

(52) CPC특허분류

A61K 38/47 (2013.01)
A61K 38/4846 (2013.01)
C12N 15/86 (2013.01)
C12N 9/16 (2013.01)
C12N 9/2402 (2013.01)
C12N 9/644 (2013.01)
C12Y 301/06013 (2013.01)
C12Y 302/01076 (2013.01)
C12Y 304/21022 (2013.01)

(72) 발명자

밀러 제프리 씨.

미국 캘리포니아 94005 브리즈번 마리나 불러바드
7000 상가모 테라퓨틱스, 인코포레이티드 내

장 레이

미국 캘리포니아 94005 브리즈번 마리나 불러바드
7000 상가모 테라퓨틱스, 인코포레이티드 내

(30) 우선권주장

62/728,226	2018년09월07일	미국(US)
62/758,786	2018년11월12일	미국(US)
62/795,937	2019년01월23일	미국(US)
62/802,092	2019년02월06일	미국(US)

명세서

청구범위

청구항 1

제1 및 제2 아연 핑거 뉴클레아제(ZFN)을 포함하는 아연 핑거 뉴클레아제로서, 제1 ZFN은 71557로 표시된 ZFN을 포함하고 제2 ZFN은 71728로 표시된 ZFN을 포함하는, 아연 핑거 뉴클레아제.

청구항 2

제1항의 제1 및 제2 ZFN을 암호화하는 하나 이상의 폴리뉴클레오타이드로서, 선택적으로, 하나 이상의 폴리뉴클레오타이드는 AAV 벡터로 운반되는 하나 이상의 폴리뉴클레오타이드.

청구항 3

제2항에 있어서, 제1 ZFN을 암호화하는 제1 폴리뉴클레오타이드는 표 4에 나타난 서열을 포함하는 AAV 벡터를 포함하고 제2 ZFN을 암호화하는 제2 폴리뉴클레오타이드는 표 5에 나타난 서열을 포함하는 AAV 벡터를 포함하는 것을 특징으로 하는 하나 이상의 폴리뉴클레오타이드.

청구항 4

제2항 또는 제3항의 AAV 벡터로서, 제1 ZFN을 암호화하는 AAV 벡터는 SEQ ID NO:43에 제시된 서열을 포함하는, AAV 벡터.

청구항 5

제2항 또는 제3항에 있어서, 제2 ZFN을 암호화하는 AAV 벡터는 SEQ ID NO:56에 제시된 서열을 포함하는 것을 특징으로 하는 AAV 벡터.

청구항 6

제1항의 하나 이상의 ZFN, 제2항 또는 제3항의 하나 이상의 폴리뉴클레오타이드, 및/또는 제4항 및 제5항의 하나 이상의 AAV 벡터를 포함하는 세포, 선택적으로 줄기 세포 또는 전구체 세포.

청구항 7

제1항의 하나 이상의 ZFN, 제2항 또는 제3항의 하나 이상의 폴리뉴클레오타이드, 제4항 및 제5항의 하나 이상의 AAV 벡터 및/또는 제6항의 하나 이상의 세포를 포함하는 제약학적 조성물.

청구항 8

대상체에서 내인성 알부민 유전자를 절단하기 위한 제1항 내지 제7항 중 어느 한 항의 하나 이상의 ZFN, 하나 이상의 폴리뉴클레오타이드, 하나 이상의 AAV 벡터, 하나 이상의 세포 및/또는 하나 이상의 제약학적 조성물의 용도.

청구항 9

제8항에 있어서, 선택적으로 AAV 벡터로 운반되는 도너를, 도너가 대상체의 절단된 알부민 유전자에 통합되도록 대상체에게 투여하는 것을 추가로 포함하는 것을 특징으로 하는 용도.

청구항 10

대상체의 세포에서 내인성 알부민 유전자를 절단하는 방법으로서, 제1항의 하나 이상의 ZFN, 제2항 또는 제3항의 하나 이상의 폴리뉴클레오타이드, 제4항 및 제5항의 하나 이상의 AAV 벡터, 제6항의 하나 이상의 세포 및/또는 제7항에 따르는 하나 이상의 제약학적 조성물을 대상체에게 투여하는 단계를 포함하는 방법.

청구항 11

제1항의 하나 이상의 ZFN, 제2항 또는 제3항의 하나 이상의 폴리뉴클레오타이드, 제4항 및 제5항의 하나 이상의 AAV 벡터, 제6항의 하나 이상의 세포 및/또는 제7항에 따르는 하나 이상의 제약학적 조성물을 포함하는 키트.

청구항 12

조성물로서,

(a) 표 4 또는 SEQ ID NO:43에 제시된 서열을 포함하는 AAV 벡터를 선택적으로 포함하는, 71557로 표시된 제1 ZFN을 암호화하는 제1 폴리뉴클레오타이드;

(b) 표 5 또는 SEQ ID NO:56에 제시된 서열을 포함하는 AAV 벡터를 선택적으로 포함하는, 71728로 표시된 제2 ZFN을 암호화하는 제2 폴리뉴클레오타이드; 및

(c) 인자 IX (FIX) 서열을 암호화하는 서열을 포함하는 도너 폴리뉴클레오타이드를 포함하는 조성물.

청구항 13

제12항에 있어서, 도너 폴리뉴클레오타이드는 표 6에 제시된 서열, 선택적으로 SEQ ID NO:59에 제시된 서열을 포함하는 벡터인 것을 특징으로 하는 조성물.

청구항 14

제12항 또는 제13항에 있어서, 제1, 제2 및 도너 폴리뉴클레오타이드는 별도의 AAV 벡터로 운반되는 것을 특징으로 하는 조성물.

청구항 15

그것을 필요로 하는 대상체에서 FIX 도입유전자를 발현시키기 위한 제12항 내지 제14항 중 어느 한 항의 조성물의 용도로서, ZFN이 대상체의 내인성 알부민 유전자를 절단하고, FIX 서열이 절단된 알부민 유전자에 통합되며 FIX 단백질이 대상체에서 발현되도록 조성물이 대상체에게 투여되는 것인 용도.

청구항 16

제15항에 있어서, 혈우병은 FIX 단백질의 발현 후에 대상체에서 치료되는 것을 특징으로 하는 용도.

청구항 17

그것을 필요로 하는 대상체에서 FIX 단백질을 발현시키는 방법으로서, 제12항 내지 제14항 중 어느 한 항의 하나 이상의 조성물을, FIX 단백질이 세포에서 발현되도록 대상체에게 투여하는 단계를 포함하는 방법.

청구항 18

제17항에 있어서, 대상체는 혈우병을 가지며 FIX 단백질의 발현은 그 질환을 치료 및/또는 예방하는 것을 특징으로 하는 방법.

청구항 19

제12항 내지 제14항 중 어느 한 항의 하나 이상의 조성물을 포함하는 키트.

청구항 20

조성물로서,

(a) 표 4 또는 SEQ ID NO:43에 제시된 서열을 포함하는 AAV 벡터를 선택적으로 포함하는, 71557로 표시된 제1 ZFN을 암호화하는 제1 폴리뉴클레오타이드;

(b) 표 5 또는 SEQ ID NO:56에 제시된 서열을 포함하는 AAV 벡터를 선택적으로 포함하는, 71728로 표시된 제2 ZFN을 암호화하는 제2 폴리뉴클레오타이드; 및

(c) 이두로네이트-2-설파타제 (IDS) 서열을 암호화하는 서열을 포함하는 도너 폴리뉴클레오타이드

를 포함하는 조성물.

청구항 21

제19항에 있어서, 도너 폴리뉴클레오타이드는 표 7에 제시된 서열, 선택적으로 SEQ ID NO:65에 제시된 서열을 포함하는 AAV 벡터인 것을 특징으로 하는 조성물.

청구항 22

제20항 또는 제21항에 있어서, 제1, 제2 및 도너 폴리뉴클레오타이드는 별도의 AAV 벡터로 운반되는 것을 특징으로 하는 조성물.

청구항 23

그것을 필요로 하는 대상체에서 IDS 도입유전자를 발현시키기 위한 제20항 내지 제22항 중 어느 한 항의 조성물의 용도로서, ZFN이 대상체의 내인성 알부민 유전자를 절단하고, IDS 서열이 절단된 알부민 유전자로 통합되며 IDS 단백질이 대상체에서 발현되도록 조성물이 대상체에게 투여되는 것인 용도.

청구항 24

제23항에 있어서, MPS II는 IDS 서열의 발현 후에 대상체에서 처리되는 것을 특징으로 하는 용도.

청구항 25

그것을 필요로 하는 대상체에서 IDS 단백질을 발현시키는 방법으로서, 제20항 내지 제22항 중 어느 한 항의 하나 이상의 조성물을, IDS 단백질이 세포에서 발현되도록 대상체에게 투여하는 단계를 포함하는 방법.

청구항 26

제25항에 있어서, 대상체는 MPS II 질환을 가지며 IDS 단백질의 발현은 그 질환을 치료 및/또는 예방하는 것을 특징으로 하는 방법.

청구항 27

제20항 내지 제22항 중 어느 한 항의 하나 이상의 조성물을 포함하는 키트.

청구항 28

조성물로서,

(a) 표 4 또는 SEQ ID NO:43에 제시된 서열을 포함하는 AAV 벡터를 선택적으로 포함하는, 71557로 표시된 제1 ZFN을 암호화하는 제1 폴리뉴클레오타이드;

(b) 표 5 또는 SEQ ID NO:56에 제시된 서열을 포함하는 AAV 벡터를 선택적으로 포함하는, 71728로 표시된 제2 ZFN을 암호화하는 제2 폴리뉴클레오타이드; 및

(c) 알파-L 이두로니다제 (IDUA) 서열을 암호화하는 서열을 포함하는 도너 폴리뉴클레오타이드

를 포함하는 조성물.

청구항 29

제28항에 있어서, 도너는 표 8에 제시된 서열, 선택적으로 SEQ ID NO:72에 제시된 서열을 포함하는 것을 특징으로 하는 조성물.

청구항 30

제28항 또는 제29항에 있어서, 제1, 제2 및 도너 폴리뉴클레오타이드는 별도의 벡터로 운반되는 것을 특징으로 하는 조성물.

청구항 31

그것을 필요로 하는 대상체에서 IDUA 도입유전자를 발현시키기 위한, 제28항 내지 제30항 중 어느 한 항의 조성

물의 용도로서, ZFN이 대상체의 내인성 알부민 유전자를 절단하고, IDUA 서열이 절단된 알부민 유전자로 통합되며 IDUA 단백질이 대상체에서 발현되도록 조성물이 대상체에게 투여되는 것인 용도.

청구항 32

제31항에 있어서, MPS I은 IDUA 서열의 발현 후에 대상체에서 처리되는 것을 특징으로 하는 용도.

청구항 33

그것을 필요로 하는 대상체에서 IDUA 단백질을 발현시키는 방법으로서, 제28항 내지 제30항 중 어느 한 항의 하나 이상의 조성물을, IDUA 단백질이 세포에서 발현되도록 대상체에게 투여하는 단계를 포함하는 방법.

청구항 34

제33항에 있어서, 대상체는 MPS I 질환을 가지며 IDUA 단백질의 발현은 그 질환을 치료 및/또는 예방하는 것을 특징으로 하는 방법.

청구항 35

제28항 내지 제30항 중 어느 한 항의 하나 이상의 조성물을 포함하는 키트.

발명의 설명

기술 분야

[0001] 관련 출원에 대한 상호 참조

[0002] 본 출원은 2018년 2월 8일에 출원된 미국 가출원 번호 62/628,016; 2018년 9월 7일에 출원된 미국 가출원 번호 62/728,226; 2018년 11월 12일에 출원된 미국 가출원 번호 62/758,786; 2019년 1월 23일에 출원된 미국 가출원 번호 62/795,937; 및 2019년 2월 9일에 출원된 미국 가출원 번호 62/802,092의 이익을 주장하며, 이들 출원의 개시내용은 그 전문이 본원에 참조로 포함된다.

[0003] 연방정부 지원 연구 하에 이루어진 발명에 대한 권리의 진술

[0004] 적용되지 않음.

[0005] 서열 목록

[0006] 본 출원은 ASCII 포맷으로 전자 출원된 서열 목록을 포함하며 본원에 그 전문이 참조로 포함된다. 2019년 2월 8일에 생성된 상기 ASCII 복사물은 8325016940SL.txt로 명명되고, 크기가 54,302 바이트이다.

[0007] 기술 분야

[0008] 본 개시는 폴리펩타이드 및 게놈 조작 및 상동성 재조합 분야에 있다.

배경 기술

[0009] 인공적인 뉴클레아제, 예컨대, 조작된 아연 핑거 뉴클레아제 (ZFN), 전사-활성자 유사 이펙터 뉴클레아제 (TALEN), RNA 가이드된 뉴클레아제로도 불리는, 조작된 crRNA/tracr RNA ('단일 가이드 RNA')를 가진 CRISPR/Cas 시스템, 및/또는 아르곤노트(Argonaute) 시스템을 기반으로 한 뉴클레아제 (예컨대, 'TtAgo'로서 알려져 있고, 티. 썬모필루스(*T. thermophilus*)로부터 유래됨, (Swartz *et al* (2014) *Nature* 507(7491): 258-261)는 절단 도메인과 회합된 또는 그것에 작동 가능하게 연결된 DNA 결합 도메인 (뉴클레오타이드 또는 폴리펩타이드)을 포함하고, 게놈 서열의 표적화된 변경에 사용되어 왔다. 예를 들어, 뉴클레아제는 외인성 서열을 삽입하고, 하나 이상의 내인성 유전자를 비활성화하며, 변경된 유전자 발현 패턴을 가진 유기체 (예컨대, 농작물) 및 세포주를 생성하는 등의 목적을 위해 사용되어 왔다. 예컨대, 미국 특허 제 9,255,250호; 9,200,266호; 9,045,763호; 9,005,973호; 8,956,828호; 8,945,868호; 8,703,489호; 8,586,526호; 6,534,261호; 6,599,692호; 6,503,717호; 6,689,558호; 7,067,317호; 7,262,054호; 7,888,121호; 7,972,854호; 7,914,796호; 7,951,925호; 8,110,379호; 8,409,861호; 미국 특허출원 공개공보 20030232410; 20050208489; 20050026157; 20050064474; 20060063231; 20080159996; 201000218264; 20120017290; 20110265198; 20130137104; 20130122591; 20130177983 및 20130177960 및 20150056705 참조. 예를 들어, 한 쌍의 뉴클레아제

(예컨대, 아연 핑거 뉴클레아제, TALEN, dCas-Fok 융합)가 게놈 서열을 절단하기 위해 사용될 수 있다. 그 쌍의 각 구성원은 일반적으로 뉴클레아제의 하나 이상의 절단 도메인 (또는 절반 도메인)에 연결된 조작된 (자연적으로 발생하지 않는) DNA-결합 단백질을 포함한다. DNA-결합 단백질이 그것의 표적 부위에 결합할 때, 그런 DNA 결합 단백질에 결합된 절단 도메인은 게놈의 다imer화 및 후속적인 절단이 일어날 수 있게 되도록 배치된다.

[0010] 아연 핑거 단백질과 관련하여, 표적 DNA 서열에 대한 ZFP의 특이성은 아연 핑거 도메인과 특정 DNA 염기 사이의 서열 특이적 접촉에 따라 좌우된다. 더불어, 아연 핑거 도메인은 또한 DNA 골격의 포스페이트와의 비특이적 이온 쌍 상호작용에 참여하는 아미노산 잔기를 포함한다. Elrod-Erickson 등 ((1996) *Structure* 4:1171)은 아연 핑거 단백질과 그것의 동계 DNA 표적의 공동 결정화를 통해 수소 결합을 통한 DNA 골격 상의 포스페이트와 상호작용할 수 있는 특정 아미노산이 있음을 입증하였다. 잘 알려져 있는 Zif268 골격을 사용하는 아연 핑거 단백질은 전형적으로, 또한 두 번째의 변하지 않는 시스테인에 대해 두 번째 카르복실-말단인 β -시트의 제2 가닥의 아미노산 말단 잔기로서 아르기닌을 가진다. 이 위치는 그것이 α -나선의 시작에 선행하는 5번째 잔기이기 때문에, 각각의 아연 핑거 도메인 내의 (-5)로서 언급될 수 있다. 이 위치에서의 아르기닌은 DNA 골격 상의 포스페이트와, 그것의 측쇄 구아니디늄 기와 대전된 수소 결합의 형성을 통해 상호작용할 수 있다. Zif268 골격의 아연 핑거 단백질은 또한 빈번하게 첫 번째의 변하지 않는 시스테인에 대해 아미노-말단으로 4개 잔기의 위치에서 리신을 가진다. 이 위치는, 그것이 아연을 배위하는 시스테인 잔기 사이에 두 잔기를 가지는 아연 핑거에 대한 α -나선의 시작에 선행하는 14번째 잔기이기 때문에, 각 핑거 내에서 (-14)로서 언급될 수 있다. 리신은 DNA 골격 상의 포스페이트와, 그것의 측쇄 아미노기와 물-매개 대전된 수소 결합의 형성을 통해 상호작용할 수 있다. 포스페이트 기는 DNA 골격을 따라 모든 곳에서 발견되기 때문에, 아연 핑거와 DNA 분자 사이의 이런 유형의 상호작용은 일반적으로 비-서열 특이적인 것으로 여겨진다 (J. Miller, Massachusetts Institute of Technology Ph.D. Thesis, 2002).

[0011] 표적-외 절단 사건을 감소시키기 위하여, 조작된 절대 헤테로다이머 절단 절반 도메인(half-domain)이 개발되었다. 예컨대, 미국 특허 제 7,914,796호; 8,034,598호; 8,961,281호 및 8,623,618호; 미국 특허출원 공개 번호 20080131962 및 20120040398 참조. 이들 절대 헤테로다이머는 그것의 표적을 단지 상이한 조작된 절단 도메인이 ZFP에 의한 적절한 표적 부위에 배치될 때에만 다imer화하여 절단함으로써, 모노머 표적 외 절단을 감소 및/또는 제거할 수 있다.

[0012] 가장 효율적인 인공 뉴클레아제를 생성하기 위해 탐색될 수 있는 또 다른 구역은 인공 뉴클레아제를 암호화하는 유전자에 포함될 수 있는 비-코딩 서열에 있다. 예를 들어, mRNA 분자의 3' 미번역 영역 ("UTR")은 전사후 수준에서 유전자 발현을 조절하는 데 중요한 역할을 할 수 있다. 3' UTR은 mRNA 및 특이적 트랜스-작용 RNA 결합 단백질 및 비-코딩 RNA의 구조적 구성요소 사이의 조화로운 상호작용에 의해 mRNA의 발현을 제어하고 (Vislovukh *et al* (2014) *World J Biol Chem* 5(1): 40-57), 또한 폴리아데닐화 서열을 포함한다. 통상적으로 사용된 3' UTR의 예는 SV40 바이러스 폴리A 단편, 소 성장 호르몬 (BGH) 유전자로부터의 폴리 A 영역 및 토끼 베타-글로빈 UTR이다 (Ludwig, Dale (2006) *BioProcess International*, supplement). 5' UTR은 고등 진핵세포에서는 길이가 100 내지 200 bp이며 높은 GC 함량 (자주 >60%)을 포함한다. 이들 서열은 리보솜 결합을 위한 Kozak 공통 서열 및 cap 부착을 위한 서열과 같은 요소를 포함할 수 있다. 높은 GC 함량은 번역 효율에 영향을 미칠 수 있고 내부 리보솜 유입 부위 (IRES)로서 알려져 있는 복잡한 헤어핀 구조 (시스-작용 조절 서열로서 알려져 있음)를 초래할 수 있다. 5' UTR은 또한 발현을 조절하기 위하여 유전자-특이적 조절 단백질 (예를 들어 철 조절 단백질)에 결합하기 위한 서열을 가질 수 있고 또한 번역 기계와의 상호작용을 제공하는 것과 같이 다른 기능해서도 역할을 할 수 있다 (Araujo *et al* (2012) *Comp and Funct Genom* (2012) doi:10.1155/2012/475731). 통상적으로 사용되는 5' UTR 서열의 예는 베타-글로빈 5'UTR이다. UTR은 또한, 자주 3' UTR의 시스 작용 요소에 의해 매개되는, 전사 후 수준에서 유전자 조절의 공간적 제어에 역할을 할 수 있다 (Mignone *et al* (2002) *Genome Biol* 3(3): PMID: PMC139023).

발명의 내용

해결하려는 과제

[0013] 그러나, 증강된 전사/번역 효율을 제공하고 뉴클레아제 활성 및/또는 특이성을 증가시키기 위하여 조작된 뉴클레아제 절단 시스템에 대한 추가적인 방법 및 조성물에 대한 필요성이 있다.

과제의 해결 수단

[0014] 본 개시는 인공 뉴클레아제의 발현을 증가시킬뿐만 아니라, 그것의 의도된 표적에 대한 뉴클레아제 (예컨대, 뉴클레아제 쌍)의 효율 (활성) 및/또는 특이성을 증가시키기 위한 방법 및 조성물을 제공한다. 그러므로, 본원에는 인공 뉴클레아제의 발현을 증가시키는 5' 및/또는 3' 미번역 서열의 요소를 포함하는 조작된 프로모터를 포함하는 인공 뉴클레아제 (예컨대, 아연 핑거 뉴클레아제 (ZFN), TALEN, CRISPR/Cas 뉴클레아제)를 발현시키기 위한 폴리뉴클레오타이드 (예컨대, DNA 발현 벡터 또는 mRNA)가 기술된다. 선택적으로, 뉴클레아제-암호화 폴리뉴클레오타이드는 작은 펩타이드 (한정하는 것은 아니지만 펩타이드 태그 및/또는 핵 국지화 서열과 같은 다가 양이온성 펩타이드)를 암호화하는 서열을 추가로 포함하거나, 및/또는 하나 이상의 DNA 결합 도메인 영역 (예컨대, 아연 핑거 단백질 또는 TALE의 골격)에 돌연변이 및/또는 *FokI* 뉴클레아제 절단 도메인 또는 절단 절반 도메인에 하나 이상의 돌연변이를 포함한다. 이들 폴리뉴클레오타이드 구성요소가 개별적으로 또는 임의의 조합으로 사용될 때 (예컨대, FLAG (예컨대, 3X FLAG), NLS, WPRE 및/또는 폴리 A 신호와 같은 펩타이드 신호가 임의의 조합으로), 발명의 방법 및 조성물은 시험관내에서 또는 생체내에서 도입유전자의 절단 및/또는 표적화된 통합의 증가된 효율 (예컨대, 본원에서 기술된 서열/변형이 없는 뉴클레아제와 비교하여 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20배 또는 그 이상의 절단) 및/또는 표적화 특이성 (감소된 표적 외 효과)으로 인공 뉴클레아제의 발현에서 놀랍고 예상하지 못했던 증가를 제공한다. 개시는 또한 관심 있는 영역에서 세포의 크로마틴의 표적화된 절단 및/또는 세포에서 관심 있는 예정된 영역에서의 표적화된 통합을 통한 도입유전자의 통합을 위해 이들 조성물을 사용하는 방법을 제공한다.

[0015] 그러므로, 본원에는 뉴클레아제 (예컨대, ZFN, TALEN, CRISPR/Cas 뉴클레아제, 등)를 암호화하는 폴리뉴클레오타이드 (mRNA, 플라스미드, AAV와 같은 바이러스 벡터)가 기술되며, 폴리뉴클레오타이드는 다음 요소 중 적어도 하나, 2, 3, 4, 5, 6, 7 또는 8개를 임의의 조합으로 추가로 포함한다: (i) 선택적으로 뉴클레아제-암호화 서열에 대해 5'인, 다가양이온성 서열 (예컨대, 3X FLAG 서열)을 암호화하는 서열; (ii) 선택적으로 뉴클레아제-암호화 서열에 대해 5'인, 5'UTR 서열 (예컨대, SEQ ID NO:1에 제시된 것과 같은 제노푸스(*Xenopus*) 베타-글로빈 서열); (iii) 뉴클레아제 코딩 서열에 대해 3' 및/또는 5'인 WPRE 서열; (iv) 뉴클레아제 도메인의 골격 (비-DNA 결합 잔기) (예컨대, ZFN의 포스페이스트 골격 잔기)를 암호화하는 서열에 대한 변형; (v) 뉴클레아제의 절단 도메인 서열 (예컨대, 조작된 *FokI* 도메인)에 대한 변형; (vi) 선택적으로 뉴클레아제를 암호화하는 서열에 작동 가능하게 연결된 조직 특이적 프로모터 및/또는 인핸서 (예컨대, hAAT, ApoE, 등); (vii) NLS 서열 (뉴클레아제 암호화 서열에 대해 5' 또는 3'); 및/또는 (viii) 폴리A 서열. 특정 구체예에서, 폴리뉴클레오타이드는 mRNA이다. 다른 구체예에서, 폴리뉴클레오타이드는 추가로 선택적으로 ITR을 포함하는 AAV 벡터, 예를 들어 첨부된 도면 및/또는 표의 임의의 구성물에서 나타낸 것과 같은 AAV 벡터이다. 단일 폴리뉴클레오타이드는 뉴클레아제의 일부 또는 모든 구성요소, 예컨대, 한 쌍의 ZFN, 단일 가이드 RNA, 등을 암호화할 수 있다. 대안으로, 별도의 폴리뉴클레오타이드 (동일하거나 상이한 유형임)는 뉴클레아제의 구성요소를 암호화할 수 있는데, 예를 들어, 별도의 뉴클레오타이드가 각각 한 ZFN 또는 ZFN의 TALEN 또는 TALEN 쌍을 암호화한다. 그러므로, 본원에는 하나 이상의 뉴클레아제 (예컨대, ZFN)를 암호화하는 하나 이상의 폴리뉴클레오타이드 (예컨대, AAV 벡터)가 제공된다. 본원에 기술된 폴리뉴클레오타이드는 적어도 하나의 도너의 표적화된 절단 및/또는 통합을 위한 시험관내, 생체의 및/또는 생체내 방법에 사용될 수 있고 뉴클레아제 활성 (절단 및/또는 통합) 및/또는 특이성 (표적 외 활성에 비교하여 표적-상)을 1 내지 50배 (또는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20배, 등을 포함하여 그 사이의 임의의 값) 증가시킬 수 있다.

[0016] 한 측면으로, 발명은 원하는 조직에서 인공 뉴클레아제를 발현시키기 위한 조직 특이적 프로모터를 포함하는 폴리뉴클레오타이드를 기술한다. 일부 구체예에서, 조직 특이적 프로모터는 간-특이적 프로모터이다. 추가의 구체예에서, 간 특이적 프로모터는 인간- $\alpha 1$ 항-트립신 프로모터 (hAAT) 또는 트랜스티레틴 최소 프로모터이다 (미국 특허 공보 20170119906 참조). 일부 경우에, 간 특이적 프로모터는 ApoE 인핸서 서열을 포함한다 (Shachter *et al.* (1993) *J. Lipid Res* 34(10):1699-707). 일부 구체예에서, 간 특이적 프로모터는 하나 이상의 ApoE 인핸서 서열을 포함한다 (예컨대, 1, 2, 3 및/또는 4개; Okuyama *et al.* (1996) *Hum Gen Ther* 7(5):637-45 참조). 추가의 구체예에서, 프로모터는 인트론에 연결된다. 바람직한 구체예에서, 인트론은 인간 β -글로빈 유전자 및 가지의 제1 인트론으로부터의 5' 도너 부위 및 면역글로불린 유전자 중쇄 가변 영역의 인트론으로부터의 3' 엑서터 부위를 포함하는 HGG-IGG 키메라 인트론이다. 본원에서 기술된 폴리뉴클레오타이드는 cDNA 구성물 (예를 들어 AAV와 같은 바이러스 벡터로 운반된), mRNA, 플라스미드 DNA, 또는 게놈에 삽입하기 위한 발현 카세트의 일부분일 수 있다.

[0017] 그러므로, 한 측면으로, 본원에는 뉴클레아제를 암호화하는 mRNA 또는 AAV 벡터가 기술되며, AAV 벡터는 전사 및 번역 효율을 증가시키기 위한 요소들을 포함한다. 일부 구체예에서, 요소는 자연적인 또는 인공적인 5' 및/

또는 3' UTR 서열과 같은 미변역 서열을 포함한다. 일부 측면으로, 5' UTR 서열이 발현 카세트에 포함되며, 한편으로 다른 측면에서는, 3' UTR 서열이 사용된다. 바람직한 구체예에서, 인공 뉴클레아제를 암호화하는 mRNA 또는 AAV는 5' UTR 및 3'UTR 둘 다를 포함한다. 한 구체예에서, 5' UTR은 제노푸스 β -글로빈 UTR이다 (Falcone and Andrews (1991) *Mol Cell Bio* 11(5): 2656-2664; Krieg and Melton (1994) *Nuc Acid Res* 22(18):7057 참조). 바람직한 구체예에서, 제노푸스 β -글로빈 UTR을 암호화하는 DNA 서열은 5' [TG]CTGTCTCTTTTGCAGAAGCTCAGAATAAACGCTCAACTTGGCAGAT (SEQ ID NO:1) (TG는 선택적임)이다. 일부 측면으로, 뉴클레아제를 암호화하는 mRNA 또는 AAV는 3' WPRE 서열을 포함한다 (미국 특허출원 공보 20160326548 참조). 추가 구체예에서, WPRE 요소는 단백질 X의 발현을 방지하기 위하여 'X' 영역에서 돌연변이된다 (미국 특허 제 7,419,829호 참조). 일부 구체예에서, 돌연변이된 WPRE 서열 is a truncated WPRE 요소. 일부 구체예에서 돌연변이된 WPRE 서열은 J02442 또는 J04514 우드척 간염 바이러스의 X 영역에서 돌연변이된다 (Galibert *et al* (1982) *J. Virol* 41(1): 51-65; Zanta-Boussif *et al.* (2009) *Gene Ther* 16(5):605-619). 본원에서 기술된 폴리뉴클레오타이드에 사용될 수 있는 WPRE 서열의 비제한적인 예는 하기 실시예에서 제시된다. 추가 측면으로, 3' UTR은 폴리 A 신호 서열을 포함한다. 폴리 A 신호는 WPRE 서열에 대해, 이들 요소가 조합하여 사용될 때 3' 또는 5'일 수 있다. 바람직한 구체예에서, 폴리 A 신호 서열은 소 성장 호르몬 신호 서열이다 (Woychik *et al* (1984) *Proc Natl Acad Sci* 81(13):3944-8 참조). 본원에서 기술된 뉴클레아제-암호화 폴리뉴클레오타이드 (mRNA, AAV 벡터)는 추가로 뉴클레아제 암호화 서열에 대한 변형, 예를 들어 뉴클레아제의 ZFP DNA-결합 도메인의 골격 영역에 대한 변형 및/또는 뉴클레아제(들)의 절단 도메인 (또는 절단 절반 도메인)에 대한 변형을 포함할 수 있다.

[0018] 또 다른 측면으로, 본원에는 하나 이상의 돌연변이가 유래되는 모(parent) (예컨대, 야생형) 절단 도메인과 비교하여 이들 돌연변이를 포함하는 조작된 뉴클레아제 절단 절반 도메인 및/또는 이들 절단 도메인을 포함하는 폴리뉴클레오타이드 (mRNA)가 기술된다. 본원에서 기술된 돌연변이로는, 한정하는 것은 아니지만, 절단 도메인의 전하를 변경시키는 돌연변이, 예를 들어 양전하 잔기의 양전하가 아닌 잔기로의 돌연변이 (예컨대, K 및 R 잔기의 돌연변이 (예컨대, S로 돌연변이됨); N 잔기의 돌연변이 (예컨대, D로 돌연변이됨), 및 Q 잔기의 돌연변이 (예컨대, E로 돌연변이됨)); 분자 모델링을 토대로 한 DNA 골격에 가까운 것으로 예측되고 *FokI* 상동체에서 변이를 보이는 잔기로의 돌연변이; 및/또는 다른 잔기에서의 돌연변이 (예컨대, 미국 특허 제 8,623,618호 및 Guo *et al.*, (2010) *J. Mol. Biol.* 400(1):96-107)를 들 수 있다.

[0019] 특정 구체예에서, 조작된 절단 절반 도메인은 *FokI* 또는 *FokI* 상동체로부터 유래되고 SEQ ID NO:2에 제시된 야생형 전장(full length) *FokI* 또는 *FokI* 상동체의 상응하는 잔기와 비교하여 넘버링된, 아미노산 잔기 416, 422, 447, 448, 및/또는 525 중 하나 이상에서의 돌연변이를 포함한다. 다른 구체예에서, *FokI*로부터 유래된 절단 절반 도메인은 야생형 *FokI* 또는 임의의 *FokI* 상동체의 상응하는 잔기와 관련하여 넘버링된, 387, 393, 394, 398, 400, 416, 418, 422, 427, 434, 439, 441, 442, 444, 446, 448, 472, 473, 476, 478, 479, 480, 481, 487, 495, 497, 506, 516, 523, 525, 527, 529, 534, 542, 559, 569, 570, 및/또는 571 중 하나 이상을 포함하는, 아미노산 잔기 414-426, 443-450, 467-488, 501-502, 및/또는 521-531 중 하나 이상의 잔기에서의 돌연변이를 포함한다. 돌연변이는 상응하는 위치에서 *FokI*에 상동하는 자연적인 제한 효소에서 발견된 잔기에 대한 돌연변이를 포함할 수 있다. 특정 구체예에서, 돌연변이는 치환, 예를 들어 야생형 잔기의 임의의 상이한 아미노산, 예를 들어 알라닌 (A), 시스테인 (C), 아스파르트산 (D), 글루탐산 (E), 히스티딘 (H), 페닐알라닌 (F), 글리신 (G), 아스파라긴 (N), 세린 (들) 또는 트레오닌 (T)으로의 치환이다. 특정 구체예에서, *FokI* 뉴클레아제 도메인은 416, 422, 447, 479 및/또는 525 (야생형, SEQ ID NO:2와 관련하여 넘버링됨) 중 하나 이상에서의 돌연변이를 포함한다. 뉴클레아제 도메인은 또한, 한정하는 것은 아니지만 ELD, KKR, ELE, KKS를 포함한다. 위치 418, 432, 441, 448, 476, 481, 483, 486, 487, 490, 496, 499, 523, 527, 537, 538 및 559에서의 하나 이상의 돌연변이를 포함할 수 있다. 예컨대, 미국 특허 제 8,623,618호 참조. 또한 추가의 구체예에서, 절단 도메인은 하나 이상의 잔기 (예컨대, 419, 420, 425, 446, 447, 470, 471, 472, 475, 478, 480, 492, 500, 502, 521, 523, 526, 530, 536, 540, 545, 573 및/또는 574)에서의 돌연변이를 포함한다. 특정 구체예에서, 본원에서 기술된 변이체 절단 도메인은 뉴클레아제 다이머화에 관여하는 잔기에 대한 돌연변이 (다이머화 도메인 돌연변이), 및 하나 이상의 추가적인 돌연변이; 예를 들어 포스페이트 접촉 잔기에 대한 돌연변이: 예컨대, 다이머화 도메인 밖의 아미노산 위치에서의, 예를 들어 포스페이트 접촉에 참여할 수 있는 아미노산 잔기에서의 1, 2, 3, 4, 5, 6개 또는 그 이상의 돌연변이와 조합된 다이머화 돌연변이 (예컨대 ELD, KKR, ELE, KKS, 등)를 포함한다. 바람직한 구체예에서, 416, 422, 447, 448 및/또는 525에서의 돌연변이는 양전하 아미노산의 대전되지 않은 또는 음전하 아미노산으로의 대체를 포함한다. 다른 구체예에서, 위치 446, 472 및/또는 478 (및 선택적으로 예를 들어 다이머화 또는 촉매적 도메인의 추가적인 잔기)에서의 돌연변이가 만들어진다. 특정 구체예에서, 조작

된 절단 절반 도메인은 위치 542 (예컨대, N542D) 및/또는 478 (예컨대, P478S)에서의 돌연변이를 포함한다. 또한 조작된 절단 도메인의 헤테로다이어머, 예를 들어, 한 개의 조작된 절단 도메인 (예컨대, N542D)을 포함하는 제1 (좌측) 뉴클레아제 및 상이한 조작된 절단 도메인 (예컨대, P478S)을 포함하는 제2 뉴클레아제가 기술된다.

[0020] 상기에서 기술된 임의의 조작된 절단 절반 도메인은 인공 뉴클레아제 (및 이들 인공 뉴클레아제를 발현하는 폴리뉴클레오타이드)에, 이것들과, 예를 들어, 한정하는 것은 아니지만 아연 핑거 뉴클레아제, TALEN, CRISPR/Cas 뉴클레아제, 등을 포함한, DNA-결합 도메인과의 회합에 의해 통합될 수 있다. 아연 핑거 뉴클레아제의 아연 핑거 단백질은 비정식 아연-배위 잔기를 포함할 수 있다 (예컨대, 정식 C2H2 구성보다는 CCHC, 미국 특허 제 9,234,187호 참조).

[0021] 또 다른 측면으로, 인공 뉴클레아제를 생성하는, 본원에서 기술된 DNA 결합 도메인 및 조작된 FokI 또는 그것의 상동체 절단 절반 도메인을 포함하는 융합 분자가 제공된다. 특정 구체예에서, 융합 분자의 DNA-결합 도메인은 아연 핑거 결합 도메인 (예컨대, 조작된 아연 핑거 결합 도메인)이다. 다른 구체예에서, DNA-결합 도메인은 TALE DNA-결합 도메인이다. 또한 추가의 구체예에서, DNA 결합 도메인은 DNA 결합 분자 (예컨대, 가이드 RNA) 촉매적으로 비활성인 Cas9 또는 Cfp1 단백질 (dCas9 또는 dCfp1)을 포함한다. 일부 구체예에서, 조작된 융합 분자는 다이머 뉴클레아제가 이중 가닥 분자의 한 가닥만 절단하여, 니카아제를 형성할 수 있도록 촉매적으로 비활성인 조작된 절단 절반 도메인과 뉴클레아제 복합체를 형성한다 (미국 특허 제 9,200,266 참조).

[0022] 발명의 방법 및 조성물은 또한 DNA 골격 상의 포스페이트와 비특이적으로 상호작용할 수 있는 표적 서열의 뉴클레오타이드를 인식하는 잔기 밖의 DNA 결합 도메인 내에 있는 하나 이상의 아미노산에 대한 돌연변이 (예컨대, 'ZFP 골격' (DNA 인식 나선 영역 밖의) 또는 'TALE 골격' (RVD 밖의)에 대한 하나 이상의 돌연변이)를 포함할 수 있다. 그러므로, 특정 구체예에서, 발명은 뉴클레오타이드 표적 특이성에 필요하지 않은 ZFP 골격의 양이온 아미노산 잔기의 돌연변이를 포함한다. 일부 구체예에서, ZFP 골격의 이들 돌연변이는 양이온 아미노산 잔기를 중성 또는 음이온 아미노산 잔기로 돌연변이시키는 것을 포함한다. 일부 구체예에서, ZFP 골격의 이들 돌연변이는 극성 아미노산 잔기를 중성 또는 비극성 아미노산 잔기로 돌연변이시키는 것을 포함한다. 바람직한 구체예에서, 돌연변이는 DNA 결합 나선과 관련하여 위치 (-5), (-9) 및/또는 위치 (-14)에서 만들어진다. 일부 구체예에서, 아연 핑거는 (-5), (-9) 및/또는 (-14)에서 하나 이상의 돌연변이를 포함할 수 있다. 추가의 구체예에서, 다중-핑거 아연 핑거 단백질의 하나 이상의 아연 핑거는 (-5), (-9) 및/또는 (-14)에 돌연변이를 포함할 수 있다. 일부 구체예에서, (-5), (-9) 및/또는 (-14)에서의 아미노산 (예컨대, 아르기닌 (R) 또는 리신 (K))은 알라닌 (A), 류신 (L), Ser (S), Asp (N), Glu (E), Tyr (Y) 및/또는 글루타민 (Q)으로 돌연변이된다. 예컨대, 미국 공개 번호 US-2018-0087072 참조.

[0023] 또 다른 측면으로, 본원에 기술된 조작된 절단 절반 도메인 또는 융합 분자 (인공 뉴클레아제를 포함함) 중 임의의 것을 암호화하는 폴리뉴클레오타이드가 제공된다. 적합한 폴리뉴클레오타이드의 비제한적인 예로는 mRNA, cDNA, 바이러스 벡터 (AAV, Ad, LV), 및/또는 비-바이러스 벡터 (플라스미드 벡터)를 들 수 있다.

[0024] 일부 측면으로, 발명의 방법 및 조성물은 진핵성 도입유전자 서열에 융합된 외인성 펩타이드 서열을 암호화하는 서열의 사용을 포함한다. 일부 구체예에서, 외인성 펩타이드는 번역 후 단백질 서열에 융합되고, 다른 구체예에서, 외인성 펩타이드를 암호화하는 서열은 인공 뉴클레아제 (예컨대, 융합 단백질)를 암호화하는 서열에 대한 프레임 내에서 (3' 및/또는 5'에) 연결된다. 외인성 펩타이드는 정제 또는 표지화, 예컨대, 친화성 정제 또는 면역조각화학에 용이한 서열을 암호화할 수 있다. 이러한 펩타이드의 예는 폴리히스티딘 태그 ("His 태그", Hochuli *et al* (1988), *Bio/Technol* 6(11):1321-5) 또는 Flag 태그와 같은 양이온성 펩타이드 태그 (Hopp *et al* (1988) *Bio/Technol* 6(10):1204-10; Hernan *et al*. (2000) *BioTechniques* 28(4), 789-793)이다. 이들 펩타이드 태그 중 하나 이상 (1, 2, 3, 4, 5개 또는 그 이상)이 임의의 조합으로 사용될 수 있다. 일부 구체예에서, 서열 N-말단 DYKDDDK (SEQ ID NO:3)를 포함하는 외인성 Flag 펩타이드를 암호화하는 서열은 인공 뉴클레아제를 암호화하는 서열의 C-말단 또는 N-말단에서 프레임 내로 융합된다. 바람직한 구체예에서, 3개의 FLAG 서열을 암호화하는 서열 (3x FLAG 펩타이드)이 사용되는데 (미국 특허 제 6,379,903호 참조), 여기서 아미노산 서열은 N-말단 (M)DYKDHDG-DYKDHDG-DYKDDDDK (SEQ ID NO:4)이고, N-말단 메티오닌 (M)은 선택적이다. 하나 이상의 이러한 펩타이드 서열 (예컨대, 3x FLAG와 같은 다가양이온성 서열)의 포함은 그 펩타이드 서열이 없는 뉴클레아제와 비교하여 뉴클레아제 (절단) 활성을 2, 3, 4, 5, 6, 7, 8, 9, 10, 11배 또는 그 이상 증가시킬 수 있다.

[0025] 일부 측면으로, 공 뉴클레아제를 암호화하는 mRNA는 핵 국지화 펩타이드 서열 (NLS)을 포함한다. 일부 구체예에서, NLS는 SV40 바이러스 큰 T 유전자로부터의 서열 PKKKRKV (SEQ ID NO:5)를 포함하고 (Kalderon *et al* (1984) *Nature* 311(5981):33-8 참조), 한편 다른 예에서, NLS는 c-myc 단백질로부터의 서열 PAAKRVKLD (SEQ ID

NO:6)를 포함한다 (Dang and Lee (1988) *Mol Cell Biol* 8(10):4048-54 참조). 일부 구체예에서, NLS는 간염 델타 바이러스로부터의 서열 EGAPPAKRAR (SEQ ID NO:7) (Alves *et al* (2008) *Virology* 370: 12-21 참조) 또는 폴리오마 T 단백질로부터의 VSRKRPRP (SEQ ID NO:8)을 포함한다 (Richardson *et al* (1986) *Cell* 44(1):77-85). 다른 구체예에서, NLS는 뉴클레오플라스민 카르복시 꼬리로부터 유래된 서열 KRPAATKKAGQAKKKLD (SEQ ID NO:9)를 포함하는 한편 (Dingwall (1988) *J Cell Biol* 107:841-849 및 Robbins *et al* (1991) *Cell* 64(3):615-23), 일부 구체예에서, NLS는 Siomi 및 Dreyfuss에 의해 처음으로 기술된 서열 NQSSNFGPMKGGNFGRRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:10)를 포함한다 (Siomi and Dreyfus (1995) *J Cell Biol* 129(3):551-560). 추가의 구체예에서, NLS는 HTLV-1의 Rex 단백질로부터의 서열 PKTRRRRPRRSQRKRPT (SEQ ID NO:11)를 포함한다 (Siomi *et al* (1988) *Cell* 55(2):197-209). 본원에서 기술된 것과 같이 하나 이상의 NLS 서열의 포함은 그 펩타이드 서열이 없는 뉴클레아제와 비교하여 뉴클레아제 (절단) 활성을 2, 3, 4, 5, 6, 7, 8, 9, 10, 11배 또는 그 이상 증가시킬 수 있다.

[0026] 또 다른 측면으로, 본원에서 기술된 뉴클레아제, 폴리펩타이드 (예컨대, 융합 분자 또는 융합 폴리펩타이드) 및/또는 폴리뉴클레오타이드 중 임의의 것을 포함하는 세포가 제공된다. 한 구체예에서, 세포는 한 쌍의 융합 폴리펩타이드를 포함하는데, 한 융합 폴리펩타이드는 아미노산 잔기 393, 394, 398, 416, 421, 422, 442, 444, 447, 448, 473, 480, 530 및/또는 525에서의 하나 이상의 돌연변이에 더불어 ELD 또는 ELE 절단 절반 도메인을 포함하고, 한 융합 폴리펩타이드는 잔기 393, 394, 398, 416, 421, 422, 442, 444, 446, 447, 448, 472, 473, 478, 480, 530 및/또는 525에서의 하나 이상의 돌연변이에 더불어 KKK 또는 KKR 절단 절반 도메인을 포함한다 (미국 특허 제 8,962,281호 참조). 일부 구체예에서, 한 융합 단백질은 N542D와 같은 FokI의 잔기 542 (절단 도메인의 잔기 159)에서의 돌연변이를 포함하고, 한 융합 폴리펩타이드는 P478S와 같은 FokI의 잔기 478 (절단 도메인의 잔기 95)에서의 돌연변이를 포함한다.

[0027] 본원에서 기술된 이들 융합 폴리펩타이드 중 임의의 것에서, ZFP 파트너는 추가로 (-5), (-9) 및/또는 (-14) 위치에서 아연 핑거 DNA 결합 도메인의 돌연변이를 포함할 수 있다. 일부 구체예에서, 위치 -5에서의 Arg (R)은 Tyr (Y), Asp (N), Glu (E), Leu (L), Gln (Q), 또는 Ala (A)로 변경된다. 다른 구체예에서, 위치 (-9)에서의 Arg (R)은 Ser (S), Asp (N), 또는 Glu (E)로 대체된다. 추가의 구체예에서, 위치 (-14)에서의 Arg (R)은 Ser (S) 또는 Gln (Q)으로 대체된다. 또 다른 구체예에서, 융합 폴리펩타이드는 (-5), (-9) 및/또는 (-14) 위치에서의 아미노산이 상기 열거된 아미노산 중 어느 것으로든 임의의 조합으로 변경되는, 아연 핑거 DNA 결합 도메인에서의 돌연변이를 포함할 수 있다.

[0028] 본원에는 또한 본원에 기술된 하나 이상의 인공 뉴클레아제를 포함하는 세포로부터 유래된 및/또는 분화된 세포를 포함하는, 발명의 폴리펩타이드 및/또는 폴리뉴클레오타이드에 의해 변형된 세포가 제공된다. 일부 구체예에서, 세포는 도입유전자의 뉴클레아제-매개 삽입, 또는 유전자의 뉴클레아제-매개 녹아웃(knock out)을 포함한다. 변형된 세포, 및 변형된 세포로부터 유래된 임의의 세포는 일시적인 것 이상으로 발명의 뉴클레아제를 반드시 포함하지 않지만, 이러한 뉴클레아제에 의해 매개된 게놈 변형은 남아 있다.

[0029] 또 다른 측면으로, 관심 있는 영역에서 세포 크로마틴의 표적화된 절단 방법; 세포에서 상동성 재조합을 유발시키는 방법; 감염을 치료하는 방법; 및/또는 질환을 치료하는 방법이 제공된다. 이들 방법은 시험관내에서, 생체 외에서 또는 생체내에서 또는 이것들의 조합으로 실행될 수 있다. 방법은 세포 크로마틴을 세포의 관심있는 예정된 영역에서 본원에 기술된 하나 이상의 뉴클레아제를 사용하여 절단하는 단계를 포함한다. 특정 구체예에서, 본원에 기술된 한 쌍의 융합 폴리펩타이드 (즉, 하나 또는 두 융합 폴리펩타이드(들)가 모두 본원에 기술된 조각된 절단 절반 도메인을 포함하는 한 쌍의 융합 폴리펩타이드). 특정 구체예에서, 표적 부위 상의 표적화된 절단은 본원에서 기술된 돌연변이가 없는 절단 도메인과 비교하여, 50%-60% (또는 그 사이의 임의의 값), 60%-70% (또는 그 사이의 임의의 값), 70%-80% (또는 그 사이의 임의의 값), 80%-90% (또는 그 사이의 임의의 값), 90% 내지 200% (또는 그 사이의 임의의 값)를 포함하여, 적어도 50 내지 200% (또는 그 사이의 임의의 값) 또는 그 이상만큼 증가된다. 유사하게, 본원에 기술된 방법 및 조성물을 사용하여, 표적-외 부위 절단은, 한정하는 것은 아니지만 1 내지 50배 (또는 그 사이의 임의의 값)를 포함하여, 1 내지 100배 또는 그 이상 감소된다. 특정 구체예에서, 뉴클레아제 활성의 표적화된 절단은 본원에 기술된 뉴클레아제를 암호화하는 구성물이 변형을 포함하지 않을 때와 비교하여, 50%-60% (또는 그 사이의 임의의 값), 60%-70% (또는 그 사이의 임의의 값), 70%-80% (또는 그 사이의 임의의 값), 80%-90% (또는 그 사이의 임의의 값), 90% 내지 200% (또는 그 사이의 임의의 값)를 포함하여 적어도 50 내지 200% (또는 그 사이의 임의의 값) 또는 한정하는 것은 아니지만 1 내지 50배 (또는 그 사이의 임의의 값)를 포함하여 1 내지 100배 또는 그 이상 증가된다 (증대).

[0030] 본원에 기술된 인공 뉴클레아제 (및 그것을 암호화하는 폴리뉴클레오타이드)는 세포의 관심 있는 영역에서 세포

크로마틴의 표적화된 절단 방법 및/또는 관심 있는 예정된 영역에서의 상동성 재조합 방법에 사용될 수 있다. 세포는 배양된 세포, 세포주, 유기체의 세포, 세포 및/또는 그것의 자손이 치료 후 유기체로 복귀되는 경우에 치료를 위해 유기체로부터 제거되었던 세포, 및 유기체로부터 제거되고, 발명의 융합 분자를 사용하여 변형된 후, 치료 방법 (세포 요법)으로 유기체로 복귀되는 세포를 포함한다. 세포 크로마틴에서 관심 있는 영역은, 예를 들어, 게놈 서열 또는 그것의 일부분일 수 있다. 조성물은 DNA 결합 분자 (예컨대, 조작된 아연 핑거 또는 TALE 결합 도메인 또는 조작된 CRISPR 가이드 RNA) 및 기술된 것과 같은 절단 절반 도메인을 포함하는 융합 분자 또는 융합 분자를 암호화하는 폴리뉴클레오타이드를 포함한다.

[0031] 융합 분자는 세포에서, 예컨대, 융합 분자를 세포에 폴리펩타이드로서 전달함으로써, 또는 융합 분자를 암호화하는 폴리뉴클레오타이드를 세포에 전달함으로써 세포에서 발현될 수 있고, 여기서 폴리뉴클레오타이드는, 만약 DNA라면, 융합 분자를 생성하기 위하여 전사되고 번역된다. 추가로, 만약 폴리뉴클레오타이드가 융합 분자 (또는 그것의 구성요소)를 암호화하는 mRNA라면 mRNA가 세포에 전달된 후, mRNA는 번역되고, 그로써 융합 분자가 생성된다.

[0032] 발명의 다른 측면으로 조작된 뉴클레아제 특이성을 증가시키기 위한 방법 및 조성물이 제공된다. 한 측면으로, 표적 외 절단 활성을 감소시킴으로써 전체 표적-상 절단 특이성을 증가시키는 방법이 제공된다. 일부 구체예에서, 조작된 뉴클레아제 복합체의 구성요소인 조작된 절단 절반 도메인-함유 파트너는 세포와 접촉하기 위해 사용되고, 이때 복합체의 각 파트너는 다른 파트너에 대해 1 대 1 이외의 비율로 제공된다. 일부 구체예에서, 두 파트너 (절반 절단 도메인)의 비율은 1:2, 1:3, 1:4, 1:5, 1:6, 1:8, 1:9, 1:10 또는 1:20 비율, 또는 그 사이의 임의의 값으로 제공된다. 다른 구체예에서, 두 파트너의 비율은 1:30보다 크다. 다른 구체예에서, 두 파트너는 1:1과는 상이하도록 선택된 비율로 사용된다. 일부 측면으로, 각 파트너는 세포에 mRNA로서 전달되거나 또는 각 파트너를 암호화하는 mRNA 또는 벡터의 상이한 양이 전달되는 바이러스 또는 비-바이러스 벡터로 전달된다. 추가의 구체예에서, 뉴클레아제 복합체의 각 파트너는 단일 바이러스 또는 비-바이러스 벡터 상에 포함될 수 있지만, 한 파트너가 다른 것보다 높거나 낮은 값으로 발현되어, 결국 1 대 1과는 다른 절단 절반 도메인 비율로 세포에 전달되도록 의도적으로 발현된다. 일부 구체예에서, 각각의 절단 절반 도메인은 상이한 발현 효율로 상이한 프로모터를 사용하여 발현된다. 다른 구체예에서, 두 절단 도메인은 둘 다 동일한 오픈 리딩 프레임으로부터 발현되지만, 두 파트너를 암호화하는 유전자는 3' 파트너가 더 낮은 속도로 발현되는 것을 초래하는 서열 (예컨대, 자기 절단 2A 서열 또는 IRES)에 의해 분리되어서, 두 파트너의 비율이 1:2, 1:3, 1:4, 1:5, 1:6, 1:8, 1:9, 1:10 또는 1:20 비율, 또는 그 사이의 임의의 값이 되는 바이러스 또는 비-바이러스 벡터를 사용하여 세포에 전달된다. 다른 구체예에서, 두 파트너는 1:1과는 상이하도록 선택된 비율로 배치된다.

[0033] 따라서, 또 다른 측면으로, 관심 있는 영역에서 세포 크로마틴을 절단하는 방법은 (a) 관심 있는 영역에서 제1 서열을 선택하는 단계; (b) 제1 서열에 특이적으로 결합하도록 제1 DNA-결합 분자를 조작하는 단계; (c) 세포에서 제1 융합 분자를 발현시키는 단계로, 제1 융합 분자는 제1 DNA-결합 분자 (예컨대, 아연 핑거, TALE, sgRNA), 및 절단 도메인 (또는 절반 도메인)을 포함하는 것인 단계; 및 (d) 세포에서 제2 융합 분자를 발현시키는 단계로, 제2 융합 분자는 제2 DNA-결합 도메인, 및 제2 절단 도메인 (또는 절반 도메인)을 포함하는 것인 단계를 포함할 수 있고, 여기서, 융합 분자의 적어도 하나는 본원에 기술된 링커를 포함하고, 추가로 제1 융합 분자는 제1 서열에 결합하며, 제2 융합 분자는 제1 서열로부터 2 내지 50개 뉴클레오타이드에 위치한 제2 서열에 결합하여서, 활성 뉴클레아제 복합체가 형성되고 세포 크로마틴이 관심 있는 영역에서 절단된다. 특정 구체예에서, 두 융합 분자는 모두 DNA 결합 도메인과 촉매적 뉴클레아제 도메인 사이에 본원에 기술된 링커를 포함한다.

[0034] 또한 예를 들어 표적화된 돌연변이를 도입하기 위하여 세포 크로마틴의 영역 (예컨대, 내인성 유전자)을 변경시키는 방법이 제공된다. 특정 구체예에서, 세포 크로마틴을 변경시키는 방법은 예정된 부위에서 세포 크로마틴에 이중 가닥 절단을 생성하기 위하여 하나 이상의 표적화된 뉴클레아제, 및 절단 영역의 세포 크로마틴의 뉴클레오타이드 서열에 대한 상동성을 가진 도너 폴리뉴클레오타이드를 세포에 도입하는 단계를 포함한다. 세포 DNA 수복 과정은 이중 가닥 절단의 존재에 의해 활성화되고 도너 폴리뉴클레오타이드는 절단의 수복을 위한 주형으로서 사용되어, 세포 크로마틴에 도너의 뉴클레오타이드 서열의 전부 또는 일부가 도입되는 결과를 초래한다. 그러므로, 세포 크로마틴의 서열은 변경될 수 있고, 특정 구체예에서, 도너 폴리뉴클레오타이드에 존재하는 서열로 전환될 수 있다. 하나 이상의 표적이 본원에 기술된 방법 및 조성물을 사용하여 변경될 수 있다.

[0035] 표적화된 변경으로는, 한정하는 것은 아니지만, 점 돌연변이 (즉, 단일 염기쌍의 상이한 염기쌍으로의 전환), 치환 (즉, 다수의 염기쌍의 동일한 길이의 상이한 서열로의 전환), 하나 이상의 염기쌍의 삽입, 하나 이상의 염기쌍의 결실 및 상기 언급된 서열 변경의 임의의 조합을 들 수 있다. 변경은 또한 암호화된 아미노산이 변경되

도록 코딩 서열의 일부인 염기쌍의 전환을 포함할 수 있다.

