



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2025-0044471
(43) 공개일자 2025년03월31일

- (51) 국제특허분류(Int. Cl.)
C12N 15/113 (2010.01) A01K 67/0275 (2024.01)
A61K 31/7105 (2006.01) A61P 21/00 (2006.01)
C12N 15/10 (2017.01) C12N 15/85 (2006.01)
C12N 15/86 (2006.01) C12N 9/22 (2006.01)
- (52) CPC특허분류
C12N 15/113 (2013.01)
A01K 67/0275 (2024.01)
- (21) 출원번호 10-2025-7009481(분할)
- (22) 출원일자(국제) 2016년11월30일
심사청구일자 2025년03월21일
- (62) 원출원 특허 10-2018-7018218
원출원일자(국제) 2016년11월30일
심사청구일자 2021년11월30일
- (85) 번역문제출일자 2025년03월21일
- (86) 국제출원번호 PCT/US2016/064285
- (87) 국제공개번호 WO 2017/095967
국제공개일자 2017년06월08일
- (30) 우선권주장
62/260,712 2015년11월30일 미국(US)
62/330,336 2016년05월02일 미국(US)
- (71) 출원인
듀크 유니버시티
미국, 노스캐롤라이나 27705, 더럼, 스위트 406,
2812 에르윈 로드
- (72) 발명자
게르스바치, 찰스, 에이.
미국 27712 노스캐롤라이나주 더럼 와일더니스 로드 3220
로빈슨-햄, 재클린, 엔.
미국 27705 노스캐롤라이나주 더럼 슈트 306 에르
윈 로드 2812 듀크 유니버시티 내
- (74) 대리인
양영준, 이상남

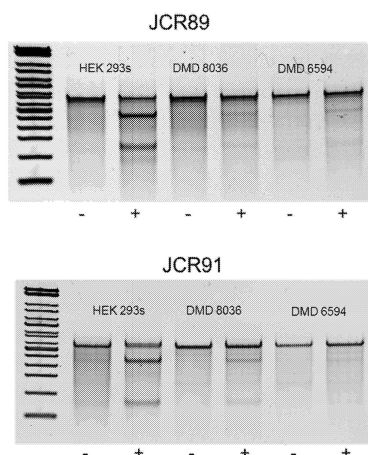
전체 청구항 수 : 총 62 항

(54) 발명의 명칭 유전자 편집에 의한 인간 디스트로핀 유전자의 교정을 위한 치료용 표적 및 사용 방법

(57) 요약

유전자 편집에 의해 인간 디스트로핀 유전자를 교정하기 위한 치료용 표적 및 사용 방법이 본원에 개시된다.

대표도



(52) CPC특허분류

A61K 31/7105 (2013.01)

A61P 21/00 (2018.01)

C12N 15/102 (2013.01)

C12N 15/8509 (2013.01)

C12N 15/86 (2013.01)

C12N 9/22 (2013.01)

C12N 2015/8527 (2013.01)

C12N 2310/20 (2017.05)

C12N 2830/008 (2013.01)

명세서

청구범위

청구항 1

제1 gRNA 분자 및 제2 gRNA 분자를 포함하는 DNA 표적화 조성물로서, 상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가:

(i) SEQ ID NO: 1의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ii) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iii) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iv) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(v) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vi) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(viii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ix) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(x) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(xi) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및

(xii) SEQ ID NO: 41의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 42의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자

로 구성된 그룹으로부터 선택되는, DNA 표적화 조성물.

청구항 2

제1항에 있어서,

클러스터링된, 규칙적으로 산재된, 짧은 회문 반복(Clustered Regularly Interspaced Short Palindromic Repeats) 관련(associated)(Cas) 단백질을 추가로 포함하는, DNA 표적화 조성물.

청구항 3

제2항에 있어서,

상기 Cas 단백질이 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 Cas9 분자를 포함하는, DNA 표적화 조성물.

청구항 4

제2항에 있어서,

상기 Cas 단백질이 SEQ ID NO: 45의 아미노산 서열을 갖는 *스타필로코커스 아우레우스*(*Staphylococcus aureus*) Cas9 분자를 포함하는, DNA 표적화 조성물.

청구항 5

제1항에 있어서,

상기 DNA 표적화 조성물이 SEQ ID NO: 83의 뉴클레오타이드 서열, 또는 SEQ ID NO: 84의 뉴클레오타이드 서열, 또는 SEQ ID NO: 37의 뉴클레오타이드 서열, 또는 SEQ ID NO: 38의 뉴클레오타이드 서열을 포함하는, DNA 표적화 조성물.

청구항 6

제1항의 DNA 표적화 조성물을 코딩하는 단리된 폴리뉴클레오타이드.

청구항 7

(a) 제1 가이드 RNA(gRNA) 분자,

(b) 제2 gRNA 분자, 및

(c) NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 적어도 하나의 Cas9 분자를 코딩하는 벡터로서,

상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가:

(i) SEQ ID NO: 1의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ii) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iii) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iv) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(v) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화

도메인을 포함하는 제2 gRNA 분자;

(vi) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(viii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ix) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(x) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(xi) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및

(xii) SEQ ID NO: 41의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 42의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자

로 구성된 그룹으로부터 선택되는, 벡터.

청구항 8

제7항에 있어서,

상기 벡터는 인간 *DMD* 유전자의 엑손 51 측면의 제1 및 제2 인트론에 제1 및 제2 이중 가닥 절단부를 형성하도록 구성되는, 벡터.

청구항 9

제7항에 있어서,

상기 벡터가 바이러스 벡터인, 벡터.

청구항 10

제9항에 있어서,

상기 벡터가 아데노-연관 바이러스(Adeno-associated virus, AAV) 벡터인, 벡터.

청구항 11

제10항에 있어서,

상기 AAV 벡터가 AAV8 벡터 또는 AAV9 벡터인, 벡터.

청구항 12

제7항에 있어서,

상기 벡터는 제1 gRNA 분자를 코딩하는 뉴클레오타이드 서열, 또는 제2 gRNA 분자를 코딩하는 뉴클레오타이드 서열, 또는 Cas9 분자를 코딩하는 뉴클레오타이드 서열, 또는 이들의 조합에 작동가능하게 연결된 조직-특이적

프로모터를 포함하는, 벡터.

청구항 13

제12항에 있어서,

상기 조직-특이적 프로모터가 근육 특이적 프로모터인, 벡터.

청구항 14

제1항의 DNA 표적화 조성물을 포함하는 세포.

청구항 15

제1항의 DNA 표적화 조성물을 포함하는 키트.

청구항 16

세포내 돌연변이체 디스트로핀 유전자를 교정하기 위한 조성물로서, 상기 세포에 투여되는 제1항의 DNA 표적화 조성물을 포함하는, 조성물.

청구항 17

대상체내 돌연변이체 디스트로핀 유전자를 게놈 편집하기 위한, 제1항의 DNA 표적화 조성물을 포함하는 게놈 편집 조성물로서, 상기 대상체에 투여되는, 게놈 편집 조성물.

청구항 18

제17항에 있어서, 대상체에 근육 내, 정맥 내 또는 이들의 조합으로 투여되는, 게놈 편집 조성물.

청구항 19

제16항에 있어서, 돌연변이체 디스트로핀 유전자를 교정하는 것이 뉴클레아제-매개된 비-상동성 말단 결합을 포함하는, 조성물.

청구항 20

돌연변이체 디스트로핀 유전자를 갖고 치료를 필요로 하는 대상체를 치료하기 위한 조성물로서, 상기 대상체에 투여되는 제1항의 DNA 표적화 조성물을 포함하는, 조성물.

청구항 21

대상체내 돌연변이체 디스트로핀 유전자를 게놈 편집하기 위한, 변형된 아데노-연관 바이러스 벡터로서, 제1항의 DNA 표적화 조성물을 코딩하는 제1 폴리뉴클레오타이드 서열, 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 Cas9 분자를 코딩하는 제2 폴리뉴클레오타이드 서열을 포함하는, 변형된 아데노-연관 바이러스 벡터.

청구항 22

제21항에 있어서,

상기 변형된 아데노-연관 바이러스 벡터는 SEQ ID NO: 39 또는 SEQ ID NO: 40의 뉴클레오타이드 서열을 포함하는, 변형된 아데노-연관 바이러스 벡터.

청구항 23

엑손 51을 포함하는 디스트로핀 유전자의 절편을 결실시키기 위한 조성물로서, 상기 조성물은:

(a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및

(b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO:

25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터

를 포함하며,

제1 벡터 및 제2 벡터는 인간 *DMD* 유전자의 엑손 51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 *디스트로핀* 유전자의 절편을 결실시키도록 구성되고,

상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가:

(i) SEQ ID NO: 1의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ii) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iii) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iv) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(v) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vi) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(viii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ix) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(x) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및

(xi) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자

로 구성된 그룹으로부터 선택되는, 조성물.

청구항 24

제23항에 있어서,

상기 절편은 50 염기쌍 내지 2,000 염기쌍의 길이를 갖는, 조성물.

청구항 25

제24항에 있어서,

상기 절편은 118 염기쌍, 233 염기쌍, 326 염기쌍, 766 염기쌍, 805 염기쌍 또는 1611 염기쌍의 길이를 갖는, 조성물.

청구항 26

제23항에 있어서,

상기 제1 Cas9 분자와 상기 제2 Cas9 분자가 동일한, 조성물.

청구항 27

제26항에 있어서,

상기 제1 Cas9 분자 및 상기 제2 Cas9 분자가 *스타필로코커스 아우레우스*(*Staphylococcus aureus*) Cas9 분자인, 조성물.

청구항 28

제27항에 있어서,

상기 제1 Cas9 분자 및 상기 제2 Cas9 분자가 돌연변이체 *스타필로코커스 아우레우스*(*Staphylococcus aureus*) Cas9 분자인, 조성물.

청구항 29

제23항에 있어서,

상기 제1 Cas9 분자와 상기 제2 Cas9 분자가 상이한, 조성물.

청구항 30

제29항에 있어서,

상기 제1 Cas9 분자 또는 상기 제2 Cas9 분자가 *스타필로코커스 아우레우스*(*Staphylococcus aureus*) Cas9 분자인, 조성물.

청구항 31

제23항에 있어서,

상기 제1 Cas9 분자 및/또는 제2 Cas9 분자가 SEQ ID NO: 45의 아미노산 서열을 포함하는 SaCas9 분자를 포함하는, 조성물.

청구항 32

제23항에 있어서,

상기 제1 벡터 및/또는 제2 벡터가 바이러스 벡터인, 조성물.

청구항 33

제32항에 있어서,

상기 제1 벡터 및/또는 상기 제2 벡터가 아데노-연관 바이러스(Adeno-associated virus, AAV) 벡터인, 조성물.

청구항 34

제33항에 있어서,

상기 AAV 벡터가 AAV8 벡터 또는 AAV9 벡터인, 조성물.

청구항 35

제23항에 있어서,

상기 *디스트로핀* 유전자가 인간 *디스트로핀* 유전자인, 조성물.

청구항 36

제23항에 있어서,

상기 제1 벡터는 SEQ ID NO: 39의 뉴클레오타이드 서열을 포함하고, 상기 제2 벡터는 SEQ ID NO: 40의 뉴클레오타이드 서열을 포함하는, 조성물.

청구항 37

제23항에 있어서,

약제에 사용하기 위한, 조성물.

청구항 38

제23항에 있어서,

뒤시엔느 근이영양증(Duchenne Muscular Dystrophy)의 치료에 사용하기 위한, 조성물.

청구항 39

제23항의 조성물을 포함하는 세포.

청구항 40

세포내 돌연변이체 *디스트로핀* 유전자를 교정하기 위한 조성물로서, 상기 조성물은:

(a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및

(b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터

를 포함하며,

상기 벡터는 인간 *디스트로핀* 유전자의 엑손 51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 *디스트로핀* 유전자의 절편을 결실시키도록 구성되고, 상기 벡터는 상기 세포에 투여되고,

상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가:

(i) SEQ ID NO: 1의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ii) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iii) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iv) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화

도메인을 포함하는 제2 gRNA 분자;

(v) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vi) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(viii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ix) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(x) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및

(xi) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자

로 구성된 그룹으로부터 선택되는, 조성물.

청구항 41

제40항에 있어서, 상기 돌연변이체 *디스트로핀* 유전자가 조기 중지 코돈, 손상된 해독 틀, 비정상 스플라이스 수신체 부위 또는 비정상 스플라이스 공여체 부위를 포함하는, 조성물.

청구항 42

제40항에 있어서, 상기 돌연변이체 *디스트로핀* 유전자가 조기 중지 코돈 및 절단된 유전자 생성물을 유발하는 틀 이동 돌연변이를 포함하는, 조성물.

청구항 43

제40항에 있어서, 상기 돌연변이체 *디스트로핀* 유전자가 해독 틀을 손상시키는 하나 이상의 엑손의 결실을 포함하는, 조성물.

청구항 44

제40항에 있어서, 상기 돌연변이체 *디스트로핀* 유전자의 교정이 조기 중지 코돈의 결실, 손상된 해독 틀의 교정, 또는 스플라이스 수신체 부위의 손상 또는 스플라이스 공여체 서열의 손상에 의한 스플라이싱의 조절을 포함하는, 조성물.

청구항 45

제40항에 있어서, 상기 돌연변이체 *디스트로핀* 유전자의 교정이 엑손 51의 결실을 포함하는, 조성물.

청구항 46

제40항에 있어서, 상기 돌연변이체 *디스트로핀* 유전자의 교정이 상동성-유도된 보수를 포함하는, 조성물.

청구항 47

제46항에 있어서, 세포에 투여되는 공여체 DNA를 추가로 포함하는, 조성물.

청구항 48

제40항에 있어서, 상기 돌연변이체 *디스트로핀* 유전자의 교정이 뉴클레아제 매개된 비-상동성 말단 결합을 포함하는, 조성물.

청구항 49

제40항에 있어서, 상기 세포가 근모세포 세포인, 조성물.

청구항 50

제40항에 있어서, 상기 세포가 뒤시엔느 근이영양증을 앓고 있는 대상체로부터의 세포인, 조성물.

청구항 51

제40항에 있어서, 상기 세포가 뒤시엔느 근이영양증을 앓고 있는 인간 대상체로부터의 근모세포인, 조성물.

청구항 52

제40항에 있어서, 상기 제1 gRNA 분자 및 제2 gRNA 분자가:

(i) SEQ ID NO: 1의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ii) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및

(iii) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자

로 구성된 그룹으로부터 선택되는, 조성물.

청구항 53

돌연변이체 *디스트로핀* 유전자를 갖고 치료를 필요로 하는 대상체를 치료하기 위한 조성물로서, 상기 조성물은:

(a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및

(b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터

를 포함하며,

상기 벡터는 인간 *디스트로핀* 유전자의 엑손 51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 *디스트로핀* 유전자의 절편을 결실시키도록 구성되고, 상기 벡터는 상기 대상체에 투여되고,

상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가:

(i) SEQ ID NO: 1의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ii) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iii) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(iv) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(v) SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vi) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(vii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(viii) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(ix) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자;

(x) SEQ ID NO: 14의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및

(xi) SEQ ID NO: 11의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18의 뉴클레오타이드 서열을 포함하는 폴리뉴클레오타이드에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자

로 구성된 그룹으로부터 선택되는, 조성물.

청구항 54

제53항에 있어서, 상기 대상체가 뒤시엔느 근이영양증을 앓고 있는, 조성물.

청구항 55

제53항에 있어서, 상기 제1 벡터 및 제2 벡터가 대상체에게 근육 내, 정맥 내 또는 이들의 조합으로 투여되는, 조성물.

청구항 56

엑손 52 결실($\Delta 52$)을 갖는 인간 *디스트로핀* 유전자(hDMD)를 갖는 트랜스제닉 설치류 배아를 생성하는 방법으로서, 상기 방법은 제1항의 DNA 표적화 조성물을 설치류 배아에 투여하는 단계를 포함하는, 방법.

청구항 57

제56항에 있어서,

상기 설치류 배아가 마우스 배아인 방법.

청구항 58

제56항에 있어서,

상기 트랜스제닉 설치류 배아가 이중접합체 hDMD 또는 이중접합체 hDMD-Δ52인, 방법.

청구항 59

제56항에 있어서,

SEQ ID NO: 41의 뉴클레오타이드 서열에 결합하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 42의 뉴클레오타이드 서열에 결합하는 표적화 도메인을 포함하는 제2 gRNA 분자가 설치류 배아에 투여되어, 인간 디스트로핀 유전자의 엑손 52를 결실시키는, 방법.

청구항 60

제56항에 있어서,

SEQ ID NO: 27의 아미노산 서열을 포함하는 Cas 단백질을 설치류 배아에 투여하는 단계를 추가로 포함하는, 방법.

청구항 61

제56항의 방법에 의해 생성된 트랜스제닉 설치류 배아.

청구항 62

제56항의 트랜스제닉 설치류 배아로부터 생성된 트랜스제닉 설치류.

발명의 설명

기술 분야

[0001]

관련 출원에 대한 교차 참조

[0002]

본 출원은 2015년 11월 30일에 출원된 U.S. 가출원 번호 62/260,712 및 2016년 5월 2일에 출원된 U.S. 가출원 번호 62/330,336을 우선권으로 주장하며, 본원에 그 전문이 참조로 포함된다.

[0003]

정부 이해관계 진술

[0004]

본 발명은 국립 과학 재단(National Science Foundation) 대학원 연구 펠로우십 프로그램(Graduate Research Fellowship Program)이 수여하는 정부 지원으로 수행되었다. 미국 정부는 본 발명에 대해 소정 권리를 갖는다.

[0005]

기술분야

[0006]

본 발명은 클러스터링된, 규칙적으로 산재된, 짧은 회문 반복(Clustered Regularly Interspaced Short Palindromic Repeats, CRISPR)/CRISPR-관련(associated)(Cas) 9-기반 시스템 및 바이러스 전달 시스템을 사용한, 유전자 발현 변형, 게놈 조작 및 유전자의 게놈 변형 분야에 관한 것이다. 본 발명은 또한 골격근 및 심장 근육과 같은 근육내 유전자의 게놈 조작 및 게놈 변형 분야에 관한 것이다.

배경 기술

[0007]

합성 전사 인자는 조직 재생 촉진, 약물 스크리닝, 유전자 결함 보상, 침묵된 종양 억제제 활성화, 줄기 세포 분화 조절, 유전자 스크린 수행 및 합성 유전자 회로 생성을 포함하는, 포유류 시스템에서 많은 다양한 의학적 및 과학적 용도를 위한 유전자 발현을 제어하도록 조작되어왔다. 이러한 전사 인자는 내인성 유전자의 프로모터 또는 인핸서를 표적화하거나, 또는 전이유전자 조절을 위한 포유류 게놈에 직교하는 서열을 인식하도록 의도적으로 고안될 수 있다. 사용자-정의 서열을 표적화하는 새로운 전사 인자를 조작하기 위한 가장 일반적인 전략은 징크 핑거 단백질 및 전사-활성화제 유사 이펙터(TALE)의 프로그래밍 가능한 DNA-결합 도메인에 기초한다. 이 두 접근법은 고유한 DNA-결합 특이성을 가진 새로운 단백질을 조작하기 위해 상기 도메인의 단백질-DNA 상호작

용 원리를 적용하는 단계를 포함한다. 이러한 방법이 많은 용도에서 널리 성공했지만, 단백질-DNA 상호작용을 조작하는데 필요한 단백질 조작은 힘들 수 있으며, 전문 지식을 필요로 한다.

[0008] 또한, 상기 새로운 단백질은 항상 효과적이지는 않다. 그 이유는 아직 알려져 있지 않지만, 게놈 표적 부위에 결합하는 단백질에 대한 후성적 변형 및 염색질 상태의 효과와 연관될 수 있다. 또한, 상기 새로운 단백질뿐만 아니라 다른 성분이 각 세포에 전달되도록 하는 데는 어려움이 있다. 상기 새로운 단백질 및 그들의 다중 성분을 전달하기 위한 기존의 방법은 복제 수의 차이로 인해 각 세포에서 매우 가변적인 발현 수준을 유도하는 별도의 플라스미드 또는 벡터 상의 세포로의 전달을 포함한다. 또한, 형질감염 후 유전자 활성화는 플라스미드 DNA의 희석으로 인해 일시적이며, 임시적 유전자 발현은 치료 효과를 유도하기에 충분하지 않을 수 있다. 또한, 이 접근법은 쉽게 형질감염되지 않는 세포 유형에는 적용할 수 없다. 따라서, 상기 새로운 단백질의 또 다른 한계는 전사 활성화의 효능이다.

[0009] CRISPR/Cas9-기반 유전자 편집 시스템은 표적화된 유전자좌에서 부위-특이적 이중 가닥 절단을 도입하는데 사용될 수 있다. 이 DNA 절단은 자연적인 DNA-보수 기구를 자극하여 가능한 두 가지 보수 경로 중 하나를 만든다. 공여체 템플릿이 없는 경우, DNA의 작은 삽입 또는 결실로 이어지는 오류가-발생하기 쉬운 보수 경로인 비-상동성 말단 연결(nonhomologous end joining, NHEJ)에 의해 절단이 보수될 것이다. 이 방법은 표적화된 유전자 서열의 해독 틀을 의도적으로 손상, 결실 또는 변경시키는데 사용될 수 있다. 그러나, 공여체 주형이 뉴클레아제와 함께 제공된다면, 세포 기구는 상동성 재조합에 의해 절단을 보수할 것이며, 이것은 DNA 절단의 존재 하에 수십배 증가된다. 이 방법은 표적 부위에서 DNA 서열의 특정 변화를 도입하는데 사용될 수 있다. 조작된 뉴클레아제는 다양한 인간 줄기 세포 및 세포주에서의 유전자 편집 및 마우스 간에서의 유전자 편집에 사용되어왔다. 그러나, 상기 기술을 구현하기 위한 주요 장애물은 효과적이고 효율적이며 성공적인 게놈 수정을 용이하게 하는 방식으로 *생체내* 특정 조직에 전달되는 것이다.

[0010] 유전성 유전 질환은 미국 어린이들에게 치명적인 영향을 미친다. 이 질환은 현재 치료법이 없으며, 증상을 완화하려는 시도에 의해서만 관리될 수 있다. 수십 년 동안 유전자 치료 분야에서는 이러한 질병을 치료할 것을 약속했다. 그러나, 세포와 환자에게 치료 유전자를 안전하고 효율적으로 전달하는 것과 관련된 기술적인 장애물은 이러한 접근 방식을 제한했다. 뒤시엔느 근디스트로피(Duchenne Muscular Dystrophy, DMD)는 치명적인 유전 질환으로, 임상적으로 근육 소모, 보행 장애 및 기능적 디스트로핀 소실로 인하여 전형적으로 인생에서 세번째 십년에 사망하는 것을 특징으로 한다. DMD는 *디스트로핀* 유전자의 유전적 돌연변이 또는 자발적인 돌연변이의 결과이다. DMD를 유발하는 대부분의 돌연변이는 번역 해독 틀을 틀 밖으로 밀어내는, 엑손(exon) 결실의 결과이다.

[0011] 디스트로핀은 근육 세포 온전성 및 기능의 조절에 관여하는 단백질 복합체의 핵심 구성요소이다. DMD 환자는 전형적으로 아동기 동안 스스로 신체적으로 지지하는 능력을 잃고, 침대 시절 동안 점차 약해져서, 20대에 사망한다. DMD에 대한 현재의 실험적인 유전자 치료 전략은 일시적인 유전자 전달 비히클의 반복 투여를 필요로 하거나, 또는 게놈 DNA 내 외래 유전 물질의 영구적인 통합에 의존한다. 이 두 가지 방법 모두 심각한 안전성 문제가 있다. 또한, 이러한 전략은 크고 복잡한 *디스트로핀* 유전자 서열을 전달할 수 없기 때문에 제한적이었다. *디스트로핀* 유전자의 돌연변이를 가진 환자를 교정하고 치료하기 위한 보다 정확하고 효율적인 유전자 편집 도구가 필요하다.

발명의 내용

해결하려는 과제

과제의 해결 수단

[0012] 본 발명은 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 42에 개시된 뉴클레오타이드 서열 또는 이의 상보체를 포함하는 표적화 도메인을 포함하는 가이드 RNA(gRNA)에 관한 것이다.

[0013] 본 발명은 또한 제1 gRNA 및 제2 gRNA를 포함하는 DNA 표적화 조성물에 관한 것이다. 제1 gRNA 분자 및 제2 gRNA 분자는 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 42에 개시된 뉴클레오타이드 서열 또는 이의 상보체를 포함하는 표적화 도메인을 포함하는 가이드 RNA(gRNA)에 관한 것이다.

NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 42에 개시된 뉴클레오타이드 서열 또는 이의 상보체를 포함하는 표적화 도메인을 포함한다. 제1 gRNA 분자 및 제2 gRNA 분자는 상이한 표적화 도메인을 포함한다.

[0014] 본 발명은 또한 상기 기술된 gRNA 분자 또는 상기 기술된 DNA 표적화 조성물을 포함하는 단리된 폴리뉴클레오타이드에 관한 것이다.

[0015] 본 발명은 상기 기술된 gRNA, 상기 기술된 DNA 표적화 조성물 또는 상기 기술된 단리된 폴리뉴클레오타이드를 포함하는 벡터에 관한 것이다.

[0016] 본 발명은 또한 상기 기술된 DNA 표적화 조성물을 포함하는 벡터에 관한 것이다.

[0017] 본 발명은 또한 (a) 제1 가이드 RNA(gRNA) 분자, (b) 제2 gRNA 분자, 및 (c) NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 적어도 하나의 Cas9 분자를 코딩하는 벡터에 관한 것이다. 제1 gRNA 분자 및 제2 gRNA 분자는 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 42에 개시된 뉴클레오타이드 서열 또는 이의 상보체를 포함하는 표적화 도메인을 포함한다. 제1 gRNA 분자 및 제2 gRNA 분자는 상이한 표적화 도메인을 포함한다.

[0018] 본 발명은 또한 상기 기술된 gRNA, 상기 기술된 DNA 표적화 조성물, 상기 기술된 단리된 폴리뉴클레오타이드 또는 상기 기술된 벡터를 포함하는 세포에 관한 것이다.

[0019] 본 발명은 또한 상기 기술된 gRNA, 상기 기술된 DNA 표적화 시스템, 상기 기술된 단리된 폴리뉴클레오타이드, 상기 기술된 벡터, 또는 상기 기술된 세포 및 임의로 사용 지침서를 포함하는 키트에 관한 것이다.

[0020] 본 발명은 또한 세포내 돌연변이체 디스트로핀 유전자를 교정하는 방법에 관한 것이다. 상기 방법은 세포에 상기 기술된 gRNA, 상기 기술된 DNA 표적화 시스템, 상기 기술된 단리된 폴리뉴클레오타이드 또는 상기 기술된 벡터를 투여하는 단계를 포함한다.

[0021] 본 발명은 또한 대상체내 돌연변이체 디스트로핀 유전자를 게놈 편집하는 방법에 관한 것이다. 상기 방법은 상기 기술된 gRNA, 상기 기술된 DNA 표적화 시스템, 상기 기술된 단리된 폴리뉴클레오타이드, 상기 기술된 벡터 또는 상기 기술된 세포를 포함하는 게놈 편집 조성물을 대상체에게 투여하는 단계를 포함한다.

[0022] 본 발명은 또한 돌연변이체 디스트로핀 유전자를 갖는, 치료를 필요로 하는 대상체를 치료하는 방법에 관한 것이다. 상기 방법은 상기 기술된 gRNA, 상기 기술된 DNA 표적화 시스템, 상기 기술된 단리된 폴리뉴클레오타이드, 상기 기술된 벡터 또는 상기 세포를 상기 대상체에게 투여하는 단계를 포함한다.

