mutA
PAPAPAPAPAP	PERV_Q4VFZ2_3mutA_WS
PAPEAAAK	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	MLVMS_P03355_PLV919
GSSPAP	MLVMS_P03355_3mutA_WS
GGGGSS	MLVMS_P03355_PLV919
GGGEAAAKPAP	AVIRE_P03360_3mutA
EAAAKPAPGGS	MLVAV_P03356_3mutA
EAAAKGGGPAP	MLVAV_P03356_3mutA
PAPGGSEAAAK	BAEVM_P10272_3mutA
PAPGGSGSS	MLVMS_P03355_3mutA_WS
PAPGGSGSS	AVIRE_P03360_3mutA
GGSGGGPAP	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
GGGGSSPAP	MLVCB_P08361_3mutA
GSSGGGPAP	MLVFF_P26809_3mutA
GGGGSSGGS	MLVMS_P03355_PLV919
GGSGGG	MLVCB_P08361_3mutA
GSSGGGGGS	MLVMS_P03355_PLV919
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	XMRV6_A1Z651_3mutA
GGGGGSGSS	KORV_Q9TTC1_3mut
GGCEAAAKGGS	BAEVM_P10272_3mutA
GGSGGG	BAEVM_P10272_3mutA
PAPAPAP	KORV_Q9TTC1-Pro_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV3L_P27401_2mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	MLVMS_P03355_3mutA_WS
GSSGGSEAAAK	MLVFF_P26809_3mutA
PAP	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS

[0829]

링커 아미노산 서열	RT 도메인 명칭
PAPAP	AVIRE_P03360_3mutA
PAP	MLVFF_P26809_3mutA
GSSGGG	MLVMS_P03355_3mut
GSSPAPGGS	MLVFF_P26809_3mutA
PAPAPAPAP	XMRV6_A1Z651_3mutA
EAAAKGSSGGS	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	KORV_Q9TTC1-Pro_3mutA
PAPGGS	MLVCB_P08361_3mutA
EAAAKGGG	MLVCB_P08361_3mutA
GSSEAAAKPAP	MLVMS_P03355_PLV919
PAPGGS	MLVFF_P26809_3mutA
EAAAKGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	FLV_P10273_3mutA
PAPGGSEAAAK	MLVAV_P03356_3mutA
GSS	MLVCB_P08361_3mutA
GSSGSSGSSGSS	AVIRE_P03360_3mutA
GSSGSSGSS	MLVFF_P26809_3mutA
GSSGGG	MLVMS_P03355_PLV919
EAAAK	MLVFF_P26809_3mutA
GGSPAPEAAAK	MLVCB_P08361_3mutA
GGSGSS	MLVCB_P08361_3mutA
GSSPAPGGG	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVAV_P03356_3mutA
EAAAKGSSPAP	FLV_P10273_3mutA
GGGGSS	XMRV6_A1Z651_3mutA
GGSPAPGSS	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
PAPEAAAKGGG	FLV_P10273_3mutA
EAAAKPAPGGS	XMRV6_A1Z651_3mut
PAPAP	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GSSPAPGGG	MLVMS_P03355_PLV919
EAAAKGGGPAP	KORV_Q9TTC1_3mutA
PAPEAAAK	MLVMS_P03355_PLV919
PAPGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKGSSGGS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GSSEAAAK	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS

[0830]

[0831]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGGGGSEAAAK	MLVMS_P03355_3mut
GGS	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mutA
GGSGSSPAP	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGS	XMRV6_A1Z651_3mutA
PAPAPAPAPAP	BAEVM_P10272_3mutA
PAPAPAPAPAP	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mutA
GGSGGSGGS	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	MLVMS_P03355_PLV919
GSS	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	MLVMS_P03355_3mutA_WS
GGGPAPGGS	MLVMS_P03355_3mutA_WS
EAAAKGGGGSS	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	MLVFF_P26809_3mut
SGSETPGTSESATPES	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	AVIRE_P03360_3mutA
PAPAPAPAPAPAP	AVIRE_P03360_3mut
GGSGGS	XMRV6_A1Z651_3mutA
PAPGSSEAAAK	MLVCB_P08361_3mut
GGSPAPEAAAK	PERV_Q4VFZ2_3mut
EAAAKGGGGS	MLVCB_P08361_3mutA
GGSGGSGGSGGS	MLVMS_P03355_PLV919
GGGGSSEAAAK	MLVMS_P03355_PLV919
GSSEAAAKGGG	MLVFF_P26809_3mutA
PAPGGS	MLVMS_P03355_3mutA_WS
EAAAKGGSGGG	MLVCB_P08361_3mutA
EAAAKGGG	PERV_Q4VFZ2_3mut
PAPGGS	XMRV6_A1Z651_3mutA
GSSPAPGGG	XMRV6_A1Z651_3mutA
PAPEAAAKGGG	MLVMS_P03355_3mutA_WS
GSSEAAAKGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	XMRV6_A1Z651_3mutA
GGGGGS	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	MLVMS_P03355_3mutA_WS
GGGPAP	MLVFF_P26809_3mutA
PAPGSSGGG	XMRV6_A1Z651_3mutA
PAPGSSGGG	MLVBM_Q7SVK7_3mutA_WS

[0832]

링커 아미노산 서열	RT 도메인 명칭
GGGEAAAKGSS	MLVMS_P03355_3mutA_WS
GSSEAAAKGGS	MLVCB_P08361_3mutA
PAPGGSGSS	MLVCB_P08361_3mutA
EAAAKGGGGSEAAAK	BAEVM_P10272_3mutA
PAPAPAP	PERV_Q4VFZ2_3mutA_WS
GGGGGG	MLVAV_P03356_3mutA
GSSPAPEAAAK	MLVCB_P08361_3mutA
GGSGGSGGS	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	XMRV6_A1Z651_3mut
GGGPAPGGS	XMRV6_A1Z651_3mutA
GGGPAPEAAAK	BAEVM_P10272_3mutA
GGSGGG	AVIRE_P03360_3mutA
SGSETPGTSESATPES	PERV_Q4VFZ2_3mutA_WS
EAAAKGSSPAP	MLVMS_P03355_PLV919
GSSEAAAK	XMRV6_A1Z651_3mut
GSSGGSGGG	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mutA
GGGGSEAAAKGGGGS	MLVMS_P03355_PLV919
PAPGGGGSS	MLVMS_P03355_3mutA_WS
SGSETPGTSESATPES	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGG	MLVMS_P03355_3mutA_WS
GSSEAAAK	WMSV_P03359_3mutA
GGGGSEAAAKGGGGS	AVIRE_P03360_3mutA
GSS	WMSV_P03359_3mutA
PAPGGSEAAAK	MLVFF_P26809_3mutA
GGGGS	MLVMS_P03355_3mutA_WS
GGGPAP	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	PERV_Q4VFZ2_3mut
EAAAKPAPGSS	MLVCB_P08361_3mutA
GGGGGG	WMSV_P03359_3mutA
EAAAKPAPGGS	MLVMS_P03355_PLV919
PAPGGGEAAAK	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	AVIRE_P03360_3mutA
GSSEAAAKPAP	XMRV6_A1Z651_3mutA
PAPGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPGSS	MLVCB_P08361_3mutA
EAAAKGGG	MLVMS_P03355_3mutA_WS
EAAAKPAP	MLVCB_P08361_3mutA

링커 아미노산 서열	RT 도메인 명칭
PAPEAAAKGGS	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	MLVCB_P08361_3mutA
PAPGGSGSS	WMSV_P03359_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GGSGGGPAP	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355
PAPEAAAKGSS	MLVCB_P08361_3mutA
EAAAKGSS	MLVMS_P03355_3mutA_WS
GGSGGS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	FLV_P10273_3mutA
GGSEAAAKGGG	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSS	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVFF_P26809_3mutA
EAAAKGGG	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	MLVCB_P08361_3mutA
EAAAKPAPGGS	MLVMS_P03355_3mutA_WS
GGGGGSGSS	XMRV6_A1Z651_3mutA
PAPGSSEAAAK	MLVMS_P03355_3mutA_WS
GSSEAAAKPAP	MLVCB_P08361_3mutA
EAAAKGSSPAP	MLVAV_P03356_3mutA
GGGPAPGGS	WMSV_P03359_3mutA
GGSPAP	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	MLVMS_P03355_3mutA_WS
GGGGGGGG	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGG	MLVAV_P03356_3mutA
GGGGGG	AVIRE_P03360_3mutA
GSSGGS	MLVMS_P03355_3mutA_WS
GGSPAPGSS	MLVFF_P26809_3mutA
PAPEAAAKGGG	PERV_Q4VFZ2_3mut
EAAAKGGGPAP	MLVFF_P26809_3mutA
GGGEAAAKGGS	MLVMS_P03355_PLV919
GGSGSSPAP	MLVFF_P26809_3mutA
SGSETPGTSESATPES	WMSV_P03359_3mutA
PAPGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGGG	MLVMS_P03355_PLV919
GGGGSSPAP	PERV_Q4VFZ2_3mut
GGGEAAAKGSS	MLVAV_P03356_3mutA

[0833]

링커 아미노산 서열	RT 도메인 명칭
PAPAPAPAPAP	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	PERV_Q4VFZ2
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GGGGGSEAAAK	PERV_Q4VFZ2_3mut
PAPGSSEAAAK	MLVCB_P08361_3mutA
GSAGSAAGSGEF	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSEAAAK	MLVFF_P26809_3mutA
GGSPAPGGG	PERV_Q4VFZ2_3mutA_WS
GSSEAAAKGGG	AVIRE_P03360_3mutA
GGGEAAAKPAP	MLVMS_P03355_3mutA_WS
GGGPAP	AVIRE_P03360_3mutA
GGSEAAAK	MLVCB_P08361_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	PERV_Q4VFZ2_3mut
EAAAKPAPGGS	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	XMRV6_A1Z651_3mut
GGGGGGGG	MLVCB_P08361_3mutA
PAPGSS	PERV_Q4VFZ2_3mut
EAAAK	PERV_Q4VFZ2_3mut
GSAGSAAGSGEF	MLVMS_P03355_3mutA_WS
PAPGGGEAAAK	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	MLVFF_P26809_3mut
GGGGSEAAAKGGGGS	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	BAEVM_P10272_3mut
PAPGGGEAAAK	MLVMS_P03355_3mutA_WS
GGSEAAAKPAP	MLVMS_P03355_3mutA_WS
PAPAP	MLVCB_P08361_3mutA
PAPAP	MLVFF_P26809_3mutA
GGSPAP	AVIRE_P03360_3mutA
EAAAKGSSGGS	MLVCB_P08361_3mutA
PAPGSSGGS	AVIRE_P03360_3mutA
EAAAKGGGGSEAAAK	XMRV6_A1Z651_3mutA
PAPAPAP	BAEVM_P10272_3mutA
GGSGGSGGSGGSGGSGGS	MLVMS_P03355_PLV919
GGGGGSGSS	MLVMS_P03355_PLV919
PAPGSSEAAAK	XMRV6_A1Z651_3mut
GGSEAAAKPAP	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAK	XMRV6_A1Z651_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	WMSV_P03359_3mut
GGSGGGEAAAK	XMRV6_A1Z651_3mutA

[0834]

링커 아미노산 서열	RT 도메인 명칭
GGGEAAAK	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	MLVFF_P26809_3mutA
GSSGGGGGS	MLVMS_P03355_3mut
PAPGGSEAAAK	MLVMS_P03355_3mutA_WS
GSSGGSPAP	MLVMS_P03355_3mutA_WS
SGSETPGTSESATPES	XMRV6_A1Z651_3mutA
GGGGSGGGGS	MLVMS_P03355_PLV919
PAPAPAPAPAP	MLVMS_P03355_3mut
GSSGSS	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	PERV_Q4VFZ2_3mut
GGSGSSGGG	MLVMS_P03355_3mutA_WS
EAAAKEAAAK	MLVCB_P08361_3mutA
GSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GSSPAPGGG	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV1_P23074_2mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_PLV919
GSAGSAAGSGEF	MLVMS_P03355_PLV919
PAPGSSEAAAK	MLVMS_P03355_3mutA_WS
GGSEAAAK	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGS	MLVCB_P08361_3mutA
EAAAKGGSGSS	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
GGSGSS	WMSV_P03359_3mutA
PAPEAAAK	MLVCB_P08361_3mutA
EAAAKPAP	BAEVM_P10272_3mutA
GSSPAP	PERV_Q4VFZ2_3mutA_WS
GGGPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	AVIRE_P03360_3mutA
GGSGGG	KORV_Q9TTC1-Pro_3mutA
GSSPAP	MLVFF_P26809_3mutA
GGSGSSEAAAK	BAEVM_P10272_3mutA

[0835]

링커 아미노산 서열	RT 도메인 명칭
PAPGSSGGS	BAEVM_P10272_3mutA
GGGGGG	MLVFF_P26809_3mutA
PAPGGSEAAAK	MLVMS_P03355_PLV919
PAPGGS	MLVMS_P03355_PLV919
GGSGGSGGSGGS	BAEVM_P10272_3mutA
GSSPAP	MLVCB_P08361_3mutA
PAPAPAPAP	MLVMS_P03355_3mutA_WS
GGGGGG	MLVCB_P08361_3mutA
GSSGSSGSSGSSGSSGSS	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGS	BAEVM_P10272_3mutA
GGSEAAAK	FLV_P10273_3mutA
GGSGGSGGSGGSGGS	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	XMRV6_A1Z651_3mutA
EAAAKPAPGGS	MLVMS_P03355_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	FLV_P10273_3mut
GGSPAPEAAAK	XMRV6_A1Z651_3mut
EAAAKGGSGGG	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mutA
GSSPAP	WMSV_P03359_3mutA
PAPAPAPAP	MLVAV_P03356_3mutA
PAPGGSEAAAK	KORV_Q9TTC1_3mut
GGSGSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GSSGGG	MLVCB_P08361_3mutA
GGGEAAAKGSS	PERV_Q4VFZ2_3mut
PAPGGSGGG	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FFV_093209
PAPGGGGSS	MLVMS_P03355_3mutA_WS
EAAAKGGS	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGS	WMSV_P03359_3mutA
PAPAP	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	MLVAV_P03356_3mutA
GGGGSSEAAAK	MLVFF_P26809_3mutA
EAAAKGSSGGS	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	MLVMS_P03355_3mutA_WS
GGGGGGGG	MLVMS_P03355_PLV919
GSSGSSGSS	MLVMS_P03355_PLV919
GGGEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GGGGGSGSS	MLVMS_P03355_3mutA_WS

[0836]

링커 아미노산 서열	RT 도메인 명칭
GGGGGGG	MLVMS_P03355_PLV919
GGS	MLVMS_P03355_PLV919
GSSGGG	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	PERV_Q4VFZ2_3mutA_WS
PAPGSSEAAAK	MLVMS_P03355_PLV919
GSSEAAAKPAP	MLVMS_P03355_PLV919
GGSPAPGSS	BAEVM_P10272_3mutA
GSAGSAAGSGEF	MLVCB_P08361_3mut
GGSPAPGGG	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mut
GSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
GGGGSEAAAKGGGGS	MLVCB_P08361_3mutA
GGSEAAAKGSS	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	MLVCB_P08361_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	XMRV6_A1Z651_3mutA
PAPGGGEAAAK	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
GSSGSS	MLVCB_P08361_3mut
PAPAPAPAPAP	PERV_Q4VFZ2_3mut
GGSPAPGGG	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGS	MLVCB_P08361_3mutA
EAAAKEAAAK	MLVFF_P26809_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	GALV_P21414_3mut
PAPAPAPAPAP	WMSV_P03359_3mutA
GGGEAAAKGGS	KORV_Q9TTC1_3mutA
EAAAKGGGPAP	KORV_Q9TTC1_3mut
PAPEAAAKGSS	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGSS	FLV_P10273_3mutA
PAPGGSEAAAK	MLVMS_P03355_3mut
GSSPAPGGG	BAEVM_P10272_3mutA
GGGEAAAKPAP	KORV_Q9TTC1-Pro_3mutA
GGGGSGGGGS	MLVMS_P03355_PLV919
GGGEAAAKGSS	MLVFF_P26809_3mutA
PAPGGGGSS	MLVBM_Q7SVK7_3mutA_WS
GSSEAAAK	BAEVM_P10272_3mutA
GGGGGGGG	MLVMS_P03355_PLV919
PAPGSSGGS	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mutA
PAP	MLVMS_P03355_3mut

[0837]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGSSPAP	XMRV6_A1Z651_3mutA
PAPEAAAKGGS	MLVFF_P26809_3mutA
GSSGGGEAAAK	BAEVM_P10272_3mutA
PAPAPAP	MLVMS_P03355_3mutA_WS
GGSEAAAKGGG	MLVMS_P03355_PLV919
GSSEAAAK	PERV_Q4VFZ2_3mut
GGGG	MLVMS_P03355_3mutA_WS
GGGGGS	MLVMS_P03355_3mut
GGGGSSEAAAK	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro_2mutA
GGSEAAAKGSS	MLVMS_P03355_3mutA_WS
PAPGSSGGS	XMRV6_A1Z651_3mutA
GGSPAP	MLVMS_P03355_3mutA_WS
GGGGSSEAAAK	BAEVM_P10272_3mut
GSSGGSGGSGGS	AVIRE_P03360_3mutA
PAPGSSGGS	MLVFF_P26809_3mutA
GSSPAPGGG	MLVMS_P03355_3mutA_WS
GGGGGGG	MLVMS_P03355_3mutA_WS
EAAAKGGGGS	MLVMS_P03355_3mutA_WS
EAAAKGGSGGG	MLVMS_P03355_PLV919
GGGGSSEAAAK	XMRV6_A1Z651_3mutA
GGGGSEAAAKGGGGS	MLVBM_Q7SVK7_3mutA_WS
GSSGSS	MLVMS_P03355_PLV919
GGSGGG	MLVMS_P03355_PLV919
PAPEAAAKGGG	AVIRE_P03360_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FOAMV_P14350-Pro_2mutA
GGGGGSGSS	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	KORV_Q9TTC1-Pro_3mut
GGGGSEAAAKGGGGS	MLVMS_P03355_3mutA_WS
GGGGSPAP	FLV_P10273_3mut
GGGEAAAK	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	FLV_P10273_3mutA
GGG	MLVMS_P03355_PLV919
GGSPAPEAAAK	BAEVM_P10272_3mutA
EAAAKEAAAK	FLV_P10273_3mutA
GGGEAAAKPAP	BAEVM_P10272_3mutA
GGGEAAAKGGS	PERV_Q4VFZ2_3mut
GGSGGSGGS	PERV_Q4VFZ2_3mut
EAAAKGGGPAP	XMRV6_A1Z651_3mutA
EAAAK	MLVBM_Q7SVK7_3mutA_WS

[0838]

[0839]

링커 아미노산 서열	RT 도메인 명칭
PAPEAAAKGGG	PERV_Q4VFZ2_3mut
EAAAKGSS	MLVCB_P08361_3mutA
GGSEAAAKGGG	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mut
GGGGSSPAP	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mut
GGGEAAAKPAP	PERV_Q4VFZ2_3mut
EAAAKEAAAK	BAEVM_P10272_3mutA
GGSGSSEAAAK	XMRV6_A1Z651_3mutA
PAPEAAAKGSS	WMSV_P03359_3mutA
PAPAPAPAPAP	XMRV6_A1Z651_3mutA
GSSGGGEAAAK	MLVMS_P03355_PLV919
GSSPAPGGG	MLVFF_P26809_3mutA
GGSPAPEAAAK	MLVFF_P26809_3mut
PAPGGSEAAAK	PERV_Q4VFZ2_3mut
GGGGSS	MLVFF_P26809_3mutA
GGSGSSGGG	BAEVM_P10272_3mutA
GSSGGGEAAAK	MLVMS_P03355_3mutA_WS
EAAAKGGS	MLVBM_Q7SVK7_3mutA_WS
GGGPAPGGS	MLVMS_P03355_PLV919
EAAAKEAAAK	MLVMS_P03355_PLV919
GSSGSSGSS	MLVMS_P03355_PLV919
GGGEAAAKPAP	MLVAV_P03356_3mutA
SGSETPGTSESATPES	FLV_P10273_3mutA
PAPAPAPAPAP	KORV_Q9TTC1-Pro_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	BAEVM_P10272_3mutA
PAPGSSGGG	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	XMRV6_A1Z651_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mutA
GGGGSSPAP	MLVFF_P26809_3mutA
GGSGGGPAP	PERV_Q4VFZ2_3mutA_WS
GSS	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	MLVMS_P03355_3mut
EAAAKGGG	XMRV6_A1Z651_3mutA
GSSGSSGSSGSS	WMSV_P03359_3mutA
PAPEAAAKGSS	MLVMS_P03355_PLV919
GSSEAAAK	AVIRE_P03360_3mutA
EAAAKGGSGSS	AVIRE_P03360_3mutA
GSSEAAAK	MLVMS_P03355_3mut

링커 아미노산 서열	RT 도메인 명칭
GGSGSSEAAAK	MLVMS_P03355_PLV919
GGSEAAAKGGG	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	MLVAV_P03356_3mutA
PAPAPAPAPAPAP	MLVFF_P26809_3mut
EAAAKPAPGSS	KORV_Q9TTC1-Pro_3mut
PAPGSSEAAAK	MLVAV_P03356_3mutA
GGGGSSPAP	WMSV_P03359_3mutA
EAAAKGGGGGS	MLVMS_P03355_3mutA_WS
GGGEAAAKGGS	MLVMS_P03355_3mut
GGSGSSGGG	MLVMS_P03355_3mut
GGGPAPGGS	MLVAV_P03356_3mutA
PAPGGGGGS	MLVMS_P03355_PLV919
GGGPAPGSS	PERV_Q4VFZ2_3mut
GGGGGGG	MLVFF_P26809_3mutA
GGSGGGGSS	MLVCB_P08361_3mutA
GGGGGG	FLV_P10273_3mutA
GGSEAAAKGSS	PERV_Q4VFZ2_3mut
GGSPAPGGG	BAEVM_P10272_3mutA
GGSPAPGSS	AVIRE_P03360_3mutA
GGSGGSGGSGGS	KORV_Q9TTC1_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mut
PAPGSSGGS	XMRV6_A1Z651_3mut
EAAAKGGGGSS	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	MLVMS_P03355_PLV919
PAPGSSGGG	PERV_Q4VFZ2_3mutA_WS
GSSGSS	BAEVM_P10272_3mutA
EAAAKGSS	MLVFF_P26809_3mutA
GGGPAP	MLVMS_P03355_PLV919
EAAAKGGGGGS	MLVFF_P26809_3mutA
EAAAKGGSPAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mutA
GSSPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGGEAAAKGSS	AVIRE_P03360_3mutA
GGGGSSEAAAK	AVIRE_P03360_3mutA
GGGGGGGG	PERV_Q4VFZ2_3mutA_WS
PAPGSSEAAAK	BAEVM_P10272_3mutA
EAAAKGSS	MLVFF_P26809_3mut
GSSEAAAKGGG	MLVCB_P08361_3mutA
GGSEAAAK	MLVBM_Q7SVK7_3mutA_WS

[0840]

링커 아미노산 서열	RT 도메인 명칭
GSSEAAAKGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	WMSV_P03359_3mutA
GSGGGSGGG	MLVCB_P08361_3mutA
EAAAKGSSGGG	FLV_P10273_3mutA
GSSEAAAK	MLVCB_P08361_3mutA
GSSGGGEAAAK	MLVMS_P03355_3mut
GGGGSGGGGS	MLVCB_P08361_3mutA
EAAAKGGGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGG	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSPAP	MLVMS_P03355_PLV919
GGGPAPGGS	AVIRE_P03360_3mutA
GSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GSSGGGEAAAK	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	MLVMS_P03355_PLV919
GGSGSSPAP	MLVMS_P03355_3mut
GGGGGG	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	XMRV6_A1Z651_3mutA
GGSGSS	PERV_Q4VFZ2_3mutA_WS
PAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGSS	MLVMS_P03355_PLV919
EAAAKGGG	MLVMS_P03355_3mut
GSSEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GGGGSS	MLVMS_P03355_3mutA_WS
GGSGSSEAAAK	PERV_Q4VFZ2_3mut
GGGGSS	BAEVM_P10272_3mutA
PAPAP	MLVFF_P26809_3mut
PAPEAAAKGGG	BAEVM_P10272_3mutA
EAAAKGGS	MLVMS_P03355_PLV919
PAPAPAPAPAP	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	MLVMS_P03355_3mut
PAPGGS	PERV_Q4VFZ2_3mut
GGGGSS	MLVCB_P08361_3mutA
GGGGS	MLVAV_P03356_3mutA
GSSPAPEAAAK	MLVMS_P03355_PLV919
GGGGSSGGS	MLVFF_P26809_3mutA
PAPEAAAKGSS	MLVMS_P03355_PLV919
GGSGSSEAAAK	MLVMS_P03355_3mutA_WS
EAAAKGGG	MLVAV_P03356_3mutA
PAPGSSEAAAK	FLV_P10273_3mutA
EAAAKGSSGGG	MLVCB_P08361_3mutA

[0841]

[0842]