- [0036] 도너 폴리뉴클레오타이드는 DNA 또는 RNA일 수 있고, 선형 또는 원형일 수 있으며, 단일 가닥 또는 이중 가닥일 수 있다. 그것은 네이키드 핵산으로서, 하나 이상의 전달 작용제를 가진 복합체 (예컨대, 리포솜, 나노입자, 폴록사머)로서, 또는 바이러스 전달 비히클, 예컨대, 예를 들어, 아데노바이러스, 렌티바이러스 또는 아데노-관련 바이러스 (AAV)에 함유되어 세포에 전달될 수 있다. 도너 서열은 10 내지 1,000개 뉴클레오타이드 (또는 그 사이의 임의의 정수 값의 뉴클레오타이드)의 범위의 길이 또는 더 긴 길이일 수 있다. 일부 구체예에서, 도너는 표적화된 절단 부위를 가진 상동성 영역이 옆에 있는 전장 유전자를 포함한다. 일부 구체예에서, 도너는 상동성 영역이 결핍되고 상동성 무관한 메커니즘 (즉 NHEJ)을 통해 표적 유전자좌에 통합된다. 다른 구체예에서, 도너는 세포에서 사용하기 위한 (즉 유전자 교정을 위해) 상동성 영역이 옆에 있는 핵산의 더 작은 조각을 포함한다. 일부 구체예에서, 도너는 shRNA, RNAi, miRNA 등과 같은 기능적 또는 구조적 구성요소를 암호화하는 유전자를 포함한다. 또 다른 구체예에서, 도너는 관심 있는 유전자에 결합하는 및/또는 그것의 발현을 조절하는 조절 요소를 암호화하는 서열을 포함한다. 또 다른 구체예에서, 도너는 관심 있는 유전자에 결합하는 및/또는 그것의 발현을 조절하는 관심 있는 조절 단백질 (예컨대, ZFP TF, TALE TF 또는 CRISPR/Cas TF)이다.
- [0037] 본원에 기술된 특정 방법 및 조성물에서, 뉴클레아제 및 도너는 하나 이상의 mRNA 및/또는 AAV 벡터를 사용하여 전달된다. 임의의 용량의 mRNA (ng) 또는 AAV 벡터 (vg/용량)가 사용될 수 있다. mRNA가 뉴클레아제(들) 및/또는 선택적 도너를 전달하는 구체예에서, mRNA의 투여량은 전형적으로 세포 또는 대상체당 10 내지 5000 ng 범위이다 (예컨대, 2000 ng, 62.5 ng, 31.3 ng, 15.6 ng). AAV 벡터가 뉴클레아제 및/또는 선택적 도너를 운반하기 위해 사용되는 구체예에서, 투여량은 전형적으로 각 뉴클레아제 (예컨대, 좌측 및 우측 ZFN)에 대해 대상체 또는 세포당 1.00E+9 내지 1.00E+13 vg의 범위이고 선택적 도너는 1.00E+10 내지 1.00E+13으로 제공된다. 특정 구체예에서, 한 쌍의 각각의 뉴클레아제는 별도의 AAV 벡터로 운반되고 세포 또는 대상체당 2.00E+10, 6.00E+10 또는 2.00E+11 vg으로 제공되며 도너는 또 다른 AAV 벡터로 운반되고 세포 또는 대상체당 1.60E+11, 4.8E+11 또는 1.6E+12 vg으로 제공된다.
- [0038] 상기 언급한 방법 중 임의의 방법에 대해, 세포 크로마틴은 염색체, 예피솜 또는 세포기관 계층에 있을 수 있다. 세포 크로마틴은, 한정하는 것은 아니지만, 원핵 및 진핵 세포, 진균 세포, 식물 세포, 동물 세포, 포유류 세포, 영양류 세포 및 인간 세포를 포함한, 모든 유형의 세포에 존재할 수 있다.
- [0039] 한 측면으로, 본원에는 제1 및 제2 (또한 좌측 및 우측 또는 ZFN 파트너로도 언급됨) ZFN 및/또는 좌측 및 우측 ZFN 중 하나 또는 둘 다를 암호화하는 하나 이상의 폴리뉴클레오타이드를 포함하는 아연 핑거 뉴클레아제가 기술되며, 여기서 제1 ZFN은 71557로 표시된 ZFN (SBS 42875에 대해 표 1에서 제시된 인식 나선 영역을 가지며 표 3 및 표 4에서 제시된 추가적인 특징 (예컨대, FokI 서열 및 ZFP 골격의 돌연변이, 5' UTR 서열, 등)을 가지는 ZFP를 포함함)을 포함하고 제2 ZFN은 표시된 71728 (SBS 47874에 대해 표 1에서 제시된 인식 나선 영역을 가지며 표 3 및 표 5에서 제시된 다른 특징 (예컨대, FokI 서열 및 ZFP 골격의 돌연변이, 5' UTR 서열, 등)을 가지는 ZFP를 포함함)을 포함한다. 특정 구체예에서, 제1 및 제2 (좌측 및 우측) ZFN은 별도의 폴리뉴클레오타이드에 의해 암호화되며, 별도의 폴리뉴클레오타이드는 동일하거나 상이한 유형의 것일 수 있다 (예컨대, 한 AAV가 좌측 ZFN을 암호화하는 서열을 포함하고 한 AAV가 우측 ZFN을 암호화하는 서열을 포함하는 2개의 AAV 벡터, 한 mRNA가 좌측 ZFN을 암호화하고 다른 mRNA가 우측 ZFN을 암호화하는 2개의 mRNA, 자기 절단 펩타이드 서열 (예를 들어 2A)에 의해 함께 연결된 두 ZFN을 포함하는 1개의 AAV 및 다른 ZFN을 암호화하는 서열을 포함하는 한 개의 AAV와 함께 사용된 한 개의 ZFN을 암호화하는 1개의 mRNA, 등). 특정 구체예에서, 벡터는 본원에서 제시된 것과 같이 "71557 AAV" 또는 "SB71557 AAV" (SEQ ID NO:43) 및 "71728 AAV" 또는 "SB71728 AAV" (SEQ ID NO:56)로 표시된 완전한 AAV 서열을 포함하여, 표 4 및/또는 표 5에 제시된 요소 (서열)를 포함하는 AAV 벡터이다. 다른 구체예에서, 표 4 및 5에 제시된 하나 이상의 요소는 임의의 유사한 서열로 대체되는데, 예를 들어 이들 표의 WPRE 서열은 기술분야에 공지된 또는 본원의 실시예 4에서 제시된 WPRE 서열로 대체될 수 있다 (예컨대, SEQ ID NO:53 대신 SEQ ID NO:68 또는 SEQ ID NO:69 또는 다른 WPRE 서열). 일부 구체예에서, 많은 아미노산 변형이 ZFN에 대해 만들어질 수 있다. 일부 구체예에서, 3, 6, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20개 또는 그 이상의 아미노산 치환이 만들어진다. 일부 구체예에서, 우측 및 좌측 측면에서 핑거 중 6개에 대한 13개의 조합된 아미노산 치환이 만들어진다.
- [0040] 그러므로, 본원에는 제1 및 제2 ZFN을 포함하는 아연 핑거 뉴클레아제가 기술되며, 제1 ZFN은 71557로 표시된 ZFN을 포함하고 제2 ZFN은 71728로 표시된 ZFN을 포함한다. 본원에 기술된 제1 및 제2 ZFN을 암호화하는 서열을 포함하는 하나 이상의 폴리뉴클레오타이드가 또한 제공된다. 특정 구체예에서, 단일 폴리뉴클레오타이드는 제1 및 제2 ZFN을 암호화하고, 다른 구체예에서, 2개의 별도의 폴리뉴클레오타이드는 제1 및 제2 ZFN을 암호화하는

서열을 포함한다. ZFN을 암호화하는 하나 이상의 폴리뉴클레오타이드는 동일하거나 상이한 AAV 벡터로 운반될 수 있다. 특정 구체예에서, 본원에 2개의 폴리뉴클레오타이드가 제공되며, 제1 폴리뉴클레오타이드 (예컨대, AAV 벡터)는 표 4에 제시된 서열 또는 SEQ ID NO:43에 제시된 서열을 포함하고 제2 폴리뉴클레오타이드 (예컨대, AAV 벡터)는 표 5에 제시된 서열 또는 SEQ ID NO:56에 제시된 서열을 포함한다. 일부 구체예에서, 좌측 ZFN (SB-71557)을 포함하는 AAV는 SB-A6P-ZL2로서 언급된다. 일부 구체예에서, 우측 ZFN (SB-71728)을 포함하는 AAV는 SB-A6P-ZR2로서 언급된다.

[0041] 또 다른 측면으로, 본원에는 본원에 기술된 하나 이상의 ZFN 및/또는 폴리뉴클레오타이드 (예컨대, AAV 벡터)를 포함하는 세포 (예컨대, 대상체의 줄기 세포, 전구 세포 또는 간 세포)가 기술된다. 한정하는 것은 아니지만, 줄기 세포, 전구 세포, 간 세포, 혈액 세포, 등을 포함한, 모든 세포 또는 세포주가 사용될 수 있다. 세포는 추가로 도너 폴리뉴클레오타이드, 전형적으로 치료 단백질 또는 그것의 단백을 암호화하는 도입유전자와 같은 외인성 서열을 암호화하는 폴리뉴클레오타이드를 포함할 수 있고, 외인성 서열은 내인성 알부민 유전자의 절단 후에 세포의 게놈으로 통합된다. 도너는 ZFN 파트너의 하나 또는 둘 다와 동일한 벡터로 운반되거나, 또는 대안으로, ZFN 파트너의 하나 또는 둘 다를 운반하는 벡터(들)와 동일하게 상이한 유형일 수 있는 별도의 벡터를 사용하여 투여될 수 있다. 특정 구체예에서, 세포는 3개의 별도의 AAV 벡터를 포함하며, 제1은 좌측 ZFN을 암호화하는 서열을 포함하고, 제2는 우측 ZFN을 암호화하는 서열을 포함하며, 제3은 도너 폴리뉴클레오타이드를 포함한다. ZFN (예컨대, 통합된 도너 폴리뉴클레오타이드)에 의해 만들어진 유전자 변형을 포함하는, ZFN 및 도너 폴리뉴클레오타이드를 포함하는 세포로부터 유래되는 딸 세포가 또한 기술된다. 이러한 유전자 변형은 딸 세포의 게놈 DNA의 차세대 서열분석을 포함한, 기술분야에 알려져 있는 표준 방법을 통해 확인될 수 있고 이러한 서열 결과가 ZFN 및 도너 폴리뉴클레오타이드로 처리되지 않은 야생형 세포와 비교된다.

[0042] 또 다른 측면으로, 본원에는 본원에 기술된 하나 이상의 ZFN, 하나 이상의 폴리뉴클레오타이드 (예컨대, AAV 벡터) 및/또는 하나 이상의 세포를 포함하는 제약학적 조성물이 기술된다. 특정 구체예에서, 제약학적 조성물은 3개의 별도의 AAV 벡터: ZFN 71557을 포함하는 제1 AAV (예컨대, "71557 AAV"); ZFN 71728을 포함하는 제2 AAV (예컨대, "71728 AAV"); 및 도너 폴리뉴클레오타이드를 포함하는 제3 AAV를 포함한다.

[0043] 또 다른 측면으로, 내인성 알부민 유전자를 절단하기 위하여 본원에 기술된 하나 이상의 of the ZFN, 하나 이상의 폴리뉴클레오타이드 (예컨대, AAV 벡터), 하나 이상의 세포 및/또는 하나 이상의 제약학적 조성물 (예컨대, 3개의 별도의 AAV 벡터를 포함하는 제약학적 조성물)을 사용하는 방법이 기술되며, 선택적으로 방법 (용도)은 분리된 세포에서 또는 대상체의 세포에서 절단된 알부민 유전자로 외인성 서열이 통합되도록 외인성 서열 (예컨대, AAV 벡터에 의해 운반됨)을 포함하는 도너 폴리뉴클레오타이드를 투여하는 단계를 추가로 포함한다. 일부 구체예에서, 본원에 기술된 이러한 하나 이상의 ZFN, 하나 이상의 폴리뉴클레오타이드, 하나 이상의 세포 및/또는 하나 이상의 제약학적 조성물은 인간 질환을 예방 또는 치료하기 위해 사용된다.

[0044] 본원에 기술된 하나 이상의 아연 핑거 뉴클레아제, 하나 이상의 폴리뉴클레오타이드, 하나 이상의 세포 및/또는 하나 이상의 제약학적 조성물 뿐만 아니라 그것들의 사용을 위한 선택적 설명서를 포함하는 키트가 또한 제공된다.

[0045] 또 다른 측면으로, 본원에는 (a) 선택적으로 표 4에 제시된 서열 또는 SB71557 AAV로 표시된 서열 (SEQ ID NO:43)을 포함하는, 71557로 표시된 제1 ZFN을 암호화하는 서열을 포함하는 제1 폴리뉴클레오타이드 (예컨대, AAV); (b) 선택적으로 표 5에 제시된 서열 또는 SB71728 AAV로 표시된 서열 (SEQ ID NO:56)을 포함하는, 71728로 표시된 제2 ZFN을 암호화하는 서열을 포함하는 제2 폴리뉴클레오타이드 (예컨대, AAV); 및 (c) 인자 IX (FIX) 단백질을 암호화하는 서열을 포함하는 도너 폴리뉴클레오타이드 (예컨대, AAV)를 포함하는 조성물 (또한 "FIX 조성물"로도 언급됨)이 기술된다. 특정 구체예에서, 도너는 표 6에 제시된 서열, 선택적으로 SEQ ID NO:59에 제시된 서열을 포함한다. 본원에 기술된 임의의 FIX 조성물에서, 제1, 제2 및 도너 폴리뉴클레오타이드는 3개의 별도의 AAV 벡터로 운반될 수 있다. 그것을 필요로 하는 대상체에서 FIX를 발현시키기 위해 본원에 기술된 조성물을 사용하는 방법이 또한 제공된다. 특정 구체예에서, 조성물은 ZFN (71557 및 71728)이 대상체에서 내인성 알부민 유전자를 절단하도록 대상체에게 투여되고, FIX 서열이 절단된 알부민 유전자에 통합되며 FIX 단백질이 대상체에서 발현된다. 본원에 기술된 방법 및 조성물은 그것을 필요로 하는 대상체에서 혈우병을 치료 및/또는 예방하기 위해 사용될 수 있다. 하나 이상의 FIX 조성물 및 선택적으로 그것의 사용을 위한 설명서를 포함하는 키트가 또한 제공된다.

[0046] 또한 추가의 측면으로, 본원에는 (a) 선택적으로 표 4에 제시된 서열 또는 SB 71557 AAV로 표시된 서열 (SEQ ID NO:43)을 포함하는, 71557로 표시된 제1 ZFN을 암호화하는 서열을 포함하는 제1 폴리뉴클레오타이드 (예컨대,

AAV); (b) 선택적으로 표 5에 제시된 서열 또는 SB 71728로 표시된 서열 (SEQ ID NO:56)을 포함하는, 71728로 표시된 제2 ZFN을 암호화하는 서열을 포함하는 제2 폴리뉴클레오타이드 (예컨대, AAV); 및 (c) 이두로네이트-2-설파타제 (IDS) 서열을 암호화하는 도너 폴리뉴클레오타이드 (예컨대, AAV)를 포함하는 조성물 (또한 "MPS II 조성물" 또는 "IDS 조성물"로도 언급됨)이 기술된다. 특정 구체예에서, 도너는 표 7에 제시된 서열, 선택적으로 SEQ ID NO:65에 제시된 서열을 포함한다. 본원에 기술된 임의의 MPS II 조성물에서, 제1, 제2 및 도너 폴리뉴클레오타이드는 3개의 별도의 AAV 벡터로 운반될 수 있다. 그것을 필요로 하는 대상체에서 IDS를 발현시키기 위해 본원에 기술된 조성물을 사용하는 방법이 또한 제공된다. 특정 구체예에서, 조성물은 ZFN (71557 및 71728)이 대상체에서 내인성 알부민 유전자를 절단하도록 대상체에게 투여되고, IDS 서열이 절단된 알부민 유전자에 통합되며 IDS 단백질이 대상체에서 발현된다. 본원에 기술된 방법 및 조성물은 그것을 필요로 하는 대상체에서 MPS II를 치료 및/또는 예방하기 위해 사용될 수 있다. 하나 이상의 MPS II 조성물 및 선택적으로 그것의 사용을 위한 설명서를 포함하는 키트가 또한 제공된다.

[0047] 또한 추가의 측면으로, 본원에는 (a) 선택적으로 표 4에 제시된 서열 또는 SB 71557 AAV로 표시된 서열 (SEQ ID NO:43)을 포함하는, 71557로 표시된 제1 ZFN을 암호화하는 서열을 포함하는 제1 폴리뉴클레오타이드 (예컨대, AAV); (b) 선택적으로 표 5에 제시된 서열 또는 SB SEQ ID NO:56으로 표시된 서열을 포함하는, 71728로 표시된 제2 ZFN을 암호화하는 서열을 포함하는 제2 폴리뉴클레오타이드 (예컨대, AAV); 및 (c) 알파-L 이두로니다제 (IDUA) 서열을 암호화하는 서열을 VHGM하는 도너 폴리뉴클레오타이드 (예컨대, AAV)를 포함하는 조성물 (또한 "MPS I 조성물" 또는 "IDUA 조성물"로도 언급됨)이 기술된다. 특정 구체예에서, 도너는 표 8에 제시된 서열, 선택적으로 SEQ ID NO:72에 제시된 서열을 포함한다. 본원에 기술된 임의의 MPS I 조성물에서, 제1, 제2 및 도너 폴리뉴클레오타이드는 3개의 별도의 AAV 벡터로 운반될 수 있다. 그것을 필요로 하는 대상체에서 IDUA를 발현시키기 위해 본원에 기술된 조성물을 사용하는 방법이 또한 제공된다. 특정 구체예에서, 조성물은 ZFN (71557 및 71728)이 대상체에서 내인성 알부민 유전자를 절단하도록 대상체에게 투여되고, IDUA 서열이 절단된 알부민 유전자에 통합되며 IDUA 단백질이 대상체에서 발현된다. 본원에 기술된 방법 및 조성물은 그것을 필요로 하는 대상체에서 MPS I을 치료 및/또는 예방하기 위해 사용될 수 있다. 하나 이상의 MPS I 조성물 및 선택적으로 그것의 사용을 위한 설명서를 포함하는 키트가 또한 제공된다.

[0048] 일부 구체예에서, 본원에서 개시된 임의의 조성물은 단일 용량으로 그것을 필요로 하는 대상체에게 투여된다. 다른 구체예에서, 조성물은 1 용량 이상으로 투여된다. 일부 구체예에서, 조성물은 용량 사이에 기간을 두고 1 용량 이상으로 투여된다. 일부 구체예에서, 기간은 1, 2, 3, 4, 5, 또는 6개월을 포함한다. 일부 구체예에서, 기간은 1년의 절반, 1년, 2년, 3년, 4년, 5년 또는 그 이상을 포함한다.

[0049] 또 다른 측면으로, 본원에 기술된 임의의 폴리펩타이드 (예컨대, 융합 분자) 및/또는 폴리뉴클레오타이드를 포함하는 세포가 또한 제공된다. 한 구체예에서, 세포는 각각 본원에 기술된 절단 도메인을 포함하는, 한 쌍의 융합 분자를 포함한다. 세포는 배양된 세포, 유기체의 세포 및 세포 및/또는 그것의 자손이 치료 후 유기체로 복귀될 경우에 치료를 위해 유기체로부터 제거되었던 세포를 포함한다. 세포 크로마틴에서 관심 있는 영역은, 예를 들어, 게놈 서열 또는 그것의 일부분일 수 있다.

[0050] 또 다른 측면으로, 본원에는 본원에 기술된 융합 단백질 또는 본원에 기술된 하나 이상의 아연 핑거 단백질, 절단 도메인 및/또는 융합 단백질을 암호화하는 폴리뉴클레오타이드; 보조 시약; 및 선택적으로 설명서 및 적합한 용기를 포함하는 키트가 기술된다. 키트는 또한 하나 이상의 뉴클레아제 또는 이러한 뉴클레아제를 암호화하는 폴리뉴클레오타이드를 포함할 수 있다.

[0051] 이들 및 다른 측면은 전체적으로 개시내용의 관점에서 숙련된 전문가에게 쉽게 드러날 것이다.

도면의 간단한 설명

[0052] **도 1A 및 1B**는 인간 알부민 게놈 서열 (SEQ ID NO:41)의 부분 서열을 도시하고 예시의 ZFN 47171-FLAG 및 47898-FLAG (표적 부위는 밑줄 또는 윗줄로 표시됨)의 결합 부위를 도시한다. 도 1A는 또한 대략 20%의 개체에서 발생하는 A/T로부터 G/C로의 단일 뉴클레오타이드 다형성 (SNP)을 나타낸다 (박스). 도 1B는 좌측 및 우측 ZFN이 표적 알부민 서열과 회합하는 방식을 보여주는 개략도이다.

도 2는 표시된 예시의 뉴클레아제 및 표시된 용량에 의한 게놈 변형 (% indels)을 보여주는 그래프이다. 각 조건 아래의 좌측 막대는 야생형 A/T 서열의 변형을 보여주고 우측 막대는 G/C SNP의 변형을 보여준다.

도 3A 내지 3H는 본원에 기술된 ZFN-암호화 벡터를 사용하는 게놈 변형의 예시 결과를 도시한다. 도 3A는 본원에 기술된 변형을 포함하지 않은 ("모(parent)") 뉴클레아제와 비교하여 본원에 기술된 예시의 인공 뉴클레아제

를 사용하여 표시된 조건 하에서 표적-외 부위 (SMCHD1)의 변형과 비교한 의도된 표적 (알부민)에서의 변형의 결과를 보여준다. 나타난 것과 같이, 본원에 기술된 뉴클레아제는 모 뉴클레아제와 비교하여 증가된 활성 및 특이성을 나타낸다. 도 3B는 의도된 표적 ("알부민 - 표적 상") (2000 ng, 62.5 ng, 31.3 ng, 15.6 ng)에 대한 또는 표적-외 부위 ("표적 외") (2000 ng)에 대한 표시된 투여량에서 표시된 모 또는 최적화된 ZFN을 사용하는 심층 서열분석에 의해 측정된 절단 활성 (% indels)을 보여준다. 도에서 나타난 조작된 FokI 도메인 N159D는 또한 N542D로도 언급되고 P95S로 표시된 FokI 도메인은 또한 P478S로도 언급된다. 도 3B는 또한 개선된 버전으로 만들어진 변형을 가진 ZFN 발현 카세트의 개략도를 보여준다. 도 3C는 본원에 기술된 ZFN으로 처리된 K562 및 HepG2 세포에서 표시된 표적-상 부위 (알부민) 및 표적-외 부위 (1 내지 26줄)에서의 활성 (% indels로 나타난 절단 활성 및 "포획 사건"으로 표시된 표적화된 통합)의 결과를 보여준다. "ns"는 유의미하지 않은 것을 나타내고; "ns*"는 ZFN 절단과 일치하지 않는 indels을 나타내며; "^"는 ZFN 절단과 일치하는 indels 및 유의미하지 않은 p-값을 나타내며; "ND"는 데이터 없음을 나타낸다. 도 3D는 모 47171/47898 ZFN 쌍 ("ZFN Std", 각 조건의 좌측 막대) 또는 최적화된 71557/71728 ZFN 쌍 ("ZFN 2.0", 각 조건의 우측 막대)을 사용하여 IDS 도입유전자를 운반하는 AAV 벡터 및 AAV 도너에 의해 운반된 알부민 ZFN의 표시된 투여량에서 절단 (% indels)을 보여주는 그래프이다 (낮음 = 30/240 ZFN/도너 MOI; 중간 = 100/800 ZFN/도너 MOI; 및 높음 = 300/2400 ZFN/도너 MOI). 이 실험에서 사용된 71557/71728 ZFN-표적화 AAV 구성물은 5' β-글로빈 미번역 영역 (UTR), 3xFLAG 및 우드척 간염 바이러스 (WHV) 전사 후 조절 요소 돌연변이 6 (WPREmut6)을 포함하였다. 도 3E(i)는 모 ("ZFN Std", 각 조건의 좌측 막대) 또는 최적화된 ZFN ("ZFN 2.0", 각 조건의 우측 막대)을 사용하여 IDS 도입유전자를 운반하는 AAV 벡터 및 AAV 도너에 의해 운반된 알부민 ZFN의 표시된 투여량에서 도너 도입유전자 (IDS)에 의해 암호화된 단백질의 활성을 보여주는 그래프이다 (낮음 = 30/240 ZFN/도너 MOI; 중간 = 100/800 ZFN/도너 MOI; 및 높음 = 300/2400 ZFN/도너 MOI). 도 3E(ii)는 표시된 투여량에서의 IDS 활성을 보여주는 그래프이다. 좌측으로부터 우측으로 각 조건 하에서 다음이 나타난다 (좌측에서 우측으로): 제5일째의 표준 ZFN; 제7일째의 표준 ZFN; 제5일째의 ZFN 2.0 및 제7일째의 ZFN 2.0. 도 3F(i) 및 3F(ii)는 47171/47898 쌍 및 71557/47898 쌍의 비교로부터의 결과를 도시한다. 표시된 양의 ZFN-표적화 mRNA가 WT (A:T) 및 SNP (G:C)-함유 ZFN 표적 부위에 대해 이형접합성인 일차 인간 간세포로 3개 한 벌로 트랜스펙션되었다. 이 실험에서 사용된 71557 ZFN-표적화 mRNA는 5' β-글로빈 미번역 영역 (UTR) 및 우드척 간염 바이러스 (WHV) 전사 후 조절 요소 (WPRE)를 포함하였다. ZFN 활성의 수준은 트랜스펙션 후 24시간 후에 심층 서열분석에 의해 결정되었다. 도 3F(i)은 A:T WT 대립유전자 (진한 회색) 및 G:C SNP 대립유전자 (연한 회색) 표적 부위에서의 ZFN 활성 (% indels로서 제시됨)을 나타낸다. 도 3F(ii)는 A:T WT 대립유전자 표적 부위 대비 G:C SNP 대립유전자 표적 부위에서의 ZFN 활성의 비율을 나타내며, 여기서 1.0의 값은 각 대립유전자에서의 동등한 절단을 나타낸다 (47171/47898 ZFN 쌍은 연한 회색으로, 71557/47898 ZFN 쌍은 진한 회색임). 도 3G는 47171/47898 및 71557/71728 ZFN 쌍의 일차 인간 간세포에서의 절단 동역학을 보여주는 그래프이며, 여기서 ZFN은 AAV에 의해 세포에 전달되었다. 이 실험에서 사용된 71557/71728 ZFN-표적화 AAV 구성물은 5' β-글로빈 미번역 영역 (UTR), 3xFLAG 및 우드척 간염 바이러스 (WHV) 전사 후 조절 요소 돌연변이 6 (WPREmut6)을 포함하였다. 도 3H는 일차 인간 간세포에서 47171/47898 및 71557/71728 ZFN 쌍에 대한 표적-상 및 표적 외 절단 데이터의 비교를 보여준다. 이 실험에서 사용된 71557/71728 ZFN-표적화 AAV 구성물은 5' β-글로빈 미번역 영역 (UTR), 3xFLAG 및 우드척 간염 바이러스 (WHV) 전사 후 조절 요소 돌연변이 6 (WPREmut6)을 포함하였다. 상부 줄은 3 K 내지 600 K의 MOI 농도에서 알부민 유전자좌 및 SMCHD1 유전자좌 상의 71557/71728 ZFN 쌍의 활성을 보여준다. 또한 상부의 좌측 그래프 상에는 예상된 임상 용량 범위가 표시된다. 제2 세대 ZFN을 암호화하는 AAV2/6으로 형질전환된 인간 일차 간세포가 MiSeq 심층 서열분석에 의해 평가되었다. NS - 양측 t-테스트에 의해 통계학적으로 유의미하지 않음, * - 양측 t-테스트에 의해 p-값 < 0.05. 하부 줄은 SMCHD1 유전자좌에 대한 별도의 실험과 비교된, 알부민 유전자좌에 대한 제1 및 제2 세대 ZFN 쌍 둘 다에 대한 용량의 폭발을 보여준다. 100 K MOI 용량에서, 제1 세대 ZFN은 17% indels의 평균 표적-상 활성 및 0.11% indels의 표적 외 활성을 보였고 제2 세대 ZFN은 35% indels의 평균 표적-상 활성 및 0.08% indels의 표적 외 활성을 보였다. 2개의 표적-상 대 표적 외 비율을 비교하면, 제2 세대 ZFN은 제1 세대 ZFN보다 약 2.8배 더 선택적이다. 600 K MOI 용량에서, 제1 세대 ZFN은 25% indels의 평균 표적-상 활성 및 0.36% indels의 표적 외 활성을 보였고; 제2 세대 ZFN은 44%의 평균 표적-상 활성 및 0.34%의 표적 외 활성을 보였다. 두 비율을 비교하면, 제2 세대 ZFN이 제1 세대 ZFN보다 약 1.9배 더 선택적이다. 100 K 및 600 K MOI에서, 47171/47898 및 71557/71728 ZFN 쌍에 대한 % indels은 각각 17% 및 35%, 및 25% 및 44%이고, 이것은 71557/71728 ZFN 쌍이 47171/47898 ZFN 쌍보다 약 2배 더 강력한 것을 시사한다.

도 4A 내지 4C는 뉴클레아제가 추가로 하나 이상의 FLAG 서열을 포함하는 폴리뉴클레오타이드로부터 발현될 때 증가된 뉴클레아제 활성 (% indels, 차세대 서열분석에 의해 측정됨)을 도시한다. 도 4A는 3xFLAG 서열이 있거

나 (y-축) 없이 (x-축) 인공 뉴클레아제를 암호화하는 폴리뉴클레오타이드의 도입 후의 활성 (% indels)을 보이는 그래프이다. 대각선의 좌측에 있는 데이터 점들은, 3xFLAG 서열이 해로운 것을 나타내는 대각선의 우측에 있는 데이터 점들과 비교하여 3x FLAG가 뉴클레아제 활성화에 대해 유익한 것을 나타낸다. 도 4B는 FLAG 서열이 없는 폴리뉴클레오타이드와 비교하여 평균 4배 이상 증가된 활성을 보여준다. 도 4C는 FLAG 펩타이드가 있거나 ("+" 펩타이드"로 나타낸 연한 음영 원형) 없이 ("No 펩타이드"로 나타낸 진한 음영 원형) 표시된 mRNA 양에서 알부민-표적화 ZFN을 암호화하는 mRNA로 트랜스펙션된 K562 세포에서의 뉴클레아제 활성 (% indels, 심층 서열 분석에 의해 측정됨)을 보여주는 그래프이다. 세포는 트랜스펙션 후 24시간 후에 ZFN 활성화에 대해 평가되었다. 연한 음영 원형 (5' 펩타이드를 포함하는 mRNA)위의 숫자는 5' 펩타이드가 없는 mRNA에 비교한 배수 증가를 나타낸다.

도 5A 내지 5C는 미변형 3' UTR과 비교하여 뉴클레아제가 변형된 3' UTR (예컨대, WPRE 서열)을 추가로 포함하는 폴리뉴클레오타이드로부터 발현될 때 증가된 뉴클레아제 활성 (%indels, 차세대 서열분석에 의해 측정됨)을 도시하는 그래프이다. 도 5A는 인공 뉴클레아제를 암호화하는 폴리뉴클레오타이드 (mRNA)가 3' UTR에 WPRE를 포함했을 때 분리된 세포에서의 증가된 절단 활성을 보여준다. 도 5B는 표시된 mRNA의 (LNP를 사용한) 마우스에의 투여 후에 생체내에서 (마우스 간에서) WPRE-포함 ZFN을 사용하여 증가된 활성을 보여주었다. 도 5B는 표시된 AAV의 마우스에의 투여 후에 생체내에서 (마우스 간에서) WPRE-포함 ZFN을 사용하여 증가된 절단 활성을 보여주었다.

도 6은 표시된 MOI에서 폴리뉴클레오타이드 구성요소의 표시된 조합을 사용하여 증가된 뉴클레아제 활성 (% indels, 차세대 서열분석에 의해 측정됨)을 보여주는 그래프이다. 각 조건 아래에서 가장 좌측의 막대 ("표준")는 뉴클레아제-표적화 폴리뉴클레오타이드가 3x FLAG, WPRE 서열 또는 폴리 A 서열을 포함하지 않았을 경우의 결과를 보여준다. 좌측으로부터 두 번째 막대 ("3xFlag, WPRE")는 뉴클레아제-표적화 폴리뉴클레오타이드가 3x FLAG 펩타이드 서열 및 WPRE 서열을 포함했을 경우의 결과를 보여준다. 우측으로부터 두 번째 막대 ("5'XBG, WPRE")는 뉴클레아제-표적화 폴리뉴클레오타이드가 소 성장 호르몬 ("BG") 폴리 A 서열 및 WPRE 서열을 포함했을 경우의 결과를 보여준다. 가장 우측의 막대 ("5XBG, 3xFLAG, WPRE")는 뉴클레아제-표적화 폴리뉴클레오타이드가 BG 폴리 A 서열, 3x FLAG 펩타이드 서열 및 WPRE 서열을 포함했을 경우의 결과를 보여준다. 각 조건 (MOI) 위에는 관찰된 뉴클레아제 활성의 배수 변화가 도시된다.

도 7A 및 7B는 만들어지고 테스트된 예시의 상이한 변이체 구성물을 예시한다. 도 7A는 V1이 초기 발현 구조이고 V2-V8이 다양한 변이체 구조인 것을 도시한다. 약어는 다음과 같다: "ApoE"는 Apo E 인핸서이고; "hAAT"는 인간- $\alpha 1$ 항-트립신 프로모터이며; "HBB-IGG"는 인간 β -글로빈 유전자의 제1 인트론으로부터의 5' 도너 부위 및 분지 및 번역글로불린 유전자 중쇄 가변 영역의 인트론으로부터의 3' 엑셉터를 포함하는 인간 베타 키메라 인트론이고; "NLS"는 핵 국지화 서열이며; "폴리A"는 폴리 A 서열이고; "WPRE"는 우드척 간염 바이러스 전사 후 조절자 요소이며; "3xFLAG"는 SEQ ID NO:4 및/또는 SEQ ID NO: 71로서 기술된 펩타이드이고; 그리고 " β -g1b"는 제노푸스 베타-글로빈 유전자의 5' 미변역 영역이다. 도 7B는 ZFN 47171, 47898, 71557 및 71728을 암호화하는 AAV의 개략도를 보여준다.

도 8은 표시된 MOI에서 폴리뉴클레오타이드 구성요소의 표시된 조합을 사용하여 뉴클레아제 활성 (% indels, 차세대 서열분석에 의해 측정됨)을 보여주는 그래프이다. 각 조건 아래에서 가장 좌측의 막대 ("표준 (V1)")는 뉴클레아제-표적화 폴리뉴클레오타이드가 3x FLAG, 변이체 1 (V1) 또는 모(parent)로도 언급되는 WPRE 서열 또는 폴리 A 서열을 포함하지 않았을 경우의 결과를 보여준다. 좌측으로부터 두 번째 막대 ("WPRE (V2)")는 뉴클레아제-표적화 폴리뉴클레오타이드가 WPRE 서열을 포함했을 경우의 결과를 보여준다. 중간 막대 ("3xFlag, WPRE (V4)")는 3x FLAG 서열 및 WPRE 서열을 포함한 변이체 (V4로 표시됨)의 결과를 보여준다. 우측으로부터 두 번째 막대 ("5'XBG, WPRE (V6)")는 뉴클레아제-표적화 폴리뉴클레오타이드 (변이체 6 또는 V6으로 표시됨)가 소 성장 호르몬 ("BG") 폴리 A 서열 및 WPRE 서열을 포함했을 경우의 결과를 보여준다. 가장 우측의 막대 ("5XBG, 3xFLAG, WPRE")는 뉴클레아제-표적화 폴리뉴클레오타이드 (변이체 8 또는 V8로 표시됨)가 BG 폴리 A 서열, 3x FLAG 펩타이드 서열 및 WPRE 서열을 포함했을 경우의 결과를 보여준다.

도 9는 표시된 mRNA 양에서 5'UTR 서열 (SEQ ID NO:1에 제시된 것과 같은 제노푸스 β -글로빈 UTR)이 있거나 ("+" Novel 5' UTR"로 나타낸 연한 음영 원형) 없이 ("No 5'UTR"로서 나타낸 진한 음영 원형) 알부민-표적화 ZFN을 암호화하는 mRNA로 트랜스펙션된 K562 세포에서의 뉴클레아제 활성 (% indels, 심층 서열분석에 의해 측정됨)을 보여주는 그래프이다. 세포는 트랜스펙션 후 24시간 후에 ZFN 활성화에 대해 평가되었다. 연한 음영 원형 (5' UTR을 포함하는 mRNA) 위의 숫자는 5' UTR이 없는 mRNA와 비교하여 배수 증가를 나타낸다.

도 10A 내지 10C는 생체내 또는 시험관내 절단 효율 및 ZFN 발현을 보여주는 그래프이다. 도 10A는 표시된 ZFN 구성물의 주입 후 56일 후에 마우스 간 세포에서 % indels에 의해 측정된 생체내 절단을 보여주는 그래프이고 도 10B는 ZFN 발현 수준을 나타낸다. 야생형 수컷 마우스가 5'UTR, FLAG 펩타이드 및 WPRE 서열 ("ZFN 개선된" 또는 "ZFP 2.0") 및 IDS 도너가 있거나 없이 ("ZFN" 표준) 알부민 ZFN-표적화 AAV6 구성물로 3가지 용량: 2.0E+11 vg/마우스 (저용량), 6.0E+11 vg/마우스 (중간 용량) 및 또는 2.0E+12 vg/마우스 (고 마우스)으로 정맥내 주사되었다. 좌측으로부터 우측으로, 제제 완충액, 저용량 미변형 (표준) ZFN-표적화 벡터, 본원에 기술된 것과 같이 변형된 저용량 AAV ZFN-표적화 벡터, 미변형 (표준) 중간 용량, 본원에 기술된 것과 같이 변형된 중간 용량 AAV ZFN-표적화 벡터 ZFN, 미변형 (표준) 고용량, 및 본원에 기술된 것과 같이 변형된 고용량 AAV ZFN-표적화 벡터 ZFN의 투여 후 indels이 도시된다. 도시된 것과 같이, 본원에 기술된 변형 (5'UTR, 5' 펩타이드, WPRE)을 포함하는 AAV 벡터가 미변형 AAV 벡터와 비교하여 절단 효율에서 7배 증가를 제공하였다. * - p<0.05, ** - p<0.01, 양측 학생 t-테스트에 의함. 도 10C는 71557/71728 또는 47171/47898 ZFN 쌍을 사용하는 FIX 도너의 증가된 발현을 나타낸다. ZFN 쌍은 제1일에 HepG2 세포를 처리하기 위해 사용되었고, 그런 후 제2일에는 FIX 도입유전자가 사용되었다. 제9일에, 배지에 ELISA가 수행되어 발현된 FIX 단백질의 양이 측정되었다. 데이터는 71557/71728 ZFN 쌍의 사용이 47171/47898 쌍과 비교하여 배지에서 거의 3배 더 많은 FIX 발현을 초래한 것을 입증한다.

도 11A 내지 11C는 이들 연구에서 사용된 도너 설계를 도시한다. 도 11A는 도너 AAV에 의해 포함된 요소 및 도입유전자를 보여주는 3개의 도너를 도시하는 그래프이다. 도 11B는 일차 인간 간세포에 대한 편집을 위한 변형의 결과 (좌측 그래프) 및 개선된 ZFN (표시된 용량 아래에 우측 막대로 나타낸 "ZFN 2.0")과 비교하여 표준 (표시된 용량 아래에 좌측 막대로 나타낸 "현재") ZFN 쌍을 사용하여 ZFN-구동된 표적화된 통합이 수행된 간세포의 상층액에서 검출된 증가된 활성을 보여준다. 도 11C는 도 10 및 실시예 7에서 기술된 것과 같이 처리된 마우스 대상체의 생체내 도입유전자 발현 (IDS)을 보여주는 그래프이다. 야생형 수컷 마우스가 5'UTR, FLAG 펩타이드 및 WPRE 서열 ("ZFN 개선된" 또는 "ZFP 2.0") 및 IDS 도너가 있거나 없이 ("ZFN" 표준) 알부민 ZFN-표적화 AAV6 구성물로 3가지 용량: 2.0E+11 vg/마우스 (저용량), 6.0E+11 vg/마우스 (중간 용량) 및 또는 2.0E+12 vg/마우스 (고 마우스)으로 정맥내 주사되었다. 간 샘플이 주사 후 56일 후에 수집되었고 상대적인 도입유전자 발현이 실시예에서 기술되는 것과 같이 측정되었다. 좌측으로부터 우측으로, 제제 완충액, 저용량 미변형 (표준) ZFN-표적화 벡터, 본원에 기술된 것과 같이 변형된 저용량 AAV ZFN-표적화 벡터, 미변형 (표준) 중간 용량, 본원에 기술된 것과 같이 변형된 중간 용량 AAV ZFN-표적화 벡터, 미변형 (표준) 고용량, 및 본원에 기술된 것과 같이 변형된 고용량 AAV ZFN-표적화 벡터의 투여 후의 결과가 도시된다. 도시된 것과 같이, 본원에 기술된 변형 (5'UTR, 5' 펩타이드, WPRE)을 포함하는 AAV 벡터가 미변형 AAV 벡터와 비교하여 도너 (IDS) 발현에서 18배 증가를 제공하였다.

도 12는 실시예 7 및 도 10 및 11에서 기술된 것과 같이 처리된 대상체의 간 샘플에서 IDS 발현의 웨스턴 블롯 분석의 결과를 도시한다.

도 13은 표적화된 통합 (실시예 7 참조)을 위해 변형된 및 미변형 ZFN (저용량, 중간 용량 및 고용량에서)으로 투여된 도너에 의해 암호화된 IDS 단백질의 효소적 활성을 보여주는 그래프이다. 효소적 활성은 실시예에서 기술된 것과 같이 측정되었다. 좌측으로부터 우측으로, 제제 완충액, 저용량 미변형 (표준) ZFN-표적화 벡터, 본원에 기술된 것과 같이 변형된 저용량 AAV ZFN-표적화 벡터, 미변형 (표준) 중간 용량, 본원에 기술된 것과 같이 변형된 중간 용량 AAV ZFN-표적화 벡터 ZFN, 미변형 (표준) 고용량, 및 본원에 기술된 것과 같이 변형된 고용량 AAV ZFN-표적화 벡터 ZFN의 투여 후의 결과가 도시된다.

도 14A 및 14B는 표준 47171/47898 ZFN 쌍 또는 the 71557/71728 ZFN 쌍을 사용한 HepG2 세포에 IDUA 도너의 삽입을 도시한다. 도 14A는 ZFN 및 도너로 처리된 HepG2 세포의 상층액에서 시간 경과에 따른 IDUA 활성을 나타낸다. ZFN 용량은 600 K의 MOI였고 도너는 1200 K MOI의 용량이었다. 도 14B는 세포에서 각각의 테스트 조건에 대한 퍼센트 indels을 나타낸다. 데이터는 ZFN의 두 쌍이 활성이고 IDUA 도입유전자의 ZFN-지시된 표적화된 통합을 유발하는 것을 입증한다.

발명을 실시하기 위한 구체적인 내용

[0053] 본원에는 조각된 뉴클레아제의 발현 효율 (절단 활성)을 증가시키기 위한, 뿐만 아니라 표적-상 조각된 뉴클레아제 절단 활성의 특이성을 증가시키기 위한 방법 및 조성물이 개시된다. 방법은 뉴클레아제 발현 벡터의 발현 요소의 조합을 최적화하는 단계 및 *FokI* 절단 도메인과 DNA 사이의 비특이적 상호작용을 감소시키고 아연 핑거 골격과 DNA 사이의 비특이적 상호작용을 감소시키는 단계를 포함한다. 추가로, 발명의 방법 및 조성물은 높은

특이성으로 인간 알부민 유전자좌를 절단할 수 있는 최적화된 ZFN 시약을 제공하고, 최적화된 알부민 시약은 또한 야생형 알부민 표적 서열 및 SNP를 포함하는 동일한 표적 서열을 절단할 수 있다. 본원에 기술된 ZFN 시약은, 단백질(들)이 대상체에서 발현되고 대상체의 질환 또는 장애를 감소, 예방, 및/또는 치료 (예컨대, 증상을 완화)하도록 하나 이상의 치료용 단백질-암호화 서열 (예컨대, 질환 또는 장애를 가진 대상체에서 없거나 결핍된 단백질)을 절단된 알부민 유전자에 뉴클레아제-매개 통합하기 위한 것을 포함하여, 알부민 유전자의 효율적이고 고도로 표적화된 절단을 위해 사용될 수 있다.

[0054] 혈우병 B는 혈액 응고 인자 IX (FIX)를 암호화하는 유전자의 돌연변이에 의해 유발된 X-연계 열성(recessive) 출혈 장애이다. 또한 그것은 크리스마스 질환으로서 알려져 있고 혈우병 A 또는 인자 VIII 결핍증 다음으로, 두 번째로 흔한 형태의 혈우병이다. 그것은 미국에서 약 4,000명의 유병률로 약 25,000명의 남성 중 한 명에서 발생한다. 질환 징후는 인자 IX 응고 활성의 수준에 따라 달라진다. 혈우병 B를 가진 대부분의 대상체는 중증 형태의 질환을 나타낸다 (<1% FIX 활성). 그들은 보통 자발적인 관절 또는 심부 근육 출혈이 발생한 후 생후 첫 2년 동안에 진단된다. 중간 정도의 질환 (1-5% FIX 활성)은 상대적으로 미미한 외상 후에 연장된 또는 지연된 출혈로 존재하며 6세 이전에 진단된다. 비교하자면, 경미한 혈우병 (>5-30% FIX 활성)은 생애 후기에 진단되고 자발적인 출혈로 고생하지는 않지만 수술 또는 발치 후에 과도한 출혈이 발생할 것이다. 마지막으로, 대략 10%의 여성 보균자가 30% 아래의 FIX 활성을 가지며 주요 외상 또는 수술 후에 과도한 출혈의 위험이 있다.

[0055] 혈우병 B의 현재 치료는 1960년대 후반에 도너 혈장으로부터 초기에 유래된 FIX 농축물의 사용으로 이루어진다. 바이러스 비활성화 및 도너 스크리닝과 같은 후속 개선은 보다 정제된 농축물로 이어졌고 1997년에 재조합 FIX의 도입으로 절정을 이루었다. 보다 최근에, 매주 또는 격주 투여를 허용하는 재조합 FIX Fc 융합 단백질이 미국에서 판매가 승인되었다. 정상 약 5% (즉 약 250 ng/mL)까지 인자 IX의 수준을 증가시키는 것은 증상의 현저한 개선을 초래하고 자발적이고 생명을 위협하는 출혈 에피소드를 방지하기에 충분하다 (Scriver, CR *et al.* The Metabolic and Molecular Bases of Inherited Disease. New York: McGraw-Hill (2001); Lofqvist, T *et al.* (1997) *J. Intern. Med.* 241(5): 395-400; Ljung, RC. (1998) *Br. J. Haematol.* 101(2): 215-219). 이들 치료적 진전은 혈장 유래 FIX의 도입 전 11년으로부터 재조합 단백질로 63년까지 중간 기대 수명을 증가시켰다 (Darby, SC *et al.* (2007) *Blood.* 110(3): 815-25).

[0056] 혈우병 B에 대한 현재 치료는 정제된 재조합 인자 IX의 만성적이고, 반복된 정맥내 주입에 의존하며 많은 결점을 겪고 있다. 이것은 억제제 형성과 관련된 반복된 정맥내 주입에 대한 필요성을 포함하며, 치료적이기보다는 예방적이다. 제자리 (대상체의 간 내에서) 치료적 도입유전자로부터의 합성을 토대로 한, 응고 인자 전달에 대한 대안 접근법은 이들 문제를 제거할 전망을 제거한다. 본원에 개시된 방법 및 조성물은 교정적인 FIX 도입유전자를 게놈 안에, 및 대상체 자체의 내인성 알부민 유전자좌 또는 고도로 발현된 외인성 프로모터의 제어 하에 놓음으로써, 인자 IX의 간-특이적 합성을 초래하는 신규한 전략을 통해 혈우병 B를 치료하는 것을 기술한다. 특히, 본원에 기술된 조작된 아연 핑거 뉴클레아제 (ZFN)는 FIX 도입유전자를 대상체 자신의 간세포의 게놈에 생체내에서 부위-특이적으로 통합시키기 위해 사용된다. 본원에 기술된 뉴클레아제를 사용한 FIX 도입유전자의 통합은 안정적이고, 높은 수준의, 간-특이적 발현 및 FIX의 대상체 혈액으로의 분비를 초래한다.

[0057] 헐러/헐러-샤이에/샤이에(Hurler/Hurler-Scheie/Scheie) 증후군으로도 언급되는 뮤코다당질축적증(Mucopolysaccharidosis) 유형 I (MPS I)은 열성 리소좀성 축적 장애이다. MPS I에 대한 미국 국립신경질환뇌졸중 연구소 (NINDS) 자료표에 따르면, 추정된 발병률은 중증 MPS I의 경우 약 100,000건의 출생 당 1명이고, 약화된 MPS I의 경우 약 500,000건의 출생 당 1명이며, 중증과 약화된 샤이에 속하는 질환의 경우 약 115,000건의 출생 당 1명이다.

[0058] MPS I은 글리코사미노글리칸 (황산화된 탄수화물 중합체; GAG)을 분해하는 이두로니다제 (IDUA) 효소를 암호화하는 유전자의 돌연변이와 관련된다. IDUA 유전자의 돌연변이는 IDUA 효소 활성을 감소시키거나 제거하고, 이것은 소변, 혈장, 및 체액에서 독성 GAG의 축적을 초래한다.

[0059] 특정 유형의 IDUA 돌연변이 (100개 이상의 상이한 돌연변이가 기술되어 있음) 및 결과적으로 생성되는 잔류하는 IDUA 효소의 수준에 따라, 환자는 헐러 증후군 (MPS I H) 또는 약화된 변이체 (MPS I H/S 및 MPS I S)를 발생시킬 것이다. 모든 MPS I 환자의 50% 내지 80%가 중증 형태를 가진 채로 존재하는 것으로 추정되었는데, 이것은 상대적으로 용이한 진단에 부분적으로 기인하는 것일 수 있다 (Muenzer *et al.* (2009) *Pediatrics.* 123(1): 19-29). MPS I H 환자는 그들이 만 1살이 되기 전에 발달 지체의 증상을 보일뿐만 아니라 2세에서 4세 사이에 성장 정지 및 점진적인 정신적 쇠퇴를 보인다. 다른 증상들로는 장기 비대, 각막 혼탁, 관절 경직 및 골격 기형 (비정상적인 척수 뼈 포함), 혀가 커지는 거친 얼굴 특징, 청력 상실 및 탈장을 들 수 있다. 이들 MPS I H 환자

의 기대 수명은 10년 미만이다. 약화된 형태를 가진 환자는 이들 임상적 표시의 대부분을 공유하지만 중증 증상은 덜하다. 더불어, CNS는 포함되지 않고 따라서 그들은 정신 지체를 겪지는 않는다.

[0060] 이들 환자 중 많은 수가 성인이 되어도 생존할 수 있지만 상당한 병적 상태를 나타낸다. MPS I에 대한 현재 치료법은 조혈 줄기 세포 이식 (HSCT) 및 효소 대체 요법 (ERT)을 포함한다. 만약 중증 MPS I 형태 ((MPS I-H)를 앓고 있는 환자가 조기에 (2.5세 미만) 진단될 수 있다면, HSCT (골수 또는 제대 줄기 세포)에 의한 치료적 개입이 신경인지를 포함한 대부분의 임상적 특징을 방지 또는 반전시킬 수 있다. 현재, MPS I H를 가진 거의 모든 환자가 HSCT를 진행한다. MPS I의 경우 HSCT 후의 사망률은 15%이고 성공적인 생착(successful engraftment)으로의 생존율은 56% ERT이며, 2003년 이후 차이나이즈 햄스터 난소 세포에서 생성된 다형성 재조합 단백질, Aldurazyme[®]이 사용되고 있다. 이 효소는 폐 기능, 간비장비대(hepatosplenomegaly), 및 운동 능력을 개선시키고 개선된 건강 관련 삶의 질로 이어지는 것으로 나타났다. ERT는 가능한 초기에 시작되어야 한다. 효소 대체 요법의 한계로는 평생 치료에 대한 필요성, 중화 항체 개발, 혈액 뇌 장벽을 넘지 못함, 지속적인 심장, 정형 외과, 안구 합병증 및 매주 정맥내 주입의 불편함을 들 수 있다. 이들 한계는 함께MPS I에 대한 광범위한 치료 요법을 개발할 급박한 필요성을 강조한다.

[0061] 본원에서 개시된 방법 및 조성물의 사용에 대한 목적 및 근거는 생체내 게놈 편집에 의한 효소 대체 요법에 대한 필요성을 무효화하는 것이다. 특히, 본원에 기술된 조각된 아연 핑거 뉴클레아제 (ZFN)는 이두로니다제 효소 (hIDUA) 도입유전자의 교정 복사물을 대상체 자신의 간세포 게놈에 생체내에서 부위-특이적으로 통합시키기 위해 사용된다. hIDUA 도입유전자의 통합은 알부민 유전자좌의 인트론 1로 표적화될 수 있어서, 안정적인, 고수준의, 간-특이적 발현 및 이두로니다제의 혈액으로의 분비를 초래할 수 있다. 고도로 발현된 내인성 알부민 유전자좌의 제어 하에 huIDUA 도입유전자의 배치는 MPS I 환자의 평생 동안 이두로니다제의 영구적인, 간-특이적 발현을 제공할 것으로 예상된다.

[0062] X-연계된 열성, 리소좀성 축적 장애로도 언급되는 뮤코다당질축적증 II (MPS II)는 남성에서 우세하다. MPS II의 발생률은 100,000명의 생존 출생당 0.3 내지 0.71인 것으로 기록된다 (Burton & Giugliani (2012) *Eur J Pediatr.* 171(4):631-9. doi: 10.1007/s00431-012-1703-y. Epub 2012 Mar 1.). 연간 발생률에 질환의 약화된 형태에 대한 21.7년의 보다 보존된 중간 기대 수명 (질환의 중증 형태에 대한 중간 기대 수명은 11.8년임) (Burrow & Leslie (2008) *Biologics.* 2008 Jun;2(2):311-20; Young & Harper (1982) *J Med Genet.* 19(6):408-11)을 적용하면 미국에서 현재 살고 있는 MPS II를 가진 약 629명의 개인이라는 추정된 유병률이 계산된다.

[0063] MPS II는 뮤코다당 글리코사미노글리칸 (GAGA)의 리소좀성 분해에 관여하는 효소를 암호화하는 이두로네이트-2-실과타제 (IDS) 유전자의 돌연변이에 의해 유발된다. 이것은 소변, 혈장 및 조직에서 GAG의 축적을 초래하고 다중-전신성, 진행성 질환을 유발한다. 헌터 증후군(Hunter syndrome)은 신체적 및 인지적으로 포함된 초기 발병, 중증의 질환 (환자의 2/3)에서 신체적 질환의 후기 발병 및 중추신경계 (CNS) 질환이 거의 없거나 없는 것을 특징으로 하는 약화된 MPS II에 걸쳐있는 질환 스펙트럼을 나타낸다. IDS 돌연변이의 특정 유형 (150개를 넘는 유전자 돌연변이가 확인되었음) 및 결과적으로 생성되는 잔류하는 IDS 효소의 수준이 질환의 중증도를 대부분 결정하는 것으로 보인다. 약화된 형태에서 잔류하는 IDS 활성은 야생형 IDS 활성 및 활성이 없는 중증 표현형을 가진 것들의 0.2 내지 2.4%로 측정되었다 (Sukegawa-Hayasaka *et al.* (2006) *J Inherit Metab Dis* 29(6): 755-61). IDS 유전자는 Xq28에 대해 지도화되고, 24 kb에 걸쳐 확산된 9개의 엑손을 함유한다. 주요 결실 및 재배열은 언제나 질환의 중증 형태와 관련된다.

[0064] 중증의 MPS II 환자는 전형적으로 생후 18개월에서 3세 사이에 지연된 언어 및 발달 지연을 나타내기 시작한다. 질환은 중증의 MPS II 환자에서 장기 비대, 과잉행동 및 공격성, 신경학적 악화, 관절 경직 및 골격 기형 (비정상적인 척수 뼈 포함), 혀가 확대되는 거친 얼굴 특징, 심장 판막 비후, 청력 상실 및 탈장을 특징으로 한다. 치료되지 않은 중증 헌터 증후군 환자의 기대 수명은 신경학적 악화 및/또는 심폐 기능 부전으로 인해 사망하는 10대 중반 까지이다. 약화된 형태를 가진 환자는 전형적으로 중증 환자보다는 나중에 진단된다. 신체적 임상 특징은 중증 환자와 유사하지만, 전체적인 질환 중증도는 더 가볍고, 일반적으로, 질환 진행이 더 느리며, 인지적 손상은 없거나 가벼운 정도이다. 치료되지 않은 약화된 형태에서 사망은 자주 심장 및 호흡기 질환으로 20 내지 30세 사이의 연령대에서 일어난다.

[0065] MPS II에 대한 유일한 승인된 치료법은 효소 대체 요법 (ERT)이다. 정맥내 (IV) ERT 재조합 IDS 단백질 (이두르실과제(idursulfase); Elaprased[®])이 2006년 이후 승인되었다. 이두르실과제를 사용하는 ERT는 간비장비대, 폐 기능 (FVC) 및 운동 능력 (6분 도보)을 개선시키고 개선된 건강 관련 삶의 질로 이어지는 것으로 나타났다. ERT에 대한 반응은 치료 시작 시점에서의 대상체의 질환 중증도에 좌우된다. ERT에 대한 한계는 평생 치료에 대한

필요성, 중화 항체의 개발, 효소가 혈액 뇌 장벽을 넘지 못함, 및 매주 정맥내 주입의 불편함을 포함한다. 혈러 증후군 (MPS I의 중증 형태)과 대조적으로, 조혈 줄기 세포 이식 (HSCT)은 MPS II의 중증 형태에 대해서는 권장되지 않는다. 이들 한계점은 함께, MPS II에 대한 광범위한 치료 요법을 개발할 급박한 필요성을 강조한다.

[0066] 그러므로, 본원에서 개시된 방법 및 조성물은 MPS II를 가진 대상체에서 생체내 게놈 편집에 의한 효소 대체 치료법에 대한 필요성을 모호화 또는 감소시킨다. 특히, 본원에 기술된 조작된 아연 핑거 뉴클레아제 (ZFN)는 효소 이두로네이트-2-설파타제 (hIDS) 도입유전자의 교정적 복사물을 대상체 자신의 간세포의 게놈에 생체내에서 부위-특이적으로 통합시키기 위해 사용된다. hIDS 도입유전자의 통합은 알부민 유전자좌의 인트론 1로 표적화되어, 안정적인, 고수준의, 간-특이적 발현 및 이두로네이트-2-설파타제의 혈액으로의 분비를 초래한다. 고도로 발현된 내인성 알부민 유전자좌의 제어 하에 hIDS 도입유전자의 배치는 MPS II 환자의 평생 동안 이두로네이트-2-설파타제의 영구적인, 간-특이적 발현을 제공할 것으로 예상된다.

[0067] **일반적 사항**

[0068] 방법의 실시, 뿐만 아니라 본원에 개시된 조성물의 제조 및 사용은, 다르게 표시되지 않는 한, 기술분야의 숙련된 기술 내에 있는 것과 같이 분자 생물학, 생화학, 크로마틴 구조 및 분석, 컴퓨터 화학, 세포 배양, 재조합 DNA 및 관련 분야의 종래 기법들을 사용한다. 이들 기법은 문헌에서 전체적으로 설명된다. 예를 들어, Sambrook *et al.* MOLECULAR CLONING: A LABORATORY MANUAL, Second edition, Cold Spring Harbor Laboratory Press, 1989 and Third edition, 2001; Ausubel *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, 1987 and periodic updates; the series METHODS IN ENZYMOLOGY, Academic Press, San Diego; Wolffe, CHROMATIN STRUCTURE AND FUNCTION, Third edition, Academic Press, San Diego, 1998; METHODS IN ENZYMOLOGY, Vol. 304, "Chromatin" (P.M. Wassarman and A. P. Wolffe, eds.), Academic Press, San Diego, 1999; 및 METHODS IN MOLECULAR BIOLOGY, Vol. 119, "Chromatin Protocols" (P.B. Becker, ed.) Humana Press, Totowa, 1999.

[0069] **정의**

[0070] 용어 "핵산", "폴리뉴클레오타이드", 및 "올리고뉴클레오타이드"는 상호교환적으로 사용되며 선형 또는 원형 형태의, 및 단일- 또는 이중-가닥 형태의 데옥시리보뉴클레오타이드 또는 리보뉴클레오타이드를 나타낸다. 본 개시의 목적에 대해, 이들 용어는 중합체의 길이와 관련하여 제한하는 것으로 해석되지 않아야 한다. 용어는 자연적인 뉴클레오타이드의 공지된 유사체, 뿐만 아니라 염기, 당 및/또는 포스페이트 모이어티 (예컨대, 포스포티오에이트 골격)에서 변형된 뉴클레오타이드를 포함할 수 있다. 일반적으로, 특정 뉴클레오타이드의 유사체는 동일한 염기쌍 형성 특이성을 가진다; 즉, A의 유사체는 T와 염기쌍을 형성할 것이다.

[0071] 용어 "폴리펩타이드", "펩타이드" 및 "단백질"은 아미노산 잔기의 중합체를 나타내기 위해 상호교환적으로 사용된다. 용어는 또한 하나 이상의 아미노산이 상응하는 자연적으로 발생하는 아미노산의 화학적 유사체이거나 변형된 유도체인 아미노산 중합체에도 적용된다.

[0072] "결합"은 거대분자 사이 (예컨대, 단백질과 핵산 사이)의 서열-특이적, 비공유 상호작용을 나타낸다. 결합 상호작용의 모든 구성요소는, 상호작용이 전체적으로 서열-특이적인 한, 서열-특이적일 필요는 없다 (예컨대, DNA 골격의 포스페이트 잔기와의 접촉). 이러한 상호작용은 일반적으로 10^{-6} M^{-1} 이하의 해리 상수 (K_d)에 의해 특성화된다. "친화성"은 결합의 강도를 나타낸다: 증가된 결합 친화성은 더 낮은 K_d 와 상관이 있다. "비특이적 결합"은 관심 있는 임의의 분자 (예컨대, 조작된 뉴클레아제)와 의존적인 표적-상 서열이 아닌 거대 분자 (예컨대, DNA) 사이에서 일어나는 비공유 상호작용을 나타낸다.

[0073] "결합 단백질"은 또 다른 분자에 비공유적으로 결합할 수 있는 단백질이다. 결합 단백질은, 예를 들어, DNA 분자 (DNA-결합 단백질), RNA 분자 (RNA-결합 단백질) 및/또는 단백질 분자 (단백질-결합 단백질)에 결합할 수 있다. 단백질-결합 단백질의 경우에, 그것은 자체에 결합할 수 있고 (호모다imer, 호모트라이머, 등을 형성함) 및/또는 그것은 상이한 단백질 또는 단백질들의 하나 이상의 분자에 결합할 수 있다. 결합 단백질은 한 가지 이상의 유형의 결합 활성을 가질 수 있다. 예를 들어, 아연 핑거 단백질은 DNA-결합, RNA-결합 및 단백질-결합 활성을 가진다. RNA-가이드된 뉴클레아제 시스템의 경우에, RNA 가이드는 뉴클레아제 구성요소 (Cas9 또는 Cpf1)에 대해 이중성이고 둘 다 조작될 수 있다.

[0074] "DNA 결합 분자"는 DNA에 결합할 수 있는 분자이다. 이러한 DNA 결합 분자는 폴리펩타이드, 단백질의 도메인, 더 큰 단백질 또는 폴리뉴클레오타이드 내의 도메인일 수 있다. 일부 구체예에서, 폴리뉴클레오타이드는 DNA인

한편, 다른 구체예에서, 폴리뉴클레오타이드는 RNA이다. 일부 구체예에서, DNA 결합 분자는 뉴클레아제의 단백질 도메인 (예컨대, *FokI* 도메인)인 한편, 다른 구체예에서, DNA 결합 분자는 RNA-가이드된 뉴클레아제의 가이드 RNA 구성요소 (예컨대, Cas9 또는 Cfp1)이다.

[0075] "DNA 결합 단백질" (또는 결합 도메인)은 서열-특이적 방식으로, 예를 들어 아연 핑거 단백질 또는 TALE의 하나 이상의 아연 핑거를 통해 또는 하나 이상의 RVD과의 상호작용을 통해, 각각, DNA에 결합하는 단백질, 또는 더 큰 단백질 내의 도메인이다. 용어 아연 핑거 DNA 결합 단백질은 자주 아연 핑거 단백질 또는 ZFP로서 축약된다.