[0023] 본 발명은 또한 상기 기술된 gRNA를 코딩하는 제1 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 Cas9 분자를 코딩하는 제2 폴리뉴클레오타이드 서열을 포함하는, 대상체내 돌연변이체 디스트로핀 유전자를 게놈 편집하기 위한 변형된 아데노-연관 바이러스 벡터에 관한 것이다.

[0024] 본 발명은 또한 엑손 51을 포함하는 디스트로핀 유전자의 절편을 결실시키기 위한 조성물에 관한 것으로서, 상기 조성물은: (a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및 (b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터를 포함한다. 제1 및 제2 gRNA 분자 각각은 19 내지 24개의 뉴클레오타이드 길이의 표적화 도메인을 가지며, 제1 벡터 및 제2 벡터는 인간 DMD 유전자의 엑손 51 측면의 제1 및 제2 인트론에 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 디스트로핀 유전자의 절편을 결실시키도록 구성된다.

[0025] 본 발명은 또한, 상기 기술된 조성물을 포함하는 세포에 관한 것이다.

[0026] 본 발명은 또한, 세포내 돌연변이체 디스트로핀 유전자를 교정하는 방법에 관한 것으로서, 상기 방법은: (a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터를 포함한다. 제1 및 제2 gRNA 분자 각각은 19 내지 24개의 뉴클레오타이드 길이의 표적화 도메인을 가지며, 제1 벡터 및 제2 벡터는 인간 DMD 유전자의 엑손 51 측면의 제1 및 제2 인트론에 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 디스트로핀 유전자의 절편을 결실시키도록 구성된다.

NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및 (b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터를 세포에 투여하는 단계를 포함한다. 제1 gRNA 및 제2 gRNA 분자 각각은 19 내지 24개의 뉴클레오타이드 길이의 표적화 도메인을 가지며, 상기 벡터는 인간 *디스트로핀* 유전자의 엑손 51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 *디스트로핀* 유전자의 절편을 결실시키도록 구성된다.

[0027] 본 발명은 또한 돌연변이체 *디스트로핀* 유전자를 갖는, 치료를 필요로 하는 대상체를 치료하는 방법에 관한 것이다. 상기 방법은: (a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및 (b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터를 대상체에 투여하는 단계를 포함한다. 제1 gRNA 및 제2 gRNA 분자 각각은 19 내지 24개의 뉴클레오타이드 길이의 표적화 도메인을 가지며, 상기 벡터는 인간 *디스트로핀* 유전자의 엑손 51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 *디스트로핀* 유전자의 절편을 결실시키도록 구성된다.

[0028] 본 발명은 또한 엑손 52 결실($\Delta 52$)을 갖는 인간 *디스트로핀* 유전자(hDMD)를 갖는 트랜스제닉 설치류 배아를 생성하는 방법에 관한 것이다. 상기 방법은 상기 기술된 gRNA, 상기 DNA 표적화 시스템, 상기 기술된 단리된 폴리뉴클레오타이드, 상기 기술된 벡터, 상기 기술된 변형된 아데노-연관 바이러스 벡터 또는 상기 조성물을 설치류 배아에 투여하여 인간 *디스트로핀* 유전자의 엑손 52가 결실된 단계, 및 인간 *디스트로핀* 유전자의 엑손 52가 결실된 트랜스제닉 설치류 배아를 선별하는 단계를 포함하며, 상기 설치류 배아는 정상적인 인간 *디스트로핀* 유전자를 포함한다.

[0029] 본 발명은 또한 상기 기술된 방법에 의해 생성된 트랜스제닉 설치류 배아에 관한 것이다.

[0030] 본 발명은 또한 상기 기술된 트랜스제닉 설치류 배아에서 생성된 트랜스제닉 설치류에 관한 것이다.

도면의 간단한 설명

[0031] 도 1은 측량 검정에 의해 측정되는, HEK293T 세포(야생형 *디스트로핀* 유전자) 및 DMD 환자 근모세포 세포주(DMD 8036 및 DMD 6594, 각각 돌연변이 형태의 *디스트로핀* 유전자를 가짐)에서 인간 *디스트로핀* 유전자를 표적화하는 개별 gRNA JCR89 및 JCR91의 활성을 나타낸다.

도 2A 및 2B는 SaCas9 및 gRNAs JCR89 및 JCR91과의 동시 처리에 의한, HEK293T 세포 및 DMD 근모세포(DMD 8036 및 DMD 6594)의 게놈 DNA(도 2A) 및 DMD 근모세포 근모세포(도 2B)로부터의 cDNA에서 엑손 51의 결실을 보여준다.

도 3은 2개의 바이러스 벡터상의 SaCas9 및 gRNA JCR89 및 JCR91의 근육 조직으로의 동시 전달을 위한 AAV-기반의 생체내 시스템을 도시한다.

도 4는 SaCas9 및 gRNA를 보유하는 바이러스 벡터를 앞 정강근(TA)에 국소 AAV8 전달한 후 인간 DMD 유전자(hDMD/*mdx* 마우스)를 갖는 트랜스제닉 마우스에서 인간 엑손 51의 결실 검출을 나타낸다.

도 5는 꼬리 혈관정맥 주사를 통한 전신성 AAV8 전달 후 hDMD 유전자를 보유하는 트랜스제닉 마우스에서 인간 엑손 51의 결실 검출을 나타낸다.

도 6은 인간과 붉은털 원숭이 게놈 사이에 보존되는 다양한 gRNA 표적을 나타낸다(표 2에서 gRNA의 서열 참조). 각 gRNA의 위치는 인간 *디스트로핀* 유전자의 엑손 51과 관련되어 표시된다.

도 7은 측량 검정에 의해 측정된 바와 같은, 인간 HEK293T 세포의 형질감염 후 개별적인 gRNA의 활성을 나타낸다.

도 8은 CasOFFinder 프로그램(Bae et al. (2014) *Bioinformatics* 30: 1473-1475)을 사용하여 예측된 후보 gRNA의 특이성을 나타낸다.

도 9는 게놈 DNA의 PCR에 의해 측정된 바와 같은, HEK293T 세포 및 DMD 6594 세포에서 gRNA JCR157 및 JCR160

에 의한 엑손 51의 결실을 나타낸다.

도 10은 측량 검정에 의해 측정된, gRNA JCR157의 다양한 표적 길이: 19, 20, 21, 22 및 23개의 뉴클레오타이드의 활성을 나타낸다.

도 11은 측량 검정에 의해 측정된, gRNA JCR160의 다양한 표적 길이: 19, 20, 21, 22 및 23개의 뉴클레오타이드의 활성을 나타낸다.

도 12는 게놈 DNA의 PCR에 의해 측정된, 다양한 길이(21, 22 또는 23개의 뉴클레오타이드)의 JCR157 및 JCR160을 조합하여 생성된 결실을 나타낸다.

도 13은 측량 검정에 의한 시험관내 표적-상 뉴클레아제 활성을 나타낸다.

도 14는 게놈 DNA에서 엑손 51의 시험관내 결실을 나타낸다.

도 15는 7일 동안 분화된 인간 DMD 근모세포내 cDNA에서 엑손 51의 시험관내 결실을 나타낸다.

도 16은 DMD 환자 근모세포의 cDNA에서 시험관내 엑손 47 내지 52 접합을 나타낸다.

도 17은 건강한 hDMD/mdx 마우스로부터 시작한 $\Delta 52/mdx$ 마우스에 대한 설계를 도시한다.

도 18은 시험관내 가이드 검증을 나타낸다: 개별(측량 검정).

도 19는 시험관내 가이드 검증을 나타낸다: 페어드(paired): HEK293T 세포의 게놈 DNA에서 엑손 51의 결실은 gRNA 쌍을 사용하여 생성되었다.

도 20은 DNA 미세주입 프로토콜의 개략도를 도시한다.

도 21은 마우스 육종의 개략도를 나타낸다.

도 22는 파운더 마우스의 지노타이핑(genotyping) 결과를 나타낸다.

도 23은 파운더 마우스 7, 63 및 76으로부터의 시퀀싱 결과의 일부를 나타낸다.

도 24는 추가 마우스 육종의 개략도를 도시한다.

도 25는 파운더 수컷 76 + mdx/mdx 육종 결과로부터의 한배자손 5(수컷만)의 지노타이핑을 나타낸다. "63"은 파운더 수컷이다(그러나, 이 경우 부모가 아님). "293"은 HEK293T 세포 게놈 DNA 대조군을 나타낸다.

도 26은 파운더 수컷 63 + mdx/mdx 육종 결과로부터의 한배자손 1의 지노타이핑을 나타낸다.

도 27은 새끼 54497 및 54498의 392bp 시퀀싱 판독의 일부를 도시한다.

도 28은 새끼 54497 및 54498로부터의 심장 및 TA의 면역조직화학 염색을 나타낸다.

도 29는 $\Delta 52/mdx$ 마우스에 디스트로핀 단백질이 결핍된 것을 나타낸다.

도 30은 $\Delta 52/mdx$ 마우스에 DMD 유전자형과 일치하는 디스트로핀 단백질이 결핍된 반면, 건강한 hDMD/mdx 마우스가 디스트로핀을 발현한다는 것을 나타내는 웨스턴 블롯을 나타낸다.

도 31은 이동 및 탐색에 의해 나타난 바와 같이, mdx 마우스 및 hDMD/mdx 마우스와 비교된 $\Delta 52/mdx$ 마우스의 전반적인 활성을 나타낸다.

도 32는 제거를 위해 인트론 부위의 엑손 51의 상류 및 하류에 gRNA를 표적화하여 엑손 51을 스킵하기 위해 SaCas9 및 gRNA를 사용하는 $\Delta 52/mdx$ 마우스에 대한 교정 전략을 도시한다.

도 33은 SaCas9 및 gRNA JCR179 및 JCR183을 사용하여 DMD 환자 근모세포(DMD 6594 세포)에서 엑손 51 결실로부터 디스트로핀 단백질의 시험관내 복원을 나타낸다.

도 34는 gRNA 및 SaCas9 시스템을 사용하여 $\Delta 52/mdx$ 마우스를 치료하기 위한 실험적 설계를 나타낸다.

도 35는 오른쪽 TA 근육에서 생체내 엑손 51 결실을 나타낸다.

도 36은 오른쪽 TA 근육에서 생체내 엑손 51 결실을 나타낸다.

도 37은 처리된 TA 근육에서의 생체내 디스트로핀 단백질 복원을 나타낸다.

도 38은 처리된 TA 근육에서의 생체내 디스트로핀 단백질 복원을 나타낸다.

도 39는 이동된 총 거리에 대한 모든 시간 지점의 평균을 나타낸다.

도 40은 전체 양육 자세에 대한 모든 시간 지점의 평균을 나타낸다.

도 41은 16주간 처리되지않은 및 처리된 마우스의 그립 강도를 나타낸다.

도 42는 심장 조직의 cDNA PCR 결과를 나타낸다.

도 43은 도 42의 증폭된 cDNA PCR 밴드의 시퀀싱을 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0032] 본 명세서에 기술된 바와 같이, 특정 방법 및 조작된 gRNA는 유전 질환, 예를 들어 DMD에 관여하는 디스트로핀 유전자에서의 발현 변경, 게놈 조작 및 돌연변이의 효과를 수정하거나 감소시키기 위한 CRISPR/CRISPR-관련 (Cas) 9-기반 유전자 편집 시스템과 함께 유용하다는 것이 밝혀졌다. 상기 개시된 gRNA는 임상 번역에 보다 순응할 수 있는 표적 부위로 생성되었다. 예를 들어, *에스. 피오게네스*(*S. pyogenes*) Cas9(SpCas9)를 코딩하는 유전자는 다른 모든 필요한 조절 서열이 포함될 때 근육에 대한 전신 유전자 전달에 사용되는 벡터인, 아데노-연관 바이러스(AAV)에 의해 전달되기에는 너무 크다. 대신, SpCas9보다 약 1kb 작은, *에스. 아우레우스*(*S. aureus*) Cas9(SaCas9)와 함께 사용하기 위해 상기 개시된 gRNA를 선별하고, 스크리닝하였다. 인간과 붉은털 원숭이 게놈 사이에 보존된 서열상의 SaCas9-호환가능 표적이 되도록 표적 선택이 스크리닝되었으며, 이는 가능한 유전자 표적의 수를 크게 제한한다. 이 선택 기준은 인간과 붉은털 원숭이 모두에서 활성화된 gRNA 후보 물질을 허용하도록 선택하여, 비-인간 영장류 모델에서 전임상 시험을 용이하게 하였다. 인간과 붉은털 원숭이 디스트로핀 유전자 서열 모두를 표적화하는 개시된 gRNA는 CRISPR/Cas9-기반 시스템과 함께 사용하여 인간 디스트로핀 유전자의 엑손 51을 둘러싼 인트론 영역을 표적화하여, DMD 환자의 세포에서 기능성 디스트로핀의 발현을 복원시키기 위해 이 영역의 게놈 결실을 일으킬 수 있다.

[0033] 또한, 디스트로핀 유전자를 표적화하기 위해 CRISPR/Cas9-기반 유전자 편집 시스템 및 다수의 gRNA를 전달하기 위한 유전체 구축물, 조성물 및 방법이 본원에 기술되어 있다. 본원에 개시된 주제는 또한 골격근 및 심장 근육에 유전자 구축물(예를 들어, 벡터) 또는 그의 조성물을 전달하는 방법을 제공한다. 벡터는 변경된 AAV 벡터를 포함하는 AAV일 수 있다. 본원에 개시된 주제는 효과적이고, 효율적이며, 성공적인 게놈 변형을 용이하게 하는 골격근 또는 심장 근육에 상기 부류의 치료제의 활성 형태를 전달하는 방법을 개시할 뿐만 아니라, 치료 용도를 위한 인간 게놈을 재기록하는 수단 및 기초 과학 응용을 위한 모델 종을 표적화하는 수단을 제공한다.

[0034] 본 섹션에서 사용된 섹션 제목과 본원에 있는 모든 내용은 단지 유기적 목적을 위한 것이며, 제한하려는 것은 아니다.

[0035] 1. 정의

[0036] 달리 정의되지 않는 한, 본원에서 이용되는 모든 기술적 및 과학적 용어는 당업자에 의해 일반적으로 이해되는 바와 동일한 의미를 갖는다. 상충 시, 정의를 포함하여 본 문헌이 우선이 될 것이다. 바람직한 방법 및 물질이 아래에 기술되지만, 본원에 기술된 것과 유사하거나 동등한 방법 및 물질이 본 발명의 실시 또는 평가에서 이용될 수 있다. 본원에서 언급되는 모든 공보, 특허 출원, 특허 및 다른 참고문헌은 이의 전문이 참조로 포함된다. 본원에 개시된 물질, 방법, 및 실시예는 단지 예시적인 것이며 제한하려는 것이 아니다.

[0037] 본원에서 이용되는 "포함한다", "포함된다", "갖는", "갖는다", "할 수 있다", "함유한다"는 용어 및 이의 변형은 추가적인 행위 또는 구조의 가능성을 배제하지 않는, 개방형 말단의 접속 어구, 용어, 또는 단어인 것으로 의도된다. 단수 형태에는 문맥 상 명확히 달리 나타내지 않는 한, 복수의 참조물이 포함된다. 본 개시는 또한 명시적으로 나타내진 아니건, 본원에 나타난 구현에 또는 요소를 "포함하는", "이로 구성되는" 및 "이로 본질적으로 구성되는" 다른 구현예를 고려한다.

[0038] 본원에서 수치 범위의 열거에 있어서, 동일한 정도의 정확도로 이들 사이의 각각의 개재 수치가 명시적으로 고려된다. 예를 들어, 6~9의 범위에 있어서, 6 및 9에 부가하여 수치 7 및 8이 고려되며, 6.0~7.0의 범위에 있어서, 수치 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 및 7.0이 명시적으로 고려된다.

[0039] 본원에 사용된 바와 같이, 용어 "약" 또는 "대략"은 당 기술 분야의 당업자에 의해 결정된 바와 같은 특정 값에 대한 허용가능한 오차 범위 내에 있음을 의미하며, 이는 값이 측정되거나 결정되는 방식, 즉 측정 시스템의 제

한에 부분적으로 의존할 것이다. 예를 들어, "약"은 당 분야의 관행에 따라 3 이내의 또는 3을 초과하는 표준 편차를 의미할 수 있다. 대안적으로, "약"은 주어진 값의 20% 이하, 바람직하게는 10% 이하, 보다 바람직하게는 5% 이하, 및 더욱 바람직하게는 1% 이하의 범위를 의미할 수 있다. 대안적으로, 특히 생물학적 시스템 또는 공정과 관련하여, 이 용어는 값의 규모 순 이내, 바람직하게는 5-배 이내, 보다 바람직하게는 2-배 이내를 의미할 수 있다.

[0040] 본원에서 상호 교환적으로 이용되는 "아데노-연관 바이러스" 또는 "AAV"는 인간 및 일부 다른 영장류 종을 감염시키는 파보비리데(Parvoviridae)과의 데펜도바이러스(Dependovirus)속에 속하는 소형 바이러스를 나타낸다. AAV는 현재 질환을 유도하는 것으로 알려져 있지 않으며, 결과적으로 바이러스는 매우 온화한 면역 반응을 유도한다.

[0041] 본원에서 이용되는 "결합 영역"은 뉴클레아제에 의해 인식되고 결합되는 뉴클레아제 표적 영역 내 영역을 나타낸다.

[0042] 본원에서 상호 교환적으로 사용되는 "심근" 또는 "심장 근육"은 심장, 심근층의 벽 및 조직학적 기초에서 발견되는 비자발적인 줄무늬 근육의 유형을 의미한다. 심장 근육은 심장근육 세포 또는 심근 세포로 구성된다. 심근 세포는 골격근 세포와 유사한 줄무늬를 보여 주지만, 다핵 골격 세포와 달리 단 하나의 유일한 핵을 함유한다. 특정 구현예에서, "심장 근육 질환"은 심장 근육과 관련된 질환, 예컨대 심근병증, 심부전, 부정맥 및 염증성 심장 질병과 같은 질환을 나타낸다.

[0043] 본원에서 이용되는 "코딩 서열" 또는 "인코딩 핵산"은 단백질을 인코딩하는 뉴클레오타이드 서열을 포함하는 핵산(RNA 또는 DNA 분자)을 의미한다. 코딩 서열에는 핵산이 투여되는 개체 또는 포유류의 세포에서 발현을 유도할 수 있는 프로모터 및 폴리아데닐화 신호를 포함하는 조절 요소에 작동 가능하게 연결된 개시 및 종결 신호가 추가로 포함될 수 있다. 코딩 서열은 코돈 최적화될 수 있다.

[0044] 본원에서 이용되는 "상보체" 또는 "상보적인"은 핵산이 핵산 분자의 뉴클레오타이드 또는 뉴클레오타이드 유사체 간 왓슨-크릭(Watson-Crick, 예컨대, A-T/U 및 C-G) 또는 호그스틴(Hoogsteen) 염기쌍 형성을 의미할 수 있음을 의미한다. "상보성"은 두 핵산 서열이 서로 역평형으로 정렬되는 경우, 각각의 위치에서의 뉴클레오타이드 염기가 상보적이도록 하는 이들 간에 공유되는 특성을 나타낸다.

[0045] 본원에서 이용되는 "교정", "게놈 편집" 및 "복원"은 전장 기능적 또는 부분적인 전장 기능적 단백질 발현이 수득되도록 하는, 절단된 단백질을 인코딩하거나 단백질을 전혀 인코딩하지 않는 돌연변이체 유전자의 변화를 나타낸다. 돌연변이체 유전자의 교정 또는 복원에는 보수 기전, 예컨대 상동성-유도 보수(HDR)를 이용한 돌연변이를 갖는 유전자 영역의 대체 또는 전체 돌연변이체 유전자의 돌연변이를 갖지 않는 유전자 카피로의 대체가 포함될 수 있다. 돌연변이체 유전자의 교정 또는 복원에는 또한 유전자에 이중 가닥 절단을 생성한 후 비-상동성 말단 연결(NHEJ)을 이용해서 보수하는, 조기 중지 코돈, 비정상 스플라이스 수신체 부위 또는 비정상 스플라이스 공여체 부위를 유도하는 틀 이동 돌연변이의 보수가 포함될 수 있다. NHEJ는 적절한 해독 틀을 복원하고 조기 중지 코돈을 제거할 수 있는 보수 동안 적어도 하나의 염기쌍을 부가하거나 결실시킬 수 있다. 돌연변이체 유전자의 교정 또는 복원에는 또한 비정상 스플라이스 수신체 부위 또는 스플라이스 공여체 서열의 손상이 포함될 수 있다. 돌연변이체 유전자의 교정 또는 복원에는 또한 두 뉴클레아제 표적 부위 간 DNA의 제거 및 NHEJ에 의한 DNA 절단의 보수에 의해 적절한 해독 틀을 복원하기 위한 동일한 DNA 가닥 상의 두 뉴클레아제의 동시 작용에 의한 비-필수 유전자 절편의 결실이 포함될 수 있다.

[0046] 본원에서 상호 교환적으로 이용되는 "공여체 DNA", "공여체 주형" 및 "보수 주형"은 관심 유전자의 적어도 일부가 포함되는 이중-가닥 DNA 단편 또는 분자를 나타낸다. 공여체 DNA는 전체-기능적 단백질 또는 부분-기능적 단백질을 인코딩할 수 있다.

[0047] 본원에서 상호 교환적으로 이용되는 "뒤시엔느 근디스트로피" 또는 "DMD"는 근육 퇴행 그리고 궁극적으로 사망을 일으키는 열성, 치사, X-연관 장애를 나타낸다. DMD는 일반적인 선천성 단일유전자 질환으로, 남성 3500명 중 1명꼴로 일어난다. DMD는 디스트로핀 유전자에서 논센스 또는 틀 이동 돌연변이를 유도하는 유전된 또는 자연적 돌연변이의 결과이다. DMD를 유도하는 대부분의 디스트로핀 돌연변이는 디스트로핀 유전자에서 해독 틀을 손상시키고 조기 번역 종결을 유도하는 엑손 결실이다. DMD 환자는 전형적으로 아동기 동안 스스로 신체적으로 지지하는 능력을 잃고, 십대 시절 동안 점차 약해져서, 20대에 사망한다.

[0048] 본원에서 이용되는 "디스트로핀"은 세포막을 통해 주위 세포의 매트릭스에 근육 섬유의 세포골격을 연결하는 단백질 복합체의 일부인 막대형 세포질 단백질을 나타낸다. 디스트로핀은 근육 세포 온전성 및 기능의 조절에 관

여하는 세포막의 디스트로글리칸 복합체에 구조적 안정성을 제공한다. 본원에서 상호 교환적으로 이용되는 디스트로핀 유전자 또는 "DMD 유전자"는 유전자위 Xp21에 있는 2.2메가베이스이다. 일차 전사는 약 2,400 kb로 측정되며 성숙 mRNA는 약 14 kb이다. 79개 엑손이 3500개 아미노산을 초과하는 단백질을 코딩한다.

[0049] 본원에서 이용되는 "엑손 51"은 디스트로핀 유전자의 51번째 엑손을 나타낸다. 엑손 51은 DMD 환자에서의 프레임-손상 결실에 자주 인접하며, 올리고뉴클레오타이드-기반 엑손 건너뛰를 위한 임상 시험에서 표적화되었다. 엑손 51 건너뛰 화합물 에테플리르센에 대한 임상 시험은 최근에 기준선 대비 평균 47% 디스트로핀 양성 섬유를 보이며, 48주에 걸쳐 유의미한 기능적 이익을 보고하였다. 엑손 51에서의 돌연변이는 NHEJ-기반 게놈 편집에 의한 영구 교정에 이상적으로 적합하다.

[0050] 본원에서 상호 교환적으로 이용되는 "틀 이동" 또는 "틀 이동 돌연변이"는 하나 이상의 뉴클레오타이드의 부가 또는 결실이 mRNA에서 코돈 해독 틀의 이동을 유도하는 유전자 돌연변이의 한 유형을 나타낸다. 해독 틀에서의 이동은 단백질 번역에서 아미노산 서열 변형, 예컨대 미스센스 돌연변이 또는 조기 중지 코돈으로 이어질 수 있다.

[0051] 본원에서 이용되는 "기능적" 및 "전체-기능적"은 생물학적 활성을 갖는 단백질을 설명한다. "기능적 유전자"는 기능적 단백질로 번역되는 mRNA로 전사되는 유전자를 나타낸다.

[0052] 본원에서 이용되는 "융합 단백질"은 원래 별도 단백질을 코딩하는 2개 이상의 유전자의 연결을 통해 생성되는 키메라성 단백질을 나타낸다. 융합 유전자의 번역은 각각의 원래 단백질로부터 유래되는 기능적 특성을 갖는 단일 폴리펩타이드를 생성한다.

[0053] 본원에서 이용되는 "유전자 구축물"은 단백질을 인코딩하는 뉴클레오타이드 서열을 포함하는 DNA 또는 RNA 분자를 나타낸다. 코딩 서열에는 핵산 분자가 투여되는 대상체의 세포에서 발현을 유도할 수 있는 프로모터 및 폴리 아데닐화 신호를 포함하는 조절 요소에 작동 가능하게 연결된 개시 및 종결 신호가 포함된다. 본원에서 이용되는 용어 "발현 가능한 형태"는 대상체의 세포에 존재하는 경우, 코딩 서열이 발현되도록 하는 단백질을 인코딩하는 코딩 서열에 작동 가능하게 연결된 필요한 조절 요소를 함유하는 유전자 구축물을 나타낸다.

[0054] 본원에서 이용되는 "유전 질환"은 부분적으로 또는 전체적으로, 직접적으로 또는 간접적으로, 게놈에서 하나 이상의 비정상에 의해 유도되는 질환, 특히 출생시부터 존재하는 질병을 나타낸다. 비정상성은 돌연변이, 삽입 또는 결실일 수 있다. 비정상성은 유전자 또는 그 조절 서열의 코딩 서열에 영향을 미칠 수 있다. 유전 질환은 비제한적으로 DMD, 베커형 근이영양증(becker muscular dystrophy, BMD), 혈우병, 낭성 섬유증, 헌팅턴 무도병, 선천성 고콜레스테롤혈증(LDL 수용체 결함), 간모세포종, 윌슨병, 선천성 간 포르피린증, 유전된 간 대사 장애, 레슈-니한(Lesch Nyhan) 증후군, 낮 적혈구 빈혈, 지중해빈혈, 색소성 피부건조증, 판코니(Fanconi) 빈혈, 망막색소 변성, 모세관확장질환, 블루움(Bloom) 증후군, 망막모세포종, 및 테이-삭스(Tay-Sachs) 질환일 수 있다.

[0055] 본원에서 상호 교환적으로 이용되는 "상동성-유도 보수" 또는 "HDR"은 DNA의 상동성 절편이 핵에, 주로 세포 주기의 G2 및 S상에 존재하는 경우, 이중 가닥 DNA 병소를 보수하기 위한 세포 내 기전을 나타낸다. HDR은 보수를 가이드하기 위해 공여체 DNA 주형을 이용하며, 전체 유전자의 표적화된 부가를 포함하는 게놈에 대한 특정 서열 변화를 생성하기 위해 이용될 수 있다. 공여체 주형이 CRISPR/Cas9-기반 유전자 편집 시스템과 함께 제공되는 경우, 세포 기구는 상동성 재조합에 의해 절단을 보수할 것이며, 이는 DNA 절단의 존재 하에 수 백 배 이상으로 증강된다. 상동성 DNA 절편이 부재하는 경우, 비상동성 말단 연결이 대신 일어날 수 있다.