링커 아미노산 서열	RT 도메인 명칭
PAPEAAAK	KORV_Q9TTC1-Pro_3mutA
GGSPAPEAAAK	KORV_Q9TTC1-Pro_3mut
GGSGGSGGSGGSGGSGGS	MLVAV_P03356_3mutA
GSSEAAAKPAP	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	XMRV6_A1Z651_3mut
PAPGGSGGG	AVIRE_P03360_3mutA
PAPGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGS	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mut
GSSGGSEAAAK	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGS	WMSV_P03359_3mutA
EAAAKGSSGGG	WMSV_P03359_3mutA
EAAAKGGG	PERV_Q4VFZ2_3mutA_WS
SGSETPGTSESATPES	PERV_Q4VFZ2_3mut
PAPGSSGGS	MLVMS_P03355_3mutA_WS
PAPEAAAKGSS	PERV_Q4VFZ2_3mut
PAPEAAAK	AVIRE_P03360_3mutA
GSSEAAAKGGG	BAEVM_P10272_3mutA
GSSPAP	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mut
PAPGGSGSS	MLVAV_P03356_3mutA
GGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	MLVCB_P08361_3mutA
EAAAKGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKGGS	MLVFF_P26809_3mutA
GGSPAP	MLVMS_P03355_PLV919
GGSGSS	MLVMS_P03355_PLV919
SGSETPGTSESATPES	WMSV_P03359_3mut
GGGGGGG	WMSV_P03359_3mut
GGSPAPGSS	MLVCB_P08361_3mutA
GGGGSSGGS	WMSV_P03359_3mut
PAPGGS	MLVMS_P03355_PLV919
PAPGSSGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	PERV_Q4VFZ2_3mut
GGSGGSGGSGGSGGS	BAEVM_P10272_3mutA
GSSEAAAK	PERV_Q4VFZ2_3mut

[0843]

링커 아미노산 서열	RT 도메인 명칭
EAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGSGGSGGS	MLVMS_P03355_3mut
PAPAPAPAPAPAP	MLVMS_P03355_3mut
GGSPAPEAAAK	MLVMS_P03355_PLV919
EAAAK	WMSV_P03359_3mutA
EAAAKGSSGGS	MLVBM_Q7SVK7_3mutA_WS
GGSGGGGSS	MLVMS_P03355_3mutA_WS
GGGEAAAKPAP	MLVMS_P03355_3mut
EAAAKGGSGGG	XMRV6_A1Z651_3mutA
GGGGGSEAAAK	KORV_Q9TTC1-Pro_3mutA
GGGGGG	BAEVM_P10272_3mutA
GGGGGG	MLVMS_P03355_3mut
GGGGGGG	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	AVIRE_P03360
PAPGSSGGS	PERV_Q4VFZ2_3mut
GGGGGS	XMRV6_A1Z651_3mut
EAAAKPAP	XMRV6_A1Z651_3mutA
GGG	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FLV_P10273_3mut
EAAAKGSSPAP	MLVMS_P03355_3mut
SGSETPGTSESATPES	BAEVM_P10272_3mutA
GGSPAPEAAAK	MLVMS_P03355_3mut
GSSGSSGSSGSS	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVMS_P03355_3mut
GGSPAP	MLVCB_P08361_3mutA
GGGGGSEAAAK	MLVMS_P03355_3mutA_WS
GGGGG	MLVFF_P26809_3mutA
GSSEAAAK	MLVAV_P03356_3mutA
GGS	BAEVM_P10272_3mut
EAAAKGGSPAP	MLVCB_P08361_3mutA
PAPAPAPAP	FLV_P10273_3mutA
PAPGGGEAAAK	MLVCB_P08361_3mutA
GGGGSSEAAAK	MLVMS_P03355_3mutA_WS
GGGGG	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mut
GGGGG	MLVMS_P03355_3mut
PAPEAAAKGGG	MLVBM_Q7SVK7_3mutA_WS
GSSGGGPAP	XMRV6_A1Z651_3mutA
GSSGSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSPAP	PERV_Q4VFZ2_3mut

[0844]

링커 아미노산 서열	RT 도메인 명칭
GSSGGSEAAAK	MLVMS_P03355_PLV919
GSS	PERV_Q4VFZ2_3mut
EAAAKGGS	WMSV_P03359_3mutA
GGGGGSPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKGSS	MLVMS_P03355_PLV919
EAAAKGGGSS	KORV_Q9TTC1-Pro_3mutA
PAPGSSGGG	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	MLVFF_P26809_3mut
PAPAPAP	MLVMS_P03355_3mut
GSSGGSEAAAK	XMRV6_A1Z651_3mut
PAPEAAAKGSS	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	MLVMS_P03355_3mutA_WS
GGSGSSPAP	XMRV6_A1Z651_3mutA
GGGGSSPAP	MLVMS_P03355_PLV919
GGGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAK	KORV_Q9TTC1_3mutA
PAPGGGEAAAK	BAEVM_P10272_3mutA
GSSGGSEAAAK	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	FLV_P10273_3mut
GSSEAAAKPAP	MLVMS_P03355_3mutA_WS
EAAAKPAPGSS	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	XMRV6_A1Z651_3mutA
GSSEAAAKGGG	PERV_Q4VFZ2_3mut
GGGEAAAKGGS	WMSV_P03359_3mutA
GSSEAAAKGGG	MLVFF_P26809_3mut
PAPAPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKGGSPAP	MLVMS_P03355_3mutA_WS
PAPGGSEAAAK	PERV_Q4VFZ2_3mut
GGGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSGGG	KORV_Q9TTC1_3mut
EAAAKGGGPAP	MLVCB_P08361_3mutA
EAAAKGSS	BAEVM_P10272_3mutA
GGSPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGGGSEAAAKGGGGS	MLVMS_P03355_3mutA_WS
GGGEAAAKGGS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSS	MLVMS_P03355_3mutA_WS
EAAAKGGGPAP	MLVFF_P26809_3mut
GSSPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	MLVMS_P03355_3mut

[0845]

링커 아미노산 서열	RT 도메인 명칭
GGGGSS	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSPAP	MLVMS_P03355_3mutA_WS
GGGPAP	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	XMRV6_A1Z651_3mutA
PAPGGG	MLVAV_P03356_3mutA
GSSPAPEAAAK	BAEVM_P10272_3mutA
GGGPAP	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	AVIRE_P03360_3mutA
SGSETPGTSESATPES	MLVMS_P03355_PLV919
GGGPAP	MLVFF_P26809_3mut
EAAAKGGGGSS	XMRV6_A1Z651_3mutA
GGGGSSPAP	XMRV6_A1Z651_3mut
GGGGSEAAAKGGGGS	MLVMS_P03355_3mut
GSSPAP	MLVBM_Q7SVK7_3mutA_WS
GGSGSSEAAAK	FLV_P10273_3mutA
SGSETPGTSESATPES	MLVBM_Q7SVK7_3mutA_WS
PAPGGG	AVIRE_P03360_3mutA
GGGEAAAKPAP	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	PERV_Q4VFZ2_3mut
GGSPAPGGG	MLVAV_P03356_3mutA
PAPGGSGSS	BAEVM_P10272_3mutA
GSSGGSPAP	MLVFF_P26809_3mutA
EAAAKGSSGGG	PERV_Q4VFZ2_3mut
GGGGSGGGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	BAEVM_P10272_3mutA
GGGGSSGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGS	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSS	MLVMS_P03355_3mut
GGS	MLVMS_P03355_3mutA_WS
GSSGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	XMRV6_A1Z651
GGGGG	FLV_P10273_3mutA
PAPEAAAKGSS	PERV_Q4VFZ2_3mut
GGGGGG	WMSV_P03359_3mut
EAAAKGGG	BAEVM_P10272_3mutA
GGGGSS	MLVMS_P03355_3mutA_WS
GSSGGGEAAAK	KORV_Q9TTC1_3mut
GGSGSS	AVIRE_P03360_3mutA
EAAAKPAP	MLVMS_P03355_3mut
EAAAKEAAAKEAAAK	FLV_P10273_3mutA

링커 아미노산 서열	RT 도메인 명칭
GGGG	XMRV6_A1Z651_3mutA
GSSPAPGGS	BAEVM_P10272_3mutA
GSSGGGGGS	MLVFF_P26809_3mutA
GGGGSSGGS	MLVAV_P03356_3mutA
GGS	PERV_Q4VFZ2_3mut
GGGGG	WMSV_P03359_3mutA
GSSGSSGSSGSSGSSGSS	FLV_P10273_3mutA
PAPGGGGSS	MLVAV_P03356_3mutA
GGGGGGGG	BAEVM_P10272_3mutA
SGSETPGTSESATPES	MLVCB_P08361_3mutA
PAPGGG	BAEVM_P10272_3mutA
GSSGSSGSS	MLVCB_P08361_3mutA
GGSGSS	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	WMSV_P03359_3mutA
GGGGGGGG	FLV_P10273_3mutA
GSSGSS	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	XMRV6_A1Z651_3mutA
EAAAKEAAAK	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	BAEVM_P10272_3mutA
EAAAKGSSPAP	MLVMS_P03355_PLV919
GGGGSSEAAAK	MLVMS_P03355_3mut
GGGGSSEAAAK	BAEVM_P10272_3mutA
PAPGGSGSS	PERV_Q4VFZ2_3mut
GGSGGGEAAAK	MLVFF_P26809_3mut
PAPEAAAKGGS	PERV_Q4VFZ2_3mut
GGGPAPGSS	AVIRE_P03360_3mut
PAPGGSGGG	PERV_Q4VFZ2_3mutA_WS
GGGGGGGG	PERV_Q4VFZ2_3mutA_WS
GSSEAAAK	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	MLVMS_P03355_3mut
GGGGGSGSS	MLVCB_P08361_3mut
GGGPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGGG	MLVCB_P08361_3mut
GSSGGSPAP	MLVCB_P08361_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVMS_P03355_3mut
PAPAPAPAP	MLVMS_P03355_3mut
GSSGGS	XMRV6_A1Z651_3mutA
GSSEAAAKGGG	MLVMS_P03355_3mut
GGSGSSPAP	MLVMS_P03355_3mutA_WS

[0846]

링커 아미노산 서열	RT 도메인 명칭
GSSEAAAKGGS	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mut
PAPGGGGSS	KORV_Q9TTC1_3mutA
EAAAKGSS	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FFV_O93209_2mut
GGSGGSGGSGGSGGSGGS	BAEVM_P10272_3mutA
GGGGGG	MLVMS_P03355_PLV919
PAPEAAAK	BAEVM_P10272_3mutA
GGSGSSEAAAK	MLVAV_P03356_3mutA
GGC	MLVCB_P08361_3mutA
GGGGG	MLVCB_P08361_3mutA
GGSGGSGGSGGS	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSSGSS	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	FLV_P10273_3mutA
GGGEAAAKPAP	MLVCB_P08361_3mutA
GSSGSSGSS	MLVMS_P03355_3mutA_WS
PAPAPAPAP	MLVMS_P03355_PLV919
EAAAKGGG	MLVMS_P03355_PLV919
PAPAPAPAPAP	FLV_P10273_3mutA
EAAAKGGSGSS	MLVMS_P03355_3mut
GGCGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGG	MLVCB_P08361_3mutA
GGGGGSGSS	KORV_Q9TTC1_3mutA
GGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mut
GGSGGSGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGGG	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	XMRV6_A1Z651
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	FLV_P10273_3mutA
EAAAKGGGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGPAPGSS	AVIRE_P03360_3mutA
GGCGG	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mut
GGGGSGGGGS	MLVMS_P03355_3mutA_WS
EAAAKGGSPAP	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	AVIRE_P03360_3mutA
PAPGGSGSS	KORV_Q9TTC1-Pro_3mutA
GSS	MLVBM_Q7SVK7_3mutA_WS
GSS	WMSV_P03359_3mut
GGGPAPGSS	MLVFF_P26809_3mutA
EAAAKPAP	MLVMS_P03355_3mut

[0847]

링커 아미노산 서열	RT 도메인 명칭
GSSPAPEAAAK	FLV_P10273_3mutA
GGSPAPGSS	MLVBM_Q7SVK7_3mutA_WS
GGGGGSEAAAK	XMRV6_A1Z651_3mut
PAPEAAAKGGG	WMSV_P03359_3mutA
PAPGGG	PERV_Q4VFZ2_3mut
GGSPAPEAAAK	WMSV_P03359_3mutA
GGSGGGGSS	PERV_Q4VFZ2_3mut
EAAAKGGGGSS	PERV_Q4VFZ2_3mut
EAAAKGGSPAP	AVIRE_P03360_3mut
GGSGGGGSS	WMSV_P03359_3mutA
PAPGSSEAAAK	MLVFF_P26809_3mut
GSSEAAAK	MLVMS_P03355_PLV919
GSAGSAAGSGEF	AVIRE_P03360_3mutA
EAAAKGGSGSS	MLVMS_P03355_3mut
GGSEAAAKPAP	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVFF_P26809_3mutA
PAPGSSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGSSPAP	MLVMS_P03355_3mutA_WS
PAPAPAP	MLVCB_P08361_3mutA
EAAAKPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGCPAPGSS	BAEVM_P10272_3mutA
PAP	MLVMS_P03355_3mutA_WS
PAPGGSGGG	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAP	XMRV6_A1Z651_3mut
GSSPAPGGG	MLVMS_P03355_3mutA_WS
GSSPAPGGG	MLVMS_P03355_3mut
PAPGGG	MLVMS_P03355_PLV919
GGGEAAAKGSS	WMSV_P03359_3mut
EAAAKGSS	KORV_Q9TTC1-Pro_3mutA
EAAAKGGS	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	MLVMS_P03355_PLV919
EAAAKGSSGGG	MLVFF_P26809_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVAV_P03356_3mutA
GSSGGSGGG	MLVFF_P26809_3mut
GSSGSSGSSGSS	PERV_Q4VFZ2_3mutA_WS
GGSPAPGGG	MLVMS_P03355_PLV919
GSS	BAEVM_P10272_3mut

[0848]

[0849]

링커 아미노산 서열	RT 도메인 명칭
GGGPAPGSS	MLVMS_P03355_3mutA_WS
GGCGSS	KORV_Q9TTC1_3mutA
GSSGGSGGG	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAK	MLVCB_P08361_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	FLV_P10273_3mutA
PAPGGGGGS	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	KORV_Q9TTC1-Pro_3mutA
EAAAK	MLVMS_P03355_3mutA_WS
GGG	MLVCB_P08361_3mut
GGSEAAAKGGG	BAEVM_P10272_3mutA
GGGGGSGSS	MLVAV_P03356_3mutA
EAAAKGSSPAP	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGS	XMRV6_A1Z651_3mut
EAAAKPAPGGG	KORV_Q9TTC1-Pro_3mutA
GGGPAPEAAAK	FLV_P10273_3mutA
GGSPAPEAAAK	MLVMS_P03355_3mutA_WS
GGSGGSGGSGGSGGS	MLVFF_P26809_3mut
EAAAKGGSGSS	MLVMS_P03355_PLV919
GGGEAAAKGGS	MLVBM_Q7SVK7_3mutA_WS
PAPAPAPAP	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mut
EAAAKPAP	XMRV6_A1Z651_3mut
EAAAKEAAAK	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGG	BAEVM_P10272_3mut
EAAAKGSS	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVFF_P26809_3mut
GGGPAPGSS	PERV_Q4VFZ2_3mutA_WS
GGGG	PERV_Q4VFZ2_3mut
EAAAKGGSGSS	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
EAAAK	MLVMS_P03355_3mutA_WS
GGGGSS	PERV_Q4VFZ2
PAPEAAAKGGS	MLVCB_P08361_3mut
GSS	MLVMS_P03355_3mut
GSAGSAAGSGEF	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mut
GGGGSGGGGS	AVIRE_P03360_3mutA
EAAAK	MLVMS_P03355_3mut
GGGPAPGGS	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGS	MLVMS_P03355_PLV919

링커 아미노산 서열	RT 도메인 명칭
PAPGGG	MLVMS_P03355_3mutA_WS
GGGEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGSS	KORV_Q9TTC1-Pro_3mutA
PAPGSS	KORV_Q9TTC1_3mutA
GSAGSAAGSGEF	PERV_Q4VFZ2_3mut
PAPGGGGSS	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	MLVCB_P08361_3mutA
GSS	AVIRE_P03360_3mutA
GSSGSSGSSGSS	XMRV6_A1Z651_3mutA
PAPEAAAKGGG	MLVMS_P03355_PLV919
GGGPAPEAAAK	MLVCB_P08361_3mutA
PAPGGGGGS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	MLVFF_P26809_3mutA
GSSGSSGSSGSSGSS	PERV_Q4VFZ2
GSSPAPEAAAK	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
GGSPAPEAAAK	MLVAV_P03356_3mutA
GSSGGG	BAEVM_P10272_3mut
EAAAKGSSGGS	KORV_Q9TTC1-Pro_3mutA
GGSGSSEAAAK	MLVMS_P03355_3mutA_WS
GGGPAPEAAAK	MLVFF_P26809_3mutA
GGGPAPGGS	MLVMS_P03355_3mutA_WS
GGGGG	MLVMS_P03355_PLV919
GGGEAAAKPAP	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGS	WMSV_P03359_3mut
GGGPAPEAAAK	PERV_Q4VFZ2_3mut
GGSGSSEAAAK	MLVMS_P03355_PLV919
EAAAKGGGPAP	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	KORV_Q9TTC1-Pro_3mutA
PAPAP	WMSV_P03359_3mutA
GGSPAPGSS	MLVAV_P03356_3mutA
GGSGGGPAP	MLVMS_P03355_3mut
GGSPAP	MLVMS_P03355_PLV919
EAAAKGGSPAP	PERV_Q4VFZ2_3mut
GSSPAPGGG	KORV_Q9TTC1-Pro_3mutA
GSAGSAAGSGEF	MLVMS_P03355_3mut
GGSPAP	PERV_Q4VFZ2_3mut
GSSGSS	KORV_Q9TTC1-Pro_3mut

[0850]

링커 아미노산 서열	RT 도메인 명칭
GGGPAPGSS	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FOAMV_P14350
PAPGSSGGG	MLVMS_P03355_PLV919
GGSEAAAKPAP	BAEVM_P10272_3mutA
GGGGGS	MLVCB_P08361_3mutA
PAPEAAAKGGS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mutA
GGSEAAAK	BAEVM_P10272_3mutA
GSSPAPEAAAK	MLVMS_P03355_3mutA_WS
PAPGGG	WMSV_P03359_3mut
EAAAKPAP	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	WMSV_P03359_3mut
PAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGSGGGEAAAK	BAEVM_P10272_3mutA
PAPGGS	MLVMS_P03355_3mut
GGSGGSGGSGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	WMSV_P03359_3mutA
GGGPAP	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651_3mut
GGSPAPGSS	KORV_Q9TTC1_3mut
GGGPAPGSS	MLVMS_P03355_3mut
GGGGSSGGS	BAEVM_P10272_3mutA
GGGEAAAKGSS	KORV_Q9TTC1-Pro_3mutA
PAPAP	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGGG	PERV_Q4VFZ2_3mut
PAPGSS	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGGSEAAAK	PERV_Q4VFZ2_3mut
GSSEAAAKGGS	KORV_Q9TTC1-Pro_3mut
PAPAPAPAP	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKPAP	WMSV_P03359_3mutA
PAPGGS	FLV_P10273_3mutA
EAAAKGGGPAP	PERV_Q4VFZ2_3mut
GGSGSSGGG	AVIRE_P03360_3mutA
EAAAKGGSGSS	BAEVM_P10272_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVCB_P08361_3mutA
GSSEAAAKGGS	XMRV6_A1Z651_3mutA
GGGGG	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	SFV3L_P27401_2mutA

[0851]

[0852]

링커 아미노산 서열	RT 도메인 명칭
GGGEAAAKGSS	MLVMS_P03355_PLV919
EAAAKGGGGSEAAAK	KORV_Q9TTC1_3mutA
EAAAKGGG	AVIRE_P03360_3mut
GGSGGG	MLVMS_P03355_3mutA_WS
GGSGSSGGG	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	KORV_Q9TTC1_3mut
GGGGSEAAAKGGGGS	KORV_Q9TTC1_3mutA
PAPAPAPAPAP	FLV_P10273_3mutA
GGS	MLVBM_Q7SVK7_3mutA_WS
GGGGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSSGSS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mut
GGSGSSGGG	PERV_Q4VFZ2_3mut
PAP	MLVFF_P26809_3mut
GSSPAPEAAAK	MLVAV_P03356_3mutA
EAAAKGGGGSS	MLVMS_P03355_3mut
GGGEAAAKGGS	XMRV6_A1Z651_3mut
GGSGGGPAP	MLVBM_Q7SVK7_3mutA_WS
GSAGSAAGSGEF	BAEVM_P10272_3mutA
GSSEAAAK	MLVCB_P08361_3mut
PAPGSS	MLVMS_P03355_3mut
EAAAKEAAAKEAAAK	MLVAV_P03356_3mutA
GSAGSAAGSGEF	XMRV6_A1Z651_3mutA
GSSGSSGSSGSS	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	KORV_Q9TTC1-Pro_3mut
GGCGSSEAAAK	WMSV_P03359_3mut
GSSGGGEAAAK	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAP	MLVFF_P26809_3mutA
GGSPAPGGG	KORV_Q9TTC1_3mutA
PAPEAAAK	FLV_P10273_3mutA
GSSGSSGSS	MLVBM_Q7SVK7_3mutA_WS
GSSGGGEAAAK	FLV_P10273_3mutA
GGSPAP	MLVBM_Q7SVK7_3mutA_WS
GSAGSAAGSGEF	KORV_Q9TTC1-Pro_3mutA
PAPGGSEAAAK	MLVMS_P03355_PLV919
GGSPAPEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGGGGSPAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSPAP	WMSV_P03359_3mut
EAAAKGGGPAP	MLVBM_Q7SVK7_3mutA_WS
PAPGSS	KORV_Q9TTC1-Pro_3mutA

링커 아미노산 서열	RT 도메인 명칭
GGSGSSGGG	BAEVM_P10272_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	FFV_O93209-Pro_2mut
GGSGGSGGSGGSGGSGGS	WMSV_P03359_3mutA
GGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
GGGGG	PERV_Q4VFZ2_3mutA_WS
GGGPAP	FLV_P10273_3mutA
PAPGGSGGG	XMRV6_A1Z651_3mutA
GGGGSEAAAKGGGGS	XMRV6_A1Z651_3mut
EAAAKGSSGGG	KORV_Q9TTC1-Pro_3mutA
GSSGGSEAAAK	WMSV_P03359_3mut
EAAAKGGSGSS	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mutA_WS
GGGGGGG	KORV_Q9TTC1_3mutA
EAAAK	KORV_Q9TTC1-Pro_3mutA
GGGEAAAKGGS	KORV_Q9TTC1-Pro_3mutA
GGGEAAAKGGS	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	XMRV6_A1Z651_3mut
GGGGSGGGGSGGGGSGGGGS	MLVFF_P26809_3mut
GGGGGGG	MLVFF_P26809_3mut
PAPAPAPAPAPAP	AVIRE_P03360_3mutA
GSSPAPGGG	FLV_P10273_3mutA
GGGGGSPAP	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGS	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	KORV_Q9TTC1_3mut
GSSEAAAKGGS	MLVAV_P03358_3mutA
GSSGSSGSSGSSGSS	MLVMS_P03355_3mut
EAAAKGGGGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	PERV_Q4VFZ2_3mut
GGGEAAAKPAP	MLVMS_P03355_3mut
GSSGGSPAP	PERV_Q4VFZ2_3mutA_WS
GSSGGGPAP	BAEVM_P10272_3mutA
GGGGGSGSS	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	BAEVM_P10272_3mut
PAPEAAAK	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	FLV_P10273_3mutA
GGSGSSGGG	WMSV_P03359_3mutA
EAAAKGGS	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	MLVCB_P08361_3mut
EAAAKGGSGSS	WMSV_P03359_3mutA

[0853]

링커 아미노산 서열	RT 도메인 명칭
GSSGSS	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAP	MLVMS_P03355_PLV919
GGSGGG	PERV_Q4VFZ2_3mutA_WS
GSS	MLVBM_Q7SVK7_3mutA_WS
PAP	KORV_Q9TTC1-Pro_3mutA
GGSGSSEAAAK	MLVFF_P26809_3mut
PAPEAAAKGSS	KORV_Q9TTC1-Pro_3mutA
GGSGGS	MLVCB_P08361_3mutA
GGGGGGG	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1_3mutA
GGSPAP	MLVMS_P03355_3mut
GGSEAAAKGGG	PERV_Q4VFZ2_3mut
GGGGSGGGGS	FLV_P10273_3mutA
GGGEAAAK	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	SFV3L_P27401_2mut
GGSEAAAKPAP	KORV_Q9TTC1-Pro_3mutA
GSSGGGEAAAK	MLVMS_P03355_PLV919
GGGGGSEAAAK	MLVMS_P03355_PLV919
EAAAKGGSGGG	MLVMS_P03355_3mutA_WS
GGGGSSPAP	MLVAV_P03356_3mutA
EAAAKEAAAK	MLVMS_P03355_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV3L_P27401_2mut
GSSGSSGSSGSSGSS	MLVMS_P03355_PLV919
GSSGGG	KORV_Q9TTC1-Pro_3mutA
GSSGGS	MLVFF_P26809_3mutA
GGGGSGGGGS	XMRV6_A1Z651_3mutA
PAPGSS	MLVBM_Q7SVK7_3mutA_WS
GGGPAPEAAAK	XMRV6_A1Z651_3mutA
EAAAKGGS	MLVFF_P26809_3mut
GSS	KORV_Q9TTC1_3mutA
GGGG	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	AVIRE_P03360_3mutA
GSSGSSGSSGSSGSS	MLVMS_P03355_PLV919
PAPGGSGGG	PERV_Q4VFZ2_3mut
GGGPAP	PERV_Q4VFZ2_3mut
GGGPAPEAAAK	AVIRE_P03360_3mutA
GGGEAAAK	MLVCB_P08361_3mut
GGG	MLVFF_P26809_3mutA
EAAAKPAPGSS	XMRV6_A1Z651_3mutA