[0076] "아연 핑거 DNA 결합 단백질" (또는 결합 도메인)은 서열-특이적 방식으로, 그것의 구조가 아연 이온의 배위결합을 통해 안정화되는 결합 도메인 내의 아미노산 서열 영역인 하나 이상의 아연 핑거를 통해 DNA에 결합하는 단백질, 또는 더 큰 단백질 내의 도메인이다. 용어 아연 핑거 DNA 결합 단백질은 자주 아연 핑거 단백질 또는 ZFP로서 축약된다. 인공 뉴클레아제 및 전사 인자로는 ZFP DNA-결합 도메인 및 기능성 도메인 (ZFN의 경우 뉴클레아제 도메인 또는 ZFP-TF의 경우 전사 조절 도메인)을 들 수 있다. 용어 "아연 핑거 뉴클레아제"는 표적 유전자를 절단하기 위해 다이머화하는 한 쌍의 ZFN (쌍의 구성원들은 "좌측 및 우측" 또는 "제1 및 제2" 또는 "쌍"으로 언급됨)뿐만 아니라 한 개의 ZFN을 포함한다.

[0077] "TALE DNA 결합 도메인" 또는 "TALE"은 하나 이상의 TALE 반복부 도메인/단위를 포함하는 폴리펩타이드이다. 반복부 도메인은 TALE의 그것의 동족 표적 DNA 서열에 대한 결합에 관여한다. 단일 "반복부 단위" (또한 "반복부"로도 언급됨)는 전형적으로 33 내지 35개 아미노산 길이며 자연적으로 발생하는 TALE 단백질 내의 다른 TALE 반복부 서열과 적어도 약간의 서열 상동성을 나타낸다. 예컨대, 미국 특허 제 8,586,526호 및 9,458,205호를 참조한다. 인공 뉴클레아제 및 전사 인자는 TALE DNA-결합 도메인 및 기능성 도메인 (TALEN의 경우 뉴클레아제 도메인 또는 TALEN-TF의 경우 전사 조절 도메인)을 포함한다. 용어 "TALEN"은 표적 유전자를 절단하기 위해 다이머화하는 한 쌍의 TALEN (쌍의 구성원들은 "좌측 및 우측" 또는 "제1 및 제2" 또는 "쌍"으로 언급됨)뿐만 아니라 한 개의 TALEN을 포함한다.

[0078] 아연 핑거 및 TALE DNA-결합 도메인은 예정된 뉴클레오타이드 서열에 결합하기 위하여, 예를 들어 자연적으로 발생하는 아연 핑거 단백질의 인식 나선 영역의 조작 (하나 이상의 아미노산의 변경)을 통해 또는 DNA 결합에 관여하는 아미노산 ("반복부 가변 2잔기" 또는 RVD 영역)의 조작에 의해 "조작될" 수 있다. 그러므로, 조작된 아연 핑거 단백질 또는 TALE 단백질은 자연적으로 발생하지 않는 단백질이다. 아연 핑거 단백질 및 TALE을 조작하는 방법의 비제한적인 예는 설계 및 선택이다. 설계된 단백질은 그것의 설계/조성이 주로 합리적 기준으로부터 나오는 자연에서 발생하지 않는 단백질이다. 설계를 위한 합리적 기준은 치환 규칙 및 기존의 ZFP 또는 TALE 설계 및 결합 데이터의 정보를 저장하는 데이터베이스에 정보를 처리하기 위한 전산화된 알고리즘의 적용을 포함한다. 예를 들어, 미국 특허 제 8,586,526호; 6,140,081호; 6,453,242호; 및 6,534,261호; 또한 WO 98/53058; WO 98/53059; WO 98/53060; WO 02/016536 및 WO 03/016496 참조.

[0079] "선택된" 아연 핑거 단백질, TALE 단백질 또는 CRISPR/Cas 시스템은 그것의 생성이 주로 파지 디스플레이, 상호작용 트랩, 합리적 설계 또는 하이브리드 선택과 같은 실험적 과정으로부터 나오는 자연계에서 발견되지 않는다. 예컨대, US 5,789,538; US 5,925,523; US 6,007,988; US 6,013,453; US 6,200,759; WO 95/19431; WO 96/06166; WO 98/53057; WO 98/54311; WO 00/27878; WO 01/60970; WO 01/88197 및 WO 02/099084 참조.

[0080] "TtAgo"는 유전자 침묵화에 관여하는 것으로 여겨지는 전핵 아르코노트 단백질이다. TtAgo는 박테리아 썬무스 썬모필루스(*Thermus thermophilus*)로부터 유래된다. 예컨대, Swarts *et al.*, *상기 동일*; G. Sheng *et al.*, (2013) *Proc. Natl. Acad. Sci. U.S.A.* 111, 652를 참조한다. "TtAgo 시스템"은 예컨대, TtAgo 효소에 의한 절단을 위한 가이드 DNA를 포함한 모든 필요한 구성요소이다.

[0081] "재조합"은, 한정하는 것은 아니지만 비상동성 단부 연결 (NHEJ) 및 상동성 재조합에 의한 포획을 포함한, 2개의 폴리뉴클레오타이드 사이의 유전자 정보의 교환 과정을 나타낸다. 본 개시의 목적에 대해, 상동성 재조합 (HR)"은, 예를 들어, 상동성-지정 수복 메커니즘을 통해 세포에서 이중 가닥 절단이 수복되는 중에 일어나는 이러한 교환의 특수화된 형태를 나타낸다. 이 과정은 뉴클레오타이드 서열 상동성을 필요로 하고, "표적" 분자 (즉, 이중 가닥 절단을 경험한 것)의 주형 수복에 대한 "도너" 분자를 사용하고, 그것이 도너로부터 표적까지의 유전자 정보의 전달로 이어지기 때문에, 다양하게 "비-교차 유전자 전환" 또는 "짧은 거리 유전자 전환"으로서 알려진다. 임의의 특정 이론에 얽매이기를 바라지는 않지만, 이러한 전달은 절단된 표적과 도너 사이에 형성되는 헤테로듀플렉스 DNA의 미스매치 교정, 및/또는 도너가 표적의 일부가 될 유전자 정보를 재합성하는 데 사용되는 "합성-의존성 가닥 어닐링", 및/또는 관련된 과정을 포함할 수 있다. 이러한 특수화된 HR은 자주 도너 폴리뉴클레오타이드의 서열의 일부 또는 전부가 표적 폴리뉴클레오타이드로 통합되도록 표적 분자의 서열의 변경

을 초래한다.

- [0082] 개시의 특정 방법에서, 본원에 기술된 하나 이상의 표적화된 뉴클레아제는 표적 서열 (예컨대, 세포 크로마틴)의 예정된 부위 (예컨대, 관심 있는 유전자 또는 유전자좌)에서 이중 가닥 절단 (DSB)을 생성한다. DSB는 본원에 기술된 구성물 (예컨대, 도너)의 통합을 매개한다. 선택적으로, 구성물은 절단 영역의 뉴클레오타이드 서열에 대한 상동성을 가진다. 발현 구성물은 물리적으로 통합될 수 있거나, 또는 대안으로, 발현 카세트는 상동성 재조합을 통해 절단의 수복을 위한 주형으로서 사용되어, 뉴클레오타이드 서열의 전부 또는 일부가 발현 카세트에서와 같이 세포 크로마틴으로 도입되는 결과를 초래한다. 그러므로, 세포 크로마틴의 제1 서열은 변경될 수 있고, 특정 구체예에서는, 발현 카세트에 존재하는 서열로 전환될 수 있다. 그러므로, 용어 "대체하다" 또는 "대체"의 사용은 한 뉴클레오타이드 서열의 다른 서열에 의한 대체 (즉, 정보의 의미에서 서열의 대체)를 나타내는 것으로 이해할 수 있고, 본질적으로 한 폴리뉴클레오타이드의 다른 것에 의한 물리적 또는 화학적 대체를 필요로 하지 않는다.
- [0083] 본원에 기술된 임의의 방법에서, 추가적인 조작된 뉴클레아제는 세포 내에서 추가적인 표적 부위의 추가적인 이중 가닥 절단을 위해 사용될 수 있다.
- [0084] 세포 크로마틴에서 관심 있는 영역의 서열의 표적화된 재조합 및/또는 대체 및/또는 변경을 위한 방법의 특정 구체예에서, 염색체 서열은 외인성 "도너" 뉴클레오타이드 서열로의 상동성 재조합에 의해 변경된다. 이러한 상동성 재조합은, 만약 절단 영역에 상동하는 서열이 존재한다면, 세포 크로마틴에서 이중 가닥 절단의 존재에 의해 자극된다.
- [0085] 본원에 기술된 임의의 방법에서, 제1 뉴클레오타이드 서열 ("도너 서열")은 관심 있는 영역의 게놈 서열에 상동하지만 동일하지 않은 서열을 함유할 수 있고, 그로써 관심 있는 영역에 동일하지 않은 서열을 삽입하기 위해 상동성 재조합을 자극할 수 있다. 그러므로, 특정 구체예에서, 관심 있는 영역의 서열에 상동하는 도너 서열의 부분들은 대체되는 게놈 서열에 대해 약 80 내지 90% (또는 그 사이의 임의의 정수)의 서열 동일성을 나타낸다. 다른 구체예에서, 도너와 게놈 서열 사이의 상동성은, 예를 들어 도너와 100개의 연속 염기쌍 이상의 게놈 서열 사이에서 단지 1개의 뉴클레오타이드만이 상이하다면, 99% 이상이다. 특정한 경우에, 도너 서열의 비상동성 부분은 관심 있는 영역에 존재하지 않는 서열을 함유할 수 있어서, 새로운 서열이 관심 있는 영역에 도입된다. 이들 경우에, 비상동성 서열은 일반적으로, 관심 있는 영역의 서열에 상동하거나 동일한, 50 내지 1,000 염기쌍 (또는 그 사이의 임의의 정수 값)의 서열 또는 1,000개 이상의 임의의 수의 염기쌍이 옆에 있다. 다른 구체예에서, 도너 서열은 제1 서열에 대해 상동하지 않고, 비상동성 재조합 메커니즘에 의해 게놈에 삽입된다.
- [0086] 본원에 기술된 임의의 방법은 도너 서열의 표적화된 통합에 의해 또는 표적 서열(들)의 절단과 이어서 관심 있는 유전자(들)의 발현을 파괴하는 예러가 발생하기 쉬운 NHEJ-매개 수복을 통해 세포에서 하나 이상의 표적 서열의 부분적인 또는 완전한 비활성화를 위해 사용될 수 있다. 부분적으로 또는 완전하게 비활성화된 유전자를 가진 세포주가 또한 제공된다.
- [0087] 나아가, 본원에 기술된 표적화된 통합 방법은 또한 하나 이상의 외인성 서열을 통합시키기 위해 사용될 수 있다. 외인성 핵산 서열은, 예를 들어, 하나 이상의 유전자 또는 cDNA 분자, 또는 임의의 유형의 코딩 또는 비코딩 서열, 뿐만 아니라 하나 이상의 제어 요소 (예컨대, 프로모터)를 포함할 수 있다. 더불어, 외인성 핵산 서열은 하나 이상의 RNA 분자 (예컨대, 작은 헤어핀 RNA (shRNA), 억제 RNA (RNAi), microRNAs (miRNAs), 등)를 생성할 수 있다.
- [0088] "절단"은 DNA 분자의 공유 골격의 절단을 나타낸다. 절단은, 한정하는 것은 아니지만, 포스포다이에스테르 결합의 효소적 또는 화학적 가수분해를 포함한 다양한 방법에 의해 시작될 수 있다. 단일 가닥 절단 및 이중 가닥 절단 둘 다 가능하며, 이중 가닥 절단은 2개의 구별되는 단일 가닥 절단 사건의 결과로서 발생할 수 있다. DNA 절단은 블런트 단부(blunt end) 또는 스테거 단부(staggered end)의 생성을 초래할 수 있다. 특정 구체예에서, 융합 폴리펩타이드는 표적화된 이중 가닥 DNA 절단을 위해 사용된다.
- [0089] "절단 절반 도메인"은 제2 폴리펩타이드 (동일하거나 상이함)와 함께 절단 활성 (바람직하게는 이중 가닥 절단 활성)을 가지는 복합체를 형성하는 폴리펩타이드 서열이다. 용어 "제1 및 제2 절단 절반 도메인"; "+ 및 - 절단 절반 도메인" 및 "우측 및 좌측 절단 절반 도메인"은 다이머화하는 절단 절반 도메인의 쌍을 나타내기 위해 상호교환적으로 사용된다. 용어 "절단 도메인"은 용어 "절단 절반 도메인"과 상호교환적으로 사용된다. 용어 "FokI 절단 도메인"은 임의의 FokI 상동체뿐만 아니라 SEQ ID NO:2에 제시된 FokI 서열을 포함한다.
- [0090] "조작된 절단 절반 도메인"은 또 다른 절단 절반 도메인 (예컨대, 또 다른 조작된 절단 절반 도메인)과 필수 핵

테로다이어머를 형성하기 위해 변형된 절단 절반 도메인이다.

- [0091] 용어 "서열"은 DNA 또는 RNA일 수 있고; 선형, 원형 또는 분지형일 수 있고 단일 가닥 또는 이중 가닥일 수 있는, 임의의 길이의 뉴클레오타이드 서열을 나타낸다. 용어 "도입유전자"는 게놈에 삽입되는 뉴클레오타이드 서열을 나타낸다. 도입유전자는 임의의 길이, 예를 들어 2 내지 100,000,000개 뉴클레오타이드 길이 (또는 그 사이의 또는 그 이상의 임의의 정수 값), 바람직하게는 약 100 내지 100,000개 뉴클레오타이드 길이 (또는 그 사이의 임의의 정수), 보다 바람직하게는 약 2000 내지 20,000개 뉴클레오타이드 길이 (또는 그 사이의 임의의 값), 보다 더 바람직하게는 약 5 내지 15 kb (또는 그 사이의 임의의 값)의 것일 수 있다.
- [0092] "염색체"는 세포의 게놈 전부 또는 일부를 포함하는 크로마틴 복합체이다. 세포의 게놈은 자주, 세포의 게놈을 포함하는 모든 염색체의 집합체인, 그것의 핵형에 의해 특성화된다. 세포의 게놈은 하나 이상의 염색체를 포함할 수 있다.
- [0093] "에피솜"은 복제하는 핵산, 핵단백질 복합체 또는 세포의 염색체 핵형의 일부가 아닌 핵산을 포함하는 다른 구조체이다. 에피솜의 예로는 플라스미드, 미니서플 및 특정 바이러스 게놈을 들 수 있다. 본원에 기술된 간 특이적 구성물은 에피솜에 유지될 수 있거나, 또는 대안으로, 세포에 안정적으로 통합될 수 있다.
- [0094] "외인성" 분자는 세포에 정상적으로 존재하지 않지만, 하나 이상의 유전자적, 생화학적 또는 다른 방법에 의해 세포에 도입될 수 있는 분자이다. "세포에서의 정상적인 존재"는 세포의 특정 발달 단계 및 환경 조건에 관련하여 결정된다. 그러므로, 예를 들어, 근육의 배아 발달 중에만 존재하는 분자는 성인 근육 세포와 관련하여 외인성 분자이다. 유사하게, 열 충격에 의해 유도된 분자는 비-열 충격 세포와 관련하여 외인성 분자이다. 외인성 분자는, 예를 들어, 오작동 내인성 분자의 기능화 버전 또는 정상적으로 기능하는 내인성 분자의 오작동 버전을 포함할 수 있다.
- [0095] 외인성 분자는, 다른 것들 중에서도, 예컨대 조합적인 화학적 과정에 의해 생성된 소분자, 또는 단백질, 핵산, 탄수화물, 지질, 당단백질, 리포단백질, 다당, 상기 분자들의 임의의 변형된 유도체, 또는 상기 분자들의 하나 이상을 포함하는 임의의 복합체와 같은 거대분자일 수 있다. 핵산은 DNA 및 RNA를 포함하고 단일- 또는 이중 가닥일 수 있으며; 선형, 분지형 또는 원형일 수 있고; 임의의 길이의 것일 수 있다. 핵산은 듀플렉스를 형성할 수 있는 것들, 뿐만 아니라 트리플렉스-형성 핵산을 포함한다. 예를 들어, 미국 특허 제 5,176,996호 및 5,422,251호를 참조한다. 단백질은, 한정하는 것은 아니지만, DNA-결합 단백질, 전사 인자, 크로마틴 리모델링 인자, 메틸화된 DNA-결합 단백질, 중합효소, 메틸화제, 탈메틸화제, 아세틸화제, 탈아세틸화제, 키나제, 포스파타제, 리가아제, 탈유비퀴닌화효소, 인티그라제, 재조합효소, 리가아제, 토포아이소머라제, 기라아제 및 헬리카제를 포함한다.
- [0096] 외인성 분자는 내인성 분자와 동일한 유형의 분자, 예컨대, 외인성 단백질 또는 핵산일 수 있다. 예를 들어, 외인성 핵산은 감염시키는 바이러스 게놈, 세포에 도입된 내인성 분자, 또는 세포에 정상적으로 존재하지 않는 염색체를 포함할 수 있다. 외인성 분자를 세포에 도입하는 방법은 기술분야의 숙련된 사람들에게 알려져 있고, 한정하는 것은 아니지만, 지질-매개 전달 (즉, 중성 및 양이온성 지질을 포함한 리포솜), 전기천공, 직접 주입, 세포 융합, 입자 폭발, 칼슘 포스페이트 공동 침전, DEAE-텍스트란-매개 전달 및 바이러스 벡터-매개 전달을 포함한다. 외인성 분자는 또한 내인성 분자와 동일한 유형의 분자일 수 있지만 세포가 유래되는 것과 상이한 종으로부터 유래된다. 예를 들어, 인간 핵산 서열은 원래 마우스 또는 햄스터로부터 유래된 세포주에 도입될 수 있다. 외인성 분자의 식물 세포로의 도입 방법은 기술분야의 숙련된 사람들에게 알려져 있고, 한정하는 것은 아니지만, 원형질체 형질전환, 탄화 구조 (예컨대, WHISKERS™), 아그로박테리움(*Agrobacterium*)-매개 형질전환, 지질-매개 전달 (즉, 중성 및 양이온성 지질을 포함한 리포솜), 전기천공, 직접 주입, 세포 융합, 입자 폭발 (예컨대, "유전자 총"을 사용함), 칼슘 포스페이트 공동 침전, DEAE-텍스트란-매개 전달 및 바이러스 벡터-매개 전달을 포함한다.
- [0097] 대조적으로, "내인성" 분자는 특정 환경 조건 하의 특정 발달 단계에서 특정 세포에 정상적으로 존재하는 것이다. 예를 들어, 내인성 핵산은 염색체, 미토콘드리아의 게놈, 염색체 또는 다른 소기관, 또는 자연적으로 발생하는 에피솜 핵산을 포함할 수 있다. 추가적인 내인성 분자로는 단백질, 예를 들어, 전사 인자 및 효소를 들 수 있다.
- [0098] 본원에서 사용되는 바, 용어 "외인성 핵산의 생성물"은 폴리뉴클레오타이드 및 폴리펩타이드 생성물 둘 다, 예를 들어, 전사 생성물 (RNA와 같은 폴리뉴클레오타이드) 및 번역 생성물 (폴리펩타이드)을 포함한다.
- [0099] "융합" 분자는 둘 이상의 하위유닛 분자가 바람직하게는 공유적으로 연결되어 있는 분자이다. 하위유닛 분자는

분자의 동일한 화학적 유형이거나, 또는 분자의 상이한 화학적 유형일 수 있다. 융합 분자의 예로는, 한정하는 것은 아니지만, 융합 단백질 (예를 들어, 단백질 DNA-결합 도메인과 ZFN 또는 TALEN과 같은 절단 도메인 사이의 융합), 절단 도메인과 작동 가능하게 회합된 폴리뉴클레오타이드 DNA-결합 도메인 (예컨대, sgRNA) 사이의 융합, 및 융합 핵산 (예를 들어, 융합 단백질을 암호화하는 핵산)을 들 수 있다.

[0100] 세포에서 융합 분자의 발현은 융합 분자의 구성요소의 세포로의 전달로부터 또는 융합 분자의 하나 이상의 구성요소를 암호화하는 하나 이상의 폴리뉴클레오타이드의 세포로의 전달에 의해 초래될 수 있고, 여기서 필요한 폴리뉴클레오타이드는 전사되며, 전사체는 번역되어 융합 분자가 생성된다. 트랜스-스플라이싱(Trans-splicing), 폴리펩타이드 절단 및 폴리펩타이드 결합이 또한 세포에서 단백질의 발현에 포함될 수 있다. 폴리뉴클레오타이드 및 폴리펩타이드의 세포로의 전달 방법은 본 개시의 다른 곳에서 제시된다.

[0101] 본 개시의 목적에 대해, "유전자"는 유전자 생성물을 암호화하는 DNA 영역 (상기 문헌 참조) 뿐만 아니라, 유전자 생성물의 생성을 조절하는 모든 DNA 영역을, 이러한 조절 서열이 코딩 및/또는 전사된 서열에 인접하든 아니든 간에 포함한다. 따라서, 유전자는 프로모터 서열, 종결자, 리보솜 결합 부위 및 내부 리보솜 유입 부위와 같은 번역 조절 서열, 인핸서, 사일런서(silencer), 인슐레이터(insulator), 경계선 요소, 복제 기원, 매트릭스 부착 부위 및 유전자와 제어 영역을 포함하며, 반드시 이것들에 한정되지는 않는다.

[0102] "유전자 발현"은 유전자에 함유된 정보가 유전자 생성물로 전환되는 것을 나타낸다. 유전자 생성물은 유전자의 직접적인 전사 생성물 (예컨대, mRNA, tRNA, rRNA, 안티센스 RNA, 리보자임, 구조적 RNA 또는 임의의 다른 유형의 RNA) 또는 mRNA의 번역에 의해 생성된 단백질일 수 있다. 유전자 생성물은 또한 캡핑(capping), 폴리아데닐화, 메틸화, 및 편집과 같은 과정에 의해 변형된 RNA, 및 예를 들어, 메틸화, 아세틸화, 인산화, 유비퀴틴화, ADP-리보실화, 미리스틸화, 및 글리코실화에 의해 변형된 단백질을 포함한다.

[0103] 유전자 발현의 "조절"은 유전자의 활성의 변화를 나타낸다. 발현의 조절은, 한정하는 것은 아니지만, 유전자 활성화 및 유전자 억제체를 포함할 수 있다. 게놈 편집 (예컨대, 절단, 변경, 비활성화, 무작위 돌연변이)은 발현을 조절하기 위해 사용될 수 있다. 유전자 비활성화는 본원에 기술된 ZFP, TALE 또는 CRISPR/Cas 시스템을 포함하지 않는 세포와 비교하여 유전자 발현의 어떠한 감소를 나타낸다. 그러므로, 유전자 비활성화는 부분적이거나 완전한 것일 수 있다.

[0104] "관심 있는 영역"은 외인성 분자에 결합하는 것이 바람직한, 세포 크로마틴의 임의의 영역, 예컨대, 예를 들자면, 유전자 또는 유전자 내에 있거나 유전자에 인접한 비코딩 서열이다. 결합은 표적화된 DNA 절단 및/또는 표적화된 재조합의 목적을 위한 것일 수 있다. 관심 있는 영역은 예를 들어, 염색체, 에피솜, 소기관 게놈 (예컨대, 미토콘드리아, 엽록체), 또는 감염 바이러스 게놈에 존재할 수 있다. 관심 있는 영역은 유전자의 코딩 영역 내에, 예를 들어 리더 서열, 트레일러 서열 또는 인트론과 같은 전사된 비코딩 영역 내에, 또는 전사되지 않은 영역 내에, 코딩 영역의 상류 또는 하류에 있을 수 있다. 관심 있는 영역은 단일 뉴클레오타이드 쌍처럼 작거나 길이로 최대 2,000개 뉴클레오타이드 쌍, 또는 어떠한 적분값의 뉴클레오타이드 쌍일 수 있다.

[0105] "안전한 은닉" 유전자좌는 유전자가 숙주 세포에 어떠한 유해 영향을 미치지 않으면서 삽입될 수 있는 게놈 내의 유전자좌이다. 가장 좋은 것은 삽입된 유전자 서열이 이웃한 유전자들로부터의 임의의 통독 발현(read-through expression)에 의해 방해받지 않는 안전한 은닉 유전자좌이다. 뉴클레아제(들)에 의해 표적화되는 안전한 은닉 유전자좌의 비제한적인 예로는 CR5, HPRT, AAVS1, *Rosa* 및 알부민을 들 수 있다. 예컨대, 미국 특허 제 7,951,925호; 8,771,985호; 8,110,379호; 7,951,925호; 미국 공개공보 번호 20100218264; 20110265198; 20130137104; 20130122591; 20130177983; 20130177960; 20150056705 및 20150159172를 참조한다.

[0106] "리포터 유전자" 또는 "리포터 서열"은, 비록 반드시 아니어도 바람직하게 기본적인 검정으로 쉽게 측정되는 단백질 생성물을 생성하는 모든 서열을 나타낸다. 적합한 리포터 유전자로는, 한정하는 것은 아니지만, 항생물질 내성 (예컨대, 암피실린 내성, 네오마이신 내성, G418 내성, 퓨로마이신 내성)을 매개하는 단백질을 암호화하는 서열, 착색된 또는 형광 또는 발광 단백질 (예컨대, 녹색 형광 단백질, 향상된 녹색 형광 단백질, 적색 형광 단백질, 루시페라제)을 암호화하는 서열, 및 향상된 세포 성장 및/또는 유전자 증폭을 매개하는 단백질 (예컨대, 다이하이드로폴레이트 환원효소)을 들 수 있다. 에피토프 태그는, 예를 들어, FLAG, His, myc, Tap, HA 또는 임의의 검출 가능한 아미노산 서열의 하나 이상의 복사물을 포함한다. "발현 태그"는 관심 있는 유전자의 발현을 모니터링하기 위해 원하는 유전자 서열에 작동 가능하게 연결될 수 있는 리포터를 암호화하는 서열을 포함한다.

[0107] "WPRE" 서열은 우드척 간염 바이러스로부터 유래된 우드척 간염 전사 후 조절 요소이다. WPRE는 감마, 알파, 및

베타 요소를 주어진 순서로 함유한 600 bp 길이의 트라파타이트 요소 (Donello *et al* (1992) *J Virol* 72:5085-5092)이고 AAV 시스템에서 도입유전자의 강력한 발현에 기여한다 (Loeb *et al* (1999) *Hum Gene Ther* 10:2295-2305). 이것은 또한 인트론이 없는 도입유전자의 발현을 향상시킨다. WPRE는 그것의 천연 형태에서 WHV-X 단백질을 위한 부분 오픈 리딩 프레임 (ORF)을 함유한다. WHV (We2) 인헨서와 같은 다른 바이러스 요소의 맥락에서 완전히 발현된 WHV-X 단백질은 우드척과 마우스에서 간암종의 고도 위험과 관련이 되었다 (Hohne *et. al* (1990) *EMBO J* 9(4):1137-45; Flajolet *et. al* (1998) *J Virol* 72(7):6175-80). WHV-X 단백질은 직접적으로 발암성인 것으로 나타나진 않지만, 일부 연구는 특정 환경 하에서 그것이 헤파드나바이러스 (인간의 경우 B형 간염 바이러스; 우드척의 경우 우드척 간염 바이러스)에 의한 감염과 관련된 간암의 생성에 대해 약한 보조인자로서 작용할 수 있음을 시사한다. 여러 번, "야생형" WPRE의 언급은 그것의 3' 영역에 WHV X 단백질 오픈 리딩 프레임 (ORF)의 일부를 함유하는 591 bp 서열 (GenBank 수탁 번호 J02442의 뉴클레오타이드 1094-1684)을 나타내는 것이다. 이 요소에서, 위치 1502에 WHV-X에 대한 초기 ATG 시작 코돈 및 위치 1488에 서열 GCTGA를 가진 프로모터 영역이 있다. Zanta-Boussif (상기 동일)에서, 위치 1488에 있는 프로모터 서열이 ATCAT로 변형되고 위치 1502에 있는 시작 코돈이 TTG로 변형되어, WHV-X의 발현이 효과적으로 금지되는 mut6WPRE 서열이 개시되었다. J04514.1 WPRE 변이체에서는, ATG WHV X 시작 부위가 위치 1504이고, mut6형 변이체는 이 J04514.1 스트레인에 서 만들어질 수 있다. 또 다른 WPRE 변이체는 WHV X 서열이 결핍된, 야생형 WPRE로부터의 최소 감마 및 알파 요소만을 포함하는 247 bp WPRE3 변이체이다 (Choi *et al* (2014) *Mol Brain* 7:17).

[0108] "진핵" 세포는, 한정하는 것은 아니지만, 진균 세포 (예컨대 효모), 식물 세포, 동물 세포, 포유류 세포 및 줄기 세포 (만능 및 다능을 포함함)를 포함한 인간 세포 (예컨대, T-세포)를 포함한다.

[0109] 용어 "작동 가능한 연결" 및 "작동 가능하게 연결된" (또는 "작동적으로 연결되")은 둘 이상의 구성요소 (예컨대 서열 요소)의 병렬위치를 참조로 상호교환적으로 사용되며, 이때 구성요소는 두 구성요소가 정상적으로 기능하고 구성요소 중 적어도 하나가 다른 구성요소 중 적어도 하나에 대해 발휘되는 기능을 매개할 수 있는 가능성을 허용하도록 배열된다. 예를 들자면, 전사 조절 서열, 예컨대, 프로모터는 만약 전사 조절 서열이 하나 이상의 전사 조절 인자의 존재 또는 부재에 대한 반응으로 코딩 서열의 전사 수준을 제어한다면 코딩 서열에 작동 가능하게 연결된 것이다. 전사 조절 서열은 일반적으로 코딩 서열과 시스로 작동 가능하게 연결되지만, 그것에 직접적으로 인접할 필요는 없다. 예를 들어, 인헨서는 비록 코딩 서열이 연속적이지 않더라도, 코딩 서열에 작동 가능하게 연결된 전사 조절 서열이다.

[0110] 단백질, 폴리펩타이드 또는 핵산의 "기능적 단편"은 그것의 서열이 전장 단백질, 폴리펩타이드 또는 핵산에 동일하지는 않지만, 여전히 전장 단백질, 폴리펩타이드 또는 핵산과 동일한 기능을 보유한 단백질, 폴리펩타이드 또는 핵산이다. 기능적 단편은 상응하는 천연 분자보다 많거나, 더 적거나, 또는 동일한 수의 잔기를 가질 수 있고, 및/또는 하나 이상의 아미노산 또는 뉴클레오타이드 치환을 함유할 수 있다. 핵산 또는 단백질의 기능 (예컨대, 코딩 기능, 또 다른 핵산에 혼성화하는 능력, 효소적 활성 검정)을 측정하는 방법은 기술 분야에 잘 알려져 있다.

[0111] 폴리뉴클레오타이드 "벡터" 또는 "구성물"은 표적 세포로 유전자 서열을 전달할 수 있다. 전형적으로, "벡터 구성물", "발현 벡터", "발현 구성물", "발현 카세트", 및 "유전자 전달 벡터"는 관심 있는 유전자의 발현을 지시할 수 있고 유전자 서열을 표적 세포로 전달할 수 있는 임의의 핵산 구성물을 의미한다. 그러므로, 용어는 클로닝, 및 발현 비히클뿐만 아니라 통합 벡터를 포함한다.

[0112] 용어 "대상체" 및 "환자"는 상호교환적으로 사용되며 인간 환자 및 비인간 영장류와 같은 포유류뿐만 아니라 토끼, 개, 고양이, 래트, 마우스, 및 다른 동물과 같은 실험 동물을 나타낸다. 따라서, 본원에서 사용되는 용어 "대상체" 또는 "환자"는 발명의 발현 카세트가 투여될 수 있는 임의의 포유류 환자 또는 대상체를 의미한다. 본 발명의 대상체는 장애를 가진 대상체를 포함한다.

[0113] 본원에서 사용되는 용어 "치료하는" 및 "치료"는 증상의 중증도 및/또는 빈도의 감소, 증상 및/또는 기저 원인의 제거, 증상 및/또는 그것의 기저 원인의 발생의 방지, 및 손상의 개선 또는 복원을 나타낸다. 암, 일원성 질환(monogenic diseases) 및 이식편 대 숙주 질환은 본원에 기술된 조성물 및 방법을 사용하여 치료될 수 있는 질병의 비제한적인 예이다.

[0114] "크로마틴"은 세포의 게놈을 포함하는 핵단백질 구조체이다. 세포 크로마틴은 핵산, 주로 DNA, 및 히스톤 및 비히스톤 염색체 단백질을 포함한, 단백질을 포함한다. 대부분의 진핵 세포 크로마틴은 뉴클레오솜의 형태로 존재하는데, 뉴클레오솜 코어는 히스톤 H2A, H2B, H3 및 H4 중 각각 2개를 포함하는 옥타머와 회합된 DNA의 대략 150 염기쌍을 포함하고; 링커 DNA (유기체에 따라 길이가 달라짐)는 뉴클레오솜 코어 사이로 뻗어 있다. 히스톤

H1의 분자는 일반적으로 링커 DNA와 회합된다. 본 개시의 목적에 대해, 용어 "크로마틴"은 모든 유형의 세포 핵 단백질, 원핵 및 진핵 둘 다를 포함하는 것을 의미한다. 세포 크로마틴은 염색체 및 에피솜 크로마틴을 전부 포함한다.

[0115] "접근 가능한 영역"은 핵산에 존재하는 표적 부위가 표적 부위를 인식하는 외인성 분자에 의해 결합될 수 있는 세포 크로마틴의 부위이다. 임의의 특정 이론에 의해 얽매이기를 바라지 않지만, 접근 가능한 영역은 뉴클레오솜 구조체로 포장되지 않는 영역인 것으로 여겨진다. 접근 가능한 영역의 뚜렷한 구조는 자주 화학적 및 효소적 프로브, 예를 들어, 뉴클레아제에 대한 그것의 민감성에 의해 검출될 수 있다.

[0116] "표적 부위" 또는 "표적 서열"은 충분한 결합 조건이 존재한다면, 결합 분자가 결합할 핵산의 부분을 규정하는 핵산 서열이다. 예를 들어, 서열 5'-GAATTC-3'는 Eco RI 제한 엔도뉴클레아제에 대한 표적 부위이다. "의도된" 또는 "표적-상" 서열은 결합 분자가 결합하는 것으로 의도된 서열이고 "의도되지 않은" 또는 "표적 외" 서열은 의도된 표적이 아닌 결합 분자에 의해 결합된 임의의 서열을 포함한다.

[0117] **DNA-결합 분자/도메인**

[0118] 본원에는 관심 있는 임의의 유전자 또는 유전자좌에서 표적 부위에 특이적으로 결합하는 DNA-결합 분자/도메인을 포함하는 조성물이 기술된다. 한정하는 것은 아니지만, 아연 핑거 DNA-결합 도메인, TALE DNA 결합 도메인, CRISPR/Cas 뉴클레아제의 DNA-결합 부분 (가이드 또는 sgRNA), 또는 메가뉴클레아제로부터의 DNA-결합 도메인을 포함한, 임의의 DNA-결합 분자/도메인이 본원에서 개시된 조성물 및 방법에 사용될 수 있다.

[0119] 특정 구체예에서, DNA 결합 도메인은 아연 핑거 단백질을 포함한다. 바람직하게, 아연 핑거 단백질은 그것이 선택된 표적 부위에 결합하도록 조작된다는 점에서 자연적으로 발생하지 않는다. 예를 들어, 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 제 6,453,242호; 6,534,261호; 6,599,692호; 6,503,717호; 6,689,558호; 7,030,215호; 6,794,136호; 7,067,317호; 7,262,054호; 7,070,934호; 7,361,635호; 7,253,273호; 및 미국 특허출원 공개 번호 2005/0064474; 2007/0218528; 2005/0267061 (모두 그 전문이 본원에 참조로 포함됨)을 참조한다. 특정 구체예에서, DNA-결합 도메인은 미국 특허출원 공개 번호 2012/0060230 (예컨대, 표 1) (그 전문이 본원에 참조로 포함됨)에서 개시된 아연 핑거 단백질을 포함한다.

[0120] 조작된 아연 핑거 결합 도메인은 자연적으로 발생하는 아연 핑거 단백질에 비교하여 신규한 결합 특이성을 가질 수 있다. 조작 방법은, 한정하는 것은 아니지만, 합리적인 설계 및 다양한 선택 유형을 포함한다. 합리적인 설계는, 예를 들어, 트리플렛 (또는 콰드러플렛) 뉴클레오타이드 서열 및 개별적인 아연 핑거 아미노산 서열을 포함하는 데이터베이스를 사용하는 것을 포함하고, 여기서 각각의 트리플렛 또는 콰드러플렛 뉴클레오타이드 서열은 특정 트리플렛 또는 콰드러플렛 서열에 결합하는 아연 핑거의 하나 이상의 아미노산 서열과 회합된다. 예를 들어, 미국 특허 제 6,453,242호 및 6,534,261호 (그 전문이 본원에 참조로 포함됨)를 참조한다.

[0121] 파지 디스플레이 및 2-하이브리드 시스템을 포함한, 예시의 선택 방법이 미국 특허 제 5,789,538호; 5,925,523호; 6,007,988호; 6,013,453호; 6,410,248호; 6,140,466호; 6,200,759호; 및 6,242,568호; 뿐만 아니라 WO 98/37186; WO 98/53057; WO 00/27878; WO 01/88197 및 GB 2,338,237에서 개시된다. 더불어, 아연 핑거 결합 도메인에 대한 결합 특이성의 향상이 예를 들어, 미국 특허 제 6,794,136호에서 기술되었다.

[0122] 더불어, 이들 및 다른 참고문헌에서 개시된 것과 같이, 아연 핑거 도메인 및/또는 다중-핑거 아연 핑거 단백질이, 예를 들어 길이가 5개 이상의 아미노산의 링커를 포함한, 임의의 적합한 링커 서열을 사용하여 함께 연결될 수 있다. 또한 길이가 6개 이상인 아미노산의 예시의 링커 서열에 대해 미국 특허 제 6,479,626호; 6,903,185호; 및 7,153,949호를 참조한다. 본원에 기술된 단백질은 단백질의 개별 아연 핑거 사이에 적합한 링커의 임의의 조합을 포함할 수 있다. 더불어, 아연 핑거 결합 도메인에 대한 결합 특이성의 향상이 예를 들어, 미국 특허 제 6,794,136호에서 기술되었다.

[0123] 표적 부위의 선택; ZFP 및 융합 분자 (및 그것을 암호화하는 폴리뉴클레오타이드)의 설계 및 구성 방법은 기술 분야의 숙련된 기술을 가진 사람들에게 알려져 있고 미국 특허 제 6,140,081호; 5,789,538호; 6,453,242호; 6,534,261호; 5,925,523호; 6,007,988호; 6,013,453호; 6,200,759호; 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 03/016496에서 상세하게 기술된다.

[0124] 더불어, 이들 및 다른 참고문헌에서 개시된 것과 같이, 아연 핑거 도메인 및/또는 다중-핑거 아연 핑거 단백질

은 예를 들어 길이가 5개 이상의 아미노산의 링커를 포함한, 임의의 적합한 링커 서열을 사용하여 함께 연결될 수 있다. 또한 길이가 6개 이상인 아미노산의 예시의 링커 서열에 대해 미국 특허 제 6,479,626호; 6,903,185호; 및 7,153,949호를 참조한다. 본원에 기술된 단백질은 단백질의 개별 아연 핑거 사이에 적합한 링커의 임의의 조합을 포함할 수 있다.

[0125] 보통, ZFP는 적어도 3개의 핑거를 포함한다. 특정 ZFP는 4, 5 또는 6개의 핑거를 포함한다. 3개의 핑거를 포함하는 ZFP는 전형적으로 9 또는 10개 뉴클레오타이드를 포함하는 표적 부위를 인식하며; 4개의 개의 핑거를 포함하는 ZFP는 전형적으로 12 내지 14개 뉴클레오타이드를 포함하는 표적 부위를 인식하는 한편; 6개의 개의 핑거를 포함하는 ZFP는 전형적으로 18 내지 21개 뉴클레오타이드를 포함하는 표적 부위를 인식한다. ZFP는 또한 하나 이상의 조절 도메인을 포함하는 융합 단백질일 수 있고, 그 도메인은 전사 활성화 또는 억제 도메인일 수 있다.

[0126] 일부 구체예에서, DNA-결합 도메인은 뉴클레아제로부터 유래될 수 있다. 예를 들어, I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII 및 I-TevIII과 같은 호우밍(homing) 엔도뉴클레아제 및 메가뉴클레아제의 인식 서열이 알려져 있다. 또한 미국 특허 제 5,420,032호; 미국 특허 제 6,833,252호; Belfort *et al.* (1997) *Nucleic Acids Res.* **25**:3379-3388; Dujon *et al.* (1989) *Gene* **82**:115-118; Perler *et al.* (1994) *핵산s Res.* **22**, 1125-1127; Jasin (1996) *Trends Genet.* **12**:224-228; Gimble *et al.* (1996) *J. Mol. Biol.* **263**:163-180; Argast *et al.* (1998) *J. Mol. Biol.* **280**:345-353 및 New England Biolabs 카탈로그를 참조한다. 더불어, 호우밍 엔도뉴클레아제 및 메가뉴클레아제의 DNA-결합 특이성은 비천연 표적 부위에 결합하도록 조작될 수 있다. 예를 들어, Chevalier *et al.* (2002) *Molec. Cell* **10**:895-905; Epinat *et al.* (2003) *Nucleic Acids Res.* **31**:2952-2962; Ashworth *et al.* (2006) *Nature* **441**:656-659; Paques *et al.* (2007) *Current 유전자 요법* **7**:49-66; 미국 특허출원 공개 번호 20070117128을 참조한다.

[0127] 특정 구체예에서, 본원에 기술된 아연 핑거 단백질 (예컨대, 야생형 또는 돌연변이 절단 도메인을 가진 융합 분자에서 사용됨)은 골격 영역 (예컨대, -1 내지 6으로 넘버링된 7개-아미노산 인식 나선 영역 밖의 영역)에 대해, 예를 들어 at 하나 이상의 of 위치 -14, -9 및/또는 -5 중 하나 이상의 위치에서 하나 이상의 돌연변이 (치환, 결실, 및/또는 삽입)를 포함한다 (예컨대, 도 5A 참조). 이들 위치 중 하나 이상에서 야생형 잔기가 결실되고, 임의의 아미노산 잔기로 대체되고 및/또는 하나 이상의 추가적인 잔기를 포함할 수 있다. 일부 구체예에서, 위치 -5에서의 Arg (R)은 Tyr (Y), Asp (N), Glu (E), Leu (L), Gln (Q), 또는 Ala (A)로 변경된다. 다른 구체예에서, 위치 (-9)에서의 Arg (R)은 Ser (S), Asp (N), 또는 Glu (E)로 대체된다. 추가의 구체예에서, 위치 (-14)에서의 Arg (R)은 Ser (S) 또는 Gln (Q)로 대체된다. 다른 구체예에서, 융합 폴리펩타이드는 아연 핑거 DNA 결합 도메인에 돌연변이를 포함하는데, (-5), (-9) 및/또는 (-14) 위치에서의 아미노산이 상기 열거된 아미노산 중 어느 것으로 임의의 조합으로 변경된다.

[0128] 특정 구체예에서, ZFN은 첨부된 표 또는 도면 중 어느 것에서 기술된 바대로 제1 및 제2 (좌측 및 우측) ZFN을 포함한다. 특정 구체예에서, 제1 ZFN은 71557로 표시된 ZFN을 포함하고 제2 ZFN은 71728로 표시된 ZFN을 포함한다. 특정 구체예에서, 71557로 표시된 ZFN은 AAV 벡터, 예를 들어 표 4에 제시된 서열 및/또는 SEQ ID NO:43에 제시된 서열을 포함하는 AAV 벡터로 운반된다. 다른 구체예에서, 71728로 표시된 ZFN은 AAV 벡터, 예를 들어 표 5에 제시된 서열 및/또는 SEQ ID NO:56에 제시된 서열을 포함하는 AAV 벡터로 운반된다.

[0129] 다른 구체예에서, DNA 결합 도메인은 식물 병원체 잔토모나스(*Xanthomonas*)로부터 (Boch *et al.*, (2009) *Science* 326: 1509-1512 및 Moscou and Bogdanove, (2009) *Science* 326: 1501 참조) 및 랄스토니아(*Ralstonia*) (Heuer *et al.* (2007) *Applied and Environmental Microbiology* 73(13): 4379-4384; 미국 특허출원 공개 번호 20110301073 및 20110145940 참조)로부터 유래된 것과 유사한 전사 활성화자-유사 (TAL) 이펙터 (TALE)로부터의 조작된 도메인을 포함한다. 잔토모나스 속의 식물 병원성 박테리아는 중요한 농작물에 많은 질환을 유발하는 것으로 알려져 있다. 잔토모나스의 병원성은 25개 이상의 이펙터 단백질을 식물 세포로 분출하는 보존된 유형 III 분비 (T3S) 시스템에 좌우된다. 이들 분출된 단백질 중에는 식물 전사 활성화제를 모방하고 식물 전사체를 다루는 전사 활성화자-유사 이펙터 (TALE)가 있다 (Kay *et al.* (2007) *Science* 318:648-651 참조). 이들 단백질은 DNA 결합 도메인 및 전사 활성화 도메인을 함유한다. 가장 잘 특성화되어 있는 TALE 중 하나는 잔토모나스 캄페스트그리스 pv. 베시카토리아 (*Xanthomonas campestris* pv. *Vesicatoria*)로부터의 AvrBs3이다 (Bonas *et al.* (1989) *Mol Gen Genet* 218: 127-136 및 W02010079430 참조). TALE은 탠덤 반복부의 집중화된 도메인을 함유하는데, 각각의 반복부는 약 34개의 아미노산을 함유하고, 이들 단백질의 DNA 결합 특이성에 대한 핵심이다. 더불어, 이것은 핵 국지화 서열 및 산성 전사 활성화 도메인을 함유한다 (개관을 위해 Schornack S,

et al (2006) *J Plant Physiol* 163(3): 256-272 참조). 더불어, 식물병원성 박테리아 랄스토니아 솔라나세아룸 (*Ralstonia solanacearum*)에서 *brg11* 및 *hpx17*로 표시된 2개의 유전자가 발견되었고, 이것은 랄스토니아 솔라나세아룸 생태형(biovar) 1 스트레인 GMI1000 및 생태형 4 스트레인 RS1000에서 잔토모나스의 *AvrBs3* 패밀리와 상동한다 (Heuer *et al* (2007) *Appl and Envir Micro* 73(13): 4379-4384 참조). 이들 유전자는 서로 뉴클레오타이드 서열에서 98.9% 동일하지만 *hpx17*의 반복부 도메인에서 1,575 염기쌍이 결실된 것만큼 상이하다. 그러나, 두 유전자 생성물은 전부 잔토모나스의 *AvrBs3* 패밀리 단백질과 40% 미만의 서열 동일성을 가진다.

[0130] 이들 TAL 이펙터의 특이성은 텐덤 반복부에서 발견되는 서열에 좌우된다. 반복된 서열은 대략 102개 염기쌍을 포함하고 반복부는 전형적으로 서로와 91 내지 100% 상동한다 (Bonas *et al*, 상기 동일). 반복부의 다형성은 보통 위치 12 및 13에 위치하고 TAL-이펙터의 표적 서열에서 인접한 뉴클레오타이드의 동일성을 가진 위치 12 및 13에서 추가변성 2잔기 (반복 가변성 2잔기 또는 RVD 영역)의 동일성 사이에는 1-대-1 상응성이 있는 것으로 여겨진다 (Moscou and Bogdanove, (2009) *Science* 326:1501 및 Boch *et al* (2009) *Science* 326:1509-1512 참조). 실험적으로, 이들 TAL-이펙터의 DNA 인식을 위한 자연적인 코드는 위치 12 및 13 (반복 가변성 2잔기 또는 RVD 영역)에서의 HD 서열이 시토신 (C)에 대한 결합으로 이어지고, NG는 T에 결합하며, NI는 A, C, G 또는 T에 결합하고, NN은 A 또는 G에 결합하며, ING는 T에 결합한다. 이들 DNA 결합 반복부는 새로운 서열과 상호작용할 수 있고 식물 세포에서 비내인성 리포터의 발현을 활성화할 수 있는 인공 전사 인자를 만들기 위하여, 새로운 조합 및 수의 반복부를 가진 단백질로 조립되었다 (Boch *et al*, 상기 동일). 조작된 TAL 단백질은 FokI 절단 절반 도메인에 연결되어 비전형적인 RVD를 가진 TALEN을 포함하여, TAL 이펙터 도메인 뉴클레아제 융합 (TALEN)을 유발한다. 예컨대, 미국 특허 제 8,586,526호를 참조한다.

[0131] 일부 구체예에서, TALEN은 엔도뉴클레아제 (예컨대, *FokI*) 절단 도메인 또는 절단 절반 도메인을 포함한다. 다른 구체예에서, TALE-뉴클레아제는 메가 TAL이다. 이들 메가 TAL 뉴클레아제는 TALE DNA 결합 도메인 및 메가뉴클레아제 절단 도메인을 포함하는 융합 단백질이다. 메가뉴클레아제 절단 도메인은 모노머로서 활성을 나타내며 활성을 위해 다이머화를 필요로 하지 않는다 (Boissel *et al.*, (2013) *Nucl Acid Res*: 1-13, doi: 10.1093/nar/gkt1224 참조).

[0132] 또한 추가의 구체예에서, 뉴클레아제는 압축 TALEN을 포함한다. 이것들은 TALE DNA 결합 도메인을 *TevI* 뉴클레아제 도메인에 결합시키는 단일 사슬 융합 단백질이다. 융합 단백질은 TALE DNA 결합 도메인이 *TevI* 뉴클레아제 도메인과 관련하여 위치하는 곳에 따라, TALE 영역에 의해 국지화된 니카아제로서 작용하거나, 또는 이중 가닥 절단을 생성할 수 있다 (Beurdeley *et al* (2013) *Nat Comm*: 1-8 DOI: 10.1038/ncomms2782 참조). 더불어, 뉴클레아제 도메인은 또한 DNA-결합 기능성을 나타낼 수 있다. 임의의 TALEN은 하나 이상의 메가-TALE를 가진 추가적인 TALEN (예컨대, 하나 이상의 TALEN (cTALEN 또는 *FokI*-TALEN)과 함께 사용될 수 있다.

[0133] 이에 더불어, 이들 및 다른 참고문헌에서 개시된 것과 같이, 아연 핑거 도메인 및/또는 다중-핑거 아연 핑거 단백질 또는 TALEN, 예를 들어, 5개 이상의 아미노산 길이의 링커를 포함한, 임의의 적합한 링커 서열을 사용하여 함께 연결될 수 있다. 또한, 6개 이상의 아미노산 길이의 예시의 링커 서열의 경우 미국 특허 제 6,479,626호; 6,903,185호; 및 7,153,949호를 참조한다. 본원에 기술된 단백질은 단백질의 개별 아연 핑거 사이에 적합한 링커들의 임의의 조합을 포함할 수 있다. 더불어, 아연 핑거 결합 도메인에 대한 결합 특이성의 향상은, 예를 들어, 미국 특허 제 6,794,136호에서 기술된 바 있다. 특정 구체예에서, DNA-결합 도메인은 DNA에 결합하는 단일 가이드 RNA (sgRNA) DNA 결합 분자를 포함하는, CRISPR/Cas 뉴클레아제 시스템의 일부이다. 예컨대, 미국 특허 제 8,697,359호 및 미국 특허출원 공개 번호 20150056705 및 20150159172를 참조한다. 시스템의 RNA 구성요소를 암호화하는 CRISPR (군집된 규칙적인 간격을 갖는 짧은 회문구조 반복부) 유전자좌, 및 단백질을 암호화하는 cas (CRISPR-회합된) 유전자좌 (Jansen *et al.*, 2002. *Mol. Microbiol.* 43: 1565-1575; Makarova *et al.*, 2002. *Nucleic Acids Res.* 30: 482-496; Makarova *et al.*, 2006. *Biol. Direct* 1: 7; Haft *et al.*, 2005. *PLoS Comput. Biol.* 1: e60)는 CRISPR/Cas 뉴클레아제 시스템의 유전자 서열을 구성한다. 미생물 숙주에서 CRISPR 유전자좌는 CRISPR-매개 핵산 절단의 특이성을 프로그래밍할 수 있는 비코딩 RNA 요소뿐만 아니라 CRISPR-회합된 (Cas) 유전자의 조합을 함유한다.

[0134] 일부 구체예에서, DNA 결합 도메인은 TtAgo 시스템의 일부이다 (Swarts *et al*, 상기 동일; Sheng *et al*, 상기 동일 참조). 진핵 세포에서, 유전자 침묵은 단백질의 아르고노트 (Ago) 패밀리에 의해 매개된다. 이 패러다임에서, Ago는 작은 (19 내지 31개 뉴클레오타이드) RNA에 결합된다. 이 단백질-RNA 침묵 복합체는 표적 RNA를 작은 RNA와 표적 사이의 왓슨-크릭 염기쌍을 통해 인식하고 표적 RNA를 핵산내분해적으로(endonucleolytically) 절단한다 (Vogel (2014) *Science* 344:972-973). 대조적으로, 원핵 Ago 단백질은 작은 단일 가닥 DNA 단편에 결합하여 외래 (자주 바이러스) DNA를 검출하고 제거하는 기능을 할 가능성이 있다 (Yuan *et al.*, (2005) *Mol. Cell*

19, 405; Olovnikov, *et al.* (2013) *Mol. Cell* 51, 594; Swarts *et al.*, 상기 동일). 예시의 원핵 Ago 단백질로는 아퀴팩스 아에올리쿠스(*Aquifex aeolicus*), 로도박터 스파에로이데스(*Rhodobacter sphaeroides*), 및 썬무스 썬모필루스(*Thermus thermophilus*)로부터 유래된 것들을 들 수 있다.

[0135] 가장 잘 특성화되어 있는 원핵 Ago 단백질 중 하나는 썬무스 썬모필루스로부터의 Ago 단백질이다 (TtAgo; Swarts *et al.* 상기 동일). TtAgo는 5' 포스페이트 기를 가진 15 nt 또는 13-25 nt 단일 가닥 DNA 단편과 회합한다. TtAgo가 결합된 이 "가이드 DNA"는 단백질-DNA 복합체가 DNA의 제3자(third-party) 분자의 왓슨-크릭 상보하는 DNA 서열에 결합하도록 한다. 이들 가이드 DNA의 서열 정보가 표적 DNA의 확인을 허용하게 되면, TtAgo-가이드 DNA 복합체는 표적 DNA를 절단한다. 이러한 메커니즘은 또한 TtAgo-가이드 DNA 복합체의 구조에 의해 지지되는 한편 그것의 표적 DNA에 결합한다 (G. Sheng *et al.*, 상기 동일). 로도박터 스파에로이데스로부터의 Ago (RsAgo)는 유사한 특성을 가진다 (Olovnikov *et al.* 상기 동일).

[0136] 임의의 DNA 서열의 외인성 가이드 DNA는 TtAgo 단백질 상에 로딩될 수 있다 (Swarts *et al.* 상기 동일). TtAgo 절단의 특이성이 가이드 DNA에 의해 지시되기 때문에, 외인성, 조사자-특정 가이드 DNA와 형성된 TtAgo-DNA 복합체는 그러므로 상보하는 조사자-특정 표적 DNA에 대한 TtAgo 표적 DNA 절단을 지시할 것이다. 이런 방식으로, 당업자는 DNA에 표적화된 이중 가닥 절단을 생성할 수 있다. TtAgo-가이드 DNA 시스템 (또는 다른 유기체로부터의 병렬상동성 Ago-가이드 DNA 시스템)은 세포 내에서 게놈 DNA의 표적화된 절단을 허용한다. 이러한 절단은 단일- 또는 이중 가닥일 수 있다. 포유류 게놈 DNA의 절단의 경우, 포유류 세포에서의 발현을 위해 코돈 최적화된 TtAgo의 버전을 사용하는 것이 바람직할 것이다. 추가로, TtAgo 단백질이 세포-침투 펩타이드에 융합되는 경우 시험관내에서 형성된 TtAgo-DNA 복합체로 세포를 치료하는 것이 바람직할 것이다. 추가로, 37°C에서 개선된 활성을 가지도록 돌연변이생성을 통해 변경되어 있는 TtAgo 단백질의 버전을 사용하는 것이 바람직할 것이다. Ago-RNA-매개 DNA 절단은 DNA 절단의 외삽을 위해 기술분야의 기법 표준을 사용하여 유전자 녹아웃, 표적화된 유전자 첨가, 유전자 교정, 표적화된 유전자 결실을 포함하는 결과의 모음에 영향을 미치기 위해 사용될 수 있을 것이다.

[0137] 그러므로, 임의의 DNA-결합 분자/도메인이 사용될 수 있다.

[0138] **융합 분자**

[0139] 본원에 기술된 DNA-결합 도메인 (예컨대, ZFP 또는 TALE, 단일 가이드 RNA와 같은 CRISPR/Cas 구성요소)을 포함하는 융합 분자 및 이중성 조절 (기능적) 도메인 (또는 이것의 기능적 단편)이 또한 제공된다. 공통 도메인으로는, 예컨대, 전사 인자 도메인 (활성자, 억제자, 공동-활성자, 공동-억제자), 사일런서, 중앙 유전자 (예컨대, myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos 패밀리 구성원 등); DNA 수복 효소 및 그것과 관련된 인자 및 변형자; DNA 재배열 효소 및 그것과 관련된 인자 및 변형자; 크로마틴 회합 단백질 및 이것의 변형자 (예컨대, 키나제, 아세틸라제 및 탈아세틸화제); 및 DNA 변형 효소 (예컨대, 메틸트랜스퍼라제, 토포아이스머라제, 헬리카제, 결찰효소, 키나아제, 포스포타제, 중합효소, 엔도뉴클레아제) 및 그것과 관련된 인자 및 변형자를 들 수 있다. DNA-결합 도메인과 뉴클레아제 절단 도메인의 융합과 관련한 상세한 설명에 대해 미국 특허출원 공개 번호 20050064474; 20060188987 및 2007/0218528 (그 전문이 본원에 참조로 포함됨)을 참조한다.

[0140] 활성화를 이루기에 적합한 도메인은 HSV VP16 활성화 도메인 (예컨대, Hagmann *et al.*, *J. Virol.* 71, 5952-5962 (1997) 참조) 핵 호르몬 수용체 (예컨대, Torchia *et al.*, *Curr. Opin. Cell. Biol.* 10:373-383 (1998) 참조); 핵 인자 카파 B의 p65 하위유닛 (Bitko & Barik, *J. Virol.* 72:5610-5618 (1998) 및 Doyle & Hunt, *Neuroreport* 8:2937-2942 (1997)); Liu *et al.*, *Cancer Gene Ther.* 5:3-28 (1998)), 또는 VP64와 같은 인공 키메라 기능성 도메인 (Beerli *et al.*, (1998) *Proc. Natl. Acad. Sci. USA* 95:14623-33), 및 데그론(degron) (Molinari *et al.*, (1999) *EMBO J.* 18, 6439-6447)을 들 수 있다. 추가적인 예시의 활성화 도메인으로, Oct 1, Oct-2A, Sp1, AP-2, 및 CTF1 (Seipel *et al.*, *EMBO J.* 11, 4961-4968 (1992), 뿐만 아니라 p300, CBP, PCAF, SRC1 PvALF, AtHD2A 및 ERF-2를 들 수 있다. 예를 들어, Robyr *et al.* (2000) *Mol. Endocrinol.* 14:329-347; Collingwood *et al.* (1999) *J. Mol. Endocrinol.* 23:255-275; Leo *et al.* (2000) *Gene* 245:1-11; Manteuffel-Cymborowska (1999) *Acta Biochim. Pol.* 46:77-89; McKenna *et al.* (1999) *J. Steroid Biochem. Mol. Biol.* 69:3-12; Malik *et al.* (2000) *Trends Biochem. Sci.* 25:277-283; 및 Lemon *et al.* (1999) *Curr. Opin. Genet. Dev.* 9:499-504를 참조한다. 추가적인 예시의 활성화 도메인으로는, 한정하는 것은 아니지만, OsGAI, HALF-1, C1, AP1, ARF-5,-6,-7, 및 -8, CPRF1, CPRF4, MYC-RP/GP, 및 TRAB1을 들 수 있다. 예를 들어, Ogawa *et al.* (2000) *Gene* 245:21-29; Okanami *et al.* (1996) *Genes Cells* 1:87-99; Goff *et al.* (1991) *Genes Dev.* 5:298-309; Cho *et al.* (1999) *Plant Mol. Biol.* 40:419-429; Ulmason *et al.* (1999) *Proc. Natl. Acad. Sci.*

USA 96:5844-5849; Sprenger-Haussels *et al.* (2000) *Plant J.* 22:1-8; Gong *et al.* (1999) *Plant Mol. Biol.* 41:33-44; 및 Hobo *et al.* (1999) *Proc. Natl. Acad. Sci. USA* 96:15,348-15,353을 참조한다.