[0056] 본원에서 이용되는 "게놈 편집"은 유전자 변화를 나타낸다. 게놈 편집에는 돌연변이체 유전자의 교정 또는 복원이 포함될 수 있다. 게놈 편집에는 유전자, 예컨대 돌연변이체 유전자 또는 정상 유전자의 녹아웃이 포함될 수 있다. 게놈 편집은 관심 유전자를 변화시킴으로써 질환을 치료하거나 근육 보수를 증강시키기 위해 이용될 수 있다.

[0057] 2개 이상의 핵산 또는 폴리펩타이드 서열의 맥락에서 본원에서 이용되는 "동일한" 또는 "동일성"은 서열이 특정된 영역에 걸쳐 동일한 특정된 백분율의 잔기를 가짐을 의미한다. 백분율은 두 서열을 최적으로 정렬하고, 특정된 영역에 걸쳐 두 서열을 비교하고, 동일한 잔기가 두 서열 모두에서 일어나는 위치의 수를 결정해서 매치되는 위치의 수를 산출하고, 매치되는 위치의 수를 특정된 영역에서의 총 위치수로 나누고, 결과에 100을 곱해서 서열 동일성 백분율을 산출함으로써 계산될 수 있다. 두 서열이 상이한 길이어거나 정렬이 하나 이상의 엇갈린 말단을 생성하고 특정된 비교 영역에 단일 서열만 포함되는 경우, 단일 서열의 잔기는 계산의 분모에는 포함되지 않지만, 분자에는 포함되지 않는다. DNA 및 RNA를 비교하는 경우, 티민(T) 및 우라실(U)은 동등하게 간주될 수

있다. 동일성은 수동으로 또는 컴퓨터 서열 알고리즘, 예컨대 BLAST 또는 BLAST 2.0을 이용해서 수행될 수 있다.

[0058] 본원에서 상호 교환적으로 이용되는 "돌연변이체 유전자" 또는 "돌연변이된 유전자"는 검출 가능한 돌연변이를 거친 유전자를 나타낸다. 돌연변이체 유전자는 유전 물질의 변화, 예컨대 손실, 획득, 또는 교환을 거쳤으며, 이는 유전자의 정상적 전달 및 발현에 영향을 미친다. 본원에서 이용되는 "손상된 유전자"는 조기 중지 코돈을 유도하는 돌연변이를 갖는 돌연변이체 유전자를 나타낸다. 손상된 유전자 산물은 손상되지 않은 전장 유전자 산물에 비해 절단된다.

[0059] 본원에서 이용되는 "비-상동성 말단 연결(NHEJ) 경로"는 상동성 주형에 대한 필요 없이 절단 말단을 직접 결합함으로써 DNA에서 이중-가닥 절단을 보수하는 경로를 나타낸다. NHEJ에 의한 DNA 말단의 주형-독립적 재-결합은 DNA 절단 지점에 무작위 마이크로-삽입 및 마이크로-결실(indel)을 도입하는 확률적, 오차가 생기기 쉬운 보수 과정이다. 상기 방법은 표적화된 유전자 서열의 해독 틀을 의도적으로 손상시키거나, 결실시키거나, 변형하기 위해 이용될 수 있다. NHEJ는 전형적으로 보수를 가이드하기 위해 마이크로상동성으로 불리는 짧은 상동성 DNA 서열을 이용한다. 이들 마이크로상동성은 종종 이중-가닥 절단의 말단 상의 단일-가닥 오버행에 존재한다. 오버행이 완벽하게 상용성인 경우, NHEJ는 보통 절단을 정확하게 보수하지만, 뉴클레오타이드의 손실로 이어지는 부정확한 보수도 일어날 수 있고, 그러나 오버행이 상용성이 아닌 경우가 훨씬 더 일반적이다.

[0060] 본원에서 이용되는 "정상 유전자"는 유전 물질의 변화, 예컨대 손실, 획득, 또는 교환을 거치지 않은 유전자를 나타낸다. 정상 유전자는 정상 유전자 전달 및 유전자 발현을 거친다.

[0061] 본원에서 이용되는 "뉴클레아제 매개된 NHEJ"는 뉴클레아제, 예컨대 Cas9 분자가 이중 가닥 DNA를 절단한 후 개시되는 NHEJ를 나타낸다.

[0062] 본원에서 이용되는 "핵산" 또는 "올리고뉴클레오타이드" 또는 "폴리뉴클레오타이드"는 함께 공유 연결된 2개 이상의 뉴클레오타이드를 의미한다. 단일 가닥의 표시는 또한 상보적 가닥의 서열을 정의한다. 따라서, 핵산은 또한 표시된 단일 가닥의 상보적 가닥을 포괄한다. 핵산의 여러 변이체는 주어진 핵산과 동일한 목적을 위해 이용될 수 있다. 따라서, 핵산은 또한 실질적으로 동일한 핵산 및 이의 상보체를 포괄한다. 단일 가닥은 엄격한 혼성화 조건 하에 표적 서열에 혼성화할 수 있는 탐침을 제공한다. 따라서, 핵산은 또한 엄격한 혼성화 조건 하에 혼성화하는 탐침을 포괄한다.

[0063] 핵산은 단일 가닥이거나 이중 가닥일 수도 있고, 또는 이중 가닥 및 단일 가닥 서열 모두인 부분을 함유할 수도 있다. 핵산은 게놈 및 cDNA 둘 다인 DNA, RNA, 또는 핵산이 테옥시리보-및 리보-뉴클레오타이드의 조합을 함유할 수 있는 하이브리드, 및 우라실, 아데닌, 티민, 시토신, 구아닌, 이노신, 잔틴 하이포잔틴, 이소시토신, 및 이소구아닌을 포함하는 염기의 조합일 수 있다. 핵산은 화학적 합성 방법에 의해 또는 제조합 방법에 의해 수득될 수 있다.

[0064] 본원에서 이용되는 "작동 가능하게 연결된"은 유전자의 발현이 공간적으로 연결되는 프로모터의 제어 하에 있음을 의미한다. 프로모터는 그 제어 하 유전자의 5'(상류) 또는 3'(하류)에 배치될 수 있다. 프로모터 및 유전자 간 거리는 프로모터 및 프로모터가 유도되는 유전자에서 이것이 제어하는 유전자 간 거리와 대략 동일할 수 있다. 당분야에 공지된 바와 같이, 상기 거리에서의 변이는 프로모터 기능의 손실 없이 수용될 수 있다.

[0065] 본원에서 이용되는 "부분-기능적인"은 돌연변이체 유전자에 의해 인코딩되고 기능적 단백질보다 작은, 그러나 비-기능적 단백질보다는 큰 생물학적 활성을 갖는 단백질을 설명한다.

[0066] 본원에서 상호 교환적으로 이용되는 "조기 중지 코돈" 또는 "프레임-변경 중지 코돈"은 DNA 서열에서의 논센스 돌연변이를 나타내며, 이는 보통 야생형 유전자에서 확인되지 않는 위치에 중지 코돈을 생성한다. 조기 중지 코돈은 단백질의 전장 버전에 비해 절단되거나 더 짧은 단백질을 유도할 수 있다.

[0067] 본원에서 이용되는 "프로모터"는 세포에서 핵산의 발현을 부여하거나, 활성화하거나, 증강시킬 수 있는 합성 또는 천연-유래 분자를 의미한다. 프로모터는 발현을 추가 증강시키고/시키거나 그 공간적 발현 및/또는 시간적 발현을 변형하기 위해 하나 이상의 특이적인 전사 조절 서열을 포함할 수 있다. 프로모터는 또한 원위 인핸서 또는 억제인자 요소를 포함할 수 있고, 이는 전사 시작 부위부터 수 천 염기쌍만큼 멀리 배치될 수 있다. 프로모터는 바이러스, 박테리아, 진균, 식물, 곤충, 및 동물을 포함하는 원천으로부터 유래될 수 있다. 프로모터는 발현이 일어나는 세포, 조직 또는 기관에 대해, 또는 발현이 일어나는 발달 단계에 대해, 또는 생리적 스트레스, 병원체, 금속 이온, 또는 유도 제제와 같은 외부 자극에 반응하여 구성적으로 또는 차별적으로 유전자 성분의 발현을 조절할 수 있다. 프로모터의 대표예에는 박테리오파지 T7 프로모터, 박테리오파지 T3 프로모터

터, SP6 프로모터, lac 오퍼레이터-프로모터, tac 프로모터, SV40 후기 프로모터, SV40 조기 프로모터, RSV-LTR 프로모터, CMV IE 프로모터, SV40 조기 프로모터 또는 SV40 후기 프로모터, 인간 U6(hU6) 프로모터 및 CMV IE 프로모터가 포함된다.

[0068] 본원에서 이용되는 "골격근"은 가로무늬 근육의 한 유형을 나타내며, 이는 신체 신경계의 제어 하에 있고 힘줄로 알려져 있는 콜라겐 섬유 번들에 의해 뼈에 부착된다. 골격근은 근세포 또는 "근육 세포"로 알려진, 때때로 "근육 섬유"로 총칭되는 개별 성분으로 이루어진다. 근세포는 근육발생으로 알려진 과정에서 발달 근모세포(근육 세포를 생성하는 배아 전구체 세포의 한 유형)의 융합으로부터 형성된다. 이러한 긴, 실린더형, 다핵 세포는 근섬유로도 불린다.

[0069] 본원에서 이용되는 "골격근 질병"은 골격근에 관련된 질병, 예컨대 근디스트로피, 노화, 근육 퇴행, 상처 치유, 및 근육 쇠약 또는 위축을 나타낸다.

[0070] 본원에서 상호 교환적으로 이용되는 "대상체" 및 "환자"는 비제한적으로 포유류(예컨대, 소, 돼지, 낙타, 라마, 말, 염소, 토끼, 양, 햄스터, 기니아피크, 고양이, 개, 래트, 및 마우스, 비-인간 영장류(예를 들어, 원숭이, 예컨대 게잡이 또는 붉은털 원숭이, 침팬지 등) 및 인간)를 포함하는 임의의 척추동물을 나타낸다. 일부 구현예에서, 대상체는 인간 또는 비-인간일 수 있다. 대상체 또는 환자는 다른 형태의 치료를 받고 있을 수 있다.

[0071] 본원에 사용된 "표적 유전자"는 공지되거나 추정된 유전자 산물을 코딩하는 임의의 뉴클레오타이드 서열을 나타낸다. 표적 유전자는 유전병과 관련된 돌연변이 유전자일 수 있다. 특정 구현예에서, 표적 유전자는 인간 *디스트로핀* 유전자이다. 특정 구현예에서, 표적 유전자는 돌연변이 인간 *디스트로핀* 유전자이다.

[0072] 본 명세서에서 사용되는 "표적 부위"는 CRISPR/Cas9-기반 유전자 편집 시스템이 결합 및 절단하도록 고안된 표적 유전자의 영역을 나타낸다.

[0073] 본원에서 이용되는 "트랜스유전자"는 하나의 유기체로부터 단리되었고 상이한 유기체 내로 도입되는 유전자 또는 유전자 서열을 함유하는 유전 물질을 나타낸다. DNA의 상기 비-원상태 절편은 트랜스제닉 유기체에서 RNA 또는 단백질을 생성하는 능력을 보유할 수 있거나, 트랜스제닉 유기체의 유전 코드의 정상 기능을 변형할 수 있다. 트랜스유전자의 도입은 유기체의 표현형을 변화시킬 가능성을 갖는다.

[0074] 핵산에 대해 본원에서 이용되는 "변이체"는 (i) 참조되는 뉴클레오타이드 서열의 일부 또는 단편; (ii) 참조되는 뉴클레오타이드 서열 또는 이의 일부의 상보체; (iii) 참조되는 핵산 또는 이의 상보체와 실질적으로 동일한 핵산; 또는 (iv) 참조되는 핵산, 이의 상보체, 또는 이와 실질적으로 동일한 서열에 대해 엄격한 조건 하에 혼성화하는 핵산을 의미한다.

[0075] 펩타이드 또는 폴리펩타이드에 대한 "변이체"는 아미노산의 삽입, 결실 또는 보존적 치환에 의해 아미노산 서열이 상이하지만, 적어도 하나의 생물학적 활성을 보유한다. 변이체는 또한 적어도 하나의 생물학적 활성을 보유하는 아미노산 서열을 갖는 참조되는 단백질과 실질적으로 동일한 아미노산 서열을 갖는 단백질을 의미할 수 있다. 아미노산의 보존적 치환, 즉 유사한 특성(예컨대, 친수성, 하전된 영역의 정도 및 분포)을 갖는 상이한 아미노산을 이용한 아미노산의 대체는 당분야에서 전형적으로 소소한 변화가 관여되는 것으로 인식된다. 이러한 소소한 변화는 부분적으로 당분야에서 이해되는 바와 같은 아미노산의 수치 지수를 고려함으로써 확인될 수 있다[Kyte *et al.*, *J. Mol. Biol.* 157:105-132(1982)]. 아미노산의 수치 지수는 그 소수성 및 전하의 고려에 기반한다. 유사한 수치 지수의 아미노산이 치환되어 여전히 단백질 기능을 보유할 수 있음이 당분야에 공지되어 있다. 하나의 양태에서, ± 2 의 수치 지수를 갖는 아미노산이 치환된다. 아미노산의 친수성도 생물학적 기능을 보유하는 단백질을 생성할 치환을 드러내기 위해 이용될 수 있다. 펩타이드의 맥락에서 아미노산의 친수성의 고려는 해당 펩타이드의 최대 국소 평균 친수성의 계산을 허용한다. 치환은 서로 ± 2 이내의 친수성값을 갖는 아미노산으로 수행될 수 있다. 아미노산의 소수성 지수 및 친수성값은 모두 그 아미노산의 특정 측쇄에 의해 영향을 받는다. 그 관찰과 일치하게, 생물학적 기능과 상용성인 아미노산 치환은 소수성, 친수성, 전하, 크기, 및 다른 특성에 의해 드러나는 바와 같은, 아미노산의 상대적 유사성, 그리고 특히 이들 아미노산의 측쇄에 의존하는 것으로 이해된다.

[0076] 본원에서 이용되는 "벡터"는 복제 기원을 함유하는 핵산 서열을 의미한다. 벡터는 바이러스 벡터, 박테리오파지, 박테리아 인공 염색체 또는 효모 인공 염색체일 수 있다. 벡터는 DNA 또는 RNA 벡터일 수 있다. 벡터는 자가-복제성 염색체의 벡터일 수 있고, 바람직하게는 DNA 플라스미드이다. 예를 들어, 벡터는 Cas9 단백질 및 적어도 하나의 gRNA 분자, 예컨대 SEQ ID NO: 1~19, 41, 42 중 어느 하나 또는 그의 상보체의 표적화 도메인을 포함하는 gRNA를 인코딩할 수 있다. 일부 구현예에서, Cas9 단백질은 SEQ ID NO: 27, SEQ ID NO: 33,

또는 SEQ ID NO: 45의 아미노산 서열을 가질 수 있다. 일부 구현예에서, Cas9 단백질은 *एस. 아우레우스* Cas9, 예컨대 SEQ ID NO: 33, 또는 SEQ ID NO: 45의 아미노산 서열을 갖는 SaCas9일 수 있다. 일부 구현예에서, Cas9 단백질은 SEQ ID NO: 26, SEQ ID NO: 28, SEQ ID NO: 29, SEQ ID NO: 30, SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 43, 또는 SEQ ID NO: 44의 핵산 서열을 포함하는 핵산 서열에 의해 코딩된다.

[0077] 본원에서 달리 정의되지 않는 한, 본 개시와 함께 이용되는 과학적 및 기술적 용어는 당업자에 의해 일반적으로 이해되는 의미를 가질 것이다. 예를 들어, 본원에 기술되는 세포 및 조직 배양, 분자 생물학, 면역학, 미생물학, 유전학 그리고 단백질 및 핵산 화학 및 혼성화에 관해 이용되는 임의의 명명법 및 이의 기술은 당분야에 널리 공지되어 있고 일반적으로 이용되는 것이다. 용어의 의미 및 범위는 명확할 것이다; 그러나 임의의 잠재적인 모호성이 있는 경우, 본원에서 제공되는 정의가 임의의 사전적 또는 외적 정의에 우선한다. 또한, 문맥 상 달리 요구되지 않는 한, 단수 용어에는 복수가 포함되며, 복수 용어에는 단수가 포함되어야 한다.

[0078] 2. 디스트로핀 유전자의 게놈 편집을 위한 유전자 구축물

[0079] 본 발명은 디스트로핀 유전자(예를 들어, 인간 디스트로핀 유전자)의 게놈 편집, 게놈 변형 또는 유전자 발현 변경을 위한 유전자 구축물에 관한 것이다. 유전자 구축물은 SaCas9-양립성 표적과 같은 인간과 붉은털 원숭이 디스트로핀 유전자 서열을 표적화하는 적어도 하나의 gRNA를 포함한다. 개시된 gRNA는 DMD 환자의 세포내 기능성 디스트로핀의 발현을 복원하기 위해 상기 영역의 게놈 결실을 일으키는, 인간 디스트로핀 유전자의 엑손 51을 둘러싼 인트론 영역을 표적화하기 위해 SaCas9를 사용하는 시스템을 포함하는 CRISPR/Cas9-기반 유전자 편집 시스템에 포함될 수 있다.

[0080] a. 디스트로핀 유전자

[0081] 디스트로핀은 세포막을 통해 주위 세포의 매트릭스에 근육 섬유의 세포골격을 연결하는 단백질 복합체의 일부인 막대형 세포질 단백질이다. 디스트로핀은 세포막의 디스트로글리칸 복합체에 구조적 안정성을 제공한다. 디스트로핀 유전자는 유전자위 Xp21에 있는 2.2메가베이스이다. 일차 전사는 약 2,400 kb로 추정되며 성숙 mRNA는 약 14 kb이다. 79개 엑손이 3500개 아미노산을 초과하는 단백질을 코딩한다. 정상 골격 근육 조직에는 소량의 디스트로핀만 함유되어 있지만 비정상적인 발현이 없으면 중증의 불치의 증상이 나타난다. 디스트로핀 유전자의 일부 돌연변이는 감염된 환자에서 디스트로핀의 결함 및 중증의 이영양증 표현형의 생산을 유발한다. 디스트로핀 유전자의 일부 돌연변이는 부분-기능성 디스트로핀 단백질과 영향을 받는 환자에서 훨씬 가벼운 이영양증 표현형을 유도한다.

[0082] DMD는 유전성 또는 자발적인 돌연변이의 결과로 디스트로핀 유전자의 논센스 또는 틀 이동 돌연변이를 유발한다. DMD에 대하여 자연적으로 발생하는 돌연변이와 그의 결과물은 비교적 잘 알려져 있다. 막대 도메인 내에 함유된 엑손 45-55 영역(예를 들어, 엑손 51)에서 발생하는 인-프레임 결실은 매우 기능성인 디스트로핀 단백질을 생성할 수 있고, 많은 담체가 무증상이거나 가벼운 증상을 나타낸다는 것이 알려져 있다. 또한, 이론적으로 환자들 중 60% 이상이 디스트로핀 유전자의 이 영역에서 엑손을 표적화함으로써(예를 들어, 엑손 51을 표적화함으로써) 치료될 수 있다. DMD 환자에서 mRNA 스플라이싱 동안 비-필수 엑손을 스킵함으로써(예를 들어, 엑손 51을 스킵함으로써) 손상된 디스트로핀 해독 틀을 복원하여 내부적으로 결실되었지만 기능성 디스트로핀 단백질을 생성하려는 노력이 이루어졌다. 내부 디스트로핀 엑손(들)의 결실(예를 들어, 엑손 51의 결실)은 적절한 해독 틀을 유지하지만, 덜 심각한 베커형 근이영양증 또는 BMD를 야기한다. 베커형 근이영양증 또는 BMD 유전자형은 디스트로핀 유전자에 결실이 존재한다는 점에서 DMD와 유사하다. 그러나, 이러한 결실은 해독 틀을 온전하게 유지한다. 따라서, 내부가 절단되지만, 부분적으로 기능적인 디스트로핀 단백질이 생성된다. BMD는 다양한 표현형을 가지고 있지만, 종종 디스트로핀의 엑손 45-55 사이에 결실이 있으면 DMD에 비해 표현형이 훨씬 더 약하다. 따라서, DMD 유전자형을 BMD 유전자형으로 변경하는 것은 디스트로핀을 교정하는 일반적인 전략이다. 디스트로핀을 교정하기 위한 많은 전략이 있으며, 그 중 많은 것은 내인성 디스트로핀의 해독 틀 복원에 의존한다. 이것은 질병 유전자형을 DMD에서 베커형 근이영양증으로 이동시킨다. 많은 BMD 환자는 번역가능한 해독 틀을 유지하는 유전자내 결실을 가지고있어, 짧지만 주로 기능성인 디스트로핀 단백질로 이어진다.

[0083] 특정 구현예에서, 해독 틀을 복원시키는 엑손 51의 변형(예를 들어, NHEJ에 의한 엑손 51의 결실 또는 절제)은 결실 돌연변이를 갖는 DMD 대상체를 포함하는 표현형 DMD 대상체를 개선한다. 특정 구현예에서, 디스트로핀 유전자의 엑손 51은 디스트로핀 유전자의 51번째 엑손을 나타낸다. 엑손 51은 DMD 환자에서의 프레임-손상 결실에 자주 인접하며 올리고뉴클레오타이드-기반 엑손 건너뛰를 위한 임상 시험에서 표적화되었다. 엑손 51 건너뛰 화합물 에테플리르센(etepilrisen)에 대한 임상 시험은 최근에 기준선 대비 평균 47% 디스트로핀 양성 섬유를 보이

며, 48주에 걸쳐 유의미한 기능적 이익을 보고하였다. 엑손 51에서의 돌연변이는 NHEJ-기반 게놈 편집에 의한 영구 교정에 이상적으로 적합하다.

[0084] 본원에 개시된 벡터는 *디스트로핀* 유전자, 예를 들어 인간 *디스트로핀* 유전자에서 결실을 생성할 수 있다. 특정 구현예에서, 벡터는 *디스트로핀* 유전자의 표적 위치에 인접한 2개의 인트론(제1 인트론 및 제2 인트론)에서 2개의 이중 가닥 절단(제1 이중 가닥 절단 및 제2 이중 가닥 절단)을 형성하여, *디스트로핀* 표적 위치를 포함하는 *디스트로핀* 유전자의 절편을 결실시키도록 구성된다. "디스트로핀 표적 위치"는 본원에 기술된 바와 같이, 디스트로핀 엑손 표적 위치 또는 디스트로핀 엑손-내 표적 위치일 수 있다. *디스트로핀* 엑손 표적 위치의 결실은 뒤시엔느 근이영양증으로 고통받는 대상체의 *디스트로핀* 서열을 최적화할 수 있는데, 예를 들어, 이는 인코딩된 디스트로핀 단백질의 기능 또는 활성을 증가시킬 수 있거나 대상체의 질병 상태를 개선시킬 수 있다. 특정 구현예에서, 상기 *디스트로핀* 엑손 표적 위치의 절제는 해독 틀을 복원시킨다. 상기 *디스트로핀* 엑손 표적 위치는 상기 *디스트로핀* 유전자의 하나 이상의 엑손을 포함할 수 있다. 특정 구현예에서, 디스트로핀 표적 위치는 *디스트로핀* 유전자(예를 들어, 인간 *디스트로핀* 유전자)의 엑손 51을 포함한다.

[0085] 본원에 개시된 유전자 구축물(예를 들어, 벡터)은 *디스트로핀* 유전자(예를 들어, 인간 *디스트로핀* 유전자)의 엑손 51에서 매우 효율적인 유전자 편집을 매개할 수 있다. 본원에 개시된 유전자 구축물(예를 들어, 벡터)은 DMD 환자의 세포에서 디스트로핀 단백질 발현을 회복시킨다.

[0086] 엑손 51은 종종 DMD에서 프레임-손상 결실에 인접한다. 엑손 건너뛰에 의한 *디스트로핀* 전사체에서 엑손 51의 제거는 모든 DMD 환자의 약 15%를 치료하는데 사용될 수 있다. 이 종류의 *디스트로핀* 돌연변이는 NHEJ-기반 게놈 편집 및 HDR에 의한 영구적인 교정에 이상적으로 적합하다. 본원에 기술된 유전자 구축물(예를 들어, 벡터)은 인간 *디스트로핀* 유전자에서 엑손 51의 표적화된 변형을 위해 개발되었다. 본원에 개시된 유전자 구축물(예를 들어, 벡터)은 인간 DMD 세포 내로 형질감염되어 효율적인 유전자 변형 및 정확한 해독 틀로의 전환을 매개한다. 단백질 복원은 프레임 복원과 병행하고, CRISPR/Cas9-기반 유전자 편집 시스템으로 처리된 세포의 대량 집단에서 검출된다.

[0087] b. CRISPR 시스템

[0088] 본원에 개시된 유전자 구축물(예를 들어, 벡터)은 *디스트로핀* 유전자(예를 들어, 인간 *디스트로핀* 유전자)에 특이적인 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩한다. 본원에서 상호교환적으로 이용되는 "클러스터링된, 규칙적으로 산재된, 짧은 회문 반복" 및 "CRISPR"은 대략 40%의 시퀀싱된 박테리아 및 90%의 시퀀싱된 원시세균(archaea)의 게놈에서 확인되는 여러 짧은 직접 반복을 함유하는 유전자위를 나타낸다. CRISPR 시스템은 획득 면역성의 한 형태를 제공하는 침입하는 파지 및 플라스미드에 대한 방어에 관여되는 미생물 뉴클레아제 시스템이다. 미생물 숙주에서의 CRISPR 유전자위는 CRISPR-연관(Cas) 유전자뿐만 아니라 CRISPR-매개 핵산 절단의 특이성을 프로그래밍할 수 있는 비-코딩 RNA 요소의 조합을 함유한다. 스페이서로 불리는 외래 DNA의 짧은 절편이 CRISPR 반복 사이의 게놈 내에 포함되고, 과거 노출의 '메모리'로 작용한다. Cas9는 sgRNA(본원에서 "gRNA"로 상호교환적으로 언급됨)의 3' 말단과 복합체를 형성하고, 단백질-RNA 페어는 sgRNA 서열의 5' 말단 및 프로토스페이서로 알려진 사전 정의된 20 bp DNA 서열 간 상보적 염기쌍 형성에 의해 그의 게놈 표적을 인식한다. 상기 복합체는 crRNA 내에 인코딩된 영역, 즉 프로토스페이서 및 병원체 게놈 내의 프로토스페이서-인접 모티프(PAM)를 통해 병원체 DNA의 상동성 유전자위로 보내진다. 비-코딩 CRISPR 어레이가 전사되고 직접 반복 내에서 개별 스페이서 서열을 함유하는 짧은 crRNA로 절단되며, 이는 Cas 뉴클레아제를 표적 부위(프로토스페이서)로 보낸다. 발현된 sgRNA의 20 bp 인식 서열을 단순 교환함으로써, Cas9 뉴클레아제는 새로운 게놈 표적으로 보내질 수 있다. CRISPR 스페이서는 진핵생물 유기체에서의 RNAi와 유사한 방식으로 외인성 유전 요소를 인식하고 침묵화하기 위해 이용된다.