[0854]

링커 아미노산 서열	RT 도메인 명칭
GGSGSSEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKGSS	MLVMS_P03355_3mut
GGSGSSEAAAK	BAEVM_P10272_3mut
GGSGGG	MLVBM_Q7SVK7_3mutA_WS
GGGPAP	MLVMS_P03355_PLV919
GGSPAPGGG	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	MLVFF_P26809_3mutA
EAAAKGSSGGS	MLVBM_Q7SVK7_3mut
PAPAP	XMRV6_A1Z651_3mut
GSSPAPGGS	MLVBM_Q7SVK7_3mutA_WS
GSSEAAAKGGG	WMSV_P03359_3mutA
EAAAKGGGGGS	PERV_Q4VFZ2_3mut
GSSGSSGSSGSSGSS	MLVCB_P08361_3mutA
EAAAKGGGGSS	PERV_Q4VFZ2_3mut
EAAAKGSS	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	AVIRE_P03360_3mutA
EAAAKGGS	MLVCB_P08361_3mut
GSSGGSEAAAK	MLVAV_P03356_3mutA
EAAAKPAPGGS	PERV_Q4VFZ2_3mut
GGSGGS	MLVAV_P03356_3mutA
EAAAKGSSGGG	AVIRE_P03360_3mutA
GGSGGSGGSGGS	PERV_Q4VFZ2_3mut
GGGGGGGG	KORV_Q9TTC1_3mutA
GGSGSSEAAAK	MLVCB_P08361_3mutA
EAAAKGGG	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGSGGGGS	MLVCB_P08361_3mut
GGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAP	WMSV_P03359_3mut
EAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
GGSGGSGGS	XMRV6_A1Z651_3mutA
PAPGGGSS	BAEVM_P10272_3mutA
GSSEAAAKGGS	MLVCB_P08361_3mut
GSSGGGPAP	MLVCB_P08361_3mutA
GGSGSS	MLVBM_Q7SVK7_3mutA_WS
GGGGGSEAAAK	MLVAV_P03356_3mutA
GSSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGGGSGSS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGSGSS	MLVFF_P26809_3mut
PAP	FLV_P10273_3mutA
GGGGG	MLVMS_P03355_3mutA_WS

[0855]

[0856]

링커 아미노산 서열	RT 도메인 명칭
EAAAK	PERV_Q4VFZ2_3mut
GSS	FLV_P10273_3mutA
PAPAPAPAPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mut
EAAAKGGGGSEAAAK	XMRV6_A1Z651_3mut
PAPGGSGGG	MLVBM_Q7SVK7_3mutA_WS
GGSGGGPAP	WMSV_P03359_3mutA
GGGGSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
PAPGGGGSS	MLVCB_P08361_3mut
GGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
PAPGGSGGG	MLVMS_P03355_3mutA_WS
GSSPAPGGS	MLVCB_P08361_3mutA
GSSGSSGSS	MLVFF_P26809_3mut
PAPGGGGGS	MLVBM_Q7SVK7_3mutA_WS
GSSPAP	PERV_Q4VFZ2_3mut
GGSGGG	KORV_Q9TTC1-Pro_3mut
EAAAKGGGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKPAP	BAEVM_P10272_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mut
EAAAKGGGGSS	MLVFF_P26809_3mut
EAAAKEAAAK	MLVCB_P08361_3mut
GSSEAAAKGGS	PERV_Q4VFZ2_3mut
GGSPAP	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	BAEVM_P10272_3mut
PAPEAAAK	MLVMS_P03355_3mut
GSSGGSPAP	PERV_Q4VFZ2
GGGPAPGGS	BAEVM_P10272_3mutA
EAAAKPAPGGS	MLVMS_P03355_PLV919
GGGGSGGGGS	PERV_Q4VFZ2
GGGEAAAK	KORV_Q9TTC1-Pro_3mut
EAAAKGGGGGS	FLV_P10273_3mutA
GGSPAPGSS	MLVMS_P03355_3mut
GSSPAPEAAAK	MLVMS_P03355_3mutA_WS
GSAGSAAGSGEF	MLVBM_Q7SVK7_3mutA_WS
EAAAK	BAEVM_P10272_3mutA
EAAAKGGGGSS	BAEVM_P10272_3mutA
GGG	WMSV_P03359_3mut
GGSGGSPAP	BAEVM_P10272_3mut

링커 아미노산 서열	RT 도메인 명칭
GGSEAAAKPAP	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGSGSS	MLVCB_P08361_3mut
PAPGSS	MLVAV_P03356_3mutA
PAPEAAAKGGG	MLVCB_P08361_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FOAMV_P14350-Pro_2mut
GSSGSSGSS	PERV_Q4VFZ2_3mut
PAPGGG	MLVMS_P03355_3mut
PAPGGS	PERV_Q4VFZ2_3mut
GSSGGG	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	WMSV_P03359_3mut
PAP	AVIRE_P03360_3mutA
EAAAKGSSPAP	MLVBM_Q7SVK7_3mutA_WS
GSSGSSGSSGSS	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGS	AVIRE_P03360
GGGGS	PERV_Q4VFZ2_3mut
EAAAKGSSGGG	MLVBM_Q7SVK7_3mutA_WS
GGGGGG	KORV_Q9TTC1-Pro_3mut
GGSGSSEAAAK	PERV_Q4VFZ2_3mut
GSSPAPEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGGGSGGGGS	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	MLVAV_P03356_3mutA
GSAGSAAGSGEF	WMSV_P03359_3mutA
GGGEAAAKGSS	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FFV_O93209-Pro_2mut
PAPGGSGGG	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401_2mut
GGSGSSPAP	MLVMS_P03355_PLV919
GGCGGG	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mut
EAAAKGSSPAP	MLVFF_P26809_3mut
GGGPAPGGS	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV3L_P27401
PAP	PERV_Q4VFZ2_3mut
EAAAKGGS	MLVMS_P03355_PLV919
GSSGGSEAAAK	WMSV_P03359_3mutA
GGSGSSEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAK	PERV_Q4VFZ2
GGSGGGEAAAK	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	BAEVM_P10272_3mut
EAAAKGSS	XMRV6_A1Z651_3mutA

[0857]

링커 아미노산 서열	RT 도메인 명칭
GSSGGGGGS	WMSV_P03359_3mutA
GSSGSSGSSGSSGSSGSS	MLVFF_P26809_3mutA
GGSGSS	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	MLVMS_P03355_PLV919
EAAAKGGGPAP	PERV_Q4VFZ2
GGSEAAAKGGG	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	MLVFF_P26809_3mutA
GGGGSEAAAKGGGGS	PERV_Q4VFZ2_3mut
GSSGSSGSSGSS	PERV_Q4VFZ2_3mut
GGSEAAAK	MLVFF_P26809_3mutA
GGGGGGGG	MLVMS_P03355_3mut
GSSGGG	XMRV6_A1Z651_3mutA
EAAAKGGS	BAEVM_P10272_3mutA
GGGGS	BAEVM_P10272_3mutA
GGSEAAAKGGG	KORV_Q9TTC1-Pro_3mutA
GGSGSSGGG	KORV_Q9TTC1_3mutA
GGSGSSEAAAK	WMSV_P03359_3mut
EAAAKGGSGSS	MLVBM_Q7SVK7_3mutA_WS
GGS	BAEVM_P10272_3mutA
GGGPAPGSS	WMSV_P03359_3mutA
GSSGSSGSSGSSGSS	AVIRE_P03360_3mut
GGGEAAAKPAP	XMRV6_A1Z651_3mut
GSSGGG	MLVFF_P26809_3mutA
GGSPAPGSS	PERV_Q4VFZ2_3mut
PAPGGS	MLVCB_P08361_3mut
PAPAPAPAP	KORV_Q9TTC1_3mutA
GSSGGS	MLVCB_P08361_3mutA
GSSGGSEAAAK	PERV_Q4VFZ2_3mut
EAAAKGSSGGS	MLVMS_P03355_PLV919
EAAAKGGG	WMSV_P03359_3mut
PAPGGGGGS	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	WMSV_P03359_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mutA_WS
GGS	KORV_Q9TTC1-Pro_3mutA
GSSGGSPAP	BAEVM_P10272_3mutA
GGG	MLVMS_P03355_PLV919
PAPGSS	KORV_Q9TTC1-Pro_3mut
GGSEAAAKGGG	FLV_P10273_3mutA

[0858]

링커 아미노산 서열	RT 도메인 명칭
GGSEAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GGGGSSPAP	XMRV6_A1Z651_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGG	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	MLVMS_P03355_3mut
PAPGSSGGG	MLVMS_P03355_3mutA_WS
PAPEAAAKGGS	AVIRE_P03360_3mut
GGGGSSPAP	MLVMS_P03355_3mutA_WS
GGGGSGGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mut
GGGEAAAK	MLVMS_P03355_3mut
GGGGSS	MLVFF_P26809_3mut
GGSPAPGSS	XMRV6_A1Z651_3mut
GGGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGS	FLV_P10273_3mutA
GSS	MLVMS_P03355_PLV919
GGGG	MLVMS_P03355_PLV919
GSSGGS	MLVMS_P03355_PLV919
GGSGGSGGSGGS	MLVMS_P03355_3mut
PAPEAAAKGGS	MLVMS_P03355_3mut
EAAAKGSSGGG	BAEVM_P10272_3mutA
GSSEAAAK	KORV_Q9TTC1-Pro_3mutA
GSAGSAAGSGEF	KORV_Q9TTC1_3mutA
GGGGGSEAAAK	MLVCB_P08361_3mut
GGGG	WMSV_P03359_3mut
GGGGSSEAAAK	MLVMS_P03355_PLV919
PAPGGG	WMSV_P03359_3mutA
EAAAKGGSGGG	MLVAV_P03356_3mutA
GGGPAPGGS	MLVMS_P03355_3mut
EAAAKPAP	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSS	KORV_Q9TTC1-Pro_3mutA
GSSPAPGGS	XMRV6_A1Z651_3mut
GGGGGSPAP	BAEVM_P10272_3mutA
GGSGSSGGG	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKGSS	AVIRE_P03360_3mut
GSSEAAAK	FLV_P10273_3mutA
EAAAK	MLVMS_P03355_3mut
EAAAKGGSGSS	WMSV_P03359_3mut
GSSEAAAKGGG	PERV_Q4VFZ2_3mut
PAPGSSGGG	BAEVM_P10272_3mutA
EAAAKGGGGS	MLVMS_P03355_3mut

[0859]

[0860]

링커 아미노산 서열	RT 도메인 명칭
GGSEAAAKPAP	AVIRE_P03360_3mut
GGGPAPGGS	XMRV6_A1Z651_3mut
GGGGS	KORV_Q9TTC1_3mutA
GGSGGSGGSGGSGGS	XMRV6_A1Z651_3mut
GGGPAP	KORV_Q9TTC1-Pro_3mut
EAAAKPAP	MLVBM_Q7SVK7_3mutA_WS
GGSEAAAK	MLVMS_P03355_PLV919
GSSEAAAKPAP	KORV_Q9TTC1-Pro_3mutA
GGSGSS	MLVMS_P03355_3mut
EAAAKPAPGGG	PERV_Q4VFZ2_3mut
GGSPAPEAAAK	KORV_Q9TTC1_3mutA
GGSEAAAKGGG	AVIRE_P03360_3mutA
GGGGSEAAAKGGGGS	MLVMS_P03355_PLV919
GSSGGGEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAKGGGPAP	WMSV_P03359_3mut
GSSPAP	XMRV6_A1Z651_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV3L_P27401-Pro
GGSEAAAKGSS	MLVMS_P03355_PLV919
GSSGGSEAAAK	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKGSS	KORV_Q9TTC1-Pro_3mutA
EAAAKGGG	AVIRE_P03360_3mutA
GSSGGSEAAAK	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	KORV_Q9TTC1-Pro_3mut
PAPGSSEAAAK	MLVMS_P03355_3mut
PAPEAAAK	WMSV_P03359_3mut
PAPGGSGSS	PERV_Q4VFZ2_3mutA_WS
PAPGSS	BAEVM_P10272_3mut
PAPGGGGS	MLVMS_P03355_3mut
EAAAKPAPGSS	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGS	MLVMS_P03355_PLV919
GGSGSSEAAAK	MLVMS_P03355_3mut
GGGGGG	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAKEAAAK	MLVBM_Q7SVK7_3mut
GGSPAPGSS	MLVMS_P03355_PLV919
PAPAPAPAPAP	MLVCB_P08361_3mut
GGSGSSPAP	WMSV_P03359_3mutA
EAAAKGGSGGG	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	KORV_Q9TTC1_3mutA
GSSGGGEAAAK	WMSV_P03359_3mutA

[0861]

링커 아미노산 서열	RT 도메인 명칭
GSSGGSEAAAK	FLV_P10273_3mutA
GGGGGGGG	PERV_Q4VFZ2_3mut
PAPGGSEAAAK	FLV_P10273_3mutA
GGGGSSPAP	BAEVM_P10272_3mutA
PAPAPAPAP	WMSV_P03359_3mut
GGSEAAAKPAP	PERV_Q4VFZ2_3mut
PAPGGSGGG	BAEVM_P10272_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_3mut
GGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mut
GGSGGGPAP	PERV_Q4VFZ2_3mut
GGGPAPEAAAK	MLVFF_P26809_3mut
GGGGGSGSS	MLVMS_P03355_3mutA_WS
GSS	MLVCB_P08361_3mut
GGGGGSPAP	MLVMS_P03355_PLV919
GGSPAP	MLVAV_P03356_3mutA
GGGPAPGGS	KORV_Q9TTC1-Pro_3mutA
PAPGSSGGG	FLV_P10273_3mutA
PAPGSSGGG	WMSV_P03359_3mutA
PAPGGS	MLVBM_Q7SVK7_3mutA_WS
GGGEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKGSS	MLVBM_Q7SVK7_3mutA_WS
PAPGGSEAAAK	MLVCB_P08361_3mut
GGSEAAAKGGG	XMRV6_A1Z651_3mutA
GGSGGGGSS	WMSV_P03359_3mut
GGGEAAAKPAP	KORV_Q9TTC1_3mutA
EAAAKGSS	KORV_Q9TTC1-Pro_3mut
PAPEAAAKGSS	MLVFF_P26809_3mut
GSAGSAAGSGEF	PERV_Q4VFZ2_3mut
EAAAKGGGGS	WMSV_P03359_3mut
EAAAKGSSPAP	WMSV_P03359_3mutA
GGGGSEAAAKGGGGS	XMRV6_A1Z651_3mutA
GSSEAAAKPAP	SFV3L_P27401-Pro_2mutA
GGGGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGS	BAEVM_P10272_3mut
PAP	AVIRE_P03360_3mut
PAPAPAP	MLVBM_Q7SVK7_3mutA_WS
GGGG	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	MLVBM_Q7SVK7_3mut
GGSGGGGSS	MLVFF_P26809_3mut
GGGGSSGGS	AVIRE_P03360_3mutA

링커 아미노산 서열	RT 도메인 명칭
GSSPAPGGG	PERV_Q4VFZ2_3mutA_WS
GGSEAAAKPAP	MLVMS_P03355_PLV919
PAP	KORV_Q9TTC1-Pro_3mut
GSSGGS	PERV_Q4VFZ2_3mut
GGGGG	PERV_Q4VFZ2_3mut
GSSGGGPAP	FLV_P10273_3mutA
GSSEAAAKGGG	KORV_Q9TTC1-Pro_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MLVCB_P08361_3mut
GGSEAAAKPAP	MLVCB_P08361_3mut
PAPAPAPAPAP	BAEVM_P10272_3mutA
GGGGSEAAAKGGGGS	MLVMS_P03355_3mut
EAAAKPAPGSS	MLVMS_P03355_3mut
GSSGSSGSSGSSGSS	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGSS	MLVAV_P03356_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	AVIRE_P03360_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	PERV_Q4VFZ2_3mutA_WS
GGSGGGGSS	MLVFF_P26809_3mutA
PAPEAAAKGSS	MLVCB_P08361_3mut
GGG	PERV_Q4VFZ2_3mutA_WS
GGSGGGEAAAK	MLVMS_P03355_3mut
EAAAKGGGGSS	WMSV_P03359_3mut
GSSPAPGGG	WMSV_P03359_3mutA
EAAAKGSSGGG	PERV_Q4VFZ2_3mut
GGSGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGSGGS	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGS	PERV_Q4VFZ2_3mutA_WS
GGGGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GSSPAP	MLVFF_P26809_3mut
GGGEAAAKPAP	AVIRE_P03360_3mut
GSSGGSEAAAK	MLVMS_P03355_PLV919
EAAAKPAPGGS	WMSV_P03359_3mutA
PAPGGG	KORV_Q9TTC1_3mutA
EAAAKGSSPAP	KORV_Q9TTC1-Pro_3mut
GSSPAPEAAAK	MLVFF_P26809_3mut
GGSGGGEAAAK	MLVFF_P26809_3mutA
GSSGSSGSS	WMSV_P03359_3mutA
EAAAKGGS	BAEVM_P10272_3mut
EAAAKPAPGGS	KORV_Q9TTC1_3mutA
EAAAKPAPGGS	BAEVM_P10272_3mutA

[0862]

[0863]

링커 아미노산 서열	RT 도메인 명칭
GSSGGGGGS	PERV_Q4VFZ2_3mut
PAPGGGGSS	PERV_Q4VFZ2_3mut
GSSGSSGSS	WMSV_P03359_3mut
EAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mut
GGS	AVIRE_P03360_3mut
EAAAKPAPGSS	MLVFF_P26809_3mut
EAAAKGGG	KORV_Q9TTC1_3mut
PAPGSSEAAAK	MLVMS_P03355_3mut
PAPGSSGGS	MLVMS_P03355_PLV919
GSSPAPEAAAK	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	WMSV_P03359_3mutA
GGGGS	BAEVM_P10272_3mut
GSSPAP	MLVMS_P03355_3mut
EAAAKGGGGSEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAK	WMSV_P03359_3mutA
GGGGSSGGS	MLVCB_P08361_3mutA
PAPGGSEAAAK	BAEVM_P10272_3mut
EAAAKGGSPAP	MLVFF_P26809_3mut
GSSGGSGGG	MLVBM_Q7SVK7_3mutA_WS
GSSGGS	PERV_Q4VFZ2_3mut
PAPGGSGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	KORV_Q9TTC1-Pro_3mutA
PAPAP	MLVCB_P08361_3mut
EAAAKGSSPAP	PERV_Q4VFZ2_3mutA_WS
EAAAKPAPGGG	MLVMS_P03355_PLV919
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mut
EAAAKGGGGSS	MLVMS_P03355_PLV919
PAPEAAAK	PERV_Q4VFZ2_3mut
EAAAKPAPGSS	BAEVM_P10272_3mutA
GGSPAP	PERV_Q4VFZ2_3mutA_WS
GGSGGS	BAEVM_P10272_3mutA
PAPEAAAKGSS	KORV_Q9TTC1_3mut
PAPGSS	MLVMS_P03355_PLV919
PAPAPAPAPAP	MLVAV_P03356_3mutA
GGG	XMRV6_A1Z651_3mutA
GGGPAP	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	KORV_Q9TTC1_3mutA
PAP	BAEVM_P10272_3mutA
GGSPAP	BAEVM_P10272_3mutA
PAPEAAAKGGS	MLVMS_P03355_PLV919

[0864]

링커 아미노산 서열	RT 도메인 명칭
PAPGSSGGS	PERV_Q4VFZ2_3mutA_WS
PAPAPAPAPAP	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAK	MLVCB_P08361_3mut
GGSGGSGGSGGSGGS	MLVMS_P03355_PLV919
EAAAKPAPGGS	MLVMS_P03355_3mut
GGSGGS	MLVMS_P03355_PLV919
EAAAKPAP	MLVMS_P03355_3mutA_WS
GGSEAAAK	XMRV6_A1Z651_3mutA
GGSGGG	KORV_Q9TTC1_3mut
GGSGGGEAAAK	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	AVIRE_P03360
PAPAP	PERV_Q4VFZ2_3mut
GSS	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGG	MLVAV_P03356_3mutA
GGSPAPGSS	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAK	MLVAV_P03356_3mut
EAAAKGGSPAP	BAEVM_P10272_3mutA
PAPAPAPAP	WMSV_P03359_3mutA
PAPGGSEAAAK	MLVMS_P03355_3mut
GGSGGSGGSGGS	WMSV_P03359_3mut
GGGGGSGSS	XMRV6_A1Z651_3mut
PAPGGSGGG	KORV_Q9TTC1_3mutA
GGS	MLVMS_P03355_3mut
EAAAK	WMSV_P03359_3mut
GGGEAAAKGSS	MLVBM_Q7SVK7_3mutA_WS
GGSPAPGSS	MLVCB_P08361_3mut
GGSEAAAKPAP	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVCB_P08361_3mutA
GGSGSS	BAEVM_P10272_3mutA
GGGEAAAKGSS	WMSV_P03359_3mutA
EAAAKGGSPAP	WMSV_P03359_3mut
GSSPAPEAAAK	MLVMS_P03355_3mut
GGSGGSGGSGGS	MLVMS_P03355_PLV919
GSSPAPEAAAK	WMSV_P03359_3mut
GSSGSSGSSGSS	PERV_Q4VFZ2
GGSGSSEAAAK	WMSV_P03359_3mutA
GGSGGG	MLVFF_P26809_3mut
GGSPAPGGG	MLVFF_P26809_3mut
GGSGGSGGS	BAEVM_P10272_3mutA
GGGGSSEAAAK	MLVBM_Q7SVK7_3mut

링커 아미노산 서열	RT 도메인 명칭
GGSPAPGSS	MLVMS_P03355_3mut
EAAAKPAPGSS	AVIRE_P03360_3mut
GGGGSSGGS	FLV_P10273_3mutA
GGSPAPEAAAK	PERV_Q4VFZ2_3mut
GGSEAAAK	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MLVMS_P03355_PLV919
GGGGG	PERV_Q4VFZ2_3mut
GGSEAAAKGSS	MLVCB_P08361_3mutA
GSSGGG	MLVBM_Q7SVK7_3mutA_WS
PAPGSSGGG	KORV_Q9TTC1-Pro_3mutA
GGSGGS	BAEVM_P10272_3mut
EAAAKGGGGS	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	MLVCB_P08361_3mut
PAPGSSGGG	KORV_Q9TTC1
PAPGGSGGG	MLVMS_P03355_3mut
GGGG	WMSV_P03359_3mutA
EAAAKGGSPAP	MLVCB_P08361_3mut
GSSGSS	FLV_P10273_3mutA
GGSEAAAKPAP	SFV3L_P27401_2mut
EAAAKGSSGGS	MLVAV_P03356_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVAV_P03356_3mutA
EAAAKGGSGSS	PERV_Q4VFZ2_3mutA_WS
GGGGG	MLVCB_P08361_3mut
GGGEAAAK	BAEVM_P10272_3mut
GGSGGSGGSGGS	MLVCB_P08361_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	PERV_Q4VFZ2
PAPAPAPAPAP	MLVMS_P03355_3mutA_WS
EAAAKEAAAK	XMRV6_A1Z651_3mut
GSSGGSEAAAK	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAKGGGPAP	MLVBM_Q7SVK7_3mutA_WS
PAPGGSGSS	PERV_Q4VFZ2
SGSETPGTSESATPES	MLVMS_P03355_3mut
GGSGGS	MLVMS_P03355_PLV919
EAAAKGGS	FLV_P10273_3mut
GGSPAPGSS	MLVMS_P03355_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	FFV_O93209_2mut
GSSGGSGGG	MLVMS_P03355_3mutA_WS
PAPGSSEAAAK	WMSV_P03359_3mut

[0865]

[0866]

링커 아미노산 서열	RT 도메인 명칭
PAPAPAPAPAP	KORV_Q9TTC1_3mutA
GGGGSS	BAEVM_P10272_3mut
GGGGSEAAAKGGGGS	AVIRE_P03360_3mut
GSSPAPEAAAK	KORV_Q9TTC1-Pro_3mutA
PAPEAAAKGGG	MLVBM_Q7SVK7_3mut
EAAAKEAAAK	WMSV_P03359_3mut
EAAAK	SFV3L_P27401-Pro_2mutA
GSSGGSGGG	XMRV6_A1Z651_3mutA
GGGEAAAKPAP	WMSV_P03359_3mutA
GGSGGS	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	FOAMV_P14350_2mutA
GGGGG	MLVAV_P03356_3mutA
GSSGGSEAAAK	BAEVM_P10272_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SFV1_P23074
GGSGGGPAP	MLVCB_P08361_3mut
GGSGSS	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	MLVFF_P26809_3mut
EAAAKGGSPAP	MLVMS_P03355_3mut
PAPAP	PERV_Q4VFZ2_3mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MLVBM_Q7SVK7_3mut
GGGGGS	BAEVM_P10272_3mutA
EAAAKEAAAK	AVIRE_P03360_3mut
GSSGGSEAAAK	PERV_Q4VFZ2_3mut
GGGEAAAK	WMSV_P03359_3mut
GSSGGGEAAAK	AVIRE_P03360_3mutA
GGG	XMRV6_A1Z651_3mut
GGGGSEAAAKGGGGS	BAEVM_P10272_3mut
GGGG	MLVMS_P03355_3mut
GGSGGS	MLVMS_P03355_3mutA_WS
GGSGGGGSS	MLVBM_Q7SVK7_3mutA_WS
GSSPAPGGS	PERV_Q4VFZ2_3mut
GSSPAPEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKGGS	WMSV_P03359_3mut
GGSGGSGGSGGS	PERV_Q4VFZ2_3mut
GGGGSSEAAAK	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAP	MLVAV_P03356_3mut
EAAAKGSSGGG	MLVMS_P03355_PLV919
GGGGG	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	FFV_O93209_2mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	KORV_Q9TTC1-Pro_3mut