- [0141] 기술분야의 숙련된 사람들에게는, DNA-결합 도메인과 기능성 도메인 사이의 융합 분자 (또는 그것을 암호화하는 핵산)의 형성에서, 활성화 도메인 또는 활성화 도메인과 상호작용하는 분자가 기능성 도메인으로서 적합한 것이 명백할 것이다. 본질적으로 표적 유전자에 대해 활성화 복합체 및/또는 활성화 활성 (예를 들어, 히스톤 아세틸화)을 증진할 수 있는 임의의 분자가 융합 단백질의 활성화 도메인으로서 유용하다. 융합 분자에서 기능성 도메인으로서 사용하기에 적합한 인슐레이터 도메인, 국지화 도메인, 및 크로마틴 리모델링 단백질, 예컨대 ISWI-함유 도메인 및/또는 메틸 결합 도메인 단백질은, 예를 들어, 미국 특허출원 공개공보 2002/0115215 및 2003/0082552 및 WO 02/44376에서 기술된다.
- [0142] 예시의 억제 도메인으로는, 한정하는 것은 아니지만, KRAB A/B, KOX, TGF-베타-유도성 초기 유전자 (TIEG), *v-erbA*, SID, MBD2, MBD3, DNMT 패밀리 구성원 (예컨대, DNMT1, DNMT3A, DNMT3B), Rb, 및 MeCP2를 들 수 있다. 예를 들어, Bird *et al.* (1999) *Cell* 99:451-454; Tyler *et al.* (1999) *Cell* 99:443-446; Knoepfler *et al.* (1999) *Cell* 99:447-450; 및 Robertson *et al.* (2000) *Nature Genet.* 25:338-342를 참조한다. 추가적인 예시의 억제 도메인으로는, 한정하는 것은 아니지만, ROM2 및 AtHD2A를 들 수 있다. 예를 들어, Chem *et al.* (1996) *Plant Cell* 8:305-321; 및 Wu *et al.* (2000) *Plant J.* 22:19-27을 참조한다.
- [0143] 융합 분자는 기술 분야의 숙련자들에게 잘 알려져 있는 클로닝 및 생화학적 컨주게이션 방법에 의해 구성된다. 융합 분자는 DNA-결합 도메인 및 기능성 도메인 (예컨대, 전사 활성화 또는 억제 도메인)을 포함한다. 융합 분자는 또한 선택적으로 핵 국지화 신호 (예컨대, 예를 들어, SV40 배지 T-항원으로부터 유래됨) 및 에피토프 태그 (예컨대, 예를 들어, FLAG 및 헤마글루티닌)를 포함한다. 융합 분자 (및 그것을 암호화하는 핵산)는 번역 리딩 프레임이 융합의 구성요소 중에서 보존되도록 설계된다.
- [0144] 한편으로 기능성 도메인 (또는 그것의 기능성 단편)의 폴리펩타이드 구성요소, 및 다른 한편으로 비단백질 DNA-결합 도메인 (예컨대, 항생물질, 인터칼레이터, 작은 홈 결합체, 핵산)이 기술분야의 숙련자들에게 알려져 있는 생화학적 컨주게이션 방법에 의해 구성된다. 예를 들어, Pierce Chemical Company (Rockford, IL) 카탈로그를 참조한다. 작은 홈 결합체와 폴리펩타이드 사이의 융합을 만들기 위한 방법 및 조성물이 기술되어 있다. Mapp *et al.* (2000) *Proc. Natl. Acad. Sci. USA* 97:3930-3935. 나아가, CRISPR/Cas 시스템의 단일 가이드 RNA는 활성 전사 조절자 및 뉴클레아제를 형성하기 위해 기능성 도메인과 회합한다.
- [0145] 특정 구체예에서, 표적 부위는 세포 크로마틴의 접근 가능한 영역에 존재한다. 접근 가능한 영역은, 예를 들어, 미국 특허 제 7,217,509호 및 7,923,542호에 기술된 것과 같이 결정될 수 있다. 만약 표적 부위가 세포 크로마틴의 접근 가능한 영역에 존재하지 않는다면, 하나 이상의 접근 가능한 영역은 미국 특허 제 7,785,792호 및 8,071,370호에서 기술된 것과 같이 생성될 수 있다. 추가적인 구체예에서, 융합 분자의 DNA-결합 도메인은 그것의 표적 부위가 접근 가능한 영역에 있거나 있지 않거나와 관계없이 세포 크로마틴에 결합할 수 있다. 예를 들어, 이러한 DNA-결합 도메인은 링커 DNA 및/또는 뉴클레오솜 DNA에 결합할 수 있다. 이런 유형의 "개척자" DNA 결합 도메인의 예는 특정 스테로이드 수용체에서 및 간세포 핵 인자 3 (HNF3)에서 발견된다 (Cordingley *et al.* (1987) *Cell* 48:261-270; Pina *et al.* (1990) *Cell* 60:719-731; 및 Cirillo *et al.* (1998) *EMBO J.* 17:244-254). 본원에 기술된 융합 분자 (예컨대, 인공 뉴클레아제)에 대한 표적 부위는 길이가 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30개 또는 그 이상 연속적이거나 비연속적인 염기쌍일 수 있다.
- [0146] 융합 분자는 기술분야에 숙련된 사람들에게 알려져 있는 것과 같이, 제약학적으로 허용되는 담체와 함께 제제화될 수 있다. 예를 들어, Remington's Pharmaceutical Sciences, 17th ed., 1985; 및 미국 특허 제 6,453,242호 및 6,534,261호를 참조한다.
- [0147] 융합 분자의 기능성 구성요소/도메인은 융합 분자가 그것의 DNA 결합 도메인을 통해 표적 서열에 결합한 후 유전자의 전사의 영향을 미칠 수 있는 다양한 상이한 구성요소 중 임의의 것으로부터 선택될 수 있다. 그러므로, 기능성 구성요소는, 한정하는 것은 아니지만, 다양한 전사 인자 도메인, 예컨대, 활성화자, 억제자, 공동-활성자, 공동-억제자, 및 사일런서를 포함한다.
- [0148] 추가적인 예시의 기능성 도메인은, 예를 들어, 미국 특허 제 6,534,261호 및 6,933,113호에서 개시된다.
- [0149] 외인성 소분자 또는 리간드에 의해 조절되는 기능성 도메인이 또한 선택될 수 있다. 예를 들어, 기능성 도메인만이 외부 RheoChem™ 리간드의 존재 하에 그것의 활성 형태를 취하는 RheoSwitch® 기술이 사용될 수 있다 (예

를 들어 US 20090136465 참조). 그러므로, ZFP는 ZFP-TF의 결과적인 활성이 외부 리간드에 의해 제어되는 조절 가능한 기능성 도메인에 작동 가능하게 연결될 수 있다.

[0150] **뉴클레아제**

[0151] 특정 구체예에서, 융합 분자는 인공 뉴클레아제를 형성하기 위하여 DNA-결합 결합 도메인 및 절단 (뉴클레아제) 도메인을 포함한다. 그로써, 유전자 변형은 뉴클레아제, 예를 들어 조작된 뉴클레아제를 사용하여 이루어질 수 있다. 조작된 뉴클레아제 기술은 자연적으로 발생하는 DNA-결합 단백질의 조작을 기반으로 한다. 예를 들어, 맞춤형 DNA-결합 사양을 가진 호우밍 엔도뉴클레아제의 조작이 기술되어 있다. Chames *et al.* (2005) *Nucleic Acids Res* 33(20):e178; Arnould *et al.* (2006) *J. Mol. Biol.* 355:443-458. 더불어, ZFP의 조작이 또한 기술되어 있다. 예컨대, 미국 특허 제 6,534,261호; 6,607,882호; 6,824,978호; 6,979,539호; 6,933,113호; 7,163,824호; 및 7,013,219호를 참조한다.

[0152] 더불어, ZFP 및/또는 TALE은 ZFN 및 TALEN - 그것의 조작된 (ZFP 또는 TALE) DNA 결합 도메인을 통해 그것의 의도된 핵산 표적을 인식할 수 있고 DNA가 뉴클레아제 활성을 통해 DNA 결합 부위 가까이에서 절단되도록 유발하는 기능성 실체를 생성하기 위하여 뉴클레아제 도메인에 융합되었다. 예컨대, Kim *et al.* (1996) *Proc Nat'l Acad Sci USA* 93(3):1156-1160을 참조한다. 보다 최근에, 이러한 뉴클레아제는 다양한 유기체에서 게놈 변형을 위해 사용되었다. 예를 들어, 미국 특허출원 공개공보 20030232410; 20050208489; 20050026157; 20050064474; 20060188987; 20060063231; 및 국제출원 공개공보 WO 07/014275를 참조한다.

[0153] 그러므로, 본원에 기술된 방법 및 조성물은 폭넓게 적용될 수 있고 관심 있는 임의의 뉴클레아제를 포함할 수 있다. 뉴클레아제의 비제한적인 예로는 메가뉴클레아제, TALEN 및 아연 핑거 뉴클레아제를 들 수 있다. 뉴클레아제는 이중성 DNA-결합 및 절단 도메인 (예컨대, 이중성 절단 도메인을 가진 아연 핑거 뉴클레아제; 메가뉴클레아제 DNA-결합 도메인)을 포함할 수 있거나 또는, 대안으로, 자연적으로 발생하는 뉴클레아제의 DNA-결합 도메인은 선택된 표적 부위에 결합하도록 변경될 수 있다 (예컨대, 동족 결합 부위와 상이한 부위에 결합하도록 조작된 메가뉴클레아제).

[0154] 특정 구체예에서, ZFN은 첨부된 표 또는 도면 중 어느 것에서 기술된 제1 및 제2 (좌측 및 우측) ZFN을 포함한다. 특정 구체예에서, 제1 ZFN은 71557로 표시된 ZFN을 포함하고 제2 ZFN은 71728로 표시된 ZFN을 포함한다. 특정 구체예에서, 71557로 표시된 ZFN은 AAV 벡터, 예를 들어 표 4에 제시된 서열 및/또는 SEQ ID NO:43에 제시된 서열을 포함하는 AAV 벡터로 운반된다. 다른 구체예에서, 71728로 표시된 ZFN은 AAV 벡터, 예를 들어 표 5에 제시된 서열 및/또는 SEQ ID NO:56에 제시된 서열을 포함하는 AAV 벡터로 운반된다.

[0155] 본원에 기술된 임의의 뉴클레아제에서, 뉴클레아제는 TALEN으로도 언급되는, 조작된 TALE DNA-결합 도메인 및 뉴클레아제 도메인 (예컨대, 엔도뉴클레아제 및/또는 메가뉴클레아제 도메인)을 포함할 수 있다. 사용자가 선택한 표적 서열과 왕성한, 부위 특이적 상호작용을 위해 이들 TALEN 단백질을 조작하기 위한 방법 및 조성물은 공개되었다 (미국 특허 제 8,586,526호 참조). 일부 구체예에서, TALEN은 엔도뉴클레아제 (예컨대, *FokI*) 절단 도메인 또는 절단 절반 도메인을 포함한다. 다른 구체예에서, TALE-뉴클레아제는 메가 TAL이다. 이들 메가 TAL 뉴클레아제는 TALE DNA 결합 도메인 및 메가뉴클레아제 절단 도메인을 포함하는 융합 단백질이다. 메가뉴클레아제 절단 도메인은 노모머로서 활성을 나타내며 활성을 위해 다이머화하는 것을 필요로 하지 않는다 (See Boissel *et al.*, (2013) *Nucl Acid Res*: 1-13, doi: 10.1093/nar/gkt1224). 더불어, 뉴클레아제 도메인은 또한 DNA-결합 기능성을 나타낼 수 있다.

[0156] 또한 추가의 구체예에서, 뉴클레아제는 압축 TALEN (cTALEN)을 포함한다. 이것들은 TALE DNA 결합 도메인을 *TevI* 뉴클레아제 도메인에 연결시키는 단일 사슬 융합 단백질이다. 융합 단백질은 TALE DNA 결합 도메인이 *TevI* 뉴클레아제 도메인과 관련하여 위치하는 곳에 따라, TALE 영역에 의해 국지화된 니카아제로서 작용하거나, 또는 이중 가닥 절단을 생성할 수 있다 (Beurdeley *et al* (2013) *Nat Comm*: 1-8 DOI: 10.1038/ncomms2782 참조). 임의의 TALEN이 추가적인 TALEN (예컨대, 하나 이상의 메가-TALE를 가진 하나 이상의 TALEN (cTALEN 또는 *FokI*-TALEN)) 또는 다른 DNA 절단 효소와 함께 사용될 수 있다.

[0157] 특정 구체예에서, 뉴클레아제는 절단 활성을 나타내는 메가뉴클레아제 (호우밍 엔도뉴클레아제) 또는 그것의 일부분을 포함한다. 자연적으로 발생하는 메가뉴클레아제는 15-40 염기쌍 절단 부위를 인식하고 보통 4개의 패밀리로 분류된다: LAGLIDADG 패밀리 (SEQ ID NO: 70으로 개시된 "LAGLIDADG"), GIY-YIG 패밀리, His-Cyst 박스 패밀리 및 HNH 패밀리. 예시의 호우밍 엔도뉴클레아제로는 I-SceI, I-CeuI, PI-PspI, PI-Sce, I-SceIV, I-CsmI, I-PanI, I-SceII, I-PpoI, I-SceIII, I-CreI, I-TevI, I-TevII 및 I-TevIII을 들 수 있다. 이들의 인식

서열은 알려져 있다. 또한 미국 특허 제 5,420,032호; 미국 특허 제 6,833,252호; Belfort *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388; Dujon *et al.* (1989) *Gene* 82:115-118; Perler *et al.* (1994) *Nucleic Acids Res.* 22, 1125-1127; Jasin (1996) *Trends Genet.* 12:224-228; Gimble *et al.* (1996) *J. Mol. Biol.* 263:163-180; Argast *et al.* (1998) *J. Mol. Biol.* 280:345-353 및 New England Biolabs 카탈로그를 참조한다.

[0158] 자연적으로 발생하는 메가뉴클레아제로부터, 주로 LAGLIDADG 패밀리 (SEQ ID NO: 70으로 개시된 "LAGLIDADG")로부터의 DNA-결합 도메인이, 식물, 효모, 초파리, 포유류 세포 및 마우스에서 부위-특이적 게놈 변형을 촉진하기 위해 사용되어 왔지만, 이 접근법은 메가뉴클레아제 인식 서열을 보존하는 상동성 유전자의 변형 (Monet *et al.* (1999), *Biochem. Biophys. Res. Commun.* 255: 88-93) 또는 인식 서열이 도입된 사전-조작된 게놈에 한정되었다 (Route *et al.* (1994), *Mol. Cell. Biol.* 14: 8096-106; Chilton *et al.* (2003), *Plant Physiology.* 133: 956-65; Puchta *et al.* (1996), *Proc. Natl. Acad. Sci. USA* 93: 5055-60; Rong *et al.* (2002), *Genes Dev.* 16: 1568-81; Gouble *et al.* (2006), *J. Gene Med.* 8(5):616-622). 따라서, 의학적으로 또는 생물공학적으로 관련된 부위에서 신규한 결합 특이성을 나타내도록 메가뉴클레아제를 조작하기 위한 시도들이 이루어졌다 (Porteus *et al.* (2005), *Nat. Biotechnol.* 23: 967-73; Sussman *et al.* (2004), *J. Mol. Biol.* 342: 31-41; Epinat *et al.* (2003), *Nucleic Acids Res.* 31: 2952-62; Chevalier *et al.* (2002) *Molec. Cell* 10:895-905; Epinat *et al.* (2003) *Nucleic Acids Res.* 31:2952-2962; Ashworth *et al.* (2006) *Nature* 441:656-659; Paques *et al.* (2007) *Current 유전자 요법* 7:49-66; 미국 특허출원 공개 번호 20070117128; 20060206949; 20060153826; 20060078552; 및 20040002092). 더불어, 자연적으로 발생하는 또는 메가뉴클레아제로부터 조작된 DNA-결합 도메인은 이중성 뉴클레아제 (예컨대, *FokI*)로부터의 절단 도메인과 작동 가능하게 연결될 수 있고 및/또는 메가뉴클레아제로부터의 절단 도메인은 이중성 DNA-결합 도메인 (예컨대, ZFP 또는 TALE)과 작동 가능하게 연결될 수 있다.

[0159] 다른 구체예에서, 뉴클레아제는 아연 핑거 뉴클레아제 (ZFN) 또는 TALE DNA 결합 도메인-뉴클레아제 융합 (TALEN)이다. ZFN 및 TALEN은 선택된 유전자 및 절단 도메인 또는 절단 절반 도메인 (예컨대, 제한 및/또는 본원에 기술된 메가뉴클레아제로부터 유래됨)에 결합하도록 조작된 DNA 결합 도메인 (아연 핑거 단백질 또는 TALE DNA 결합 도메인)을 포함한다.

[0160] 상기에서 상세하게 기술된 것과 같이, 아연 핑거 결합 도메인 및 TALE DNA 결합 도메인은 선택된 서열에 결합하도록 조작될 수 있다. 예를 들어, 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.* 10:411-416을 참조한다. 조작된 아연 핑거 결합 도메인 또는 TALE 단백질은 자연적으로 발생하는 단백질과 비교하여 신규한 결합 특이성을 가질 수 있다. 조작 방법은, 한정하는 것은 아니지만, 합리적 설계 및 다양한 유형의 선택을 포함한다. 합리적 선택은, 예를 들어, 트리플렛 (또는 콰드라플렛) 뉴클레오타이드 서열 및 개별적인 아연 핑거 또는 TALE 아미노산 서열을 포함하는 데이터베이스를 사용하는 것을 포함하는데, 여기서 트리플렛 또는 콰드라플렛 뉴클레오타이드 서열은 특정 트리플렛 또는 콰드라플렛 서열에 결합하는 아연 핑거 또는 TALE 반복부 단위의 하나 이상의 아미노산 서열과 회합된다. 예를 들어, 미국 특허 제 6,453,242호 및 6,534,261호 (그 전문이 본원에 참조로 포함됨)를 참조한다.

[0161] 표적 부위의 선택; 및 융합 분자 (및 그것을 암호화하는 폴리뉴클레오타이드)의 설계 및 구성 방법은 기술분야의 숙련된 사람들에게 알려져 있고 미국 특허 제 7,888,121호 및 8,409,861호 (그 전문이 본원에 참조로 포함됨)에서 상세하게 기술된다.

[0162] 더불어, 이들 및 다른 참고문헌에서 기술된 것과 같이, 아연 핑거 도메인, TALE 및/또는 다중-핑거 아연 핑거 단백질은, 예를 들어, 길이가 5개 이상의 아미노산인 링커를 포함한, 임의의 적합한 링커 서열을 사용하여 함께 연결될 수 있다. 예컨대, 길이가 6개 이상의 아미노산인 예시의 링커 서열에 대해 미국 특허 제 6,479,626호; 6,903,185호; 및 7,153,949호를 참조한다. 본원에 기술된 단백질은 단백질의 개별적인 아연 핑거 사이에 및/또는 DNA-결합 도메인과 뉴클레아제 도메인 사이에 적합한 링커들의 임의의 조합을 포함할 수 있다. 또한, 미국 특허 제 8,772,453호 및 9,567,609호를 참조한다.

[0163] 그러므로, ZFN, TALEN 및/또는 메가뉴클레아제와 같은 뉴클레아제는 임의의 DNA-결합 도메인 및 임의의 뉴클레아제 (절단) 도메인 (절단 도메인, 절단 절반 도메인)을 포함할 수 있다. 상기에서 주지된 것과 같이, 절단 도메인은 DNA-결합 도메인, 예를 들어 아연 핑거 또는 TAL-이펙터 DNA-결합 도메인 및 뉴클레아제로부터의 절단

도메인 또는 메가뉴클레아제 DNA-결합 도메인 및 상이한 뉴클레아제로부터의 절단 도메인에 대해 이중성일 수 있다. 이중성 절단 도메인은 임의의 엔도뉴클레아제 또는 엑소뉴클레아제로부터 얻어질 수 있다. 절단 도메인이 유래될 수 있는 예시의 엔도뉴클레아제로는, 한정하는 것은 아니지만, 제한 엔도뉴클레아제 및 호우밍 엔도뉴클레아제를 들 수 있다. 예를 들어, 2002-2003 Catalogue, New England Biolabs, Beverly, MA; 및 Belfort *et al.* (1997) *Nucleic Acids Res.* 25:3379-3388을 참조한다. DNA를 절단하는 추가적인 효소가 알려져 있다 (예컨대, S1 뉴클레아제; 녹두 뉴클레아제; 췌장 DNase I; 미세구균 뉴클레아제; 효모 HO 엔도뉴클레아제; 또한 Linn *et al.* (eds.) *Nucleases*, Cold Spring Harbor Laboratory Press, 1993 참조). 이들 효소 (또는 이들의 기능성 단편) 중 하나 이상이 절단 도메인 및 절단 절반 도메인의 공급원으로서 사용될 수 있다.

[0164] 유사하게, 절단 절반 도메인은 절단 활성을 위해 다이머화를 필요로 하는, 상기에서 제시된 것과 같은 임의의 뉴클레아제 또는 그것의 일부로부터 유래될 수 있다. 일반적으로, 만약 융합 분자가 절단 절반 도메인을 포함한다면, 2개의 융합 분자가 절단에 필요하다. 대안으로, 2개의 절단 절반 도메인을 포함하는 단일 단백질이 사용될 수 있다. 2개의 절단 절반 도메인은 동일한 엔도뉴클레아제 (또는 그것의 기능성 단편)로부터 유래되거나, 또는 각각의 절단 절반 도메인은 상이한 엔도뉴클레아제 (또는 그것의 기능성 단편)로부터 유래될 수 있다. 더불어, 두 융합 분자에 대한 표적 부위는 바람직하게, 서로와 관련하여 배치되어서, 두 융합 분자의 그들의 각각의 표적 부위에 대한 결합으로 서로에 대해 공간적인 방향으로 절단 절반 도메인이 놓이게 되어 절단 절반 도메인이, 예컨대 다이머화에 의해 기능성 절단 도메인을 형성하는 것을 허용하게 된다. 그러므로 특정 구체예에서, 쌍을 형성한 표적 부위의 거의 가장자리는 5-10개의 뉴클레오타이드에 의해 또는 15-18개의 뉴클레오타이드에 의해 분리된다. 그러나, 임의의 정수의 뉴클레오타이드 또는 뉴클레오타이드 쌍은 두 표적 부위 사이에 개재할 수 있다 (예컨대, 2 내지 50개 또는 그 이상의 뉴클레오타이드 쌍). 일반적으로 절단 부위는 표적 부위 사이에 놓여 있다.

[0165] 제한 엔도뉴클레아제 (제한 효소)는 많은 종에 존재하고 DNA에 서열-특이적 결합할 수 있고 (인식 부위에서), 결합 부위 또는 근처에서 DNA를 절단할 수 있다. 특정 제한 효소 (예컨대, 유형 IIS)는 인식 부위로부터 제거된 부위에서 DNA를 절단하고 분리 가능한 결합 및 절단 도메인을 가진다. 예를 들어, 유형 IIS 효소 *FokI*은 한 가닥 상에서 그것의 인식 부위로부터 9개 뉴클레오타이드 떨어진 곳에서 및 다른 가닥 상의 그것의 인식 부위로부터 13개 뉴클레오타이드 떨어진 곳에서 DNA의 이중 가닥 절단을 촉매한다. 예를 들어, 미국 특허 제 5,356,802호; 5,436,150호 및 5,487,994호; 또한 Li *et al.* (1992) *Proc. Natl. Acad. Sci. USA* 89:4275-4279; Li *et al.* (1993) *Proc. Natl. Acad. Sci. USA* 90:2764-2768; Kim *et al.* (1994a) *Proc. Natl. Acad. Sci. USA* 91:883-887; Kim *et al.* (1994b) *J. Biol. Chem.* 269:31,978-31,982를 참조한다. 그러므로, 한 구체예에서, 융합 분자는 적어도 한 유형 IIS 제한 효소로부터의 절단 도메인 (또는 절단 절반 도메인) 및 조작되었거나 조작되지 않았을 수 있는 하나 이상의 아연 핑거 결합 도메인을 포함한다.

[0166] 그것의 절단 도메인이 결합 도메인으로부터 분리 가능한 예시의 유형 IIS 제한 효소는 *FokI*이다. 이 특정 효소는 다이머로서 활성을 나타낸다. Bitinaite *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95: 10,570-10,575. 따라서, 본 개시의 목적에 대해, 개시된 융합 분자에 사용된 *FokI* 효소의 일부가 절단 절반 도메인으로 간주된다. 그러므로, 아연 핑거-*FokI* 융합을 사용한 세포 서열의 표적화된 이중 가닥 절단 및/또는 표적화된 대체를 위해, 각각이 *FokI* 절단 절반 도메인을 포함한 두 융합 분자가 촉매적으로 활성인 절단 도메인을 재구성하기 위해 사용될 수 있다. 대안으로, 아연 핑거 결합 도메인 및 2개의 *FokI* 절단 절반 도메인을 함유하는 단일 폴리펩타이드 분자가 또한 사용될 수 있다. 아연 핑거-*FokI* 융합을 사용하는 표적화된 절단 및 표적화된 서열 변경에 대한 파라미터는 본 개시의 다른 곳에서 제공된다.

[0167] 절단 도메인 또는 절단 절반 도메인은 절단 활성을 보유한, 또는 기능성 절단 도메인을 형성하기 위하여 멀티머화하는 (예컨대, 다이머화하는) 능력을 보유한 단백질의 임의의 부분일 수 있다.

[0168] 예시의 유형 IIS 제한 효소는 국제 공개공보 WO 07/014275 (그 전문이 본원에 포함됨)에서 기술된다. 추가적인 제한 효소는 또한 분리 가능한 결합 및 절단 도메인을 함유하고, 이것들은 본 개시에 의해 포함된다. 예를 들어, Roberts *et al.* (2003) *Nucleic Acids Res.* 31:418-420을 참조한다.

[0169] 특정 구체예에서, 절단 도메인은 결정 구조 1FOK.pdb 및 2FOK.pdb를 생성하기 위하여 사용된 *FokI* 절단 도메인을 포함한다 (Wah *et al.* (1997) *Nature* 388:97-100 참조). 전장 *FokI*의 서열은 하기에 제시된다. 본원에 기술된 뉴클레아제에 사용된 절단 도메인은 이탤릭체 및 밑줄로 제시되며 (전장 단백질의 위치 384 내지 579) ho10 단백질 서열이 아래에 제시된다 (SEQ ID NO:2):

[0170] M V S K I R T F G W V Q N P G K F E N L K R V V Q V F D R N S K V H N E V K N I K I P T L V K E S K I Q K E L V A I M N Q H D L I Y T Y K E L

[0171] VGTGTSIRSEAPCDAI IQATIADQGNKKG YIDNWS SDGFLRWAHALGF IEYINKSDSFVITDVGLAYSKSA
 [0172] DGSAIEKEIL IEA I SSYP 쌍 ILTLEDGQH LTKFDL GKNLGFSGESGFTSLPEG ILLD TLANAMPKDKG
 [0173] EIRNNWEGSSDKYARMIGGWLDKLG LVKQGGKEFI IPTLGKPDNKEFI SHAFKITGEGLKVLRRAKGSTKF
 [0174] TRVPKRIVYWEMLATNLTDKEYVRTRRALILE ILIKAGSLKIEQIQDNLKKLGFDEVIETIENDIKGLINTG
 [0175] IFIEIKGRFYQLKDHILQFVIPNRGVTK QLVKSELEEKKSELRHKLKYPHEYTEL IEIARNSTQDRILEM
 [0176] KVMEFFMKVYGRGKHLGGSRKPDGAIYTVGSPIDYGVIVDTKAYSGGYNLP IGOADEMQRYVEENQTRNK
 [0177] HINPNEWKVPSSVTEFKFLFVSGHFKGNYKAQLTRLNHITNCNGAVLSVEELL IGGEMIKAGTLTLEEY
 [0178] RRKFNNGEINF (SEQ ID NO:2)

[0179] *FokI*로부터 유래된 절단 절반 도메인은 SEQ ID NO:2에 제시된 아미노산 잔기 중 하나 이상에서 돌연변이를 포함할 수 있다. 돌연변이는 (상이한 잔기에 대한 야생형 아미노산 잔기의) 치환, (하나 이상의 아미노산 잔기의) 삽입 및/또는 (하나 이상의 아미노산 잔기의) 결실을 포함한다. 특정 구체예에서, 잔기 414-426, 443-450, 467-488, 501-502, 및/또는 521-531 (SEQ ID NO:2와 관련하여 넘버링됨) 중 하나 이상은, 이들 잔기가 Miller 등에서 기술된 그것의 표적 부위에 결합된 ZFN의 분자 모델에서 DNA 골격에 가까이 위치하기 때문에 (Miller *et al.* ((2007) *Nat Biotechnol* 25:778-784) 돌연변이된다. *FokI* 돌연변이의 비제한적인 예로는, 한정하는 것은 아니지만, 위치 416, 421, 422, 424, 472, 478, 480, 525 또는 542에서 하나 이상의 KSRL가 돌연변이된 것을 포함하여, 미국 특허출원 공개 번호 20180087072의 본원에 기술된 하나 이상의 돌연변이를 포함한다. 특정 구체예에서, 돌연변이는 야생형 잔기의 임의의 상이한 잔기, 예를 들어 알라닌 (A) 잔기, 시스테인 (C) 잔기, 아스파르트산 (D) 잔기, 글루탐산 (E) 잔기, 히스티딘 (H) 잔기, 페닐알라닌 (F) 잔기, 글리신 (G) 잔기, 아스파라긴 (N) 잔기, 세린 (들) 잔기 또는 트레오닌 (T) 잔기로의 치환을 포함한다. 다른 구체예에서, 위치 416, 418, 421, 422, 424, 446, 448, 472, 476, 478, 479, 480, 481, 525 및/또는 542 중 하나 이상에서의 야생형 잔기가 임의의 다른 잔기로 대체되며, R416D, R416E, S418E, S418D, R422H, S446D, K448A, N476D, P478S, I479Q, I479T, G480D, Q481A, Q481E, K525S, K525A, N527D, N542D, R416E+R422H, R416D+R422H, R416E+K448A, R416D+R422H, K448A+I479Q, K448A+Q481A, K448A+K525A, R416E, R416D, R416H, R416N, S418D, S418E, D421S, L424F, S446D, K448A, S472D, N476E, N476G, N476K, P478D, I479Q, I479T, G480D, Q481A, Q481C, Q481D, Q481S, Q481E Q481H, K525A, K525C, K525AE, K525I, K525S, K525T, K525V, 및/또는 N542D를 포함하지만, 이에 한정되지는 않는다.

[0180] 특정 구체예에서, 절단 도메인은 예를 들어, 미국 특허 제 7,914,796호; 8,034,598호 및 8,623,618호; 및 미국 특허출원 공개 번호 20110201055 (모든 문서의 개시내용은 그 전문이 본원에 참조로 포함됨)에서 기술된 것과 같이, 호모다имер화를 최소화 또는 방지하는 하나 이상의 조각된 절단 절반 도메인 (또한 다имер화 도메인 돌연변이로도 언급됨)을 포함한다. *FokI*의 위치 446, 447, 478, 479, 483, 484, 486, 487, 490, 491, 496, 498, 499, 500, 531, 534, 537, 538 및 542 (SEQ ID NO:2와 관련하여 넘버링됨)에 있는 아미노산 잔기는 모두 *FokI* 절단 절반 도메인의 다имер화에 영향을 미치기 위한 표적이다. 돌연변이는 *FokI*에 상동성인 자연적인 제한 효소에서 발견된 잔기들에 대한 돌연변이를 포함한다. 바람직한 구체예에서, 위치 416, 422, 447, 448, 478, 525 및/또는 542 (SEQ ID NO:2와 관련하여 넘버링됨)에 있는 돌연변이는 양전하 아미노산의 대전되지 않은 또는 음전하 아미노산으로의 대체를 포함한다. 다른 구체예에서, 조각된 절단 절반 도메인은 하나 이상의 아미노산 잔기 416, 422, 447, 448, 또는 525에서의 돌연변이에 더불어 아미노산 잔기 499, 496 및 486에서의 돌연변이를 포함하며, 이때 잔기는 모두 SEQ ID NO:2와 관련하여 넘버링된다.

[0181] 특정 구체예에서, 본원에 기술된 조성물은 예를 들어, 미국 특허 제 7,914,796호; 8,034,598호; 8,961,281호 및 8,623,618호; 미국 특허출원 공개 번호 20080131962 및 20120040398에서 기술된 것과 같이 필수 헤테로다имер을 형성하는 *FokI*의 조각된 절단 절반 도메인을 포함한다. 그러므로, 한 바람직한 구체예에서, 발명은 조각된 절단 절반 도메인이 위치 416, 422, 447, 448, 또는 525 (SEQ ID NO:2와 관련하여 넘버링됨)에서의 하나 이상의 돌연변이에 더불어 위치 486에서의 야생형 Gln (Q) 잔기가 Glu (E) 잔기로 대체되고, 위치 499에서 야생형 Ile (I) 잔기가 Leu (L) 잔기로 대체되며 위치 496에서 야생형 Asn (N) 잔기가 Asp (D) 또는 Glu (E) 잔기로 대체된 ("ELD" 또는 "ELE") 폴리펩타이드를 포함한다. 다른 구체예에서, 조각된 절단 절반 도메인은 야생형 *FokI* 절단 절반 도메인으로부터 유래되고 아미노산 잔기 416, 422, 447, 448, 또는 525에서의 하나 이상의 돌연변이에 더불어 야생형 *FokI* (SEQ ID NO:2)과 관련하여 넘버링된, 아미노산 잔기 490, 538 및 537에서의 돌연변이를 포

함한다. 바람직한 구체예에서, 발명은 융합 분자를 제공하며, 여기서 조작된 절단 절반 도메인은 위치 416, 422, 447, 448, 또는 525에서의 하나 이상의 돌연변이에 더불어 위치 490에서의 야생형 Glu (E) 잔기가 Lys (K) 잔기로 대체되고, 위치 537에서의 야생형 Ile (I) 잔기가 Lys (K) 잔기로 대체되며, 위치 537에서의 야생형 His (H) 잔기가 Lys (K) 잔기 또는 Arg (R) 잔기로 대체된 ("KKK" 또는 "KKR") (본원에 참조로 포함되는 U.S. 8,962,281 참조) 폴리펩타이드를 포함한다. 예컨대, 미국 특허 제 7,914,796호; 8,034,598호 및 8,623,618호 (이것들의 개시 내용은 모든 목적에 대해 그 전문이 참조로 포함됨)를 참조한다. 다른 구체예에서, 위치 542에서의 야생형 Asn (N) 잔기가 Asp (D) 잔기로 대체되거나 또는 위치 478에서의 야생형 Pro (P) 잔기가 Ser (S) 잔기로 대체된다. 다른 구체예에서, 조작된 절단 절반 도메인은 "샤키(Sharkey)" 및/또는 "샤키" 돌연변이를 포함한다 (Guo *et al.*, (2010) *J. Mol. Biol.* 400(1):96-107 참조).

[0182] 또 다른 구체예에서, 조작된 절단 절반 도메인은 야생형 *FokI* 절단 절반 도메인으로부터 유래되고 아미노산 잔기 416, 422, 447, 448, 또는 525에서의 하나 이상의 돌연변이에 더불어 야생형 *FokI* 또는 *FokI* 상동체와 관련하여 넘버링된 아미노산 잔기 490, 및 538에서의 돌연변이를 포함한다. 바람직한 구체예에서, 발명은 융합 분자를 제공하며, 여기서 조작된 절단 절반 도메인은 위치 416, 422, 447, 448, 또는 525에서의 하나 이상의 돌연변이에 더불어 위치 490에서의 야생형 Glu (E) 잔기가 Lys (K) 잔기로 대체되고, 위치 538에서의 야생형 Ile (I) 잔기가 Lys (K) 잔기로 대체된 ("KK") 폴리펩타이드를 포함한다. 바람직한 구체예에서, 발명은 융합 분자를 제공하며, 여기서 조작된 절단 절반 도메인은 위치 416, 422, 447, 448, 또는 525에서의 하나 이상의 돌연변이에 더불어 위치 486에서의 야생형 Gln (Q) 잔기가 Glu (E) 잔기로 대체되고, 위치 499에서의 야생형 Ile (I) 잔기가 Leu (L) 잔기로 대체된 ("EL") 폴리펩타이드를 포함한다 (본원에 참조로 포함되는 U.S. 8,034,598 참조).

[0183] 한 측면으로, 발명은 조작된 절단 절반 도메인이 *FokI* 촉매성 도메인의 위치 387, 393, 394, 398, 400, 402, 416, 422, 427, 434, 439, 441, 447, 448, 469, 478, 487, 495, 497, 506, 516, 525, 529, 534, 542, 559, 569, 570, 571 중 하나 이상에 있는 야생형 아미노산 잔기가 돌연변이된 폴리펩타이드를 포함하는 융합 분자를 제공한다. 첨부된 표 및 도면 중 어느 것에 제시된 하나 이상의 돌연변이를 포함하는 뉴클레아제 도메인이 제공된다. 일부 구체예에서, 하나 이상의 돌연변이는 양전하 잔기로부터의 야생형 아미노산을 중성 잔기 또는 음전하 잔기로 변경시킨다. 이들 구체예 중 어느 것에서, 기술된 돌연변이는 또한 하나 이상의 추가적인 돌연변이를 포함하는 *FokI* 도메인에서 만들어질 수 있다. 바람직한 구체예에서, 이들 추가적인 돌연변이는, 예컨대, 위치 418, 432, 441, 481, 483, 486, 487, 490, 496, 499, 523, 527, 537, 538 및/또는 559에서의 다이머화 도메인에 있다. 돌연변이의 비제한적인 예는 위치 393, 394, 398, 416, 421, 422, 442, 444, 472, 473, 478, 480, 525 또는 530에 있는 임의의 절단 도메인 (예컨대, *FokI* 또는 *FokI*의 상동체)의 야생형 잔기의 임의의 아미노산 잔기로의 돌연변이 (예컨대, 치환) (예컨대, K393X, K394X, R398X, R416S, D421X, R422X, K444X, S472X, G473X, S472, P478X, G480X, K525X, A530X 및/또는, N542X, 여기서 첫 번째 잔기는 야생형을 나타내고 X는 야생형 잔기에 대해 치환되는 임의의 잔기를 나타냄)를 포함한다. 일부 구체예에서, X는 E, D, H, A, K, S, T, D 또는 N이다. 다른 예의 돌연변이로는 S418E, S418D, S446D, K448A, P478S, I479Q, I479T, Q481A, Q481N, Q481E, A530E, A530K 및/또는 N542D를 들 수 있고, 여기서 아미노산 잔기는 전장 *FokI* 야생형 절단 도메인 및 그것의 상동체와 관련하여 넘버링된다. 특정 구체예에서, 조합은 416 및 422, 위치 416에서의 돌연변이 및 K448A, K448A 및 I479Q, K448A 및 Q481A 및/또는 K448A 및 위치 525에서의 돌연변이를 포함한다. 한 구체예에서, 위치 416에서의 야생형 잔기는 Glu (E) 잔기로 대체될 수 있고 (R416E), 위치 422에서의 야생형 잔기는 His (H) 잔기로 대체되며 (R422H), 위치 525에서의 야생형 잔기는 Ala (A) 잔기로 대체된다. 본원에 기술된 절단 도메인은 추가로, 한정하는 것은 아니지만 위치 432, 441, 483, 486, 487, 490, 496, 499, 527, 537, 538 및/또는 559에서의 추가적인 돌연변이, 예를 들어 다이머화 도메인 돌연변이 (예컨대, ELD, KKR) 및 또는 니카아제 돌연변이 (촉매성 도메인에 대한 돌연변이)를 포함할 수 있다. 본원에 기술된 돌연변이를 가진 절단 절반 도메인은 기술분야에 알려진 것과 같이 헥테로다이머를 형성한다.

[0184] 대안으로, 뉴클레아제는 생체내에서 핵산 표적 부위에서 소위 "스플릿-효소" 기술을 사용하여 조립될 수 있다 (예컨대, 미국 특허출원 공개 번호 20090068164 참조). 이러한 스플릿 효소의 구성요소는 별도의 발현 구성물 상에서 발현되거나, 또는 개별적인 구성요소가, 예를 들어, 자가-절단 2A 펩타이드 또는 IRES 서열에 의해 분리되어 있는 한 오픈 리딩 프레임에서 연결될 수 있다. 구성요소는 개별적인 아연 핑거 결합 도메인 또는 메가뉴클레아제 핵산 결합 도메인의 도메인일 수 있다.

[0185] 뉴클레아제 (예컨대, ZFN 및/또는 TALEN)는, 예를 들어 미국 특허 제 8,563,314호에서 기술된 것과 같은 효모-기반 염색체 시스템에서 사용 전에 활성화에 대해 스크리닝될 수 있다.

[0186] 특정 구체예에서, 뉴클레아제는 CRISPR/Cas 시스템을 포함한다. 시스템의 RNA 구성요소를 암호화하는 CRISPR

(군집된 규칙적인 간격을 갖는 짧은 회문구조 반복부) 유전자좌, 및 단백질을 암호화하는 Cas (CRISPR-회합) 유전자좌 (Jansen *et al.*, 2002. *Mol. Microbiol.* 43: 1565-1575; Makarova *et al.*, 2002. *Nucleic Acids Res.* 30: 482-496; Makarova *et al.*, 2006. *Biol. Direct* 1: 7; Haft *et al.*, 2005. *PLoS Comput. Biol.* 1: e60)는 CRISPR/Cas 뉴클레아제 시스템의 유전자 서열을 구성한다. 미생물 숙주에서 CRISPR 유전자좌는 CRISPR-매개 핵산 절단의 특이성을 프로그래밍할 수 있는 비코딩 RNA 요소뿐만 아니라 CRISPR-회합 (Cas) 유전자의 조합을 함유한다.

[0187] 유형 II CRISPR은 가장 잘 특성화된 시스템 중 하나이고 4개의 순차적인 단계로 표적화된 DNA 이중 가닥 절단을 수행한다. 첫째, 2개의 비코딩 RNA, 사전-crRNA 어레이 및 tracrRNA가 CRISPR 유전자좌로부터 전사된다. 둘째, tracrRNA는 사전-crRNA의 반복부 영역에 혼성화하여 사전-crRNA의 개별 스페이서 서열을 함유하고 있는 성숙한 crRNA로의 프로세싱을 매개한다. 셋째, 성숙한 crRNA:tracrRNA 복합체는, 표적 인식을 위한 추가적인 필요조건인, crRNA 상의 스페이서와 프로토스페이서 인접 모티프 (PAM) 다음에 있는 표적 DNA 상의 프로토스페이서 사이의 왓슨-크릭 염기쌍 형성을 통해 표적 DNA로 Cas9를 향하게 한다. 마지막으로, Cas9는 표적 DNA의 절단을 매개하여 프로토스페이서 내에서 이중 가닥 절단을 생성한다. CRISPR/Cas 시스템의 활성화는 3 단계로 구성된다: (i) '적응'으로 불리는 과정에서, 미래의 공격을 방지하기 위한 CRISPR 어레이로의 외래 DNA의 삽입, (ii) 관련 단백질의 발현, 뿐만 아니라 어레이의 발현 및 프로세싱, 이어서 (iii) 외래 핵산에 대한 RNA-매개 간섭. 그러므로, 박테리아 세포에서, 소위 'Cas' 단백질은 여러 개가 CRISPR/Cas 시스템의 자연적 기능에 관여하며 외래 DNA 등의 삽입과 같은 기능에서 역할을 담당한다.

[0188] 일부 구체예에서, CRISPR-Cpf1 시스템이 사용된다. 프란시셀라 종(*Francisella spp*)에서 확인된 CRISPR-Cpf1 시스템은 인간 세포에서 활발한 DNA 간섭을 매개하는 부류 2 CRISPR-Cas 시스템이다. 비록 기능적으로 보존되긴 했지만, Cpf1 및 Cas9는 그것들의 가이드 RNA 및 기질 특이성을 포함하는 많은 측면에서 상이하다 (Fagerlund *et al.*, (2015) *Genom Bio* 16:251 참조). Cas9 및 Cpf1 단백질 사이의 주요 차이점은 Cpf1이 tracrRNA를 활용하지 않고, 따라서 crRNA만을 필요로 하는 것이다. FnCpf1 crRNA는 42-44개 뉴클레오타이드 길이 (19개의 뉴클레오타이드 반복부 및 23-25개의 뉴클레오타이드 스페이서)이며 이차 구조를 보유하는 서열 변화를 감당하는 단일 스템 루프를 함유한다. 더불어, Cpf1 crRNA는 Cas9가 필요로 하는 약 100개 뉴클레오타이드 조작된 sgRNA보다 상당히 더 짧고, FnCpf1에 대한 PAM 필요조건은 대체된 가닥 상의 5'TTN-3' 및 5'-CTA-3'이다. 비록 Cas9 및 Cpf1 둘 다 표적 DNA에서 이중 가닥 절단을 생성하지만, Cas9는 가이드 RNA의 시드(seed) 서열 내에서 블런트-단부 절단을 생성하기 위해 그것의 RuvC- 및 HNH-유사 도메인을 사용하는 반면, Cpf1은 시드의 외부에 엇갈린 절단을 생성하기 위해 RuvC-유사 도메인을 사용한다. Cpf1이 결정적인 시드 영역으로부터 먼 곳을 절단하기 때문에, NHEJ가 표적 부위를 파괴하지 못할 것이고, 따라서 Cpf1은 원하는 HDR 재조합 사건이 일어날 때까지 동일한 부위를 계속해서 절단할 수 있다. 그러므로, 본원에 기술된 방법 및 조성물에서, 용어 "Cas"는 Cas9 및 Cpf1 단백질 둘 다를 포함하는 것으로 이해된다. 그러므로, 본원에서 사용되는 바, "CRISPR/Cas 시스템"은 뉴클레아제 및/또는 전사 인자 시스템 둘 다를 포함한, CRISPR/Cas 및/또는 CRISPR/Cpf1 시스템 둘 다를 지칭한다.

[0189] 특정 구체예에서, Cas 단백질은 자연적으로 발생하는 Cas 단백질의 "기능성 유도체"일 수 있다. 천연 서열 폴리펩타이드의 "기능성 유도체"는 천연 서열 폴리펩타이드와 공통되는 정성적 생물학적 특성을 가진 화합물이다. "기능성 유도체"로는, 한정하는 것은 아니지만, 천연 서열의 단편 및, 상응하는 천연 서열 폴리펩타이드와 공통된 생물학적 활성을 가진다면, 천연 서열 폴리펩타이드의 유도체 및 그것의 단편을 들 수 있다. 본원에서 고려되는 생물학적 활성은 DNA 기질을 단편으로 가수분해하는 기능성 유도체의 능력이다. 용어 "유도체"는 폴리펩타이드의 아미노산 서열 변이체, 공유성 변형, 및 유도체 Cas 단백질과 같은 그것의 융합을 포함한다. Cas 폴리펩타이드 또는 그것의 단편의 적합한 유도체는, 한정하는 것은 아니지만, Cas 단백질 또는 그것의 단편의 돌연변이, 융합, 공유성 변형을 포함한다. Cas 단백질 또는 그것의 단편, 뿐만 아니라 Cas 단백질 또는 그것의 단편의 유도체를 포함하는 Cas 단백질은 세포로부터 얻을 수 있거나 또는 화학적으로 합성할 수 있거나 또는 이들 두 과정의 조합에 의해 얻을 수 있다. 세포는 자연적으로 Cas 단백질을 생성하는 세포, 또는 자연적으로 Cas 단백질을 생성하고 내인성 Cas 단백질을 더 높은 발현 수준으로 생성하기 위하여 또는 내인성 Cas와 동일하거나 상이한 Cas를 암호화하는, 외래에서 도입된 핵산으로부터 Cas 단백질을 생성하기 위하여 유전자 조작되는 세포일 수 있다. 일부 경우에, 세포는 Cas 단백질을 자연적으로 생성하지 않고 Cas 단백질을 생성하기 위해 유전자 조작된다. 일부 구체예에서, Cas 단백질은 AAV 벡터를 통한 전달을 위한 작은 Cas9 병렬상동체이다 (Ran *et al* (2015) *Nature* 510, p. 186).

[0190] 뉴클레아제(들)는 표적 부위에서 하나 이상의 이중 가닥 및/또는 단일 가닥 절단을 생성할 수 있다. 특정 구체예에서, 뉴클레아제는 촉매적으로 비활성인 절단 도메인 (예컨대, *FokI* 및/또는 Cas 단백질)을 포함한다. 예컨대

대, 미국 특허 제 9,200,266호; 8,703,489호 및 Guillinger *et al.* (2014) *Nature Biotech.* 32(6):577-582를 참조한다. 촉매적으로 비활성인 절단 도메인은, 촉매적으로 활성인 도메인과 함께, 니카아제로서 작용하여 단일 가닥을 절단할 수 있다. 그러므로, 2개의 니카아제가 조합하여 사용되어 특정 영역에서 이중 가닥을 절단할 수 있다. 추가적인 니카아제가 또한 기술분야에, 예를 들어, McCaffery *et al.* (2016) *Nucleic Acids Res.* 44(2):e11. doi: 10.1093/nar/gkv878. Epub 2015 Oct 19에서 공지된다.

[0191] 특정 구체예에서, 뉴클레아제는 각각 아연 핑거 DNA-결합 도메인 및 절단 도메인 (예컨대, 조작된 FokI)을 포함하는 제1 및 제2 (also referred to as "좌측 및 우측" 및 "partner") 아연 핑거 뉴클레아제를 포함하는 아연 핑거 뉴클레아제이다. ZFN은 by 하나 이상의 AAV 벡터에 의해 운반될 수 있다. 특정 구체예에서, 별도의 AAV 벡터가 뉴클레아제의 좌측 및 우측 ZFN을 운반한다. AAV 벡터(들)는, 한정하는 것은 아니지만, 5' ITR, 하나 이상의 인핸서 서열 (예컨대, ApoE 인핸서), 하나 이상의 프로모터 서열 (예컨대, hAAT 프로모터), 5' UTR, 하나 이상의 인트론 서열 (예컨대, 인간 β 글로빈/IgG 키메라 인트론), N-말단 펩타이드 코딩 서열, NLS 신호, 하나 이상의 WPRE 서열 (예컨대, WPREmut6), 폴리A 신호 및/또는 3' ITR을 포함하여, 추가적인 코딩 및/또는 비코딩 서열을 포함할 수 있다. 예시의 뉴클레아제 AAV는 하기 표 4 및 5에서 제시된다. 열거된 요소 중 하나 이상 (ZFN-암호화 서열을 배제함)이 생략될 수 있고; 유사한 서열 (예컨대, 상이한 프로모터 서열, 상이한 WPRE 서열, 예컨대 기술분야에 알려져 있는 또는 실시예 4에서 기술되는 것들), 상이한 인트론 서열, 등)로 대체될 수 있고; 및/또는 추가적인 요소가 첨가될 수 있는 것이 분명할 것이다. 뉴클레아제를 암호화하는 AAV 벡터(들)가 도너를 가진 시스템에서, 예를 들어 2개의 ZFN AAV (예컨대, 표 4 및 5에서 개시된 것과 같은 좌측 및 우측 ZFN AAV)가, 전형적으로 치료용 펩타이드를 암호화하는 도너 AAV와 함께 사용될 수 있다. AAV 도너는 다음 요소 중 하나 이상을 포함할 수 있다: 임의의 공급원으로부터의 5' 및/또는 3' ITR; 임의의 길이의 도입유전자 (임의의 단백질 또는 그것의 기능성 단편을 암호화하는 임의의 길이의 치료용 단백질-암호화 서열) 옆에 있는 (알부민에 대해) 좌측 및/또는 우측 상동성 팔; 스플라이스 엑셉터 서열; 및/또는 폴리아데닐화 (폴리아) 신호. 특정 구체예에서, AAV 도너는 인자 IX, IDS 또는 IDUA 단백질을 암호화하고, 예를 들어 도너는 하기 표 6 내지 8에서 제시된다.

[0192] 전달

[0193] 본원에 기술된 단백질 (예컨대, 뉴클레아제), 폴리뉴클레오타이드 및/또는 단백질 및/또는 폴리뉴클레오타이드를 포함하는 조성물은 임의의 적합한 수단에 의해, 예를 들어, 단백질 및/또는 mRNA 구성요소의 주입에 의해 표적 세포에 전달될 수 있다.

[0194] 적합한 세포는, 한정하는 것은 아니지만 진핵 및 원핵 세포 및/또는 세포주를 포함한다. 이러한 세포 또는 이러한 세포로부터 생성된 세포주의 비제한적인 예로는 T-세포, COS, CHO (예컨대, CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11, CHO-DUKX, CHOK1SV), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NSO, SP2/0-Ag14, HeLa, HEK293 (예컨대, HEK293-F, HEK293-H, HEK293-T), 및 스포도프테라 푸지퍼다(*Spodoptera fugiperda*)(Sf)와 같은 곤충 세포뿐만 아니라 perC6 세포, 또는 사카로마이세스(*Saccharomyces*), 피치아(*Pichia*) 및 스킴조사카로마이세스(*Schizosaccharomyces*)와 같은 진균 세포를 들 수 있다. 특정 구체예에서, 세포주는 CHO-K1, MDCK 또는 HEK293 세포주이다. 적합한 세포는 또한, 예를 들면, 배아 줄기 세포, 유도된 만능 줄기 세포 (iPS 세포), 조혈 줄기 세포, 뉴런 줄기 세포 및 중간엽 줄기 세포와 같은 줄기 세포를 포함할 수 있다.

[0195] 본원에 기술된 DNA-결합 도메인을 포함하는 단백질을 전달하는 방법은, 예를 들어, 미국 특허 제 6,453,242호; 6,503,717호; 6,534,261호; 6,599,692호; 6,607,882호; 6,689,558호; 6,824,978호; 6,933,113호; 6,979,539호; 7,013,219호; 및 7,163,824호 (이들 모두의 개시내용은 그 전문이 본원에 참조로 포함됨)에서 기술된다.

[0196] 본원에 기술된 DNA 결합 도메인 및 이들 DNA 결합 도메인을 포함하는 융합 분자는 또한 하나 이상의 DNA-결합 단백질(들)을 암호화하는 서열을 함유하는 벡터를 사용하여 전달될 수 있다. 추가적으로, 추가적인 핵산 (예컨대, 도너)이 또한 이들 벡터를 통해 전달될 수 있다. 한정하는 것은 아니지만, 플라스미드 벡터, 레트로바이러스 벡터, 렌티바이러스 벡터, 아데노바이러스 벡터, 폭스바이러스 벡터; 헤르페스바이러스 벡터 및 아데노-관련 바이러스 벡터, 등을 포함한 임의의 벡터 시스템이 사용될 수 있다. 또한, 미국 특허 제 6,534,261호; 6,607,882호; 6,824,978호; 6,933,113호; 6,979,539호; 7,013,219호; 및 7,163,824호 (그 전문이 본원에 참조로 포함됨)를 참조한다. 나아가, 이들 벡터 중 어느 것이든지 하나 이상의 DNA-결합 단백질-암호화 서열 및/또는 추가적인 핵산을 필요에 따라 포함할 수 있다는 것이 명백해질 것이다. 그러므로, 본원에 기술된 하나 이상의 DNA-결합 단백질, 및 필요에 따라 추가적인 DNA가 세포에 도입될 때, 그것들은 동일한 벡터로 또는 상이한

백터로 운반될 수 있다. 다중 백터가 사용되는 경우, 각 백터는 하나 또는 다중 DNA-결합 단백질 및 원하는 대로 추가적인 핵산을 암호화하는 서열을 포함할 수 있다.

- [0197] 종래의 바이러스 및 비-바이러스 기반 유전자 전달 방법이 세포 (예컨대, 포유류 세포) 및 표적 조직에서 조작된 DNA-결합 단백질을 암호화하는 핵산을 도입하기 위하여 및 원하는 대로 추가적인 뉴클레오타이드 서열을 동시 도입하기 위하여 사용될 수 있다. 이러한 방법은 또한 시험관내에서 핵산 (예컨대, DNA-결합 단백질 및/또는 도너를 암호화하는 핵산)을 세포에 투여하기 위해 사용될 수 있다. 특정 구체예에서, 핵산은 생체내 또는 생체외 유전자 요법 용도에 대해 투여된다. 비-바이러스 백터 전달 시스템은 DNA 플라스미드, 네이키드 핵산, 및 리포솜 또는 폴록사머와 같은 전달 비히클과 복합체를 형성한 핵산을 포함한다. 바이러스 백터 전달 시스템은 세포에 전달된 후 에피솜에 있거나 게놈에 통합되는 DNA 및 RNA 바이러스를 포함한다. 유전자 요법 과정의 개관에 대해서는, Anderson, *Science* 256:808-813 (1992); Nabel & Felgner, *TIBTECH* 11:211-217 (1993); Mitani & Caskey, *TIBTECH* 11:162-166 (1993); Dillon, *TIBTECH* 11:167-175 (1993); Miller, *Nature* 357:455-460 (1992); Van Brunt, *Biotechnology* 6(10):1149-1154 (1988); Vigne, *Restorative Neurology and Neuroscience* 8:35-36 (1995); Kremer & Perricaudet, *British Medical Bulletin* 51(1):31-44 (1995); Haddada *et al.*, in *Current Topics in Microbiology and Immunology* Doerfler and Bohm (eds.) (1995); 및 Yu *et al.*, 유전자 요법 1:13-26 (1994)을 참조한다.
- [0198] 핵산의 비-바이러스 전달 방법으로 전기천공, 리포펙션, 미세주입, 입자사출(biolistics), 비로솜, 리포솜, 면역리포솜, 다가양이온 또는 지질:핵산 컨주게이트, 네이키드 DNA, mRNA, 인공 비리온, 및 DNA의 작용제-항상 흡수를 들 수 있다. 예컨대, Sonitron 2000 시스템 (Rich-Mar)을 사용하는 초음파천공(Sonoporation)이 또한 핵산의 전달을 위해 사용될 수 있다. 바람직한 구체예에서, 하나 이상의 핵산은 mRNA로서 전달된다. 또한 바람직한 것은 번역 효율 및/또는 mRNA 안정성을 증가시키기 위한 캡핑된 mRNA의 사용이다. 특히 바람직한 것은 ARCA (항-역방향 캡 유사체) 캡 또는 이것의 변이체이다. 미국 특허 제 7,074,596호 및 8,153,773호 (본원에 참조로 포함됨)를 참조한다.
- [0199] 추가적인 예시의 핵산 전달 시스템은 Amaxa Biosystems (Cologne, Germany), Maxcyte, Inc. (Rockville, Maryland), BTX Molecular Delivery Systems (Holliston, MA) 및 Copernicus Therapeutics Inc. (예를 들어 US6008336 참조)에 의해 제공되는 것들을 포함한다. 피로펙션은 예컨대, US 5,049,386, US 4,946,787; 및 US 4,897,355에서 기술되고, 리포펙션 시약은 상업적으로 판매중이다 (예컨대, TransfectamTM, LipofectinTM, 및 LipofectamineTM RNAiMAX). 폴리뉴클레오타이드의 효율적인 수용체-인식 리포펙션에 적합한 다가양이온 및 중성 지질로는 Felgner, WO 91/17424, WO 91/16024의 것들을 들 수 있다. 전달은 세포에 대한 것 (생체외 투여)이거나 표적 조직에 대한 것 (생체내 투여)일 수 있다.
- [0200] 면역지질 복합체와 같은 표적화된 리포솜을 포함한 지질:핵산 복합체의 제조는 기술분야의 숙련된 사람에게 잘 알려져 있다 (예컨대, Crystal, *Science* 270:404-410 (1995); Blaese *et al.*, *Cancer Gene Ther.* 2:291-297 (1995); Behr *et al.*, *Bioconjugate Chem.* 5:382-389 (1994); Remy *et al.*, *Bioconjugate Chem.* 5:647-654 (1994); Gao *et al.*, 유전자 요법 2:710-722 (1995); Ahmad *et al.*, *Cancer Res.* 52:4817-4820 (1992); 미국 특허 제 4,186,183호, 4,217,344호, 4,235,871호, 4,261,975호, 4,485,054호, 4,501,728호, 4,774,085호, 4,837,028호, 및 4,946,787호 참조).
- [0201] 추가적인 전달 방법으로는 EnGeneIC 전달 비히클 (EDV)로 전달될 핵산의 포장을 들 수 있다. 이들 EDV는 항체의 한 팔은 표적 조직에 대한 특이성을 가지며 다른 팔은 EDV에 대한 특이성을 가진 이중특이적 항체를 사용하여 표적 조직에 특이적으로 전달된다. 항체는 EDV를 표적 세포 표면으로 가져간 후 EDV는 세포내이입에 의해 세포안으로 들어간다. 세포에 있게 되면, 내용물이 방출된다 (MacDiarmid *et al* (2009) *Nature Biotechnology* 27(7) p. 643 참조).
- [0202] 조작된 DNA-결합 단백질, 및/또는 원하는 도너 (예컨대, CAR 또는 ACTR)를 암호화하는 핵산의 전달을 위한 RNA 또는 DNA 바이러스 기반 시스템의 사용은 바이러스를 신체의 특정 세포로 표적화하고 바이러스 페이로드를 핵으로 수송하기 위해 고도로 진화된 과정의 장점을 취한다. 바이러스 백터는 환자에게 직접 투여되거나 (생체내) 또는 그것은 시험관내에서 세포를 처리하기 위해 사용될 수 있고 변형된 세포가 환자에게 투여된다 (생체외). 핵산의 전달을 위한 종래의 바이러스 기반 시스템으로는, 한정하는 것은 아니지만, 유전자 전달을 위한 레트로바이러스, 렌티바이러스, 아데노바이러스, 아데노-관련, 백시니아 및 헤르페스 단순 바이러스 백터를 들 수 있다. 숙주 게놈에서의 통합은 자주 삽입된 도입유전자의 장기간 발현을 초래하는, 레트로바이러스, 렌티바이러스, 및 아데노-관련 바이러스 유전자 전달 방법으로 가능하다. 추가적으로, 높은 형질도입 효율이 많

은 상이한 세포 유형 및 표적 조직에서 관찰되었다.