[0089] 3 클래스의 CRISPR 시스템(I, II 및 III형 효과기 시스템)이 알려져 있다. II형 효과기 시스템은 단일 효과기 효소, Cas9를 이용해서 4개의 순차적 단계로 표적화된 DNA 이중-가닥 절단을 수행하여 dsDNA를 절단한다. 복합체로 작용하는 여러 별도의 효과기를 필요로 하는 I형 및 III형 효과기 시스템에 비해, II형 효과기 시스템은 진핵생물 세포와 같은 대안적 맥락에서 기능할 수 있다. II형 효과기 시스템은 스페이서-함유 CRISPR 유전자위로부터 전사되는 긴 프리-crRNA, Cas9 단백질, 및 프리-crRNA 가공에 관여되는 tracrRNA로 구성된다. tracrRNA는 프리-crRNA의 스페이서를 구분하는 반복 영역에 혼성화함으로써, 내인성 RNase III에 의해 dsRNA 절단을 개시한다. 상기 절단에는 Cas9에 의해 각각의 스페이서 내에서 두 번째 절단 이벤트가 뒤따라서, tracrRNA 및 Cas9와 연관된 채 유지되는 성숙 crRNA를 생성하여, Cas9:crRNA-tracrRNA 복합체를 형성한다.

[0090] Cas9:crRNA-tracrRNA 복합체는 DNA 듀플렉스를 풀고 crRNA와 매치되는 서열을 탐색하여 절단한다. 표적 DNA 내

의 "프로토스페이서" 서열 및 crRNA 내의 나머지 스페이서 서열 간 상보성 검출 시 표적 인식이 일어난다. Cas9는 정확한 프로토스페이서-인접 모티프(PAM)가 또한 프로토스페이서의 3' 말단에 존재하는 경우 표적 DNA의 절단을 매개한다. 프로토스페이서 표적화를 위해, 서열에는 DNA 절단을 위해 요구되는, Cas9 뉴클레아제에 의해 인식되는 짧은 서열인 프로토스페이서-인접 모티프(PAM)가 바로 뒤따라야 한다. 상이한 II형 시스템은 상이한 PAM 요건을 갖는다. *예스. 피오게네스(S. pyogenes)* CRISPR 시스템은 상기 Cas9(SpCas9)에 대한 PAM 서열을 5'-NRG-3'로 가질 수 있고, 여기서 R은 A 또는 G이며, 인간 세포에서 상기 시스템의 특이성을 특징으로 한다. CRISPR/Cas9-기반 유전자 편집 시스템의 고유한 능력은 단일 Cas9 단백질과 2개 이상의 sgRNA의 공동-발현에 의해 여러 별도의 게놈 유전자위를 동시에 표적화하는 간단한 능력이다. 예를 들어, *스트렙토코커스 피오게네스(Streptococcus pyogenes)* II형 시스템은 자연적으로 "NGG" 서열을 이용하는 것을 선호하며, 여기서 "N"은 임의의 뉴클레오타이드일 수 있지만, 다른 PAM 서열, 예컨대 조작된 시스템 내의 "NAG"도 허용한다(Hsu et al., *Nature Biotechnology* (2013) doi:10.1038/nbt.2647). 유사하게, *나이세리아 메닝지티디스(Neisseria meningitidis)*에서 유래되는 Cas9(NmCas9)는 보통 NNNNGATT의 원상태 PAM을 갖지만, 고충돌성 NNNNGNNN PAM을 포함하는 다양한 PAM에 걸쳐 활성을 갖는다(Esvelt et al. *Nature Methods*(2013) doi:10.1038/nmeth.2681).

[0091] *예스. 아우레우스*의 Cas9 분자는 서열 모티프 NNGRR(R = A 또는 G)(SEQ ID NO: 22)을 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다. 특정 구현예에서, *예스. 아우레우스*의 Cas9 분자는 서열 모티프 NNGRRN(R = A 또는 G)(SEQ ID NO: 23)을 인식하고, 표적 핵산 서열 1 내지 10, 예를 들어 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다. 특정 구현예에서, *예스. 아우레우스*의 Cas9 분자는 서열 모티프 NNGRR(R = A 또는 G)(SEQ ID NO: 24)를 인식하고 표적 핵산 서열 1 내지 10, 예를 들어 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다. 특정 구현예에서, *예스. 아우레우스*의 Cas9 분자는 서열 모티프 NNGRRV(R = A 또는 G)(SEQ ID NO: 25)를 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다. 상기 구현예에서, N은 임의의 뉴클레오타이드 잔기, 예를 들어, A, G, C 또는 T 중 임의의 것일 수 있다. Cas9 분자는 Cas9 분자의 PAM 특이성을 변화시키도록 조작될 수 있다.

[0092] (1) CRISPR/Cas9-기반 유전자 편집 시스템

[0093] *스트렙토코커스 피오게네스(Streptococcus pyogenes)*의 II형 효과기 시스템의 하나의 조작된 형태는 게놈 조작을 위해 인간 세포에서 기능하는 것으로 나타났다. 상기 시스템에서, Cas9 단백질은 합성적으로 재구성된 "가이드 RNA"("gRNA", 또한 본원에서 키메라성 단일 가이드 RNA("sgRNA")와 상호 교환적으로 이용됨)에 의해 게놈 표적 부위로 보내지며, 일반적으로 RNase III 및 crRNA 가공에 대한 필요성을 없애는 것은 crRNA-tracrRNA 융합이다. 게놈 편집 및 유전 질환의 치료에서 이용하기 위한 CRISPR/Cas9-기반 조작된 시스템이 본원에 제공된다. CRISPR/Cas9-기반 조작된 시스템은 유전 질환, 노화, 조직 재생, 또는 상처 치유에 관여되는 유전자를 포함하는 임의의 유전자를 표적화하도록 설계될 수 있다. CRISPR/Cas9-기반 유전자 편집 시스템에는 Cas9 단백질 또는 Cas9 융합 단백질 및 적어도 하나의 gRNA가 포함될 수 있다. 특정 구현예에서, 상기 시스템은 2개의 gRNA 분자를 포함한다. Cas9 융합 단백질에는, 예를 들어, Cas9에 내인성인 상이한 활성을 갖는 도메인, 예컨대 트랜스활성화 도메인이 포함될 수 있다.

[0094] 표적 유전자(예를 들어, *디스트로핀* 유전자, 예를 들어 인간 *디스트로핀* 유전자)은 세포의 분화 또는 유전자의 활성화가 요구될 수 있는 임의의 다른 과정에 관여할 수 있거나, 또는 틀 이동 돌연변이 또는 논센스 돌연변이와 같은 돌연변이를 가질 수 있다. 표적 유전자가 조기 중지 코돈, 비정상 스플라이스 수신체 부위 또는 비정상 스플라이스 공여체 부위를 유도하는 돌연변이를 갖는 경우, CRISPR/Cas9-기반 유전자 편집 시스템은 조기 중지 코돈, 비정상 스플라이스 수신체 부위 또는 비정상 스플라이스 공여체 부위로부터 상류 또는 하류의 뉴클레오타이드 서열을 인식하고 결합하도록 설계될 수 있다. CRISPR-Cas9-기반 시스템은 또한 조기 중지 코돈의 건너뛰음을 유도하거나 손상된 해독 틀을 복원하기 위해 스플라이스 수신체 및 공여체를 표적화함으로써 정상적인 유전자 스플라이싱을 손상시키기 위해 이용될 수 있다. CRISPR/Cas9-기반 유전자 편집 시스템은 게놈의 단백질-코딩 영역에 대한 표적-이외 변화를 매개할 수도 매개하지 않을 수도 있다.

[0095] (a) Cas9 분자 및 Cas9 융합 단백질

[0096] CRISPR/Cas9-기반 유전자 편집 시스템에는 Cas9 단백질 또는 Cas9 융합 단백질이 포함될 수 있다. Cas9 단백질은 핵산을 절단하는 엔도뉴클레아제이며, CRISPR 유전자위에 의해 인코딩되고 II형 CRISPR 시스템에 관여된다. Cas9 단백질은 임의의 박테리아 또는 원시세균 중, 예컨대 *스트렙토코커스 피오게네스(Streptococcus pyogenes)*, *스타필로코커스 아우레우스(Staphylococcus aureus, S. aureus)*, *아시도보락스 아베나에(Acidovorax avenae)*, *악티노바실러스 플레오루모니아에(Actinobacillus pleuropneumoniae)*, *악티노바실러스 속*

시노케네스(*Actinobacillus succinogenes*), 악티노바실러스 수이스(*Actinobacillus suis*), 악티노마이세스 종(*Actinomyces* sp.), 사이클리필러스 데니트리피칸스(*cycliphilus denitrificans*), 아미노모나스 파우시보란스(*Aminomonas paucivorans*), 바실러스 세레우스(*Bacillus cereus*), 바실러스 스미티(*Bacillus smithii*), 바실러스 트루링기엔시스(*Bacillus thuringiensis*), 박테로이데스 종(*Bacteroides* sp.), 블라스토피렐라 마리나(*Blastopirellula marina*), 브라디리조븀 종(*Bradyrhizobium* sp.), 브레비바실러스 라테로스포르스(*Brevibacillus laterosporus*), 캄필로박터 콜라이(*Campylobacter coli*), 캄필로박터 제주니(*Campylobacter jejuni*), 캄필로박터 라리(*Campylobacter lari*), 칸디다투스 푸니세이스피릴룸(*Candidatus Puniceispirillum*), 클로스트리디움 셀룰로리티쿰(*Clostridium cellulolyticum*), 클로스트리디움 페르프링겐스(*Clostridium perfringens*), 코리네박테리움 악콜렌스(*Corynebacterium accolens*), 코리네박테리움 디프테리아(*Corynebacterium diphtheria*), 코리네박테리움 마트루코티(*Corynebacterium matruchotii*), 디노로세오박터 쉬배(*Dinoroseobacter shibae*), 유박테리움 돌리쑈(*Eubacterium dolichum*), 감마 프로테오박테리움(*gamma proteobacterium*), 글루코나세토박터 디아조트로피쿠스(*Gluconacetobacter diazotrophicus*), 헤모필루스 파라인플루엔자(*Haemophilus parainfluenzae*), 헤모필루스 스푸토룸(*Haemophilus sputorum*), 헬리코박터 카나덴시스(*Helicobacter canadensis*), 헬리코박터 시나에디(*Helicobacter cinaedi*), 헬리코박터 무스텔라(*Helicobacter mustelae*), 일리오박터 폴리트로푸스(*Ilyobacter polytropus*), 킨겔라 킹가에(*Kingella kingae*), 락토바실러스 크리스파투스(*Lactobacillus crispatus*), 리스테리아 이바노비(*Listeria ivanovii*), 리스테리아 모노사이토게네스(*Listeria monocytogenes*), 리스테리아세 박테리움(*Listeriaceae bacterium*), 메틸로시스티스 종(*Methylocystis* sp.), 메틸로시너스 트리코스포리움(*Methylosinus trichosporium*), 모빌룬커스 물리에리스(*Mobiluncus mulieris*), 네이세리아 바실리포르미스(*Neisseria bacilliformis*), 네이세리아 시네레아(*Neisseria cinerea*), 네이세리아 플라베센스(*Neisseria flavescens*), 네이세리아 락타미카(*Neisseria lactamica*), 네이세리아 종(*Neisseria* sp.), 네이세리아 와드워티(*Neisseria wadsworthii*), 니트로소모나스 종(*Nitrosomonas* sp.), 파르비바쿨룸 라바멘티보란스(*Parvibaculum lavamentivorans*), 파스테우렐라 물토시다(*Pasteurella multocida*), 파스콜락토박테리움 숙시나투텐스(*Phascolarctobacterium succinatutens*), 랄스토니아 시지기(*Ralstonia syzygii*), 로도슈모나스 팔루스트리스(*Rhodospseudomonas palustris*), 로도볼룸 종(*Rhodovulum* sp.), 시몬시엘라 무엘레리(*Simonsiella muelleri*), 스펡고모나스 종(*Sphingomonas* sp.), 스포로락토바실러스 비니아(*Sporolactobacillus vineae*), 스타필로코커스 루그두넨시스(*Staphylococcus lugdunensis*), 스트렙토코커스 종(*Streptococcus* sp.), 서브돌리그라눌룸 종(*Subdoligranulum* sp.), 티스트렐라 모빌리스(*Tistrella mobilis*), 트레포네마 종(*Treponema* sp.), 또는 베르미네프로박터 에이세니아(*Verminephrobacter eiseniae*)에서 유래될 수 있지만, 이에 한정되지 않는다. 특정 구현예에서, Cas9 분자는 이다. 상기 Cas9 단백질은 스트렙토코커스 피오게네스(*Streptococcus pyogenes*) Cas9 분자(본원에서 "SpCas9"로도 언급됨)이다. 특정 구현예에서, Cas9 분자는 스타필로코커스 아우레우스(*Staphylococcus aureus*) Cas9 분자(본원에서 "SaCas9"로도 언급됨)이다.

[0097] Cas9 분자 또는 Cas9 융합 단백질은 하나 이상의 gRNA 분자와 상호작용할 수 있고, gRNA 분자(들)와 협력하여 표적 도메인을 포함하는 부위 및, 특정 구현예에서 PAM 서열에 국한된다. Cas9 분자 또는 Cas9 융합 단백질이 PAM 서열을 인식하는 능력은 예를 들어 상기 기술된 바와 같은 형질전환 검정법(Jinek 2012)을 사용하여 측정될 수 있다.

[0098] 특정 구현예에서, Cas9 분자 또는 Cas9 융합 단백질이 표적 핵산과 상호작용하고 표적 핵산을 절단하는 능력은 PAM 서열 의존적이다. PAM 서열은 표적 핵산의 서열이다. 특정 구현예에서, 표적 핵산의 절단은 PAM 서열의 상류에서 일어난다. 상이한 세균 종으로부터의 Cas9 분자는 상이한 서열 모티프(예를 들어, PAM 서열)를 인식할 수 있다. 특정 구현예에서, 에스. 피오게네스(*S. pyogenes*)의 Cas9 분자는 서열 모티프 NGG를 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다(예를 들어, Mali, 2013 참조). 특정 구현예에서, 에스. 썬모필루스(*S. thermophilus*)의 Cas9 분자는 서열 모티프 NGGNG(SEQ ID NO: 36) 및/또는 NNAGAAW(W = A 또는 T)(SEQ ID NO: 20)를 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다(예를 들어, Horvath 2010; Deveau 2008 참조). 특정 구현예에서, 에스. 뮈탄스(*S. mutans*)의 Cas9 분자는 서열 모티프 NGG 및/또는 NAAR(R = A 또는 G)(SEQ ID NO: 21)을 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다(예를 들어, Deveau 2008 참조). 특정 구현예에서, 에스. 아우레우스(*S. aureus*)의 Cas9 분자는 서열 모티프 NGR(R = A 또는 G)(SEQ ID NO: 22)를 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다. 특정 구현예에서, 에스. 아우레우스(*S. aureus*)의 Cas9 분자는 서열 모티프 NGRN(R = A 또는 G)(SEQ ID NO: 23)을 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절

단을 유도한다. 특정 구현예에서, *에스. 아우레우스*(*S. aureus*)의 Cas9 분자는 서열 모티프 NNGRRT(R = A 또는 G)(SEQ ID NO: 24)를 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다. 특정 구현예에서, *에스. 아우레우스*(*S. aureus*)의 Cas9 분자는 서열 모티프 NNGRRV(R = A 또는 G; V = A 또는 C 또는 G)(SEQ ID NO: 25)를 인식하고 표적 핵산 서열 1 내지 10, 예를 들어, 그 서열로부터 3 내지 5 bp 상류의 절단을 유도한다. 상기 구현예에서, N은 임의의 뉴클레오타이드 잔기, 예를 들어, A, G, C 또는 T 중 임의의 것일 수 있다. Cas9 분자는 Cas9 분자의 PAM 특이성을 변화시키도록 조작될 수 있다.

[0099] 특정 구현예에서, 벡터는 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25) 중 하나의 프로토스페이서 인접 모티프(PAM)를 인식하는 적어도 하나의 Cas9 분자를 코딩한다. 특정 구현예에서, 적어도 하나의 Cas9 분자는 *에스. 아우레우스*(*S. aureus*) Cas9 분자이다. 특정 구현예에서, 적어도 하나의 Cas9 분자는 돌연변이체 *에스. 아우레우스*(*S. aureus*) Cas9 분자이다.

[0100] Cas9 단백질은 뉴클레아제 활성이 불활성화되도록 돌연변이될 수 있다. 엔도뉴클레아제 활성이 없는 불활성화된 Cas9 단백질("iCas9", 또한 "dCas9"로 나타냄)은 최근에 입체 장애를 통해 유전자 발현을 침묵화하기 위해 gRNA에 의해 박테리아, 효모, 및 인간 세포에서의 유전자로 표적화되었다. *에스. 피오게네스*(*S. pyogenes*) Cas9 서열과 관련하여 예시적인 돌연변이는: D10A, E762A, H840A, N854A, N863A 및/또는 D986A를 포함한다. *에스. 아우레우스*(*S. aureus*) Cas9 서열과 관련하여 예시적인 돌연변이는 D10A 및 N580A를 포함한다. 특정 구현예에서, Cas9 분자는 돌연변이체 *에스. 아우레우스*(*S. aureus*) Cas9 분자이다. 특정 구현예에서, 돌연변이체 *에스. 아우레우스*(*S. aureus*) Cas9 분자는 D10A 돌연변이를 포함한다. 상기 돌연변이체 *에스. 아우레우스*(*S. aureus*) Cas9를 코딩하는 뉴클레오타이드 서열은 하기에 제공되는 SEQ ID NO: 34에 개시되어있다:

```
atgaaaagga actacattct ggggctggcc atcgggatta caagcgtggg gtatgggatt
attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac
gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgctgaa acgacggaga
aggcacagaa tcagaggggt gaagaaactg ctgttcgatt acaacctgct gaccgacat
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg agtgcataac
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc
aatagcaaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa
gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc
aagcagctgc tgaagtgcga gaaggcttac caccagctgg atcagagctt catcgatact
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc
ttcggtatga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatttt
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat
gacctgaaca acctgggtcat caccagggat gaaaacgaga aactggaata ctatgagaag
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct
aaggagatcc tgggtcaacga agaggacatc aagggtacc gggtgacaag cactggaaaa
ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgacca ggaagagatc
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg
ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg
gtggacgatt tcattctgtc acccggtgtc aagcggagct tcattccagag catcaaagtg
atcaacgcca tcattcaagaa gtacggcctg cccaatgata tcattatcga gctggctag
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
```

[0101]

accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg
 attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtattc tctggaggcc
 atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgcata tattatcccc
 agaagcgtgt ccttcgacaa ttcttttaac aacaagggtgc tggtaagca ggaagagaac
 tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct
 tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag
 accaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat
 tttattaacc ggaatctggt ggacacaaga tacgctactc gcggcctgat gaatctgctg
 cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc
 acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac
 catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag
 ctggacaaag ccaagaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct
 atgcccgaag tcgagacaga acaggagtac aaggagattt tcatcactcc tcaccagatc
 aagcatatca aggatttcaa ggactacaag tactctcacc ggggtggataa aaagcccaac
 agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
 attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac
 aacaaaagtc ccgagaagct gctgatgtac caccatgatc ctcagacata tcagaaactg
 aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag
 actgggaact acctgaccaa gtatagcaaa aaggataatg gcccctgat caagaagatc
 aagtactatg ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt
 cgcaacaagg tggtaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac
 ggctgtgata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat
 gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca
 gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg
 gtcatcgggg tgaacaatga tctgtgaac cgcattgaag tgaatatgat tgacatcact
 taccgagagt atctggaaaa catgaatgat aagcgccccc ctgaattat caaaacaatt
 gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag
 gtgaagagca aaaagcacc tcagattatc aaaaagggc [SEQ ID NO: 34].

[0102]

[0103]

특정 구현예에서, 돌연변이체 에스. 아우레우스(*S. aureus*) Cas9 분자는 N580A 돌연변이를 포함한다. 상기 돌연
 변이체 에스. 아우레우스(*S. aureus*) Cas9 분자를 코딩하는 뉴클레오타이드 서열은 하기에 제공되는 SEQ ID NO:
 35에 개시되어있다:

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt
 attgactatg aaacaaggga cgtgatcgac gcaggcgta gactgttcaa ggaggccaac
 gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgctgaa acgacggaga
 aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgaccat
 tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg
 tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg agtgcataac
 gtcaatgagg tggaaagga caccggcaac gagctgtcta caaaggaaca gatctcacgc
 aatagcaaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa
 gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc
 aagcagctgc tgaagtgc gaaggcttac caccagctgg atcagagctt catcgatact
 tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc
 ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatttt
 ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat
 gacctgaaca acctggtcat caccagggt gaaaacgaga aactggaata ctatgagaag
 ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct
 aaggagatcc tggtaacga agaggacatc aagggtacc gggtgacaag cactggaaaa
 ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaagaa
 atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc
 tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgaccca ggaagagatc
 gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
 aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg
 ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg
 gtggacgatt tcattctgtc acccgtggtc aagcggagct tcattccagag catcaaagtg
 atcaacgcca tcataagaa gtacggcctg cccaatgata tcattatcga gctggctagg
 gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
 accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg
 attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtattc tctggaggcc
 atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc
 agaagcgtgt ccttcgacaa ttctttaac aacaagggtc tggtaagca ggaagaggcc
 tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct
 tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag
 accaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat
 ttatttaacc ggaatctggt ggacacaaga tacgtactc gcggcctgat gaatctgctg
 cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cgcggggttc
 acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac
 catgccgaag atgctctgat tatcgaaat gccgacttca tctttaagga gtggaaaaag
 ctggacaaaag ccaagaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct
 atgccgaaa tcgagacaga acaggagtag aaggagattt tcactactcc tcaccagatc

[0104]

aagcatatca aggatattcaa ggactacaag tactctcacc ggggtggataa aaagcccaac
 agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
 attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac
 aacaaaagtc cggagaagct gctgatgtac caccatgatc ctacagacata tcagaaactg
 aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag
 actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc
 aagtactatg ggaacaagct gaatgcccac ctggacatca cagacgatta cctaacagt
 cgcaacaagg tggtaacgct gtactgaag ccatacagat tcgatgtcta tctggacaac
 ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat
 gaagtgaata gcaagtgtc cgaagaggct aaaagctga aaaagattag caaccaggca
 gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg
 gtcatcgggg tgaacaatga tctgctgaac cgattgaag tgaatatgat tgacatcact
 taccgagagt atctggaaaa catgaatgat aagcgcccc ctcgaattat caaacaatt
 gcctctaaga ctacagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag
 gtgaagagca aaaagcacc tcagattatc aaaaagggc [SEQ ID NO: 35].

[0105]

[0106]

Cas9 분자를 코딩하는 핵산은 합성 핵산 서열일 수 있다. 예를 들어, 합성 핵산 분자는 화학적으로 변형될 수 있다. 합성 핵산 서열은 코돈 최적화될 수 있으며, 예를 들어, 적어도 하나의 비-공통 코돈 또는 덜-공통 코돈이 공통 코돈으로 치환된다. 예를 들어, 합성 핵산은 최적화된 메신저 mRNA의 합성을 유도할 수 있으며, 예를 들어 본원에 기술된 포유류 발현 시스템에서의 발현에 최적화된다.

[0107]

부가적으로 또는 대안적으로, Cas9 분자 또는 Cas9 폴리펩타이드를 코딩하는 핵산은 핵 위치화 서열(NLS)을 포함할 수 있다. 핵 위치화 서열은 당 분야에 공지되어있다.

[0108]

예스. *피오게네스(S. pyogenes)*의 Cas9 분자를 코딩하는 예시적인 코돈 최적화 핵산 서열은 하기에 제공되는 SEQ ID NO: 26에 개시되어있다:

atggataaaa agtacagcat cgggctggac atcggtagaa actcagtggg gtgggcccgtg
 attacggacg agtacaaggt accctccaaa aaatttaag tgctgggtaa cacggacaga
 cactctataa agaaaaatct tattggagcc ttgctgttcg actcaggcga gacagccgaa
 gccacaaggt tgaagcggac cgccaggagg cggtagacca ggagaaagaa ccgcatatgc
 tacctgcaag aaatcttcag taacgagatg gcaaagggtg acgatatgctt tttccatcgc
 ctggaagaat cctttcttgt tgaggaagac aagaagcagc aacggcacc ccatcttggc
 aatattgtcg acgaagtggc ataccacgaa aagtaccgca ctatctacca cctcaggaag

[0109]

aagctgggtgg actctaccga taaggcggac ctcagactta tttatttggc actcgccac
atgattaaat ttagaggaca tttcttgatc gagggcgacc tgaaccgga caacagtgc
gtcgataagc tgtcatcca acttgtgcag acctacaatc aactgttcga agaaaaccct
ataaatgctt caggagtcga cgctaaagca atcctgtccg cgcgcctctc aaaatctaga
agacttgaga atctgattgc tcagttgccc ggggaaaaga aaaatggatt gtttgcaac
ctgatcgccc tcagtctcgg actgaccca aatttcaaaa gtaacttcga cctggccgaa
gacgctaagc tccagctgtc caaggacaca tacgatgacg acctcgaca tctgctggcc
cagattgggg atcagtagc cgatctctt ttggcagcaa agaactgtc cgacgccatc
ctgttgagcg atatcttgag agtgaacacc gaaattacta aagcaccct tagcgcatct
atgatcaagc ggtacgacga gcatcatcag gatctgacc tgetgaaggc tcttgtgagg
caacagctcc ccgaaaaata caaggaaatc ttctttgacc agagcaaaaa cggctacgct
ggctatatag atgggtgggc cagtcaggag gaattctata aattcatcaa gccattctc
gagaaaatgg acggcacaga ggagttgtg gtcaactta acagggagga cctgctgcgg
aagcagcggc cctttgaca cggtctatc cccaccaga ttcactctggg cgaactgcac
gcaatcctga ggaggcagga ggatttttat ccttttctta aagataaccg cgagaaaata
gaaaagattc ttacattcag gatcccgtag tacgtgggac ctctcgcccg gggcaattca
cggtttgccct ggatgacaag gaagtcagag gagactatta cacottggaa cttcgaagaa
gtggtggaca aggggtgcatc tgcccagctc ttcactgagc ggatgacaaa ttttgacaag
aacctcccta atgagaaggt gctgcccata cattctctgc tctacgagta ctttaccgtc
tacaatgaac tgactaaagt caagtacgtc accgagggaa tgaggaaggc ggcattcctt
agtggagaac agaagaaggc gattgtagac ctgttggtca agaccaacag gaaggtgact
gtgaagcaac ttaaagaaga ctactttaag aagatcgaat gttttgacag tgtggaaatt
tcaggggttg aagaccgctt caatgcgtca ttggggactt accatgatct tctcaagatc
ataaaggaca aagacttcct ggacaacgaa gaaaatgagg atattctcga agacatcgtc
ctcaccctga ccctgttcga agacagggaa atgatagaag agcgcttgaa aacctatgcc
cacctcttcg acgataaagt tatgaagcag ctgaagcgca ggagatacac aggatgggga
agattgtcaa ggaagctgat caatggaatt agggataaac agagtggcaa gaccatactg
gatttcctca aatctgatgg cttcgccaat aggaacttca tgcaactgat tcacgatgac
tctcttacct tcaaggagga cattcaaaag gctcaggtga gcgggcaggg agactccctt
catgaacaca tcgcgaattt ggcagggtcc cccgctatta aaaaggcat cttcaaact
gtcaagggtg tgatgaatt ggtcaaggta atgggcagac ataagccaga aaatattgtg
atcgagatgg cccgcgaaaa ccagaccaca cagaagggcc agaaaaatag tagagagcgg
atgaagagga tcgaggagg catcaaagag ctgggatctc agattctcaa agaaccctc
gtagaaaaca cacagctgca gaacgaaaaa ttgtacttgt actatctgca gaacggcaga
gacatgtacg tcgaccaaga acttgatatt aatagactgt ccgactatga cgtagaccat

[0110]

atcgtgcccc agtccttctt gaaggacgac tccattgata acaaagtctt gacaagaagc
gacaagaaca ggggtaaaag tgataatgtg cctagcgagg aggtggtgaa aaaaatgaag
aactactggc gacagctgct taatgcaaag ctcattacac aacggaagtt cgataatctg
acgaaagcag agagaggtgg ctgtgtctgag ttggacaagg cagggtttat taagcggcag
ctggtggaac ctaggcagat cacaaagcac gtggcgacag ttttgacag ccggatgaac
acaaatacag acgaaaatga taaactgata cgagaggtca aagttatcac gctgaaaagc
aagctggtgt ccgattttcg gaaagacttc cagttctaca aagttcgcga gattaataac
taccatcatg ctcacgatgc gtacctgaac gctgttgcg ggaccgcctt gataaagaag
taccctaaagc tggaatccga gttcgtatag ggggattaca aagtgtacga tgtgaggaac
atgatagcca agtcgagca ggagattgga aaggccacag ctaagtactt cttttattct
aacatcatga atttttttta gacggaaatt accctggcca acggagagat cagaaaagcg
ccccttatag agacaaatgg tgaacaggt gaaatcgtct gggataaggg cagggatttc
gctactgtga ggaaggtgct gagtatgcca caggtaaata tcgtgaaaaa aaccgaagta
cagaccggag gattttccaa ggaagcatt ttgcctaaaa gaaactcaga caagctcatc
gcccgaaga aagattggga ccctaagaaa tacgggggat ttgactcacc caccgtagcc
tattctgtgc tgggtgtagc taaggtgga aaaggaaagt ctaagaagct gaagtccgtg
aaggaaactc tgggaatcac tatcatgga agatcatcct ttgaaaagaa ccctatcgat
ttcctggagg ctaagggta caaggaggtc aagaaagacc tcatcattaa actgcaaaaa
tactctctct tcgagctgga aaatggcagg aagagaatgt tggccagcgc cggagagctg
caaaagggaa acgagcttgc tctgccctcc aaatatgtta attttctcta tctcgcttcc
cactatgaaa agctgaaagg gtctcccgaa gataacgagc agaagcagct gttcgtcgaa
cagcacaagc actatctgga tgaataatc gaacaaataa gcgagttcag caaaaggggt
atcctggcgg atgctaattt ggacaaagta ctgtctgctt ataacaagca ccgggataag
cctattaggg aacaagccga gaatataatt cacctcttta cactcacgaa tctcgagacc
cccgccgct tcaataactt tgatacgact atcgaccgga aacggtatag cagtacaaa
gaggtcctcg atgccacct catccaccag tcaattactg gcctgtacga
aacacggatcgacctctctc aactgggcgg cgactag [SEQ ID NO: 26].