[0867]

링커 아미노산 서열	RT 도메인 명칭
GGSPAPGGG	MLVMS_P03355_3mutA_WS
GGGEAAAKGGS	MLVMS_P03355_3mut
GGGEAAAK	PERV_Q4VFZ2_3mut
PAPEAAAKGGG	MLVMS_P03355_3mut
GSSGSSGSSGSSGSSGSS	BAEVM_P10272_3mutA
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	GALV_P21414_3mutA
EAAAKGGSPAP	FFV_O93209-Pro
EAAAKEAAAK	MLVFF_P26809_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGSGGS	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401_2mutA
GSSGSSGSSGSSGSSGSS	BAEVM_P10272_3mut
GGGGS	MLVMS_P03355_PLV919
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SFV1_P23074
GGGGSGGGGS	KORV_Q9TTC1-Pro_3mutA
GGGGSGGGGS	MLVMS_P03355_3mut
GGSGSS	KORV_Q9TTC1_3mutA
GSSPAPGGG	PERV_Q4VFZ2_3mut
GSSGGSPAP	PERV_Q4VFZ2_3mutA_WS
PAPGGS	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	FOAMV_P14350_2mutA
GGGPAPGGS	SFV3L_P27401_2mut
PAPGSSGGG	MLVCB_P08361_3mut
GSSGGGEAAAK	AVIRE_P03360_3mut
GSSGGG	XMRV6_A1Z651_3mut
GSSGSS	PERV_Q4VFZ2_3mut
GSSGGG	MLVAV_P03356_3mutA
PAPGGGGGS	PERV_Q4VFZ2_3mut
GSSEAAAK	MLVMS_P03355_3mut
PAPGGG	FLV_P10273_3mutA
GGGGSGGGGS	PERV_Q4VFZ2_3mut
GSSGGS	MLVMS_P03355_PLV919
GGGGSGGGGS	SFV3L_P27401_2mut
EAAAKGGSGSS	FLV_P10273_3mutA
GSSEAAAKGGS	MLVMS_P03355_3mutA_WS
PAPGSSEAAAK	SFV3L_P27401_2mutA
GGGGSGGGGS	SFV3L_P27401-Pro_2mutA
PAPGSSEAAAK	PERV_Q4VFZ2_3mut
PAPGSSEAAAK	PERV_Q4VFZ2
GGSPAPGGG	AVIRE_P03360_3mut

[0868]

링커 아미노산 서열	RT 도메인 명칭
GGGGGS	PERV_Q4VFZ2_3mutA_WS
GGGGSSGGS	PERV_Q4VFZ2_3mut
PAPAPAPAP	AVIRE_P03360_3mutA
GGSGGS	WMSV_P03359_3mutA
GGGPAPGGS	PERV_Q4VFZ2_3mut
GGSGGSGGSGGSGGS	MLVMS_P03355_PLV919
GGSGGG	PERV_Q4VFZ2_3mut
EAAAKEAAAK	SFV3L_P27401_2mut
PAPGSS	XMRV6_A1Z651_3mut
GSSEAAAK	MLVFF_P26809_3mut
GGSPAPGGG	MLVMS_P03355_3mut
EAAAKGGG	WMSV_P03359_3mutA
GSSEAAAKGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGSPAP	FFV_O93209
GGGGGS	KORV_Q9TTC1-Pro_3mut
GSSGGG	MLVCB_P08361_3mut
GSSGSS	MLVCB_P08361_3mutA
GGSEAAAKPAP	BAEVM_P10272_3mut
EAAAKGGGGSS	MLVCB_P08361_3mut
EAAAKPAPGGS	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSS	MLVAV_P03358_3mutA
GGGGSEAAAKGGGGS	PERV_Q4VFZ2_3mutA_WS
GGSGSS	KORV_Q9TTC1-Pro_3mut
GSS	SFV3L_P27401-Pro_2mutA
PAPAP	BAEVM_P10272_3mut
EAAAKPAP	BAEVM_P10272
EAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mut
GGGGGGG	PERV_Q4VFZ2_3mutA_WS
GGGGS	MLVMS_P03355_3mut
GSSGGG	FLV_P10273_3mutA
PAPAPAPAPAP	FLV_P10273_3mut
EAAAKEAAAKEAAAK	WMSV_P03359_3mutA
GSSGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGGG	MLVMS_P03355_3mut
GSSPAPGGS	WMSV_P03359_3mut
PAPGSSGGG	PERV_Q4VFZ2_3mutA_WS
GSSGGG	AVIRE_P03360_3mutA
PAPGGSGSS	MLVFF_P26809_3mut
PAPGSS	PERV_Q4VFZ2_3mut
GGGGGSGSS	WMSV_P03359_3mutA

[0869]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGGGGSS	MLVBM_Q7SVK7_3mutA_WS
GGGGGGG	BAEVM_P10272_3mut
PAPEAAAKGSS	MLVMS_P03355_3mut
GGSGGGEAAAK	MLVMS_P03355_PLV919
EAAAKGGGGGS	MLVCB_P08361_3mut
PAPGGS	KORV_Q9TTC1-Pro_3mut
GGGG	FLV_P10273_3mutA
EAAAKGGSGSS	MLVBM_Q7SVK7_3mutA_WS
GGGGSSGGS	MLVMS_P03355_3mutA_WS
GGGGGGGG	WMSV_P03359_3mut
GGSGSSGGG	MLVMS_P03355_PLV919
GSSEAAAKGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKPAPGSS	MLVCB_P08361_3mut
GGSPAPGSS	KORV_Q9TTC1_3mutA
PAPGSSGGG	BAEVM_P10272_3mut
EAAAKPAPGSS	WMSV_P03359_3mut
GGSPAPEAAAK	XMRV6_A1Z651_3mutA
GSSPAP	FLV_P10273_3mutA
GSS	BAEVM_P10272_3mutA
EAAAKPAPGGS	FLV_P10273_3mutA
GGSGSSPAP	FLV_P10273_3mutA
PAPGSSGGS	MLVMS_P03355_3mut
GSAGSAAGSGEF	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	KORV_Q9TTC1_3mutA
GSSGGS	MLVMS_P03355_3mutA_WS
EAAAKGGGGSEAAAK	SFV3L_P27401_2mut
GSSGGS	PERV_Q4VFZ2_3mutA_WS
GGSPAPEAAAK	FLV_P10273_3mut
GGSEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	PERV_Q4VFZ2_3mutA_WS
GGSGSSGGG	PERV_Q4VFZ2_3mut
GGGG	AVIRE_P03360_3mutA
GGSEAAAKPAP	WMSV_P03359_3mut
GSSGGSPAP	MLVAV_P03356_3mutA
GSSGGSEAAAK	MLVMS_P03355_3mut
PAPEAAAKGGS	KORV_Q9TTC1-Pro_3mut
GGSPAP	PERV_Q4VFZ2_3mutA_WS
GGSEAAAK	MLVAV_P03356_3mutA
EAAAKGGGGSEAAAK	KORV_Q9TTC1-Pro_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVMS_P03355_PLV919

링커 아미노산 서열	RT 도메인 명칭
GSSEAAAK	KORV_Q9TTC1_3mutA
GGG	AVIRE_P03360
GGSEAAAKGSS	MLVBM_Q7SVK7_3mut
GGSEAAAKGSS	MLVMS_P03355_3mut
GGSPAPEAAAK	MLVCB_P08361_3mut
GGSGGGEAAAK	MLVCB_P08361_3mut
GGSEAAAKPAP	MLVMS_P03355_3mutA_WS
EAAAKGGSGSS	KORV_Q9TTC1-Pro_3mut
GGGEAAAKGGS	MLVCB_P08361_3mut
EAAAKGGGGSEAAAK	FLV_P10273_3mutA
GGSPAP	MLVFF_P26809_3mut
GGGGSSGGS	XMRV6_A1Z651_3mutA
PAP	MLVCB_P08361_3mut
GGS	SFV3L_P27401-Pro_2mutA
GGGGSGGGGS	MLVMS_P03355_3mut
GGGEAAAKGGS	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSSGSS	MLVMS_P03355_PLV919
PAPGSS	MLVCB_P08361_3mut
GGSGGSGGS	MLVMS_P03355_PLV919
PAPGGSGGG	FLV_P10273_3mutA
GGGGSGGGGSGGGGS	FLV_P10273_3mut
GGSGSSGGG	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGS	GALV_P21414_3mutA
GGGEAAAKGGS	WMSV_P03359_3mut
SGSETPGTSESATPES	KORV_Q9TTC1_3mutA
EAAAKGGGGS	KORV_Q9TTC1-Pro_3mut
EAAAKGSSPAP	BAEVM_P10272_3mut
GGGG	MLVCB_P08361_3mut
GGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mut
GSSGGSGGG	MLVMS_P03355_PLV919
GGSGSS	MLVFF_P26809_3mut
EAAAKGGS	AVIRE_P03360_3mutA
GSSEAAAKGGS	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGGG	WMSV_P03359_3mut
PAPGSSGGG	MLVCB_P08361_3mutA
GGGGSSEAAAK	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKPAP	BAEVM_P10272_3mutA
PAPGGGEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGGGEAAAK	MLVCB_P08361_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	FFV_093209

[0870]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGGGGS	GALV_P21414_3mutA
GGSPAPGGG	MLVMS_P03355_3mut
GSSGSSGSS	FLV_P10273_3mutA
EAAAK	MLVBM_Q7SVK7_3mut
GGGGSSGGS	MLVMS_P03355_3mut
GGSGSSPAP	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAK	BAEVM_P10272_3mut
GGGPAPGSS	MLVMS_P03355_3mut
GSSPAPGGS	PERV_Q4VFZ2_3mutA_WS
PAPAP	FLV_P10273_3mutA
PAPAPAPAP	PERV_Q4VFZ2_3mut
GGGGGSEAAAK	GALV_P21414_3mutA
GGGGGSGSS	BAEVM_P10272_3mutA
GGGEAAAKGSS	KORV_Q9TTC1_3mutA
GGGGGSPAP	AVIRE_P03360_3mut
GGGGGSEAAAK	SFV3L_P27401_2mutA
GGS	KORV_Q9TTC1_3mutA
GGGGGGG	PERV_Q4VFZ2_3mut
SGSETPGTSESATPES	SFV3L_P27401_2mutA
EAAAKGGSGGG	MLVMS_P03355_3mut
GGGGS	MLVFF_P26809_3mut
EAAAKGSSGGG	BAEVM_P10272_3mut
EAAAKPAPGGS	MLVF5_P26810_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SFV3L_P27401_2mutA
GGSPAPGGG	WMSV_P03359_3mutA
GSAGSAAGSGEF	MLVFF_P26809_3mut
GGGGSSGGS	MLVMS_P03355_3mutA_WS
GGGGGGG	MLVCB_P08361_3mut
GSSEAAAK	WMSV_P03359_3mut
PAPGSS	FLV_P10273_3mutA
GSSGGG	PERV_Q4VFZ2_3mutA_WS
PAPGGG	MLVFF_P26809_3mut
GGGGGSPAP	MLVMS_P03355_3mut
GGSEAAAK	XMRV6_A1Z651_3mut
GSSGGG	PERV_Q4VFZ2_3mut
GGSGGSGGSGGS	MLVMS_P03355_3mut
PAPAP	AVIRE_P03360_3mut
GGSEAAAK	PERV_Q4VFZ2_3mut
GGGGS	MLVMS_P03355_PLV919
GGGG	BAEVM_P10272_3mutA

[0871]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGGGGSS	MLVCB_P08361_3mutA
EAAAKEAAAKEAAAK	GALV_P21414_3mutA
PAPGGGEAAAK	KORV_Q9TTC1
EAAAKGGSPAP	MLVMS_P03355_3mut
GGSGSSEAAAK	MLVMS_P03355_3mut
GGSPAPEAAAK	FLV_P10273_3mutA
GGGGGGG	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SFV1_P23074_2mutA
EAAAKGSSGGS	MLVMS_P03355_3mut
GSSEAAAKPAP	MLVFF_P26809_3mut
GGGGSS	FLV_P10273_3mutA
EAAAKGGSGGG	AVIRE_P03360_3mutA
GGSGGS	PERV_Q4VFZ2_3mutA_WS
GGGGGSPAP	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAK	XMRV6_A1Z651_3mut
PAPEAAAKGGS	FLV_P10273_3mutA
GSSGGSEAAAK	MLVCB_P08361_3mut
EAAAKGGSGGG	MLVMS_P03355
GGSGGGPAP	MLVMS_P03355_3mut
GGS	XMRV6_A1Z651_3mut
GGSEAAAKPAP	MLVFF_P26809_3mut
EAAAKGGG	MLVMS_P03355_PLV919
GSSGSSGSSGSS	WMSV_P03359_3mut
GGSGSSPAP	PERV_Q4VFZ2_3mut
GGGEAAAK	MLVMS_P03355_3mutA_WS
GSSPAPGGS	KORV_Q9TTC1-Pro_3mutA
GSSEAAAKGGG	SFV3L_P27401_2mut
EAAAKPAPGGS	MLVCB_P08361_3mut
GGSGGGEAAAK	PERV_Q4VFZ2
GGSGSS	MLVCB_P08361_3mut
GGSGGGEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGSGGSGGSGGSGGSGGS	FLV_P10273_3mut
PAPEAAAKGSS	MLVMS_P03355_3mut
EAAAKGSSGGS	WMSV_P03359_3mutA
GGSGSSEAAAK	MLVCB_P08361_3mut
GGSGSSEAAAK	KORV_Q9TTC1_3mutA
GSSGGSGGG	MLVMS_P03355_PLV919
EAAAKGGSGGG	SFV3L_P27401-Pro_2mutA
GGSGGS	AVIRE_P03360_3mutA
GSAGSAAGSGEF	MLVMS_P03355_PLV919

[0872]

링커 아미노산 서열	RT 도메인 명칭
GGSGSS	GALV_P21414_3mutA
GGGG	MLVFF_P26809_3mutA
GGGGSGGGGSGGGGSGGGGS	WMSV_P03359_3mut
SGSETPGTSESATPES	BAEVM_P10272_3mut
EAAAKEAAAKEAAAKEAAAK	FOAMV_P14350_2mutA
GGGEAAAKGGS	FLV_P10273_3mutA
GSSGGSEAAAK	MLVFF_P26809_3mut
EAAAKGGGGSS	MLVAV_P03356_3mut
PAPGGSEAAAK	KORV_Q9TTC1-Pro_3mut
EAAAK	XMRV6_A1Z651_3mut
GSSGSSGSSGSSGSSGSS	PERV_Q4VFZ2_3mut
GGGG	MLVCB_P08361_3mutA
GSSGSS	WMSV_P03359_3mutA
GSSGGSPAP	AVIRE_P03360_3mut
GGSGGSGGS	MLVCB_P08361_3mut
EAAAKGGGPAP	FLV_P10273_3mutA
GGGGSGGGGS	MLVCB_P08361_3mut
GGSEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401_2mutA
GGSGSSEAAAK	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro_2mutA
GSSEAAAKGGS	FLV_P10273_3mutA
GGSGSS	PERV_Q4VFZ2
GGSGSSEAAAK	SFV3L_P27401-Pro_2mutA
GSSGSSGSS	XMRV6_A1Z651_3mutA
EAAAKGSSPAP	KORV_Q9TTC1_3mutA
EAAAKPAP	FLV_P10273_3mutA
GGSGSSEAAAK	KORV_Q9TTC1-Pro_3mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	KORV_Q9TTC1_3mutA
GGGGSGGGGSGGGGS	KORV_Q9TTC1-Pro_3mutA
GGGGGGG	FLV_P10273_3mut
EAAAKGSS	WMSV_P03359_3mut
EAAAKGGGPAP	MLVCB_P08361_3mut
GSSGSS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGGGGGS	MLVFF_P26809_3mut
GGSGGGEAAAK	FLV_P10273_3mutA
PAPGSS	MLVFF_P26809_3mutA
PAPGSS	BAEVM_P10272_3mutA
GGSPAPGSS	AVIRE_P03360_3mut
GGGGSSEAAAK	MLVMS_P03355_3mut

[0873]

[0874]

링커 아미노산 서열	RT 도메인 명칭
GSSGGGGGS	FFV_O93209-Pro
EAAAKGSSPAP	PERV_Q4VFZ2_3mut
GSSPAPGGS	PERV_Q4VFZ2_3mut
GGGGGG	BAEVM_P10272_3mut
EAAAKGGGGSS	PERV_Q4VFZ2_3mutA_WS
PAPGGSEAAAK	KORV_Q9TTC1_3mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSS	MLVMS_P03355_3mut
EAAAKGSSGGG	MLVMS_P03355_PLV919
GGSEAAAKPAP	AVIRE_P03360_3mutA
GSSGSSGSSGSSGSS	WMSV_P03359_3mutA
GGGEAAAKPAP	FLV_P10273_3mutA
PAPGSSGGG	KORV_Q9TTC1_3mutA
GSSGSS	MLVMS_P03355_3mutA_WS
PAPEAAAK	BAEVM_P10272_3mut
GGGPAPGSS	PERV_Q4VFZ2
GSSGGSPAP	MLVFF_P26809_3mut
GGGGSS	SFV3L_P27401_2mut
PAPEAAAKGSS	SFV3L_P27401_2mut
GGSGGGPAP	XMRV6_A1Z651_3mutA
PAPGGS	BAEVM_P10272_3mutA
EAAAKGGGGS	AVIRE_P03360_3mut
GSSGGSPAP	KORV_Q9TTC1-Pro_3mutA
GSSGGGGGS	WMSV_P03359_3mut
GGGEAAAKGGS	AVIRE_P03360_3mut
GGGEAAAKGSS	BAEVM_P10272_3mut
PAPEAAAKGSS	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	MLVCB_P08361_3mut
GGSPAPGSS	FLV_P10273_3mutA
EAAAKGSSPAP	BAEVM_P10272_3mutA
GGSGGSGGSGGSGGSGGS	PERV_Q4VFZ2
GGGGSSEAAAK	FLV_P10273_3mutA
GGGGSSPAP	FFV_O93209
GSSGGSPAP	MLVMS_P03355_3mut
GGGPAPGSS	MLVMS_P03355_PLV919
PAPGSSGGS	PERV_Q4VFZ2_3mut
GGGGGSPAP	MLVFF_P26809_3mut
SGSETPGTSESATPES	MLVMS_P03355_3mutA_WS
GSSGSSGSSGSSGSS	KORV_Q9TTC1_3mutA
GSSPAPGGG	WMSV_P03359_3mut

링커 아미노산 서열	RT 도메인 명칭
PAPAPAPAPAP	SFV3L_P27401_2mutA
GGGPAPGGS	MLVMS_P03355_3mut
PAPGGSEAAAK	WMSV_P03359_3mut
GGGGSSEAAAK	FFV_O93209-Pro
GGSPAPGGG	FLV_P10273_3mutA
GSSPAPEAAAK	AVIRE_P03360_3mut
GGGEAAAK	FLV_P10273_3mutA
PAPEAAAKGGG	MLVCB_P08361_3mut
GGSPAPGGG	MLVCB_P08361_3mut
GGSGGGGSS	BAEVM_P10272_3mutA
GSSPAPEAAAK	MLVCB_P08361_3mut
GGSPAPGGG	KORV_Q9TTC1-Pro_3mutA
PAPGGSGSS	KORV_Q9TTC1_3mutA
GSSPAP	KORV_Q9TTC1-Pro_3mutA
SGSETPGTSESATPES	MLVMS_P03355
GSSGSSGSS	MLVAV_P03356_3mutA
PAPGSSGGS	PERV_Q4VFZ2_3mutA_WS
PAPGGS	KORV_Q9TTC1-Pro_3mutA
PAPEAAAKGGG	SFV3L_P27401-Pro_2mutA
GGSGGSGGS	BAEVM_P10272_3mut
PAPGGS	MLVFF_P26809_3mut
GSSGGSPAP	MLVMS_P03355_PLV919
GSSGGGGGS	FLV_P10273_3mutA
GGGGGSPAP	KORV_Q9TTC1-Pro_3mut
EAAAKPAPGSS	SFV3L_P27401-Pro_2mutA
EAAAKGGSPAP	KORV_Q9TTC1-Pro
GGGPAPEAAAK	MLVMS_P03355_PLV919
GGSEAAAKGSS	MLVMS_P03355
PAPEAAAKGSS	KORV_Q9TTC1_3mutA
PAPEAAAKGGS	WMSV_P03359_3mutA
GSSGGG	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGSS	MLVMS_P03355_PLV919
EAAAKGGSPAP	AVIRE_P03360_3mutA
GGGGSSGGS	MLVMS_P03355_PLV919
PAPEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGGGGS	BAEVM_P10272_3mut
GSSGGGGGS	MLVMS_P03355_3mut
PAPAPAPAP	KORV_Q9TTC1_3mutA
GGSGGSGGSGGS	MLVAV_P03356_3mut
PAPAPAPAP	SFV3L_P27401_2mut

[0875]

링커 아미노산 서열	RT 도메인 명칭
GSSEAAAKPAP	MLVMS_P03355_3mut
GGSGGGEAAAK	SFV3L_P27401_2mutA
GSSGGSGGG	MLVMS_P03355_3mutA_WS
GGGGGSPAP	MLVCB_P08361_3mutA
GGGEAAAKGSS	XMRV6_A1Z651_3mutA
GGGGSSPAP	BAEVM_P10272_3mut
GGSGGG	PERV_Q4VFZ2_3mut
GGGGSS	MLVBM_Q7SVK7_3mutA_WS
EAAAKGSSGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGGGGS	PERV_Q4VFZ2
EAAAKGSSGGS	PERV_Q4VFZ2_3mut
EAAAKEAAAK	MLVAV_P03356_3mut
GSSGGGEAAAK	MLVAV_P03356_3mut
GSSPAPGGG	XMRV6_A1Z651_3mut
GGGGSGGGGSGGGGS	PERV_Q4VFZ2_3mut
EAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1_3mutA
EAAAKGGSGSS	MLVBM_Q7SVK7_3mut
PAPEAAAK	BLVJ_P03361
GSSGGG	FFV_O93209-Pro
GGSGGGEAAAK	KORV_Q9TTC1-Pro_3mutA
EAAAK	FLV_P10273_3mutA
GGGGSSPAP	MLVMS_P03355_3mut
GSS	SFV3L_P27401-Pro_2mut
PAPEAAAKGSS	BAEVM_P10272_3mut
GGGGGSPAP	PERV_Q4VFZ2_3mut
GSSGSSGSS	BAEVM_P10272_3mutA
GGGGSGGGGSGGGGSGGGGS	SFV1_P23074_2mut
GGGGSSEAAAK	SFV3L_P27401_2mutA
GGGGSGGGGSGGGGSGGGGS	FOAMV_P14350-Pro_2mut
PAPGSSEAAAK	MLVBM_Q7SVK7_3mutA_WS
GGGGGSGSS	MLVFF_P26809_3mutA
GGSEAAAKGGG	MLVBM_Q7SVK7_3mut
PAPGSSGGG	PERV_Q4VFZ2
GGS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	FLV_P10273_3mut
GGGEAAAK	WMSV_P03359_3mutA
GGSEAAAKPAP	MLVBM_Q7SVK7_3mut
SGSETPGTSESATPES	FOAMV_P14350-Pro_2mutA
EAAAKPAPGGS	AVIRE_P03360_3mut
EAAAKGGGGGS	KORV_Q9TTC1-Pro_3mutA

[0876]

[0877]

링커 아미노산 서열	RT 도메인 명칭
GGGGS	PERV_Q4VFZ2_3mut
GGSEAAAKGSS	MLVFF_P26809_3mutA
GGSEAAAKGGG	AVIRE_P03360
GGSGGSGGSGGSGGSGGS	SFV3L_P27401_2mut
GGSEAAAKGSS	SFV3L_P27401-Pro_2mutA
GGGEAAAKPAP	MLVCB_P08361_3mut
GGSEAAAK	MLVMS_P03355_PLV919
GGSPAPGSS	KORV_Q9TTC1-Pro_3mutA
GSSPAPEAAAK	WMSV_P03359_3mutA
GGSGSS	KORV_Q9TTC1-Pro_3mutA
PAPGGGGGS	AVIRE_P03360_3mut
PAPEAAAKGSS	FFV_O93209-Pro
GGSGGGEAAAK	WMSV_P03359_3mut
PAPGGG	MLVMS_P03355_3mut
EAAAKGGG	FLV_P10273_3mutA
GSSGSSGSSGSS	MLVCB_P08361_3mut
EAAAKGGSGGG	FFV_O93209
GSSPAPGGS	PERV_Q4VFZ2_3mutA_WS
GSSPAPGGS	MLVCB_P08361_3mut
GGGPAP	WMSV_P03359_3mutA
GGGPAP	KORV_Q9TTC1_3mutA
GGSPAPGSS	KORV_Q9TTC1-Pro_3mut
PAPAP	MLVMS_P03355_3mut
GGGGGGG	MLVMS_P03355_3mut
GGGGG	KORV_Q9TTC1-Pro_3mut
GSAGSAAGSGEF	FOAMV_P14350_2mutA
PAPAP	KORV_Q9TTC1-Pro_3mutA
GGSEAAAKGGG	SFV3L_P27401-Pro_2mutA
PAPAP	WMSV_P03359_3mut
GGGGSGGGGSGGGGS	SFV3L_P27401_2mut
PAPGGS	KORV_Q9TTC1_3mutA
GGGEAAAKPAP	FLV_P10273_3mut
GGGGGS	MLVAV_P03356_3mutA
GSSEAAAKGGG	WMSV_P03359_3mut
EAAAKGGGGSS	GALV_P21414_3mutA
GSSGGS	MLVAV_P03356_3mutA
GSSGGG	MLVBM_Q7SVK7_3mut
PAPAPAP	SFV3L_P27401-Pro_2mutA
GGGG	KORV_Q9TTC1_3mutA
EAAAKPAPGGS	MLVFF_P26809_3mut