- [0203] 레트로바이러스의 항성은 표적 세포의 잠재적 표적 집단으로 뻗어있는 외래 엔벨로프 단백질을 통합함으로써 변경될 수 있다. 렌티바이러스 벡터는 비분할 세포를 형질도입시키거나 감염시킬 수 있고 전형적으로 높은 바이러스 역가를 유발하는 레트로바이러스 벡터이다. 레트로바이러스 유전자 전달 시스템의 선택은 표적 조직에 좌우된다. 레트로바이러스 벡터는 6 내지 10 kb까지의 외래 서열에 대한 포장 능력을 가진 시스-작용 긴 말단 반복 부로 구성된다. 최소 시스-작용 LTR이면 벡터의 복제 및 포장에 충분하고, 그것은 후에 영구적인 도입유전자 발현을 제공하기 위하여 표적 세포에 치료제를 통합시키는 데 사용된다. 광범위하게 사용되는 레트로바이러스 벡터로는 쥐과 백혈병 바이러스 (MuLV), 긴팔 원숭이 백혈병 바이러스 (GaLV), 유인원 면역결핍 바이러스 (SIV), 인간 면역결핍 바이러스 (HIV), 및 이것들의 조합을 들 수 있다 (예컨대, Buchscher *et al.*, *J. Virol.* 66:2731-2739 (1992); Johann *et al.*, *J. Virol.* 66:1635-1640 (1992); Sommerfelt *et al.*, *Virology* 176:58-59 (1990); Wilson *et al.*, *J. Virol.* 63:2374-2378 (1989); Miller *et al.*, *J. Virol.* 65:2220-2224 (1991); PCT/US94/05700을 참조한다).
- [0204] 일시적인 발현이 바람직한 적용에서, 아데노바이러스 기반 시스템이 사용될 수 있다. 아데노바이러스 기반 벡터는 많은 세포 유형에서 매우 높은 형질도입 효율을 보일 수 있고 세포 분할을 필요로 하지 않는다. 이러한 벡터를 사용하여, 고역가 및 고수준의 발현이 얻어졌다. 이 벡터는 상대적으로 간단한 시스템에서 대량으로 제조될 수 있다. 아데노-관련 바이러스 ("AAV") 벡터는 또한, 예컨대, 시험관내 핵산 및 펩타이드 제조에서, 그리고 생체내 및 생체의 유전자 요법 과정에서, 표적 핵산으로 세포를 형질도입시키기 위해 사용된다 (예컨대, West *et al.*, *Virology* 160:38-47 (1987); 미국 특허 제 4,797,368호; WO 93/24641; Kotin, *Human 유전자 요법* 5:793-801 (1994); Muzyczka, *J. Clin. Invest.* 94:1351 (1994) 참조). 재조합 AAV 벡터의 구성은 많은 출판물, 이를테면 미국 특허 제 5,173,414호; Tratschin *et al.*, *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, *et al.*, *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS USA* 81:6466-6470 (1984); 및 Samulski *et al.*, *J. Virol.* 63:03822-3828 (1989)에서 기술된다.
- [0205] 적어도 6가지 바이러스 벡터 접근법이 현재 형질도입제(transducing agent)를 생성하기 위하여 헬퍼 세포주에 삽입된 유전자에 의한 결핍 벡터의 보완을 포함하는 접근법을 활용하는 임상 실험에서 유전자 전달에 활용 가능하다.
- [0206] pLASN 및 MFG-S는 임상 실험에서 사용되어 온 레트로바이러스의 예이다 (Dunbar *et al.*, *Blood* 85:3048-305 (1995); Kohn *et al.*, *Nat. Med.* 1:1017-102 (1995); Malech *et al.*, *PNAS USA* 94:22 12133-12138 (1997)). PA317/pLASN은 유전자 요법 실험에서 사용된 첫 번째 치료용 벡터였다 (Blaese *et al.*, *Science* 270:475-480 (1995)). 50% 이상의 형질도입 효율이 MFG-S 포장된 벡터에 대해 관찰되었다. (Ellem *et al.*, *Immunol Immunother.* 44(1):10-20 (1997); Dranoff *et al.*, *Hum. Gene Ther.* 1:111-2 (1997)).
- [0207] 재조합 아데노-관련 바이러스 벡터 (rAAV)는 결핍성 및 비병원성 파코바이러스 아데노-관련 유형 2 바이러스를 기반으로 한 유망한 대체 유전자 전달 시스템이다. 모든 벡터는 도입유전자 발현 카세트 옆에 있는 AAV 145 bp 반전된 말단 반복부만을 보유한 플라스미드로부터 유래된다. 형질도입된 세포의 계승으로의 통합으로 인한 효율적인 유전자 전달 및 안정적인 도입유전자 전달이 이 벡터 시스템에 대한 핵심 특징이다. (Wagner *et al.*, *Lancet* 351:9117 1702-3 (1998), Kearns *et al.*, *Gene Ther.* 9:748-55 (1996)). AAV1, AAV3, AAV4, AAV5, AAV6, AAV8, AAV8.2, AAV9 및 AAVrh10을 포함한 다른 AAV 혈청형 및 AAV2/8, AAV2/5 및 AAV2/6과 같은 위험 AAV도 또한 본 발명에 따라 사용될 수 있다.
- [0208] 복제-결핍 재조합 아데노바이러스 벡터 (Ad)는 고역가로 제조될 수 있고 많은 상이한 세포 유형을 쉽게 감염시킬 수 있다. 대부분의 아데노바이러스 벡터는 도입유전자가 Ad E1a, E1b, 및/또는 E3 유전자를 대체하고; 계속해서 복제 결핍 벡터가 트랜스로 고갈된 유전자 기능을 공급하는 인간 293 세포에서 증식되도록 조작된다. Ad 벡터는 간, 신장 및 근육에서 발견되는 것과 같은 비분할, 분화된 세포를 포함하여, 생체내 다수의 조직 유형을 형질도입시킬 수 있다. 종래의 Ad 벡터는 큰 수송 용량을 가진다. 임상 실험에서 Ad 벡터의 사용의 예는 근육내 주사를 이용한 항종양 면역화를 위한 폴리뉴클레오타이드 요법을 포함하였다 (Sternan *et al.*, *Hum. Gene Ther.* 7:1083-9 (1998)). 임상 실험에서 유전자 전달을 위한 아데노바이러스 벡터의 사용의 추가적인 예는 Rosenecker *et al.*, *Infection* 24:1 5-10 (1996); Sternan *et al.*, *Hum. Gene Ther.* 9:7 1083-1089 (1998); Welsh *et al.*, *Hum. Gene Ther.* 2:205-18 (1995); Alvarez *et al.*, *Hum. Gene Ther.* 5:597-613 (1997); Topf *et al.*, *Gene Ther.* 5:507-513 (1998); Sternan *et al.*, *Hum. Gene Ther.* 7:1083-1089 (1998)을 포함한다.
- [0209] 포장 세포는 숙주 세포를 감염시킬 수 있는 바이러스 입자를 형성하기 위해 사용된다. 이러한 세포로는 아데노

바이러스를 포장하는 293 세포, 레트로바이러스를 포장하는 ψ 2 세포 또는 PA317 세포를 들 수 있다. 유전자 요법에 사용된 바이러스 벡터는 보통 핵산 벡터를 바이러스 입자에 포장하는 생성자 셀라인에 의해 생성된다. 벡터는 전형적으로 숙주에서 포장 및 후속되는 통합 (적용되는 경우)에 필요한 최소 바이러스 서열을 함유하며, 다른 바이러스 서열은 발현될 단백질을 암호화하는 발현 카세트에 의해 대체된다. 누락 바이러스 기능은 포장 세포주에 의해 트랜스로 공급된다. 예를 들어, 유전자 요법에 사용된 AAV 벡터는 전형적으로 숙주 게놈으로의 포장 및 통합에 필요한 AAV 게놈으로부터의 반전된 말단 반복부 (ITR) 서열을 가진다. 바이러스 DNA는 다른 AAV 유전자, 즉 rep 및 cap을 암호화하지만, ITR 서열이 없는 헬퍼 플라스미드를 함유하는 세포주에서 포장된다. 세포주는 또한 헬퍼로서 아데노바이러스로 감염된다. 헬퍼 바이러스는 AAV 벡터의 복제 및 헬퍼 플라스미드로부터의 AAV 유전자의 발현을 촉진한다. 헬퍼 플라스미드는 ITR 서열의 결핍으로 인해 상당한 양으로 포장되지 않는다. 아데노바이러스로의 오염은, 예컨대, 아데노바이러스가 AAV보다 더 민감한 열 처리에 의해 감소될 수 있다.

[0210] 많은 유전자 요법 적용에서, 유전자 요법 벡터는 특정 조직 유형에 고도의 특이성으로 전달되는 것이 바람직하다. 따라서, 바이러스 벡터는 바이러스의 외면 상의 바이러스 코트 단백질과의 융합으로서 리간드를 발현함으로써 주어진 세포 유형에 대한 특이성을 갖도록 변형될 수 있다. 리간드는 관심 있는 세포 유형에 존재하는 것으로 알려져 있는 수용체에 대한 친화성을 갖도록 선택된다. 예를 들어, Han 등 (*Proc. Natl. Acad. Sci. USA* 92:9747-9751 (1995))은, 몰로니(Moloney) 쥐과 백혈병 바이러스가 gp70에 융합된 인간 헤레굴린을 발현하도록 변형될 수 있고, 아데노바이러스는 인간 상피 성장 인자 수용체를 발현하는 특정 인간 유방암 세포를 감염시키는 것을 보고하였다. 이 원칙은 표적 세포가 수용체를 발현하고 바이러스가 세포-표면 수용체에 대한 리간드를 포함하는 융합 분자를 발현하는 다른 바이러스-표적 세포 쌍으로 확장될 수 있다. 예를 들어, 섬유상 파지는 실제로 임의의 선택된 세포 수용체에 대한 특정 결합 친화성을 갖는 항체 단편 (예컨대, F_{ab} 또는 F_v)을 나타내도록 조작될 수 있다. 비록 상기 설명이 주로 바이러스 벡터에 적용되긴 하지만, 동일한 원리가 비바이러스 벡터에 적용될 수 있다. 이러한 벡터는 특정 표적 세포에 의한 흡수를 선호하는 특정 흡수 서열을 함유하도록 조작될 수 있다.

[0211] CRISPR/Cas 시스템에 대한 전달 방법은 상기 기술된 방법들을 포함할 수 있다. 예를 들어, 동물 모델에서, mRNA 또는 재조합 Cas 단백질을 암호화하는 시험관내 전사된 Cas는 유리 바늘을 사용하여 게놈-편집된 동물로 1-세포기 배아로 직접 주입될 수 있다. 시험관내에서 세포에서 Cas 및 가이드 RNA를 발현하기 위하여, 전형적으로 그것들을 암호화하는 플라스미드가 리포펙션 또는 전기천공을 통해 세포 안으로 트랜스펙션된다. 또한, 재조합 Cas 단백질이, Cas-가이드 RNA 리보핵단백질이 관심 있는 세포에 의해 차지되는 시험관내 전사된 가이드 RNA와 복합체를 형성할 수 있다 (Kim *et al* (2014) *Genome Res* 24(6):1012). 치료 목적으로, Cas 및 가이드 RNA는 바이러스 및 비바이러스 기법의 조합에 의해 전달될 수 있다. 예를 들어, Cas를 암호화하는 mRNA는 나노입자 전달을 통해 전달될 수 있는 한편, 가이드 RNA 및 임의의 원하는 도입유전자 또는 수복 주형은 AAV를 통해 전달된다 (Yin *et al* (2016) *Nat Biotechnol* 34(3) p. 328).

[0212] 유전자 요법 벡터는 하기에서 기술되는 것과 같이, 전형적으로 전신성 투여 (예컨대, 정맥내, 복강내, 근육내, 피하, 또는 두개내 주입)에 의해 또는 국소 적용에 의해 개별 환자로의 투여에 의해 생체내에서 전달될 수 있다. 대안으로, 벡터는 생체외에서 세포, 예컨대, 개별 환자 (예컨대, 림프구, 골수 흡인물, 조직 생검)로부터 외식된 세포 또는 보편적인 도너 조혈 줄기 세포에 전달될 수 있고, 이어서, 보통 벡터를 통합한 세포에 대한 선택 후에 세포가 환자에게 재이식될 수 있다.

[0213] 진단, 연구, 이식을 위한 또는 유전자 요법을 위한 생체의 세포 트랜스펙션 (예컨대, 트랜스펙션된 세포의 숙주 유기체로의 재주입을 통한)은 기술분야의 숙련된 지식을 가진 사람들에게 잘 알려져 있다. 바람직한 구체예에서, 세포는 대상 유기체로부터 분리되고, DNA-결합 단백질 핵산 (유전자 또는 cDNA)으로 트랜스펙션되며, 대상 유기체 (예컨대, 환자)로 다시 재주입된다. 생체의 트랜스펙션에 적합한 다양한 세포 유형이 기술분야의 숙련된 지식을 가진 사람들에게 잘 알려져 있다 (예컨대, 환자로부터 세포를 분리하고 배양하는 방법의 논의에 대해 Freshney *et al.*, *Culture of Animal Cells, A Manual of Basic Technique* (3rd ed. 1994) 및 여기에 인용된 참고문헌 참조)

[0214] 한 구체예에서, 줄기 세포는 세포 트랜스펙션 및 유전자 요법을 위한 생체의 과정에 사용된다. 줄기 세포를 사용하는 것에 대한 장점은 그것이 시험관내에서 다른 세포 유형으로 분화되거나, 또는 그것이 골수에서 생착하게 될 포유류 (예컨대 세포의 도너)로 도입될 수 있다는 것이다. CD34+ 세포를 시험관내에서 GM-CSF, IFN- γ 및 TNF- α 와 같은 사이토카인을 사용하여 임상적으로 중요한 면역 세포 유형으로 분화시키는 방법은 알려져 있다 (Inaba *et al.*, *J. Exp. Med.* 176:1693-1702 (1992) 참조).

- [0215] 줄기 세포는 공지된 방법을 사용하여 형질도입 및 분화를 위해 분리된다. 예를 들어, 줄기 세포는 골수 세포를 원치 않는 세포, 예컨대, CD4+ 및 CD8+ (T 세포), CD45+ (panB 세포), GR-1 (과립구), 및 Iad (분화 항원 제시 세포)에 결합하는 항체로 패닝함으로써 골수로부터 분리된다 (Inaba *et al.*, *J. Exp. Med.* 176:1693-1702 (1992) 참조).
- [0216] 변형된 줄기 세포는 또한 일부 구체예에서 사용될 수 있다. 예를 들어, 세포자멸사에 대해 내성으로 만들어진 뉴런 줄기 세포는, 줄기 세포가 또한 발명의 ZFP TF를 함유한 치료 조성물로서 사용될 수 있다. 세포자멸사에 대한 내성은, 예를 들어, 줄기 세포에서 BAX- 또는 BAK-특이적 ZFN (미국 특허 제 8,597,912호 참조), 또는 예를 들어 다시 카스파제-6 특이적 ZFN을 사용하여 카스파제에서 파괴되는 것을 사용하여 BAX 및/또는 BAK를 녹아웃시킴으로써 일어날 수 있다.
- [0217] 치료용 DNA-결합 단백질 (또는 이들 단백질을 암호화하는 핵산)을 함유하는 벡터 (예컨대, 레트로바이러스, 아데노바이러스, 리포솜, 등)는 또한 생체내에서 세포의 형질도입을 위해 유기체에 직접 투여될 수 있다. 대안으로, 네이키드 DNA가 투여될 수 있다. 투여는, 한정하는 것은 아니지만, 주사, 주입, 국소 적용 및 전기천공을 포함한, 혈액 또는 조직 세포와의 궁극적인 접촉으로 분자를 도입하기 위해 일반적으로 사용된 임의의 경로에 의해 일어난다. 이러한 핵산을 투여하는 적합한 방법은 기술분야에서 이용 가능하고 숙련된 지식을 가진 사람들에게 잘 알려져 있지만, 한 가지 이상의 경로가 특정 조성물을 투여하기 위해 사용될 수 있고, 특정 경로는 자주 다른 경로보다 더 즉각적이고 더 효과적인 반응을 제공할 수 있다.
- [0218] 조혈 줄기 세포로 DNA를 도입하기 위한 방법은, 예를 들어, 미국 특허 제 5,928,638호에서 개시된다. 도입유전자의 조혈 줄기 세포, 예컨대, CD34+ 세포로의 도입에 유용한 벡터로 아데노바이러스 유형 35를 들 수 있다.
- [0219] 도입유전자의 면역 세포 (예컨대, T-세포)로의 도입에 적합한 벡터로 비통합 렌티바이러스 벡터를 들 수 있다. 예를 들어, Ory *et al.* (1996) *Proc. Natl. Acad. Sci. USA* 93:11382-11388; Dull *et al.* (1998) *J. Virol.* 72:8463-8471; Zuffery *et al.* (1998) *J. Virol.* 72:9873-9880; Follenzi *et al.* (2000) *Nature Genetics* 25:217-222를 참조한다.
- [0220] 제약학적으로 허용되는 담체는 부분적으로 투여되는 특정 조성물에 의해, 또한 조성물을 투여하기 위해 사용된 특정 방법에 의해 결정된다. 따라서, 하기에 기술되는 것과 같이 이용 가능한 제약학적 조성물의 광범위한 적합한 제제가 있다 (예컨대, Remington's Pharmaceutical Sciences, 17th ed., 1989 참조).
- [0221] 상에서 주지된 것과 같이, 개시된 방법 및 조성물은 한정하는 것은 아니지만, 전핵 세포, 진균 세포, 고세균 세포, 식물 세포, 곤충 세포, 동물 세포, 척추동물 세포, 포유류 세포 및 임의의 유형의 T-세포 및 줄기 세포를 포함한 인간 세포를 포함한 임의의 유형의 세포에서 사용될 수 있다. 단백질 발현에 적합한 세포주는 기술분야의 숙련된 지식을 가진 사람들에게 알려져 있고, 한정하는 것은 아니지만, COS, CHO (예컨대, CHO-S, CHO-K1, CHO-DG44, CHO-DUXB11), VERO, MDCK, WI38, V79, B14AF28-G3, BHK, HaK, NSO, SP2/0-Ag14, HeLa, HEK293 (예컨대, HEK293-F, HEK293-H, HEK293-T), perC6, 스포도프테라 푸지페르다(*Spodoptera fugiperda*, Sf)와 같은 곤충 세포, 및 사카로마이세스, 피치아 및 스킴조사카로마이세스와 같은 진균 세포를 포함한다. 이들 세포주의 자손, 변이체 및 유도체 또한 사용될 수 있다.
- [0222] **적용**
- [0223] 질환의 치료 및 예방에서 조작된 뉴클레아제의 사용은 앞으로 몇 년 동안 의학에서 가장 중요한 발전 중 하나이다. 본원에 기술된 방법 및 조성물은 원하는 표적 부위가 절단의 주요 장소가 될 것임을 보장하기 위해 이들 신규한 도구의 특이성을 증가시키는 작용을 한다. 표적 외 절단을 최소화하거나 제거하는 것이, 모든 시험관내, 생체내 및 생체의 적용에 대해 이 기술의 전체 잠재력을 실현시키기 위해 필요할 것이다.
- [0224] 예시의 유전자 질환으로는, 한정하는 것은 아니지만, 연골 무형성증, 색도증, 산성 말타제 결핍증, 아데노신 탈아민효소 결핍증 (OMIM No. 102700), 부신 백질 이영양증, 아이카르디 증후군, 알파-1 항트립신 결핍증, 알파-지중해 빈혈, 안드로젠 무감각 증후군, 아퍼트 증후군, 부정맥 유발성 우심실, 이형성증, 운동 실조증, 바르스 증후군, 베타-지중해 혈증, 청색 고무 수포 모반 증후군, 카나반병, 만성 육아종성 질환 (CGD), 크리 뒤샤 증후군, 낭포성 섬유증, 진피병, 외배엽 이형성증, 판코니 빈혈, 섬유 이형성증 진행성, 연약한 X 증후군, 갈락토세미, 고셔병, 전신성 신경절증 (예컨대, GM1), 혈색소 침착증, 베타-글로빈의 제6 코돈의 헤모글로빈 C 돌연변이 (HbC), 혈우병, 헌팅턴병, 혈러 증후군, 인산 저하증, 클라인플레터 증후군, 크라베스병, 랑게르-지디온 증후군, 백혈구 유착 결핍증 (LAD, OMIM 번호 116920), 백혈구 이영양증, 긴 QT 증후군, 마르판 증후군, 피비우스 증후군, 뮤코다당질축적증 (MPS), 손발톱 슬개 증후군, 신성 요붕증, 신경 섬유종증, 네이만-픽병, 골형성

불완전증, 페닐케톤뇨증 (PKU), 포르피린증, 프라더-윌리 증후군, 조로증, 프로테우스 증후군, 망막모세포종, 레트 증후군, 루빈스타인-타이비 증후군, 산필리포 증후군, 중증 복합 면역결핍증 (SCID), 슈바크만 증후군, 겸상 적혈구 질환 (겸상 적혈구 빈혈), 스미스-마제니스 증후군, 스티클러 증후군, 데이-삭스병, 혈소판감소 요골 무형성(Thrombocytopenia Absent Radius, TAR) 증후군, 트리처 콜린스 증후군, 삼염색체 성, 결절성 경화증, 터너 증후군, 요소 주기 장애, 폰 히펠-란다우 질병, 바르텐부르크 증후군, 윌리엄스 증후군, 윌슨병, 비스코트-올드리치 증후군, X-결합 림프증식성 증후군 (XLP, OMIM No. 308240)을 들 수 있다.

[0225] 표적화된 DNA 절단 및/또는 상동성 재조합에 의해 치료될 수 있는 추가적인 예시의 질환으로는 후천성 면역결핍 증, 리소좀 축적 질환 (예컨대, 고셔병, GM1, 파브리병 및 데이-삭스병), 뮤코다당질축적증 (예컨대, MPSII (헌터병), MPSI (헐러병), 혈색소병증 (예컨대, 겸상 적혈구병, HbC, α -지중해빈혈, β -지중해빈혈) 및 혈우병을 들 수 있다. 예컨대, 미국 특허 제 9,877,988호 및 9,956,247호 참조. 특히, 글루코세레브로시다제 (GBA)가 고셔병에서 결핍이고, 알파-갈락토시다제 (GLA)가 파브리병에서 결핍이며, 이두로네이트-2-설패타제 결핍증 (IDS)은 MPS II (헌터병)에서 결핍이고, 알파-L 이두로니다제 결핍증 (IDUA)은 MPS I (헐러병)에서 결핍이며, 스팅고미엘린 포스포다이에스테라제 1 결핍증 (SMPD1)은 니에만-픽병에서 결핍이다. 그러므로, 이들 질환에서 하나 이상의 단백질 부족 또는 결핍을 발현하는 도너가 이들 질환에 대한 치료 및/또는 예방을 제공하기 위하여 본원에 기술된 뉴클레아제를 사용하여 도입될 수 있다.

[0226] 이러한 방법은 또한 유전자 질환을 치료하기 위하여 숙주에서 감염 (바이러스 또는 박테리아)의 치료를 허용한다 (예컨대, 바이러스 또는 박테리아 수용체의 발현을 차단함으로써, 그로써 숙주 유기체에서 감염 및/또는 확산이 방지된다).

[0227] 감염 또는 통합된 바이러스 게놈의 표적화된 절단은 숙주에서 바이러스 감염을 치료하기 위해 사용될 수 있다. 추가적으로, 바이러스에 대한 수용체를 암호화하는 유전자의 표적화된 절단은 이러한 수용체의 발현을 차단하기 위해 사용될 수 있고, 그로써 숙주 유기체에서 감염 및/또는 확산이 방지될 수 있다. 바이러스 수용체 (예컨대, HIV의 경우 CCR5 및 CXCR4 수용체)를 암호화하는 유전자의 표적화된 돌연변이는 수용체가 바이러스에 결합할 수 없게 만들기 위하여 사용됨으로써, 새로운 감염을 방지하고 기존 감염의 확산을 차단할 수 있다. 미국 특허출원 공개 번호 2008/015996을 참조한다. 표적화될 수 있는 바이러스 또는 바이러스 수용체의 비제한적인 예로는 헤르페스 단순 바이러스 (HSV), 예컨대, HSV-1 및 HSV-2, 수두 대상포진 바이러스 (VZV), 엡스타인-바 바이러스 (EBV) 및 사이토메갈로바이러스 (CMV), HHV6 및 HHV7을 들 수 있다. 바이러스의 간염 패밀리는 A형 간염 바이러스 (HAV), B형 간염 바이러스 (HBV), C형 간염 바이러스 (HCV), 델타 간염 바이러스 (HDV), E형 간염 E 바이러스 (HEV) 및 G형 간염 바이러스 (HGV)를 포함한다. 다른 바이러스 또는 그것들의 수용체가 표적화될 수 있는데, 이를테면, 한정하는 것은 아니지만, 피코르나비리다에 (Picornaviridae) (예컨대, 폴리오바이러스, 등); 칼리시비리다에 (Caliciviridae); 토가비리다에 (Togaviridae) (예컨대, 루벨라 바이러스, 탕기 바이러스, 등); 플라비비리다에 (Flaviviridae); 코로나비리다에 (Coronaviridae); 레오비리다에 (Reoviridae); 비르나비리다에 (Birnaviridae); 라보도비리다에 (Rhabdoviridae) (예컨대, 광견병 바이러스, 등); 필로비리다에 (Filoviridae); 파라믹소비리다에 (Paramyxoviridae) (예컨대, 유행성 이하선염 바이러스, 홍역 바이러스, 호흡기 세포융합 바이러스, 등); 오르토믹소비리다에 (Orthomyxoviridae) (예컨대, 인플루엔자 바이러스 타입 A, B 및 C, 등); 분야비리다에 (Bunyaviridae); 아레나비리다에 (Arenaviridae); 레트로비리다에 (Retroviridae); 렌티 바이러스 (예컨대, HTLV-I; HTLV-II; HIV-1 (또한 HTLV-III, LAV, ARV, hTLR, 등으로도 알려짐) HIV-II); 유인원 면역결핍증 바이러스 (SIV), 인간 유두종바이러스 (HPV), 인플루엔자 바이러스 및 및 진드기 매개 뇌염 바이러스를 들 수 있다. 이들 및 다른 바이러스의 설명에 대해서는, 예컨대, Virology, 3rd Edition (W. K. Joklik ed. 1988); Fundamental Virology, 2nd Edition (B. N. Fields 및 D. M. Knipe, eds. 1991)을 참조한다. HIV에 대한 수용체는, 예를 들어, CCR-5 및 CXCR-4를 들 수 있다.

[0228] 그러므로, 본원에 기술된 헤테로다имер 절단 도메인 변이체는 유전자 변형 적용에서 ZFN 특이성을 개선하기 위해 광범위한 이용성을 제공한다. 이들 변이 절단 도메인은 임의의 기존의 ZFN에 부위 지정 돌연변이생성 또는 임의의 ZFN 다이머의 생체내 특이성을 개선하기 위한 하위클로닝에 의해 쉽게 통합될 수 있다.

[0229] 상기에서 주지된 것과 같이, 본원에 기술된 조성물 및 방법은 유전자 변형, 유전자 수정, 및 유전자 파괴를 위해 사용될 수 있다. 유전자 변형의 비제한적인 예로는 상동성 지정 수복 (HDR)-기반 표적화된 통합; HDR-기반 유전자 수정; HDR-기반 유전자 변형; HDR-기반 유전자 파괴; NHEJ-기반 유전자 파괴 및/또는 HDR, NHEJ, 및/또는 단일 가닥 어닐링 (SSA)의 조합을 들 수 있다. 단일 가닥 어닐링 (SSA)은 2개의 상보하는 영역을 노출시키기 위하여 5'-3' 엑소뉴클레아제에 의한 DSB의 절제에 의해 동일한 방향으로 일어나는 2개의 반복된 서열 사이의 이중 가닥 절단의 수복을 나타낸다. 그런 후 2개의 직접 반복부를 암호화하는 단일 가닥은 서로 어닐링하고, 어

닐링된 중간체는 단일 가닥 꼬리 (임의의 서열에 어닐링되지 않은 단일 가닥 DNA의 부분)가 소화되고, 그 갭이 DNA 중합효소에 의해 채워지며, DNA 단부가 다시 연결되도록 처리될 수 있다. 이것은 직접적인 반복부 사이에 위치한 서열의 결실을 초래한다.

[0230] 본원에 기술된, 절단 도메인 (예컨대, ZFN, TALEN, CRISPR/Cas 시스템)을 포함하는 조성물 및 방법은 또한 다양한 유전자 질환 및/또는 감염성 질환의 치료에 사용될 수 있다.

[0231] 조성물 및 방법은 또한, 한정하는 것은 아니지만 다음을 포함하는 줄기 세포 기반 치료법에 적용될 수 있다: 짧은 패치 유전자 변환 또는 단일유전자 유전자 요법을 위한 표적화된 통합에 의한 체세포성 세포 돌연변이의 수정; 우세한 네거티브 대립유전자의 파괴; 병원체의 세포로의 유입 또는 증식성 감염에 필요한 유전자의 파괴; 예를 들어, 유전자 활성을 기능성 조직의 분화 또는 형성을 촉진하도록 변형시킴으로써 증강된 조직 조작; 및/또는 기능성 조직의 분화 또는 형성을 촉진하는 유전자 활성의 파괴; 예를 들어, 줄기 세포가 특정 계통 경로로 분화하는 것을 촉진하기 위하여 분화를 차단하는 유전자를 파괴함으로써 분화 차단 또는 유도, 줄기 세포 분화를 자극할 수 있는 유전자 또는 siRNA 발현 카세트의 표적화된 삽입, 줄기 세포 분화를 차단할 수 있고 만능성의 더 좋은 팽창 및 유지에 허용하는 유전자 또는 siRNA 발현 카세트의 표적화된 삽입, 및/또는 용이한 마커가 줄기 세포의 분화 상태 및 매질, 사이토카인, 성장 조건, 유전자의 발현, siRNA, shRNA 또는 miRNA 분자의 발현, 세포 표면 마커에 대한 항체에 대한 노출에서 어떻게 변화하는지, 또는 약물이 이런 상태를 변경시키는 것을 기록하는 것을 허용할 만능성 또는 분화 상태의 마커인 내인성 유전자와 한 프레임의 리포터 유전자의 표적화된 삽입; 체세포성 세포 핵 전달, 예를 들어, 환자 자신의 체세포성 세포가 분리되고, 적절한 방식으로 의도된 표적이 변형되고, 세포 클론이 생성되며 (그리고 품질이 게놈 안전성을 보장하기 위하여 제어되며), 이들 세포로부터 핵이 분리되고, 직접 주사될 수 있거나 환자에게 생착되기 전에 분화될 수 있는 환자-특이적 hES 세포를 생성하기 위하여 수정되지 않은 난에 전달됨으로써, 조직 거부반응이 감소되거나 제거될 수 있는 방식; (예컨대, 감소된 또는 모두 폐지된 면역학적 동일성의 세포를 생성하기 위해) MHC 수용체를 녹아웃시킴에 의한 보편적 줄기 세포. 이런 과정을 위한 세포 유형으로는, 한정하는 것은 아니지만, T-세포, B 세포, 조혈 줄기 세포, 및 배아 줄기 세포를 들 수 있다. 추가적으로, 또한 환자 자신의 체세포성 세포로부터 생성된 유도된 만능 줄기 세포 (iPSC)가 사용될 수 있다. 그러므로, 이들 줄기 세포 또는 이것의 유도체 (분화된 세포 유형 또는 조직)는 그것의 기원 또는 조직적합성과 관계 없이 임의의 사람에게 잠재적으로 생착될 수 있을 것이다.

[0232] 조성물 및 방법은 또한 체세포성 세포 치료법에 사용될 수 있음으로써, 그것의 생물학적 특성을 향상시키도록 변형된 세포의 스톱의 생성이 허용된다. 이러한 세포는 세포의 도너 공급원 및 수령체에 대한 그것의 조직적합성과 관계없이, 다양한 환자에 주입될 수 있다.

[0233] 치료적 적용에 더불어, 본원에 기술된 변이체에 의해 제공된 증가된 특이성은 조작된 뉴클레아제에 사용될 때 작물 조작, 세포주 조작 및 질환 모델의 구성에 사용될 수 있다. 절대 헤테로다이며 절단 절반 도메인은 뉴클레아제 특성을 개선하기 위한 직접적인 수단을 제공한다.

[0234] 기술된 조작된 절단 절반 도메인은 또한 개체 영역을 결실시키기 위해 또는 2개의 특이적 유전자좌를 한번에 변경시키기 위하여 다중 표적에서 동시 절단을 필요로 하는 유전자 변형 프로토콜에 사용될 수 있다. 두 표적에서의 절단은 잠재적으로 10개의 상이한 활성 ZFN TALEN 조합을 생성할 수 있는, 4개의 ZFN 또는 TALEN의 세포 발현을 필요로 할 것이다. 이러한 적용을 위해, 야생형 뉴클레아제 도메인에 대한 이들 신규한 변이체의 치환은 원하지 않는 조합의 활성을 제거하고 표적 외 절단의 기회를 감소시킬 것이다. 만약 특정한 원하는 DNA 표적에서의 절단이 뉴클레아제 쌍 A+B의 활성을 필요로 하고, 제2의 원하는 DNA 표적에서의 동시 절단이 뉴클레아제 쌍 X+Y의 활성을 필요로 한다면, 본원에 기술된 돌연변이의 사용은 A의 A와의 쌍 형성, A와 X와의 쌍 형성, A와 Y와의 쌍 형성 등을 방지할 수 있다. 그러므로, 이들 *FokI* 돌연변이는 "부조리한" 쌍 형성의 결과로서 비특이적 절단 활성을 감소시키고 뉴클레아제의 보다 효율적인 병렬상동성 돌연변이 쌍의 생성을 허용한다 (본 건과 공동 소유의 특허 미국 특허출원 공개 번호 20080131962 및 20090305346 참조).

[0235] 실시예

[0236] 실시예 1: ZFN의 제조

[0237] ZFN 표적화된 to 인간 알부민 유전자에 대해 표적화된 ZFN을 설계하여 본질적으로 Urnov *et al.* (2005) *Nature* 435(7042):646-651; Perez *et al* (2008) *Nature Biotechnology* 26(7):808-816, 및 미국 특허 제 9,394,545호에서 기술된 것과 같이 플라스미드 벡터에 통합하였다.

[0238] 실시예 2: 알부민-특이적 ZFN의 최적화

[0239] 좌측 손 ZFN 파트너 (SBS47171-FLAG, 표 1 참조)의 결합 부위는 20%의 인간에서 SNP를 포함한다 (도 1 참조). 야생형 서열에서, 서열은 AT 염기쌍 (타원형으로 표시됨)을 포함하는 한편, SNP를 포함하는 서열에는, 이 위치에 GC 염기쌍이 있다 (서열 위 직사각형으로 표시됨). 야생형 및 SNP 알부민 서열에 대해 이형접합성인 인간 간 세포에서, 47171-FLAG/47898-FLAG 쌍은 야생형 서열에 대해 3-4배의 선호도를 가진다 (도 2 참조). 야생형 알부민 서열 및 동등한 활성을 가진 SNP-함유 서열을 절단하는 것으로 나타난 제2 좌측 손 파트너 (42875)를 확인하였지만, 42875/47898 쌍 또한 SMCHD1 표적 외 부위에서 일부 절단 활성을 나타냈다.

[0240] 그러므로, ZFP 골격 내에서 아미노산을 접촉하는 포스페이트에 대해 변형이 이루어져 있는 추가적인 후보 ZFN으로 연구를 수행하였다. 사용한 단백질을 하기 표 1에 나타낸다.

표 1

알부민-특이적 ZFN 설계

[0241]

ZFN 명칭 표적 서열	F1	F2	F3	F4	F5	F6	도메인 링커	기타
SBS47171-FLAG ttTGGGATAGTTAT GAAttcaatcttca (SEQ ID NO:12)	QSGNLSR(SE Q ID NO:14)	LKQNLCM (SEQ ID NO:15)	WADNLQN(SE Q ID NO:16)	TSGNLTR (SEQ ID NO:17)	RQSHLCL (SEQ ID NO:18)	N/A	5,6	N-말단 3x FLAG
SBS42875-FLAG ttTGGGATAGTTAT GAAttcaatcttca (SEQ ID NO:12)	QSGNLSR(SE Q ID NO:19)	LKQNLCM(SEQ ID NO:15)	WQSNLQN(SE Q ID NO:20)	TSGNLTR(SE Q ID NO:17)	RRSHLTS(SEQ ID NO:21)	N/A	5,6	N-말단 3x FLAG
42877-FLAG ttTGGGATAGTTAT GAAttcaatcttca (SEQ ID NO:12)	QSGNLSR(SE Q ID NO:19)	LKQNLCM(SEQ ID NO:15)	LITTLRN(SE Q ID NO:22)	TSGNLTR(SE Q ID NO:17)	RQSHLCL(SEQ ID NO:18)	N/A	5,6	N-말단 3x FLAG
42908-FLAG ttTGGGATAGTTAT GAAttcaatcttca (SEQ ID NO:12)	QSGNLSR(SE Q ID NO:19)	LKQNLCM(SEQ ID NO:15)	WASNLQN(SE Q ID NO:23)	TSGNLTR(SE Q ID NO:17)	RQSHLCL(SEQ ID NO:18)	N/A	5,6	N-말단 3x FLAG
47898-FLAG ccTATCCATTGCAC TATGCTttatttaa (SEQ ID NO:13)	TPQLLDR(SE Q ID NO:24)	LKHNLTLT(SEQ ID NO:25)	DQSNLNA(SE Q ID NO:26)	RNFSLTM(SE Q ID NO:27)	LRHDLDR(SEQ ID NO:28)	HRSNLNK(SE Q ID NO:29)	5,6	N-말단 3x FLAG
47874-FLAG ccTATCCATTGCAC TATGCTttatttaa (SEQ ID NO:13)	QSSDLSR(SE Q ID NO:30)	LKHNLTLT(SEQ ID NO:25)	DQSNLRA(SE Q ID NO:31)	RNFSLTM(SE Q ID NO:27)	LRHDLER(SEQ ID NO:32)	HRSNLNK(SE Q ID NO:29)	5,6	N-말단 3x FLAG
47931-FLAG ccTATCCATTGCAC TATGCTttatttaa (SEQ ID NO:13)	TPQLLDR(SE Q ID NO:24)	LKWNLRT(SEQ ID NO:33)	DQSNLNA(SE Q ID NO:26)	RNFSLTM(SE Q ID NO:27)	LRHDLDR(SEQ ID NO:28)	HRSNLNK(SE Q ID NO:29)	5,6	N-말단 3x FLAG

[0242] 다음으로, 상기 열거한 ZFN을 ZFP와 DNA 포스페이트 골격 사이의 임의의 잠재적인 비특이적 접촉을 감소시키기 위하여 그것의 ZFP 골격에 변경을 포함하도록 변형시켰다 (미국 특허출원 공개 번호 US-2018-0087072-A1 참조). 하기 표 2A 및 2B에서, 예시의 ZFP 골격 변화를 표 1로부터의 모 ZFN의 앞부분 아래에 각각의 변이체에 대해 배정된 새로운 SBS 특유한 숫자 식별자와 함께 나타낸다.

표 2

포스페이트 접착 변이체 ZFN

표 2A: 좌측 파트너

m5Q 변이체					47171-FLAG	42875-FLAG	42877-FLAG	42908-FLAG
F1	F2	F3	F4	F5	SBS#	SBS#	SBS#	SBS#
					70949	70957	70965	70973
Q					70950	70958	70966	70974
		Q			70951	70959	70967	70975
			Q		70952	70960	70968	70976
Q	Q				70953	70961	70969	70977
Q			Q		70954	70962	70970	70978
		Q	Q		70955	70963	70971	70979
Q		Q	Q		70956	70964	70972	70980

표 2B: 우측 파트너

m5Q 변이체					47898-FLAG	47874-FLAG
F1	F2	F3	F4	F5	SBS#	SBS#
					70981	70989
Q					70982	70990
		Q			70983	70991
				Q	70984	70992
Q		Q			70985	70993
Q				Q	70986	70994
		Q		Q	70987	70995
Q		Q		Q	70988	70996

[0243]

[0244]

그런 후 이들 단백질을 알부민 유전자좌 (ALB) 또는 SMCHD1 표적 외에 대한 활성에 대해 테스트하였고, 여기서 변이체는 원래의 우측 (47898-FLAG)- 또는 좌측-손 파트너 (47171-FLAG)와 쌍을 형성하였다. K562 세포를 ZFN mRNA로 Amaxa 전기천공에 의해 제조사의 설명서를 따라 전기천공하였다. 세포를 전기천공 후 16시간 후에 수득하였다. Hdna를 QuickExtract™ DNA 추출 용액 (Lucigen)을 사용하여 제조사의 설명서를 따라 추출하였다. indels의 백분율을 알부민 ZFN 절단 부위 또는 SMCHD1 표적 외를 둘러싼 프라이머로 얻어진 PCR 생성물의 MiSeq 서열분석에 의해 측정하였다. 고효성은 모든 변이체에서 볼 수 있었고, 증가된 특이성 (표적-외 부위와 비교한)을 대부분의 변이체에서, 특히 F1 및 F3 변이체에서 관찰하였다 (도 3A).

[0245]

다음으로, ELD/KKR FokI 뉴클레아제 도메인 (미국 특허 제 8,962,281호 참조)에서 만들어진 포스페이트 접착 아미노산 측쇄 돌연변이를 상기에서 설명된 ZFP 골격 돌연변이와 쌍을 형성하였다. 이들 실험에서, 표 1에 나타낸 ZFN을 골격 및 뉴클레아제 도메인 둘 다에 돌연변이를 포함하는 것으로 요약 설명하였고, 각각에 새로운 특유한 숫자 식별자를 제시하였다 (표 3 참조).

표 3

알부민 특이적 ZFN 변이체

SBS 번호	ZFN 모	골격 돌연변이(들)	Fok*
71545	42875-FLAG	F1m5Q, F3m5Q	ELD-R416H
71557	42875-FLAG	F1m5Q, F3m5Q	ELD-N542D
71673	47898-FLAG	F3m5Q	KKR-R416H
71721	47874-FLAG	F3m5Q	KKR-R416H
71739	47874-FLAG	F1m5Q, F3m5Q	KKR-S418P
71741	47874-FLAG	F1m5Q, F3m5Q	KKR-L424F

[0246]

71721	47874-FLAG	F3m5Q	KKR-R416H
71728	47874-FLAG	F3m5Q	KKR-P478S
71639	42908-FLAG	F1m5Q, F3m5Q	ELD
71641	42908-FLAG	F1m5Q, F3m5Q	ELD-R416H
71653	42908-FLAG	F1m5Q, F3m5Q	ELD-N542D
71696	47898-FLAG	F1m5Q, F3m5Q	KKR-P478S

- [0247] *FokI 도메인은 표에서 제시된 것과 같이 또는 FokI 단독의 절단 도메인과 관련하여 전장과 관련하여 숫자화될 수 있다 (예컨대, N542D 및 N159D는 동일한 조작된 FokI 도메인을 나타내고 P478S 및 P95S는 동일한 조작된 FokI 도메인을 나타냄).
- [0248] 그런 후 이들 ZFN을 포함하는 쌍을 알부민 유전자좌 및 표적-외 부위 SMCHD1에서의 ZFN 활성을 관찰하기 위해 K562 세포에서 테스트하였다. 간단히 설명하면, K562 세포를 표지된 바와 같은 알부민-표적화 ZFN으로 트랜스펙션하였다. 세포를 ZFN 활성 (% indels)에 대해 상기에서 기술한 것과 같이 트랜스펙션 후 24시간 후에 심층 서열분석에 의해 평가하였다.
- [0249] 그 결과 (도 3A-3C에 도시됨)는 알부민-특이적 활성이 크게 개선된 한편 표적 외 활성은 바탕값 수준으로 떨어진 것을 입증한다. 도 3A 및 3B는 표적-외 부위 SMCHD1에서의 결과를 보여준다. 비편향 포획 평가를 또한 사용하여 K562 세포 및 HepG2 세포 둘 다에서 및 도시된 것과 같이 (도 3C 참조) 이들 ZFN 쌍에 대한 잠재적 표적 외 유전자좌를 확인하였고 (미국 특허 출원 62/675,435 참조), 변형된 ZFN 쌍은 이들 부위에서 활성을 거의 검출할 수 없었다.
- [0250] 그러므로, 알부민 유전자좌에서 고수준의 표적 상 절단을 유지하는 한편 A 대 G SNP에 대해 내성이며 고도의 특이성을 가진 최적화된 인간 알부민-특이적 ZFN을 설계하였다.
- [0251] 최적화된 ZFN을 또한 iPSC-유래 인간 간세포에서의 활성 (절단 및 도너의 표적화된 통합)에 대해 테스트하였다. 간단히 설명하면, iPSC-유래 인간 간세포를 Cellular Dynamics international로부터 구입하여, 제조사의 프로토콜을 따라 플레이팅하고 배양하였다. 세포를 플레이팅 후 4일째에 인간 ZFN AAV로 형질도입시켰다: 저 - 30 MOI, 중간 - 100 MOI 및 고 - 300 MOI. 다음날 세포를 인간 도너 AAV로 형질도입시켰다: 저 - 240 MOI, 중간 - 800 MOI 및 고 2400 MOI. 세포 및 조건부 배지를 ZFN AAV 형질도입 후 7일째에 분석을 위해 수득하였다.
- [0252] 도 3D 및 3E(i)에서 나타난 것과 같이, 최적화된 ZFN은 모 ZFN과 비교하여 최적화된 ZFN을 사용하여 통합된 도입유전자로부터 최대 12배 더 높은 수준의 절단 효율 및 13배 더 높은 수준의 도입유전자 (IDS) 생성을 보였다.
- [0253] 시간 경과에 따른 두 ZFN 쌍으로의 절단 후 포획된 도입유전자의 발현을 평가하기 위한 연구를 수행하였다. 간단히 설명하면, 인간 iPSC-유래 간세포를 인간 IDS 도입유전자 도너 (SB-IDS)와 함께 제1 (47171:47898) 또는 제2 (71557:71728) 세대 ZFN을 암호화하는 rAAV2/6 벡터로 3개 한 벌로 형질도입시켰다. 형질도입 후 제5일 및 7일째에 IDS 효소 활성 (세포 배양 상층액의 mL당 시간당 생성물의 nmol로서 표시함, [nmol/hr/mL])을 IDS 효소 활성 검정에 의해 측정하였다. ZFN 및 SB-IDS 도너를 1:1:2의 좌측 ZFN:우측 ZFN:도너 비율로 전달하였다. 100:100:200 (MOI of ZFN:ZFN:도너)의 용량에서, 세포의 제2 세대 ZFN 쌍으로의 처리는 각각 제5일 및 7일에 세포 상층액에서 제1 세대 쌍으로의 처리보다 2배 및 5배 더 많은 IDS를 초래하였다. 300:300:600의 용량에서, 세포의 제2 세대 ZFN 쌍으로의 처리는 각각 제5일 및 7일에 세포 상층액에서 제1 세대 쌍으로의 처리보다 7배 및 21배 더 많은 IDS를 초래하였다 (도 3E(ii) 참조).
- [0254] 다음으로, 좌측 ZFN 결합 부위 내에서 A 대 G SNP에 대해 이형접합성인 일차 인간 간세포를 ZFN 당 10 또는 50 ng mRNA의 농도에서 47171/47898 또는 71557/47898 쌍을 암호화하는 메신저 RNA (mRNA)로 트랜스펙션하였다. 게놈 DNA를 야생형 (A:T) 또는 SNP (G:C) 표적-상 부위에서 유전자 변형 (% 삽입 및 결실 [indels]) 수준에 대해 MiSeq 심층 서열분석에 의해 평가하였다. 결과 (도 3F(i) 및 3F(ii) 참조)는 71557/47898 쌍이 도입유전자를 함유하는 SNP 및 비SNP 둘 다에서 동등한 활성을 가졌음을 입증하였다.
- [0255] 시험관내에서 세포에서의 유전자 변형 속도를 또한 분석하였다. ZFN (100 K 및 600 K MOI)의 세포로의 AAV2/6-매개 전달 후의 유전자 변형 수준을 3개의 생물학적 복제물에서 이간 일차 간세포에서 10일 동안 노출시켜 평가하였다. 세포를 제1, 3, 5 및 10일에 수득하고, 게놈 DNA를 분리하고, PCR 증폭시키고 MiSeq 심층 분석하였다. 71557/71728 ZFN 쌍은 47171/47898 ZFN 쌍과 비교하여 10일 동안 더 빠른 동역학을 보였다 (도 3G). 71557/71728 ZFN 쌍의 두 용량 수준에 대한 더 빠른 동역학은 3일 정도로 빠르게 분명하였고 이때 이 쌍은 47171/47898 ZFN 쌍으로의 2.2% 및 3.0% indels과 비교하여, 100 K 및 600 K 용량 수준에 대해 각각 8.3% 및

17.8% indels을 유발하였다. 71557/71728 ZFN 쌍은 제10일에 40% indels 이상의 효과 포화도에 도달하는 것으로 나타났다. 71557/71728 ZFN 쌍으로의 치료는 47171/47898 ZFN 쌍과 비교하여 시간 경과에 따라 더 높은 수준의 유전자 변형을 초래하였다. 47171/47898 ZFN 쌍의 투여는 100 K 및 600 K MOI 용량 수준에 대해 각각 16.9% 및 25.4% indels의 유전자 변형 수준으로 이어진 한편, 71557/71728 ZFN 쌍은 각각 35.1% 및 44.2% indels을 유발하였다. 10일째에, 71557/71728 ZFN 쌍 활성은 저용량 및 고용량 그룹에 대해 47171/47898 ZFN 쌍 활성보다, 각각 2.1배 및 1.7배 더 높았다. 전체 용량-반응 곡선이 10-일 세포 배양 실험 중에 이루어지지 않았기 때문에, 진성 EC50 값을 계산하고 비교하는 것은 가능하지 않았다. 그러나, 상기 결과는 제2 세대 ZFN으로 10일에 걸친 ZFN 활성에서 대략 2배 증가의 합리적인 추정을 제공한다.

[0256] 평가 연구를 또한 제2 세대 ZFN 쌍과 관련된 임의의 표적 외 절단 사건을 특성화하기 위하여 인간 일차 간세포에서 수행하였다. 표적 외 절단을 이전에 개시된 방법에 의해 측정하였다 (PCT 공보 WO2018039440 참조). 게놈 DNA를 QuickExtract™ DNA 추출 용액 (Lucigen)을 사용하여 제조사 조건을 따라 추출하였다.

[0257] SMCHD1을 ZFN에 대한 검출된 표적-외 부위로서 확인하였다. 47171/47898 또는 71557/71728 ZFN 쌍을 암호화하는 AAV2/6으로 형질도입된 인간 일차 간세포를 MiSeq 심층 서열분석에 의해 평가하였다. 인간 일차 간세포를 다음의 MOI: 3 K, 10 K, 30 K, 100 K, 300 K, 600 K 및 Mock에서 10일 동안 제2 세대 ZFN 쌍을 포함하는 AAV로 처리하였다. 용량 반응을 표적-상 알부민 부위에서 ZFN 변형과 관련하여 관찰하였다. 알부민 유전자좌에서 제10 일 째에 평균 % indels은 71557/71728 ZFN 쌍으로의 처리 후에 0.16, 7, 15, 15, 21, 30 및 44%였다 (도 3H, 상부 줄 참조). NS - 양측 t-테스트에 의해 통계학적으로 유의미하지 않음, * - p-값 < 0.05, 양측 t-테스트에 의함. 100 K MOI 용량에서, 47171/47898 ZFN은 17% indels의 평균 표적-상 활성 및 0.11% indels의 표적 외 활성 (17/0.11 비율 = 154)을 보였고; 71557/71728 ZFN은 35% indels의 평균 표적-상 활성 및 0.08% indels의 표적 외 활성 (35/0.08 비율 = 438)을 보였다. 두 비율을 비교하면, 제2 세대 ZFN은 제1 세대 ZFN보다 약 2.8배 더 선택적이다. 600 K MOI 용량에서, 47171/47898 ZFN은 25% indels의 평균 표적-상 활성 및 0.36% indels의 표적 외 활성 (25/0.36 비율 = 69)을 보였고; 71557/71728 ZFN은 44% indels의 평균 표적-상 활성 및 0.34% indels의 표적 외 활성 (44/0.34 비율 = 130)을 보였다. 두 비율을 비교하면, 제2 세대 ZFN은 제1 세대 ZFN보다 약 1.9배 더 선택적이다. 100 K 및 600 K MOI에서, 제1 및 제2 세대 ZFN에 대한 % indels은 각각 17% 및 35%, 및 25% 및 44%였고, 이것은 제2 세대 ZFN (ZFN 2.0)이 제1 세대 ZFN보다 약 2배 더 강력한 것을 나타낸다.

[0258] 제1 세대 좌측 ZFN (47171)으로 처리된 세포에서, SNP-함유 대립유전자에서의 ZFN 활성은 단지 39-44%로, 야생형 대립유전자에서의 활성만큼 높았다. 비교하면, 제2 세대 좌측 ZFN (71557)으로 처리된 세포에서, SNP-함유 대립유전자에서의 ZFN 활성은 야생형 대립유전자에서의 활성만큼 높았다.

[0259] **실시예 3: 다가양이온성 펩타이드 태그의 사용은 ZFN 활성을 증가시킨다**

[0260] 다가양이온성 펩타이드 태그를 포함하는 ZFN을 절단 활성에 대해 조사하였다. 사용한 펩타이드는 아미노산 서열 Nterm- DYKDHG-DYKDHDI-DYKDDDK (SEQ ID NO:71)를 포함하는 3x Flag 태그 (PCT 공보 번호 WO2001027293 참조). 3x Flag 태그를 암호화하는 서열을 단백질의 N-말단에서 ZFN 융합 단백질에 융합한다 (표 1 참조).

[0261] 이 연구에 대한 ZFN을 테스트되는 세포의 세포 기원을 제어하기 위하여 마우스, NHP 및 인간 사이에서 관찰되는 7개의 초보존된 DNA 표적에 대해 만들었다 (Bejerano *et al* (2004) *Science* 302(5675):1321-1325 참조). 74개의 ZFN 쌍을 초보존된 DNA 표적에 대해 만들었고 3X Flag 태그가 있거나 없이 K562 세포에서의 활성에 대해 테스트하였다.

[0262] 결과 (도 4A)는 다가양이온성 3x Flag 태그의 존재가 ZFN 절단 활성에 대해 매우 유의했고 (% indels을 측정하기 위한 세포 검정에 의해 측정됨), 3x Flag 태그를 포함하는 이들 ZFN 쌍에서, 활성은 Flag 태그가 없는 ZFN과 비교하여 평균 4.1배 증가하였다 (도 4B). 도 4C에서 나타난 것과 같이, 3x Flag 태그의 첨가는 ZFN 활성에서 2-3배 증가로 이어졌다.

[0263] **실시예 4: WPRE 조절 요소의 첨가**

[0264] WPRE는 LNP에 의한 (도 5B, 미국 특허출원 공개 번호 US-2018-0185516-A1호 참조) 또는 AAV에 의한 (도 5C, 미국 특허출원 공보 2016-0326548호 참조) 전달 후에 시험관내에서 (도 5A, 미국 특허출원 공보 2016-0326548호 참조) 및 생체내에서 모두, ZFN의 활성을 증가시키는 것으로 나타났다.

[0265] 천연 형태에서 WPRE는 WHV-X 단백질에 대한 부분 오픈 리딩 프레임 (ORF)을 함유한다. 다른 바이러스 요소의 맥락에서 전체 발현된 WHV-X 단백질은 WHV (We2) 인헨서와 같이 우드척과 마우스에서 간암종의 더 높은 위험과 관

련되었다 (Hohne *et. al* (1990) *EMBO J* 9(4):1137-45; Flajolet *et. al* (1998) *J Virol* 72(7):6175-80). WHV-X 단백질은 직접적으로 발암성인 것으로 나타나지는 않지만, 일부 연구는 특정 상황 하에서 그것이 헤파드나바이러스 (남성의 경우 B형 간염 바이러스; 우드척의 경우 우드척 간염 바이러스)에 의한 감염과 관련된 간 암의 생성에 대한 약한 보조인자로서 작용할 수 있다고 시사한다. 사용한 WPRE 서열은 WHV-X 단백질에 대한 부분 오픈 리딩 프레임에 함유하지만, WHV-X 단백질의 발현을 증폭시키는 것으로 여겨지는 We2 인핸서는 함유하지 않는다. 추가로, WPRE 서열은 중단 코돈의 3'에 배치되고 프로모터 및 치료용 도입유전자와의 프레임 밖에 있다; 그러므로 이 서열의 번역은 중단 코돈 리드-스루(read-through)가 일어난다 하더라도 예상되지 않는다.

[0266] 그러므로, WPRE 요소를 본원에 기술된 폴리뉴클레오타이드에, 전형적으로 뉴클레아제 코딩 서열의 3'에 첨가하였다 (도 7). 사용한 WPRE 요소는 하기에서 나타낸, J02442 변이체로부터 유래된 WPREmut6 (Zanta-Boussif, 상기 동일)일 수 있다:

[0267] J02442 WPREmut6:

[0268] 5' AATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGATATTCTTAACTATGTTGCTCCTTTTACG
 [0269] CTGTGTGGATATGCTGCTTAAATGCCTCTGTATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTC
 [0270] CTTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCGCCAACGTGGCGTGGTGT
 [0271] GCTCTGTGTTTGTGACGCAACCCCACTGGCTGGGGCATTGCCACCACCTGTCAACTCCTTTCTGGGACT
 [0272] TTCGCTTTCCTCCCGATCGCCACGGCAGAACTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGC
 [0273] TAGGTTGCTGGGCACTGATAATTCCTGGTGTGTGCGGGAAATCATCGTCCCTTTCCTTGGCTGCTCGCCT
 [0274] GTGTTGCCAACTGGATCCTGCGCGGACGTCCTTCTGCTACGTCCTTCGGCTCTCAATCCAGCGGACCTC
 [0275] CCTTCCCGAGGCTTCTGCGGGTCTGCGGCCTCTCCCGCTCTTCGCTTTCGGCCTCCGACGAGTCGGAT
 [0276] CTCCTTTGGGCGCCTCCCGCCTG (SEQ ID NO:42).

[0277] 일부 구체예에서, WPRE 요소는 감마 및 알파 요소만을 포함하는 절단된 구성물이다. WPRE3의 서열을 하기에 나타낸다:

[0278] WPRE3:

[0279] GATAATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTAC
 [0280] GCTATGTGGATACGCTGCTTAAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTCTCCT
 [0281] CCTGTATAAATCCTGGTTAGTTCTTGCCACGGCGAACTCATCGCCGCTGCCTTGCCCGCTGCTGGACA
 [0282] GGGGCTCGGCTGTTGGGCACTGACAATTCCTGG (SEQ ID NO:68).

[0283] 일부 구체예에서, 하기에서 나타내는 J04514.1 변이체를 사용하고 mut6 변이체를 서열에 첨가한다:

[0284] J04514.1 WPREmut6:

[0285] AATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGGTATTCTTAACTATGTTGCTCCTTTTACGCT
 [0286] ATGTGGATACGCTGCTTAAATGCCTTTGTATCATGCTATTGCTTCCCGTATGGCTTTCATTTCTCCTCCT
 [0287] TGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCAGGCAACGTGGCGTGGTGTGC
 [0288] ACTGTGTTTGTGACGCAACCCCACTGGTTGGGGCATTGCCACCACCTGTGAGTCTCTTTCGGGACTTT
 [0289] CGCTTTCCTCCCTATTGCCACGGCGAACTCATCGCCGCTGCCTTGCCCGCTGCTGGACAGGGGCTC
 [0290] GGCTGTTGGGCACTGACAATTCCTGGTGTGTGCGGGAAATCATCGTCCCTTTCCTTGGCTGCTCGCCTGT
 [0291] GTTGCCACCTGGATTCTGCGGGGACGTCCTTCTGCTACGTCCTTCGGCCCTCAATCCAGCGGACCTTCC
 [0292] TTCCTCCCGGCTGCTGCGGGCTGCGGCCTTCCCGCTTTCGCTTTCGGCCTCAGACGAGTCGGATCT
 [0293] CCTTTGGGCGCCTCCCGCCTG (SEQ ID NO:69)

[0294] 모두 3개의 WPRE 변이체는 WHV X 유전자를 발현하는 능력이 없고, 본원에 기술된 발현 구성물에서 상호교환 가

능하게 사용될 수 있다.

[0295] 실시예 5: 5' UTR의 첨가

[0296] 1994년에, Krieg 및 Melton (*Nucl. Acid. Res* 12(18):7057)은 제노푸스 베타-글로빈 유전자의 5' 미번역 영역을 기술하였고 그것의 mRNA 안정화 서열로서의 잠재력을 인식하였다. 그러므로, 이 요소에 대한 서열 ([TG]CTTGTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGCAGAT (SEQ ID NO:1), 여기서 TG는 선택적으로 약칭된 β-globin)을 뉴클레아제-표적화 발현 카세트의 5' 미번역 영역에서 테스트하였다. 구성물을 상기에서 기술한 것과 같이 K562 세포에서 테스트하되, 단 세포를 트랜스펙션 후 24시간째에 검정하였다.

[0297] 도 9에서 나타낸 것과 같이, 신규한 5'UTR의 첨가는 ZFN 활성의 2-3배 증가로 이어졌다.