[0111]

[0112] 에스. 피오게네스(*S. pyogenes*) Cas9 분자의 상응하는 아미노산 서열은 하기에 제공되는 SEQ ID NO: 27에 개시되어 있다:

MDKKYSIGLDIGTNSVGWAVITDEYKVPKFKVLGNTDRHSIKKNLIGALLFDSGETAEATRLK
RTARRRYTRRKNRICYLQEIFSNEMAKVDDSFHRLEESFLVEEDKKHERHPIFGNIVDEVAYHE
KYPTIYHLRKLVDSTDKADRLRIYLALAHMIKFRGHFLIEGDLNPDNSVDKLFILVQTYNQL
FEENPINASGVDAKAILSRLSKSRLENLIAQLPGEKKNLFGNLIALSLGLTPNFKSNFDLAE
DAKLQLSKDTYDDLDNLLAQIGDQYADLFLLAKNLSDAILLSDILRVNTEITKAPLSASMIKRY
DEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELL
VKLNREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYFPLKDNREKIEKILTRIPYYVGPL
ARGNSRFAWMTRKSEETITPWNFEVVVDKGASAQSFIERMTNFDKNLPNEKVLPHKSLLEYFTV
YNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVED
RFNASLGTYHDLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKYAHLFDDKVMKQ
LKRRTYTGWRLSRKLINGIRDKQSGKTILDFLKSDFANRNFQILHDDSLTFKEDIQKAQVSG
QGDSLHEHIANLAGSPAIIKKGILQTVKVVDELVKVMGRHKPENIVIEARENQTQKGQKNSRER
MKRIEIEGKELGSQLKEHPVENTQLQNEKLYLYLQNGRDMYVDQELDINRLSDYDVDHIVPQS
FLKDDSIDNKVLRSDKNRKGSDNVPSEEVKKMKNYWRQLLNAKLITQRKFDNLTKAERGLSE
LDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSILVSDFRKDFQFYKV
REINNYHHAHDAYLNAVGTALIKKYPKLESEFVYGDYKVYDVRKMIKSEQEIGKATAKYFFYS
NIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKVLSPQVNIKKTEVQTGGF
SKESILPKRNSDKLIARKKDWDPKKGFGDSPTVAYSVLVVAKVEKGSKKLKSVKELLGITIME
RSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLEFLENKRMLASAGELQKGNELALPSKYVNF
LYLASHYEKLKSPEDNEQKQLFVQHKHYLDEIIIEQISEFSKRVLADANLDKVL SAYNKHARDK
PIREQAENI IHLFTLTNLGAPAAFKY FDTTIDRKRYTSTKEVLDTLIHQSTGLYETRIDLSQL
GGD [SEQ ID NO: 27]

[0114]

[0115] 에스. 아우레우스(*S. aureus*)의 Cas9 분자를 코딩하고, 선택적으로 핵 위치화 서열(NLSs)을 함유하는 예시적인 코돈 최적화 핵산 서열은 하기 제공되는 SEQ ID NO: 28-32, 43 및 44에 개시되어있다. 에스. 아우레우스(*S. aureus*)의 Cas9 분자를 코딩하는 또 다른 예시적인 코돈 최적화 핵산 서열은 SEQ ID NO: 83의 뉴클레오타이드 1293-4451을 포함한다.

[0116] SEQ ID NO: 28은 이하에 개시되어 있다:

```
atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt
attgactatg aaacaaggga cgtgatcgac gcaggcgta gactgttcaa ggaggccaac
gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgctgaa acgacggaga
aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgaccat
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg
tcagagggaag agttttccgc agctctgctg cacctggcta agcgccgagg agtgcataac
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc
aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa
gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc
aagcagctgc tgaagtgc gaaggcttac caccagctgg atcagagctt catcgatact
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc
```

[0117]

```
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctathtt
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat
gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct
aaggagatcc tggtaacga agaggacatc aagggtacc gggtgacaag cactggaaaa
ccagagttca ccaatctgaa agtgatcac gatattaagg acatcacagc acggaagaa
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgaccca ggaagagatc
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg
ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg
gtggacgatt tcattctgtc acccggtgtc aagcggagct tcatccagag catcaaagtg
atcaacgcca tcataagaa gtacggcctg ccaatgata tcattatcga gctggctagg
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
accaatgaac gcattgaaga gattatccga actaccgga aagagaacgc aaagtacctg
attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtatc tctggaggcc
tccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc
agaagcgtgt ccttcgacaa ttcctttaac aacaagggtc tggtaagca ggaagagaac
tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaaggccg catcagcaag
acaaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat
tttattaacc ggaatctggt ggacacaaga tacgtactc gcggcctgat gaatctgctg
cgatccattt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcggttc
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac
catgccgaag atgctctgat tatcgaaaat gccgacttca tctttaagga gtggaaaaag
ctggacaaag ccaagaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct
atgcccgaat tcgagacaga acaggagtac aaggagattt tcatcactcc tcaccagatc
aagcatatca aggatttcaa ggactacaag tactctcacc gggtgataa aaagcccaac
agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgatc
aacaaaagtc ccgagaagct gctgatgtac caccatgatc ctcagacata tcagaaactg
aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag
actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc
aagtactatg ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt
cgcaacaagg tggtaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac
```

[0118]

ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat
 gaagtgaata gcaagtgtct cgaagaggct aaaaagctga aaaagattag caaccaggca
 gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg
 gtcatcgggg tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact
 taccgagagt atctggaaaa catgaatgat aagcgccccc ctogaattat caaaacaatt
 gcctctaaga ctcaagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag
 gtgaagagca aaaagcacc tcagattatc aaaaagggc [SEQ ID NO: 28]

[0119]

[0120]

SEQ ID NO: 29는 이하에 개시되어 있다:

atgaagcgga actacatcct ggccctggac atcggcatca ccagcgtggg ctacggcatc
 atcgactacg agacacggga cgtgatcgat gccggcgtgc ggctgttcaa agaggccaac
 gtggaaaaca acgagggcag gcggagcaag agaggcgcca gaaggctgaa gcggcggagg
 cggcatagaa tccagagagt gaagaagctg ctgttcgact acaacctgct gaccgaccac
 agcgagctga gcggcatcaa cccctacgag gccagagtga agggcctgag ccagaagctg
 agcgagggaag agttctctgc cgccctgctg cacctggcca agagaagagg cgtgcacaac
 gtgaacgagg tggagagga caccggcaac gagctgtcca ccaaagagca gatcagccgg
 aacagcaagg ccttggaaga gaaatacgtg gccgaactgc agctggaacg gctgaagaaa
 gacggcgaag tgcggggcag catcaacaga ttcaagacca gcgactacgt gaaagaagcc
 aaacagctgc tgaaggtgca gaaggcctac caccagctgg accagagctt catcgacacc
 tacatcgacc tgctggaaac ccggcggaac tactatgagg gacctggcga gggcagcccc
 ttccggctgga aggacatcaa agaattgtac gagatgctga tgggccactg cacctacttc
 cccgaggaac tgcggagcgt gaagtacgcc tacaacgccg acctgtacaa cgccctgaac
 gacctgaaca atctcgtgat caccagggac gagaacgaga agctggaata ttacgagaag
 ttccagatca tcgagaacgt gttcaagcag aagaagaagc ccacctgaa gcagatcgcc
 aaagaaatcc tcgtgaacga agaggatatt aagggctaca gagtgaccag caccggcaag
 cccgagttca ccaacctgaa ggtgtaccac gacatcaagg acattaccgc ccggaagag
 attattgaga acgccgagct gctggatcag attgccaaga tcctgaccat ctaccagagc
 agcgaggaca tccaggaaga actgaccaat ctgaactccg agctgaccca ggaagagatc
 gagcagatct ctaatctgaa gggctatacc ggcaccaca acctgagcct gaaggccatc
 aacctgatcc tggacgagct gtggcacacc aacgacaacc agatcgctat cttcaaccgg
 ctgaagctgg tgcccaagaa ggtggacctg tcccagcaga aagagatccc caccacctg
 gtggacgact tcatoctgag ccccgctcgtg aagagaagct tcatccagag catcaaagtg
 atcaacgcca tcatcaagaa gtacggcctg cccaacgaca tcattatcga gctggcccg
 gagaagaact ccaaggacgc ccagaaaatg atcaacgaga tgcagaagcg gaaccggcag
 accaacgagc ggatcgagga aatcatccgg accaccggca aagagaacgc caagtacctg

[0121]

atcgagaaga tcaagctgca cgacatgcag gaaggcaagt gcctgtacag cctggaagcc
atccctctgg aagatctgct gaacaacccc ttcaactatg aggtggacca catcatcccc
agaagcgtgt ccttcgacaa cagcttcaac aacaaggtgc tcgtgaagca ggaagaaaac
agcaagaagg gcaaccggac cccattccag tacctgagca gcagcgacag caagatcagc
tacgaaacct tcaagaagca catcctgaat ctggccaagg gcaagggcag aatcagcaag
accaagaaag agtatctgct ggaagaacgg gacatcaaca ggttctccgt gcagaaagac
ttcatcaacc ggaacctggg ggataccaga tacgccacca gaggcctgat gaacctgctg
cggagctact tcagagtga caacctggac gtgaaagtga agtccatcaa tggcgcttc
accagctttc tgccggcgaa gtggaagttt aagaaagagc ggaacaaggg gtacaagcac
cacgccgagg acgccctgat cattgccaac gccgatttca tcttcaaaga gtggaagaaa
ctggacaagg ccaaaaaagt gatggaaaa cagatgttcg aggaaaagca ggccgagagc
atgcccgaga tcgaaaccga gcaggagtac aaagagatct tcatcacccc ccaccagatc
aagcacatta aggacttcaa ggactacaag tacagccacc ggggtggaca gaagcctaata
agagagctga ttaacgacac cctgtactcc acccggaagg acgacaagg caacaccctg
atcgtgaaca atctgaacgg cctgtacgac aaggacaatg acaagctgaa aaagctgac
aacaagagcc ccgaaaagct gctgatgtac caccacgacc ccagaccta ccagaaactg
aagctgatta tggaaacagta cggcgacgag aagaatcccc tgtacaagta ctacgaggaa
accgggaact acctgaccaa gtactccaaa aaggacaacg gccccgtgat caagaagatt
aagtattacg gcaacaaact gaacgcccac ctggacatca ccgacgacta cccaacagc
agaacaagg tcgtgaagct gtccctgaag ccctacagat tcgacgtgta cctggacaat
ggcgtgtaca agttcgtgac cgtgaagaat ctggatgtga tcaaaaaaga aaactactac
gaagtgaata gcaagtgcta tgaggaagct aagaagctga agaagatcag caaccaggcc
gagtttatcg cctccttcta caacaacgat ctgatcaaga tcaacggcga gctgtataga
gtgatcggcg tgaacaacga cctgctgaac cggatcgaag tgaacatgat cgacatcacc
taccgcgagt acctgaaaa catgaacgac aagaggcccc ccaggatcat taagacaatc
gcctccaaga ccagagcat taagaagtac agcacagaca ttctgggcaa cctgtatgaa
gtgaaatcta agaagcacc tcagatcatc aaaaagggc [SEQ ID NO: 29]

[0122]

[0123]

SEQ ID NO: 30은 이하에 개시되어 있다:

atgaagcgca actacatcct cggactggac atcggcatta cctccgtggg atacggcatc
atcgattacg aaactagga tgatgatcgc gctggagtca ggctgttcaa agaggcgaac
gtggagaaca acgaggggag gcgctcaaag aggggggccc gccggctgaa gcgcccggc
agacatagaa tccagcgct gaagaagctg ctgttcgact acaaccttct gaccgaccac
tccgaacttt ccggcatcaa ccatatgag gctagagtga agggattgtc ccaaaagctg
tccgaggaag agttctccgc cgcgttgctc cacctcgcca agcgcagggg agtgcacaat

[0124]

gtgaacgaag tggaagaaga taccggaac gagctgtcca ccaaggagca gatcagccgg
aactccaagg ccctggaaga gaaatactgt gcggaactgc aactggagcg gctgaagaaa
gacggagaag tgcgcggctc gatcaaccgc ttcaagacct cgactactgt gaaggaggcc
aagcagctcc tgaagtgc aaaggcctat caccaacttg accagtcctt tatcgatacc
tacctgcatc tgcctgagac tcggcggact tactacgagg gtccagggga gggctcccca
tttggttga aggatattaa ggagtgttac gaaatgctga tgggacactg cacatacttc
cctgaggagc tgcggagcgt gaaatacgca tacaacgcag acctgtacaa cgcgctgaac
gacctgaaca atctcgtgat ccccgggac gagaacgaaa agctcgagta ttacgaaaag
ttccagatta ttgagaacgt gttcaaacag aagaagaagc cgacactgaa gcagattgcc
aaggaaatcc tcgtgaacga agaggacatc aagggtctac gagtgcctc aacgggaaag
ccggagttca ccaatctgaa ggtctaccac gacatcaaag acattaccgc ccggaaggag
atcattgaga acgcggagct gttggaccag attgcgaaga ttctgacct ctaccaatcc
tccgaggata ttcaggaaga actcaccaac ctcaacagcg aactgaccca ggaggagata
gagcaaactc ccaacctgaa gggctacacc ggaactcata acctgagcct gaaggccatc
aacttgatcc tggacgagct gtggcacacc aacgataacc agatcgctat tttcaatcgg
ctgaagctgg tccccaaagaa agtggacctc tcacaacaaa aggagatccc tactaccctt
gtggacgatt tcattctgtc ccccggtgtc aagagaagct tcatacagtc aatcaaagt
atcaatgcca ttatcaagaa atacggtctg cccaacgaca ttatcattga gctcgccgc
gagaagaact cgaaggagc ccagaagatg attaacgaaa tgcagaagag gaaccgacag
actaacgaac ggatcgaaga aatcatccgg accaccggga aggaaaacgc gaagtacctg
atcgaagaaga tcaagctcca tgacatgcag gaaggaaagt gtctgtactc gctggaggcc
attccgctgg aggacttgct gaacaaccct ttaactacg aagtggatca tatcattccg
aggagcgtgt cattcgacaa ttccttcaac aacaaggtcc tcgtgaagca ggaggaaaac
tcgaagaagg gaaaccgcac gccgttccag tacctgagca gcagcgactc caagatttcc
tacgaaacct tcaagaagca catcctcaac ctggcaaaagg ggaagggtcg catctccaag
accaagaagg aatatctgct ggaagaaaga gacatcaaca gattctccgt gcaaaaggac
ttcatcaacc gcaacctcgt ggatactaga tacgtactc ggggtctgat gaacctcctg
agaagctact tttagagtga caatctggac gtgaaggtca agtcgattaa cggaggtttc
acctccttcc tgcggcgcaa gtggaagttc aagaaggaaac ggaacaaggg ctacaagcac
cacgccgagg acgccctgat cattgccaac gccgacttca tcttcaaaga atggaagaaa
cttgacaagg ctaagaaggc catggaaaac cagatgttcg aagaaaagca ggccgagttc
atgcctgaaa tcgagactga acaggagtac aaggaaatct ttattacgcc acaccagatc
aaacacatca aggatttcaa ggattacaag tactcacatc gcgtggacaa aaagccgaac
agggactga tcaacgacac cctctactcc acccggaagg atgacaaagg gaataccctc
atcgtcaaca accttaacgg cctgtacgac aaggacaacg ataagctgaa gaagctcatt

[0125]

aacaagtcgc ccgaaaagtt gctgatgtac caccacgacc ctgagactta ccagaagctc
aagctgatca tggagcagta tggggacgag aaaaaccctg tgtacaagta ctacgaagaa
actgggaatt atctgactaa gtactccaag aaagataacg gcccgtgat taagaagatt
aagtactacg gcaacaagct gaacgcccat ctggacatca ccgatgacta ccctaattcc
cgcaacaagg tcgtcaagct gagcctcaag ccctaccggt ttgatgtgta ccttgacaat
ggagtgtaca agttcgtgac tgtgaagaac ctgacgtga tcaagaagga gaactactac
gaagtcaact ccaagtgcga cgaggaagca aagaagttga agaagatctc gaaccaggcc
gagttcattg cctccttcta taacaacgac ctgattaaga tcaacggcga actgtaccgc
gtcattggcg tgaacaacga tctcctgaac cgcatcgaag tgaacatgat cgacatcact
taccgggaat acctggagaa tatgaacgac aagcgccgc ccggatcat taagactatc
gcctcaaaga ccagtcgat caagaagtac agcaccgaca tctgggcaa cctgtacgag
gtcaaactga agaagcacc ccagatcatc aagaaggga [SEQ ID NO: 30]

[0126]

[0127] SEQ ID NO: 31은 이하에 개시되어 있다:

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAAGCGGAACACAT
CCTGGGCGCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACGGGACG
TGATCGATGCCGCGTGGCGCTGTTCAAAGAGGCCAACGTGGAAAAACAACGAGGGCAGGCGGAGC
AAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCT
GTTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAG
TGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAG
AGAAGAGGCGTGACACAACCTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAGAGA
GCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACCTGCAGCTGGAACGGC
TGAAGAAAGACGGCGAAGTGGGGGCGAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAA
GCCAAACAGCTGCTGAAGGTGCAGAAGGCCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTA
CATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCCTTCGGCT
GGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTG
CGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGT
GATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCAGATCATCGAGAACGTGT
TCAAGCAGAAGAAGAAGCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGAT
ATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCGAGTTCACCAACCTGAAGGTGTACCACGA
CATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCA
AGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAG
CTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAG
CCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACAGATCGCTATCT
TCAACCGGCTGAAGCTGGTGGCCAAAGAGTGGACCTGTCCAGCAGAAAGAGATCCCCACCACC

[0128]

CTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCATCAAAGTGAT
CAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGCCCGCGAGAAGA
ACTCCAAGGACGCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGG
ATCGAGGAAATCATCCGGACACCGGCAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCT
GCACGACATGCAGGAAGGCAAGTGCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGA
ACAACCCCTTCACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCTTCGACAACAGCTTC
AACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCAGTA
CCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCA
AGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGG
TTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCT
GATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATG
GCGGCTTACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAG
CACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTTCATCTTCAAAGAGTGGAAGAACT
GGACAAGGCCAAAAAAGTGATGGAACACAGATGTTTCGAGGAAAGGCAGGCCGAGAGCATGCCCCG
AGATCGAAACCGAGCAGGAGTACAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAG
GACTTCAAGGACTACAAGTACAGCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGA
CACCCTGTACTCCACCGGAAGGACGACAAGGGCAACACCTGATCGTGAACAATCTGAACGGCC
GTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATG
TACCACCACGACCCCCAGACCTACCAGAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAA
GAATCCCCGTGTACAAGTACTACGAGGAACCGGGAACTACCTGACCAAGTACTCCAAAAGGACA
ACGGCCCCGTGATCAAGAAGATTAAAGTATTACGGCAACAACTGAACGCCCATCTGGACATCACC
GACGACTACCCCAACAGCAGAAACAGGTGCTGAAGCTGTCCCTGAAGCCCTACAGATTGACGCT
GTACCTGGACAATGGCGTGTACAAGTTCTGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAA
ACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCAGCAACCAG
GCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCTGTATAGAGT
GATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCG
AGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAAGACAATCGCCTCCAAGACC
CAGAGCATTAAGAAGTACAGCACAGACATCTCTGGCAACCTGTATGAAGTGAAATCTAAGAAGCA
CCCTCAGATCATAAAAAGGGCAAAAGGCCGGGCCACGAAAAGGCCGGCCAGGCAAAAAAGA
AAAAG [SEQ ID NO: 31]

[0129]

[0130] SEQ ID NO: 32는 이하에 개시되어 있다:

ACCGGTGCCA CCATGTACCC ATACGATGTT CCAGATTACG CTTCGCCGAA GAAAAAGCGC
AAGGTCGAAG CGTCCATGAA AAGGAACCTAC ATTCTGGGGC TGGACATCGG GATTACAAGC
GTGGGGTATG GGATTATTGA CTATGAAACA AGGGACGTGA TCGACGCAGG CGTCAGACTG

[0131]

TTCAAGGAGG CCAACGTGGA AAACAATGAG GGACGGAGAA GCAAGAGGGG AGCCAGGCGC
CTGAAACGAC GGAGAAGGCA CAGAATCCAG AGGGTGAAGA AACTGCTGTT CGATTACAAC
CTGCTGACCG ACCATTCTGA GCTGAGTGGA ATTAATCCTT ATGAAGCCAG GGTGAAAGGC
CTGAGTCAGA AGCTGTCAGA GGAAGAGTTT TCCGCAGCTC TGCTGCACCT GGCTAAGCGC
CGAGGAGTGC ATAACGTCAA TGAGGTGGAA GAGGACACCG GCAACGAGCT GTCTACAAAG
GAACAGATCT CACGCAATAG CAAAGCTCTG GAAGAGAAGT ATGTCGCAGA GCTGCAGCTG
GAACGGCTGA AGAAAGATGG CGAGGTGAGA GGGTCAATTA ATAGGTTCAA GACAAGCGAC
TACGTCAAAG AAGCCAAGCA GCTGCTGAAA GTGCAGAAGG CTTACCACCA GCTGGATCAG
AGCTTCATCG ATACTTATAT CGACCTGCTG GAGACTCGGA GAACCTACTA TGAGGGACCA
GGAGAAGGGA GCCCCTTCGG ATGGAAGAC ATCAAGGAAT GGTACGAGAT GCTGATGGGA
CATTGCACCT ATTTTCCAGA AGAGCTGAGA AGCGTCAAGT ACGCTTATAA CGCAGATCT
TACAACGCC TGAATGACCT GAACAACCTG GTCATCACC GGGATGAAAA CGAGAACTG
GAATACTATG AGAAGTTCCA GATCATCGAA AACGTGTTTA AGCAGAAGAA AAAGCCTACA
CTGAAACAGA TTGCTAAGGA GATCCTGGTC AACGAAGAGG ACATCAAGGG CTACCGGGTG
ACAAGCACTG GAAAACCAGA GTTCACCAAT CTGAAAGTGT ATCACGATAT TAAGGACATC
ACAGCACGGA AAGAAATCAT TGAGAACGCC GAACTGCTGG ATCAGATTGC TAAGATCCTG
ACTATCTACC AGAGCTCCGA GGACATCCAG GAAGAGCTGA CTAACCTGAA CAGCGAGCTG
ACCCAGGAAG AGATCGAACA GATTAGTAAT CTGAAGGGGT ACACCGGAAC ACACAACCTG
TCCCTGAAAG CTATCAATCT GATTCTGGAT GAGCTGTGGC ATACAAACGA CAATCAGATT
GCAATCTTTA ACCGGCTGAA GCTGGTCCCA AAAAAGGTGG ACCTGAGTCA GCAGAAAGAG
ATCCCAACCA CACTGGTGGA CGATTTCTATT CTGTCACCCG TGGTCAAGCG GAGCTTCATC
CAGAGCATCA AAGTGATCAA CGCCATCATC AAGAAGTACG GCCTGCCCAA TGATATCATT
ATCGAGCTGG CTAGGGAGAA GAACAGCAAG GACGCACAGA AGATGATCAA TGAGATGCAG
AAACGAAACC GGCAGACCAA TGAACGCATT GAAGAGATTA TCCGAACCTAC CGGGAAGAG
AACGCAAAGT ACCTGATTGA AAAAATCAAG CTGCACGATA TGCAGGAGGG AAAGTGTCTG
TATTCTCTGG AGGCCATCCC CCTGGAGGAC CTGCTGAACA ATCCATTCAA CTACGAGGTC
GATCATATTA TCCCCAGAAG CGTGTCTTTC GACAATTCTT TTAACAACAA GGTGCTGGTC
AAGCAGGAAG AGAACTCTAA AAAGGGCAAT AGGACTCCTT TCCAGTACCT GTCTAGTTCA
GATTCCAAGA TCTCTTACGA AACCTTTAAA AAGCACATTC TGAATCTGGC CAAAGGAAAG
GGCCGCATCA GCAAGACCAA AAAGGAGTAC CTGCTGGAAG AGCGGGACAT CAACAGATTC
TCCGTCCAGA AGGATTTTAT TAACCGGAAT CTGGTGGACA CAAGATACGC TACTCGCGGC
CTGATGAATC TGCTGCGATC CTATTTCCGG GTGAACAATC TGGATGTGAA AGTCAAGTCC
ATCAACGGCG GGTTCACATC TTTTCTGAGG CGCAAATGGA AGTTTAAAAA GGAGCGCAAC
AAAGGGTACA AGCACCATGC CGAAGATGCT CTGATTATCG CAAATGCCGA CTTCATCTTT
AAGGAGTGGA AAAAGCTGGA CAAAGCCAAG AAAGTGATGG AGAACCAGAT GTTCGAAGAG