링커 아미노산 서열	RT 도메인 명칭
GGGGSGGGGS	XMRV6_A1Z651_3mut
EAAAKGGG	MLVCB_P08361_3mut
GGGGSSPAP	KORV_Q9TTC1_3mutA
GSSEAAAKGGG	KORV_Q9TTC1-Pro_3mutA
GGGGG	BLVJ_P03361_2mutB
GGGEAAAKGSS	FFV_O93209-Pro
GSSGSSGSS	BAEVM_P10272_3mut
GSSGGSPAP	PERV_Q4VFZ2_3mut
EAAAKGGS	KORV_Q9TTC1_3mut
GGSPAPEAAAK	AVIRE_P03360_3mut
GGSEAAAK	WMSV_P03359_3mut
GSSGGS	KORV_Q9TTC1-Pro_3mutA
GGGPAPEAAAK	KORV_Q9TTC1_3mutA
PAPGSS	WMSV_P03359_3mutA
GGSEAAAKGSS	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401
GSSEAAAKGGG	SFV3L_P27401-Pro_2mutA
GGGGSEAAAKGGGGS	KORV_Q9TTC1-Pro_3mutA
GGSGGSGGS	WMSV_P03359_3mut
GGGGGSGSS	KORV_Q9TTC1-Pro
GGGGSGGGGSGGGGSGGGGS	MLVMS_P03355_3mut
EAAAKGGG	PERV_Q4VFZ2
GGSEAAAKGGG	KORV_Q9TTC1-Pro_3mut
GSSGGSGGGG	PERV_Q4VFZ2_3mutA_WS
GGGGGS	PERV_Q4VFZ2_3mut
GSAGSAAGSGEF	PERV_Q4VFZ2
PAPEAAAKGSS	BAEVM_P10272_3mutA
GSSPAPGGG	MLVCB_P08361_3mut
GGGGSSPAP	KORV_Q9TTC1-Pro_3mutA
PAPGGSGGG	MLVFF_P26809_3mut
GSSPAP	KORV_Q9TTC1_3mutA
PAPGSS	SFV3L_P27401-Pro_2mut
GGSGGGGSS	MLVMS_P03355_PLV919
GSSGGS	WMSV_P03359_3mutA
EAAAKGGGGGS	PERV_Q4VFZ2
GGGGG	KORV_Q9TTC1_3mutA
EAAAKGSS	MLVMS_P03355_PLV919
EAAAKEAAAKEAAAKEAAAKEAAAK	FLV_P10273_3mut
EAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro_2mut
GSAGSAAGSGEF	SFV3L_P27401_2mutA

[0878]

[0879]

링커 아미노산 서열	RT 도메인 명칭
GGGPAPGGS	FLV_P10273_3mutA
GGSEAAAKGGG	MLVCB_P08361_3mut
PAPGGGEAAAK	BAEVM_P10272_3mut
EAAAKPAPGSS	FOAMV_P14350_2mut
GGSEAAAK	KORV_Q9TTC1_3mutA
GGSGSS	AVIRE_P03360
GGSPAPEAAAK	MLVMS_P03355_PLV919
GGGGS	XMRV6_A1Z651_3mut
GGSPAPGGG	XMRV6_A1Z651_3mut
EAAAKPAPGGS	PERV_Q4VFZ2
GGSPAP	BAEVM_P10272_3mut
GGSGSSGGG	FLV_P10273_3mutA
PAPGGG	PERV_Q4VFZ2_3mutA_WS
GSSGGSEAAAK	MLVBM_Q7SVK7_3mut
GGSEAAAK	MLVMS_P03355_3mut
GGGPAPGGS	MLVFF_P26809_3mut
GSAGSAAGSGEF	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGGS	SFVCP_Q87040
PAPGGG	PERV_Q4VFZ2_3mutA_WS
GSSPAPEAAAK	MLVBM_Q7SVK7
PAPEAAAK	MLVBM_Q7SVK7_3mut
PAPGGGGGS	AVIRE_P03360_3mutA
GGSEAAAKPAP	MLVBM_Q7SVK7_3mut
EAAAKGSS	WMSV_P03359_3mutA
GGGEAAAK	MLVFF_P26809_3mutA
EAAAKEAAAKEAAAK	MLVMS_P03355_3mut
PAPEAAAKGGG	BAEVM_P10272_3mut
PAPAPAP	MLVCB_P08361_3mut
EAAAKPAPGGS	BAEVM_P10272_3mut
GGGGSGGGGS	FLV_P10273_3mut
GGGGSEAAAKGGGGS	KORV_Q9TTC1_3mut
EAAAK	FLV_P10273_3mut
PAPAPAP	WMSV_P03359_3mut
GGGGSEAAAKGGGGS	FFV_O93209-Pro
GGSPAPEAAAK	MLVMS_P03355_3mut
GGSGSSGGG	XMRV6_A1Z651_3mut
GGSPAPGSS	PERV_Q4VFZ2_3mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SFV3L_P27401-Pro_2mutA
EAAAKGGGPAP	BAEVM_P10272_3mutA
GSSGGSEAAAK	MLVMS_P03355_3mutA_WS

링커 아미노산 서열	RT 도메인 명칭
SGSETPGTSESATPES	PERV_Q4VFZ2_3mutA_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSS	KORV_Q9TTC1_3mutA
GSSPAPGGG	SFV3L_P27401-Pro_2mutA
GSSGGGEAAAK	KORV_Q9TTC1_3mutA
GGSGGGGSS	PERV_Q4VFZ2_3mutA_WS
GSSGGGEAAAK	MLVCB_P08361_3mut
GSSEAAAKGGG	MLVCB_P08361_3mut
GGSGGGGSS	KORV_Q9TTC1_3mutA
GGSGGSPAP	PERV_Q4VFZ2_3mutA_WS
GSSPAP	MLVMS_P03355_3mut
GGGGSSEAAAK	AVIRE_P03360
GGS	WMSV_P03359_3mut
EAAAKEAAAK	PERV_Q4VFZ2_3mut
PAPAPAPAP	MLVAV_P03356_3mut
GGSEAAAKGGG	KORV_Q9TTC1_3mutA
PAPGGG	MLVAV_P03356_3mut
EAAAKGSS	BAEVM_P10272_3mut
GGGGSGGGGS	WMSV_P03359_3mutA
GGSGGSGGS	SFV3L_P27401_2mut
EAAAK	MLVCB_P08361_3mut
GGGGSSGGS	WMSV_P03359_3mutA
GGGPAPEAAAK	MLVAV_P03356_3mutA
EAAAKEAAAKEAAAK	FFV_O93209
GSSEAAAKGGG	MLVBM_Q7SVK7_3mut
GGGPAPGGS	FLV_P10273_3mut
GGSEAAAKGGG	WMSV_P03359_3mut
EAAAKGGGGGS	XMRV6_A1Z651_3mutA
EAAAKGGSGGG	FLV_P10273_3mutA
GGSEAAAKGGG	SFV3L_P27401_2mutA
GGGGS	PERV_Q4VFZ2_3mutA_WS
GSSGGS	MLVMS_P03355_3mut
GSSGSS	MLVAV_P03356_3mutA
GGSPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GSSGGGGGS	MLVF5_P26810_3mut
PAPAPAPAP	MLVCB_P08361_3mut
PAPAP	PERV_Q4VFZ2_3mutA_WS
PAPGSSGGS	KORV_Q9TTC1_3mut
PAPGSSGGG	PERV_Q4VFZ2_3mut
GGGEAAAK	MLVMS_P03355_PLV919

[0880]

링커 아미노산 서열	RT 도메인 명칭
GGSGGSGGSGGSGGS	SFV3L_P27401-Pro_2mutA
GGSGGG	FLV_P10273_3mut
PAPEAAAKGGG	MLVFF_P26809_3mut
PAP	PERV_Q4VFZ2_3mutA_WS
PAPGGSGSS	FFV_O93209_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	FFV_O93209-Pro_2mut
GSSGSSGSSGSS	FFV_O93209-Pro
GSSGSSGSSGSSGSS	FLV_P10273_3mutA
GGGEAAAKPAP	PERV_Q4VFZ2
PAPGSSGGG	SFV3L_P27401_2mut
PAPGGSGSS	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAP	GALV_P21414_3mutA
GGSGGGEAAAK	PERV_Q4VFZ2_3mut
GSSPAP	MLVCB_P08361_3mut
EAAAKPAP	MLVF5_P26810_3mut
GGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7_3mut
GGSGGG	WMSV_P03359_3mut
GGSGGSGGS	KORV_Q9TTC1_3mut
GGGGGGGG	MLVFF_P26809_3mut
GGGGSS	MLVAV_P03356_3mut
GSSGGGGGS	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	GALV_P21414_3mutA
GSSGSSGSS	PERV_Q4VFZ2_3mut
GSSPAPGGS	MLVFF_P26809_3mut
PAPAPAP	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAK	WMSV_P03359_3mutA
PAPAPAPAP	SFV3L_P27401_2mutA
GGGGSS	MLVAV_P03356_3mutA
GSSGSSGSSGSSGSS	SFV3L_P27401_2mutA
PAPGGS	WMSV_P03359_3mutA
GSSEAAAKGGG	PERV_Q4VFZ2
GSSGGSPAP	MLVMS_P03355_PLV919
GSSGSSGSSGSSGSSGSS	SFV3L_P27401_2mutA
GGSGSSGGG	MLVCB_P08361_3mut
GGGPAPGSS	SFV3L_P27401-Pro_2mutA
GSSEAAAKGGS	WMSV_P03359_3mut
GSSEAAAKGGG	MLVAV_P03356_3mut
GGSGGGPAP	FFV_O93209-Pro
GSSGSS	PERV_Q4VFZ2_3mut
PAPGGGGGS	GALV_P21414_3mutA

[0881]

[0882]

링커 아미노산 서열	RT 도메인 명칭
EAAAKPAPGGS	MLVAV_P03356_3mut
GSSGSS	MLVMS_P03355_3mut
EAAAKPAPGGS	FFV_O93209-Pro
GGGPAPEAAAK	MLVMS_P03355_3mutA_WS
GSSEAAAKGGG	MLVBM_Q7SVK7_3mut
GGGEAAAKGGS	BAEVM_P10272_3mut
GSSGSS	KORV_Q9TTC1-Pro_3mutA
EAAAKEAAAKEAAAK	SFV1_P23074
PAPGSSGGS	KORV_Q9TTC1-Pro_3mut
PAPAPAPAPAP	MLVMS_P03355
GSSEAAAK	SFV3L_P27401_2mut
PAP	PERV_Q4VFZ2_3mut
GGSEAAAKGGG	MLVBM_Q7SVK7_3mut
GGSGGGPAP	MLVBM_Q7SVK7_3mutA_WS
GSSGSS	MLVMS_P03355_3mut
GGSEAAAK	MLVMS_P03355
GSSEAAAKGGS	MLVMS_P03355_PLV919
PAPGGGGGS	MLVFF_P26809_3mut
GSSGGG	PERV_Q4VFZ2_3mut
GSSGGS	PERV_Q4VFZ2_3mutA_WS
PAPGGG	BAEVM_P10272_3mut
PAPGSSGGG	MLVBM_Q7SVK7_3mut
GGSEAAAK	SFV3L_P27401_2mut
GSSPAPEAAAK	SFV3L_P27401-Pro_2mut
GSSGGSPAP	BAEVM_P10272_3mut
GGSPAPGSS	PERV_Q4VFZ2_3mutA_WS
GGSGGSGGS	PERV_Q4VFZ2
GGSGGGPAP	FLV_P10273_3mut
GGGPAPEAAAK	SFV3L_P27401_2mutA
GGGGS	FLV_P10273_3mutA
GSSGGSGGG	XMRV6_A1Z651_3mut
EAAAKGGGGSS	PERV_Q4VFZ2
GGSGSSGGG	SFV3L_P27401-Pro_2mutA
GGSGGSGGS	MLVFF_P26809_3mut
GGGPAPEAAAK	FLV_P10273_3mut
GSSGGGEAAAK	MLVMS_P03355_3mut
GGG	SFV3L_P27401_2mut
GSAGSAAGSGEF	WMSV_P03359_3mut
GSSGGGPAP	MLVMS_P03355_PLV919
GGGGSS	KORV_Q9TTC1-Pro_3mut

[0883]

링커 아미노산 서열	RT 도메인 명칭
GGCGSSEAAAK	KORV_Q9TTC1
PAPGGSGGG	SFV3L_P27401_2mut
GSSGSSGSSGSSGSS	FFV_O93209
GSSGGSPAP	MLVMS_P03355_3mut
GGSEAAAK	KORV_Q9TTC1-Pro_3mutA
GGCGSGGGGS	BAEVM_P10272_3mut
GSSEAAAKGGG	AVIRE_P03360_3mut
EAAAKPAPGGG	FLV_P10273_3mut
EAAAKGGSPAP	SFV3L_P27401-Pro_2mutA
GSSEAAAKPAP	MLVBM_Q7SVK7_3mut
GGCPAPGGS	MLVCB_P08361_3mut
GGG	SFV3L_P27401_2mutA
EAAAKGGGGSEAAAK	SFV3L_P27401_2mutA
GGSGSSGGG	MLVBM_Q7SVK7_3mut
GSAGSAAGSGEF	BAEVM_P10272_3mut
GGGEAAAK	FOAMV_P14350_2mutA
PAPEAAAKGGS	WMSV_P03359_3mut
PAPAPAPAPAP	MLVF5_P26810_3mutA
GGSGGGGSS	FLV_P10273_3mutA
PAPGSSGGS	BAEVM_P10272_3mut
PAPEAAAK	WMSV_P03359_3mutA
GSSGSSGSSGSSGSSGSS	FFV_O93209-Pro_2mut
GGGGGSGSS	FFV_O93209-Pro
GGGGGGGG	SFV3L_P27401-Pro_2mutA
GGGGGG	FLV_P10273_3mut
GSSGGSGGG	MLVAV_P03356_3mutA
GGGGSS	SFV3L_P27401-Pro_2mutA
GGSGGGPAP	FOAMV_P14350_2mut
GSSGSS	AVIRE_P03360_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro_2mutA
EAAAKEAAAK	BAEVM_P10272_3mut
GSSPAPEAAAK	GALV_P21414_3mutA
GGSEAAAKPAP	SFV3L_P27401_2mutA
GGSGGGEAAAK	SFV3L_P27401-Pro_2mutA
EAAAKGSSPAP	FOAMV_P14350_2mut
GGSGSSEAAAK	SFV3L_P27401_2mut
GGG	PERV_Q4VFZ2
GGGGGSGSS	FOAMV_P14350_2mut
GGSGGGEAAAK	KORV_Q9TTC1-Pro_3mut
GSSGGSGGG	AVIRE_P03360_3mutA

링커 아미노산 서열	RT 도메인 명칭
EAAAKPAPGGG	SFV3L_P27401_2mutA
PAPGGSGGG	KORV_Q9TTC1-Pro_3mut
PAPAPAP	WMSV_P03359_3mutA
GSSEAAAKPAP	SFV1_P23074
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SRV2_P51517
GSSGGSGGG	PERV_Q4VFZ2_3mutA_WS
GSSGSSGSSGSSGSSGSS	FFV_O93209
GSSGGGPAP	WMSV_P03359_3mut
PAPAPAPAPAPAP	MLVBM_Q7SVK7_3mut
GGGGGSPAP	KORV_Q9TTC1-Pro_3mutA
PAPGSS	MLVBM_Q7SVK7_3mutA_WS
PAPEAAAKGGS	SFV3L_P27401-Pro_2mut
GGGGSSPAP	MLVMS_P03355_3mut
GGSEAAAK	FFV_O93209-Pro
EAAAKPAPGGS	AVIRE_P03360_3mutA
PAPGSS	WMSV_P03359_3mut
PAPGSSGGG	SFV3L_P27401-Pro_2mutA
EAAAKEAAAKEAAAK	SFV3L_P27401_2mut
GGS	MLVRD_P11227_3mut
GGGGS	KORV_Q9TTC1-Pro_3mut
GGSGGGGSS	KORV_Q9TTC1
GGSGGG	MLVMS_P03355_3mutA_WS
GGGEAAAKPAP	BAEVM_P10272_3mut
EAAAKEAAAKEAAAKEAAAKEAAAK	FLV_P10273
PAPGGSGGG	KORV_Q9TTC1-Pro_3mutA
GSSGSSGSSGSSGSSGSS	HTL1L_P0C211
GGGEAAAKPAP	WMSV_P03359
GSSGGSPAP	FFV_O93209-Pro
PAPAPAPAPAP	SFV3L_P27401-Pro_2mutA
GSSGGSEAAAK	SFV3L_P27401_2mutA
GGSPAPGSS	SFV3L_P27401_2mut
GGSGGSGGS	KORV_Q9TTC1-Pro_3mut
PAPEAAAKGSS	KORV_Q9TTC1-Pro_3mut
EAAAKGGS	KORV_Q9TTC1_3mutA
EAAAKGGGGSEAAAK	SFV3L_P27401-Pro_2mut
GGGGSSPAP	FFV_O93209-Pro
EAAAK	SFV3L_P27401_2mut
EAAAKGGGGSS	BAEVM_P10272_3mut
GGGGGSEAAAK	MLVBM_Q7SVK7_3mut
GGGG	PERV_Q4VFZ2

[0884]

[0885]

링커 아미노산 서열	RT 도메인 명칭
GGGGGSEAAAK	FLV_P10273_3mut
EAAAKGGGPAP	KORV_Q9TTC1-Pro
GGGGSGGGGSGGGGSGGGGS	FFV_O93209_2mutA
GSSGGSGGG	PERV_Q4VFZ2_3mut
GGGGSGGGGSGGGGS	GALV_P21414_3mutA
GGSGGGEAAAK	AVIRE_P03360_3mutA
PAPEAAAKGGG	SFV3L_P27401_2mut
GGGGSGGGGS	AVIRE_P03360
GSSGGGEAAAK	SFV3L_P27401_2mutA
GGGGG	AVIRE_P03360_3mutA
GGSGSS	KORV_Q9TTC1_3mut
PAPAPAPAPAPAP	FOAMV_P14350_2mut
GGSEAAAKPAP	KORV_Q9TTC1-Pro_3mut
GGGGGG	PERV_Q4VFZ2_3mut
GSSGGGEAAAK	MLVBM_Q7SVK7
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVAV_P03356
GGSPAPGSS	BAEVM_P10272_3mut
GGGGSSPAP	BAEVM_P10272
GGGGSEAAAKGGGGS	SFV3L_P27401_2mut
GGGGGGGG	GALV_P21414_3mutA
PAPAP	MLVAV_P03356_3mut
GGGEAAAK	PERV_Q4VFZ2_3mutA_WS
GSSPAPGGG	FFV_O93209_2mut
GGSGGSGGSGGSGGS	BAEVM_P10272
GGGGGS	MLVF5_P26810_3mutA
PAPGGGGSS	FLV_P10273_3mutA
GGGEAAAK	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	WMSV_P03359_3mut
GSSEAAAK	MLVBM_Q7SVK7_3mut
EAAAKEAAAK	AVIRE_P03360
EAAAKGGGGS	MLVBM_Q7SVK7_3mut
GGGEAAAKGGS	SFV3L_P27401-Pro_2mutA
PAPAPAPAPAP	MLVF5_P26810_3mut
PAPGSSEAAAK	SFV3L_P27401-Pro_2mutA
EAAAKEAAAKEAAAK	BAEVM_P10272_3mutA
GGSPAPGSS	MLVMS_P03355
PAPGSSGGS	FLV_P10273_3mutA
EAAAKEAAAKEAAAKEAAAK	FOAMV_P14350-Pro_2mut
EAAAKGGG	KORV_Q9TTC1_3mutA
EAAAKGGSGGG	MLVBM_Q7SVK7_3mut

링커 아미노산 서열	RT 도메인 명칭
GGGGGS	KORV_Q9TTC1-Pro_3mutA
PAPGGSGGG	WMSV_P03359_3mut
GGGPAPGGS	KORV_Q9TTC1_3mutA
GSS	FFV_O93209
GGSGGSGGG	PERV_Q4VFZ2_3mut
GGGGS	GALV_P21414_3mutA
GGGG	MLVF5_P26810_3mut
GGSEAAAKPAP	FFV_O93209-Pro_2mut
PAPAPAPAP	FFV_O93209-Pro
PAP	MLVF5_P26810_3mut
EAAAKEAAAKEAAAK	FFV_O93209_2mut
EAAAKGSS	MLVCB_P08361_3mut
EAAAKGGG	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	FFV_O93209_2mut
GSSGGGEAAAK	SFV1_P23074-Pro_2mut
PAPGGGEAAAK	GALV_P21414_3mutA
GGGGSGGGGSGGGGSGGGGS	FOAMV_P14350-Pro_2mutA
GSSGGG	FOAMV_P14350_2mut
GGGGSGGGGSGGGGSGGGGS	SFV3L_P27401_2mutA
GGSGSS	AVIRE_P03360_3mut
GGSGSSEAAAK	MMTVB_P03365_WS
PAPAPAP	MLVAV_P03356_3mutA
GSSGGSPAP	SFV3L_P27401-Pro_2mut
GGSPAP	AVIRE_P03360
GGSGGGPAP	FFV_O93209
GSSEAAAK	PERV_Q4VFZ2
GSSGGGPAP	PERV_Q4VFZ2_3mutA_WS
GGGGSSEAAAK	KORV_Q9TTC1_3mutA
GGSEAAAKPAP	SFVCP_Q87040
GGSGGGPAP	FOAMV_P14350_2mutA
GGGGSGGGGSGGGGSGGGGS	BLVJ_P03361_2mutB
GGGGSSPAP	SFV3L_P27401_2mutA
EAAAKGGS	MLVF5_P26810_3mut
GGSEAAAKGSS	MLVCB_P08361_3mut
GGGGSSEAAAK	SFV3L_P27401_2mut
EAAAKGGSGGG	FOAMV_P14350_2mut
GGSGGS	FLV_P10273_3mut
EAAAKGGG	FFV_O93209-Pro
GSSGSSGSSGSSGSS	SFV3L_P27401
GSSGGGPAP	PERV_Q4VFZ2_3mutA_WS

[0886]

링커 아미노산 서열	RT 도메인 명칭
PAPGGSEAAAK	SFV3L_P27401-Pro_2mutA
GGSPAP	KORV_Q9TTC1
EAAAKPAPGSS	KORV_Q9TTC1_3mutA
SGSETPGTSESATPES	SFV1_P23074
GSSPAP	SFV3L_P27401-Pro_2mutA
GSSPAPGGG	SFV3L_P27401_2mut
GGGEAAAKGSS	SFV1_P23074_2mut
GGGPAPGGS	BAEVM_P10272_3mut
EAAAKGGG	KORV_Q9TTC1-Pro_3mutA
GSSGGG	SFV3L_P27401-Pro_2mut
GGSPAPEAAAK	BAEVM_P10272_3mut
EAAAKGSSPAP	FFV_093209
EAAAKGGGGSEAAAK	SFV3L_P27401-Pro_2mutA
GSSGSSGSSGSSGSS	SFV1_P23074_2mut
EAAAKGGSPAP	FOAMV_P14350_2mut
GGSGGS	KORV_Q9TTC1-Pro_3mutA
EAAAKGSSGGS	GALV_P21414
GSSGGGPAP	MLVAV_P03356
PAPEAAAKGGS	FOAMV_P14350_2mut
EAAAKPAPGGG	AVIRE_P03360_3mut
GGSPAP	SFV3L_P27401_2mutA
GGGGSGGGGS	SFV3L_P27401_2mutA
GGGGSS	AVIRE_P03360_3mutA
GGSPAPGGG	SFV3L_P27401-Pro_2mutA
EAAAKPAPGSS	SFV3L_P27401
EAAAKPAP	FOAMV_P14350-Pro_2mut
PAPEAAAKGSS	PERV_Q4VFZ2_3mutA_WS
EAAAKGGSGSS	SFV3L_P27401_2mutA
GGGEAAAKGSS	GALV_P21414_3mutA
GGGGSEAAAKGGGGS	PERV_Q4VFZ2_3mut
PAPGGSGSS	FFV_093209-Pro_2mutA
GGSEAAAKPAP	GALV_P21414_3mutA
GGSGGSGGSGGSGGS	FFV_093209-Pro
GSSGGSEAAAK	SFV3L_P27401-Pro_2mut
GGS	GALV_P21414_3mutA
PAPGGSEAAAK	MLVMS_P03355
PAPEAAAKGGS	BAEVM_P10272_3mutA
GGSGSSPAP	SFV3L_P27401-Pro_2mutA
GSSPAP	WMSV_P03359_3mut
GGGEAAAK	MMTVB_P03365