[0298] 실시예 6: 향상의 조합

[0299] 3x Flag 태그 ("3xFlag"), 제노푸스 β-globin ("XBG") 및/또는 WPRE 요소를 포함하는 ZFN을 iPSC 유래 인간 간세포에서 다양한 조합으로 실험하였다. 사용한 ZFN은 42877/47931이었고 그것들을 AAV6을 통해 전달하였다. 세포를 형질도입 후 6일 후에 수득하였고 알부민 표적에 대한 활성을 측정하였다.

[0300] 그 결과는 초기 벡터와 비교하여 전부 3개의 요소의 존재 하에 ZFN 활성의 증가를 입증하였다.

[0301] 추가로, 따라서 상기 기술된 상이한 요소 전부의 다양한 조합이 없거나 있도록 ZFN을 포함한 일련의 8개의 발현 구성물을 구성하였다 (변이체 V1 내지 V8을 보여주는 도 7A 참조). 모든 변이체를 AAV 6 벡터에 사용하였고 ApoE-hAAT 프로모터와 함께 사용하였다 (Okuyama *et al* (1996) *Hum Gene Ther* 7(5):637-45). 이들 벡터는 또한 기술된 인간 베타-글로빈-면역글로불린 키메라 인트론 서열을, 인간 β-글로빈 유전자의 제1 인트론으로부터의 5'-도너 부위 및 분지 및 면역글로불린 유전자 중쇄 가변 영역 (Promega)의 인트론으로부터의 3'-억셉터 부위로 구성된 키메라 인트론으로서 포함하였다. 이들 구성물에 사용한 NLS는 SV40으로부터 유래하였다.

[0302] 그런 후 벡터를 두 ZFN이 동일한 변이체 구조로 전달된, 전달된 ZFN 쌍 (42877/47931)의 절단 활성에 대해 테스트하였다.

[0303] 도 8에서 나타낸 것과 같이, 변이체 구조 중 2개, 변이체 4 및 변이체 8이 둘 다 FLAG 서열을 포함하였고 (표 1), 현재 사용 중인 벡터 구조를 포함한 다른 변이체들과 비교하여 월등한 결과를 제공하였다. 이들 실험을 HepG2 세포에서 상이한 벡터를 포함하는 AAV의 300,000 MOI를 사용하여 수행하였다.

[0304] 다양한 향상의 조합을 포함하여 만들어진 1개의 알부민-특이적 ZFN 쌍은 상기에서 기술한 것과 같이 71557/71728 쌍이다. 도 7B는 이들 ZFN을 암호화하는 AAV의 개략을 도시한다. 쌍의 요소는 하기 표 및 서열에 나타낸다. 임의의 요소가 유사한 서열, 예를 들어 하기 WPRE 서열 대신 기술분야에 알려져 있는 또는 상기 실시예 4에서 나타낸 것과 같은 WPRE 서열로 치환될 수 있는 것이 명백할 것이다.

표 4

[0305] SB71557 AAV (SEQ ID NO:43)의 요소

뉴클레오타이드 위치- 주석	특징/설명	SEQ ID NO:	서열
1-130	5' ITR [괄호 안의 일반 텍스트]	44	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGCTCAGTGAGCGAGCGAGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCT
156-476	ApoE (인헨서) 밑줄	45	AGGCTCAGAGGCACACAGGAGTTTCTGGGCTCACCCCTGCCCTTCCAACCCCTCAGTTCCATCTCCAGCAGCTGTTTGTGTGCTGCCTCTGAAGTCCACTGAACAACTTCAGCCTACTCATGTCCCTAAAATGGGCAAACATTGCAAGCAGCAAACAGCAAACACACAGCCCTCCCTGCCTGCTGACCTTGGAGCTGGGGCAGAGGTCAGAGACCTCTCTGGGCCATGCCACCTCCAACATCCACTCGACCCCTTGGAAATTCGGTGGAGAGGAGCAGAGGTTGCTCGCGTGGTTTAGGTAGTGTGAGAGGG
485-877	hAAT (프로모터) 이탤릭체	46	GATCTTGCTACCAGTGGAAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCTAAAGTGGTACTCTCCAGAGACTGTCTGACTCAGCCACCCCTCCACCTTGGACACAGGACGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTCGGTAAGTGCAGTGGAAAGCTGTACTGCCCAGGCAAAGCGTCCGGCAGCGTAGGCGGGGACTCAGATCCCAGCCAGTGGACTTAGCCCTGTTTGTCTCCTCCGATAACTGGGGTACCTTGGTTAATATTCACAGCAGCCTCCCCGTTGCCCTCTGGATCCACTGCTTAATACGGACGAGGACAGGGCCCTGCTCCTCAGCTTCAGGCACCACCACTGACCTGGGACAGT

886-933	5' UTR 진합	47	CTGTGTTCTTTTTGCAGAAGCTCAGAATAAACGCTCAACTTTGGCA GAT
943-1075	인간 β 글로빈 /IgG 키메라 인 트론 (인트론) 이중 밀출	48	GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAACTGGGCTTGTCGA GACAGAGAAGACTCTTGCGTTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTT TGCCTTTCTCTCCACAG
1086-1154	N-말단 펩타이드 물결선	49	GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAAGGATGACG ATGACAAG
1161-1181	핵 국지화 신호 진한 이텔릭체	50	CCCAAGAAGAAGAGGAAGGTC
1200-1631	ZFP 71557 DNA -결합 도메인 소문자	51	GCCGCTATGGCTGAGAGGCCCTTCCAGTGTGCAATCTGCATGCAGAAGTTTCAGTCACT CCGGCAACCTGGCCGCCACATCCGCACCCACACCGGCGAGAAGCCTTTTGCCTGTGA CATTTGTGGGAGGAAATTTGCCCTGAAGCAGAACCTGTGATGCATACCAAGATACAC ACGGGCGAGAAGCCCTTCCAGTGTGCAATCTGCATGCAGAAGTTTGCCTGGCAGTCCA ACCTGCAGAACCATAACCAAGATACACACGGGCGAGAAGCCCTTCCAGTGTGCAATCTG CATGCGTAACTTCACTACCTCCGGCAACCTGACCCGCCACATCCGCACCCACACCGGC GAGAAGCCTTTTGCCTGTGACATTTGTGGGAGGAAATTTGCCCGCGCTCCCACCTGA CCTCCCATACCAAGATACACTGCGG
1638-2237	FokI-ELD 뉴클레 아제 도메인 파선 밀출	52	CAGCTGGTGAAGAGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAAGCTGAAGT ACGTGCCCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCG CATCCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGTACGGCTACAGGGGAAAG CACCTGGGCGAAGCAGAAAGCCTGACGGCGCCATCTATACAGTGGGAGCCCATCG ATTACGGCGTGATCGTGGACACAAAGGCTACAGCGGCGGTACAATCTGCCTATCGG CCAGGCGACGAGATGGAGAGATACGTGGAGGAGAACCAGACCCGGGATAGCACCTC AACCCCAACGAGTGGTGAAGGTGATCCCTAGCAGCGTGACCGAGTTCAAGTTCTCTGT TCGTGAGCGGCCACTCAAGGGCAACTACAAGGCCAGCTGACC AGGCTGAACCACATCACTCAACTGCGACGGCGCGTGTGAGCGTGGAGGAGCTGCTGA TCGGCGGCGAGATGATCAAAAGCCGACCCCTGACACTGGAGGAGGTGCGGCGCAAGTT CAACAACGGCGAGATCAACTTCAGATCTTGATAA
2250-2841	WPREmut6 3'UTR 점선 밀출	53	AATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGATATTCTTAACTATGTTG CTCCTTTTACGCTGTGTGGATATGCTGCTTAAATGCCTCTGTATCATGCTATTGCTTC CCGTACGGCTTTCGTTTCTCCTCTGTATAAATCCTGGTGTGCTGTCTTTATGAG GAGTTGTGGCCCGTTGTCCGTCAACGTGGCGTGGTGTGCTCTGTGTTTGTGACGCAA CCCCCACTGGCTGGGCATTTGCCACCCTGTCAACTCCTTTCTGGGACTTTCGCTTT CCCCTCCGATCGCCACGGCAGAACTCATCGCCCGCTGCCTTGCCCGCTGTGGACA GGGCTAGGTTGCTGGCACTGATAAATCCGTGGTGTGTGCGGGAAATCATCGTCCCT TTCTTGGCTGCTCGCTGTGTGCAACTGGATCCTGCGCGGG ACGTCTTCTGCTACGTCCTTCCGCTCTCAATCCAGCGACCTCCCTTCCGAGGCC TTCTGCCGTTTCTGCGGCTCTCCCGCTCTCGCTTTCGGCTCCGACGAGTGGAT CTCCTTTGGGCCCTCCCGCTG
2848-3070	폴리아데닐화 신 호 물결선	54	CTGTGCCCTTCTAGTTGCCAGCCATCTGTGTGTTTGGCCCTCCCCCGTGCCTTCCTTGAC CCTGGAAGGTGCCACTCCCCTGCTTTCCTAATAAAATGAGGAAATTCATCGCAT TGTCTGAGTAGGTGTCATTCTATCTGGGGGTGGGTGGGCAGGACAGCAAGGGGG AGGATTGGGAAGACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTAT
3088-3195	3' ITR [괄호 안 의 진한 텍스트]	55	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGA GGCCGCCCGGCTTTGCCCGGGCGCCTCAGTGAGCGAGCGAGCGCGCAG

[0306] 71557 AAV의 완전한 서열:

```
[CTGCGCGCTC GCTCGCTCAC TGAGGCCGCC CGGGCAAAGC CCGGGCGTCG 50
GGCGACCTTT GGTCGCCCGG CCTCAGTGAG CGAGCGAGCG CGCAGAGAGG 100
GAGTGGCCAA CTCCATCACT AGGGTTCTCT] GCGGCCTAAG CTTGAGCTCT 150
TCGAAAGGCT CAGAGGCACA CAGGAGTTTC TGGGCTCACC CTGCCCCCTT 200
CCAACCCCTC AGTCCCATC CTCCAGCAGC TGTTTGTGTG CTGCCTCTGA 250
AGTCCACACT GAACAACTT CAGCCTACTC ATGTCCCTAA AATGGGCAAA 300
CATTGCAAGC AGCAAACAGC AAACACACAG CCCTCCCTGC CTGCTGACCT 350
TGGAGCTGGG GCAGAGGTCA GAGACCTCTC TGGGCCCATG CCACCTCCAA 400
CATCCACTCG ACCCCTTGA ATTTGCGTGG AGAGGAGCAG AGGTTGTCTT 450
GGCGTGGTTT AGGTAGTGTG AGAGGGTCC CGGGGATCTT GCTACCAGTG 500
GAACAGCCAC TAAGGATTCT GCAGTGAGAG CAGAGGGCCA GCTAAGTGGT 550
ACTCTCCAG AACTGTCTG ACTCACGCCA CCCCCTCCAC CTTGGACACA 600
```

[0307]

GGACGCTGTG GTTTCTGAGC CAGGTACAAT GACTCCTTTC GGTAAAGTGA 650
 GTGGAAGCTG TACTACTGCC AGGCAAAGCG TCCGGGCAGC GTAGGCGGGC 700
 GACTCAGATC CCAGCCAGTG GACTTAGCCC CTGTTTGCTC CTCCGATAAC 750
 TGGGGTGACC TTGGTTAATA TTCACCAGCA GCCTCCCCCG TTGCCCTCT 800
 GGATCCACTG CTTAAATACG GACGAGGACA GGGCCCTGTC TCCTCAGCTT 850
 CAGGCACCAC CACTGACCTG GGACAGTCCT AGGTGCTTGT **TCTTTTTGCA** 900
GAAGCTCAGA ATAAACGCTC AACTTTGGCA GATACTAGTC AGGTAAGTAT 950
CAAGGTTACA AGACAGGTTT AAGGAGACCA ATAGAAACTG GGCTTGTCGA 1000
GACAGAGAAG ACTCTTGCGT TTCTGATAGG CACCTATTGG TCTTACTGAC 1050
ATCCACTTTG CTTTTCTCTC CACAGGACCG GTGCCATGGA CTACAAAGAC 1100
CATGACGGTG ATTATAAGA TCATGACATC GATTACAAGG ATGACGATGA 1150
CAAGATGGCC **CCCAAGAAGA AGAGGAAGGT** CGGCATTCAT GGGGTACCCg 1200
 ccgctatggc ttagaggccc ttccagtgtc gaatctgcat gcagaacttc 1250
 agtcagtccg gcaacctggc ccgccacatc cgcaccaca ccggcgagaa 1300
 gccttttggc tgtgacattt gtgggaggaa atttgcctg aagcagaacc 1350
 tgtgtatgca taccaagata cacacggcg agaagcctt ccagtgtcga 1400
 atctgcatgc agaagtttgc ctggcagtcc aacctgcaga accataccaa 1450
 gatacacacg ggcgagaagc ccttccagtg tcgaatctgc atgcgtaact 1500
 tcagtacctc cgcaacctg acccgccaca tccgcacca caccggcgag 1550
 aagccttttg cctgtgacat ttgtgggagg aaatttgccc gccgctccca 1600
 cctgacctcc cataccaaga tacacctgcy gGGATCCCAG CTGGTGAAGA 1650
GCGAGCTGGA GGAGAAGAAG TCCGAGCTGC GGCACAAGCT GAAGTACGTG 1700
CCCCACGAGT ACATCGAGCT GATCGAGATC GCCAGGAACA GCACCCAGGA 1750
CCGCATCCTG GAGATGAAGG TGATGGAGTT CTTCATGAAG GTGTACGGCT 1800
ACAGGGGAAA GCACCTGGGC GGAAGCAGAA AGCCTGACGG CGCCATCTAT 1850
ACAGTGGGGA GCCCATCGA TTACGGCGTG ATCGTGGACA CAAAGGCCTA 1900
CAGCGCGGC TACAATCTGC CTATCGGCCA GGCCGACGAG ATGGAGAGAT 1950
ACGTGGAGGA GAACCAGACC CGGATAAGC ACCTCAACCC CAACGAGTGG 2000
TGGAAGGTGT ACCCTAGCAG CGTGACCGAG TTCAAGTTCC TGTTCTGAG 2050
CGGCCACTTC AAGGGCAACT ACAAGGCCCA GCTGACCAGG CTGAACCACA 2100
TCACCAACTG CGACGGCGCC GTGCTGAGCG TGGAGGAGCT GCTGATCGGC 2150
GGCGAGATGA TCAAAGCCGG CACCTGACA CTGGAGGAGG TGCGGCGCAA 2200
GTTCACAAC GCGGAGATCA ACTTCAGATC TTGATAACTC GAGTCTAGAA 2250
ATCAACCTCT GGATTACAAA ATTTGTGAAA GATTGACTGA TATCTTAAC 2300
TATGTGCTC CTTTTACGCT GTGTGGATAT GCTGCTTAA TGCCTCTGTA 2350
TCATGCTATT GCTTCCGTA CGGCTTTCGT TTTCTCCTCC TTGTATAAT 2400
CCTGTTGCT GTCTCTTAT GAGGAGTTGT GGCCCGTTGT CCGTCAACGT 2450
GGCGTGGTGT GCTCTGTGTT TGCTGACGCA ACCCCCACTG GCTGGGGCAT 2500
TGCCACCACC TGTCAACTCC TTTCTGGGAC TTTCGCTTTC CCCCTCCGA 2550
TCGCCACGGC AGAACTCATC GCCCCTGCC TTGCCCGCTG CTGGACAGGG 2600
GCTAGGTTGC TGGGCACTGA TAATFCCGTG GTGTTGTCGG GGAATCATC 2650

GTCCTTTCCT TGGTGTCTCG CCTGTGTTGC CACTGGATC CTGCGCGGA 2700
CGTCCTTCTG CTACGTCCCT TCGGCTCTCA ATCCAGCGGA CCTCCCTCC 2750
CGAGGCCTTC TGCCGGTCTC GCGGCTCTC CCGCTCTTC GCTTTCGGCC 2800
TCCGACGAGT CGGATCTCC TTTGGGCGC CTCCCGCCT GGCTAGCCTG 2850
TGCCTTCTAG TTGCCAGCA TCTGTTGTT GCCCTCCCC CGTGCTTCC 2900
TTGACCCTGG AAGTGCCAC TCCACTGTC CTTTCTAAT AAAATGAGGA 2950
AATTGCATCG CATGTCTGA GTAGGTGCA TTCTATTCTG GGGGTGGGG 3000
TGGGCAGGA CAGCAAGGG GAGGATTGGG AAGACAATAG CAGGCATGCT 3050
GGGATGCGG TGGCTCTAT GCGCCCGCT CGAGCGC **[AGG AACCCTAGT** 3100
GATGGAGTTG GCCACTCCCT CTCTGCGCGC TCGCTCGCTC ACTGAGGCCG 3150
CCCCGGCTTT GCCCGGGCGC CCTCAGTGAG CGAGCGAGCG CGCAG 3195

(SEQ ID NO:43)

[0308]

[0309]

표 5

SB71728 AAV (SEQ ID NO:56)의 요소

[0310]

뉴클레오타이드 위치- 주석	특징/설명	SEQ ID NO:	서열
1-130	5' ITR [괄호 안의 일반 텍스트]	44	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTGGGCGACCTTTGGTCGCCCCGCCCTCAGTGAGCGAGCGAGCGCAGAGAGGGAGTGGCCAATCCATCACTAGGGGTTCTCT
156-476	ApoE (인헨서) 밀줄	45	AGGCTCAGAGGCACACAGGAGTTCTGGGCTCACCTGCCCCCTTCCAACCCCTCAGTCCCATCCTCCAGCAGCTGTTTGTGTGCTGCCTCTGAAGTCCACACTGAACAAATTCAGCCTACTCATGTCCCTAAAATGGGCAAACATTGCAAGCAGCAAACAGCAAACACACAGCCCTCCCTGCCTGCTGACCTTGGAGCTGGGGCAGAGGTCAGAGACCTCTCTGGGCCATGCCACCTCCAACATCCACTCGACCCCTTGAATTTTCGGTGGAGAGGAGCAGAGGTTGTCCTGGCGTGGTTTAGGTAGTGTGAGAGGG
485-877	hAAT (프로모터) 이탤릭체	46	GATCTTGCTACCAGTGGAAACAGCCACTAAGGATTCTGCAGTGAGAGCAGAGGGCCAGCTAAGTGGTACTCTCCAGAGACTGTCTGACTCACGCCACCCCCTCCACCTTGGACACAGGACGCTGTGGTTTCTGAGCCAGGTACAATGACTCCTTTCGGTAAGTGCAGTGGAAAGCTGTACTGCCCAGGCAAAGCGTCCGGCAGCGTAGGCGGGGACTCAGATCCCAGCCAGTGGACTTAGCCCTGTGTTGCTCCTCGATAACTGGGGTGACCTTGGTTAATATTCACCAGCAGCCTCCCCGTTGCCCTCTGGATCCACTGCTTAAATACGGACGAGGACAGGGCCCTGTCTCCTCAGCTTCAGGCACCACCCTGACCTGGGACAGT
886-933	5' UTR 진함	47	CTTGTCTTTTTGCGAGAAGCTCAGAATAAACGCTCAACTTTGGCAGAT
943-1075	인간 β 글로빈 /IgG 키메라 인트론 (인트론) 이중 밀줄	48	GTAAGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAACTGGGCTTGTGCGAGACAGAGAAGACTCTTGCCTTCTGATAGGCACCTATTGGTCTTACTGACATCCACTTTCCTTTCTCTCCACAG
1086-1154	N-말단 펩타이드 물결선	49	GACTACAAAGACCATGACGGTGATTATAAAGATCATGACATCGATTACAA GGATGACGATGACAAG
1161-1181	핵 국지화 신호 진함 이탤릭체	50	CCCAAGAAGAAGGGAAGGTC
1200-1715	ZFP 71728 DNA-결합 도메인 소문자	57	GCCGCTATGGCTGAGAGGCCCTTCCAGTGTGCAATCTGCATGCGTAACTTCAGTCAGTCTCCGACCTGTCCCGCCACATCCGCACCCACACCGGGGAGAGCCCTTTGCCTGTGACATTTGTGGGAGGAAATTTGCCCTGAAGCACAACTGTGACCCATACCAAGATACACACGGGGGAGAAAGCCCTTCCAGTGTGCAATCTGCATGCAAGCTTCAGTGACCAGTCCAACCTGCGCGCCACATCCGCACCCACACCGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGGAAA TTTGCCCGCAACTTCTCCCTGACCATGCATACCAAGATACACACCGGAGA GCGCGGCTTCCAGTGTGCAATCTGCATGCGTAACTTCAGTCTGCGCCACGACCTGGAGCGCCACATCCGCACCCACACCGGCGAGAAGCCTTTTGCCTGTGACATTTGTGGGAGGAAATTTGCCACCGCTCCAACCTGAACAAGCATAACAAGATACACCTGCGG
1722-2315	FokI-KKR 뉴클레아제 도메인 파선 밀줄	58	CAGCTGGTGAAGGCGAGCTGGAGGAGAAGAAGTCCGAGCTGCGGCACAA GCTGAAGTACGTGCCACGAGTACATCGAGCTGATCGAGATCGCCAGGAACAGCACCCAGGACCGCATCTGGAGATGAAGGTGATGGAGTTCTTCATGAAGGTGACGGCTACAGGGGAAAGCACCTGGGGGGAAGCAGAAAGCCTGA CGGGCCATCTATACAGTGGGCGAGCCCATCGATTACGGCGTGATCGTGGACACAAAGGCCTACAGCGGGCTACAATCTGAGCATCGCCAGGCCGACGAGATGCAGAGATACGTGAAGGAGAACCAGACCCGGAATAAGCACATCAACCCCAACGAGTGGTGAAGGTGTACCCTAGCAGCGTGACCGAGTTCAAGTTCCTGTTCTGTGAGCGGCCACTTCAAGGGCAACTACAAGGCCAGCTGACCAGGCTGAACCGCAAAACCAACTGCAATGGCGCGGTGCTGAGCGTGGAGGAGCTGTGATCGGGCGGAGATGATCAAGCCGGCACCCCTGACACTGGAGGAGGTGCGGGCAAGTTCAACAACGGCGAGATCAACTTCTGATAA

2328-2919	WPREmut6 3'UTR 점선 밑줄	53	AATCAACCTCTGGATTACAAAATTTGTGAAAGATTGACTGATATTCTTAA CTATGTTGCTCCTTTTACGCTGTGTGGATATGCTGCTTAAATGCCTCTGT ATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTCCTGTATAAA TCCTGGTTGCTGTCTCTTATGAGGAGTTGTGGCCCGTTGTCCGTCAACG TGGCGTGGTGTGCTCTGTGTTTGTGACGCAACCCCACTGGCTGGGGCA TTGCCACCACCTGTCAACTCCTTTCTGGGACTTTCGCTTCCCCCTCCCG ATCGCCACGGCAGAACTCATCGCCGCTGCCTTGCCCGCTGTGGACAGG GGCTAGGTTGCTGGGCACTGATAATTCCGTGGTGTGTGCGGGAAATCAT CGTCTTTTCTGGCTGCTCGCTGTGTTGCCAACTGGATCCTGCGCGGG ACGTCTTCTGTACGTCCCTTCGGCTCTCAATCCAGCGGACCTCCCTTC CCGAGGCCTTCTGCCGTTTCTGCGGCCTCTCCCGCTTCTCGCTTTCGGC CTCCGACGAGTCGGATCTCCCTTTGGGCCGCTCCCCGCTG
2926-3148	폴리아데닐화 신 호 물결선	54	CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCTCCCCCGTGCCT TCCTTGACCCTGGAAGGTGCCACTCCCACTGTCTTTCCTAATAAAATGA GGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG GGGTGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGATGCGGTGGGCTCTAT
3166-3273	3' ITR [괄호 안 의 진한 텍스트]	55	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCG CTCACTGAGGCCGCCCCGGCTTTGCCGGGCGGCCCTCAGTGAGCGAGCGA GCGCGCAG

[0311] 71728 AAV의 완전한 서열:

```
[CTGCGCGCTC GCTCGCTCAC TGAGGCCGCC CGGGCAAAGC CCGGGCGTGC 50
GGCGACCTTT GGTCGCCCGG CCTCAGTGAG CGAGCGAGCG CGCAGAGAGG 100
GAGTGGCCAA CTCCATCACT AGGGGTTCT] GCGGCCTAAG CTTGAGCTCT 150
TCGAAAGGCT CAGAGGCACA CAGGAGTTTC TGGGCTCACC CTGCCCCCTT 200
CCAACCCCTC AGTTCCCATC CTCCAGCAGC TGTTTGTGTG CTGCCTCTGA 250
AGTCCCACT GAACAACTT CAGCCTACTC ATGTCCCTAA AATGGGCAA 300
CATTGCAAGC AGCAAACAGC AAACACACAG CCCTCCCTGC CTGCTGACCT 350
TGGAGCTGGG GCAGAGGTCA GAGACCTCTC TGGGCCCATG CCACCTCCAA 400
```

[0312]

CATCCACTCG ACCCCTTGGA ATTTCCGGTGG AGAGGAGCAG AGGTTGTCC 450
GCGGTGGTTT AGGTAGTGTG AGAGGGGTCC CGGGGATCTT GCTACCAGTG 500
GAACAGCCAC TAAGGATTCT GCAGTGAGAG CAGAGGGCCA GCTAAGTGGT 550
ACTTCTCCAG AGACTGTCTG ACTCACGCCA CCCCTCCAC CTTGGACACA 600
GGACGCTGTG GTTCTGAGC CAGGTACAAT GACTCCTTTC GGTAAGTGCA 650
GTGGAAGCTG TACTACTGCC AGGCAAAGCG TCCGGGCAGC GTAGGCGGGC 700
GACTCAGATC CCAGCCAGTG GACTTAGCCC CTGTTTGCTC CTCCGATAAC 750
TGGGGTGACC TTGGTTAATA TTCACCAGCA GCCTCCCCCG TTGCCCTCT 800
GGATCCACTG CTTAAATACG GACGAGGACA GGGCCCTGTC TCCTCAGCTT 850
CAGGCACCAC CACTGACCTG GGACAGTCTT AGGTGCTTGT TCTTTTGGCA 900
GAAGCTCAGA ATAAACGCTC AACTTTGGCA GATACTAGTC AGGTAAGTAT 950
CAAGGTTACA AGACAGGTTT AAGGAGACCA ATAGAACTG GGCTTGTCGA 1000
GACAGAGAAG ACTCTTGCCT TTCTGATAGG CACCTATTGG TCTTACTGAC 1050
ATCCACTTTG CCTTCTCTC CACAGGACCG GTGCCATGGA CTACAAAGAC 1100
CATGACGGTG ATTATAAAGA TCATGACATC GATTACAAGG ATGACGATGA 1150
CAAGATGGCC CCAAGAAGA AGAGGAAGGT CGGCATTCAT GGGGTACCCg 1200
 ccgctatggc tgagaggccc ttccagtgtc gaatctgcat gcgtaacttc 1250
 agtcagtcct ccgacctgtc ccgccacatc cgcacccaca ccggcgagaa 1300
 gccttttgcc tgtgacattt gtgggaggaa atttgccctg aagcacaacc 1350
 tgctgacceca taccaagata cacacggcgy agaagccctt ccagtgtcga 1400
 atctgcatgc agaacttcag tgaccagtcc aacctgcgcy cccacatccg 1450
 caccacacacc ggcgagaagc cttttgcctg tgacatttgt gggaggaaat 1500
 ttgcccgcaa cttctccctg accatgcata ccaagataca caccggagag 1550
 cgcggcttcc agtgtcgaat ctgcatgogt aacttcagt cgcgccacga 1600
 cctggagcgc cacatccgca cccacaccgy cgagaagcct tttgcctgtg 1650
 acatttgtgg gaggaaattt gccaccgct ccaacctgaa caagcatacc 1700
 aagatacacc tgcggGGATC CCAGCTGGTG AAGAGCGAGC TGGAGGAGAA 1750
GAAGTCCGAG CTGCGGCACA AGCTGAAGTA CGTGCCCCAC GAGTACATCG 1800
AGTGATCGA GATCGCCAGG AACAGCACCC AGGACCGCAT CCTGGAGATG 1850
AAGGTGATGG AGTTCTTCAT GAAGGTGTAC GGCTACAGGG GAAAGCACCT 1900
GGCGGAAGC AGAAAGCCTG ACGGCGCCAT CTATACAGTG GGCAGCCCA 1950
TCGATTACGG CGTGATCGTG GACACAAAGG CCTACAGCGG CGGCTACAAT 2000
CTGAGCATCG GCCAGGCCGA CGAGATGCAG AGATACGTGA AGGAGAACCA 2050
GACCCGGAAT AAGCACATCA ACCCAACGA GTGGTGGAAAG GTGTACCCTA 2100
GCAGCGTGAC CGAGTCAAG TTCTGTTCG TGAGCGGCCA CTCAAGGGC 2150
AACTACAAGG CCCAGCTGAC CAGGCTGAAC CGCAAAACCA ACTGCAATGG 2200
CGCCGTGCTG AGCGTGGAGG AGCTGTGAT CGGCGGCGAG ATGATCAAAG 2250
CCGGCACCTT GACACTGGAG GAGGTGCGGC GCAAGTTCAA CAACGGCGAG 2300
ATCAACTTCT GATAACTCGA GTCTAGAAAT CAACCTCTGG ATTACAAAAT 2350
TTGTGAAAGA TTGACTGATA TTCTTAACTA TGTTGCTCCT TTTACGCTGT 2400
GTGGATATGC TGCTTTAATG CCTCTGTATC ATGCTATTGC TTCCCCTACG 2450

[0313]

GCTTTCGTTT TCTCCTCCTT GTATAAATCC TGGTGTGCTGT CTCITTTATGA 2500
 GGAGTTGTGG CCCGTTGTCC GTCAACGTGG CGTGGTGTGC TCTGTGTTTG 2550
 CTGACGCAAC CCCCACTGGC TGGGGCATTG CCACCACCTG TCAACTCCTT 2600
 TCTGGGACTT TCGTTTTCCC CCTCCCAGTC GCCACGGCAG AACTCATCGC 2650
 CGCCTGCCTT GCCCGCTGCT GGACAGGGGC TAGGTTGCTG GGCAGTGATA 2700
 ATTCCGTGGT GTTGTTCGGG AAATCATCGT CCTTTCCTTG GCTGCTCGCC 2750
 TGTGTTGCCA ACTGGATCCT GCGCGGGACG TCCTTCTGCT ACGTCCCTTC 2800
 GGCTCTCAAT CCAGCGGACC TCCCTTCCCG AGGCCTTCTG CCGGTTCTGC 2850
 GGCCCTCTCC GCGTCTTCGC TTTTCGGCTC CGACGAGTCG GATCTCCCTT 2900
 TGGGCGCCT CCCGCTTG CTAGCCTGTG CCTTCTAGTT GCCAGCCATC 2950
 TGTGTTTGC CCCTCCCCG TGCCTTCCTT GACCCTGGAA GGTGCCACTC 3000
 CCACTGTCCT TTCCTAATAA AATGAGGAAA TTGCATCGCA TTGTCTGAGT 3050
 AGGTGTCATT CTATTCTGGG GGGTGGGGTG GGGCAGGACA GCAAGGGGGA 3100
 GGATTGGGAA GACAAATAGCA GGCATGCTGG GGATGCGGTG GGCTCTATGC 3150
 GGCCCGCTCG AGCGC[AGGAA CCCCTAGTGA TGGAGTTGGC CACTCCCTCT 3200
 CTGCGCGCTC GCTCGCTCAC TGAGGCCGCC CGGGCTTTGC CCGGGCGGCC 3250
 TCAGTGAGCG AGCGAGCGC CAG] (SEQ ID NO:56) 3273

[0314]

[0315]

상기 제시한 ZFN 쌍 (71557/71728)을 사용하여 알부민 유전자에 삽입하기 위한 일련의 도입유전자 벡터 (F9, IDS 및 IDUA)를 만들었고, hIDS 삽입에 대한 결과를 도시하는 도 11을 참조한다. 벡터는 AAV 벡터였고 모두 ZFN 절단 부위 옆에 있는 상동성 영역을 포함하였다 (좌측 상동성 팔: LA, 및 우측 상동성 팔: RA). 벡터는 추가로 스플라이스 억셉터 서열 (SA) 및 폴리A 신호 서열 (폴리아)을 포함하였다. 마지막으로, 모두 5' 및 3' AAV ITR 서열을 포함하였다. 인자 9 엑손 2-9 AAV 도입유전자 도너의 요소 및 서열을 하기에 나타낸다.

표 6

[0316]

인자 9 AAV (SEQ ID NO:59)의 요소

위치	특징 설명	SEQ ID NO	서열
1-130	5' ITR[괄호 안의 일반 텍스트]	44	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCG GGCGACCTTTGGTCGCCCGCCTCAGTGAGCGAGCGCGCAGAGAGG GAGTGCCCAACTCCATCACTAGGGGTTCT
271-550	LA: 좌측 상동성 팔 이탤릭체	60	TTTATTCTATTTTCCAGTAAAATAAAGTTTTAGTAAACTCTGCATCTTT AAAGAATTATTTTGGCATTATTTCTAAAATGGCATAGTATTTGTATTT GTGAAGTCTTACAAGTTATCTTATTAATAAAAATTCAAACATCCTAGGTA AAAAAAAAAAGGTCAGAATTGTTTAGTGACTGTAATTTCTTTTGGCG ACTAAGGAAAGTGCAAAGTAACTTAGAGTGACTGAAACTTCACAGAATAG GGTTGAAGATTGAATTCATAACTATCCCAA
557-584	SA: 스플라이스 억셉터 진하게	61	ACTAAAGAATTATTCTTTTACATTCAG

585-1882	hF9 엑손 2-8, 코 돈 최적화된 밀줄	62	TTTTCTTGATCATGAAAACGCCAACAAAATCCTGAACCGGCCAAGCGG TACAACCTCAGGCAAGCTGGAAGAGTTTCGTGCAGGGCAACCTGGAACGGGA GTGCATGGAAGAGAAGTGCAGCTTCGAGGAAGCCCGGAGGTGTTTCGAGA ACACCGAGCGGACCACCGAGTTCTGGAAGCAGTACGTGGACGGCGACCAG TGCAGTCAAACCCCTGCCTGAACGGCGGCAGCTGCAAGGACGATATCAA CAGCTACGAGTGTGGTGCCTTCGGCTTCGAGGGCAAGAACTGCGAGC TGACGTGACCTGCAACATCAAGAACGGCCGCTGCGAGCAGTTCTGCAAG AACAGCGCCGACAACAAGGTGGTGTGCTCATGCACTGAGGGCTACCGGCT GGCCGAGAACCAGAAGAGCTGCGAGCCCGCTGCCCTTCCCTGCGGCA GAGTGTCCGTGAGCCAGACCAGCAAGCTGACCAGGGCCGAGGCCGTGTTT CCTGACGTGGACTACGTGAACCTAACCGAGGCCGAGACAATCCTGGACAA CATCACCCAGAGCACCAGTCCCTCAACGACTTCACCCGGGTGGTGGCGG CGAGGACGCCAAGCCCGCCAGTTCCTTGGCAGGTGGTGTGAACGGG AAGGTGGACGCCTTCTGCGCGGCTCAATCGTGAACGAGAAGTGGATCGT GACAGCCGCCACTGCGTGGAGACAGGCGTGAAGATCACCGTGGTGGCGG GCGAACACAATATCGAGGAAACCGAGCACACCGAGCAGAAACGGAACGTG ATCCGGATTATCCCCACCACAACCTACAACGCCCATCAACAAGTACAA CCACGATATCGCCCTGCTGGAACCTGGACGAGCCTCTGGTGTGAATTCAT ACGTGACCCCATCTGTATCGCCGACAAAGAGTACACCAACATCTTTCTG AAGTTCGGCAGCGGCTACGTGTCCGGCTGGGGCAGGGTGTTCACAAGGG CCGACAGCGCCCTGGTGTGCACTGCGGGTGCCTGGTGGACAGAG CCACCTGCCTGCGGTCAACCAAGTTCACCATCTACAACAACATGTTCTGC GCCGGCTTCCACGAGGGCGGCAGGGACAGCTGCCAGGGCGACAGCGGGG ACCCACGTGACCGAGGTGGAGGGCACCAGCTTCTGACCGCATCATCT CATGGGGCGAGGAATGCGCCATGAAGGGCAAGTACGGAATCTACACTAAG GTGTCAAGATACGTGAACCTGGATCAAAGAGAAAACCAAGCTGACCTGA
1890-2114	폴리A 소문자	63	CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCTCCCCGTGCCT TCCTTGACCTGGAAGGTGCCACTCCCAGTCTCTTCCCTAATAAAATGA GAAATTCATCGCATGCTGTGAGTAGGTGTCATTCTATTCTGGGGGGTG GGGTGGGCAGGACGCAAGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGATGCGGTGGGCTCTATGG
2121-2220	RA: 우측 상동성 팔 이중 밀줄, 물결 선	64	CTATCCATTGCACTATGCTTTATTTAAAAACCACAAAACCTGTGCTGTTG ATCTCATAAATAGAACTGTATTTATTTATTTTCATTTTGTGCTGCT
2367-2474	3' ITR[괄호 안에 진하게]	55	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGGCTCGCTCG CTACTGAGGCCCGCGGGCTTTGCCCGGGCGGCTCAGTGAGCGAGCGA GCGGCAG

[0317] F9 AAV의 서열:

```
[CTGCGCGCTC GCTCGCTCAC TGAGCCGCC CGGGCAAAGC CCGGGCGTCG
50
GGCGACCTTT GGTCGCCCGG CCTCAGTGAG CGAGCGAGCG CGCAGAGAGG
100
GAGTGGCCAA CTCATCACT AGGGTTTCCT] GCGGCCTAAG CTTGAGCGGA
150
GTTCCAATTG TACTGTACAG AACCATGGTC ACATGTTTAA CGCTAGCGTG
200
```

[0318]

CCGACCTGGT AACTGATCA GTGGGTGCAC TTAGGACTGC GTCTTACGCT
 250
 AATCATATGC GTGCGGCCGC TTTATTCTAT TTTCCAGTA AAATAAAGTT
 300
TTAGTAAACT CTGCATCTTT AAAGAATTAT TTTGGCATT ATTTCTAAAA
 350
TGGCATAAGTA TTTTGTATTT GTGAAGTCTT ACAAGGTTAT CTTATTAATA
 400
AAATTCAAAC ATCCTAGGTA AAAAAAAAAA AAGGTCAGAA TTGTTTAGTG
 450
ACTGTAATTT TCTTTTGCGC ACTAAGGAAA GTGCAAAGTA ACTTAGAGTG
 500
ACTGAAACTT CACAGAATAG GGTGAAGAT TGAATTCATA ACTATCCCAA
 550
GTACCACCTA AAGAATTATT CTTTTACATT TCAGTTTTTC TTGATCATGA
 600
AAACGCCAAC AAAATCCTGA ACCGGCCCAA GCGGTACAAC TCAGGCAAGC
 650
TGGAAGAGTT CGTGCAGGGC AACCTGGAAC GGGAGTGCAT GGAAGAGAAG
 700
TGCAGCTTCG AGGAAGCCCG GGAGGTGTTT GAGAACCACG AGCGGACCAC
 750
CGAGTTCTGG AAGCAGTACG TGGACGGCGA CCAGTGCAG TCAAACCCCT
 800
GCCTGAACGG CGGCAGCTGC AAGGACGATA TCAACAGCTA CGAGTGCTGG
 850
TGCCCCCTTCG GCTTCGAGGG CAAGAACTGC GAGCTGGACG TGACCTGCAA
 900
CATCAAGAAC GGCCGCTGCG AGCAGTTCTG CAAGAACAGC GCCGACAACA
 950
AGTGGTGTG CTCATGCACT GAGGGCTACC GGCTGGCCGA GAACCAGAAG
 1000
AGCTGCGAGC CCGCCGTGCC CTTCCCCTGC GGCAGAGTGT CCGTGAGCCA
 1050
GACCAGCAAG CTGACCAGGG CCGAGGCCGT GTTCCCTGAC GTGGACTACG
 1100
TGAACTCAAC CGAGGCCGAG ACAATCCTGG ACAACATCAC CCAGAGCACC
 1150
CAGTCCTTCA ACGACTTCAC CCGGGTGGTG GCGGCGAGG ACGCCAAGCC
 1200
CGGCCAGTTC CCTTGCAGG TGGTGTGAA CGGCAAGGTG GACGCCTTCT
 1250
GCGGCGGCTC AATCGTGAAC GAGAAGTGA TCGTGACAGC CGCCCACTGC
 1300
GTGGAGACAG GCGTGAAGAT CACCGTGGTG GCCGGCGAAC ACAATATCGA
 1350
GGAAACCGAG CACACCGAGC AGAAACGGAA CGTGATCCGG ATTATCCCC
 1400
ACCACAATA CAACGCCGCC ATCAACAAGT ACAACCACGA TATCGCCCTG
 1450

[0319]

CTGGAACTGG ACGAGCCTCT GGTGCTGAAT TCATACGTGA CCCCCATCTG
 1500
TATCGCCGAC AAAGAGTACA CCAACATCTT TCTGAAGTTC GGCAGCGGCT
 1550
ACGTGTCCGG CTGGGGCAGG GTGTTCCACA AGGGCCGAG CGCCCTGGTG
 1600
CTGCAGTACC TGCGGGTGCC CCTGGTGGAC AGAGCCACCT GCCTGCGGTC
 1650
AACCAAGTTC ACCATCTACA ACAACATGTT CTGCGCCGGC TTCCACGAGG
 1700
GCGGCAGGGA CAGCTGCCAG GGCACAGCG GCGGACCCCA CGTGACCGAG
 1750
GTGGAGGGCA CCAGCTTTCT GACCGGCATC ATTCATGGG GCGAGGAATG
 1800
CGCCATGAAG GGCAAGTACG GAATCTACAC TAAGGTGTCA AGATACGTGA
 1850
ACTGGATCAA AGAGAAAACC AAGCTGACCT GAGTTTAAAc tgtgccttct
 1900
 agttgccagc catctgttgt ttgcccctcc cccgtgcctt ccttgaccct
 1950
 ggaagtgcc actcccactg tcctttccta ataaaatgag gaaattgcat
 2000
 cgcattgtct gagtaggtgt cattctattc tgggggggtgg ggtggggcag
 2050
 gacagcaagg gggaggattg ggaagacaat agcaggcatg ctgggggatgc
 2100
 ggtgggctct atggACCGGT CTATCCATTG CACTATGCTT TATTTAAAAA
 2150
CCACAAAACC TGTGCTGTTG ATTCATAAA TAGAACTTGT ATTTATATTT
 2200
ATTTTCATTT TAGTCTGTCT GGATCCACAA ATTAATCGAA CCTGCAGCTG
 2250
 ATATCGACGC TTAAGTAGGG CTTAGCAAAC GCGTCTCCAA CGTTTCGCCG
 2300
 TTAACACCCC ACATAGTGAG TGGTCTTAGT AGTCCGGGTG TTAAACTGA
 2350
 AAGATAACTC GAGCGC [AGGA ACCCCTAGTG ATGGAGTTGG
CCACTCCCTC2400
TC'TGCGCGCT CGCTCGCTCA CTGAGGCCGC CCGGGCTTTG CCCGGGCGGC
 2450
CTCAGTGAGC GAGCGAGCGC GCAG] (SEQ ID NO:59)
 2474

[0320]

[0321]

IDS AAV 도입유전자 도너의 요소 및 서열을 하기에 나타낸다:

표 7

IDS AAV (SEQ ID NO:65)의 요소

[0322]

위치	특징 설명	SEQ ID NO	서열
1-130	5' ITR[괄호 안의 일반 텍스트]	44	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCCG GGCGACCTTGGTCGCCCGGCTCAGTGAGCGAGCGAGCGCAGAGAGG GAGTGGCCAACTCCATCACTAGGGGTTCTT
271-550	LA: 좌측 상동성 팔 이탤릭체	60	TTTATCTATTTTCCAGTAAAATAAAGTTTATAGTAAACTCTGCATCTTT AAAGAATTATTTTGGCATTTATTTCTAAAATGGCATAGTATTTGTATTT GTGAAGTCTTACAAGGTATCTTATTAATAAAAATCAAACATCTAGGTA AAAAAAAAAAAAAGTCAAGTGTAGTACTGTAATTTCTTTTGGCG ACTAAGGAAAGTGCAAAGTAACTTAGAGTACTGAACTTACAGAATAG GGTTGAAGATTGAATTCATAACTATCCCAA
557-584	SA: 스플라이스 엑스온 진하게	61	ACTAAGAATTATTTTACATTTTCAG

587-2161	hIDS, 코돈 최적화 된 밑줄	66	AGCGAAACCCAGGCCAACTCAACTACAGATGCGCTTAACGTCTGCTCAT CATCGTGGACGATTTGCGGCCGTCGCTTGGCTGCTATGGAGATAAGCTCG TCCGCTCGCCGAACATCGATCAGTTGGCCTCACACTACTGCTTTTCCAA AATGCGTTTGGCGCAGCAGGCTGCTGTGACCTTCAAGAGTCTCATTCTT GACCGGGCAGCCCTGACACAACCGCGCTGTACGACTTCAACAGCTACT GGAGAGTCCACGCGGGTAACTTTCAACTATCCCACAGTACTTTAAAGAG AACGATACGTGACAATGAGCGTGGGAAAGTCTTTCACCCCGCATCTC CTCGAATCACACCGACGATTCGCCCTACTCGTGGTCGTTTCTCCCTACC ATCCTTCGAGCGAGAAGTATGAGAACACGAAAACCTGTGCGGACCCGAC GGAGAGCTGCACGCTAATCTGCTGTGTCCGGTGGATGCTTGGACGTGCC CGAGGGAACGCTCCCCGACAAGCAGTCAACGGAGCAGGCGATTCAAGTTC TGGAGAAGATGAAAACAAGCGCGTCGCTTTCTTCTCGCCGTGGGGTAT CACAAGCCCATATTCTTTCCGCTACCCGAAGGAGTTCAGAAAATTTA TCCTTTGGAAAACATCACTTTGGCACCGGACCCGGAAGTCCAGACCGTC TGCCACCCGTGGCCTACAATCCCTGGATGGATATCAGGCAGAGGGAAGAT GTGCAGGCACTCAACATCTCAGTCCCCTACGGGCTATTCCAGTCGATTT TCAACGCAAGATTGCGCAGTCGATTTTGGCTCGGTGTCTACTCGATA CGAAGTAGGTCGACTTCTGAGCGCGCTTGATGACCTTCAGCTGGCAAAT TCCACAATCATCGCTTTACGTCGGACCATGGGTGGGCGTTGGGAGAGCA TGGAGAGTGGCAAAGTATAGCAATTTGATGTAGCAACGCACGTGCCCC TGATTTTCTACGTGCCGGGTAGAACGGCCTCGCTTCCCGAGGCAGCGGAA AACTTTTTCCCTATCTCGATCCATTGACTCGGCGAGCCAGCTTATGGA ACCGGCAGACAATCCATGGACTTGGTAGAATTGGTGTCCCTTTTTCCGA CCCTCGCCGGTTGGCGGGCTTGCAAGTACCCCTAGATGCCCTGTACCG AGCTTCCATGTGGAACCTGCGCGAAGGGAAAAACCTCCTCAAACACTT TCGGTTCAGGGACCTTGAGGAGGACCCCTATCGCCAGGGAATCCGCGAG AGTTGATTGCCTATTCCAGTATCCGCGACCCAGCGATATTCTCAATGG AACTCCGATAAAGCCCTCCCTCAAAGACATCAAGATTATGGGGTACTCGAT CAGGACCATCGACTATCGCTACACAGTGTGGGTAGGGTTCAATCCTGACG AATCCTCGGAACTTTTCGGACATCCACGCTGGTGAGCTGATTTTCGTA GACTCGGACCCGTGCAAGATCACAATATGTATAATGATTTCCAAGGAGG AGATTTGTCCAGCTGCTCATGCCG
2174-2398	폴리A 소문자	63	CTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCT TCCTTGACCCTGGAAGGTGCCACTCCCCTGTCCTTTCTTAATAAAATGA GGAATTCATCGCATGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTG GGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGATGCGGTGGGCTCTATGG
2405-2504	RA: 우측 상동성 팔 이중 밑줄, 물결선	64	CTATCCATTGCACTATGCTTTATTTAAAAACACAAAACCTGTGCTGTTG ATCTCATAAATAGAACTTGTATTTATATTTATTTTCAATTTTAGTCTGTCT
2651-2758	3' ITR[괄호 안에 진하게]	55	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCG CTCACTGAGGCCGCCGGCTTTGCCGGGGCGCTCAGTGAGCGAGCGA GCGCGCAG

[0323] hIDS AAV의 서열:

```
[CTGCGCGCTC GCTCGCTCAC TGAGGCCGCC CGGGCAAAGC CCGGGCGTCG 50
GGCGACCTTT GGTGCGCCCG CCTCAGTGAG CGAGCGAGCG CGCAGAGAGG 100
GAGTGGCCAA CTCCATCACT AGGGGTTCCT]GCGGCCTAAG CTGAGCGGGA 150
GTTCCAATTG TACTGTACAG AACCATGGTC ACATGTTTAA CGCTAGCGTG 200
CCGACCTGGT AAACGTATCA GTGGGTGCAC TTAGGACTGC GTCTTACGCT 250
AATCACATGC GTGCGGCCGC TTTATTCTAT TTTCCCAGTA AAATAAAGTT 300
TTAGTAAACT CTGCATCTTT AAAGAATTAT TTTGGCATT AATTCTAAAA 350
TGGCATAGTA TTTTGTATTT GTGAAGTCTT ACAAGGTTAT CTATTAATA 400
AAATTCAAAC ATCCTAGGTA AAAAAAAAAA AAGGTCAGAA TTGTTTAGTG 450
ACTGTAATTT TCTTTTGGCG ACTAAGGAAA GTGCAAAGTA ACTTAGAGTG 500
ACTGAAACTT CACAGAATAG GGTGAAGAT TGAATTCATA ACTATCCCAA 550
GGTACCACCTA AAGAATTATT CTTTACATT TCAGTTAGCG AAACCCAGGC 600
CAACTCAACT ACAGATGCGC TTAACGTCCT GCTCATCATC GTGGACGATT 650
TGCGGCCGTC GCTTGGCTGC TATGGAGATA AGCTCGTCCG CTCGCCGAAC 700
ATCGATCAGT TGGCCTCACA CTCACTGCTT TTCAAAATG CGTTTGCGCA 750
GCAGGCTGTC TGTGCACCTT CAAGAGTCTC ATTCTTGACC GGGCGACGCC 800
CTGACACAAC GCGGTGTAC GACTTCAACA GCTACTGGAG AGTCCACGCG 850
GGTAACTTTT CAACTATCCC ACAGTACTTT AAAGAGAAGC GATACGTGAC 900
AATGAGCGTG GGAAGGTCCT TTCACCCCGG CATCTCCTCG AATCACACCG 950
ACGATTCGCC CTACTCGTGG TCGTTTCCCTC CCTACCATCC TTCGAGCGAG 1000
AAGTATGAGA ACACGAAAAC TTGTGCGGGA CCCGACGGAG AGCTGCACGC 1050
TAATCTGCTG TGTCGGGTGG ATGTCTTGA CGTGCCCGAG GGAACGCTCC 1100
CCGACAAGCA GTC AACGGAG CAGGCGATTC AGTTGCTGGA GAAGATGAAA 1150
ACAAGCGCGT CCCTTTCCTT CCTCGCCGTG GGGTATCACA AGCCCATAT 1200
TCCTTTCCCG TACCCGAAGG AGTTCAGAA ACTTTATCCT TTGAAAACA 1250
TCACTTTGGC ACCGACCCG GAAGTCCCGG ACGGTCTGCC ACCGTGGCC 1300
TACAATCCCT GGATGGATAT CAGGCAGAGG GAAGATGTC AGGCACTCAA 1350
CATCTCAGTC CCCTACGGGC CTATTCAGT CGATTTTCAA CGCAAGATTC 1400
GGCAGTCGTA TTTTGCCTCG GTGTCTACC TCGATACGCA AGTAGGTGGA 1450
CTTCTGAGCG CGCTGATGA CCTTCAGTG GCAAATCCA CAATCATCGC 1500
CTTTACGTCG GACCATGGGT GGGCGTTGGG AGAGCATGGA GAGTGGGCAA 1550
AGTATAGCAA TTTTGTATTA GCAACGCACG TGCCCTGAT TTTCTACGTG 1600
CCGGGTAGAA CGGCCTCGCT TCCCGAGGCA GGCGAAAAC TTTTCCCTA 1650
TCTCGATCCA TTGACTCGG CGAGCCAGCT TATGGAACCG GCGAGACAAT 1700
CCATGGACTT GGTAGAATTG GTGTCCCTTT TTCCGACCTT CGCCGGGTTG 1750
GCGGGCTTGC AAGTACCCCT TAGATGCCCT GTACCCGAGT TCCATGTGGA 1800
ACTCTGCCG GAAGGGAAA ACCTCCTCAA AACTTTCCG TTCAGGGACC 1850
TTGAGGAGGA CCCCTATCTG CCAGGGAATC CGCGAGAGTT GATTCCTAT 1900
TCCCAGTATC CGCAGCCAG CGATATTCCT CAATGGAAT CCGATAAGCC 1950
CTCCCTCAA GACATCAAGA TTATGGGGTA CTCGATCAGG ACCATCGACT 2000
ATCGCTACAC AGTGTTGGTA GGGTTCAATC CTGACGAAT CCTCGCGAAC 2050
TTTTCGGACA TCCACGCTGG TGAGCTGTAT TTCGTAGACT CGGACCCGTT 2100
GCAAGATCAC AATATGTATA ATGATTCCCA AGGAGGAGAT TTGTTCCAGC 2150
TGCTCATGCC GTGATAAAGA TCTctgtgcc ttctagttgc cagccatctg 2200
ttgtttgccc ctccccgtg ccttccttga ccctggaagg tgccactccc 2250
actgtccttt cctaataaaa tgaggaaatt gcatcgatt gtctgagtag 2300
```

[0324]

```
gtgtcattct attctggggg gtgggggtggg gcaggacagc aagggggagg 2350
attgggaaga caatagcagg catgctgggg atgcgggtggg ctctatggAC 2400
CGTCTATCC ATGCACTAT GCTTATTTA AAAACCACAA AACCTGTGCT 2450
GTTGATCTCA TAAATAGAAC TTGTATTTAT ATTTATTTTC ATTTAGTCT 2500
GTCTGGATCC ACAAATTAAT CGAACCTGCA GCTGATATCG ACCTTAAGT 2550
AGGGCTTAGC AAACGCTCT CCAACGTTTC GCCGTTAACA CCCACATAG 2600
TGAGTGGTCT TAGTAGTCCG GGTGTTTAAA CTGAAAGATA ACTCGAGCGC 2650
[AGGAACCCCT AGTGATGGAG TTGGCCACTC CCTCTCTGCG CGCTCGCTCG 2700
CTCACTGAGG CCGCCCGGGG TTTGCCCGGG CGGCCTCAGT GAGCGAGCGA 2750
GCGCGCAG] (SEQ ID NO:65) 2758
```

[0325]

[0326] IDUA AAV 도입유전자 도너의 요소 및 서열을 하기에 나타낸다:

표 8

[0327]

IDUA AAV (SEQ ID NO:72)의 요소

위치	특징 설명	SEQ ID NO	서열
1-130	5' ITR[팔호 안의 일반 텍스트]	44	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGCTCG GGCGACCTTTGGTGCGCCCGCTCAGTGAGCGAGCGAGCGCGAGAGAGG GAGTGGCCAACCTCCATCACTAGGGGTTCT

271-550	LA: 좌측 상동성 팔 이탈릭체	60	TTTATTCTATTTTCCCAGTAAAATAAAGTTTTAGTAAACTCTGCATCTTT AAAGAATTATTTTGGCATTATTTCTAAAATGGCATAGTATTTTGTATTT GTGAAGTCTTACAAGGTTATCTTATTAATAAAAATTCAAACATCCTAGGTA AAAAAAAAAAAGGTCAGAATTGTTTTAGTGACTGTAATTTTCTTTTGGCG ACTAAGGAAAGTGCAAAGTAACTTAGAGTGACTGAAACTTCACAGAATAG GGTTGAAGATTGAATTCATAACTATCCCAA
557-584	SA: 스플라이스 억셉터 진하게	61	ACTAAAGAATTATTCCTTTTACATTCAG
587-2458	hIDUA, 코돈 최 적화된 밑줄	67	CACTTGGTCCACGTCGACGCTGCCAGAGCCCTGTGGCCGCTTCGAAGATT TTGGAGGTCAACGGGTTTCTGTCTCCCCTTCCCACCTCGAAGCAGATC AGTATGTAAGTCTATGGGATCAACAGCTTAACCTCGCCTATGTCCGAGCA GTGCCTCACC CGGGATCAAGCAAGTAAGGACACATTGGCTCCTGAACT CGTACCACGAGAGGATCGACGGGAAGGGGGCTTCGTACAACCTCACTC ATCTCGATGGCTATTTGGATCTCTCCCGGAGAATCAGTTGTGCCAGGC TTCGAATTGATGGGATCGGCGAGCGGGCACTTACAGACTTCGAGGACAA GCAGCAAGTGTGAGTGAAGGACCTCGTGTCTGCTCGCGAGGAGAT ACATTGGTCGCTACGGTTTGGCGCATGTGTCAAAGTGGAACCTCGAAACG TGGAACGAGCCGATCATCACGATTTTGACAACGTGCAATGACCATGCA GGGTTTCCTTAACTATTACGACGCTGTTCGAGGGATTGAGGGCAGCAT CACCGCGCTTCGGTGGGAGGGCTGGTGATAGCTTTCATACACCACCT CGATCGCCACTTTCGTGGGGGCTGCTGCGCCATTGTACAGATGGTACGAA CTTCTTACC GGGAAGCGGGGTACGGCTTGATTACATCAGCCTCCACC GAAAGGGAGCGCGTCAAGCATCTCGATTCTGGAGCAGGAGAAGGTAGTC GCTCAGCAGATCCGCAACTCTTCCCAAGTTCGACAGACACCTATCTA CAATGATGAGGCAGACCCACTTGTGGGATGGTCCCTTCCCGACCATGGC GCGCAGATGTGACTTATGCCGCGATGGTAGTAAAAGTCATCGCCAGCAC CAGAATCTGCTTCTTGAATAACGACCAGCGCTTTCCTTACGCGCTTTT GTGGAACGATAATGCCTTCTGTGATATCACCCCATCCGTTTGGCAGAG GGACTCTTACGGCGGATTCCAAGTGAATAACACCAGACCGCCGACGCTG CAGCTGTGCGAAAACCCGTGTGACTGCGATGGGGCTTCTGGCGTTGCT TGATGAGGAACAACCTCTGGGCTGAAGTGTCCAGGGGGGACAGTACTTG ATAGCAATCATACAGTAGGCGTGTGGCGTCCGCGCACCCGACCGAGGA CCCGCGGATGCTTGGAGGGCAGCGGTCTGATCTACGCTCGGACGATAC TAGGGCACATCCAACAGATCGGTCGCTGTACCCTTCGCCTCAGAGGGG TCCCGCTGGTCCCGCTTGGTATACGTCAGTATCTCGACAAATGGA CTGTGCAGCCCGACGGAGAGTGGCGGAGGCTGGGACGGCCGTTTCC GACAGCCGAGCAGTTTAGACGGATGAGGGCCGCTGAGGACCCCGTGGCAG CGCACCGAGGCCCTCCCGGACGAGGTCGCCTCACTCTTCGACCGGCA CTCGCGTCCGCTCCCTTCTGCTCGTACAGTCTGCGCGGACCCGAAAA GCCGCTGGACAGGTAACCAGGCTCAGGGCGCTCCCTTGACGACGGGGC AGTTGGTACTTGTCTGGTCGACGAACAGTGGGTCCAAATGCTTGTGG ACGTATGAAATTCAGTTTCCCAAGACGGGAAAGCGTACACTCCCGTGTG GCGAAACCCTCCACGTTCAACCTTTCGTTTTCCTCCAGACACGGGAG CCGTATCAGGGTCGTACCGAGTCAGAGCCCTCGATTATTGGGCGAGGCCT GGGCGTTCCTGGACCCTGTACCATACTTGAAGTGCCGGTGCACAGGGG ACCGCCCTCGCTGGTAATCCT
2471-269 5	폴리A 소문자	63	CTGTGCCTTCTAGTTGCCAGCCATCTGTGTTTGCCCTCCCCCGTGCCT TCCTTGACCCTGGAAGGTGCCACTCCCCTGCTCTTCTTAATAAAAATGA GGAATTCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGTG GGTGGGCGAGGACGCAAGGGGAGGATTGGGAAGACAATAGCAGGCAT GCTGGGGATCGGGTGGCTCTATGG
2702-280 1	RA: 우측 상동성 팔 이중 밑줄, 물결 선	64	CTATCCATTGCACTATGCTTTATTTAAAAACCACAAAACCTGTGCTGTTG ATCTCATAAATAAGAACTTGTATTTATTTATTTTTCATTTTGTCTGTCT
2948- 3055	3' ITR[팔호 안 에 진하게]	55	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCG CTCACTGAGGCCGCCGGGCTTTGCCGGGGCGGCTCAGTGAGCGAGCGA GCGCGCAG

[0328] IDUA AAV의 서열:

[CTGCGCGCTC GCTCGCTCAC TGAGGCCGCC CGGGCAAAGC CCGGGCGTCC 50
 GCGGACCTTT GTTCGCCCGG CCTCAGTGAG CGAGCGAGCG CGCAGAGAGG 100
 GAGTGGCCAA CTCCATCACT AGGGGTTCCT] GCGGCCTAAG CTTGAGCGGA 150
 GTTCCAATTG TACTGTACAG AACCATGGTC ACATGTTTAA CGCTAGCGTG 200
 CCGACCTGGT AACTGATCA GTGGGTGCAC TTAGGACTGC GTCTTACGCT 250
 AATCACATGC GTGCGGCCGC TTTATTCTAT TTTCCAGTA AAATAAAGTT 300
 TTAGTAAACT CTGCATCTTT AAAGAATTAT TTTGGCATTT ATTTCTAAAA 350
 TGGCATAGTA TTTTGTATTT GTGAAGTCTT ACAAGGTAT CTTATTAATA 400
 AAATTCAAAC ATCTAGGTA AAAAAAAAAA AAGGTCAGAA TTGTTTAGTG 450
 ACTGTAATTT TCTTTTGC GC ACTAAGGAAA GTGCAAAGTA ACTTAGAGTG 500
 ACTGAAACTT CACAGAATAG GGTGAAGAT TGAATTCATA ACTATCCCAA 550
 GGTACCACCTA AAGAATTATT CTTTTACATT TCAGCGCACT TGGTCCACGT 600
 CGACGCTGCC AGAGCCCTGT GGCCGCTTCG AAGATTTTGG AGGTCAACGG 650
 GTTCTGTGCC TCCCCTTCCC CACTCGCAAG CAGATCAGTA TGTACTGTCA 700
 TGGGATCAAC AGCTTAACCT CGCCTATGTC GGAGCAGTGC CTCACGCGCG 750
 GATCAAGCAA GTAAGGACAC ATTGGCTCCT TGAACCTGTC ACCACGAGAG 800
 GATCGACGGG AAGGGGGCTT TCGTACAACT TCACTCATCT CGATGGCTAT 850
 TTGGATCTCC TCCGCGAGAA TCAGTTGTTG CCAGGCTTCG AATTGATGGG 900
 ATCGGCGAGC GGGCACTTTA CAGACTTCGA GGACAAGCAG CAAGTGTGTTG 950
 AGTGGAAAGG CCTCGTTCG TCGCTCGCGA GGAGATACAT TGGTCGCTAC 1000
 GGTGTGGCGC ATGTGTCAA GTGGAACCTT GAAACGTGGA ACGAGCCCGA 1050
 TCATCAGGAT TTTGACAACG TGTCAATGAC CATGCAGGGT TTCTTAACCT 1100
 ATTACGACGC CTGTTCCGAG GGATTGAGGG CAGCATCACC GCGGCTTCGG 1150
 CTGGGAGGGC CTGGTGATAG CTTTCATACA CCACCTCGAT CGCCACTTTC 1200
 GTGGGGGCTG CTGCGCCATT GTCACGATGG TACGAACTTC TTCACGCGGG 1250
 AAGCGGGGGT ACGGCTTGAT TACATCAGCC TCCACCAGAA GGGAGCGCGG 1300
 TCAAGCATCT CGATTCTGGA GCAGGAGAAG GTAGTCGCTC AGCAGATCCG 1350

[0329]

GCAACTCTTT CCAAGTTCG CAGACACACC TATCTACAAT GATGAGGCAG 1400
 ACCCACTTGT GGGATGGTCC CTTCGCGAGC CATGGCGCGC AGATGTGACT 1450
 TATGCCGCGA TGGTAGTGAA AGTCATCGCC CAGCACCAGA ATCTGCTTCT 1500
 TGCGAATACG ACCAGCGCGT TTCTTACGC GCTTTGTGTC AACGATAATG 1550
 CCTTCCTGTC ATATCACCCC CATCCGTTTG CGCAGAGGAC TCTTACGGCG 1600
 CGATTTCCAAG TGAATAACAC CAGACCGCCG CACGTGCAGC TGTTCGGA 1650
 ACCCGTGTG ACTGCGATGG GGCTTCTGGC GTTGCTTGAT GAGGAACAAC 1700
 TCTGGGCTGA AGTGTCCAG GCGGGACAG TACTTGATAG CAATCATA 1750
 GTAGGCGTGT TGGCGTCGGC GCACCGACCG CAGGGACCCG CCGATGCTTG 1800
 GAGGGCAGCG GTCCTGATCT ACGCCTCGGA CGATACTAGG GCACATCCCA 1850
 ACAGATCGGT CGCTGTCACC CTTTCCTCA GAGGGTCCG GCCTGGTCCC 1900
 GGCTTGGTAT ACGTCACTAG ATATCTCGAC AATGGACTGT GCAGCCCGA 1950
 CGGAGAGTGG CGGAGCTGG GACGGCCGGT GTTTCGACA GCCGAGCAGT 2000
 TTAGACGGAT GAGGGCCGCT GAGGACCCCG TGGCAGCGGC ACCGAGGCC 2050
 CTCCCGCAG GAGCTCGCCT CACTCTTCGA CCGGCACTGC GGCTGCCGTC 2100
 CCTTCTGCTC GTACACGCTC GCGCGGACC CGAAAAGCCG CCTGGACAGG 2150
 TAACCAAGCT CAGGGCGCTC CCCTTGACGC AGGGGCACTT GGTACTTGTG 2200
 TGGTCGGAGC AACACGTGGG GTCCAAATGC TTGTGGACGT ATGAAATTCA 2250
 GTTTTCCCAA GACGGGAAAG CGTACACTCC GGTGTGCGCG AAACCCCTCA 2300
 CGTTCAACCT CTTCGCTTTT TCCCAGACA CGGGAGCCGT ATCAGGGTCCG 2350
 TACCGAGTCA GAGCCCTCGA TTATTGGGCG AGGCCTGGGC CGTTCTCGGA 2400
 CCCTGTACCA TACTTGAAG TGCCGGTGCC CAGGGGACCG CCCTCGCCTG 2450
 GTAATCCTTG ATAAAGATCT ctgtgccttc tagttgccag ccatctgttg 2500
 tttgcccctc ccccggtgct tccttgacct tggaaaggtgc cactcccact 2550
 gtoccttctc aataaaatga ggaaattgca tcgcattgtc tgagtaggtg 2600
 tcattctatt ctggggggtg ggggtgggca ggacagcaag ggggaggatt 2650
 ggggaagacaa tagcaggcat gctggggatg cgggtgggctc tatggACCGG 2700
 TCTATCCATT GCACTATGCT TTATTTAAAA ACCACAAAAC CTGTGCTGTT 2750
 GATCTCATAA ATAGAATTG TATTTATATT TATTTTCATT TTAGTCTGTC 2800
 TGGATCCAA AATTAATCGA ACCTGCAGT GATATCGACG CTTAAGTAGG 2850
 GCTTAGCAAA CGCTCTCCA ACGTTTCGCC GTTAACACCC CACATAGTGA 2900
 GTGGTCTTAG TAGTCCGGGT GTTTAAACTG AAAGATAACT CGAGCGC [AGG 2950
 AACCCCTAGT GATGGAGTTG GCCACTCCCT CTCTGCGCGC TCGCTCGCTC 3000
 ACTGAGGCCG CCCGGGCTTT GCCCGGGCGG CCTCAGTGAG CGAGCGAGCG 3050
 CGCAG] (SEQ ID NO:72) 3055

[0330]

[0331] 알부민-특이적 71557/71728 쌍을 사용하여, 도입유전자 도너 AAV (AAV-F.IX, AAV-IDS, AAV-IDUA) 중 하나의 존재하에, 도입유전자가 상동성 지시된 표적화된 통합을 통해 알부민 유전자좌로 통합되도록 표적 인간 세포에서 알부민 유전자좌를 절단한다. 통합시, 도입유전자의 발현은 알부민 프로모터에 의해 조절된다.