[0132]

AAGCAGGCCG AATCTATGCC CGAAATCGAG ACAGAACAGG AGTACAAGGA GATTTTCATC
 ACTCCTCACC AGATCAAGCA TATCAAGGAT TTCAAGGACT ACAAGTACTC TCACCGGGTG
 GATAAAAAGC CCAACAGAGA GCTGATCAAT GACACCCTGT ATAGTACAAG AAAAGACGAT
 AAGGGGAATA CCCTGATTGT GAACAATCTG AACGGACTGT ACGACAAAGA TAATGACAAG
 CTGAAAAAGC TGATCAACAA AAGTCCCGAG AAGCTGCTGA TGTACCACCA TGATCCTCAG
 ACATATCAGA AACTGAAGCT GATTATGGAG CAGTACGGCG ACGAGAAGAA CCCACTGTAT
 AAGTACTATG AAGAGACTGG GAACTACCTG ACCAAGTATA GCAAAAAGGA TAATGGCCCC
 GTGATCAAGA AGATCAAGTA CTATGGGAAC AAGCTGAATG CCCATCTGGA CATCACAGAC
 GATTACCTTA ACAGTCGCAA CAAGGTGGTC AAGCTGTCAC TGAAGCCATA CAGATTGAT
 GTCTATCTGG ACAACGGCGT GTATAAATTT GTGACTGTCA AGAATCTGGA TGTATCAAA
 AAGGAGAACT ACTATGAAGT GAATAGCAAG TGCTACGAAG AGGCTAAAA GCTGAAAAAG
 ATTAGCAACC AGGCAGAGTT CATCGCCTCC TTTTACAACA ACGACCTGAT TAAGATCAAT
 GCGGAACCTGT ATAGGGTCAT CGGGGTGAAC AATGATCTGC TGAACCGCAT TGAAGTGAAT
 ATGATTGACA TCCTTACCG AGAGTATCTG GAAAACATGA ATGATAAGCG CCCCCCTCGA
 ATTATCAAAA CAATTGCCTC TAAGACTCAG AGTATCAAAA AGTACTCAAC CGACATTCTG
 GGAAACCTGT ATGAGGTGAA GAGCAAAAAG CACCCTCAGA TTATCAAAA GGGCTAAGAA
 TTC [SEQ ID NO: 32]

[0133]

[0134]

SEQ ID NO: 43은 이하에 개시되어 있다:

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAAGCGGAACATACAT
 CCTGGGCCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACGGGACG
 TGATCGATGCCGGCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGGAGC
 AAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCT
 GTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAGGCCAGAG
 TGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAG
 AGAAGAGGCGTGCAACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGA
 GCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACGAGCTGGAACGGC
 TGAAGAAAGACGGCGAAGTGCGGGGACAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAA
 GCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTA
 CATCGACCTGCTGGAAACCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCTTCGGCT
 GGAAGGACATCAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTG
 CGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGT
 GATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCAGATCATCGAGAACGTGT
 TCAAGCAGAAGAAGGCCACCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGAT
 ATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCGAGTTACCAACCTGAAGGTGTACCACGA

[0135]

CATCAAGGACATTACCGCCCGAAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCA
 AGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTCCGAG
 CTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAG
 CCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACAGATCGCTATCT
 TCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACC
 CTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCATCAAAGTGAT
 CAACGCCATCATCAAGAAGTACGGCTGCCCAACGACATCATTATCGAGCTGGCCCCGCGAGAAGA
 ACTCCAAGGACGCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGG
 ATCGAGGAAATCATCCGACCACCGGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCT
 GCACGACATGCAGGAAGGCAAGTGCCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGA
 ACAACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCTTCGACAAACAGCTTC
 AACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCATTCAGTA
 CCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCA
 AGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGG
 TTCTCCGTGCGAAGAACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCT
 GATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATG
 GCGGCTTCACCAGCTTCTCGCGCGGAAGTGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAG
 CACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTTCATCTCAAAGAGTGAAGAAAAT
 GGACAAGGCCAAAAAAGTGAAGAAAACAGATGTTTCGAGGAAAAGCAGGCCGAGAGCATGCCCG
 AGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAG
 GACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGA
 CACCCTGTACTCCACCCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCC
 TGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATG
 TACCACCACGACCCCCAGACCTACCAGAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAA
 GAATCCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGGACA
 ACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAACTGAACGCCCATCTGGACATCACC
 GACGACTACCCCAACAGCAGAAACAAGGTCTGTAAGCTGTCCCTGAAGCCCTACAGATTCGACGT
 GTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAA
 ACTACTACGAAGTGAATAGCAAGTCTATGAGGAAGCTAAGAAGCTGAAGAAGATCAGCAACCAG
 GCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCTGTATAGAGT
 GATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCG
 AGTACCTGGAAAACATGAACGACAAGAGGCCCCCAGGATCATTAAGACAATCGCCTCCAAGACC
 CAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAATCTAAGAAGCA
 CCCTCAGATCATCAAAAAGGGCAAAAAGCCGGCGGCCACGAAAAGGCCGGCCAGGCAAAAAGA
 AAAAG [SEQ ID NO: 43]

[0136]

[0137]

일부 구현예에서, 에스. 아우레우스(*S. aureus*) Cas9 분자를 코딩하는 뉴클레오타이드 서열은 하기에 제공되는
 SEQ ID NO: 44의 뉴클레오타이드 서열을 포함한다:

AAGCGGAAC TACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATCGACTA
CGAGACACGGGACGTGATCGATGCCGGCTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACG
AGGGCAGGCGGAGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGA
GTGAAGAAGCTGCTGTTCTGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCC
CTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGC
TGACCTGGCCAAGAGAAGAGGCGTGACAACTGAACGAGGTGGAAGAGGACACCGGCAACGAG
CTGTCCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACT
GCAGCTGGAACGCGTGAAGAAAGACGGCGAAGTGCAGGGCAGCATCAACAGATTCAAGACCAGCG
ACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGC
TTCATCGACACCTACATCGACCTGCTGGAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGG
CAGCCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGCACCTACT
TCCCCGAGGAAC TCGGAGCGTGAAGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGAC
CTGAACAATCTCGTGATCACCAGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGAT
CATCGAGAACGTGTTCAAGCAGAAGAAGAGCCACCTGAAGCAGATCGCCAAGAAATCCTCG
TGAAACGAAGAGGATATTAAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTACCAACCTG
AAGGTGTACCACGACATCAAGGACATTACCGCCCGAAAGAGATTATTGAGAACGCCGAGCTGCT
GGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCA
ATCTGAAC TCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCGGC
ACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAA
CCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCAGCAGAAAG
AGATCCCCACCACCCTGGTGAGCACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAG
AGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCCAACGACATCATTATCGAGCT
GGCCCGCAGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAACCGGC
AGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATC
GAGAAGATCAAGCTGCACGACATCGAGGAAGGCAAGTGCCGTGTACAGCCTGGAAGCCATCCCTCT
GGAAGATCTGCTGAACAACCCCTTCAACTATGAGGTGGACCACATCATCCCAGAAGCGTGTCTCT
TCGACAACAGCTTCAACAACAAGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGG
ACCCCATTCAGTACCTGAGCAGCAGCGACGCAAGATCAGCTACGAAACCTTCAAGAAGCACAT
CCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAAC
GGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATAC

[0138]

GCCACCAGAGGCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGT
GAAGTCCATCAATGGCGGCTTACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGA
ACAAGGGGTACAAGCACACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAA
GAGTGGAAGAACTGGACAAGGCCAAAAAGTGATGGAAAACAGATGTTTCGAGGAAAAGCAGGC
CGAGAGCATGCCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCACCAGA
TCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGA
GAGCTGATTACGACACCTGTACTCCACCCGAAGGACGACAAGGGCAACACCTGATCGTGAA
CAATCTGAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCG
AAAAGCTGCTGATGTACCACCAGACCCCCAGACCTACCAGAACTGAAGCTGATTATGGAACAG
TACGGCGACGAGAAGAATCCCTGTACAAGTACTACGAGGAAACCGGGAACCTGACCAAGTA
CTCCAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAACTGAACGCCC
ATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCC
TACAGATTCCGAGCTGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGT
GATCAAAAAAGAAAAC TACTACGAAGTGAATAGCAAGTGTATGAGGAAGCTAAGAAGCTGAAGA
AGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGC
GAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGA
CATCACCTACCGGAGTACCTGGAACCATGAACGACAAGAGGCCCCCAGGATCATTAAGACAA
TCGCCTCCAAGACCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTG
AAATCTAAGAAGCACCCCTCAGATCATCAAAAAGGGC [SEQ ID NO: 44]

[0139]

[0140] 에스. 아우레우스(*S. aureus*) Cas9 분자의 아미노산 서열은 하기 제공되는 SEQ ID NO: 33에 개시되어있다.

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGRARLKRRRRHRIQ
RVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGN
ELSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVKQKAYHQLDQ
SFIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALN
DLNNLVITRDENEKLEYEYKFIENVFQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTN
LKVYHDIKDITARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNLSELTQEEIEQISNLKGYT
GTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFI
QSIKVINAIKKYGLPNDIIELAREKNSKDAQKMINEMQKRNQRTNERIEEIIRTTGKENAKYL
IEKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIIPRSVSFDNSFNNKVLVKQEENSCKGN
RTPFQYLSSSDSKISYETFKKHILNLAKGGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTR
YATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALIINANADFI
KEWKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPN
RELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIME
QYGEKNPLYKYEETGNYLTYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLK
PYRFDVYLDNGVYKFVTVKNLDVICKENYEVNSKCYEEAKKLKISNQAEFIASFYNNDLIKIN
GELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYE
VSKKHPQIIKKG [SEQ ID NO: 33]

[0141]

[0142] 에스. 아우레우스(*S. aureus*) Cas9 분자의 아미노산 서열은 하기 제공되는 SEQ ID NO: 45에 개시되어있다.

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGRARLKRRRRHRIQR
VKKLLFDYNLLTDHSELGINPYEARVKGLSQKLSEEEFSAALLHLAKRRGVHNVNEVEEDTGNE
LSTKEQISRNSKALEEKYVAELQLERLKKDGEVRGSINRFKTSDYVKEAKQLLKVKQKAYHQLDQS
FIDTYIDLLETRRTYYEGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALND
LNNLVITRDENEKLEYEYKFIENVFQKKKPTLKQIAKEILVNEEDIKGYRVTSTGKPEFTNL
KVYHDIKDITARKEIENAEELLDQIAKILTIYQSSEDIQEELTNLNLSELTQEEIEQISNLKGYTG
THNLSLKAINLILDELWHTNDNQIAIFNRLKLVPKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQ
SIKVINAIKKYGLPNDIIELAREKNSKDAQKMINEMQKRNQRTNERIEEIIRTTGKENAKYLI
EKIKLHDMQEGKCLYSLEAIPLEDLLNNPFNYEVDHIIIPRSVSFDNSFNNKVLVKQEENSCKGNR
TPFQYLSSSDSKISYETFKKHILNLAKGGRISKTKKEYLLEERDINRFSVQKDFINRNLVDTRY
ATRGLMNLRSYFRVNNLDVKVKSINGGFTSFLRRKWKFKKERNKGYKHAEDALIINANADFIK
EWKLDKAKKVMENQMFEEKQAESMPEIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNR
ELINDTLYSTRKDDKGNTLIVNNLNGLYDKDNDKLLKLINKSPEKLLMYHHPQTYQKLKLIMEQ
YGDEKNPLYKYEETGNYLTYSKKDNGPVIKKIKYYGNKLNALHDITDDYPNSRNKVVKLSLKP
YRFDVYLDNGVYKFVTVKNLDVICKENYEVNSKCYEEAKKLKISNQAEFIASFYNNDLIKING
ELYRVIGVNNDLLNRIEVMIDITYREYLENMNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEV
KSKKHPQIIKKG [SEQ ID NO: 45]

[0143]

[0144] 대안적으로 또는 부가적으로, CRISPR/Cas9-기반 유전자 편집 시스템은 융합 단백질을 포함할 수 있다. 융합 단백질은 2개의 이중 폴리펩타이드 도메인을 포함할 수 있으며, 제1 폴리펩타이드 도메인은 Cas 단백질을 포함하고, 제2 폴리펩타이드 도메인은 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변경 활성, 뉴클레아제 활성, 핵산 연관 활성, 메틸라제 활성, 또는 탈메틸라제 활성과 같은 활성을 가진다. 융합 단백질은 전사 활성화 활성, 전사 억제 활성, 전사 방출 인자 활성, 히스톤 변경 활성, 뉴클레아제 활성, 핵산 연관 활성, 메틸라제 활성, 또는 탈메틸라제 활성과 같은 활성을 갖는 제2 폴리펩타이드 도메인에 융합된 Cas9 단백질 또는 돌연변이된 Cas9 단백질을 포함할 수 있다.

[0145] (a) 전사 활성화 활성

[0146] 제2 폴리펩타이드 도메인은 전사 활성화 활성, 즉 트랜스활성화 도메인을 가질 수 있다. 예를 들어, 인간 유전자와 같은 내인성 포유류 유전자의 유전자 발현은 iCas9와 트랜스활성화 도메인의 융합 단백질을 gRNA 조합을 통해 포유동물의 프로모터에 표적화함으로써 달성될 수 있다. 트랜스활성화 도메인은 VP16 단백질, 다수의 VP16 단백질, 예컨대 VP48 도메인 또는 VP64 도메인 또는 NFκB 전사 활성화제 활성의 p65 도메인을 포함할 수 있다.

예를 들어, 융합 단백질은 iCas9-VP64일 수 있다.

[0147] (b) 전사 억제 활성

[0148] 제2 폴리펩타이드 도메인은 전사 억제 활성을 가질 수 있다. 제2 폴리펩타이드 도메인은 KRAB 도메인, ERF 억제 인자 도메인 활성, Mxi1 억제인자 도메인 활성, SID4X 억제인자 도메인 활성, Mad-SID 억제인자 도메인 활성 또는 TATA 박스 결합 단백질 활성과 같은 Kruppel 관련 박스 활성을 가질 수 있다. 예를 들어, 융합 단백질은 dCas9-KRAB일 수 있다.

[0149] (c) 전사 방출 인자 활성

[0150] 제2 폴리펩타이드 도메인은 전사 방출 인자 활성을 가질 수 있다. 제2 폴리펩타이드 도메인은 진핵세포 방출 인자 1(ERF1) 활성 또는 진핵세포 방출 인자 3(ERF3) 활성을 가질 수 있다.

[0151] (d) 히스톤 변경 활성

[0152] 제2 폴리펩타이드 도메인은 히스톤 변경 활성을 가질 수 있다. 제2 폴리펩타이드 도메인은 히스톤 탈아세틸라제, 히스톤 아세틸트랜스퍼라제, 히스톤 탈메틸라제 또는 히스톤 메틸트랜스퍼라제 활성을 가질 수 있다. 히스톤 아세틸트랜스퍼라제는 p300 또는 CREB-결합 단백질(CBP) 단백질 또는 그의 단편일 수 있다. 예를 들어, 융합 단백질은 dCas9-p300일 수 있다.

[0153] (e) 뉴클레아제 활성

[0154] 제2 폴리펩타이드 도메인은 Cas9 단백질의 뉴클레아제 활성과 상이한 뉴클레아제 활성을 가질 수 있다. 뉴클레아제, 또는 뉴클레아제 활성을 갖는 단백질은 핵산의 뉴클레오타이드 서브유닛 간 포스포디에스테르 결합을 절단할 수 있는 효소이다. 뉴클레아제는 보통 엔도뉴클레아제 및 엑소뉴클레아제로 추가 구분되지만, 일부 효소는 두 범주에 모두 속할 수 있다. 널리 알려진 뉴클레아제는 데옥시리보뉴클레아제 및 리보뉴클레아제이다.

[0155] (f) 핵산 연관 활성

[0156] 제2 폴리펩타이드 도메인은 핵산 연관 활성을 가질 수 있거나, 또는 핵산 결합 단백질-DNA-결합 도메인(DBD)은 이중-가닥 또는 단일-가닥 DNA를 인식하는 적어도 하나의 모티프를 함유하는 독립적으로 접힌 단백질 도메인이다. DBD는 특정 DNA 서열(인식 서열)을 인식하거나, DNA에 대한 일반적인 친화도를 가질 수 있다. 핵산 연관 영역은 헬릭스-턴-헬릭스 영역, 류신 지퍼 영역, 날개형 헬릭스 영역, 날개형 헬릭스-턴-헬릭스 영역, 헬릭스-루프-헬릭스 영역, 면역글로불린 접합, B3 도메인, 아연 핑거, HMG-박스, Wor3 도메인, TAL 효과기 DNA-결합 도메인으로 구성된 그룹으로부터 선택된다.

[0157] (g) 메틸라제 활성

[0158] 제2 폴리펩타이드 도메인은 메틸기를 DNA, RNA, 단백질, 소분자, 시토신 또는 아데닌으로 전달하는 메틸라제 활성을 가질 수 있다. 제2 폴리펩타이드 도메인은 DNA 메틸트랜스퍼라제를 포함할 수 있다.

[0159] (h) 탈메틸라제 활성

[0160] 제2 폴리펩타이드 도메인은 탈메틸라제 활성을 가질 수 있다. 제2 폴리펩타이드 도메인은 핵산, 단백질(특히 히스톤) 및 다른 분자로부터 메틸(CH₃-)기를 제거하는 효소를 포함할 수 있다. 대안적으로, 제2 폴리펩타이드는 DNA를 탈메틸화시키는 메카니즘에서 메틸기를 히드록시메틸시토신으로 전환시킬 수 있다. 제2 폴리펩타이드는 상기 반응을 촉매할 수 있다. 예를 들어, 상기 반응을 촉매하는 제2 폴리펩타이드는 Tet1일 수 있다.

[0161] (b) 디스트로핀 유전자를 표적화하는 gRNA

[0162] CRISPR/Cas9-기반 유전자 편집 시스템은 적어도 하나의 gRNA 분자, 예를 들어 2개의 gRNA 분자를 포함한다. gRNA는 CRISPR/Cas9-기반 유전자 편집 시스템의 표적화를 제공한다. gRNA는 2가지 비코딩 RNA: crRNA 및 tracrRNA의 융합이다. sgRNA는 목적하는 DNA 표적과의 상보적 염기쌍 형성을 통해 표적화 특이성을 부여하는 20 bp 프로토스페이서를 인코딩하는 서열의 교환에 의해 임의의 목적하는 DNA 서열을 표적화할 수 있다. gRNA는 II형 효과기 시스템에 관여되는 자연 발생 crRNA:tracrRNA 듀플렉스를 모사한다. 예를 들어, 42-뉴클레오타이드 crRNA 및 75-뉴클레오타이드 tracrRNA가 포함될 수 있는 상기 듀플렉스는 Cas9에 대한 가이드로 작용하여 표적 핵산을 절단한다. 본원에서 상호 교환적으로 이용되는 "표적 영역", "표적 서열", 또는 "프로토스페이서"는 CRISPR/Cas9-기반 유전자 편집 시스템이 표적화하는 표적 유전자(예를 들어, 디스트로핀 유전자)의 영역을 나타낸다. CRISPR/Cas9-기반 유전자 편집 시스템에는 적어도 하나의 gRNA가 포함될 수 있고, 여기서 gRNA는 상이한

DNA 서열을 표적화한다. 표적 DNA 서열은 중첩될 수 있다. 표적 서열 또는 프로토스페이서에는 프로토스페이서의 3' 말단에 PAM 서열이 뒤따른다. 상이한 II형 시스템은 상이한 PAM 요건을 갖는다. 예를 들어, 스트렙토코커스 피오제네스(*Streptococcus pyogenes*) II형 시스템은 "NGG" 서열을 이용하며, 여기서 "N"은 임의의 뉴클레오타이드일 수 있다. 일부 구현예에서, PAM 서열은 "NGG"일 수 있으며, 여기서 "N"은 임의의 뉴클레오타이드일 수 있다. 일부 구현예에서, PAM 서열은 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)일 수 있다.

[0163] 본원에 개시된 유전자 구축물(예를 들어, AAV 벡터)에 의해 코딩되는 gRNA 분자의 수는 적어도 1개의 gRNA, 적어도 2개의 상이한 gRNA, 적어도 3개의 상이한 gRNA, 적어도 4개의 상이한 gRNA, 적어도 5개의 상이한 gRNA, 적어도 6개의 상이한 gRNA, 적어도 7개의 상이한 gRNA, 적어도 8개의 상이한 gRNA, 적어도 9개의 상이한 gRNA, 적어도 10개의 상이한 gRNAs, 적어도 11개의 상이한 gRNAs, 적어도 12개의 상이한 gRNAs, 적어도 13개의 상이한 gRNAs, 적어도 14개의 상이한 gRNAs, 적어도 15개의 상이한 gRNAs, 적어도 16개의 상이한 gRNAs, 적어도 17개의 상이한 gRNAs, 적어도 18개의 상이한 gRNAs, 적어도 18개의 상이한 gRNAs, 적어도 20개의 상이한 gRNAs, 적어도 25개의 상이한 gRNAs, 적어도 30개의 상이한 gRNAs, 적어도 35개의 상이한 gRNAs, 적어도 40개의 상이한 gRNAs, 적어도 45개의 상이한 gRNAs, 또는 적어도 50개의 상이한 gRNAs일 수 있다. 본원에 개시된 벡터에 의해 코딩되는 gRNA의 수는 적어도 1개의 gRNA 내지 적어도 50개의 상이한 gRNA, 적어도 1개의 gRNA 내지 적어도 45개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 40개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 35개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 30개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 25개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 20개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 16개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 12개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 8개의 상이한 gRNAs, 적어도 1개의 gRNA 내지 적어도 4개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 50개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 45개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 40개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 35개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 30개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 25개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 20개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 16개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 12개의 상이한 gRNAs, 적어도 4개의 상이한 gRNAs 내지 적어도 8개의 상이한 gRNAs, 적어도 8개의 상이한 gRNAs 내지 적어도 50개의 상이한 gRNAs, 적어도 8개의 상이한 gRNAs 내지 적어도 45개의 상이한 gRNAs, 적어도 8개의 상이한 gRNAs 내지 적어도 40개의 상이한 gRNAs, 적어도 8개의 상이한 gRNAs 내지 적어도 35개의 상이한 gRNAs, 8개의 상이한 gRNAs 내지 적어도 30개의 상이한 gRNAs, 적어도 8개의 상이한 gRNAs 내지 적어도 25개의 상이한 gRNAs, 8개의 상이한 gRNAs 내지 적어도 20개의 상이한 gRNAs, 적어도 8개의 상이한 gRNAs 내지 적어도 16개의 상이한 gRNAs, 또는 8개의 상이한 gRNAs 내지 적어도 12개의 상이한 gRNAs일 수 있다. 특정 구현예에서, 유전자 구축물(예를 들어, AAV 벡터)은 하나의 gRNA 분자, 즉 제1 gRNA 분자 및 선택적으로 Cas9 분자를 코딩한다. 특정 구현예에서, 제1 유전자 구축물(예를 들어, 제1 AAV 벡터)은 1개의 gRNA 분자, 즉 제1 gRNA 분자 및 선택적으로 Cas9 분자를 코딩하고, 제2 유전자 구축물(예를 들어, 제2 AAV 벡터)은 1개의 gRNA 분자, 즉, 제2 gRNA 분자 및 선택적으로 Cas9 분자를 코딩한다.

[0164] gRNA 분자는 표적 DNA 서열의 상보적 폴리뉴클레오타이드 서열에 이어 PAM 서열인 표적화 도메인을 포함한다. gRNA는 표적화 도메인 또는 상보적 폴리뉴클레오타이드 서열의 5' 말단에 "G"를 포함할 수 있다. gRNA 분자의 표적화 도메인은 적어도 10 염기쌍, 적어도 11 염기쌍, 적어도 12 염기쌍, 적어도 13 염기쌍, 적어도 14 염기쌍, 적어도 15 염기쌍, 적어도 16 염기쌍, 적어도 17 염기쌍, 적어도 18 염기쌍, 적어도 19 염기쌍, 적어도 20 염기쌍, 적어도 21 염기쌍, 적어도 22 염기쌍, 적어도 23 염기쌍, 적어도 24 염기쌍, 적어도 25 염기쌍, 적어도 30 염기쌍, 또는 적어도 35 염기쌍의 표적 DNA 서열의 상보적 폴리뉴클레오타이드 서열에 이어 PAM 서열을 포함할 수 있다. 특정 구현예에서, gRNA 분자의 표적화 도메인은 19 내지 25개 뉴클레오타이드 길이를 갖는다. 특정 구현예에서, gRNA 분자의 표적화 도메인은 20개 뉴클레오타이드 길이이다. 특정 구현예에서, gRNA 분자의 표적화 도메인은 21개 뉴클레오타이드 길이이다. 특정 구현예에서, gRNA 분자의 표적화 도메인은 22개 뉴클레오타이드 길이이다. 특정 구현예에서, gRNA 분자의 표적화 도메인은 23개 뉴클레오타이드 길이이다.

[0165] gRNA는 디스트로핀 유전자(DMD)의 영역을 표적화할 수 있다. 특정 구현예에서, gRNA는 엑손, 인트론, 프로모터 영역, 인핸서 영역 또는 디스트로핀 유전자의 의 전사되는 영역 중 적어도 하나를 표적화할 수 있다. 특정 구현예에서, gRNA 분자는 인간 디스트로핀 유전자의 인트론 50을 표적화한다. 특정 구현예에서, gRNA 분자는 인간 디스트로핀 유전자의 인트론 51을 표적화한다. 특정 구현예에서, gRNA 분자는 인간 디스트로핀 유전자의 엑손 51을 표적화한다. gRNA는 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19에 개

시된 뉴클레오타이드 서열 또는 그의 상보체를 포함하는 표적화 도메인을 포함할 수 있다.

[0166] 단일 또는 다중화된 gRNA는 엑손 51에서 돌연변이 핫스팟을 표적화함으로써, 또는/및 엑손 51의 엑손내 작은 삽입 및 결실, 또는 절제를 도입함으로써 디스트로핀 해독 틀을 복원하도록 설계될 수 있다. 본원에 개시된 벡터로 처리한 후, *시험관내에서* 뒤시엔느 환자 근육세포내에서 디스트로핀 발현이 복원될 수 있다. 유전적으로 교정된 환자 세포를 면역결핍 마우스에 이식한 후 *생체내에서* 인간의 디스트로핀을 검출했다. 중요하게는, CRISPR/Cas9-기반 유전자 편집 시스템의 독특한 다중 유전자 편집 기능은 보편적 또는 환자-특이적 유전자 편집 접근법에 의해 환자 돌연변이의 62%까지 교정할 수 있는 상기 돌연변이 핫스팟 영역의 큰 결실을 효율적으로 생성할 수 있게 한다. 일부 구현예에서, 후보 gRNA는 표적-이외 활성화, 측량자에 의해 측정된 표적-상의 활성화 및 엑손으로부터의 거리에 기초하여 평가되고 선택된다.