[0887]

[0888]

링커 아미노산 서열	RT 도메인 명칭
GGGGSS	PERV_Q4VFZ2_3mut
GGSPAPGSS	SFV3L_P27401-Pro_2mut
PAPGGS	MLVBM_Q7SVK7_3mut
EAAAKGSSPAP	MLVBM_Q7SVK7_3mut
GGGGSSGGS	PERV_Q4VFZ2_3mut
PAPAPAPAPAP	SFV1_P23074
GGSEAAAKGGG	SFV3L_P27401-Pro_2mut
GGSGGS	SFV1_P23074_2mut
GSSGGGGGS	MLVF5_P26810_3mutA
EAAAKGGGPAP	SFV3L_P27401
EAAAKEAAAKEAAAKEAAAK	FOAMV_P14350-Pro_2mutA
GGGPAPGSS	SFV3L_P27401_2mutA
GGGGSGGGGSGGGGSGGGGS	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAK	MMTVB_P03365_WS
PAPGSSGGS	KORV_Q9TTC1-Pro_3mutA
PAPGSSEAAAK	FOAMV_P14350-Pro_2mut
GSSPAPEAAAK	BAEVM_P10272_3mut
EAAAKGGGSEAAAK	FFV_O93209-Pro
GGSPAP	PERV_Q4VFZ2
GGSGSSEAAAK	XMRV6_A1Z651_3mut
GGSEAAAKGGG	GALV_P21414_3mutA
PAPGGGGSS	AVIRE_P03360_3mutA
GGSGGSGGSGGS	PERV_Q4VFZ2
GGGGSSGGS	PERV_Q4VFZ2_3mutA_WS
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	BAEVM_P10272_3mutA
GGGPAP	MLVAV_P03356_3mut
GGGGSGGGGSGGGGSGGGGS	FFV_O93209_2mut
GSSEAAAK	FFV_O93209
GGSPAPEAAAK	FOAMV_P14350_2mut
GGGGGSEAAAK	FOAMV_P14350_2mut
GSSPAPGGS	MLVBM_Q7SVK7_3mut
GSS	SFVCP_Q87040_2mut
EAAAKPAP	FOAMV_P14350-Pro
EAAAKGGG	SFV3L_P27401_2mut
GGGEAAAK	AVIRE_P03360_3mutA
PAPGSSGGG	WMSV_P03359_3mut
EAAAKGGSPAP	SFV3L_P27401
GSSGGSGGG	SFV3L_P27401-Pro_2mutA
GSSGGGEAAAK	GALV_P21414_3mutA
GGGPAPGSS	MLVBM_Q7SVK7_3mutA_WS

링커 아미노산 서열	RT 도메인 명칭
PAPGGGEAAAK	FFV_O93209-Pro_2mut
GSSGSSGSSGSS	SFV1_P23074_2mut
GGSEAAAK	PERV_Q4VFZ2_3mutA_WS
GGGEAAAKPAP	SFV3L_P27401_2mut
EAAAKGGGPAP	SFV3L_P27401_2mut
GGGGSSPAP	FLV_P10273_3mut
EAAAKPAPGSS	FFV_O93209_2mut
GGGGSSPAP	SFV3L_P27401_2mut
GSSGSS	KORV_Q9TTC1_3mutA
GGGGSGGGGSGGGGSGGGSGGGGS	BLVJ_P03361_2mut
GGGGSSGGS	GALV_P21414_3mutA
EAAAKGGSGSS	FFV_O93209-Pro
EAAAKPAP	PERV_Q4VFZ2
GSSGGGEAAAK	MLVBM_Q7SVK7_3mut
PAPGGSGGG	BAEVM_P10272
EAAAKGGGPAP	MLVF5_P26810
GSSGSSGSS	MLVBM_Q7SVK7_3mut
GSSGGS	AVIRE_P03360_3mutA
GGSEAAAKGGG	FOAMV_P14350_2mut
EAAAKGGS	MLVF5_P26810_3mutA
GGSGSSGGG	WMSV_P03359_3mut
EAAAK	SFV1_P23074_2mut
GSSGGSPAP	SFV3L_P27401-Pro_2mutA
GGGGSSGGS	KORV_Q9TTC1_3mut
PAPGGSGGG	FFV_O93209-Pro_2mut
GGGPAPGGS	SFV3L_P27401_2mutA
GSSPAPEAAAK	FLV_P10273_3mut
GGSGSSPAP	SFV3L_P27401_2mut
GSSEAAAKGGS	SFV3L_P27401_2mut
PAPGGG	SFV3L_P27401_2mutA
SGSETPGTSESATPES	KORV_Q9TTC1-Pro_3mut
GGGGS	SFV1_P23074-Pro_2mutA
GSSGGGEAAAK	WMSV_P03359
EAAAKGGGGSEAAAK	MLVF5_P26810_3mutA
GSSEAAAKPAP	FFV_O93209
GGGGGG	SFV1_P23074_2mutA
EAAAKEAAAKEAAAK	MMTVB_P03365-Pro
EAAAKPAPGSS	MLVBM_Q7SVK7_3mut
GGSGSSEAAAK	SFV3L_P27401_2mutA
GGSEAAAK	MLVMS_P03355_3mut

[0889]

링커 아미노산 서열	RT 도메인 명칭
GSSPAPEAAAK	SFV3L_P27401_2mut
GSGPAPGSS	SFV1_P23074
GSGGGSEAAAK	MLVBM_Q7SVK7_3mutA_WS
EAAAKPAPGSS	KORV_Q9TTC1-Pro
GSSGSSGSSGSS	SFV3L_P27401_2mut
EAAAKPAP	SFV3L_P27401_2mut
GGGEAAAK	PERV_Q4VFZ2_3mut
GSSGGS	SFV3L_P27401_2mutA
EAAAKGSSGGS	MMTVB_P03385
SGSETPGTSESATPES	SFV3L_P27401
EAAAKGSSGGG	PERV_Q4VFZ2
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03385
GGSGGGPAP	KORV_Q9TTC1_3mutA
PAPAPAPAP	SFV3L_P27401
GGGEAAAKGGS	SFV1_P23074_2mut
GSSGGSGGG	PERV_Q4VFZ2_3mut
PAPEAAAKGGS	FOAMV_P14350_2mutA
GGGEAAAKGSS	SFV3L_P27401_2mut
GGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7
PAPGSSGGG	FLV_P10273
GGSGSSGGG	FFV_O93209
EAAAKPAPGSS	MLVBM_Q7SVK7
GSSEAAAKGGG	SFV3L_P27401_2mutA
GGSGGSGGSGGSGGS	MLVF5_P26810
GGSEAAAKPAP	SFV3L_P27401-Pro_2mutA
EAAAKGGSPAP	SFV3L_P27401_2mutA
EAAAKGGGGGS	SFV3L_P27401_2mut
GSSPAPEAAAK	SFV3L_P27401_2mutA
PAPAP	MLVBM_Q7SVK7_3mut
PAPGGSEAAAK	KORV_Q9TTC1-Pro
GGSGSS	MLVF5_P26810_3mutA
GGSEAAAKPAP	FFV_O93209_2mut
GSS	MLVMS_P03355
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	SFV3L_P27401-Pro
PAPGGGEAAAK	SFV3L_P27401_2mut
PAPGGGGGS	SFV3L_P27401-Pro_2mut
PAPGGSGSS	BAEVM_P10272_3mut
GSSGGGEAAAK	FFV_O93209
GGSEAAAKPAP	SFV1_P23074_2mut
GGGGG	FLV_P10273_3mut

[0890]

[0891]

링커 아미노산 서열	RT 도메인 명칭
GGGEAAAKGSS	SFV3L_P27401
GSSGSSGSSGSSGSS	SFV1_P23074-Pro
SGSETPGTSESATPES	AVIRE_P03360
PAPGSSGGG	MLVBM_Q7SVK7_3mut
GGGGSSPAP	HTL3P_Q4U0X6_2mut
GGGEAAAK	SFV1_P23074
GGSGGG	AVIRE_P03360
EAAAKGSSGGG	SFV3L_P27401_2mutA
GSSPAPEAAAK	FOAMV_P14350-Pro_2mutA
GGGPAPGSS	WMSV_P03359
EAAAKGSSGGG	MLVMS_P03355
GGGGGSEAAAK	MLVMS_P03355
EAAAKPAPGSS	SFV3L_P27401
EAAAKGSSPAP	SFV3L_P27401
GGGGGGG	FOAMV_P14350_2mutA
EAAAKEAAAKEAAAK	SFV3L_P27401
GSSPAPGGS	FFV_O93209_2mutA
GGGGSSEAAAK	SFV3L_P27401-Pro_2mutA
GGSEAAAKGSS	GALV_P21414_3mutA
GGSEAAAKGSS	BAEVM_P10272_3mutA
EAAAKPAPGGG	MLVCB_P08361
GSSGSSGSSGSSGSSGSS	SFV1_P23074-Pro
GGGGSEAAAKGGGGS	FOAMV_P14350_2mut
GSSPAPGGS	MLVMS_P03355_PLV919
GGGGSGGGGS	FFV_O93209-Pro
GSSGGSPAP	KORV_Q9TTC1_3mutA
GGSGGS	GALV_P21414_3mutA
PAPGSSEAAAK	WMSV_P03359
PAPGGGGSS	MMTVB_P03365-Prc
GGGGSSGGS	PERV_Q4VFZ2_3mutA_WS
GGGGSGGGGS	FFV_O93209_2mut
GGGGSGGGGSGGGGSGGGGS	XMRV6_A1Z651
GGSGSSEAAAK	SFV1_P23074_2mut
GGSGGGGSS	GALV_P21414_3mutA
GGSEAAAKPAP	MLVBM_Q7SVK7
EAAAKGGSPAP	SFV1_P23074_2mutA
PAPAPAPAP	FFV_O93209
GSSGGSPAP	MMTVB_P03365-Prc
GGGGSPAP	KORV_Q9TTC1_3mutA
EAAAKGGGPAP	PERV_Q4VFZ2

[0892]

링커 아미노산 서열	RT 도메인 명칭
GSSGGSPAP	BAEVM_P10272
GGGGG	FFV_O93209
GGGGGS	FLV_P10273_3mutA
EAAAKEAAAKEAAAK	FOAMV_P14350
PAPGGG	MLVCB_P08361_3mut
GSSGGSEAAAK	FOAMV_P14350_2mutA
GGSPAPGGG	FLV_P10273_3mut
GSSGSSGSSGSSGSSGSS	SFV1_P23074-Pro_2mutA
GGSPAPEAAAK	SFV3L_P27401
PAPGGGGSS	HTL3P_Q4U0X6_2mutB
GGGGSSEAAAK	MMTVB_P03365_2mut_WS
PAPGGS	MLVRD_P11227_3mut
GGSGGSGGSGGSGGS	MMTVB_P03365
GSAGSAAGSGEF	AVIRE_P03360
GSSGGS	BAEVM_P10272_3mutA
GGSGGGGSS	MMTVB_P03365
GGSGGGGSS	WMSV_P03359
PAPEAAAKGSS	SFV1_P23074
GSSGSSGSSGSS	SFV1_P23074-Pro_2mutA
PAPAPAPAPAP	SFV3L_P27401
PAPGSSGGG	FLV_P10273_3mut
GGSGSSPAP	MLVMS_P03355
GGSGGGPAP	FOAMV_P14350
PAPGGGGGS	KORV_Q9TTC1_3mutA
EAAAKGSSPAP	GALV_P21414_3mutA
GGSGSSPAP	MLVBM_Q7SVK7_3mut
EAAAKGSS	SFV3L_P27401_2mut
GGGGGSEAAAK	WMSV_P03359
GGGGGGGG	SFV1_P23074-Pro
EAAAKEAAAK	MLVBM_Q7SVK7
GGGEAAAKGGS	MLVBM_Q7SVK7
EAAAKGGSPAP	SFV3L_P27401_2mut
GSSEAAAK	XMRV6_A1Z651
PAPGGGEAAAK	MMTVB_P03365_WS
GGSPAP	GALV_P21414_3mutA
GSSPAPGGG	MLVBM_Q7SVK7_3mutA_WS
GGSGSSPAP	SFV1_P23074_2mutA
GGS	HTL32_Q0R5R2_2mut
GGSGGGGSS	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGS	SFVCP_Q87040_2mutA

링커 아미노산 서열	RT 도메인 명칭
EAAAKGGGPAP	FOAMV_P14350_2mut
GSSGGGGEAAAK	MMTVB_P03365
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVBM_Q7SVK7_3mutA_WS
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MMTVB_P03365_WS
EAAAKEAAAK	FOAMV_P14350-Pro_2mut
GSSPAPEAAAK	FOAMV_P14350_2mutA
EAAAKPAPGGS	GALV_P21414_3mutA
GSSGGSPAP	KORV_Q9TTC1-Pro_3mut
GGGPAPEAAAK	MLVAV_P03356
GGGEAAAKPAP	SFV1_P23074-Pro_2mut
GGGGGSEAAAK	SFV3L_P27401_2mut
GGGPAPGSS	SFV3L_P27401_2mut
GGSEAAAKPAP	AVIRE_P03360
GSSGSSGSSGSSGSSGSS	SFV1_P23074-Pro_2mut
EAAAKGSSGGS	FOAMV_P14350_2mutA
GGGGGG	MLVBM_Q7SVK7_3mut
GSSPAPGGS	PERV_Q4VFZ2
GGSGSSPAP	GALV_P21414_3mutA
GGGPAPEAAAK	SFV3L_P27401
GGSGGGEAAAK	WMSV_P03359
GSAGSAAGSGEF	SFV1_P23074_2mut
GSSGGGEAAAK	MLVMS_P03355
GGG	MMTVB_P03365-Pro
PAPGSSGGS	FOAMV_P14350_2mut
GGGGSSPAP	FFV_O93209_2mut
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MMTVB_P03365_WS
GGGGGGG	XMRV6_A1Z651
PAPAPAPAPAP	FOAMV_P14350
GGGGSGGGGSGGGGSGGGGS	MMTVB_P03365_2mut_WS
GGSGGGPAP	SFV3L_P27401_2mut
GGGGGG	SFV1_P23074-Pro
EAAAKPAPGSS	SFV3L_P27401_2mut
GGGGSSGGS	HTL3P_Q4U0X6_2mut
PAPGSSEAAAK	MMTVB_P03365-Pro
GGGGSSPAP	FOAMV_P14350-Pro_2mut
PAPGSSGGS	MMTVB_P03365
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	SRV2_P51517
PAPAPAP	MMTVB_P03365_2mut_WS
PAPGGGGGS	MMTVB_P03365_2mutB
GGGGSS	SFV1_P23074-Pro_2mutA

[0893]

[0894]

링커 아미노산 서열	RT 도메인 명칭
EAAAKEAAAKEAAAKEAAAK	SFV3L_P27401-Pro
GGSGGSGGSGGSGGS	MMTVB_P03365-Pro
GGGGGGG	SFV3L_P27401_2mut
PAPGGGEAAAK	SFV3L_P27401
PAPGSS	FOAMV_P14350_2mutA
GGGGSGGGGS	SFVCP_Q87040_2mutA
GSSGGSGGG	XMRV6_A1Z651
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVBM_Q7SVK7
GSSEAAAKGGG	FFV_O93209-Pro_2mut
GGSEAAAKPAP	SFV3L_P27401-Pro
GSSGGSGGG	SFV1_P23074_2mut
EAAAKGGGGSS	FOAMV_P14350_2mutA
GGGGGG	SFV3L_P27401_2mut
GGGGG	MLVBM_Q7SVK7_3mut
PAPEAAAKGGG	SFV3L_P27401
EAAAKGGSPAP	KORV_Q9TTC1_3mutA
GGGEAAAKPAP	SFV1_P23074_2mut
GSSGSSGSSGSSGSSGSS	KORV_Q9TTC1-Pro
EAAAKEAAAKEAAAKEAAAK	SFVCP_Q87040
PAPGSSEAAAK	MLVBM_Q7SVK7
GSSGSSGSS	FFV_O93209-Pro_2mut
GSSGGGPAP	SFV3L_P27401-Pro_2mut
GGGPAPEAAAK	WMSV_P03359_3mut
GGGEAAAK	MMTVB_P03365-Pro
GSSGSSGSSGSS	SFV3L_P27401-Pro_2mutA
PAPAPAPAPAP	FFV_O93209-Pro
GGSPAPEAAAK	FFV_O93209-Pro_2mut
GSSGSSGSSGSSGSSGSS	GALV_P21414
EAAAKEAAAKEAAAKEAAAKEAAAK	FOAMV_P14350
GGGPAPEAAAK	MMTVB_P03365-Pro
PAPGGSGGG	MLVF5_P26810_3mutA
PAPGGSGGG	FLV_P10273_3mut
GGGEAAAKGGS	SFV3L_P27401
GSAGSAAGSGEF	MLVBM_Q7SVK7_3mut
GSSPAPGGG	MPMV_P07572_2mutB
GSSGSSGSSGSSGSSGSS	FOAMV_P14350
GGSGGGGSS	BLVJ_P03361_2mut
PAPEAAAKGSS	SFV1_P23074-Pro
GGG	FFV_O93209
EAAAKGGGGSS	SFV1_P23074_2mut

[0895]

링커 아미노산 서열	RT 도메인 명칭
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	SRV2_P51517
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MMTVB_P03365
GGGEAAAKGGS	MMTVB_P03365_WS
GSSGSS	SFV1_P23074
GSSGGGGGS	SFV3L_P27401
GGGGSSEAAAK	SFV1_P23074
EAAAKGSSGGS	HTL1A_P03362_2mutB
GSSEAAAKGGS	GALV_P21414_3mutA
EAAAKGSSPAP	SFV1_P23074
EAAAKPAPGSS	SFV3L_P27401_2mutA
PAPGSSGGG	SFV3L_P27401-Pro_2mut
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	SFV3L_P27401-Pro
EAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365_WS
GGGGSSEAAAK	MLVF5_P26810_3mutA
EAAAKGGSPAP	GALV_P21414
PAPEAAAKGSS	MMTVB_P03365_WS
GSSGGGGGS	SFVCP_Q87040_2mut
GGGGSSPAP	SFV1_P23074
EAAAKGGGSS	XMRV6_A1Z651
PAPAPAPAP	MMTVB_P03365
GGSEAAAKGSS	SFV3L_P27401_2mutA
GSSPAPGGG	MMTVB_P03365_WS
GGGGGG	SFV3L_P27401-Pro
GGSGGSGGS	FOAMV_P14350-Pro_2mut
PAPAPAPAPAP	WMSV_P03359
GSSPAP	MLVBM_Q7SVK7
GGGGGSGSS	MMTVB_P03365_2mut_WS
EAAAKGSSGGS	MMTVB_P03365_2mutB_WS
EAAAK	FFV_O93209_2mutA
PAPEAAAK	SFV1_P23074-Pro
EAAAKGGSGSS	SFV3L_P27401
GGSGGSGGS	FFV_O93209-Pro
GSSGGGEAAAK	MMTVB_P03365
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MLVFF_P26809_3mutA
GGSGGSGGSGGSGGSGGS	HTL1L_P0C211_2mutB
GGGEAAAK	SFV3L_P27401-Pro_2mutA
GGGGGSGSS	MMTVB_P03365
GSSPAPGGS	FOAMV_P14350_2mutA
EAAAKGSS	MLVMS_P03355
GSSGGSGGG	FFV_O93209-Pro

[0896]

링커 아미노산 서열	RT 도메인 명칭
GSSGGGGSS	MMTVB_P03365-Pro_2mut
GGSPAPGSS	FOAMV_P14350_2mut
GGSGGSGGSGGSGGSGGS	SFVCP_Q87040-Pro_2mut
GSSEAAAKGGG	FOAMV_P14350_2mutA
GGSGGSGGS	MMTVB_P03365-Pro
GSSGSSGSSGSSGSSGSS	MMTVB_P03365_2mut_WS
GSSGSSGSSGSSGSS	MMTVB_P03365-Pro
PAPEAAAK	WDSV_O92815
GSSGSSGSSGSSGSS	FFV_O93209-Pro_2mut
EAAAKGGGGSEAAAK	MMTVB_P03365-Pro
GGSPAPEAAAK	FOAMV_P14350
GSSGSS	PERV_Q4VFZ2
GGG	MMTVB_P03365-Pro
GGGGSGGGGSGGGGS	FFV_O93209_2mut
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365-Pro
GGSGSSPAP	WMSV_P03359
GGGGGGGG	SFV3L_P27401_2mut
PAPGSSEAAAK	FOAMV_P14350-Pro_2mutA
GGGGSSPAP	FOAMV_P14350_2mut
GSSGGSPAP	MLVBM_Q7SVK7_3mut
GSSGGGGGS	GALV_P21414_3mutA
EAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365
GSSGGGGGS	SFV1_P23074_2mut
GGGGSEAAAKGGGGS	SFV1_P23074
GGGEAAAKPAP	FFV_O93209
PAPGGGEAAAK	SFV1_P23074
GGSGGGEAAAK	PERV_Q4VFZ2_3mutA_WS
GSSGGG	MMTVB_P03365-Pro
EAAAKGSSGGS	FFV_O93209_2mut
GGGGG	SFV1_P23074_2mut
GGGPAP	SFV3L_P27401
GSSGGSEAAAK	FFV_O93209
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MMTVB_P03365-Pro
GSSGGGEAAAK	SFV1_P23074_2mutA
GSSGSSGSSGSSGSS	SFV3L_P27401_2mut
GGSEAAAKPAP	FLV_P10273
GGGGSGGGGS	FOAMV_P14350-Pro_2mutA
GSSEAAAKPAP	SFV3L_P27401
GGGGSEAAAKGGGGS	MMTVB_P03365-Pro
PAPGSSEAAAK	MLVF5_P26810_3mut

링커 아미노산 서열	RT 도메인 명칭
EAAAKGGSGGG	SFV3L_P27401
GGGPAPGGS	SFV3L_P27401
GSSEAAAKGGS	FOAMV_P14350_2mutA
EAAAKGGSGGG	HTL1L_P0C211
GSSGGSPAP	SFV3L_P27401_2mutA
PAPAP	FFV_O93209
PAPGGSGSS	MMTVB_P03365_WS
EAAAKGGGGGS	FOAMV_P14350_2mut
PAPEAAAKGGS	SFV3L_P27401_2mut
GSSEAAAKPAP	MMTVB_P03365-Pro
GGSGGS	PERV_Q4VFZ2_3mut
GSSEAAAKGGG	FFV_O93209-Pro_2mutA
EAAAK	HTL1L_P0C211
GSSPAP	MLVMS_P03355
EAAAKPAPGGG	FFV_O93209-Pro_2mut
GGGGSEAAAKGGGGS	SFV1_P23074-Pro_2mut
EAAAKGSSGGS	SFV3L_P27401
GSAGSAAGSGEF	FFV_O93209_2mutA
PAPEAAAKGGS	MMTVB_P03365_2mutB_WS
EAAAKEAAAKEAAAKEAAAKEAAAKEAAAK	MMTVB_P03365
GGS	MMTVB_P03365
GGSEAAAKPAP	SFV1_P23074
EAAAKGSSGGG	HTLV2_P03363_2mut
GGSEAAAKGGG	MMTVB_P03365_WS
GGSGGS	FFV_O93209-Pro
GSSEAAAKGGS	MMTVB_P03365-Pro
PAPAPAPAPAP	SFV1_P23074_2mutA
GGSEAAAKGGG	MMTVB_P03365_2mutB_WS
PAPAPAPAP	MMTVB_P03365_WS
GGGGSGGGGSGGGGSGGGGSGGGGS	HTL3P_Q4U0X6_2mut
PAPGGSEAAAK	SFV1_P23074-Pro_2mut
GGSGGGPAP	MMTVB_P03365
GSSGSSGSSGSSGSSGSS	MMTVB_P03365-Pro
GGSEAAAKPAP	SFV1_P23074-Pro
GGGEAAAKGSS	SFV3L_P27401_2mutA
GGGPAPGGS	AVIRE_P03360
PAPGGG	MLVRD_P11227
GGSEAAAKGSS	SFV3L_P27401_2mut
GGGEAAAKGSS	FOAMV_P14350_2mut
GGGEAAAKGSS	SFV1_P23074-Pro

[0897]