[0332] 71557/71728 ZFN을 표준 프로토콜에 따라 HepG2에서 알부민 유전자좌를 절단하고 FIX 도입유전자를 삽입하는 것에 대해 47171/47898 ZFN 쌍과 비교하였다. 간단히 설명하면, HepG2 세포를 ZFN 로트를, 각각 총 부피 500 μL 중의 1.25 x 10⁶ vg/세포의 MOI에서 사용하여 형질유도하였다. 세포를 37°C/5% CO₂ 인큐베이터에서 밤새 (12 내

지 24시간) 인큐베이션하였다. 다음날, FIX 도너를 총 부피 500 μ L 중의 2.5×10^6 vg/세포의 MOI에서 FIX 도너로 형질유도하였다. 9일 후에, 배지를 FIX 단백질에 대해 ELISA에 의해 VisuLize FIX 항원 키트 (Affinity Biologicals)를 제조사의 설명서를 따라 사용하여 테스트하였다.

[0333] 그 결과 (도 10C 참조)는 이 검정에서, 71557/71728 ZFN을 사용하는 FIX 도너의 삽입이 47171/47898 쌍을 사용했을 때보다 거의 3배 더 많은 FIX 생성을 초래한 것을 나타낸다.

[0334] 유사한 방식으로, 표준 프로토콜에 따라 HepG2 세포에서 알부민 유전자좌를 절단하고 IDUA 도입유전자를 삽입하는 것에 대해 71557/71728 ZFN을 47171/47898 ZFN 쌍과 비교하였다. 그 결과 (도 14)는 두 쌍 모두 알부민 유전자좌로의 IDUA 도입유전자의 ZFN-지시된 표적화된 통합을 유발할 수 있고, 도입유전자는 IDUA 활성이 세포 상층액에 존재하도록 발현될 수 있음을 입증한다.

[0335] **실시예 7: 생체내 절단 및 표적화된 통합**

[0336] 본원에 기술된 구성물을 또한 생체내에서 테스트하였다.

[0337] **동물 연구 설계**

[0338] 적어도 생후 6 내지 8주령의 42마리의 수컷 야생형 C57BL/6 마우스를 Charles River Laboratories, Inc., Wilmington, MA로부터 구입하였다. 마우스 취급, 주입 및 샘플 수집을 Experimur (Chicago IL)에 의해 축산과 관련된 표준 프로토콜을 따라 수행하였다. 마우스를 처리 개시 전 적어도 1주 동안 Experimur의 검역소에서 보관하였다.

[0339] AAV를 Sangamo Therapeutics에 의해 준비하고 -70°C 에서 사용할 때까지 받은 그대로 보관하였다. 5개의 조작된 AAV2/6 벡터를 이 연구에 사용하였다; 표준 구조를 가지는 2개의 ZFN을 암호화하는 2개의 AAV 벡터 ("ZFN 표준"); 개선된 구조를 가지는 2개의 ZFN을 암호화하는 2개의 AAV 벡터 ("ZFN 개선된", 예컨대, ZFN 표준 + 5'UTR, 3xFLAG 및 WPREmut6); 및 마우스 알부민 인트론 1 상동성 팔이 옆에 있는 프로모터 없는 hIDS 도입유전자 DNA 주형 (도너)을 암호화하는 1개의 AAV 벡터.

[0340] AAV2/6 벡터를 표 1에 제시한 용량으로 제제 완충액 (35 mM NaCl 및 5% 글리세롤이 보충된 PBS [pH 7.1])으로 희석하였다. 생후 6주 내지 9주령의 마우스를 그룹 1 내지 7에 무작위로 배정하였다. 그룹 1의 마우스를 비히클, 즉, 제제 완충액으로 정맥내 주사하였고, 그룹 2-7의 마우스를 하기 표 9에 제시한 대로 상이한 용량의 벡터 조합으로 정맥내 주사하였다. 총 용량 부피 주사는 마우스당 200 μ L였다.

표 9

그룹 표시 및 용량

[0341]

그룹	그룹 표시	동물의 수	각각의 ZFN 용량 수준 (vg/마우스)	도너 용량 수준 (vg/마우스)
1	제제 완충액	6마리	0	0
2	ZFN 표준 저용량	6마리	2.00E+10	1.60E+11
3	ZFN 표준 중간 용량	6마리	6.00E+10	4.80E+11
4	ZFN 표준 고용량	6마리	2.00E+11	1.60E+12
5	ZFN 개선된 저용량	6마리	2.00E+10	1.60E+11
6	ZFN 개선된 중간 용량	6마리	6.00E+10	4.80E+11
7	ZFN 개선된 고용량	6마리	2.00E+11	1.60E+12

[0342] **AAV 벡터 구성물 및 포장**

[0343] FokI 도메인 1에 강제 헤테로다이머 ELD/KKR 돌연변이를 함유한 마우스 알부민 유전자좌의 인트론 1을 표적화하는 헤테로다이머 ZFN. 마우스 생체내 연구를 위해, 표준 ZFN (48641 및 31523) 또는 개선된 ZFN (5'UTR, N-말단 3xFLAG 및 3'WPREmut6을 가진 48641 및 31523)을 사용하였다. 인간 시험관내 연구를 위해, 표준 ZFN (47171 및 47898) 또는 ZFN2.0 (5'UTR, N-말단 3xFLAG 및 3'WPREmut6을 가진 71557 및 71728)을 사용하였다. hIDS 도너 구성물은 이전에 기술되었다 (Sharma *et al* (2015) *Blood* 126, 1777-1784). hIDS 도너 구성물은 내인성 IDS 신호 펩타이드가 결핍된 hIDS cDNA, hF9 스플라이스 억셉터 서열, 및 대략 총 600 bp 길이의 마우스 또는 인간 알부민 표적 부위에 대한 상동성 팔을 함유한다. 재조합 AAV2/6 벡터 (AAV2 ITR 및 AAV6 캡시드로 구성됨)를 10-챔버 CELLSTACK 배양 챔버 (Corning)에서 293 세포의 삼중 트랜스펙션에 의해 제조하고, 염화 세슘밀도 구배

원심분리와 이어서 투석에 의해 정제하고, 이전에 기술된 것과 같이 적정하였다 (Sharma, 상기 동일).

[0344] 조직 수집

[0345] 마우스를 3분 동안 2 L/분의 유량의 CO₂ 흡 챔버에서 56일 제에 안락사시켰다. 간 샘플을 수집하고 3부분으로 해부하였다: 한 부분은 조직학적 분석을 위해 10% 중성-완충 포르말린에 넣고 나머지 부분은 순간 냉동시켜서 IDS 효소 활성, RNA 추출, 웨스턴 블롯팅 및 유전자 변형의 평가를 위해 처리할 때까지 -70°C에서 보관하였다.

[0346] 차세대 서열분석에 의한 간에서의 Indel 검출

[0347] 마우스 간 샘플로부터의 게놈 DNA를 AllPrep DNA/RNA/단백질 미니 키트 (Qiagen)를 제조사의 프로토콜을 따라 사용하여 추출하였다. iPS-유래 간세포 gDNA를 QIAamp DNA 마이크로 키트 (Qiagen)를 제조사의 프로토콜을 따라 사용하여 추출하였다. ZFN 표적 부위를 표 10에 기술한 프라이머를 사용하여 PCR에 의해 증폭시켰다. PCR 생성물을 MiSeq (Illumina)를 사용하여 서열분석하고 이전에 기술된 것과 같이 분석하였다 (Laoharawee, K. *et al.* (2018) *Mol. Ther.* 26, 1127-1136).

표 10

MiSeq 분석을 위한 프라이머

[0348]

표시	서열	SEQ ID NO:
마우스 Alb 전방향	ACACGACGCTCTCCGATCTNNNNTTGAGTTTGAATGCACAGAT	34
마우스 Alb 역방향	GACGTGTGCTCTCCGATCTNNNNGAACAGGGAGAGAAAAACC	35
인간 Alb 전방향	ACACGACGCTCTCCGATCTNNNNGCACTAAGGAAAGTGCAAAG	36
인간 Alb 역방향	GACGTGTGCTCTCCGATCTNNNNAACCAAGAAGACAGACTAAAATG	37

[0349] RT-qPCR

[0350] 간 샘플로부터의 총 RNA를 AllPrep DNA/RNA/단백질 미니 키트 (Qiagen)를 제조사의 프로토콜을 따라 사용하여 추출하였다. cDNA를 SuperScript™ III 제1-가닥 합성 슈퍼믹스 (Thermo Fisher Scientific)를 제조사의 프로토콜을 따라 사용하여 생성하였다. qPCR을 TaqMan Universal PCR 마스터믹스 (Thermo Fisher Scientific)를 이용하여 수행하였다. 프라이머 및 프로브 서열에 대해 표 11을 참조한다. 데이터를 엑틴에 대해 표준화하였다.

표 11

RT-qPCR 프라이머 및 프로브

[0351]

표시	서열	SEQ ID NO:
마우스 ALB ex 1 전방향	CAGGGGTGTGTTTCGTGCGAG	38
인간 IDS 역방향	ATGAGCAGGACGTTAAGCGC	39
인간 IDS TaqMan	FAM-AAACCCAGGCCAACTCAACT-BHQ	40
마우스 엑틴	NA (Thermo Fisher, Mm01205647_g1)	NA

[0352] 웨스턴 블롯팅

[0353] 총 단백질 추출물을 앞서 기술된 것과 같이 간 샘플로부터 제조하였다. 단백질 농도를 Pierce 비시초닌산 (BCA) 단백질 검정 키트 (Thermo Fisher Scientific)를 사용하여 결정한 후에 웨스턴 블롯에 의해 IDS 검출을 하였다. 사용한 항체는 IDS (AF2449; R&D Systems) 및 글리세르알데하이드 3-포스페이트 탈수소효소 (GAPDH) (A00191-40, GeneScript)였다.

[0354] IDS 검정

[0355] 1 ug의 총 간 단백질 추출물 또는 1:3 희석의 iPS-유래 간세포 조건부 배지를 이전에 기술된 검정에 사용하였다 (Laoharawee, K. *et al.*, 상기 동일).

[0356] 간 샘플을 주사 후 56일 제에 처리한 동물로부터 수집하고 알부민 절단 활성을 상기에서 기술한 것과 같이 측정하였다. 더불어, 도입유전자 발현을 역전사 간 mRNA에 의해 분석하고 그 생성물에 내인성 마우스 알부민 엑손 1 과 도입유전자 인간 IDS 사이의 접합을 포함한 TaqMan 프라이머-프로브 쌍을 사용하여 qPCR을 수행하였다. 간의

총 단백질 추출물이 인간-특이적 IDS 항체 (로딩 대조군으로서 작용한 GAPDH)와 혼성화되는 IDS의 웨스턴 블롯을 또한 수행하였다.

[0357] 도 10A 및 10B에서 나타난 것과 같이, 본원에 기술된 변형 (5'UTR, FLAG, mut6WPRE)을 포함하는 AAV 벡터는 변형이 없는 AAV 벡터와 비교하여 생체내에서 알부민 유전자를 최대 7배까지 절단하였다. 유사하게, 변형된 ZFN은 F.IX 도너의 통합의 증가를 유발할 수 있었다 (도 10C 참조).

[0358] 나아가, 도 11 내지 13에서 나타난 것과 같이, 도너 IDS 도입유전자는 변형된 ZFN-표적화 AAV 벡터가 표적화된 통합을 위해 사용되었을 때, 도너 도입유전자 (IDS)의 발현은 미변형 ZFN-표적화 AAV 벡터에 의해 매개된 통합 (도 11 및 12)과 비교하여 생체내에서 18배 증가하였고 혈장에서 (IDS의) 효소 활성 또한 증가하였다. 도 11A는 이 연구에 사용한 3개의 상이한 도너: SB-IDS, SB-F9 및 SB-IDUA의 개략도를 보여준다. 변형된 ZFN은 도 11B에서 나타난 것과 같이 일차 간세포에서 알부민 표적에 대해 증가된 활성을 유발하였고, 여기서 개선된 ZFN ("ZFN 2.0")은 원래의 ZFN 쌍 ("현재")과 비교하여 중간 용량에서 indel 퍼센트의 34배 증가 및 고용량에서 활성의 22배 증가를 유발하였다. 이들 ZFN이 IDS 도너와 쌍을 이룰 때, IDS 활성은 표준 ZFN ("현재")과 비교하여 개선된 ("ZFN 2.0") ZFN에 대해 중간 용량에서 5배 및 고용량에서 21배의 증가를 세포 상층액에서 검출하였다.

[0359] IDUA 활성 검정

[0360] IDUA 활성을 기술분야에 공지된 방법에 따라 측정한다. 예를 들어, 한 예시의 검정은 다음과 같다: α-L-이두로니다제의 활성을 확립된 검정 조건 (Whitley 1987 상기 동일, (Whitley 1986, 상기 동일)을 따라 기질로서 4-메틸움벨리페릴 α-L-이두로니다 (Glycosynth)를 사용하여 형광 검정에 의해 결정하였다. 4MU-이두로니다 기질을 좁고, 잘 확립된 최적 범위의 pH로, 및 선택된 기질 농도에서 포름산 나트륨 완충액, 0.4 M, pH 3.5로 희석하였다 (Hopwood *et al* (1979), *Clin Chim Acta*. 92:257-265, Whitley 1986 상기 동일). 그런 후 25 μL의 부분표본을 25 μL의 생물학적 샘플 (예컨대, 혈장, 백혈구, 조직 균등화물)과 혼합하였다. 혼합물을 37°C에서 30분 동안 인큐베이션하고, 200 μL의 글리신 카보네이트 완충액 (pH 10.4)을 첨가하여 반응을 순간 정지시켰다. α-L-이두로니다제는 비형광 기질 (4MU-이두로니다)의 형광 생성물 (4-MU)로의 절단을 측정하였다. 4-메틸움벨리페론 (4-MU, Sigma)을 사용하여 표준 곡선을 작성하였다. 그 결과 생성된 형광을 Bio-Tek 플레이트 판독기를 사용하여 355 nm에서의 여기 및 460 nm에서의 방출로 측정하였다. α-L-이두로니다제 효소 활성을 Pierce 단백질 검정 키트 (Fisher)로 결정된 것과 같이 mg 단백질당 단위 (시간당 생성물로 전환된 nmol)로 표시하였다. 모든 반응을 삼중으로 실시하였다.

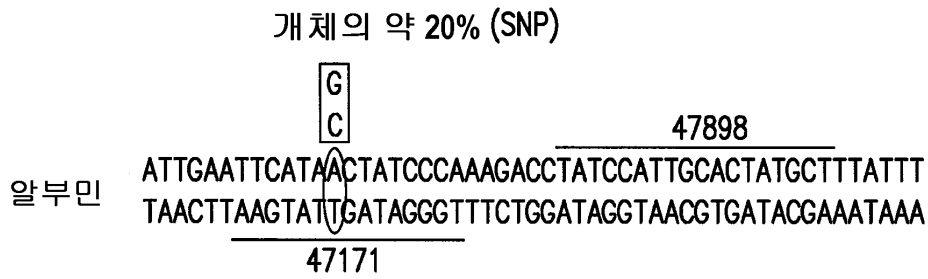
[0361] 데이터는 AAV-ZFN 발현 구성물의 최적화가, 단일유전자 질환의 교정을 위한 시험관내 및 생체내 게놈 편집 구성물에서 전부, 게놈 편집에서 놀랍고 예상하지 못했던 장점 (최대 7배의 절단의 증가 및 최대 18배의 도입유전자 발현을 포함함)을 유발한 것을 입증한다. 그러므로, ZFN 발현 벡터를 구성하는 요소들을 최적화함으로써, 전체 ZFN 활성 및/또는 특이성의 개선이 시험관내 및 생체내에서 현실화된다. 이들 방법은 알부민 유전자좌로의 삽입을 위해 어떠한 도입유전자 도너 (예를 들어, IDS, IDUA 및 F.IX)로도 사용될 수 있다. 도입유전자에 의해 암호화된 단백질의 발현 및 분비는 그것을 필요로 하는 대상체에 대한 생체내 생성을 허용한다.

[0362] 본원에서 언급된 모든 특허, 특허 출원 및 공보는 그 전문이 참조에 의해 본원에 포함된다.

[0363] 비록 개시가 이해의 명료함을 목적으로 예시 및 실시예에 의해 어느 정도 상세하게 제공되었지만, 기술분야의 숙련된 사람들에게는 다양한 수정 및 변형이 개시의 사상 또는 범주로부터 벗어나지 않으면서 실시될 수 있다는 것이 분명한 것이다. 따라서, 전술한 설명 및 실시예는 제한하는 것으로 해석되지 않아야 한다.

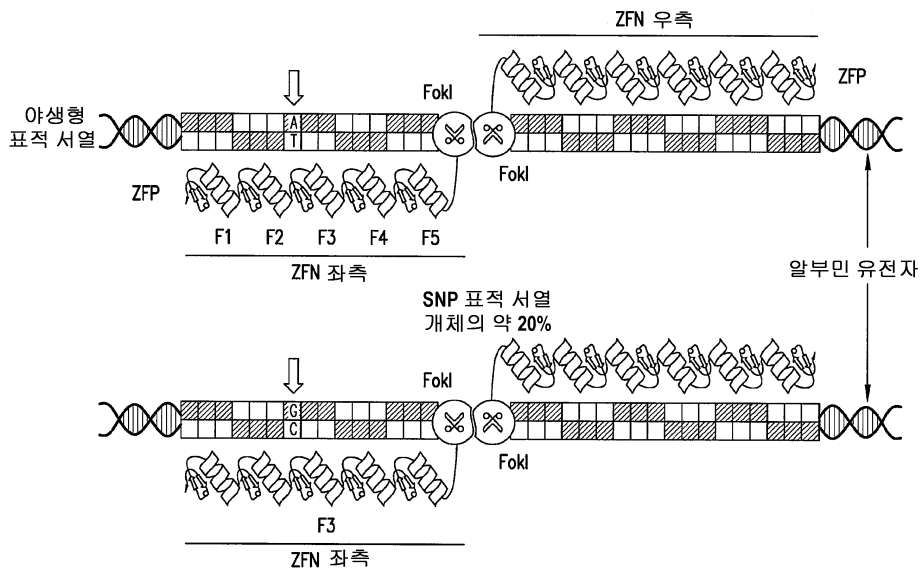
도면

도면1a



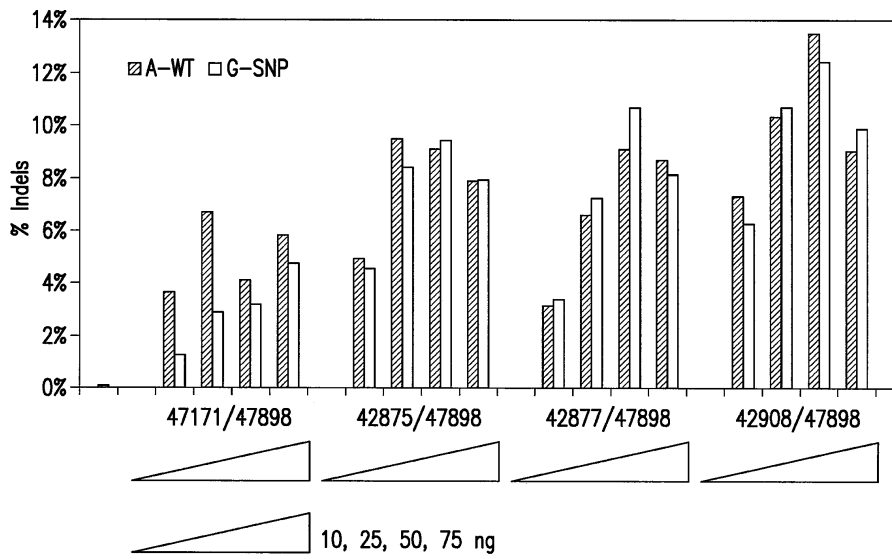
도 1A

도면1b



도 1B

도면2



도 2

도면3a

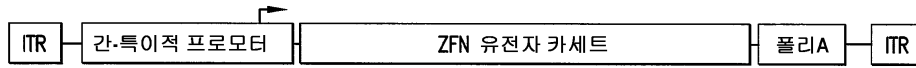
ZFN1	ZFN2	알부민 %indels				SMCHD1		
		2000 ng	62.5 ng	31.25 ng	15.6 ng	2000 ng		
71545	42875-F1F3Q-R416H	71673	47898-F3Q-R416H	82.4	39.8	6.1	0.8	0.017
71545	42875-F1F3Q-R416H	71721	47874-F3Q-R416H	82.7	44.5	9.2	2.0	0.018
71545	42875-F1F3Q-R416H	71739	47874-F1F3Q-S418P	84.2	63.3	19.6	2.8	0.011
71545	42875-F1F3Q-R416H	71741	47874-F1F3Q-L424F	85.1	61.1	23.9	4.5	0.017
71557	42875-F1F3Q-N542D	71721	47874-F3Q-R416H	80.8	46.8	8.9	1.8	0.018
71557	42875-F1F3Q-N542D	71728	47874-F3Q-P478S	80.9	53.9	20.6	4.6	0.012
71639	42908-F1F3Q-WT	71696	47898-F1F3Q-P478S	70.3	16.6	1.0	0.1	0.012
71641	42908-F1F3Q-R416H	71673	47898-F3Q-R416H	72.6	12.9	0.8	0.1	0.008
71641	42908-F1F3Q-R416H	71741	47874-F1F3Q-L424F	69.3	38.3	6.5	0.5	0.009
71653	42908-F1F3Q-N542D	71673	47898-F3Q-R416H	67.4	11.5	0.8	0.0	0.014
71653	42908-F1F3Q-N542D	71721	47874-F3Q-R416H	86.4	19.9	2.5	0.2	0.018
71653	42908-F1F3Q-N542D	71741	47874-F1F3Q-L424F	87.7	37.9	4.8	0.3	0.011
47171	Parent	47898	모	78.9	21.1	1.9	0.2	1.225

GFP	
0.098	알부민 플레이트 1
0.033	알부민 플레이트 2
0.008	SMCHD1

도 3A

도면3b

ZFN 좌측		ZFN 우측		알부민 (표적·상)				표적 외
모	표준	모	표준	2000 ng	62.5 ng	31.3 ng	15.6 ng	2000 ng
최적화된	F1F3Q(ZFP) + N159D(FokI)	최적화된	F3Q(ZFP) + P95S (FokI)	78.9	21.1	1.9	0.2	1.23
				80.9	53.9	20.6	4.6	0.01



↓
도입유전자 변형

- 5'UTR의 첨가는 ZFN 발현을 증가시킨다
- N·말단 펩타이드의 첨가는 절단 효율을 증가시킨다
- WPRE의 첨가는 ZFN mRNA 안정성 및 번역을 증가시킨다.

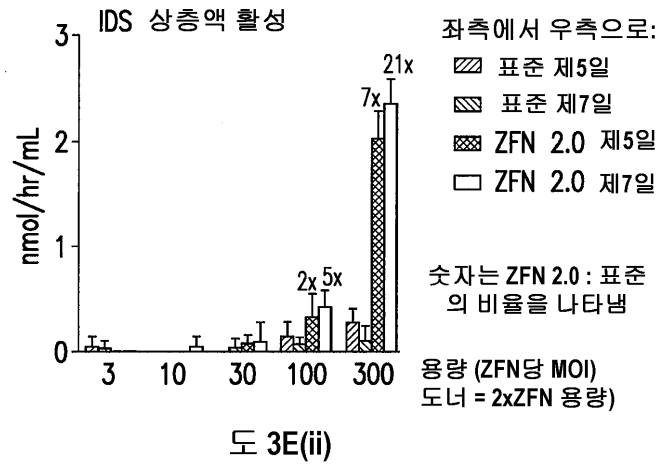
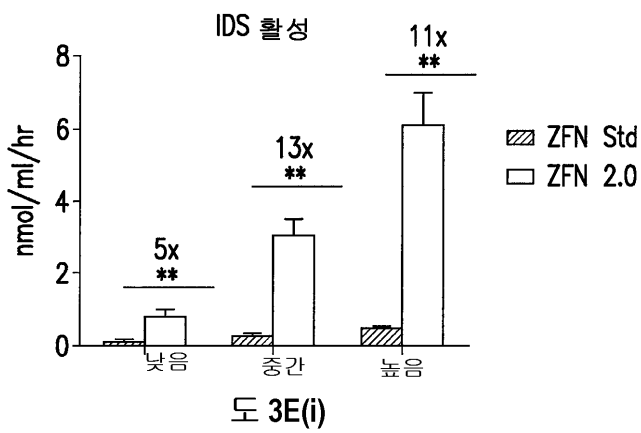
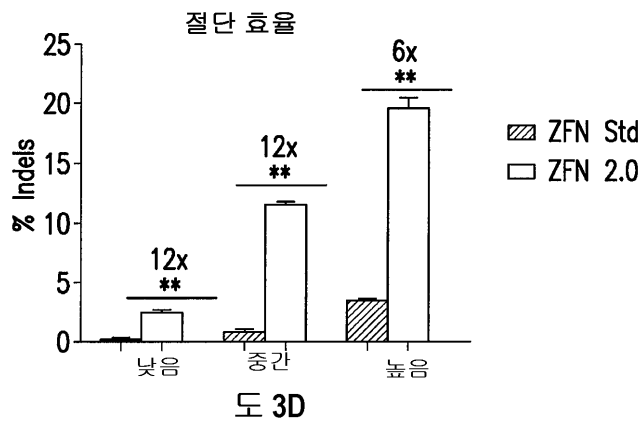
도 3B

도면3c

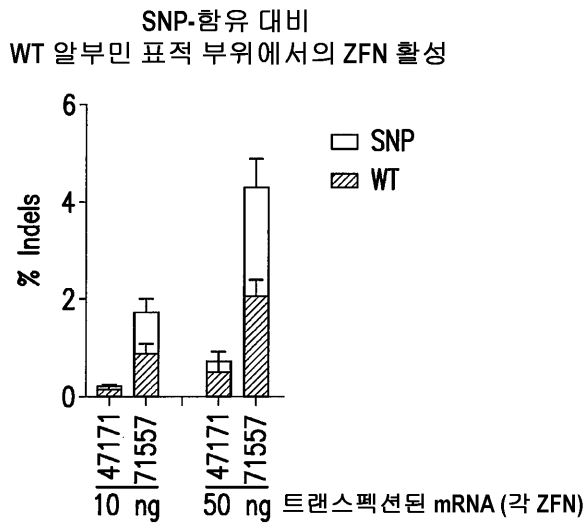
	유전자/위치	포획 사건	K562 % indels	HepG2 % indels
	ALB (표적-상)	3498	87.7%	98.2%
1	SKA2	18	ns	ns
2	Chr5 (유전자간)	12	ns	ns
3	Chr9 (유전자간)	11	ns	ns
4	AC079807.4	10	ns	ns
5	Chr1 (유전자간)	9	ns	ns
6	CSMD3	9	ns	ns
7	Chr18 (유전자간)	9	ns	ns
8	Chr19 (유전자간)	9	ns	ns
9	Chr1 (유전자간)	8	ns	0.08 [^]
10	HECTD4	8	ns*	ns
11	Chr17 (유전자간)	8	ns	ns
12	Chr4 (유전자간)	8	ND	ND
13	CSMD3	8	ns	ns
14	Chr11 (유전자간)	8	ns	ns
15	Chr1 (유전자간)	7	ns	ns
16	AKAP6	7	ND	ND
17	Chr19 (유전자간)	7	ns	ns
18	OSBPL6	7	ns	ns
19	QPCT	7	ns	ns
20	RP11-269F21.3	7	ns	ns
21	LPHN3	7	ns	ns
22	CYFIP2	7	ns	ns
23	AUTS2	7	ns	ns
24	COL4A6	7	ns	ns
25	ChrX (유전자간)	7	ns	ns
26	ACAN	7	ns	ns

도 3C

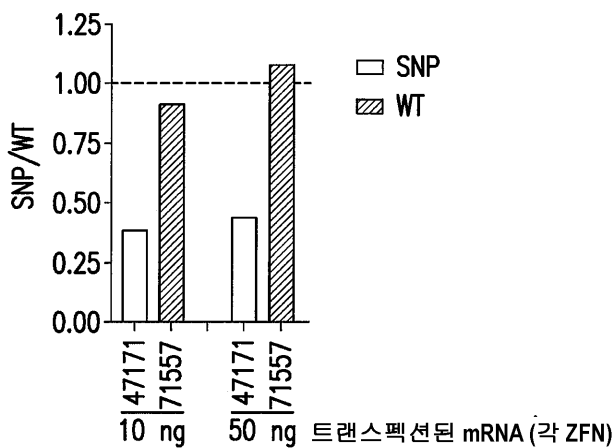
도면3de



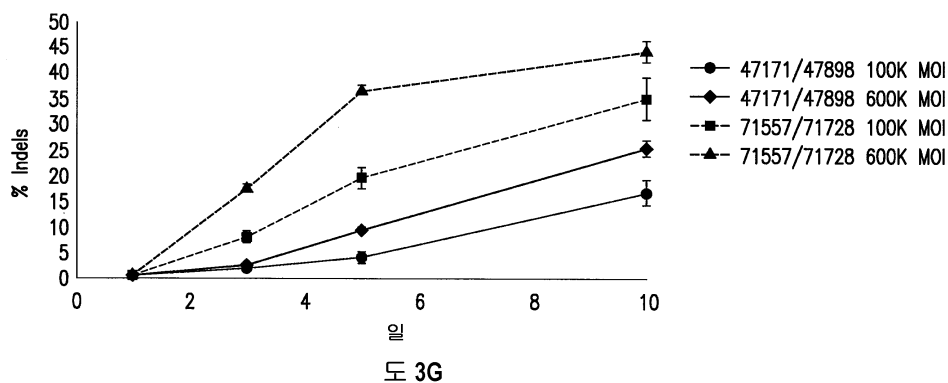
도면3f



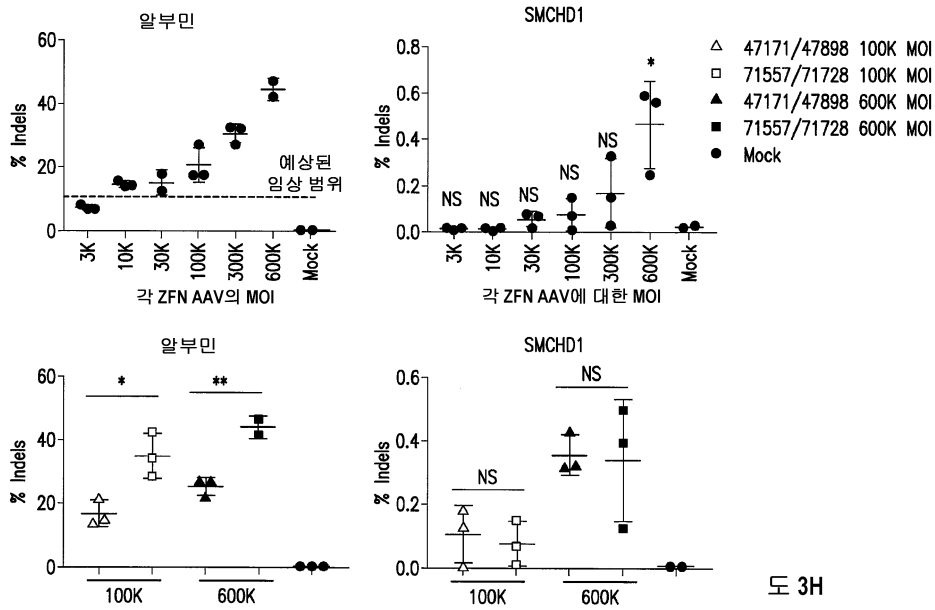
SNP 대비 WT 표적 부위에서의 활성의 비율
1의 값 → SNP & WT 표적 부위에서의 동등한 절단



도면3g

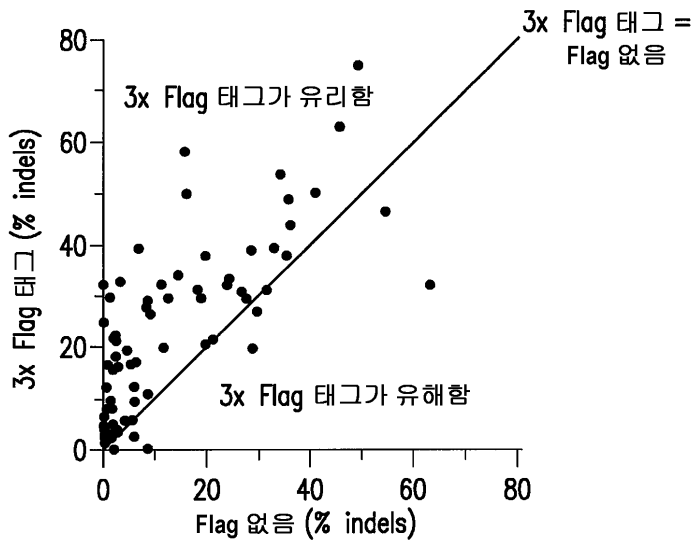


도면3h



도 3H

도면4ab



도 4A

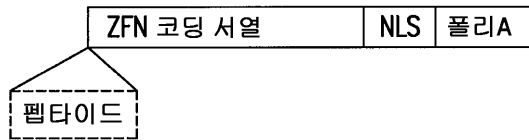
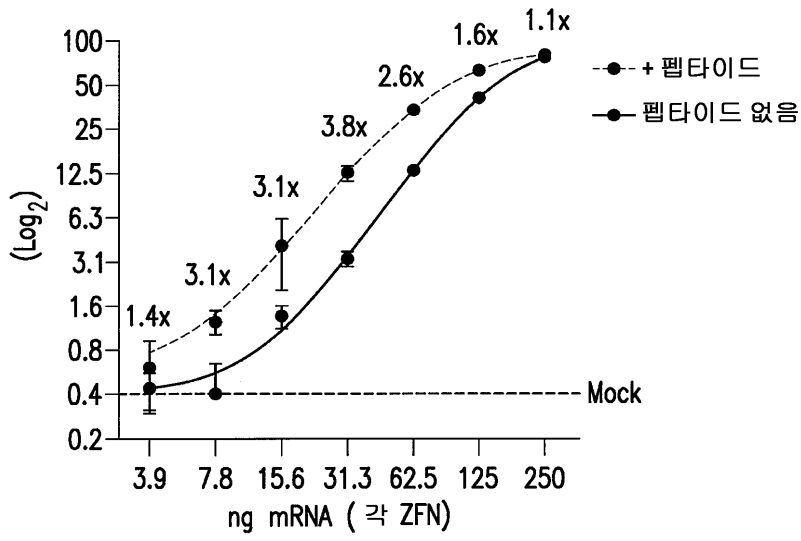
	3xFlag/Flag 없음
긍정적 효과 >10%	61
중립 (10% 이내)	7
부정적 효과 >10%	6

n=74

평균 4.1x

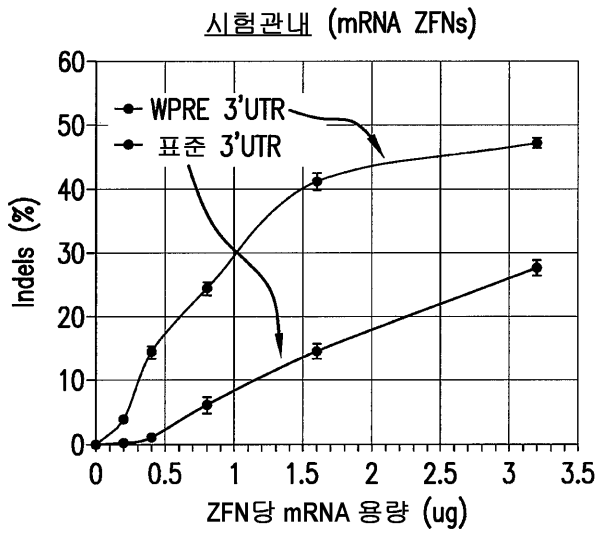
도 4B

도면4c

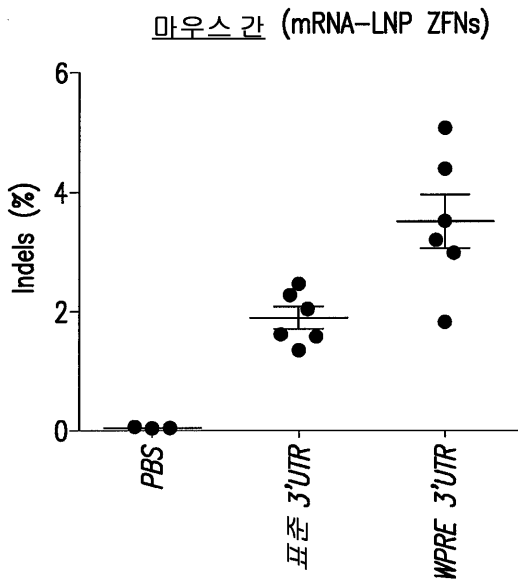


도 4C

도면5ab

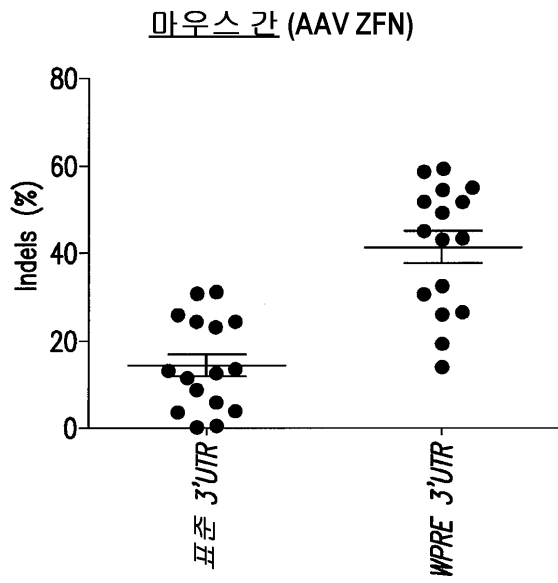


도 5A



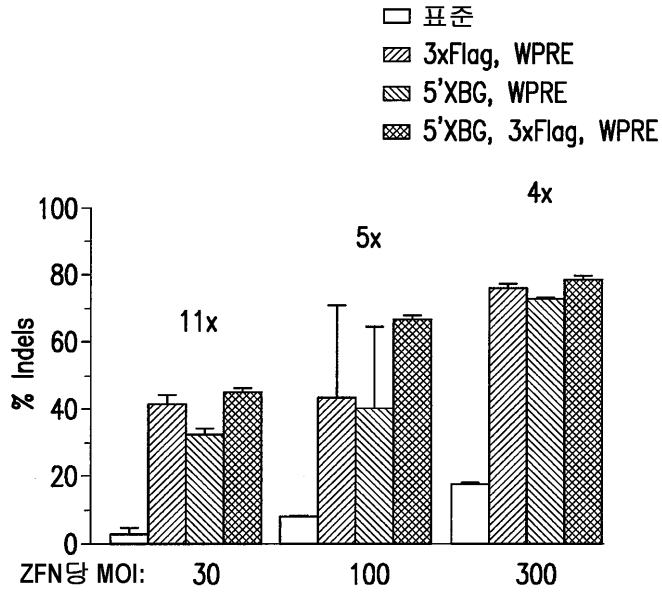
도 5B

도면5c



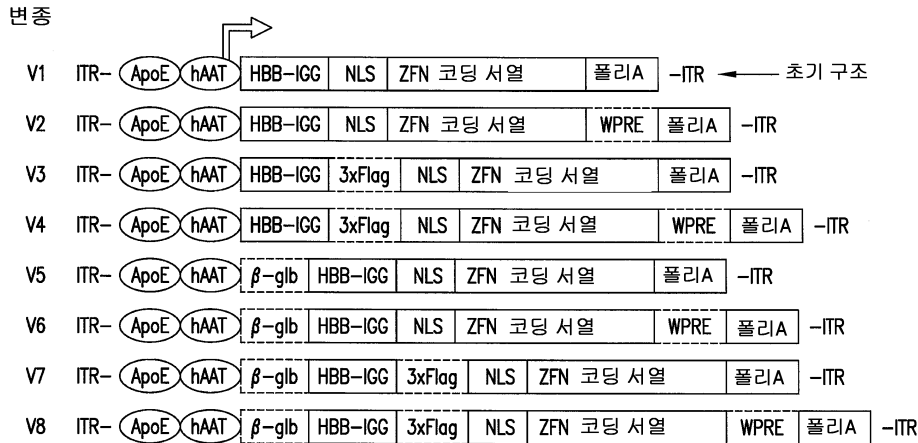
도 5C

도면6



도 6

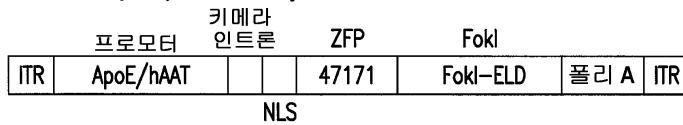
도면7a



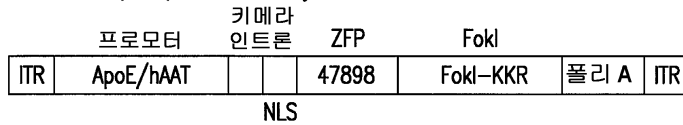
도 7A

도면7b

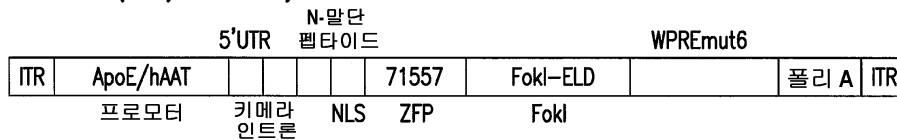
SB-47171 (AAV2/6 좌측 ZFN)



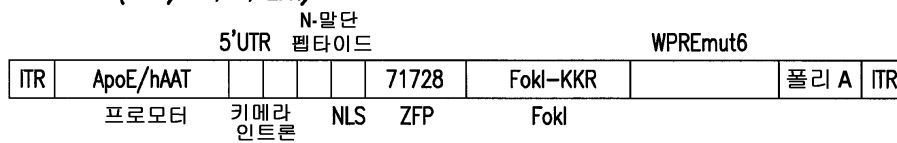
SB-47898 (AAV2/6 우측 ZFN)



SB-71557 (AAV2/6 좌측 ZFN)

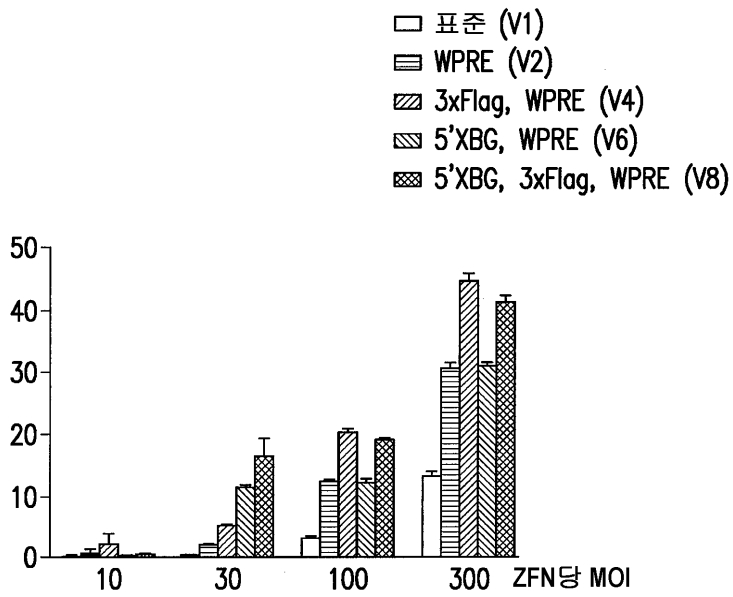


SB-71728 (AAV2/6 우측 ZFN)



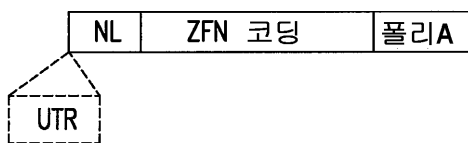
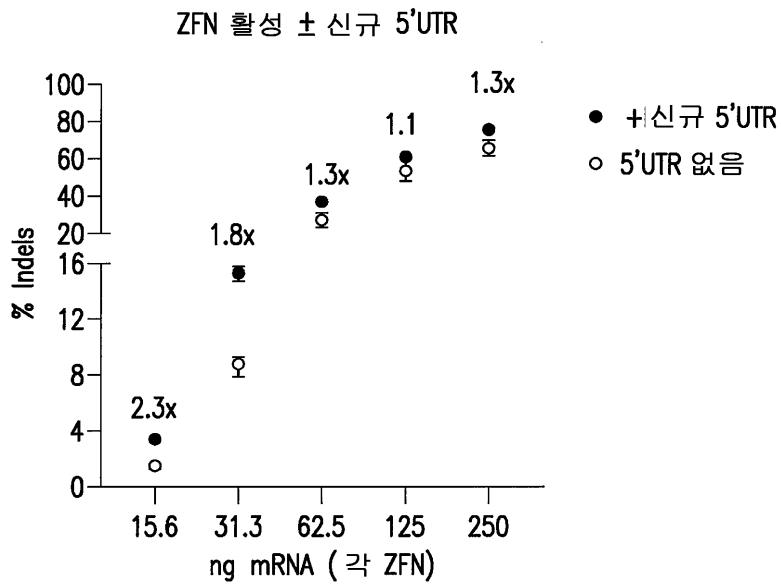
도 7B

도면8



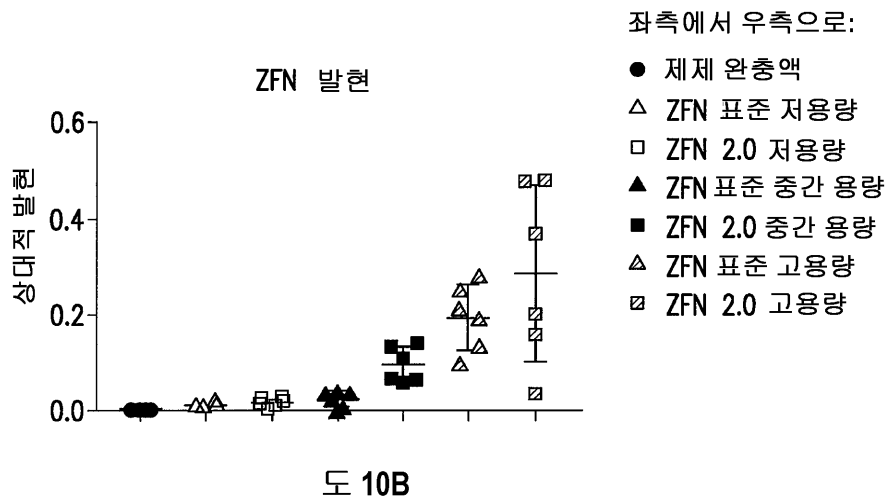
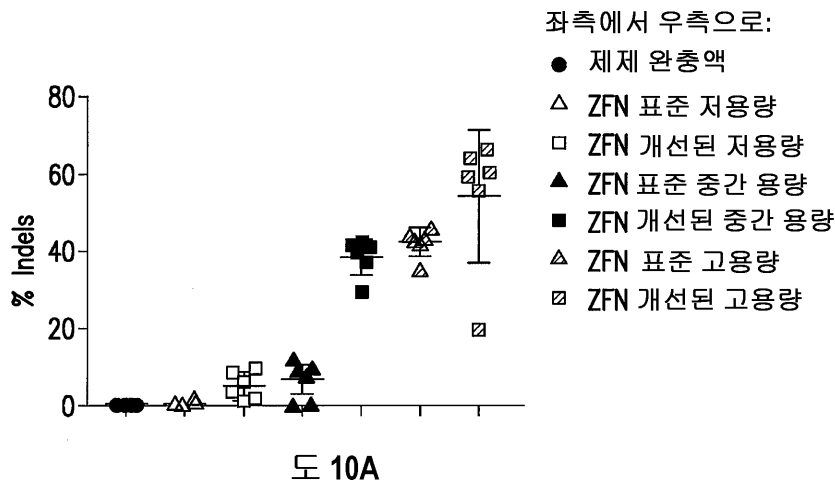
도 8

도면9

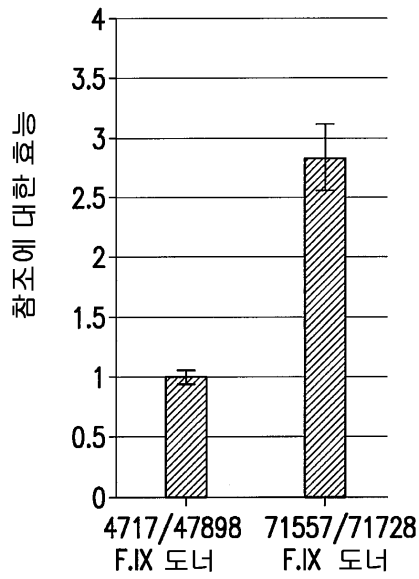


도 9

도면10ab



도면10c

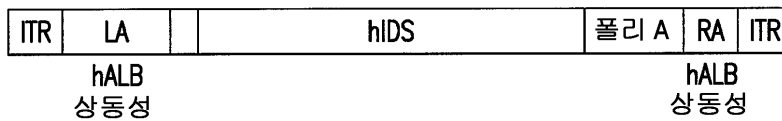


도 10C

도면11a

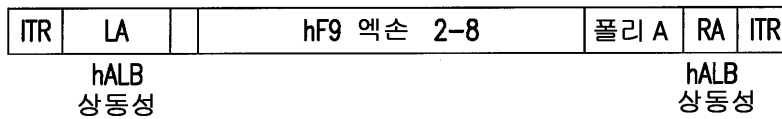
SB-IDS 도입유전자 도너 (AAV2/6 도너)

SA 인간 IDS 도입유전자



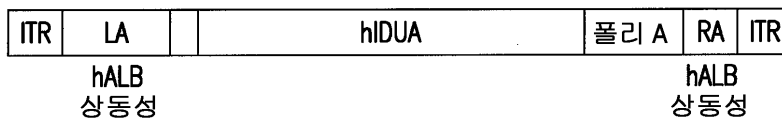
SB-F9 도입유전자 도너 (rAAV2/6 도너)

SA 인간 F9 도입유전자



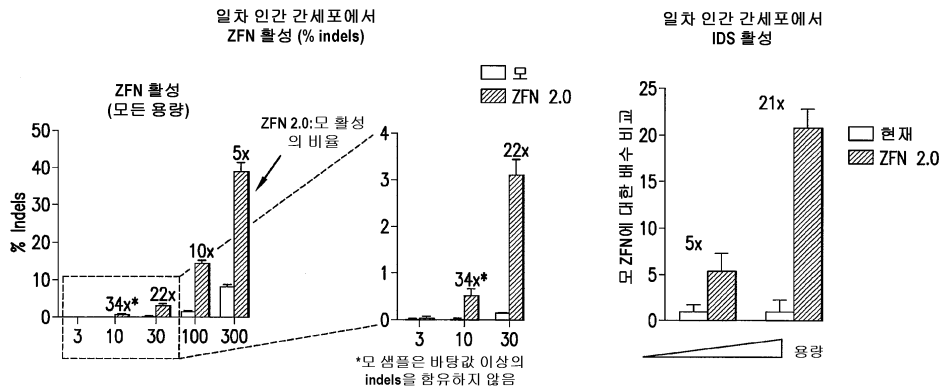
SB-IDUA 도너 (hIDUA 도너 - AAV2/6)

SA 인간 IDUA 도입유전자



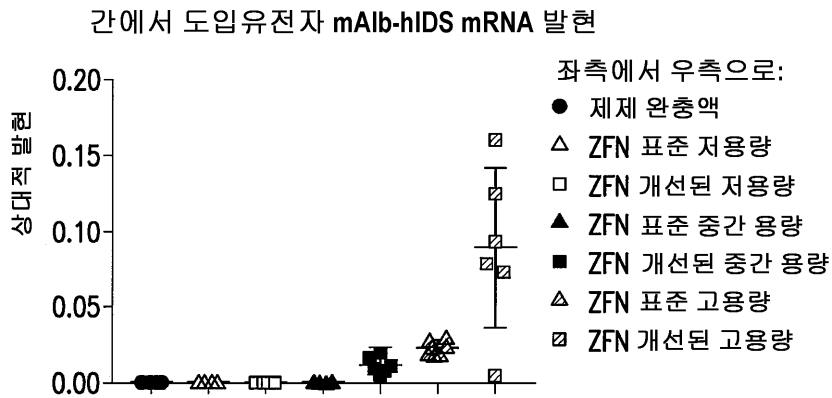
도 11A

도면11b



도 11B

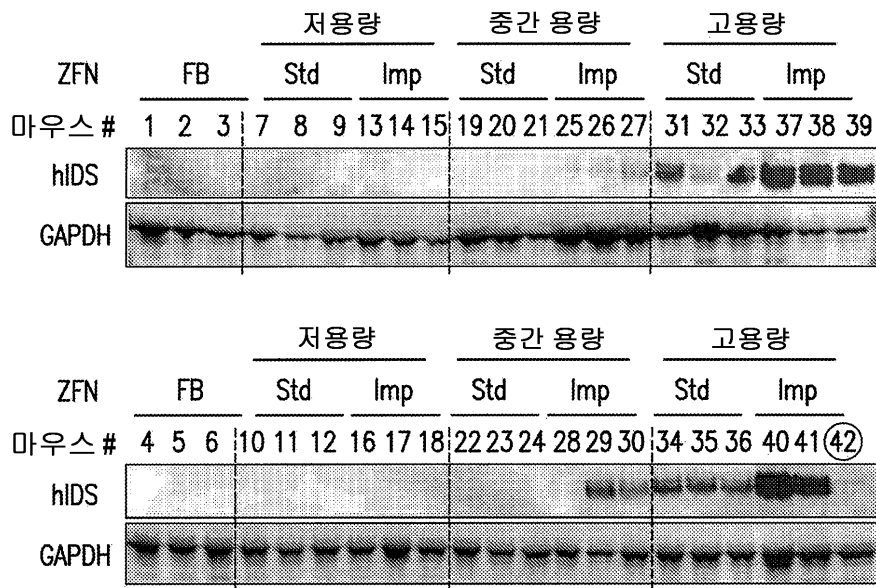
도면11c



도 11C

도면12

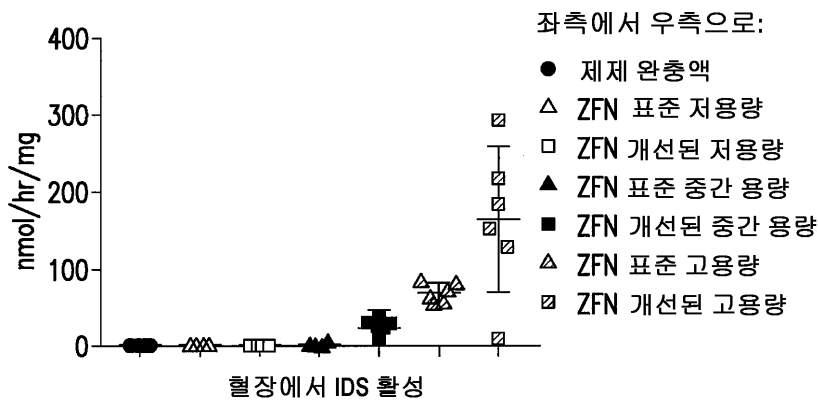
마우스 간 단백질 추출물을 사용하는
인간 IDS 웨스턴 블롯



도 12

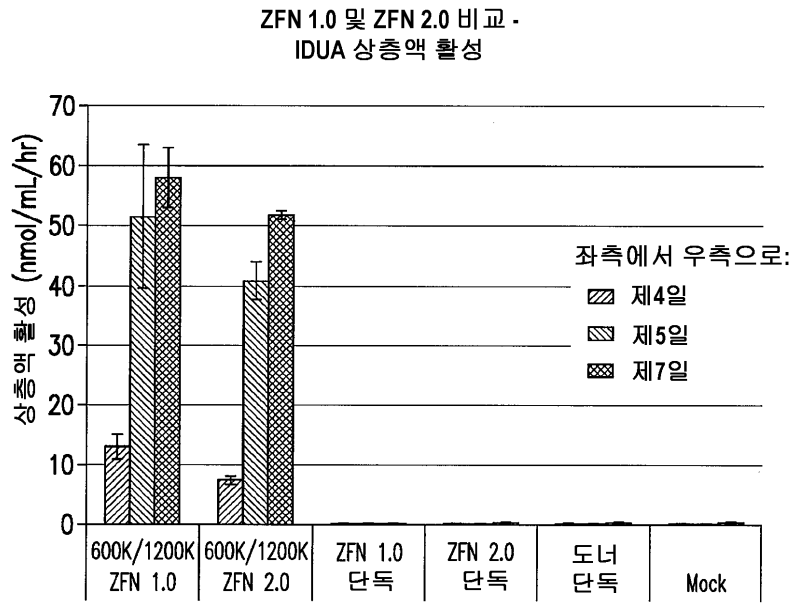
도면13

간에서 IDS 활성

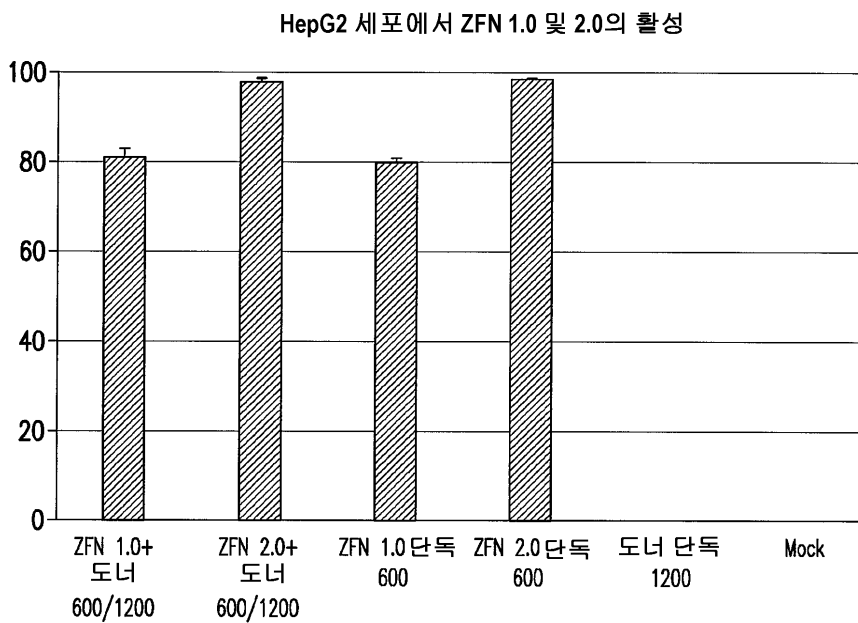


도 13

도면14



도 14A



도 14B

서열목록

SEQUENCE LISTING

<110> SANGAMO THERAPEUTICS, INC.