[0167] 3. DNA 표적화 조성물

[0168] 본 발명은 또한 상기 유전적 구축물을 포함하는 DNA 표적화 조성물에 관한 것이다. DNA 표적화 조성물은 상기 기술된 바와 같이 *디스트로핀* 유전자(예를 들어, 인간 *디스트로핀* 유전자)를 표적화하는 적어도 하나의 gRNA 분자(예를 들어, 2개의 gRNA 분자)를 포함한다. 적어도 하나의 gRNA 분자는 표적 영역을 결합하고 인식할 수 있다. 표적 영역은 가능한 프레임-변경 중지 코돈의 바로 상류에서 선택되어, 보수 과정 중 삽입 또는 결실이 디스트로핀 해독 틀을 틀 전환으로 복원시킨다. 표적 영역은 또한 스플라이스 수신체 부위 또는 스플라이스 공여체 부위일 수 있어서, 보수 과정 중 삽입 또는 결실이 스플라이스 부위를 손상하고 스플라이스 부위의 손상 및 엑손 배제에 의해 디스트로핀 해독 틀을 복원시킬 수 있다. 표적 영역은 또한 비정상적 종료 코돈이 될 수 있어서, 보수 과정동안의 삽입 또는 결실이 정지 코돈을 제거하거나 손상시킴으로써 디스트로핀 해독 틀을 복원시킨다.

[0169] 특정 구현예에서, 본 명세서에 개시된 DNA 표적화 조성물은 제1 gRNA 및 제2 gRNA를 포함하며, 상기 제1 gRNA 분자 및 제2 gRNA 분자는 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19에 개시된 뉴클레오타이드 서열 또는 그의 상보체를 포함하는 표적화 도메인을 포함한다. 특정 구현예에서, 제1 gRNA 분자 및 제2 gRNA 분자는 상이한 표적화 도메인을 포함한다. 특정 구현예에서, 제1 gRNA 분자는 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, 또는 SEQ ID NO: 15이고, 제2 gRNA 분자는 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, 또는 SEQ ID NO: 19이다. 특정 구현예에서, 상기 제1 gRNA 분자는 SEQ ID NO: 1, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14 및 SEQ ID NO: 15로 구성된 그룹으로부터 선택되며, 상기 제2 gRNA 분자는 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, 및 SEQ ID NO: 19로 구성된 그룹으로부터 선택된다.

[0170] 특정 구현예에서, 제1 gRNA 분자 및 제2 gRNA 분자는 (i) SEQ ID NO: 1에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 2에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ii) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iii) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iv) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (v) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vi) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (viii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ix) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (x) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인

을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (xi) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및 (xii) SEQ ID NO: 41에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 42에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자로 구성된 그룹으로부터 선택된다. 일부 구현예에서, DNA 표적화 조성물은 SEQ ID NO: 37에 개시된 뉴클레오타이드 서열 및/또는 SEQ ID NO: 38에 개시된 뉴클레오타이드 서열을 포함한다.

[0171] 특정 구현예에서, DNA 표적화 조성물은 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25) 중 하나의 PAM을 인식하는 적어도 하나의 Cas9 분자 또는 Cas9 융합 단백질을 추가로 포함할 수 있다. 일부 구현예에서, DNA 표적화 조성물은 SEQ ID NO: 83 또는 SEQ ID NO: 84에 개시된 뉴클레오타이드 서열을 포함한다. 특정 구현예에서, 상기 벡터는 인간 *디스트로핀* 유전자의 엑손 51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 *디스트로핀* 유전자의 절편을 결실시키도록 구성된다.

[0172] 본원에 개시된 벡터의 결실 효율은 결실 크기, 즉 벡터에 의해 결실된 절편의 크기와 관련될 수 있다. 특정 구현예에서, 특정 결실의 길이 또는 크기는 표적화될 유전자(예를 들어, *디스트로핀* 유전자) 내의 PAM 서열 간의 거리에 의해 결정된다. 특정 구현예에서, 그의 길이 및 그것이 포함하는 서열(예를 들어, 엑손 51)에 관하여 정의된 *디스트로핀* 유전자의 절편의 특이적 결실은 표적 유전자(예를 들어, *디스트로핀* 유전자) 내 특정 PAM 서열에 인접하여 만들어진 절단부의 결과이다.

[0173] 특정 구현예에서, 결실 크기는 약 50 내지 약 2,000 염기 쌍(bp), 예를 들어, 약 50 내지 약 1999 bp, 약 50 내지 약 1900bp, 약 50 내지 약 1800 bp, 약 50 내지 약 1700 bp, 약 50 내지 약 1650 bp, 약 50 내지 약 1600 bp, 약 50 내지 약 1500 bp, 약 50 내지 약 1400 bp, 약 50 내지 약 1300 bp, 약 50 내지 약 1200 bp, 약 50 내지 약 1150 bp, 약 50 내지 약 1100 bp, 약 50 내지 약 1000 bp, 약 50 내지 약 900 bp, 약 50 내지 약 850 bp, 약 50 내지 약 800 bp, 약 50 내지 약 750 bp, 약 50 내지 약 700 bp, 약 50 내지 약 600 bp, 약 50 내지 약 500 bp, 약 50 내지 약 400 bp, 약 50 내지 약 350 bp, 약 50 내지 약 300 bp, 약 50 내지 약 250 bp, 약 50 내지 약 200 bp, 약 50 내지 약 150 bp, 약 50 내지 약 100 bp, 약 100 내지 약 1999 bp, 약 100 내지 약 1900 bp, 약 100 내지 약 1800 bp, 약 100 내지 약 1700 bp, 약 100 내지 약 1650 bp, 약 100 내지 약 1600 bp, 약 100 내지 약 1500 bp, 약 100 내지 약 1400 bp, 약 100 내지 약 1300 bp, 약 100 내지 약 1200 bp, 약 100 내지 약 1150 bp, 약 100 내지 약 1100 bp, 약 100 내지 약 1000 bp, 약 100 내지 약 900 bp, 약 100 내지 약 850 bp, 약 100 내지 약 800 bp, 약 100 내지 약 750 bp, 약 100 내지 약 700 bp, 약 100 내지 약 600 bp, 약 100 내지 약 500 bp, 약 100 내지 약 400 bp, 약 100 내지 약 350 bp, 약 100 내지 약 300 bp, 약 100 내지 약 250 bp, 약 100 내지 약 200 bp, 약 100 내지 약 150 bp, 약 200 내지 약 1999 bp, 약 200 내지 약 1900 bp, 약 200 내지 약 1800 bp, 약 200 내지 약 1700 bp, 약 200 내지 약 1650 bp, 약 200 내지 약 1600 bp, 약 200 내지 약 1500 bp, 약 200 내지 약 1400 bp, 약 200 내지 약 1300 bp, 약 200 내지 약 1200 bp, 약 200 내지 약 1150 bp, 약 200 내지 약 1100 bp, 약 200 내지 약 1000 bp, 약 200 내지 약 900 bp, 약 200 내지 약 850 bp, 약 200 내지 약 800 bp, 약 200 내지 약 750 bp, 약 200 내지 약 700 bp, 약 200 내지 약 600 bp, 약 200 내지 약 500 bp, 약 200 내지 약 400 bp, 약 200 내지 약 350 bp, 약 200 내지 약 300 bp, 약 200 내지 약 250 bp, 약 300 내지 약 1999 bp, 약 300 내지 약 1900 bp, 약 300 내지 약 1800 bp, 약 300 내지 약 1700 bp, 약 300 내지 약 1650 bp, 약 300 내지 약 1600 bp, 약 300 내지 약 1500 bp, 약 300 내지 약 1400 bp, 약 300 내지 약 1300 bp, 약 300 내지 약 1200 bp, 약 300 내지 약 1150 bp, 약 300 내지 약 1100 bp, 약 300 내지 약 1000 bp, 약 300 내지 약 900 bp, 약 300 내지 약 850 bp, 약 300 내지 약 800 bp, 약 300 내지 약 750 bp, 약 300 내지 약 700 bp, 약 300 내지 약 600 bp, 약 300 내지 약 500 bp, 약 300 내지 약 400 bp, 또는 약 300 내지 약 350 bp이다. 특정 구현예에서, 결실 크기는 약 118 염기쌍, 약 233 염기쌍, 약 326 염기쌍, 약 766 염기쌍, 약 805 염기쌍 또는 약 1611 염기쌍일 수 있다.

[0174] 4. 근육내 게놈 편집을 위한 조성물

[0175] 본 발명은 대상체의 골격근 또는 심장 근육에서 표적 유전자를 게놈 편집하기 위한 유전자 구축물(예를 들어, 벡터) 또는 그의 조성물에 관한 것이다. 상기 조성물은 변경된 AAV 벡터 및 CRISPR/Cas9-기반 유전자 편집 시스템, 예를 들어, gRNA 분자 및 Cas9 분자를 코딩하는 뉴클레오타이드 서열을 포함한다. 조성물은 CRISPR/Cas9-기반 유전자 편집 시스템의 활성 형태를 골격근 또는 심장 근육에 전달한다. 본원에 개시된 유전자 구축물(예를 들어, 벡터)은 유전 질환 및/또는 다른 골격 또는 심근 질환, 예를 들어 DMD에 연관된 *디스트로핀* 유전자에 있어서 돌연변이의 영향을 교정하거나 감소시키는데 사용될 수 있다. 상기 조성물은 공여체 DNA 또는 트랜스유전

자를 추가로 포함할 수 있다. 상기 조성물은 게놈 편집, 게놈 조작, 및 유전자 질환 및/또는 다른 골격 또는 심장 근육 질환과 연관된 유전자에 있어서 돌연변이의 영향을 교정하거나 감소시키는데 사용될 수 있다.

[0176] **a. 디스트로핀을 표적화하기 위한 CRISPR/Cas9-기반 유전자 편집 시스템**

[0177] 디스트로핀 유전자에 특이적인 CRISPR/Cas9-기반 유전자 편집 시스템이 본 명세서에 개시되어있다. CRISPR/Cas9-기반 유전자 편집 시스템은 Cas9 및 디스트로핀 유전자를 표적화하는 하나 이상의 gRNA를 포함할 수 있다. CRISPR/Cas9-기반 유전자 편집 시스템은 표적 영역을 결합하고 인식할 수 있다. 표적 영역은 가능한 프레임-변경 중지 코돈의 바로 상류에서 선택되어, 보수 과정 중 삽입 또는 결실이 디스트로핀 해독 틀을 틀 전환으로 복원시킨다. 표적 영역은 또한 스플라이스 수신체 부위 또는 스플라이스 공여체 부위일 수 있어서, 보수 과정 중 삽입 또는 결실이 스플라이싱을 손상시키고, 스플라이스 부위의 손상 및 엑손 배제에 의해 디스트로핀 해독 틀을 복원시킬 수 있다. 표적 영역은 또한 비정상적 중지 코돈이 될 수 있어서, 보수 과정동안의 삽입 또는 결실이 중지 코돈을 제거하거나 손상시킴으로써 디스트로핀 해독 틀을 복원시킨다.

[0178] gRNA는 SEQ ID NO: 1-19, 41, 42로 구성된 그룹으로부터 선택된 뉴클레오타이드 서열 또는 그의 상보체를 표적화할 수 있다. 예를 들어, 개시된 CRISPR/Cas9-기반 유전자 편집 시스템은 디스트로핀 유전자의 엑손 51에서 매우 효율적인 유전자 편집을 중재하도록 조작되었다. 상기 CRISPR/Cas9-기반 유전자 편집 시스템은 DMD 환자의 세포에서 디스트로핀 단백질 발현을 복원시켰다. 일부 구현예에서, DNA 표적화 조성물은 SEQ ID NO: 37에 개시된 뉴클레오타이드 서열, SEQ ID NO: 38에 개시된 뉴클레오타이드 서열, SEQ ID NO: 83에 개시된 뉴클레오타이드 서열 및/또는 SEQ ID NO: 84에 개시된 뉴클레오타이드 서열을 포함한다. 예를 들어, DNA 표적화 조성물은 SEQ ID NO: 37에 개시된 뉴클레오타이드 서열, SEQ ID NO: 38에 개시된 뉴클레오타이드 서열 및 SEQ ID NO: 83에 개시된 뉴클레오타이드 서열을 포함하거나, 또는 DNA 표적화 조성물은 SEQ ID NO: 37에 개시된 뉴클레오타이드 서열, SEQ ID NO: 38에 개시된 뉴클레오타이드 서열 및 SEQ ID NO: 84에 개시된 뉴클레오타이드 서열을 포함한다.

[0179] **b. 아데노-연관 바이러스 벡터**

[0180] 상기 조성물은 또한 바이러스 전달 시스템을 포함할 수 있다. 특정 구현예에서, 벡터는 아데노-연관 바이러스(AAV) 벡터이다. AAV 벡터는 인간 및 일부 다른 영장류 종을 감염시키는 파보비리데(Parvoviridae)과의 데펜도 바이러스(Dependovirus)속에 속하는 소형 바이러스이다. AAV 벡터는 다양한 구축물 구성을 사용하여 CRISPR/Cas9-기반 유전자 편집 시스템을 전달하는데 사용될 수 있다. 예를 들어, AAV 벡터는 Cas9 및 gRNA 발현 카세트를 별도의 벡터 또는 동일한 벡터 상에 전달할 수 있다. 대안적으로, *스타필로코커스 아우레우스* (*Staphylococcus aureus*) 또는 *나이세리아 메닝지티디스* (*Neisseria meningitidis*)와 같은 종에서 유래되는 소형 Cas9 단백질이 이용되는 경우, Cas9 및 최대 2개의 gRNA 발현 카세트는 둘 다 4.7 kb 패키징 한계 내에서 단일 AAV 벡터에 조합될 수 있다.

[0181] 특정 구현예에서, AAV 벡터는 변경된 AAV 벡터이다. 변경된 AAV 벡터는 심장 및 골격 근육 조직 방향성을 향상시킬 수 있다. 변경된 AAV 벡터는 포유류 세포에서 CRISPR/Cas9-기반 유전자 편집 시스템을 전달하고 발현할 수 있다. 예를 들어, 변경된 AAV 벡터는 AAV-SASTG 벡터(Piacentino *et al.* (2012) Human Gene Therapy 23:635-646)일 수 있다. *생체내*에서 골격 및 심장 근육에 뉴클레아제를 전달할 수 있다. 변경된 AAV 벡터는 AAV1, AAV2, AAV5, AAV6, AAV8, 및 AAV9를 포함하는 하나 이상의 몇몇 캡시드 유형에 기반할 수 있다. 변경된 AAV 벡터는 대안적인 근육-친화성 AAV 캡시드를 갖는 AAV2 슈도타입(pseudotype), 예컨대 전신 및 국소 전달에 의해 골격근 또는 심장근육을 효율적으로 형질도입시키는 AAV2/1, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV2.5 및 AAV/SASTG 벡터에 기반할 수 있다(Seto *et al.* Current Gene Therapy (2012) 12:139-151). 변경된 AAV 벡터는 AAV2i8G9일 수 있다(Shen *et al.* J. Biol. Chem. (2013) 288:28814-28823). 일부 구현예에서, 조성물은 SEQ ID NO: 39에 개시된 뉴클레오타이드 서열 및/또는 SEQ ID NO: 40에 개시된 뉴클레오타이드 서열을 포함한다. 일부 구현예에서, 조성물은 SEQ ID NO: 39에 개시된 뉴클레오타이드 서열을 포함하는 제1 벡터, 및 SEQ ID NO: 40에 개시된 뉴클레오타이드 서열을 포함하는 제2 벡터를 포함한다.

[0182] **5. 근육내 게놈 편집 방법**

[0183] 본 발명은 대상체의 골격근 또는 심장 근육내 게놈 편집 방법에 관한 것이다. 상기 방법은 상기한 바와 같이, 골격근 또는 심장 근육내 게놈 편집을 위한 조성물을 대상체의 골격근 또는 심장 근육에 투여하는 단계를 포함한다. 게놈 편집은 돌연변이 유전자를 교정하거나 트랜스유전자를 삽입하는 것을 포함할 수 있다. 돌연변이체 유전자를 교정하는 것은 돌연변이 유전자를 결실, 재배치 또는 대체하는 것을 포함할 수 있다. 돌연변이체 유전

자를 교정하는 것은 뉴클레아제-매개된 NHEJ 또는 HDR을 포함할 수 있다.

[0184] **6. 돌연변이체 유전자를 교정하고 대상체를 치료하는 방법**

[0185] 본원에 개시된 주제는 세포내 돌연변이체 유전자(예를 들어, 돌연변이체 *디스트로핀* 유전자, 예를 들어, 돌연변이체 인간 *디스트로핀* 유전자)를 교정하고, DMD와 같은 유전자 질환을 앓고있는 대상체를 치료하는 방법을 제공한다. 이 방법은 상기 기술된 바와 같이 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 그것을 포함하는 조성물을 세포 또는 대상체에 투여하는 단계를 포함할 수 있다. 이 방법은 상기 기술된 바와 같이, 골격근 또는 심장 근육에서 게놈 편집을 위해, 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 그를 포함하는 조성물을 대상체의 골격근 또는 심장 근육에 투여하는 단계를 포함할 수 있다. 골격근 또는 심장 근육에 CRISPR/Cas9-기반 유전자 편집 시스템을 전달하기 위해 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 이의 조성물을 사용하여 보수 템플릿 또는 공여체 DNA를 갖는 전체-기능 또는 부분-기능 단백질의 발현을 복원시킬 수 있으며, 이는 전체 유전자 또는 돌연변이를 포함하는 영역을 대체할 수 있다. CRISPR/Cas9-기반 유전자 편집 시스템은 표적화된 유전자위에서 부위-특이적 이중 가닥 절단을 도입하는데 사용될 수 있다. CRISPR/Cas9-기반 유전자 편집 시스템이 표적 DNA 서열에 결합하여 표적 DNA의 절단을 허용할 때 부위-특이적 이중-가닥 절단이 생성된다. 이 DNA 절단은 자연적 DNA-보수 기구를 자극하여, 상동성-유도 보수(homology-directed repair, HDR) 또는 비-상동성 말단 연결(non-homologous end joining, NHEJ) 경로의 2개의 가능한 보수 경로 중 하나로 이어질 수 있다.

[0186] 본 발명은 보수 주형 없이 CRISPR/Cas9-기반 유전자 편집 시스템을 이용하는 게놈 편집에 대한 것이며, 이는 해독 틀을 효율적으로 교정하고 유전 질환에 관여되는 기능성 단백질의 발현을 복원할 수 있다. 개시되는 CRISPR/Cas9-기반 유전자 편집 시스템은 상동성-유도 보수 또는 뉴클레아제-매개된 비-상동성 말단 연결(NHEJ)-기반 교정 접근의 이용을 포함할 수 있고, 이는 상동성 재조합 또는 선택-기반 유전자 교정에 적합하지 않을 수 있는 증식이 제한된 일차 세포주에서의 효율적 교정을 가능케 한다. 상기 전략은 활성 CRISPR/Cas9-기반 유전자 편집 시스템의 신속하고 강력한 어셈블리를 틀 이동, 조기 중지 코돈, 비정상 스플라이스 공여체 부위 또는 비정상 스플라이스 수신체 부위를 유도하는 비필수 코딩 영역에서의 돌연변이에 의해 유도되는 유전자 질환의 치료를 위한 효율적인 유전자 편집 방법과 통합한다.

[0187] **a. 뉴클레아제 매개된 비-상동성 말단 연결**

[0188] 내인성 돌연변이된 유전자로부터 단백질 발현의 복원은 주형-비함유 NHEJ-매개된 DNA 보수를 통한 수 있다. 표적 유전자 RNA를 표적화하는 일시적 방법과 대조적으로, 일시적으로 발현되는 CRISPR/Cas9-기반 유전자 편집 시스템에 의한 게놈 내 표적 유전자의 해독 틀의 교정은 각각의 변경된 세포 및 모든 그의 자손에 의해 영구적으로 복원된 표적 유전자 발현으로 이어질 수 있다. 특정 구현예에서, NHEJ는 뉴클레아제 매개된 NHEJ이며, 특정 구현예에서, Cas9 분자에서 개시되는 NHEJ를 지칭하고, 이중 가닥 DNA를 절단한다. 상기 방법은 골격근 또는 심장 근육내 게놈 편집을 위해 대상체의 골격근 또는 심장 근육에 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 그를 포함하는 조성물을 투여하는 단계를 포함한다.

[0189] 뉴클레아제-매개된 NHEJ 유전자 교정은 돌연변이된 표적 유전자를 교정하고 HDR 경로에 비해 몇몇 잠재적인 장점을 제공할 수 있다. 예를 들어, NHEJ는 공여체 주형을 필요로 하지 않으며, 이는 비특이적인 삽입 돌연변이화를 유도할 수 있다. HDR과 대조적으로, NHEJ는 세포 주기의 모든 단계에서 효율적으로 작동하므로 분열 주기에 있는 세포 및 유사 분열-후 세포, 예컨대 근육 섬유 모두에서 효과적으로 활용될 수 있다. 이는 올리고뉴클레오타이드-기반 엑손 건너뛰기 또는 중지 코돈의 약리적으로 강제되는 해독-통과에 대한 강력하고 영구적인 유전자 복원 대안을 제공하며, 이론적으로 1회 만큼 적은 약물 처리를 필요로 할 수 있다. CRISPR/Cas9-기반 유전자 편집 시스템뿐만 아니라 메가뉴클레아제 및 아연 핑거 뉴클레아제를 포함하는 다른 조작된 뉴클레아제를 이용하는 NHEJ-기반 유전자 교정은 본원에 기술되는 플라스미드 전기전공 접근에 부가하여, 세포-기반 및 유전자-기반 치료법을 위한 다른 기존 *생체외* 및 *생체내* 플랫폼과 조합될 수 있다. 예를 들어, mRNA-기반 유전자 전달에 의한 또는 정제된 세포 투과성 단백질로서의 CRISPR/Cas9-기반 유전자 편집 시스템의 전달은 임의의 삽입 돌연변이화 가능성을 회피할 DNA-비함유 게놈 편집 접근을 가능케할 수 있다.

[0190] **b. 상동성-유도 보수**

[0191] 내인성 돌연변이된 유전자로부터 단백질 발현의 복원에는 상동성-유도 보수가 관여될 수 있다. 상술된 바와 같은 방법에는 세포에 공여체 주형을 투여하는 단계가 추가로 포함된다. 공여체 주형에는 전체-기능적 단백질 또는 부분-기능적 단백질을 인코딩하는 뉴클레오타이드 서열이 포함될 수 있다. 예를 들어, 공여체 주형에는 미니

디스트로핀("minidys")으로 명명된 소형화된 디스트로핀 구축물, 돌연변이체 디스트로핀 유전자의 복원을 위한 전체-기능적 디스트로핀 구축물 또는 상동성-유도 보수 후 돌연변이체 디스트로핀 유전자 복원으로 이어지는 디스트로핀 유전자의 단편이 포함될 수 있다.

[0192] **c. CRISPR/Cas9를 사용하여 돌연변이체 유전자를 교정하고 대상체를 치료하는 방법**

[0193] 본 개시는 또한 보수 주형 또는 공여체 DNA로 전체-기능적 또는 부분-기능적 단백질의 발현을 복원하기 위한 상술되는 CRISPR/Cas9-기반 유전자 편집 시스템을 이용한 게놈 편집에 대한 것이며, 이는 전체 유전자 또는 돌연변이를 함유하는 영역을 대체할 수 있다. CRISPR/Cas9-기반 유전자 편집 시스템은 표적화된 게놈 유전자위에서 부위-특이적 이중-가닥 절단을 도입하기 위해 이용될 수 있다. 부위-특이적 이중 가닥 절단은 CRISPR/Cas9-기반 유전자 편집 시스템이 gRNA를 이용하여 표적 DNA에 결합하는 경우 생성되며, 이에 의해 표적 DNA의 절단을 허용한다. CRISPR/Cas9-기반 유전자 편집 시스템은 이의 높은 성공적이고 효율적인 유전자 변경 비율로 인해 진보된 게놈 편집의 장점을 갖는다. 상기 DNA 절단은 자연적 DNA-보수 기구를 자극하여 2가지 가능한 보수 경로: 상동성-유도 보수(HDR) 또는 비-상동성 말단 연결(NHEJ) 경로 중 하나로 이어질 수 있다. 예를 들어, 디스트로핀 유전자를 향한 CRISPR/Cas9-기반 유전자 편집 시스템은 SEQ ID NO: 1 내지 19, 41, 42 중 어느 하나의 핵산 서열 또는 그의 상보체를 갖는 gRNA를 포함할 수 있다.

[0194] 본 개시는 보수 주형 없이 CRISPR/Cas9-기반 유전자 편집 시스템을 이용하는 게놈 편집에 대한 것이며, 이는 해독 틀을 효율적으로 교정하고 유전 질환에 관여되는 기능적 단백질의 발현을 복원할 수 있다. 개시되는 CRISPR/Cas9-기반 유전자 편집 시스템 및 방법은 상동성-유도 보수 또는 뉴클레아제-매개 비-상동성 말단 연결(NHEJ)-기반 교정 접근의 이용이 관여될 수 있고, 이는 상동성 제조합 또는 선택-기반 유전자 교정에 적합하지 않을 수 있는 증식이 제한된 일차 세포주에서의 효율적 교정을 가능케 한다. 상기 전략은 활성 CRISPR/Cas9-기반 유전자 편집 시스템의 신속하고 강력한 어셈블리를 틀 이동, 조기 중지 코돈, 비정상 스플라이스 공여체 부위 또는 비정상 스플라이스 수신체 부위를 유도하는 비필수 코딩 영역에서의 돌연변이에 의해 유도되는 유전 질환의 치료를 위한 효율적인 유전자 편집 방법과 통합한다.

[0195] 본 개시는 세포에서 돌연변이체 유전자의 교정 및 DMD와 같은 유전 질환을 겪는 대상체의 치료 방법을 제공한다. 방법에는 세포 또는 대상체에 상술된 바와 같은 CRISPR/Cas9-기반 유전자 편집 시스템, 상기 CRISPR/Cas9-기반 유전자 편집 시스템을 인코딩하는 폴리뉴클레오타이드 또는 벡터, 또는 상기 CRISPR/Cas9-기반 유전자 편집 시스템의 구성물을 투여하는 단계가 포함될 수 있다. 방법에는 CRISPR/Cas9-기반 유전자 편집 시스템을 투여하는 단계, 예컨대 뉴클레아제 활성을 갖는 제2 도메인을 함유하는 Cas9 단백질 또는 Cas9 융합 단백질, 상기 Cas9 단백질 또는 Cas9 융합 단백질을 인코딩하는 뉴클레오타이드 서열, 및/또는 적어도 하나의 gRNA를 투여하는 단계가 포함될 수 있고, 여기서 gRNA는 상이한 DNA 서열을 표적화한다. 표적 DNA 서열은 중첩될 수 있다. 세포에 투여되는 gRNA의 수는 상술된 바와 같은 적어도 1개의 gRNA, 적어도 2개의 상이한 gRNA, 적어도 3개의 상이한 gRNA, 적어도 4개의 상이한 gRNA, 적어도 5개의 상이한 gRNA, 적어도 6개의 상이한 gRNA, 적어도 7개의 상이한 gRNA, 적어도 8개의 상이한 gRNA, 적어도 9개의 상이한 gRNA, 적어도 10개의 상이한 gRNA, 적어도 15개의 상이한 gRNA, 적어도 20개의 상이한 gRNA, 적어도 30개의 상이한 gRNA, 또는 적어도 50개의 상이한 gRNA일 수 있다. gRNA에는 SEQ ID NO: 1~19, 41, 42 중 적어도 하나의 핵산 서열 또는 그의 상보체가 포함될 수 있다. 방법에는 상동성-유도 보수 또는 비-상동성 말단 연결이 관여될 수 있다.