링커 아미노산 서열	RT 도메인 명칭
EAAAKEAAAKEAAAKEAAAK	MLVAV_P03356
EAAAKGGGPAP	JSRV_P31623_2mutB
EAAAKGGGGSS	FOAMV_P14350_2mut
EAAAKEAAAKEAAAKEAAAKEAAAK	SRV2_P51517
GSSGGGGGS	FFV_O93209
PAPAPAP	FOAMV_P14350_2mutA
GGSGGSGGSGGS	FOAMV_P14350
GGGEAAAK	MMTVB_P03365_WS
GGGGGS	SFV1_P23074_2mutA
GGSGGS	WMSV_P03359_3mut
EAAAKGGS	MMTVB_P03365-Pro
GGGGSS	BLVJ_P03361_2mut
PAPAP	MMTVB_P03365-Pro_2mut
PAPGGG	SMRVH_P03364
EAAAKGGGGSS	SFV3L_P27401
PAPAPAPAP	MMTVB_P03365
GGGPAP	MMTVB_P03365-Pro
GSSGGSGGG	MMTVB_P03365
EAAAKGGGPAP	FOAMV_P14350_2mutA
GSSGSSGSSGSS	SFV1_P23074
GGGGSGGGGS	SFV3L_P27401
GSSGGSGGG	MLVF5_P26810
GGGEAAAKPAP	MMTVB_P03365-Pro
PAPEAAAK	HTLV2_P03363_2mut
GSSGSSGSSGSS	FOAMV_P14350_2mut
GSSEAAAKPAP	MMTVB_P03365-Pro
PAPEAAAKGGG	HTL3P_Q4U0X6_2mut
GGSEAAAKGSS	MMTVB_P03365-Pro
EAAAKPAPGGS	MMTVB_P03365_2mut_WS
GSSGGSEAAAK	MLVF5_P26810_3mutA
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MLVF5_P26810_3mut
EAAAKGGGGSS	MMTVB_P03365-Pro
GGGGGSGSS	HTL1A_P03362_2mutB
PAPAP	FFV_O93209-Pro_2mut
GGGGGSPAP	HTL1C_P14078_2mut
GGGPAP	HTLV2_P03363_2mut
EAAAKGGGGSEAAAK	SFVCP_Q87040
GGSEAAAKGGG	FFV_O93209-Pro_2mutA
GSSPAPGGS	FOAMV_P14350-Pro_2mut
GGGGGGG	MMTVB_P03365-Pro

[0898]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGSS	SFV3L_P27401_2mutA
EAAAKGGGGSEAAAK	MMTVB_P03365-Pro
GGGGSEAAAKGGGGS	SFV1_P23074-Pro_2mutA
EAAAKGGGGSS	MMTVB_P03365
GGGEAAAKGGS	SFV1_P23074
PAPEAAAKGGG	MLVF5_P26810
GGGGSSGGS	MMTVB_P03365
GGSGSS	MMTVB_P03365
PAPAPAPAPAPAP	KORV_Q9TTC1
EAAAKGGG	SFV1_P23074-Pro_2mut
PAPAPAPAPAPAP	SRV2_P51517
GSSGSSGSSGSSGSS	FFV_O93209-Pro_2mutA
GGGGSS	FOAMV_P14350_2mut
PAPGGGEAAAK	MMTVB_P03365_WS
GGSGGGEAAAK	FFV_O93209-Pro_2mut
PAPAPAPAPAP	MMTVB_P03365_WS
GGGEAAAKGGS	MMTVB_P03365-Pro
GGGEAAAKGSS	MMTVB_P03365_2mutB
GSSPAPEAAAK	MMTVB_P03365_WS
EAAAKEAAAKEAAAKEAAAKEAAAK	SFV1_P23074-Pro_2mutA
PAPGGG	SFV3L_P27401
GSSEAAAKGGG	MMTVB_P03365_WS
GGGGSSEAAAK	FOAMV_P14350_2mut
PAPGSSGGS	SFV1_P23074-Pro_2mut
GSSGSSGSSGSSGSSGSS	SFV3L_P27401
EAAAKGSSGGG	MMTVB_P03365
PAPGGGGSS	WDSV_O92815_2mutA
GGSPAP	MMTVB_P03365-Pro
GGSGGSGGSGGSGGS	SFVCP_Q87040-Pro_2mut
PAPAPAPAP	MMTVB_P03365-Pro
GGGGG	HTL1A_P03362
GGSGGSGGSGGS	SFV1_P23074_2mutA
GSSGSSGSSGSSGSS	FOAMV_P14350-Pro_2mut
PAPGGSEAAAK	MMTVB_P03365_2mutB_WS
PAPAPAPAP	SFV1_P23074_2mut
PAPGGGGSS	MMTVB_P03365
GGSGSS	SFV3L_P27401_2mut
EAAAKEAAAKEAAAKEAAAK	MMTVB_P03365_2mut
EAAAKGGSGGG	HTL3P_Q4U0X6_2mut
PAPGGGGSS	SFVCP_Q87040-Pro_2mutA

[0899]

링커 아미노산 서열	RT 도메인 명칭
EAAAKGGGGGS	MLVAV_P03356
GGGGGS	FOAMV_P14350_2mut
GGGEAAAKGGS	FFV_O93209-Pro_2mutA
EAAAKPAPGGG	MMTVB_P03365_2mutB
GGSGGGPAP	FFV_O93209_2mut
GGSEAAAKPAP	MMTVB_P03365
PAPAPAPAPAPAP	SFV1_P23074_2mut
GGSPAPGGG	MMTVB_P03365-Pro
GGSGGGEAAAK	MMTVB_P03365
PAPAP	SFVCP_Q87040
GSSEAAAK	SFVCP_Q87040
GGGGSGGGGSGGGGS	MMTVB_P03365-Pro
GSSGSSGSS	SFV3L_P27401
EAAAKGGSGGG	MMTVB_P03365-Pro
GSSPAP	SFV1_P23074_2mut
GGGEAAAK	SFV1_P23074-Pro
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MMTVB_P03365-Pro
PAPGGS	HTL1C_P14078_2mut
PAPGSSGGS	SFV1_P23074_2mut
PAPEAAAK	MMTVB_P03365_WS
PAPAP	MMTVB_P03365-Pro
EAAAKGGS	HTL1A_P03362_2mut
GGGGSEAAAKGGGGS	HTL1C_P14078
EAAAKGSSGGS	FOAMV_P14350-Pro
PAPGGSGSS	MMTVB_P03365-Pro
PAPGGSEAAAK	SFV1_P23074_2mut
PAPGSSEAAAK	FFV_O93209-Pro_2mut
PAPGSSGGG	FOAMV_P14350-Pro_2mutA
GSSGGGEAAAK	AVIRE_P03360
GGGGGG	SMRVH_P03364_2mut
PAPEAAAKGGG	MMTVB_P03365-Pro
GGGEAAAKGGS	SFVCP_Q87040_2mutA
PAPAPAPAPAP	SRV2_P51517
GSSGSSGSSGSSGSSGSS	MMTVB_P03365
EAAAKGGGPAP	MLVAV_P03356
PAPAPAPAPAP	FOAMV_P14350-Pro_2mutA
PAPGGSEAAAK	FOAMV_P14350
GSSGGGPAP	HTL32_Q0R5R2_2mutB
GGGGGSPAP	HTL3P_Q4U0X6_2mutB
GSSGGSGGG	MMTVB_P03365-Pro

[0900]

링커 아미노산 서열	RT 도메인 명칭
PAPAP	SFVCP_Q87040-Pro
GSSGGGPAP	MMTVB_P03365-Pro
GGSGSS	MMTVB_P03365-Pro_2mut
GGSPAPEAAAK	SFV1_P23074-Pro_2mut
EAAAKGGSGGG	SFV3L_P27401_2mut
GGGGSSEAAAK	MMTVB_P03365_WS
GGGGGSGSS	MMTVB_P03365_2mut
GGGGSSGGS	SFV1_P23074-Pro_2mutA
EAAAKGGGGSEAAAK	MMTVB_P03365_WS
PAPGGGEAAAK	SFV1_P23074-Pro
PAPEAAAKGGG	MMTVB_P03365
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	MMTVB_P03365
GSSGGSEAAAK	FOAMV_P14350-Pro_2mut
GGSPAP	MLVBM_Q7SVK7_3mut
GSSEAAAK	FOAMV_P14350
GSSEAAAK	MMTVB_P03365-Pro
EAAAKGSSGGS	HTL1A_P03362_2mut
GGGEAAAKPAP	FOAMV_P14350-Pro_2mut
EAAAKGGSPAP	FOAMV_P14350
GSSEAAAKPAP	MMTVB_P03365_WS
GSSGSSGSS	FOAMV_P14350_2mut
EAAAKEAAAKEAAAKEAAAK	MMTVB_P03365_WS
EAAAK	MMTVB_P03365
PAPGSS	BAEVM_P10272
PAPGGS	FFV_O93209-Pro_2mut
GGSGGS	SFV1_P23074-Pro_2mutA
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	HTLV2_P03363_2mut
GGSGGGEAAAK	MMTVB_P03365_WS
PAPGSSGGG	HTL1A_P03362
GGSGGS	SFV3L_P27401-Pro
GSSGSS	SFV1_P23074-Pro
PAPGGSEAAAK	MMTVB_P03365
GSAGSAAGSGEF	MMTVB_P03365-Pro
PAPGGG	FOAMV_P14350_2mut
EAAAKGGSGSS	MMTVB_P03365_WS
GSSGGGEAAAK	SFV3L_P27401-Pro
GGSGGGPAP	FOAMV_P14350-Pro_2mut
PAPAPAPAPAP	WDSV_O92815
SGSETPGTSESATPES	SFVCP_Q87040-Pro_2mutA
GGSGGSGGS	SFV1_P23074

[0901]

[0902]

링커 아미노산 서열	RT 도메인 명칭
GGGGSS	SFVCP_Q87040_2mut
GGGGGSEAAAK	MMTVB_P03365
SGSETPGTSESATPES	MMTVB_P03365_WS
PAPAPAP	SFV3L_P27401
PAPEAAAKGSS	MMTVB_P03365_2mutB_WS
GSSGSSGSSGSSGSS	SRV2_P51517
GGGPAPGSS	HTL32_Q0R5R2_2mutB
GGSGGGGSS	MMTVB_P03365-Pro
SGSETPGTSESATPES	SRV2_P51517
EAAAKGSSGGS	MMTVB_P03365-Pro
GSSPAPEAAAK	MMTVB_P03365-Pro
GSSPAPEAAAK	SRV2_P51517
GGGGSSPAP	MMTVB_P03365-Pro
PAPGGGEAAAK	SFV1_P23074-Pro_2mutA
PAPEAAAKGGS	MMTVB_P03365
GSSGSSGSSGSSGSSGSS	FOAMV_P14350-Pro
GGSPAPGSS	SFV3L_P27401
GGGPAPGGS	SFV1_P23074-Pro_2mutA
GGGPAPGSS	MMTVB_P03365-Pro
EAAAKPAP	MLVBM_Q7SVK7
EAAAKEAAAKEAAAK	HTL1C_P14078
GSSGGSEAAAK	SRV2_P51517
PAPGGGGGS	SRV2_P51517
GGGEAAAK	FFV_O93209-Pro_2mut
EAAAKGGGPAP	HTL32_Q0R5R2
GGSGSSGGG	MMTVB_P03365
PAPEAAAKGSS	MMTVB_P03365-Pro
PAPGGGGGS	MMTVB_P03365-Pro
EAAAKGGGGGS	MMTVB_P03365_WS
GGGGGS	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGSGGGGS	HTL1C_P14078
EAAAKGGSPAP	MMTVB_P03365
GGGGSSPAP	FFV_O93209-Pro_2mut
GGGGSSGGS	MMTVB_P03365-Pro
PAPGSSGGS	MMTVB_P03365-Pro
GGGGGS	SRV2_P51517
GGSGSSGGG	MMTVB_P03365
GSSGGSEAAAK	MMTVB_P03365-Pro
EAAAKEAAAKEAAAKEAAAK	GALV_P21414
GGSEAAAKGGG	MMTVB_P03365-Pro

링커 아미노산 서열	RT 도메인 명칭
SGGSSGGSSGSETPGTSESATPESSGGSSGGSS	MMTVB_P03365-Pro
GSSEAAAKGGS	MMTVB_P03365
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	HTL3P_Q4U0X6_2mutB
GGGEAAAK	MMTVB_P03365-Pro
PAPAPAPAP	MMTVB_P03365-Pro
PAPGSSGGG	MMTVB_P03365
GSSGSSGSSGSSGSS	GALV_P21414
GGSPAP	MMTVB_P03365_WS
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MMTVB_P03365-Pro
PAPEAAAK	MMTVB_P03365-Pro
PAPGSSGGG	SFV1_P23074-Pro_2mutA
GGGGGSEAAAK	MMTVB_P03365_2mutB_WS
PAPAPAPAPAP	MMTVB_P03365-Pro
EAAAKGGSGSS	MMTVB_P03365-Pro
EAAAKEAAAKEAAAKEAAAK	MLVRD_P11227_3mut
PAPAPAPAP	FOAMV_P14350_2mutA
GGGPAPGSS	SFVCP_Q87040_2mut
PAPEAAAKGSS	SFVCP_Q87040_2mut
GGSPAPGGG	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGS	MMTVB_P03365
EAAAKGGS	HTL3P_Q4U0X6_2mut
PAPGSSGGS	MMTVB_P03365_WS
GGGGSGGGGS	MMTVB_P03365
GGSGGS	FOAMV_P14350
EAAAKGGGSEAAAK	SFVCP_Q87040-Pro_2mut
EAAAKEAAAKEAAAKEAAAK	MMTVB_P03365-Pro_2mutB
PAPGGGEAAAK	SFVCP_Q87040-Pro
GSSGSS	JSRV_P31623_2mutB
EAAAKGGGGS	MMTVB_P03365_2mut_WS
GSSPAPEAAAK	MMTVB_P03365-Pro
GGGEAAAK	HTL1C_P14078
PAPEAAAKGSS	HTL32_Q0R5R2_2mutB
GGGGSSEAAAK	MMTVB_P03365-Pro
PAPGSSGGS	MMTVB_P03365-Pro
EAAAKGGGGS	MMTVB_P03365
GGGGSGGGGSGGGGSGGGGS	MMTVB_P03365
EAAAKGGGGSS	HTL3P_Q4U0X6_2mut
GGGEAAAKGGS	SFVCP_Q87040-Pro
GGGGGSPAP	MMTVB_P03365-Pro_2mutB
GGSGGGEAAAK	SFV3L_P27401-Pro

[0903]

링커 아미노산 서열	RT 도메인 명칭
PAPGGGGGS	SFV3L_P27401-Pro
EAAAKGGGGSEAAAK	MMTVB_P03365
PAPEAAAKGSS	MMTVB_P03365-Pro
GGSEAAAKGGG	MMTVB_P03365-Pro
GGSGGSGGSGGSGGS	SMRVH_P03364_2mutB
GGSGGSGGSGGSGGS	HTL1L_P0C211_2mut
GGGGGG	WDSV_O92815
GGGGGSGSS	MMTVB_P03365-Pro
GGSEAAAKPAP	SFV3L_P27401-Pro_2mut
GGGPAPGSS	MMTVB_P03365_2mut_WS
GGGGGS	MMTVB_P03365_WS
GGSPAPEAAAK	MMTVB_P03365
PAPEAAAKGGS	HTL1A_P03362
EAAAKGGSGSS	MMTVB_P03365_2mut_WS
GGGPAPEAAAK	SFV3L_P27401-Pro_2mut
PAPGGGGSS	HTL32_Q0R5R2_2mut
GSSPAPGGG	HTL3P_Q4U0X6_2mut
GGGGSSGGS	BLVAU_P25059_2mut
EAAAKGGGGGS	HTL1L_P0C211
GGSEAAAKGSS	JSRV_P31623_2mutB
GSSGGG	JSRV_P31623
GGSGGSGGSGGS	MMTVB_P03365-Pro
EAAAKPAP	SFV1_P23074-Pro_2mutA
GGGGSSGGS	MMTVB_P03365_WS
GGSGGS	MMTVB_P03365_WS
EAAAKGGGGGS	MMTVB_P03365-Pro
GGGGSGGGGSGGGGSGGGGSGGGGSGGGGS	MMTVB_P03365
GGSGGSGGS	MMTVB_P03365
GGGGGSEAAAK	MLVBM_Q7SVK7
GGSGSSPAP	MMTVB_P03365_WS
EAAAKEAAAKEAAAK	JSRV_P31623
PAPEAAAKGGS	MMTVB_P03365-Pro
GGSGSSEAAAK	FOAMV_P14350
GGGGGSGSS	MMTVB_P03365-Pro_2mut
GGGPAPGGS	MMTVB_P03365
SGSETPGTSESATPES	SFVCP_Q87040_2mut
GSSPAPGGS	SFV1_P23074-Pro_2mutA
GSSGSSGSSGSSGSS	MMTVB_P03365
EAAAKGGGPAP	MMTVB_P03365
GSSGGG	MMTVB_P03365_2mut_WS

[0904]

링커 아미노산 서열	RT 도메인 명칭
GGGEAAAKPAP	MMTVB_P03365
PAPGGSGGG	MMTVB_P03365-Pro
GSSGGSGGG	WDSV_O92815_2mut
GGSGGG	HTL32_Q0R5R2_2mut
EAAAKGGSPAP	HTLV2_P03363_2mut
GGSPAPEAAAK	MMTVB_P03365-Pro
GSSGGSEAAAK	MMTVB_P03365_2mut
GSAGSAAGSGEF	MMTVB_P03365_WS
PAPGGSGSS	FFV_O93209
GGSEAAAKGGG	MMTVB_P03365
GGSPAPGSS	MMTVB_P03365-Pro
GSSGGSGGG	SFV3L_P27401
PAPEAAAKGGG	HTL1A_P03362_2mutB
GGGEAAAKPAP	MMTVB_P03365-Pro
GGSEAAAK	HTL32_Q0R5R2_2mutB
GGGEAAAKGSS	MPMV_P07572
GGGGGSEAAAK	MMTVB_P03365-Pro
PAPAPAPAPAP	SFVCP_Q87040-Pro_2mutA
PAPAPAPAPAP	HTL1L_P0C211_2mut
GGGGSSGGS	HTL3P_Q4U0X6
PAPGGSEAAAK	MMTVB_P03365_2mut_WS
PAPAPAPAPAP	HTL1A_P03362
EAAAKPAPGGG	MMTVB_P03365_2mut_WS
GGSEAAAK	MMTVB_P03365_2mut_WS
GGGEAAAKGSS	SFV1_P23074-Pro_2mutA
GGSPAPGSS	MMTVB_P03365-Pro
GGSEAAAKPAP	MLVBM_Q7SVK7
PAPEAAAKGGG	MMTVB_P03365_2mut_WS
GSSEAAAKPAP	MMTVB_P03365-Pro_2mutB
GGGGSEAAAKGGGGS	MMTVB_P03365-Pro_2mut
GSSEAAAKGGS	MMTVB_P03365-Pro_2mutB
GSSGSSGSSGSSGSS	SRV2_P51517_2mutB
GGGGGSPAP	HTL1L_P0C211_2mut
GGSEAAAK	MMTVB_P03365
GSSPAPEAAAK	SMRVH_P03364_2mutB
GGGPAPGGS	HTL1C_P14078_2mut
GGSPAPEAAAK	MMTVB_P03365_WS
GGSEAAAKPAP	HTL1A_P03362_2mut
PAPAPAPAP	HTLV2_P03363_2mut
GSSPAPGGG	MMTVB_P03365

[0905]

링커 아미노산 서열	RT 도메인 명칭
GSSGSSGSSGSS	MMTVB_P03365-Pro
GGSEAAAKGSS	MMTVB_P03365_WS
GGSGSSGGG	MMTVB_P03365_2mutB
GSSGSSGSSGSSGSSGSS	JSRV_P31623_2mutB
GGSEAAAKPAP	MMTVB_P03365-Pro
GSSGGSGGG	HTLV2_P03363_2mut
AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA	WDSV_O92815_2mut
GGSPAPEAAAK	MMTVB_P03365
GGGGSSEAAAK	MMTVB_P03365
GGSGGGEAAAK	SFV1_P23074-Pro_2mutA
GGGGSEAAAKGGGGS	WDSV_O92815_2mut
GGSGSSEAAAK	MMTVB_P03365_2mutB_WS
GGSEAAAKPAP	MMTVB_P03365_WS
GSSGGGEAAAK	SFVCP_Q87040-Pro
GSSGGS	SFVCP_Q87040-Pro_2mut
GGSEAAAKPAP	SFVCP_Q87040_2mut
GSSGGSEAAAK	SFVCP_Q87040_2mut
GSSPAPEAAAK	SRV2_P51517_2mutB
GGSGGSGGSGGSGGSGGS	BLVAU_P25059
GSSGSSGSSGSSGSS	HTL1C_P14078_2mut
EAAAKGGGGSS	MMTVB_P03365_2mutB
GGGEAAAKGSS	SFVCP_Q87040-Pro

[0906]

[0907] **실시예 4: 인간 세포에서 정확한 편집을 가능하게 하기 위해 다양한 인간 DNA 중합효소 세타 도메인을 갖는 Cas-Pol 융합체를 포함하는 유전자 변형 시스템**

[0908] 이 실시예에서, Cas 도메인 및 중합효소 도메인을 포함하는 예시적인 유전자 변형 폴리펩티드(Cas-Pol 융합체)를 특성화하고, 인간 세포의 게놈 DNA에 돌연변이를 프로그래밍 가능한 방식으로 설치하는 능력을 결정하였다. 유전자 변형 폴리펩티드는 예시적인 인간 중합효소(인간 DNA 중합효소 세타(Polθ))(도 2, 표 23)의 다양한 절단으로 구성되었으며: 유전자 변형 폴리펩티드는 중합효소 도메인, 엑소뉴클레아제 도메인, 및 두 개의 절단된 중앙 도메인(Polθ_M 또는 Polθ) 또는 중앙 도메인이 결여된 것(Polθ_4x0q) 중 어느 하나를 포함하였다. 유전자 변형 폴리펩티드는 N863A 돌연변이(nCas9)를 포함하는 예시적인 Cas 도메인인 SpCas9로 구성되었다. 2개의 상이한 링커 서열을 사용하여 Cas 도메인과 Pol 도메인을 연결하였다.

[0909] *예시적인 링커 FL:*

[0910] *DNA 서열:*

[0911] TCTGGAGGATCTAGCGGAGGATCCTCTGGCAGCGAGACACCAGGAACAAGCGAGTCAGCAACACCAGAGAGCAGTGGCGGCAGCAGCGGCGGCAGCAGC (서열번호 4400)

[0912] *아미노산 서열:*

[0913] SGGSSGGSSGSETPGTSESATPESSGGSSGGSS (서열번호 4401)

[0914] *예시적인 링커 UL:*

[0915] *DNA 서열:*

[0916] GCCGAAGCCGCCGCCAAGGAGGCCGCCGCTAAGGAGGCTGCTGCCAAGGAAGCTGCTGCTAAGGCTTTAGAAGCTGAAGCTGCTGCCAAAGAAGCTGCCGCCAAAGAGGCTGCCGCAAAGGAGGCCGCTGCCAAGGCT (서열번호 4402)

[0917] *아미노산 서열:*

[0918] AEAAAKEAAAKEAAAKEAAAKALEAEAAAKEAAAKEAAAKEAAAKA (서열번호 4403)

[0919] 6개의 유전자 변형 폴리펩티드(Cas-Pol 융합체)를 시험하였다: (1) nCas9-FL-Polθ_M, (2) nCas9-FL-Polθ_4x0q, (3) nCas9-UL-Polθ_L, (4) nCas9-UL-Polθ_M, (5) nCas9-UL-Polθ_4x0q, 및 (6) 중합효소 도메인 대신에 역전사효소 도메인과 nCas9를 포함하는 벤치마크 유전자 변형 폴리펩티드.