<120> ENGINEERED TARGET SPECIFIC NUCLEASES

<130> 8325-0169.40

<140><141><150> 62/802,092

<151> 2019-02-06

<150> 62/795,937
 <151> 2019-01-23
 <150> 62/758,786
 <151> 2018-11-12
 <150> 62/728,226
 <151> 2018-09-07
 <150> 62/628,016
 <151> 2018-02-08
 <160> 72
 <170> PatentIn version 3.5
 <210> 1
 <211> 50
 <212> DNA
 <213> Xenopus sp.
 <220><221> misc_feature
 <222> (1)..(2)
 <223> This region may or may not be present in its entirety

<400> 1
 tgcttgttct ttttgcagaa gctcagaata aacgctcaac tttggcagat 50

<210> 2
 <211> 579
 <212> PRT
 <213> Planomicrobium okeanokoites
 <400> 2
 Met Val Ser Lys Ile Arg Thr Phe Gly Trp Val Gln Asn Pro Gly Lys
 1 5 10 15
 Phe Glu Asn Leu Lys Arg Val Val Gln Val Phe Asp Arg Asn Ser Lys
 20 25 30
 Val His Asn Glu Val Lys Asn Ile Lys Ile Pro Thr Leu Val Lys Glu
 35 40 45

 Ser Lys Ile Gln Lys Glu Leu Val Ala Ile Met Asn Gln His Asp Leu
 50 55 60
 Ile Tyr Thr Tyr Lys Glu Leu Val Gly Thr Gly Thr Ser Ile Arg Ser

Gly Ser Leu Lys Ile Glu Gln Ile Gln Asp Asn Leu Lys Lys Leu Gly
 325 330 335

Phe Asp Glu Val Ile Glu Thr Ile Glu Asn Asp Ile Lys Gly Leu Ile
 340 345 350

Asn Thr Gly Ile Phe Ile Glu Ile Lys Gly Arg Phe Tyr Gln Leu Lys
 355 360 365

Asp His Ile Leu Gln Phe Val Ile Pro Asn Arg Gly Val Thr Lys Gln
 370 375 380

Leu Val Lys Ser Glu Leu Glu Glu Lys Lys Ser Glu Leu Arg His Lys
 385 390 395 400

Leu Lys Tyr Val Pro His Glu Tyr Ile Glu Leu Ile Glu Ile Ala Arg
 405 410 415

Asn Ser Thr Gln Asp Arg Ile Leu Glu Met Lys Val Met Glu Phe Phe
 420 425 430

Met Lys Val Tyr Gly Tyr Arg Gly Lys His Leu Gly Gly Ser Arg Lys
 435 440 445

Pro Asp Gly Ala Ile Tyr Thr Val Gly Ser Pro Ile Asp Tyr Gly Val
 450 455 460

Ile Val Asp Thr Lys Ala Tyr Ser Gly Gly Tyr Asn Leu Pro Ile Gly
 465 470 475 480

Gln Ala Asp Glu Met Gln Arg Tyr Val Glu Glu Asn Gln Thr Arg Asn
 485 490 495

Lys His Ile Asn Pro Asn Glu Trp Trp Lys Val Tyr Pro Ser Ser Val
 500 505 510

Thr Glu Phe Lys Phe Leu Phe Val Ser Gly His Phe Lys Gly Asn Tyr
 515 520 525

Lys Ala Gln Leu Thr Arg Leu Asn His Ile Thr Asn Cys Asn Gly Ala
 530 535 540

Val Leu Ser Val Glu Glu Leu Leu Ile Gly Gly Glu Met Ile Lys Ala
 545 550 555 560

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

565

570

575

Ile Asn Phe

<210> 3

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 3

Asp Tyr Lys Asp Asp Asp Lys

1 5

<210> 4

<211> 23

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

Peptide

<220><221

> MOD_RES

<222> (1)..(1)

<223> May or may not be present

<400> 4

Met Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp

1 5 10 15

Tyr Lys Asp Asp Asp Asp Lys

20

<210> 5

<211> 7

<212> PRT

<213> Simian virus 40

<400> 5

Pro Lys Lys Lys Arg Lys Val

1 5

<210> 6

<211> 9

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

C-myc NLS sequence

<400> 6

Pro Ala Ala Lys Arg Val Lys Leu Asp

1 5

<210> 7

<211> 10

<212> PRT

<213> Hepatitis delta virus

<400> 7

Glu Gly Ala Pro Pro Ala Lys Arg Ala Arg

1 5 10

<210> 8

<211> 8

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

polyoma T protein NLS sequence

<400> 8

Val Ser Arg Lys Arg Pro Arg Pro

1 5

<210> 9

<211> 18

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

nucleoplasmin carboxy tail NLS sequence

<400> 9

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

1 5 10 15

Leu Asp

<210> 10

<211> 38

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

NLS sequence

<400> 10

Asn Gln Ser Ser Asn Phe Gly Pro Met Lys Gly Gly Asn Phe Gly Gly

1 5 10 15

Arg Ser Ser Gly Pro Tyr Gly Gly Gly Gly Gln Tyr Phe Ala Lys Pro

20 25 30

Arg Asn Gln Gly Gly Tyr

35

<210> 11

<211> 17

<212> PRT

<213> Human T-cell leukemia virus type 1

<400> 11

Pro Lys Thr Arg Arg Arg Pro Arg Arg Ser Gln Arg Lys Arg Pro Pro

1 5 10 15

Thr

<210> 12

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 12

tttgggatag ttatgaattc aatcttca

28

<210> 13

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
oligonucleotide

<400> 13

cctatccatt gcactatgct ttatttaa

28

<210> 14

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 14

Gln Ser Gly Asn Leu Ser Arg

1 5

<210> 15

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 15

Leu Lys Gln Asn Leu Cys Met

1 5

<210> 16

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 16

Trp Ala Asp Asn Leu Gln Asn

1 5

<210> 17

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 17

Thr Ser Gly Asn Leu Thr Arg

1 5

<210> 18

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 18

Arg Gln Ser His Leu Cys Leu

1 5

<210> 19

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 19

Gln Ser Gly Asn Leu Ala Arg

1 5

<210> 20

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 20

Trp Gln Ser Asn Leu Gln Asn

1 5

<210> 21

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 21

Arg Arg Ser His Leu Thr Ser

1 5

<210> 22

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 22

Leu Ile Thr Thr Leu Arg Asn

1 5

<210> 23

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 23

Trp Ala Ser Asn Leu Gln Asn

1 5

<210> 24

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 24

Thr Pro Gln Leu Leu Asp Arg

1 5

<210> 25

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 25

Leu Lys His Asn Leu Leu Thr

1 5

<210> 26

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 26

Asp Gln Ser Asn Leu Asn Ala

1 5

<210> 27

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 27

Arg Asn Phe Ser Leu Thr Met

1 5

<210> 28

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 28

Leu Arg His Asp Leu Asp Arg

1 5

<210> 29

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 29

His Arg Ser Asn Leu Asn Lys

1 5

<210> 30

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 30

Gln Ser Ser Asp Leu Ser Arg

1 5

<210> 31

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
peptide

<400> 31

Asp Gln Ser Asn Leu Arg Ala

1 5
 <210> 32
 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 peptide

<400> 32
 Leu Arg His Asp Leu Glu Arg

1 5
 <210> 33

 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 peptide

<400> 33
 Leu Lys Trp Asn Leu Arg Thr

1 5
 <210> 34
 <211> 44
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer

<220><221> modified_base
 <222> (21)..(24)
 <223> a, c, t, g, unknown or other

<400> 34
 acacgacgct cttccgatct nnnnttgagt ttgaatgcac agat

44

<210> 35

 <211> 44
 <212> DNA

<213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <220><221> modified_base
 <222> (21)..(24)
 <223> a, c, t, g, unknown or other
 <400> 35
 gacgtgtgct cttccgatct nnnngaaaca gggagagaaa aacc 44
 <210> 36
 <211> 44
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <220><221> modified_base
 <222> (21)..(24)
 <223> a, c, t, g, unknown or other

 <400> 36
 acacgacgct cttccgatct nnnngcacta aggaaagtgc aaag 44
 <210> 37
 <211> 47
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 primer
 <220><221> modified_base
 <222> (21)..(24)
 <223> a, c, t, g, unknown or other
 <400> 37
 gacgtgtgct cttccgatct nnnaaccaa gaagacagac taaaatg 47
 <210> 38
 <211> 20
 <212> DNA
 <213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 38

caggggtgtg ttcgtcgag

20

<210> 39

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

primer

<400> 39

atgagcagga cgtaagcgc

20

<210> 40

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

probe

<400> 40

aaaccaggc caactcaact

20

<210> 41

<211> 51

<212> DNA

<213> Homo sapiens

<400> 41

attgaattca tarctatccc aaagacctat ccattgcact atgctttatt t

51

<210> 42

<211> 592

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 42

aatcaacctc tggattacaa aatttgtgaa agattgactg atattcttaa ctatgttgct 60
 ccttttacgc tgtgtggata tgctgcttta atgcctctgt atcatgctat tgcttcccgt 120
 acggctttcg ttttctcctc cttgtataaa tcctgggtgc tgtctcttta tgaggagtgg 180

tggcccgttg tccgtcaacg tggcgtggig tgctctgtgt ttgctgacgc aacccccact 240
 ggctggggca ttgccaccac ctgtcaactc ctttctggga ctttcgcttt cccctcccg 300
 atgccacagg cagaactcat cgcccgctgc cttgcccgct gctggacagg ggctaggttg 360
 ctgggcactg ataattccgt ggtgtgtgctg gggaaatcat cgtcctttcc ttggctgctc 420
 gcctgtgttg ccaactggat cctgcgcggg acgtccttct gctacgtccc ttcggtctc 480
 aatccagcgg acctccctc ccgaggcctt ctgccggttc tgcggcctct cccgctctt 540
 cgctttcggc ctccgacgag tcggatctcc ctttgggccg cctccccgcc tg 592

<210> 43

<211> 3195

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 43

ctgcgcctc gctcgtcac tgaggccgc cgggcaaagc ccgggcgtcg ggcgacctt 60
 ggtcgcggc cctcagtgag cgagcgagcg cgcagagagg gaggggcaa ctccatcact 120
 aggggttctt gcggcctaag cttgagctct tcgaaaggct cagaggcaca caggagtttc 180
 tgggctcacc ctgccccctt ccaaccctc agttccatc ctccagcagc tgtttgtgtg 240
 ctgcctctga agtccacact gaacaaactt cagcctactc atgtccctaa aatgggcaaa 300

cattgcaagc agcaaacagc aaacacacag cctccctgc ctgctgacct tggagctggg 360
 gcagaggtea gagacctctc tgggcccattg ccacctcaa catccactcg accccttggg 420
 atttcgggtg agaggagcag aggttgtcct ggcgtggttt aggtagtgtg agaggggtcc 480
 cggggatctt gctaccagtg gaacagccac taaggattct gcagtgagag cagagggcca 540
 gctaagtggg actctcccag agactgtctg actcacgcca cccctccac cttggacaca 600
 ggacgtgtg gtttctgagc caggtacaat gactccttc ggtaagtgca gtggaagctg 660
 tacactgcc aggcaaagcg tccgggcagc gtaggcgggc gactcagatc ccagccagtg 720

gacttagccc ctgtttgctc ctccgataac tggggtgacc ttggttaata ttcaccagea 780
 gcctccccg ttgcccctct ggatccactg cttaaatag gacaggaca gggccctgctc 840

tcctcagctt caggcaccac cactgacctg ggacagtcct aggtgcttgt tctttttgca 900
 gaagctcaga ataacgctc aactttggca gatactagtc aggtaagtat caaggttaca 960
 agacaggttt aaggagacca atagaaactg ggcttgtcga gacagagaag actcttgcgt 1020
 ttctgatagg cacctattgg tcttactgac atccactttg cttttctctc cacaggaccg 1080
 gtgccatgga ctacaaagac catgacggtg attataaaga tcatgacatc gattacaagg 1140

 atgacgatga caagatggcc cccaagaaga agaggaaggt cggcattcat ggggtacccg 1200
 ccgctatggc tgagaggccc ttccagtgtc gaatctgcat gcagaacttc agtcagtccg 1260
 gcaacctggc ccgccacatc cgcaccaca ccggcgagaa gccttttgcc tgtgacattt 1320
 gtgggaggaa atttgccctg aagcagaacc tgtgtatgca taccaagata cacacgggcg 1380
 agaagccctt ccagtgtcga atctgcatgc agaagtttgc ctggcagtcc aacctgcaga 1440
 accataccaa gatacacagc ggcgagaagc cttccagtg tcgaatctgc atgcgtaact 1500
 tcagtacctc cggcaacctg acccggcaca tccgcacca caccggcgag aagccttttg 1560

 cctgtgacat ttgtgggagg aaatttgccc gccgctccca cctgacctcc cataccaaga 1620
 tacacctgcg gggateccag ctggngaaga gcgagctgga ggagaagaag tccgagctgc 1680
 ggcacaagct gaagtactg ccccacgagt acatcgagct gatcgagatc gccaggaaca 1740
 gcaccaggga ccgcatcctg gagatgaagg tgatggagt tttcatgaag gtgtacggct 1800
 acaggggaaa gcacctgggc ggaagcagaa agcctgacgg cgccatctat acagtgggca 1860
 gccccatcga ttacggcgtg atcgtggaca caaaggccta cagcggcggc tacaatctgc 1920
 ctatcggcca ggccgacgag atggagagat acgtggagga gaaccagacc cgggataagc 1980

 acctcaacc caacgagtgg tggaagggtg acctagcag cgtgaccgag ttcaagtcc 2040
 tgttcgtgag cggccacttc aaggcaact acaaggcca gctgaccagg ctgaaccaca 2100
 tcaccaactg cgacggcgcc gtgctgagcg tggaggagct gctgatcggc ggcgagatga 2160
 tcaaagccgg cacctgaca ctggaggagg tgcggcgcaa gttcaacaac ggcgagatca 2220
 acttcagatc ttgataactc gagtctagaa atcaacctct ggattacaaa atttgtgaaa 2280
 gattgactga tattctaac tatgttctc cttttacgct gtgtggatat getgctttaa 2340
 tgccctgtga tcatgctatt gcttccgta cggctttcgt tttctctcc ttgtataaat 2400

 cctggttget gtctctttat gaggagtgt ggeccgttgt ccgtcaactg ggcgtggtgt 2460
 gctctgtgtt tgctgacgca acccccactg gctggggcat tgccaccacc tgtcaactcc 2520
 tttctgggac tttcgtttc cccctcccga tcgccacggc agaactcatc gccgcctgcc 2580
 ttgcccgctg ctggacaggg gctaggttgc tgggcaactga taattccgtg gtgttgcgg 2640
 ggaaatcatc gtcttttct tggctgctcg cctgtgttgc caactggatc ctgcgcggga 2700

cgctcttctg ctacgtcctt tcggtcttca atccagcggg cctcccttcc cgaggccttc 2760
 tgccggttct gggcctctc ccgctcttc gctttcggcc tccgacgagt cggatctccc 2820

tttgggccgc ctccccgctt ggctagcctg tgccttctag ttgccagcca tetgttgttt 2880
 gccctcccc cgtgccttcc ttgaccctgg aaggtgccac tccactgtc ctttcctaat 2940
 aaaatgagga aattgcatcg cattgtctga gtaggtgtca ttctattctg gggggtgggg 3000
 tggggcagga cagcaagggg gaggattggg aagacaatag caggcatgct ggggatgcgg 3060
 tgggctctat gcggcccgct cgagcgcagg aaccctagt gatggagtg gccactcct 3120
 ctctgcgcgc tcgctcgtc actgaggccg cccgggcttt gcccgggcgg cctcagtgag 3180
 cgagcgagcg cgcag 3195

<210> 44

<211> 130

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 44

ctgcgcgctc gctcgtcac tgagccgcc cgggcaaage ccggcgtcg ggcgacctt 60
 ggtcggccgg cctcagtgag cgagcgagcg cgcagagagg gaggggcaa ctccatcact 120
 aggggttctt 130

<210> 45

<211> 321

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 45

aggctcagag gcacacagga gtttctgggc tcaccctgcc ccttccaac cctcagttc 60
 ccatctcca gcagctgttt gtgtgctgcc tctgaagtcc aactgaaca aacttcagcc 120
 tactcatgtc ctaaaatgg gcaaacattg caagcagcaa acagcaaca cacagcctc 180
 cctgcctgct gaccttggag ctggggcaga ggtcagagac ctctctgggc ccatgccacc 240
 tccaacatcc actcgacccc ttggaatttc ggtggagagg agcagaggtt gtcctggcgt 300
 ggtttaggta gtgtgagagg g 321

<210> 46

<211> 393

<212> DNA

<213>

Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 46

gatcttgcta ccagtggaac agccaactaag gattctgcag tgagagcaga gggccagcta 60
 agtggctactc tcccagagac tgtctgactc acgccacccc ctccacctg gacacaggac 120
 gctgtggttt ctgagccagg tacaatgact ctttctggta agtgcagtgg aagctgtaca 180
 ctgccaggc aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact 240
 tagcccctgt ttgctcctcc gataactggg gtgacctgg ttaatattca ccagcagcct 300
 cccccgttc ccctctggat ccaactgctta aatacggacg aggacagggc cctgtctcct 360

cagcttcagg caccaccact gacctgggac agt 393

<210> 47

<211> 48

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 47

cttgttcttt ttgcagaagc tcagaataaa cgtcaactt tggcagat 48

<210> 48

<211> 133

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 48

gtaagtatca aggttacaag acaggtttaa ggagaccaat agaaactggg cttgtcgaga 60
 cagagaagac tcttgcgttt ctgataggca cctattggtc ttactgacat ccactttgcc 120
 tttctctcca cag 133

<210> 49

<211> 66

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 49

gactacaaag accatgacgg tgattataaa gatcatgaca tcgattacaa ggatgacgat 60

gacaag 66

<210> 50

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 oligonucleotide

<400> 50

cccaagaaga agaggaaggt c 21

<210> 51

<211> 432

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 polynucleotide

<400> 51

gccgctatgg ctgagaggcc cttccagtgt cgaatctgca tgcagaactt cagtcagtcc 60

ggcaacctgg cccgccacat ccgcacccac accggcgaga agccttttgc ctgtgacatt 120

tgtgggagga aatttgcct gaagcagaac ctgtgtatgc ataccaagat acacacgggc 180

gagaagccct tccagtgtcg aatctgcatg cagaagtttg cctggcagtc caacctgcag 240

aaccatacca agatacacac gggcgagaag cccttccagt gtcgaatctg catgcgtaac 300

ttcagtacct cggcaacct gaccgcccac atccgcaccc acaccggcga gaagcctttt 360

gcctgtgaca tttgtgggag gaaatttgc cgcgctccc acctgacctc ccataccaag 420

atacacctgc gg 432

<210> 52

<211> 600

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 52

cagctggtga agagcgagct ggaggagaag aagtccgagc tgcggcacia gctgaagtac 60
gtgccccacg agtacatcga gctgatcgag atcgccagga acagcaccca ggaccgcac 120
ctggagatga aggtgatgga gttcttcatg aaggtgtacg gctacagggg aaagcacctg 180
ggcggaaagca gaaagcctga cggcgccatc tatacagtgg gcagcccat cgattacggc 240
gtgatcgtgg acacaaaggc ctacagcggc ggctacaatc tgcctatcgg ccaggccgac 300
gagatggaga gatcgtgga ggagaaccag acccgggata agcacctcaa ccccaacgag 360
tggtggaagg tgtaccctag cagcgtgacc gagttcaagt tcctgttcgt gagcggccac 420

ttcaaggga actacaaggc ccagctgacc aggctgaacc acatcaccaa ctgacgagc 480
gccgtgctga gcgtggagga gctgtgatc ggcgccgaga tgatcaaagc cggcacctg 540
acactggagg aggtgcccgc caagttcaac aacggcgaga tcaacttcag atcttgataa 600

<210> 53

<211> 592

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 53

aatcaacctc tggattacaa aatttggaa agattgactg atattcttaa ctatgttgc 60
ccttttacgc tgtgtggata tgctgcttta atgcctctgt atcatgctat tgcttcccgt 120
acggctttcg tttctcctc cttgtataaa tcttgggtgc tgtctctta tgaggagttg 180
tggcccgttg tccgtaacg tggcgtggtg tgctctgtgt ttgctgacgc aacccccact 240
ggctggggca ttgccaccac ctgtcaactc ctttctggga ctttcgcttt cccctcccg 300
atgccacgg cagaactcat cgccgcctgc cttgcccgct gctggacagg ggctaggttg 360
ctgggactg ataattccgt ggtgtgtcgg gggaaatcat cgtcctttcc ttggctgctc 420
gcctgtgttg ccaactggat cctgcgcggg acgtccttct gctacgtccc ttcggtctc 480

aatccagcgg acctcccttc ccgaggcctt ctgccggttc tgcggcctct cccgcgtctt 540

cgctttcggc ctccgacgag tcggatctcc ctttgggccg cctccccgcc tg 592

<210> 54

<211> 223

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 polynucleotide

<400> 54

ctgtgccttc tagttgccag ccatctgttg ttgcccctc ccccgctcct tccttgacct 60

tggaaggtgc cactcccact gtcctttcct aataaaatga ggaaattgca tcgattgtc 120

tgagtaggtg tcattctatt ctggggggtg ggggtgggca ggacagcaag ggggaggatt 180

gggaagacaa tagcaggcat gctggggatg cgtgggctc tat 223

<210> 55

<211> 108

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 polynucleotide

<400> 55

aggaaccct agtgatggag ttggccactc cctctctcgc cgctcgtcg ctcaactgagg 60

ccgcccgggc tttgcccggg cggcctcagt gagcgagcga gcgcgcag 108

<210> 56

<211> 3273

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic
 polynucleotide

<400> 56

ctgcgcctc gctcgtcac tgaggccgcc cgggcaaagc cggggcgtcg ggcgacctt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgagagagg gaggggcaa ctccatcact 120

aggggttctt gcggcctaag cttgagctct tcgaaaggct cagaggcaca caggagtttc 180

tgggctcacc ctgccccctt ccaacccctc agttcccatc ctccagcagc tgtttgtgtg 240
 ctgcctctga agtccacact gaacaaactt cagcctactc atgtccctaa aatgggcaaa 300
 cattgcaagc agcaaacagc aaacacacag ccctccctgc ctgctgacct tggagctggg 360
 gcagaggta gagacctctc tgggcccattg ccacctccaa catccactcg accccttggg 420
 atttcgggtg agaggagcag aggttgcctt ggcgtggttt aggtagtgtg agaggggtcc 480

 cggggatctt gctaccagtg gaacagccac taaggattct gcagtgagag cagagggcca 540
 gctaagtggg actctcccag agactgtctg actcacgcca cccctccac ctggacaca 600
 ggacgctgtg gtttctgagc caggtacaat gactcctttc ggtaagtgca gtggaagctg 660
 tacactgccc aggcaaagcg tccgggcagc gtaggggggc gactcagatc ccagccagtg 720
 gacttagccc ctgtttgtct ctccgataac tggggtgacc ttggttaata ttcaccagca 780
 gcctccccg ttgccctctt ggatccactg cttaaatacg gacgaggaca gggccctgtc 840
 tcctcagctt caggcaccac cactgacctg ggacagtcct aggtgcttgt tctttttgca 900

 gaagctcaga ataaacgctc aactttggca gatactagtc aggttaagtat caaggttaca 960
 agacaggttt aaggagacca atagaaactg ggcttgcga gacagagaag actcttgcgt 1020
 ttctgatagg cacctattgg tcttactgac atccactttg cttttctctc cacaggaccg 1080
 gtgcatgga ctacaaagac catgacgggtg attataaaga tcatgacatc gattacaagg 1140
 atgacgatga caagatggcc cccaagaaga agaggaaggt cggcattcat ggggtaccgc 1200
 ccgctatggc tgagaggccc ttccagtgtc gaatctgcat gcgtaacttc agtcagtcct 1260
 ccgacctgtc ccgccacatc cgcaaccaca ccggcgagaa gccttttgcc tgtgacattt 1320

 gtgggaggaa atttgccctg aagcacaacc tgctgacca taccaagata cacacgggcg 1380
 agaagccctt ccagtgtcga atctgcatgc agaacttcag tgaccagtcc aacctgcgcg 1440
 cccacatccg cacccacacc ggcgagaagc cttttgectg tgacatttgt gggaggaaat 1500
 ttgcccgcaa ctctccctg accatgcata ccaagataca caccggagag cgcggcttcc 1560
 agtgtcgaat ctgcatgctt aacttcagtc tgcgccacga cctggagcgc cacatccgca 1620
 cccacaccgg cgagaagcct tttgectgtg acatttgggg gaggaattt gcccaccgct 1680
 ccaacctgaa caagcatacc aagatacacc tgcggggatc ccagctggtg aagagcgagc 1740

 tggaggagaa gaagtccgag ctgcggcaca agctgaagta cgtgccccac gagtacatcg 1800
 agctgatcga gatcgccagg aacagcacc aggaccgat cctggagatg aaggtgatgg 1860
 agttcttcat gaaggtgtac ggctacaggg gaaagcacct gggcggaagc agaaagcctg 1920
 acggcgccat ctatacagtg ggcagcccca tcgattacgg cgtgatcgtg gacacaaagg 1980
 cctacagcgg cggctacaat ctgagcatcg gccaggccga cgagatgcag agatactgca 2040

aggagaacca gaccgggaat aagcacatca accccaacga gtggtggaag gtgtacccta 2100
gcagcgtgac cgagttcaag ttctgttcg tgagcgcca cttcaagggc aactacaagg 2160

cccagctgac caggctgaac cgcaaaacca actgcaatgg cgccgtgctg agcgtggagg 2220
agctgctgat cggcggcgag atgatcaaag ccggcacct gacactggag gaggtgcggc 2280
gcaagttcaa caacggcgag atcaacttct gataactcga gtctagaaat caacctctgg 2340
attacaaaat ttgtgaaaga ttgactgata ttcttaacta tgttgctcct tttacgtgt 2400
gtggatatgc tgctttaatg cctctgtatc atgctattgc ttcccgtagc gctttcgttt 2460
tctcctcctt gtataaatcc tggttgctgt ctctttatga ggagtgtgg cccgttgtcc 2520
gtcaacgtgg cgtggtgtgc tctgtgtttg ctgacgcaac ccccactggc tggggcattg 2580

ccaccactg tcaactcctt tctgggactt tcgctttccc cctcccgatc gccacggcag 2640
aactcatcgc cgctgcctt gcccgtgct ggacaggggc taggttctg ggcaactgata 2700
attccgtggt gttgtcggg aaatcatcgt cctttccttg gctgctcgcc tgtgttgcca 2760
actggatcct gcgcgggacg tcttctgct acgtccctc ggctctcaat ccagcggacc 2820
tcccttccc aggccttctg ccggttctgc ggcctctccc gcgtcttcgc tttcgccctc 2880
cgacgagtcg gatctcctt tgggcccct ccccgctgg cttagcctgtg ccttctagtt 2940
gccagccatc tgtgtttgc cctcccccg tgccttctt gaccctggaa ggtgccactc 3000

cactgtcct ttctaataa aatgagaaa ttgcatcgca ttgctgagt aggtgicatt 3060
ctattctggg ggtgggggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 3120
ggcatgctgg ggatgcggtg ggctctatgc ggccgctcg agcgcaggaa cccctagtga 3180
tggagtggc cactccctct ctgcgcgctc gctcgctcac tgaggccgcc cgggctttgc 3240
ccgggcggcc tcagtgagcg agcgagcgcg cag 3273

<210> 57
<211> 516
<212> DNA
<213> Artificial Sequence
<220><223> Description of Artificial Sequence: Synthetic
polynucleotide

<400> 57
gccgctatgg ctgagaggcc cttccagtgt cgaatctgca tgcgtaactt cagtcagtcc 60
tccgacctgt cccceacat ccgcacccac accggcgaga agccttttgc ctgtgacatt 120
tgtgggagga aattgcctt gaagcacaac ctgctgacct ataccaagat acacacgggc 180

gagaagccct tccagtgtcg aatctgcatg cagaacttca gtgaccagtc caacctgctc 240
 gccccatcc gcacccacac cggcgagaag ccttttgctt gtgacatttg tgggaggaaa 300
 ttgtcccga acttctccct gaccatgcat accaagatac acaccggaga gcgctgcttc 360
 cagtgtcgaa tctgcatgct taacttcagt ctgcccacg acctggagcg ccacatccgc 420

accacaccg gcgagaagcc ttttgctgtgacatttggt ggaggaaatt tgcccaccgc 480
 tccaacctga acaagcatac caagatacac ctgctg 516

<210> 58

<211> 594

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 58

cagctggtga agagcgagct ggaggagaag aagtccgagc tgccgcacaa gctgaagtac 60
 gtgccccacg agtacctga gctgatcgag atcgccagga acagcaccca ggaccgcatc 120
 ctggagatga aggtgatgga gttcttcatg aaggtgtacg gctacagggg aaagcacctg 180

ggcggaaaca gaaagcctga cggcgccatc tatacagtgg gcagcccat cgattacggc 240
 gtgatcgtgg acacaaaggc ctacagcggc ggctacaatc tgagcatcgg ccaggccgac 300
 gagatgcaga gatacgtgaa ggagaaccag acccggaata agcacatcaa cccaacgag 360
 tggatggaagg tgiaccctag cagcgtgacc gaggttcaagt tcctgttcgt gagcggccac 420
 ttcaaggca actacaaggc ccagctgacc aggctgaacc gcaaaacaa ctgcaatggc 480
 gccgtgctga gcgtggagga gctgctgac ggcgcgaga tgatcaaagc cggcacctg 540
 aactggagg aggtgcggcg caagttcaac aacggcgaga tcaacttctg ataa 594

<210> 59

<211> 2474

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 59

ctgcgcgctc gctcgtcac tgaggccgcc cgggcaaagc ccggcgctcg ggcgacctt 60
 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gaggggcaa ctccatcact 120

aggggttctt gcgccctaag cttgagcggg gttccaattg tactgtacag aacctatggtc 180
 acatgtttta cgctagcgtg ccgacctggt aaactgatca gtgggtgcac ttaggactgc 240
 gtcttacgct aatcacatgc gtgcggccgc tttattctat tttcccagta aaataaagtt 300

 ttagtaaacct ctgcatcttt aaagaattat tttggcattt atttctaaaa tggcatagta 360
 ttttgtatgt gtgaagtctt acaaggttat cttaataata aaattcaaac atcctaggtta 420
 aaaaaaaaaa aaggtcagaa ttgttttagtg actgtaattt tcttttgcgc actaaggaaa 480
 gtgcaaagta acttagagtg actgaaactt cacagaatag ggttgaagat tgaattcata 540
 actatcccaa ggtaccacta aagaattatt cttttacatt tcagtttttc ttgatcatga 600
 aaacgccaac aaaatcctga accggcccaa gcggtacaac tcaggcaagc tggagagtt 660
 cgtgcagggc aacctggaac gggagtgcac ggaagagaag tgcagcttcg aggaagcccg 720

 ggaggtgttc gagaacaccg agcggaccac cgagttctgg aagcagtacg tggacggcga 780
 ccagtgcgag tcaaacacct gcctgaacgg cggcagctgc aaggacgata tcaacagcta 840
 cgagtgtctg tgccccctcg gcttcgaggg caagaactgc gagctggacg tgacctgcaa 900
 catcaagaac ggccgctcgg agcagttctg caagaacagc gccgacaaca aggtggtgtg 960
 ctcatgcact gagggtacc ggctggccga gaaccagaag agctgcgagc ccgccgtgcc 1020
 cttcccctgc ggcagagtgt ccgtgagcca gaccagcaag ctgaccaggg ccgaggccgt 1080
 gttccctgac gtggactacg tgaactcaac cgaggccgag acaatcctgg acaacatcac 1140

 ccagagcacc cagtccttca acgacttcac ccgggtggtg ggcggcgagg acgccaagcc 1200
 cggccagttc ccttggcagg tgggtctgaa cggcaagggt gacgccttct gcggcggctc 1260
 aatcgtgaac gagaagtgga tcgtgacagc cgcccactgc gtggagacag gcgtgaagat 1320
 caccgtggtg gccggcgaac acaatcga gaaaccgag cacaccgagc agaaacggaa 1380
 cgtgatccgg attatecccc accacaacta caacgccgc atcaacaagt acaaccacga 1440
 tatcgccctg ctggaactgg acgagcctct ggtgctgaat tcatcgtga ccccatctg 1500
 tatcgccgac aaagagtaca ccaacatctt tctgaagttc ggcagcggct acgtgtccgg 1560

 ctggggcagg gtgtccaca agggccgag cgccctggtg ctgcagtacc tgcgggtgcc 1620
 cctggtggac agagccacct gcctgcggtc aaccaagttc accatctaca acaacatgtt 1680
 ctgcgccggc tccacgagg gcggcaggga cagctgccag ggcgacagcg gcggacccca 1740
 cgtgaccgag gtggagggca ccagcttctt gaccggcatc atctcatggg gcgaggaatg 1800
 cgccatgaag ggcaagtacg gaatctacac taaggtgtca agatacgtga actggatcaa 1860
 agagaaaacc aagctgacct gagttaaac tgtgccttct agttgccagc catctgttgt 1920

ttgcccctcc cccgtgcctt ccttgaccct ggaaggtgcc actcccactg tcctttccta 1980

ataaaatgag gaaattgcat cgcattgtct gagtagtgt cattctattc tggggggtgg 2040

ggtggggcag gacagcaagg gggaggattg ggaagacaat agcaggcatg ctggggatgc 2100

ggtgggctct atggaccggt ctatccattg cactatgctt tatttaaaaa ccacaaaacc 2160

tgtgctgttg atctcataaa tagaacttgt atttatattt attttcattt tagtctgtct 2220

ggatccacaa attaatcgaa cctgcagctg atatcgacgc ttaagtaggg cttagcaaac 2280

gcgtctccaa cgtttcgccc ttaacacccc acatagttag tggcttagt agtccgggtg 2340

tttaactga aagataactc gagcgcagga acccctagtg atggagttag ccaactcctc 2400

tctgcgctcgct cgctcgctca ctgaggccgc ccgggctttg cccgggcggc ctcagtgage 2460

gagcagcgc gcag 2474

<210> 60

<211> 280

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 60

tttattctat tttcccagta aaataaagt ttagtaaact ctgcatcttt aaagaattat 60

tttggcattt atttctaaaa tggcatagta ttttgtattt gtgaagtctt acaagttat 120

cttattaata aaattcaaac atcctaggta aaaaaaaaaa aaggtcagaa ttgtttagt 180

actgtaattt tcttttgcgc actaaggaaa gtgcaaagta acttagagtg actgaaactt 240

cacagaatag ggttgaagat tgaattcata actatcccaa 280

<210> 61

<211> 28

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

oligonucleotide

<400> 61

actaaagaat tattctttta catttcag 28

<210> 62

<211> 1298

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 62

tttttcttga tcatgaaaac gccacaacaaa tctgaaccg gcccaagcgg tacaactcag	60
gcaagctgga agagtctgtg caggcaacc tggacggga gtgcatgga gagaagtgca	120
gcttcgagga agcccgggag gtgttcgaga acaccgagcg gaccaccgag ttctggaagc	180
agtacgtgga cggcgaccag tgcgagtcaa acccctgcct gaacggcggc agctgcaagg	240
acgatataca cagctacgag tgctggtgcc ccttcggctt cgaggcaag aactgcgagc	300
tggacgtgac ctgaacatc aagaacggcc gctgcgagca gttctgcaag aacagcggcg	360
acaacaaggt ggtgtgctca tgcaactgagg gctaccggct ggccgagaac cagaagagct	420
gcgagcccgc cgtgcccttc cctgcggca gagggtccgt gagccagacc agcaagctga	480
ccagggccga ggccgtgttc cctgacgtgg actacgtgaa ctcaaccgag gccgagacaa	540
tcttgacaaa catcaccagc agcaccagc cttcaacga cttcaccggc gtggtggcgc	600
gagagcgc caagcccggc cagttccctt ggcagggtgt gctgaacggc aaggtggacg	660
ccttctgagg cggtcaatc gtgaacgaga agtggatcgt gacagccgc cactgcgtgg	720
agacaggcgt gaagatcacc gtggtggccg gcgaacacaa tatcaggaa accgagcaca	780
ccgagcagaa acggaactg atccggatta tccccacca caactacaac gccgcatca	840
acaagtacaa ccacgatatc gccctgctgg aactggacga gcctctggtg ctgaattcat	900
acgtgacccc catctgtatc gccgacaaag agtacaccaa catctttctg aagttcggca	960
gaggtactgt gtcggctgg ggcagggtgt tccacaaggc ccgagcgcct ctggtgctgc	1020
agtacctgcg ggtgccctg gtggacagag ccacctgcct gcggtcaacc aagttacca	1080
tctacaacaa catgttctgc gccgcttcc acgagggcgg caggacagc tgccaggcgc	1140
acagcggcgg accccactg accgaggtgg agggcaccag ctttctgacc ggcatcatct	1200
catggggcga ggaatgcgcc atgaaggcga agtacggaat ctacactaag gtgtcaagat	1260
acgtgaactg gatcaaaag aaaaccaagc tgacctga	1298

<210> 63

<211> 225

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 63

ctgtgcttc tagttgccag ccatctgttg ttgcccctc ccccgctcct tccttgacct 60
 tggaaggtgc cactcccact gtcctttcct aataaaatga ggaaattgca tcgattgtc 120
 tgagtaggtg tcattctatt ctggggggtg ggtggggca ggacagcaag ggggaggatt 180
 gggaagacaa tagcaggcat gctgggatg cgtgggctc tatgg 225

<210> 64

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 64

ctatccattg cactatgctt tatttaaaaa ccacaaaacc tgtgctgtg atctcataaa 60
 tagaacttgt atttatattt attttcattt tagtctgtct 100

<210> 65

<211> 2758

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 65

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccggcgctcg ggcgaccttt 60

 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120
 aggggttctt gcggcctaag cttgagcggg gttccaattg tactgtacag aacctggctc 180
 acatgtttaa cgctagcgtg ccgacctggt aaactgatca gtgggtgcac ttaggactgc 240
 gtcttacgct aatcacatgc gtgcggccgc tttattctat tttcccagta aaataaagtt 300
 ttagtaaaact ctgcatcttt aaagaattat tttggcattt atttctaaaa tggcatagta 360
 ttttgtatgt gtgaagtctt acaaggttat cttattaata aaattcaaac atcctaggtg 420
 aaaaaaaaaa aaggtcagaa ttgttttagtg actgtaattt tcttttgcgc actaaggaaa 480

 gtgcaaagta acttagagtg actgaaactt cacagaatag ggttgaagat tgaattcata 540

actatcccaa ggiaccacta aagaattatt cttttacatt tcagttagcg aaaccaggc 600
caactcaact acagatgcgc ttaacgtcct gctcatcatc gtggacgatt tgcggccgtc 660
gcttggctgc tatggagata agctcgtccg ctgcgccaac atcgatcagt tggcctcaca 720
ctcactgctt ttccaaaatg cgtttgcgca gcaggctgtc tgtgcacctt caagagtctc 780
attcttgacc gggcgacgcc ctgacacaac gcggctgtac gacttcaaca gctactggag 840
agtccacgcg ggtaactttt caactatccc acagtacttt aaagagaacg gatacgtgac 900

aatgagcgtg ggaaaggtct ttcaccccgg catctcctcg aatcacaccg acgattcgcc 960
ctactcgtgg tcgtttcctc cctaccatcc ttcgagcgag aagtatgaga acacgaaaac 1020
ttgtcgcgga cccgacggag agctgcacgc taatctgctg tgtccggtgg atgtcttggg 1080
cgtgcccag ggaacgctcc ccgacaagca gtcaacggag caggcgattc agttgctgga 1140
gaagatgaaa acaagcgcgt cgcttttctt cctcgccgtg gggtatcaca agccccatat 1200
tcctttccgc taccggaagg agttccagaa actttatcct ttggaaaaca tcactttggc 1260
accggaccgg gaagtccccg acggtctgcc acccgtggcc tacaatccct ggatggatat 1320

caggcagagg gaagatgtgc aggcaactca catctcagtc ccctacgggc ctattccagt 1380
cgattttcaa cgcaagattc ggcagtcgta ttttgcgtcg gtgtcctacc tcgatacgca 1440
agtaggtcga ctctgagcg cgcttgatga ccttcagctg gcaaattcca caatcatcgc 1500
ctttacgtcg gaccatgggt gggcgttggg agagcatgga gagtgggcaa agtatagcaa 1560
ttttgatgta gcaacgcacg tgcccctgat tttctacgtg ccgggtagaa cggcctcgt 1620
tcccaggca ggcgaaaaac tttttccta tctcgatcca ttcgactcgg cgagccagct 1680
tatggaaccg ggcagacaat ccatggactt ggtagaattg gtgtcccttt ttccgacct 1740

cgccgggttg gcgggcttgc aagtacccc tagatgcct gtaccgagct tccatgtgga 1800
actctgccgc gaagggaaaa acctctcaa acactttcgg ttcagggacc ttgaggagga 1860
cccctatctg ccagggaatc cgcgagagt gattgcctat tcccagtatc cgcgaccag 1920
cgatattcct caatggaact ccgataagcc ctccctcaaa gacatcaaga ttatgggta 1980
ctcgatcagg accatcgact atcgetacac agtgtgggta gggttcaatc ctgacgaatt 2040
cctcgcgaac ttttcggaca tccacgctgg tgagctgtat ttcglagact cggaccggtt 2100
gcaagatcac aatatgtata atgattccca aggaggagat ttgttccagc tgctcatgcc 2160

gtgataaaga tcctctgtcc ttctagtgc cagccatctg ttgtttgccc ctccccctg 2220
ccttcttga ccttgaagg tgccactccc actgtccttt cctaataaaa tgaggaaatt 2280
gcatcgatt gtctgagtag gtgtcattct attctggggg gtggggtggg gcaggacagc 2340
aagggggagg attgggaaga caatagcagg catgctgggg atgcggtggg ctctatggac 2400

cggtctatcc attgcaactat gctttattta aaaaccacaa aacctgtgct gttgatctca 2460
 taaatagaac ttgtatttat atttattttc attttagtct gtctggatcc acaaattaat 2520
 cgaacctgca gctgatatcg acgcttaagt agggccttagc aaacgcgtct ccaacgtttc 2580

 gccgttaaca cccacacatag tgagtggctct tagtagtccg ggtgtttaa ctgaaagata 2640
 actcgagcgc aggaaccctt agtgatggag ttggcactc cctctctgcg cgctcgctcg 2700
 ctcaactgagg ccgcccgggc tttgccgggc cggcctcagt gagcgagcga gcgcgcag 2758

 <210> 66
 <211> 1575
 <212> DNA
 <213> Artificial Sequence
 <220><223> Description of Artificial Sequence: Synthetic
 polynucleotide
 <400> 66
 agcgaacecc aggccaactc aactacagat gcgcttaacg tctgctcat catcgtggac 60
 gatttgcggc cgtcgcttgg ctgctatgga gataagctcg tccgctcgcc gaacatcgat 120

 cagttggcct cacactcact gcttttcaa aatgcgtttg cgcagcagc tgtctgtgca 180
 ctttcaagag tctcattctt gaccgggcga cgcctgaca caacgcggct gtacacttc 240
 aacagctact ggagagtcca cgcgggtaac tttcaacta tcccacagta ctttaagag 300
 aacggatacg tgacaatgag cgtgggaaag gtctttcacc ccggcatctc ctgcaatcac 360
 accgacgatt cgcctactc gtggtcgttt cctccctacc atccttcgag cgagaagtat 420
 gagaacacga aaacttgtcg cggaccgcag ggagagctgc acgctaactc gctgtgtccg 480
 gtggatgtct tggacgtgcc cgagggaacg ctccccgaca agcagtcaac ggagcaggcg 540

 attcagttgc tggagaagat gaaaacaagc gcgtgcctt tcttctcgc cgtggggtat 600
 cacaagcccc atattccttt ccgctaccg aaggagtcc agaaacttta tcctttggaa 660
 aacatcactt tggcaccgga cccggaagtc cccgacggtc tgccaccctg ggcctacaat 720
 ccctggatgg atatcaggca gagggaagat gtgcaggcac tcaacatctc agtcccctac 780
 gggcctattc cagtcgattt tcaacgcaag attcggcagt cgtattttgc gtcggtgtcc 840
 tacctcgata cgcaagtagg tcgacttctg agcgcgcttg atgacctca gctggcaaat 900
 tccacaatca tcgcctttac gtcggacat ggtgggcgt tgggagagca tggagagtgg 960

 gcaaagtata gcaattttga tgtagcaacg cacgtgcccc tgattttcta cgtgccgggt 1020
 agaacggcct cgcttcccga ggcaggcga aaacttttc cctatctcga tccattcgac 1080

tcggcgagcc agcttatgga accgggcaga caatccatgg acttggtaga attgggtgcc 1140
 ctttttccga ccctcgccgg gttggcgggc ttgcaagtac ccctagatg ccctgtaccg 1200
 agcttccatg tggaactctg ccgcaaggg aaaaacctcc tcaaacctt tcggttcagg 1260
 gaccttgagg aggacccta tctgccaggg aatccgcgag agttgattgc ctattcccag 1320
 tatccgcgac ccagcgatat tcctcaatgg aactccgata agccctcct caaagacatc 1380

aagattatgg ggtactcgat caggaccatc gactatcgct acacagtgtg gtagggttc 1440
 aatcctgacg aatcctcgc gaacttttcg gacatccacg ctggtgagct gtatttcgta 1500
 gactcggacc cgttgcaaga tcacaatatg tataatgatt cccaaggagg agatttgttc 1560
 cagctgctca tgccg 1575

<210> 67

<211> 1872

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 67

cacttggtec acgtcgacgc tgccagagcc ctgtggccgc ttcgaagatt ttggaggtca 60

acgggtttct gtctccct tcccactcg caagcagatc agtatgtact gtcattggat 120
 caacagctta acctcgcta tgtcggagca gtgcctcacc gcgggatcaa gcaagtaagg 180
 acacattggc tccttgaact cgtcaccacg agaggatcga cgggaagggg gctttcgtac 240
 aacttactc atctcgatgg ctatttggat ctctccgcg agaatcagtt gttgccaggc 300
 ttcgaattga tgggatcggc gagcgggcac ttacagact tcgaggacaa gcagcaagtg 360
 tttgagtgga aggacctcgt gtcgtcgctc gcgaggagat acattggtcg ctacggtttg 420
 gcgcatgtgt caaagtggaa cttcgaaacg tgaacgagc ccgatcatca cgattttgac 480

aacgtgtcaa tgaccatgca gggtttccct aactattacg acgcctgttc cgagggattg 540
 agggcagcat caccggcgt tcggctggga gggcctggtg atagctttca tacaccacct 600
 cgatcgccac tttcgtgggg gctgctgcgc cattgtcagc atggtacgaa cttcttcacc 660
 ggggaagcgg gggtaggct tgattacatc agcctccacc gaaagggagc gcggtcaagc 720
 atctcgattc tggagcagga gaaggtagtc gctcagcaga tccggcaact ctttccaag 780
 ttcgagaca cacctatcta caatgatgag gcagaccac ttgtgggatg gtcccttccg 840
 cagccatggc gcgagatgt gacttatgcc gcgatggtag tgaaagtcat cgcccagcac 900

cagaatctgc ttcttgcgaa tacgaccagc gcgtttcctt acgcgctttt gtcgaacgat 960
aatgccttcc tgcataatca cccccatccg tttgcgcaga ggactcttac ggcgcgattc 1020
caagtgaata acaccagacc gccgcacgtg cagctgttgc gaaaaccctg gttgactgcg 1080
atggggcttc tggcgttctg tgatgaggaa caactctggg ctgaagtgtc ccaggcgggg 1140
acagtacttg atagcaatca tacagtaggc gtgttgccgt cggcgcaccg accgcagga 1200
cccgcggatg cttggagggc agcggctctg atctacgcct cggacgatac tagggcacat 1260
cccaacagat cggtcgctgt cacccttcgc ctacagagggg tcccgcctgg tcccggcttg 1320

gtatactca ctagatatct cgacaatgga ctgtgcagcc cgcacggaga gtggcggagg 1380
ctgggacggc cgggtgttcc gacagccgag cagtttagac ggatgagggc cgctgaggac 1440
cccgtggcag cggcaccgag gccctcccg gcaggaggtc gcctcactct tcgaccggca 1500
ctgcggctgc cgtcccttct gctcgtacac gtctgcgcgc gaccgaaaa gccgcctgga 1560
caggtaacca ggctcagggc gctccccttg acgcaggggc agttggtact tgtctggtcg 1620
gacgaacacg tggggtccaa atgcttgtgg acgtatgaaa ttcagttttc ccaagacggg 1680
aaagcgtaca ctccggtgtc gcgcaaaccc tccacgttca acctcttcgt cttttcccca 1740

gacacgggag ccgtatcagg gtcgtaccga gtcagagccc tcgattattg ggcgaggcct 1800
gggccgttct cggaccctgt accatacttg gaagtgccgg tgcccagggg accgcctcgc 1860
cctggtaate ct 1872

<210> 68

<211> 247

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 68

gataatcaac ctctggatta caaaatttgt gaaagattga ctggtattct taactatgtt 60
gtcctttta cgctatgtgg atacgtgct ttaatgcctt tgtatcatgc tattgcttcc 120

cgtatggctt tcattttctc ctcttgtat aaatcctggt tagttcttgc cacggcggaa 180
ctcatcgccg cctgccttgc ccgctgctgg acaggggctc ggctgttggg cactgacaat 240
tccgtgg 247

<210> 69

<211> 592

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

polynucleotide

<400> 69

```

aatcaacctc tggattacaa aatttgtgaa agattgactg gtattcttaa ctatgttgc      60
ccttttacgc tatgtggata cgctgcttta atgcctttgt atcatgctat tgcttcccgt    120

atggctttca ttttctcctc cttgtataaa tcttgggtgc tgtctcttta tgaggagtgg    180
tggcccgttg tcaggcaacg tggcgtggtg tgcactgtgt ttgctgacgc aacccccact    240
ggttggggca ttgccaccac ctgtcagctc ctttccggga ctttcgcttt cccccccct    300
attgccacgg cggaaactcat cgccgcctgc cttgcccgct gctggacagg ggctcggctg    360
ttgggcactg acaattccgt ggtgttgcg gggaaatcat cgtcctttcc ttggctgctc    420
gcctgtgttg ccacctggat tctgcgcggg acgtccttct gctacgtccc ttcggccctc    480
aatccagcgg accttccttc ccgcggcctg ctgccgctc tgcggcctct tccgcgtctt    540

```

```

cgcttccgcc ctcagacgag tcggatctcc ctttgggccg cctcccgcgc tg          592

```

<210> 70

<211> 9

<212> PRT

<213> Unknown

<220><223> Description of Unknown:

"LAGLIDADG" family motif peptide

<400> 70

Leu Ala Gly Leu Ile Asp Ala Asp Gly

1 5

<210> 71

<211> 22

<212> PRT

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic

peptide

<400> 71

Asp Tyr Lys Asp His Asp Gly Asp Tyr Lys Asp His Asp Ile Asp Tyr

1 5 10 15

Lys Asp Asp Asp Asp Lys

20

<210> 72

<211> 3055

<212> DNA

<213> Artificial Sequence

<220><223> Description of Artificial Sequence: Synthetic polynucleotide

<400> 72

ctgcgcgctc gctcgcctcac tgaggccgcc cgggcaaagc cggggcgtcg ggcgaccttt	60
ggtcgcccgg cctcagttag cgagcgagcg cgcagagagg gaggggccaa ctccatcact	120
aggggttccct gcggcctaag cttgagcggga gttccaattg tactgtacag aacctatggtc	180
acatgtttaa cgctagcgtg ccgacctggt aaactgatca gtgggtgcac ttaggactgc	240
gtcttacgct aatcacatgc gtgcggccgc tttattctat tttcccagta aaataaagtt	300
ttagtaaaact ctgcatcttt aaagaattat tttggcattt atttctaaaa tggcatagta	360
ttttgtatgt gtgaagtctt acaaggttat cttattaata aaattcaaac atcctaggtta	420
aaaaaaaaaa aaggtcagaa ttgttttagtg actgtaattt tcttttgcgc actaaggaaa	480
gtgcaaagta acttagagtg actgaaactt cacagaatag ggttgaagat tgaattcata	540
actatcccaa ggtaccacta aagaattatt cttttacatt tcagcgactt tggccacgt	600
cgacgctgcc agagccctgt ggccgcttcg aagattttgg aggtcaacgg gtttctgtcc	660
tccccttccc cactcgaag cagatcagta tgtactgtca tgggatcaac agcttaacct	720
cgctatgtc ggagcagtgc ctcaccgagg gatcaagcaa gtaaggacac attggtctct	780
tgaactcgtc accacgagag gatcgacggg aagggggctt tcgtacaact tcaactcatct	840
cgatggctat ttggatctcc tccgcgagaa tcagttgttg ccaggcttcg aattgatggg	900
atcggcgagc gggcacttta cagacttcga ggacaagcag caagtgtttg agtggaaagga	960
cctcgtgtcg tcgctcgcga ggagatacat tggtcgctac ggtttggcgc atgtgtcaaa	1020
gtggaacttc gaaacgtgga acgagcccga tcatcacgat tttgacaacg tgtcaatgac	1080
catgcagggt ttcttaact attacgacgc ctgttccgag ggattgaggg cagcatcacc	1140
ggcgcttcgg ctgggagggc ctggtgatag ctttcataca ccacctgat cgccactttc	1200
gtgggggctg ctgcgccatt gtcacgatgg tacgaacttc ttcaccgggg aagcgggggt	1260
acggcttgat tacatcagcc tccaccgaaa gggagcggg tcaagcatct cgattctgga	1320
gcaggagaag gtagtcgctc agcagatccg gcaactcttt cccaagttcg cagacacacc	1380

tatctacaat gatgaggcag acccacttgt gggatgggcc cttccgcagc catggcgcgc 1440
 agatgtgact tatgccgcga tggtagtgaa agtcatcgcc cagcaccaga atctgcttct 1500

tgcgaatacg accagcgcgt ttccttacgc gcttttgcg aacgataatg cettcctgtc 1560
 atatcacccc catccgtttg cgcagaggac tcttacggcg cgattccaag tgaataacac 1620
 cagaccgccg cacgtgcagc tgttgcgaaa acccgtgttg actgcgatgg ggcttctggc 1680
 gttgcttgat gaggaacaac tctgggctga agtgtcccag gcggggacag tacttgatag 1740
 caatcataca gtaggcgtgt tggcgtcggc gcaccgaccg cagggaccgc cggatgcttg 1800
 gagggcagcg gtctgatct acgcctcgga cgatactagg gcacatcca acagatcggt 1860
 cgctgtcacc cticgcctca gaggggtccc gcctgggtccc ggcttgggat acgtcactag 1920

atatctcgac aatggactgt gcagccccga cggagagtgg cggaggctgg gacggccggt 1980
 gtttccgaca gccgagcagt ttagacggat gagggccgct gaggaccgcg tggcagcggc 2040
 accgaggccc ctcccggcag gaggtcgcct cactcttcga ccggcactgc ggctgcctgc 2100
 ctttctgctc gtacacgtct gcgcgcgacc cgaaaagccg cctggacagg taaccaggct 2160
 cagggcgcctc ccttgacgc aggggcagtt ggtacttgtc tggtcggacg aacacgtggg 2220
 gtccaaatgc ttgtggactg atgaaattca gttttccaa gacgggaaag cgtacactcc 2280
 ggtgtcgcgc aaacctcca cgttcaacct ctctgtcttt tcccagaca cgggagccgt 2340

atcagggtcg taccgagtca gagccctcga ttattgggcg aggcctgggc cgttctcgga 2400
 ccctgtacca tacttggaag tgccggtgcc caggggaccg ccctcgctg gtaatccttg 2460
 ataaagatct ctgtgccttc tagttgceag ccatctgttg tttgcccctc ccccgtgctt 2520
 tccttgacc tggaaggtgc cactcccact gtcttttct aataaaatga ggaaattgca 2580
 tcgcatgtc ttagtaggtg tcattctatt ctggggggtg ggggtgggca ggacagcaag 2640
 ggggaggatt gggaagacaa tagcaggcat gctgggatg cgggtggctc tatggaccgg 2700
 tctatccatt gcaatgct tttttaaaa accacaaaac ctgtgctgtt gatctcataa 2760

atagaacttg tttttatatt tttttcatt ttagtctgtc tggatccaca aattaatcga 2820
 acctgcagct gatatcgacg cttaagtagg gcttagcaaa cgcgtctcca acgtttcgcc 2880
 gttaacacc cacatagiga gtggcttag tagtccgggt gtttaactg aaagataact 2940
 cgagcgcagg aaccctagt gatggagtg gccactccct ctctgcgcgc tcgctcgtc 3000
 actgaggccg cccgggcttt gcccggcgg cctcagtgag cgagcgagcg cgcag 3055