[0920] **표 23: 예시적인 Cas-Pol 융합 유전자 변형 폴리펩티드**

표 23

[0921]

명칭	아미노산 서열	서열번호
Polθ_L	GASFDLSPGLQRILDKVSSPLENEKLKSMTINFSSLNRKNTELNEEQEVISNLETKQVQGISFSSNNEVKSKIEMLENNAN HDETSSLLPRKESNIVDDNGLIPPTPIPTSASKLTFPGILETPVNPWKTNNVLQPGESYLFGSPSDIKNHDLSPGSRNGFK DNSPISDTSFSLQLSQDGLQLTPASSSSESLSIIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRSLTSSKTATIGSRFK QASSPQEIPIRDDGFPIKGCDDTLVVGLAVCWGGRDAYYFSLQKEQKHSEISASLVPPSLDPSLTLKDRMWYLQSCLRKES DKECSVVIYDFIQSYKILLLSCGISLEQSYEDPKVACWLLDPDSQEPTLHSIVTSFLPHELPLLEGMETSQGIQSLGLNAG SEHSGRYRASVESILIFNSMNQLNSLLQKENLQDVFRKVEMPSQYCLALLELNGIGFSTAECESQKHIMQAKLDAIETQAY QLAGHSFSFTSSDDIAEVLFLELKLPPNREMKNQGSKKTLGSTRRGIDNGRKLRLGRQFSTSKDVLNKLKALHPLPGLILE WRRITNAITKVVFPLQREKCLNPFLGMERIYPVSQSHTATGRITFTEPNIQNVPRDFEIKMPTLVGESPPSQAVGKGLLPM GRGKYKKGFSVNPRCQAQMEERAADRGMPFSISMRHAFVPFPGGSILAADYSQLELRILAHLSHDRRLIQVLNTGADVFRS IAAEWKMIEPESVGDDLRQQAKQICYGIIYGMGAKSLGEQMGIKENDAACYIDSFKSRYTGINQFMTETVKNCKRDGFVQT ILGRRRYLPGIKDNNPYRKAHAERQAINTIVQGSAADIVKIATVNIQKQLETFHSTFKSHGHREGMLQSDQTGLSRKRKLQ GMFCPIRGGFFILQLHDELLYEVAEEDVVQVAQIVKNEMESAVKLSVKLKVKVKIGASWGELKDFDV	4404
Polθ_M	GFKDNSPISDTSFSLQLSQDGLQLTPASSSSESLSIIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRSLTSSKTATIGS RFKQASSPQEIPIRDDGFPIKGCDDTLVVGLAVCWGGRDAYYFSLQKEQKHSEISASLVPPSLDPSLTLKDRMWYLQSCLR KESDKECSVVIYDFIQSYKILLLSCGISLEQSYEDPKVACWLLDPDSQEPTLHSIVTSFLPHELPLLEGMETSQGIQSLGL NAGSEHSGRYRASVESILIFNSMNQLNSLLQKENLQDVFRKVEMPSQYCLALLELNGIGFSTAECESQKHIMQAKLDAIET QAYQLAGHSFSFTSSDDIAEVLFLELKLPPNREMKNQGSKKTLGSTRRGIDNGRKLRLGRQFSTSKDVLNKLKALHPLPGL ILEWRRITNAITKVVFPLQREKCLNPFLGMERIYPVSQSHTATGRITFTEPNIQNVPRDFEIKMPTLVGESPPSQAVGKGL LPMGRGKYKKGFSVNPRCQAQMEERAADRGMPFSISMRHAFVPFPGGSILAADYSQLELRILAHLSHDRRLIQVLNTGADV FRSIAAEWKMIEPESVGDDLRQQAKQICYGIIYGMGAKSLGEQMGIKENDAACYIDSFKSRYTGINQFMTETVKNCKRDGF VQTILGRRRYLPGIKDNNPYRKAHAERQAINTIVQGSAADIVKIATVNIQKQLETFHSTFKSHGHREGMLQSDQTGLSRKR KLQGMFCPIRGGFFILQLHDELLYEVAEEDVVQVAQIVKNEMESAVKLSVKLKVKVKIGASWGELKDFDV	4405
Polθ_4x0q	SSSSESLSIIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRSLTSSKTATIGSRFKQASSPQEIPIRDDGFPIKGCDDTL VVGLAVCWGGRDAYYFSLQKEQKHSEISASLVPPSLDPSLTLKDRMWYLQSCLRKESDKECSVVIYDFIQSYKILLLSCGI SLEQSYEDPKVACWLLDPDSQEPTLHSIVTSFLPHELPLLEGMETSQGIQSLGLNAGSEHSGRYRASVESILIFNSMNQLN SLLQKENLQDVFRKVEMPSQYCLALLELNGIGFSTAECESQKHIMQAKLDAIETQAYQLAGHSFSFTSSDDIAEVLFLELK LPPNREMKNQGSKKTLGSTRRGIDNGRKLRLGRQFSTSKDVLNKLKALHPLPGLILEWRRITNAITKVVFPLQREKCLNPF LGMERIYPVSQSHTATGRITFTEPNIQNVPRDFEIKMPTLVGESPPSQAVGKGLLPMGRGKYKKGFSVNPRCQAQMEERAA DRGMPFSISMRHAFVPFPGGSILAADYSQLELRILAHLSHDRRLIQVLNTGADVFRSIAAEWKMIEPESVGDDLRQQAKQI CYGIIYGMGAKSLGEQMGIKENDAACYIDSFKSRYTGINQFMTETVKNCKRDGFVQTILGRRRYLPGIKDNNPYRKAHAER QAINTIVQGSAADIVKIATVNIQKQLETFHSTFKSHGHREGMLQSDQTGLSRKRKLQGMFCPIRGGFFILQLHDELLYEVA EEDVVQVAQIVKNEMESAVKLSVKLKVKVKIGASWGELKDFDV	4406
Polθ_4x0q_GS	SSSSESLSIIDVASDQNLFQTFIKEWRCKKRFSISLACEKIRGSGDDTLVVGLAVCWGGRDAYYFSLGGSGGLDPSLTLKD RMWYLQSCLRKESDKECSVVIYDFIQSYKILLLSCGISLEQSYEDPKVACWLLDPDSQEPTLHSIVTSFLPHELPLLEGME TSQGIQSLGLNAGSEHSGRYRASVESILIFNSMNQLNSLLQKENLQDVFRKVEMPSQYCLALLELNGIGFSTAECESQKHI MQAKLDAIETQAYQLAGHSFSFTSSDDIAEVLFLELKLPPGGSGGQFSTSKDVLNKLKALHPLPGLILEWRRITNAITKVV FPLQREKCLNPFLGMERIYPVSQSHTATGRITFTEPNIQNVPRDFEIKMGGSGGMPFSISMRHAFVPFPGGSILAADYSQL ELRILAHLSHDRRLIQVLNTGADVFRSIAAEWKMIEPESVGDDLRQQAKQICYGIIYGMGAKSLGEQMGIKENDAACYIDS FKSRYTGINQFMTETVKNCKRDGFVQTILGRRRYLPGIKDNNPYRKAHAERQAINTIVQGSAADIVKIATVNIQKQLETFH STFKSHGHREGMLQSDGGSGGCPIRGGFFILQLHDELLYEVAEEDVVQVAQIVKNEMESAVKLSVKLKVKVKIGASWGELK DFDV	4407
nCas9	DKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLKRTARRRYTRRKNRICYL QEIFSNEMAKVDDSFFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIK FRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIAL SLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRY DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFD NGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGAS AQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDY FKKIECFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKTYAHLFDDKVMK QLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGS PAIKKGILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVENTQLQNEKL YLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLTRSDKARGKSDNVPSEEVVKKMKNYWRQLLNAKLI TQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFY KVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLAN GEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFD SPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLAS AGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKH RDKPIREQAENIIHLFTLTNLGAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD	4408

[0922] 예시적인 유전자 변형 폴리펩티드에 의해 편집물로서 도입될 때 외인성 BFP 마커를 GFP로 전환시키는 작용을 하는 직접적인 단일 염기쌍 치환을 코딩하는 시험 주형 핵산을 구축하였다(본 명세서에 기재된 바와 같음). 주형

핵산은 (i) 유전자 변형 폴리펩티드를 표적 영역으로 안내하기 위한 gRNA 스페이서 서열, 예를 들어 BFP 유전자 좌의 20개 뉴클레오티드 서열에 상보적인 서열; (ii) 표적-프라이밍된 중합을 위해 닉이 생성된 DNA의 단일 가닥과 상보적인 염기쌍을 형성할 수 있는 프라이머-결합 서열(PBS); (iii) 의도된 최종 표적 서열을 추가로 포함하는 중합을 위한 주형을 제공하는 이종 목적 서열; 및 (iv) 유전자 변형 폴리펩티드의 Cas9 도메인과 회합하는 gRNA 스캐폴드 서열을 포함한다. 예시적인 주형 핵산은 5'에서 3'으로 (i)-(iv)-(iii)-(ii) 순서로 배열되었다. 주형 구성은 표 24(RNACS-4239-DD 및 RNACS-4241-DR)에 기재되어 있으며, 여기서 (ii) 및 (iii)은 각각 RNA 또는 DNA 뉴클레오티드로 포함될 수 있다.

[0923] **표 24: 실시예 1에서 사용된 주형 핵산**

표 24

[0924]

명칭	설명	스페이서 (i)	스캐폴드 (iv)	Pol 주형 (iii)	PBS (ii)	전체 주형 분자
주형 DD	BFP_g4_11RT_13PBS_DD	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrU (서열번호 4409)	rGrUrUrUrUrArGrGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrC (서열번호 4411)	CCTGACGTACG (서열번호 4413)	GCGTGCAGTG*mC*mU*mU (서열번호 4415)	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrUrGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCCCTGACGTACGGCGTGCAGTG*mC*mU*mU (서열번호 4417)
주형 DR	BFP_g4_11RT_13PBS_DR	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrU (서열번호 4410)	rGrUrUrUrUrArGrGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrC (서열번호 4412)	CCTGACGTACG (서열번호 4414)	rGrCrGrUrGrCrArGrUrG*mC*mU*mU (서열번호 4416)	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrUrGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCCCTGACGTACGrGrCrGrUrGrCrArGrUrG*mC*mU*mU (서열번호 4418)
주형 RR	BFP_g4_11RT_13PBS_RR	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrU (서열번호 4419)	rGrUrUrUrUrArGrGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrC (서열번호 4420)	rCrCrUrGrArCrGrUrArCrG (서열번호 4421)	rGrCrGrUrGrCrArGrUrG*mC*mU*mU (서열번호 4422)	mG*mC*mC*rGrArArGrCrArCrUrGrCrArCrGrCrCrGrUrGrUrUrUrUrArGrArGrCrUrArGrArArArUrArGrCrArArGrUrUrArArArArUrArArGrGrCrUrArGrUrCrCrGrUrUrArUrCrArArCrUrUrGrArArArArArGrUrGrGrCrArCrCrGrArGrUrCrGrGrUrGrCrCrCrUrGrArCrGrUrArCrGrGrCrGrUrGrCrArGrUrG*mC*mU*mU (서열번호 4423)

[0925] RNA, 'r_'로 표현됨, 예를 들어 rA, rU; DNA, '_'로 표현됨, 예를 들어 A, T; 2' O-메틸 RNA 염기는 'm_'으로

표현됨; 포스포로티오에이티드 2'-O-메틸 RNA 염기는 'm_*'로 입력됨.

[0926] Cas-Pol 융합 유전자 변형 폴리펩티드의 게놈 편집 능력을 결정하기 위해 실험을 수행하였다.

[0927] U2OS 또는 HEK293T 세포를 예시적인 유전자 변형 폴리펩티드를 인코딩하고 항생제(퓨로마이신) 내성 카세트를 함유하는 렌티바이러스 벡터로 감염시켰다. 유전자 변형 폴리펩티드를 안정적으로 발현하는 세포를 퓨로마이신이 보충된 배지에서 세포를 배양하여 선택하였다. 250,000개 세포/웰(96-웰 플레이트)을 2.5 uM 및 5 uM의 화학적으로 합성된 예시적인 주형 핵산 분자로 뉴클레오펙션시켰다. 뉴클레오펙션 후 3일째에 세포에 대해 형광-활성화 세포 분류(FACS, 유동)를 수행하였다. 의도된 편집 빈도(GFP 양성 세포로 BFP 양성 세포의 전환)는 BFP 양성 세포의 백분율에 대한 GFP 양성 세포의 백분율로 결정되었다(도 3a 및 3b).

[0928] 결과는 유전자 변형 폴리펩티드 nCas9-UL- PolΘ_L, nCas9-UL- PolΘ_M, nCas9-UL- PolΘ_4x0q, nCas9-FL- PolΘ_M, 및 nCas9-FL- PolΘ_4x0q가 HEK293 세포에서 편집 활성, 예를 들어 최대 약 75%의 편집 효율을 나타냄을 보여준다. 유전자 변형 폴리펩티드는 DNA-인코딩된 이종 목적 서열과 "DR"로 표시되는 RNA 프라이머 결합 서열을 포함하는 주형 핵산과 함께 높은 편집 활성(대조 역전사효소를 함유하는 벤치마크 유전자 변형 폴리펩티드와 유사하거나 이보다 높음)을 나타내었다(도 3a 내지 b). 유전자 변형 폴리펩티드는 또한 DNA 이종 목적 서열과 "DD"로 표시된 DNA 프라이머 결합 서열을 둘 다 포함하는 주형 핵산에 대해 편집 활성을 나타내었지만, 이는 RNA 프라이머 결합 서열이 사용된 경우보다 적었다(도 3a 내지 b). 유전자 변형 폴리펩티드는 "RR"로 표시된 RNA 이종 목적 서열 및 RNA 프라이머 결합 서열을 포함하는 주형 핵산에 대해 편집 활성을 나타내지 않았다(도 3a 내지 b). 편집 수준은 벤치마크 유전자 변형 폴리펩티드와 비슷하거나 이보다 더 높았으며, 이는 중합효소 도메인을 사용하는 유전자 변형 폴리펩티드가 역전사효소 도메인을 사용하는 유전자 변형 폴리펩티드에 대한 유용한 대안임을 입증한다. 데이터가 시험한 임의의 절단(중앙 도메인의 완전한 절단을 포함함)을 포함하는 유전자 변형 폴리펩티드가 편집 활성을 가짐을 나타내므로, 결과는 PolΘ의 다양한 절단이 중합효소 도메인으로 사용될 수 있음을 추가로 입증한다.

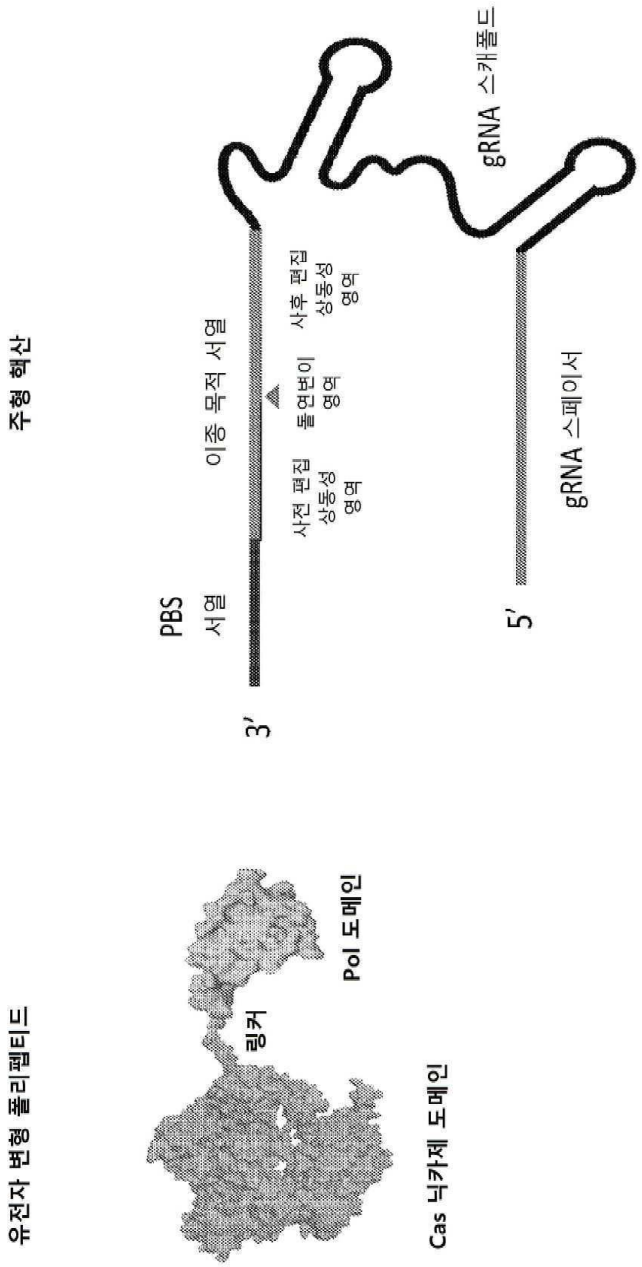
[0929] "약", "적어도", "미만", 및 "초과"와 같은 본 출원에서 일부 파라미터를 기재하는 모든 수치 범위에 대하여, 기재는 또한 언급된 값에 의해 제한되는 임의의 범위를 필수적으로 포함하는 것으로 이해되어야 한다. 따라서, 예를 들어 "적어도 1, 2, 3, 4, 또는 5"라는 기재는 특히 1 내지 2, 1 내지 3, 1 내지 4, 1 내지 5, 2 내지 3, 2 내지 4, 2 내지 5, 3 내지 4, 3 내지 5, 및 4 내지 5 등의 범위를 또한 기재한다.

[0930] 본 명세서에 인용된 모든 특허, 출원, 또는 기타 참조문헌, 예컨대 비-특허 문헌 및 참조 서열 정보에 있어서, 이들은 모든 목적을 위해 그리고 언급된 명제를 위해 그 전체가 참고로 포함되는 것으로 이해되어야 한다. 참조로 포함된 문서와 본 출원 사이에 임의의 상충이 존재하는 경우, 본 출원이 우선할 것이다. 예를 들어 게놈 유전자좌, 게놈 서열, 기능적 주석, 대립유전자 변이체, 및 참조 mRNA(예를 들어, 엑손 경계 또는 반응 요소를 포함함) 및 단백질 서열(예컨대, 보존된 도메인 구조)뿐만 아니라, 화학적 참조(예를 들어, PubChem 화합물, PubChem 물질, 또는 PubChem 생물검정 항목(구조 및 분석과 같은 그 안의 주석을 포함함) 등)를 포함하여, 유전자ID 또는 수탁 번호(통상적으로 NCBI 수탁 번호를 참조함)와 같은, 본 출원에 개시된 참조 유전자 서열과 연관된 모든 정보는 그 전체가 본 명세서에 참고로 포함된다.

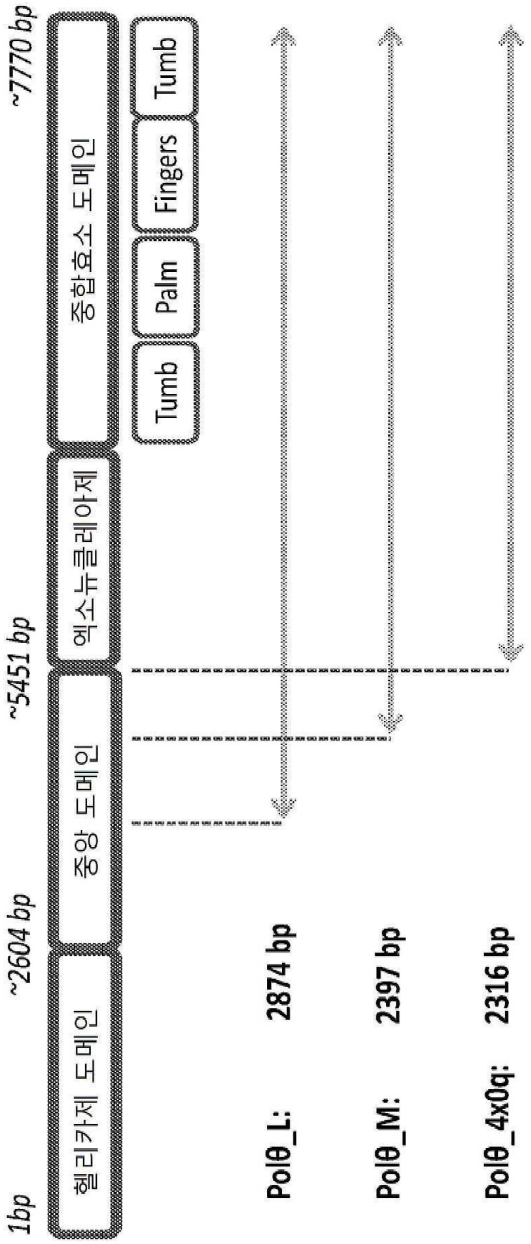
[0931] 본 출원에 사용된 제목은 단지 편의를 위한 것이며 본 출원의 해석에 영향을 미치지 않는다.

도면

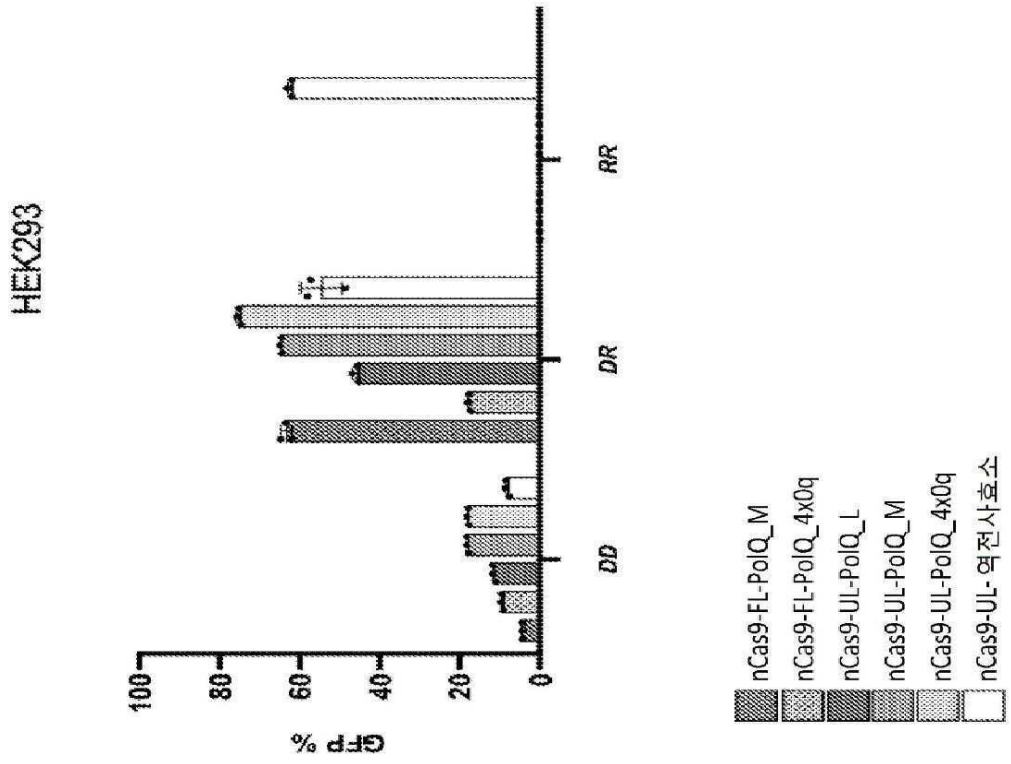
도면1


주형 핵산
이중 목적 서열
PBS
서열
3'
사전 편집 상동성 영역
돌연변이 영역
사후 편집 상동성 영역
gRNA 스캐폴드
gRNA 스페이서
5'
유전자 변형 폴리펩티드
링커
Pol 도메인
Cas 닉카제 도메인

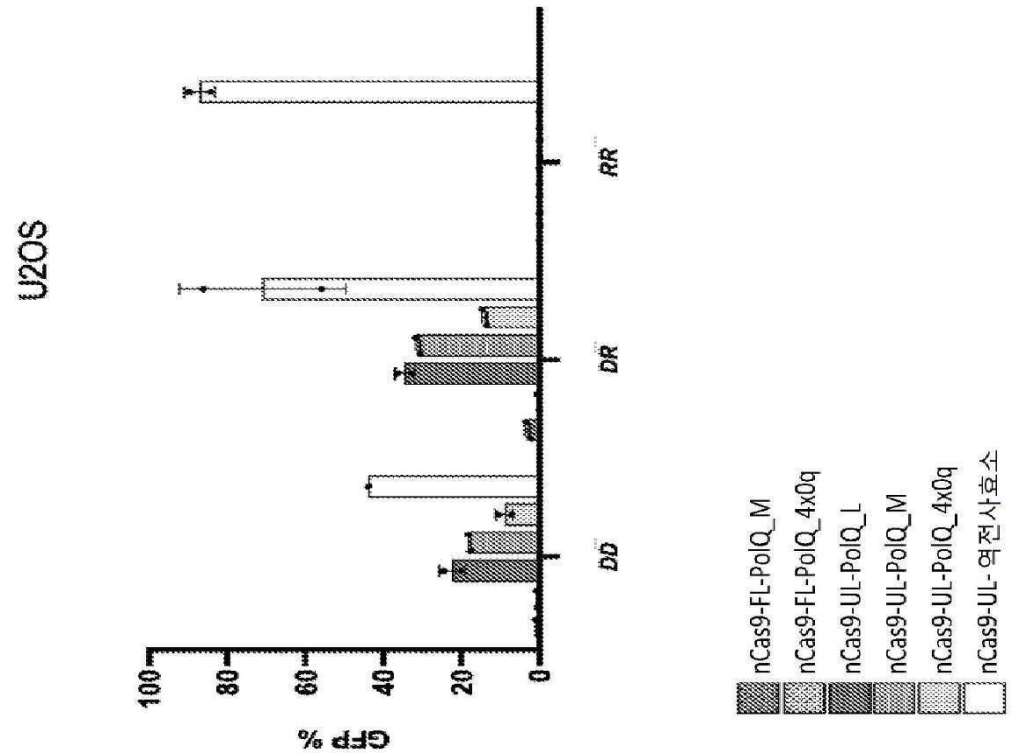
도면2


1bp
~2604 bp
~5451 bp
~7770 bp
헬리카제 도메인
중앙 도메인
엑소뉴클레아제
중합효소 도메인
Tumb
Palm
Fingers
Tumb
Polθ_L: 2874 bp
Polθ_M: 2397 bp
Polθ_4x0q: 2316 bp

도면3a


HEK293
GFP %
100
80
60
40
20
0
DD
DR
RR
nCas9-FL-PolQ_M
nCas9-FL-PolQ_4x0q
nCas9-UL-PolQ_L
nCas9-UL-PolQ_M
nCas9-UL-PolQ_4x0q
nCas9-UL-역전사효소

도면3b


U2OS
GFP %
100
80
60
40
20
0
DD
DR
RR
nCas9-FL-PolQ_M
nCas9-FL-PolQ_4x0q
nCas9-UL-PolQ_L
nCas9-UL-PolQ_M
nCas9-UL-PolQ_4x0q
nCas9-UL-역전사효소

서 열 목 록 (첨부)



아이콘을 클릭하시면 서열목록 파일이 열립니다.

본 공보 PDF는 첨부파일을 가지고 있습니다. Acrobat Reader PDF뷰어를 제공하지 않는 브라우저(크롬, 파이어폭스, 사파리 등)의 경우 첨부파일 열기가 제한되어 있으므로 Acrobat Reader PDF뷰어 설치 후 공보 PDF를 다운로드 받아 해당 뷰어에서 조회해주시기 바랍니다.