[0196] **7. 질병 치료 방법**

[0197] 본 발명은 이를 필요로 하는 대상체를 치료하는 방법에 관한 것이다. 상기 방법은 상기 기술된 바와 같이, 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 그것을 포함하는 구성물을 대상체의 조직에 투여하는 단계를 포함한다. 특정 구현예에서, 이 방법은 상기 기술된 바와 같이, 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 그를 포함하는 구성물을 대상체의 골격근 또는 심장 근육에 투여하는 단계를 포함할 수 있다. 특정 구현예에서, 이 방법은 상기 기술된 바와 같이, 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 이의 구성물을 상기 대상체의 정맥에 투여하는 단계를 포함할 수 있다. 특정 구현예에서, 대상체는 골격근 또는 심장 근육 질환의 퇴행 또는 악화 또는 유전자 질환을 앓고 있다. 예를 들어, 대상체는 상기 기술된 바와 같이, 뉘시엔느 근이영양증을 앓고 있을 수 있다.

[0198] **a. 뉘시엔느 근이영양증**

[0199] 상기 기술된 바와 같이, 상기 방법은 디스트로핀 유전자를 교정하고 상기 돌연변이된 디스트로핀 유전자의 전체-기능적 단백질 또는 부분-기능적 단백질 발현을 회복시키기 위해 사용될 수 있다. 일부 양태 및 구현예에서,

본원은 환자에서 DMD의 효과(예를 들어, 임상 증상/징후)를 감소시키는 방법을 제공한다. 일부 양태 및 구현예에서, 본원은 환자에서 DMD를 치료하는 방법을 제공한다. 일부 양태 및 구현예에서, 본원은 환자에서 DMD를 예방하는 방법을 제공한다. 일부 양태 및 구현예에서, 본원은 환자에서 DMD의 추가 진행을 예방하는 방법을 제공한다.

[0200] 8. Δ52 hDMD를 갖는 트랜스제닉 설치류를 생성하는 방법

[0201] 본 발명은 엑손 52 결실을 갖는 인간 *디스트로핀* 유전자를 갖는 트랜스제닉 설치류 배아를 생성하는 방법에 관한 것이다. 이 방법은 gRNA를 설치류 배아에 투여하여 인간 *디스트로핀* 유전자의 엑손 52를 결실시키는 단계, 설치류 배아가 정상 인간 *디스트로핀* 유전자를 포함하는 인간 *디스트로핀* 유전자의 엑손 52를 결실시킨 트랜스제닉 설치류 배아를 선별하는 단계를 포함한다. 일부 구현예에서, 설치류 배아는 마우스 배아이다. 일부 구현예에서, 트랜스제닉 설치류 배아는 이형접합체 hDMD 또는 이형접합체 hDMD-Δ52이다. 일부 구현예에서, SEQ ID NO: 41에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자 및 SEQ ID NO: 42에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자가 설치류 배아에 투여되어, 인간 *디스트로핀* 유전자의 엑손 52를 결실시킨다. 일부 구현예에서, 상기 방법은 SEQ ID NO: 27에 개시된 아미노산 서열을 포함하는 Cas 단백질을 설치류 배아에 투여하는 단계를 추가로 포함한다. 본 발명은 이 방법에 의해 생성된 트랜스제닉 설치류 배아에 관한 것이다. 본 발명은 또한 트랜스제닉 설치류 배아로부터 생성된 트랜스제닉 설치류에 관한 것이다.

[0202] 9. 구축물 및 플라스미드

[0203] 상기 기술된 바와 같은 조성물은 본원에 개시된 바와 같이 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩하는 유전자 구축물을 포함할 수 있다. 플라스미드와 같은 유전자 구축물은 Cas9 단백질 및 Cas9 융합 단백질 및/또는 적어도 하나의 gRNA와 같은 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩하는 핵산을 포함할 수 있다. 상기 기술된 바와 같은 조성물은 본원에 개시된 바와 같이, 변형된 AAV 벡터를 코딩하는 유전자 구축물 및 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩하는 핵산 서열을 포함할 수 있다. 플라스미드와 같은 유전자 구축물은 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩하는 핵산을 포함할 수 있다. 상기 기술된 바와 같은 조성물은 본원에 개시된 바와 같이 변형된 렌티바이러스 벡터를 코딩하는 유전자 구축물을 포함할 수 있다.

[0204] 유전자 구축물, 예컨대 재조합 플라스미드 또는 재조합 바이러스 입자는 Cas9-융합 단백질 및 적어도 하나의 gRNA를 코딩하는 핵산을 포함할 수 있다. 일부 구현예에서, 유전자 구축물은 Cas9-융합 단백질 및 적어도 2개의 상이한 gRNA를 코딩하는 핵산을 포함할 수 있다. 일부 구현예에서, 유전자 구축물은 Cas9-융합 단백질 및 2개 이상의 상이한 gRNA를 코딩하는 핵산을 포함할 수 있다. 일부 구현예에서, 유전자 구축물은 적어도 하나의 gRNA 분자 및/또는 Cas9 분자를 코딩하는 뉴클레오타이드 서열에 작동가능하게 연결된 프로모터를 포함할 수 있다. 일부 구현예에서, 프로모터는 제1 gRNA 분자, 제2 gRNA 분자 및/또는 Cas9 분자를 코딩하는 뉴클레오타이드 서열에 작동가능하게 연결된다. 유전자 구축물은 기능하는 염색체의 분자로서 세포에 존재할 수 있다. 유전자 구축물은 동원체, 텔로미어 또는 플라스미드 또는 코스미드를 포함하는 선형 미니염색체일 수 있다.

[0205] 유전자 구축물은 또한 재조합 렌티바이러스, 재조합 아데노바이러스, 및 재조합 아데노바이러스 연관 바이러스를 포함하는 재조합 바이러스 벡터 계통의 일부일 수 있다. 유전자 구축물은 세포에서 사는 약독화된 살아있는 미생물 또는 재조합 미생물 벡터 내 유전 물질의 일부일 수 있다. 유전자 구축물은 핵산의 코딩 서열의 유전자 발현을 위한 조절 요소를 포함할 수 있다. 조절 요소는 프로모터, 인핸서, 개시 코돈, 중지 코돈, 또는 폴리아데닐화 신호일 수 있다.

[0206] 특정 구현예에서, 유전자 구축물은 벡터이다. 벡터는 적어도 하나의 Cas9 분자 및 적어도 하나의 gRNA 분자를 코딩하는 아데노-연관 바이러스(AAV) 벡터일 수 있으며; 상기 벡터는 포유 동물의 세포에서 적어도 하나의 Cas9 분자 및 적어도 gRNA 분자를 발현할 수 있다. 벡터는 플라스미드일 수 있다. 이 벡터는 *생체내* 유전자 치료에 사용될 수 있다. 벡터는 재조합될 수 있다. 벡터는 융합 단백질, 예컨대 Cas9-융합 단백질 또는 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩하는 이중 핵산을 포함할 수 있다. 벡터는 플라스미드일 수 있다. 상기 벡터는 Cas9-융합 단백질 또는 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩하는 핵산으로 세포를 형질감염시키는데 유용할 수 있으며, 상기 형질전환된 숙주 세포는 배양되고 Cas9-융합 단백질 또는 CRISPR/Cas9-기반의 유전자 편집 시스템의 발현이 일어나는 조건하에 유지된다.

[0207] 코딩 서열은 안정성 및 높은 수준의 발현을 위해 최적화될 수 있다. 일부 경우에, 코돈은 분자내 결합으로 인해 형성된 것과 같은 RNA의 2차 구조 형성을 감소시키도록 선택된다.

- [0208] 벡터는 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩하는 이중 핵산을 포함할 수 있으며, CRISPR/Cas9-기반 유전자 편집 시스템 코딩 서열의 상류일 수 있는 개시 코돈 및, CRISPR/Cas9-기반 유전자 편집 시스템 코딩 서열의 하류일 수 있는 종결 코돈을 추가로 포함할 수 있다. 개시 및 종결 코돈은 CRISPR/Cas9-기반 유전자 편집 시스템 코딩 서열과 같은 틀에 있을 수 있다. 벡터는 또한 CRISPR/Cas9-기반 유전자 편집 시스템 코딩 서열에 작동 가능하게 연결되는 프로모터를 포함할 수 있다. CRISPR/Cas9-기반 유전자 편집 시스템 코딩 서열에 작동 가능하게 연결되는 프로모터는 시미안 바이러스 40(SV40) 유래 프로모터, 마우스 유방암 바이러스(MMTV) 프로모터, 인간 면역결핍 바이러스(HIV) 프로모터, 예컨대 소 면역결핍 바이러스(BIV) 장형 말단 반복(LTR) 프로모터, 몰로니(Moloney) 바이러스 프로모터, 조류 백혈증 바이러스(ALV) 프로모터, 사이토메갈로바이러스(CMV) 프로모터, 예컨대 CMV 최조기 프로모터, 엡스타인 바 바이러스(EBV) 프로모터, U6 프로모터, 예컨대 인간 U6 프로모터, 또는 라우스(Rous) 육종 바이러스(RSV) 프로모터일 수 있다. 프로모터는 또한 인간 유전자, 예컨대 인간 유비퀴틴 C(hUbC), 인간 액틴, 인간 미오신, 인간 헤모글로빈, 인간 근육 크레아틴, 또는 인간 메탈로티오닌 유래 프로모터일 수 있다. 프로모터는 또한 조직 특이적 프로모터, 예컨대 천연 또는 합성, 근육 또는 피부 특이적 프로모터일 수 있다. 이러한 프로모터의 예는 그 내용의 전문이 본원에 참조로 포함되는, US 특허 출원 공개 번호 US20040175727 및 제 US20040192593에 기술되어 있다. 근육-특이적 프로모터의 예는 Spc5-12 프로모터 (전체 내용이 본원에 참고로 인용된 미국 특허 출원 공개 US 20040192593에 기술되어 있음; Hakim et al. Mol. Ther. Methods Clin. Dev. (2014) 1:14002; and Lai et al. Hum Mol Genet. (2014) 23(12): 3189-3199), MHCK7 프로모터(Salva et al., Mol. Ther. (2007) 15:320-329에 기술되어 있음), CK8 프로모터(Park et al. PLoS ONE (2015) 10(4): e0124914에 기술되어 있음) 및 CK8e 프로모터(Muir et al., Mol. Ther. Methods Clin. Dev. (2014) 1:14025에 기술되어 있음)를 포함한다. 일부 구현예에서, gRNA 및/또는 Cas9 단백질의 발현은 tRNA에 의해 유도된다.
- [0209] gRNA 분자 및/또는 Cas9 분자를 코딩하는 각각의 폴리뉴클레오타이드 서열은 각각 프로모터에 작동 가능하게 연결될 수 있다. gRNA 분자 및/또는 Cas9 분자에 작동 가능하게 연결된 프로모터는 동일한 프로모터일 수 있다. gRNA 분자 및/또는 Cas9 분자에 작동 가능하게 연결된 프로모터는 상이한 프로모터일 수 있다. 프로모터는 구성성 프로모터, 유도성 프로모터, 억제성 프로모터 또는 조절가능한 프로모터일 수 있다.
- [0210] 벡터는 또한 폴리아데닐화 신호를 포함할 수 있으며, 이는 CRISPR/Cas9-기반 유전자 편집 시스템의 하류에 있을 수 있다. 폴리아데닐화 신호는 SV40 폴리아데닐화 신호, LTR 폴리아데닐화 신호, 소 성장 호르몬(bGH) 폴리아데닐화 신호, 인간 성장 호르몬(hGH) 폴리아데닐화 신호, 또는 인간 β -글로빈 폴리아데닐화 신호일 수 있다. SV40 폴리아데닐화 신호는 pCEP4 벡터(Invitrogen, San Diego, CA)로부터의 폴리아데닐화 신호일 수 있다.
- [0211] 벡터는 또한 CRISPR/Cas9-기반 유전자 편집 시스템, 즉 Cas9 단백질 또는 Cas9 융합 단백질 코딩 서열 또는 sgRNA, 또는 CRISPR/Cas9-기반 유전자 편집 시스템의 상류에 인핸서를 포함할 수 있다. 인핸서는 DNA 발현을 위해 필요할 수 있다. 인핸서는 인간 액틴, 인간 미오신, 인간 헤모글로빈, 인간 근육 크레아틴 또는 바이러스 인핸서, 예컨대 CMV, HA, RSV 또는 EBV 유래의 인핸서일 수 있다. 폴리뉴클레오타이드 기능 인핸서는 각각의 내용 전체가 참조로 포함되는, U.S. 특허 번호. 5,593,972, 5,962,428, 및 W094/016737에 기재되어 있다. 벡터는 또한 벡터를 염색체외로 유지하고 세포에서 여러 카피의 벡터를 생성하기 위해, 포유류 복제 기원을 포함할 수 있다. 벡터는 또한 조절 서열을 포함할 수 있고, 이는 벡터가 투여되는 포유류 또는 인간 세포에서의 유전자 발현을 위해 매우 적합할 수 있다. 벡터는 또한 리포터 유전자, 예컨대 녹색 형광 단백질("GFP") 및/또는 선별 마커, 예컨대 하이그로마이신("Hygro")을 포함할 수 있다.
- [0212] 벡터는 전체가 참조로 포함되는, 문헌[Sambrook et al., Molecular Cloning and Laboratory Manual, Second Ed., Cold Spring Harbor(1989)]을 포함하는 일상적 기법 및 쉽게 이용 가능한 원료에 의해 단백질을 생성하기 위한 발현 벡터 또는 시스템일 수 있다. 일부 구현예에서, 벡터는 Cas9 단백질 또는 Cas9 융합 단백질을 코딩하는 핵산 서열 및 SEQ ID NO: 1-19, 41, 42 중 적어도 하나의 핵산 서열, 또는 그의 상보체를 포함하는 적어도 하나의 gRNA를 코딩하는 핵산 서열을 포함하는 CRISPR/Cas9-기반 유전자 편집 시스템을 코딩하는 핵산 서열을 포함할 수 있다. 일부 구현예에서, Cas9 단백질 또는 Cas9 융합 단백질은 SEQ ID NO: 26 중 임의의 하나의 핵산 서열에 의해 코딩된다. 일부 구현예에서, 상기 벡터는 SEQ ID NO: 39 또는 SEQ ID NO: 40의 핵산 서열을 포함한다.
- [0213] **10. 약학적 조성물**
- [0214] 본원에 개시된 주제는 상기 기술된 유전자 구축물을 포함하는 조성물을 제공한다. 본 발명에 따른 약학적 조성물은 사용되는 투여 방식에 따라 제제화될 수 있다. 약학적 조성물이 주사용 약학적 조성물인 경우, 이들은 무

균, 발열원이 없고, 미립자가 없다. 등장성 제제가 바람직하게 사용된다. 일반적으로, 등장성을 위한 첨가제는 염화나트륨, 텍스트로스, 만니톨, 소르비톨 및 락토오스를 포함할 수 있다. 일부의 경우, 인산 완충 식염수와 같은 등장 용액이 바람직하다. 안정화제에는 젤라틴과 알부민이 포함된다. 일부 구현예에서, 혈관수축제가 제제에 첨가된다.

[0215] 조성물은 약학적으로 허용가능한 부형제를 추가로 포함할 수 있다. 약학적으로 허용가능한 부형제는 기능적 분자, 예컨대 비히클, 아주반트, 담체, 또는 희석제일 수 있다. 약학적으로 허용가능한 부형제는 전달감염 촉진제일 수 있고, 여기에는 표면 활성화제, 예컨대 면역-자극 복합체(ISCOMS), 프로인트 불완전 아주반트, 모노포스포릴 지질 A를 포함하는 LPS 유사체, 무라밀 펩타이드, 퀴논 유사체, 스쿠알렌 및 스쿠알렌과 같은 소포, 히알루론산, 지질, 리포솜, 칼슘 이온, 바이러스 단백질, 다중음이온, 다중양이온, 또는 나노입자, 또는 다른 공지된 전달감염 촉진제가 포함될 수 있다.

[0216] 전달감염 촉진제는 다중음이온, 폴리-L-글루타메이트(LGS)를 포함하는 다중양이온, 또는 지질이다. 전달감염 촉진제는 폴리-L-글루타메이트이며, 보다 바람직하게는, 폴리-L-글루타메이트는 6 mg/ml 미만의 농도로 골격근 또는 심장 근육내 게놈 편집을 위한 조성물에 존재한다. 전달감염 촉진제에는 또한 표면 활성화제, 예컨대 면역-자극 복합체(ISCOMS), 프로인트 불완전 아주반트, 모노포스포릴 지질 A를 포함하는 LPS 유사체, 무라밀 펩타이드, 퀴논 유사체 및 소포, 예컨대 스쿠알렌 및 스쿠알렌이 포함될 수 있고, 히알루론산이 또한 유전자 구축물과 함께 투여될 수 있다. 일부 구현예에서, 조성물을 인코딩하는 DNA 벡터에는 또한 전달감염 촉진제, 예컨대 지질, 레시틴 리포솜 또는 DNA-리포솜 혼합물로 당분야에 공지된 다른 리포솜을 포함하는 리포솜(예를 들어 국제 특허 공보 W09324640 참고), 칼슘 이온, 바이러스 단백질, 다중음이온, 다중양이온, 또는 나노입자, 또는 다른 공지된 전달감염 촉진제가 포함될 수 있다. 바람직하게는, 전달감염 촉진제는 다중음이온, 폴리-L-글루타메이트(LGS)를 포함하는 다중양이온, 또는 지질이다.

[0217] 11. 전달 방법

[0218] 본원에 제공된 것은 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 그의 조성물을 세포에 전달하는 방법이다. 조성물의 전달은 세포에서 발현되어 세포 표면에 전달되는 핵산 분자로서 조성물의 형질감염 또는 전기천공일 수 있다. 핵산 분자는 BioRad Gene Pulser Xcell 또는 Amaxa Nucleofector IIb 장치를 사용하여 전기천공될 수 있다. BioRad 전기천공 용액, Sigma 인산 완충 식염수 제품 #D8537(PBS), Invitrogen OptiMEM I(OM) 또는 Amaxa Nucleofector 용액 V(N.V.)를 포함하는 몇몇 상이한 완충액이 이용될 수 있다. 전달감염에는 전달감염제, 예컨대 Lipofectamine 2000이 포함될 수 있다.

[0219] 조직에 대한 본원에 개시된 유전자 구축물 또는 조성물의 전달 시, 그리고 포유류 세포 내로의 벡터의 전달 시, 형질감염된 세포는 gRNA 분자 및 Cas9 분자를 발현할 것이다. 유전자 구축물 또는 조성물을 포유류에 투여하여 유전자 발현을 변경시키거나 게놈을 재조직 또는 변형시킬 수 있다. 예를 들어, 유전자 구축물 또는 조성물은 포유동물에 투여되어, 포유동물내 디스트로핀 유전자를 교정할 수 있다. 포유동물은 인간, 비-인간 영장류, 소, 돼지, 양, 염소, 영양, 들소, 물소, 새와 동물, 사슴, 고슴도치, 코끼리, 라마, 알파카, 마우스, 래트, 또는 닭, 및 바람직하게는 인간, 소, 돼지, 또는 닭일 수 있다.

[0220] gRNA 분자(들)과 Cas9 분자를 코딩하는 유전자 구축물(예를 들어, 벡터)은 생체내 전기천공, 리포솜 매개된, 나노입자 촉진된 및/또는 재조합 벡터에 의해 또는 없이 DNA 주입(DNA 백신접종이라고도 함)에 의해 포유동물에게 전달될 수 있다. 재조합 벡터는 임의의 바이러스성 모드에 의해 전달될 수 있다. 바이러스성 모드는 재조합 렌티바이러스, 재조합 아데노바이러스 및/또는 재조합 아데노-연관 바이러스일 수 있다.

[0221] 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 이를 포함하는 조성물은 세포에 도입되어 디스트로핀 유전자(예를 들어, 인간 디스트로핀 유전자)를 유전적으로 교정할 수 있다. 특정 구현예에서, 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 이를 포함하는 조성물은 DMD 환자로부터 근모세포 세포내로 도입된다. 특정 구현예에서, 유전자 구축물(예를 들어, 벡터) 또는 이를 포함하는 조성물은 DMD 환자로부터의 섬유아세포 세포로 도입되고, 유전적으로 교정된 섬유아세포 세포가 MyoD에 의해 처리되어 근모세포로의 분화를 유도할 수 있으며, 이는 교정된 디스트로핀 단백질이 기능적인지를 검증하기 위해, 및/또는 대상체를 치료하기 위해, 대상체, 예컨대 대상체의 손상된 근육에게 이식될 수 있다. 변형된 세포는 또한 유도된 만능 줄기 세포, 골수-유래 전구 세포, 골격근 전구세포, DMD 환자의 인간 골격 근모세포, CD 133⁺ 세포, 중간혈관모세포(mesoangioblast), 및 MyoD- 또는 Pax7-형질도입된 세포, 또는 다른 근원 전구세포 세포와 같은 줄기세포일 수 있다. 예를 들어, CRISPR/Cas9-기반 유전자 편집 시스템은 유도된 만능 줄기 세포의 뉴런 또는 근원성 분화를 일으킬 수 있다.

12. 투여 경로

본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 이를 포함하는 조성물은 경구, 비경구, 설하, 경피, 직장, 경점막, 국소, 흡입을 통해, 협측 투여를 통해, 흉강내, 정맥내, 동맥내, 복강내, 피하, 근육내, 비내 경막내, 및 관절내 또는 이의 조합을 포함하는 상이한 경로에 의해 대상체에 투여될 수 있다. 특정 구현예에서, 근육내, 정맥 내 또는 이들의 조합으로 대상체(예를 들어, DMD를 앓고있는 대상체)에게 투여된다. 수의학적 이용을 위해, 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 조성물은 일반적인 수의학적 관례에 따라 적합하게 허용가능한 제형물로 투여될 수 있다. 수의사는 특정 동물에 가장 적절한 투여 요법 및 투여 경로를 쉽게 결정할 수 있다. 조성물은 전통적 주사기, 무바늘 주사 장치, "마이크로발사 충격 유전자 총" 또는 다른 물리적 방법, 예컨대 전기천공("EP"), "수력학적 방법", 또는 초음파에 의해 투여될 수 있다.

본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 조성물은 생체내 전기천공, 리포솜 매개, 나노입자 촉진된 것, 재조합 벡터, 예컨대 재조합 렌티바이러스, 재조합 아데노바이러스 및 재조합 아데노바이러스 연관 바이러스를 사용하여, 및 사용하지 않는 DNA 주사(DNA 백신접종이라고도 함)를 포함하는 여러 기술에 의해 포유동물에게 전달될 수 있다. 상기 조성물은 골격근 또는 심장 근육으로 주사될 수 있다. 예를 들어, 조성물은 경골 전방 근육 또는 꼬리에 주사될 수 있다.

일부 구현예에서, 본원에 개시된 유전자 구축물(예를 들어, 벡터) 또는 그의 조성물은 1) 꼬리 정맥 주사(전신 투여)에 의해 성체 마우스에 투여되고; 2) 근육내 주사, 예를 들어, TA 또는 비복근과 같은 근육 내로의 국소 주사에 의해 성체 마우스에 투여되고; 3) P2 마우스에 복강 내 주사에 의해 투여되고; 또는 4) P2 마우스에 안면 정맥 주사(전신)에 의해 투여된다.

13. 세포 유형

상기 전달 방법 및/또는 투여 경로는 불멸화된 근모세포 세포, 예컨대 야생형 및 DMD 환자 유도 세포주, 예를 들어 Δ48-50 DMD, DMD 6594(de148-50), DMD 8036(de148-50), C25C14 및 DMD-7796 세포주, 원생 DMD 진피 섬유아세포, 유도된 만능 줄기 세포, 골수-유래된 전구세포, 골격근 전구세포, DMD 환자의 인간 골격 근모세포, CD 133⁺ 세포, 중간혈관모세포, 심장근육세포, 간세포, 연골세포, 중간엽 전구세포 세포, 조혈 줄기세포, 평활근 세포 및 MyoD- 또는 Pax7-형질도입된 세포, 또는 다른 근원 전구세포 세포를 포함하지만, 이에 한정되지 않는 무수히 많은 세포 유형, 예를 들어, DMD의 세포-기반 요법에 대해 현재 조사중인 세포 유형에 의해 사용될 수 있다. 인간의 근원 세포의 불멸화는 유전적으로 교정된 근원 세포의 클론 유도를 위해 사용될 수 있다. 세포는 유전적으로 교정된 디스트로핀 유전자를 포함하고 게놈의 단백질 코딩 영역에 다른 뉴클레아제-도입된 돌연변이가 없는 불멸화된 DMD 근섬유의 클론 집단을 분리 및 확장하기 위해 생체외에서 변형될 수 있다. 대안적으로, 비-바이러스성 또는 비-통합 바이러스성 유전자 전달에 의한, 또는 세포-침투 모티프를 함유하는 정제된 단백질 및 gRNA의 직접 전달에 의한 CRISPR/Cas9-기반 시스템의 일시적 생체내 전달이 최소의 또는 외인성 DNA 통합의 위험이 없이, 원위치내 매우 특별한 교정을 가능하게 할 수 있다.

14. 키트

본 명세서에는 돌연변이된 디스트로핀 유전자를 교정하는데 사용될 수 있는 키트가 제공된다. 키트는 돌연변이된 디스트로핀 유전자를 교정하기 위한 적어도 gRNA 및 CRISPR/Cas9-기반 유전자 편집 시스템을 사용하기 위한 지침서를 포함한다. 또한, 골격근 또는 심장 근육에서 디스트로핀 유전자의 게놈 편집에 사용될 수 있는 키트가 본원에 제공된다. 상기 키트는 상기 기술된 바와 같이, 골격근 또는 심장 근육에서 게놈 편집을 위한 유전자 구축물(예를 들어, 벡터) 또는 이를 포함하는 조성물 및 상기 조성물의 사용 지침서를 포함한다.

키트에 포함된 사용 지침서는 포장재에 부착하거나, 포장재 삽입물로 포함될 수 있다. 사용 지침서는 일반적으로 서면 또는 인쇄물이지만, 이에 국한되지는 않다. 상기 사용 지침서를 저장하고, 이들을 최종 사용자에게 전달할 수 있는 임의의 매체가 이 개시 내용으로 고려된다. 상기 매체는 전자 저장 매체(예를 들어, 자기 디스크, 테이프, 카트리지, 칩), 광학 매체(예를 들어, CD ROM) 등을 포함하지만, 이에 한정되지 않는다. 본 명세서에 사용된 "지침서"란 용어는 설명서를 제공하는 인터넷 사이트의 주소를 포함할 수 있다.

골격근 또는 심장 근육에서 디스트로핀 유전자의 돌연변이된 디스트로핀 또는 게놈 편집을 교정하기 위한 유전자 구축물(예를 들어, 벡터) 또는 이를 포함하는 조성물은 디스트로핀 유전자의 영역에 특이적으로 결합하고 절단하는, 상기과 같이 gRNA 분자(들) 및 Cas9 분자를 포함하는 변형된 AAV 벡터를 포함할 수 있다. CRISPR/Cas9-기반의 유전자 편집 시스템은 위에 기술된 바와 같이, 돌연변이된 디스트로핀 유전자의 특정 영역에 특이적으로 결합하고 표적화하기 위해 키트에 포함될 수 있다. 상기 키트는 상기 기술된 바와 같이 공여체 DNA, 상이한

gRNA 또는 트랜스유전자를 추가로 포함할 수 있다.

15. 실시예

본원에 기술된 본 발명의 방법의 다른 적당한 변형 및 개조는 용이하게 적용 가능하고 인식할 수 있으며, 본 명세서의 범주 또는 본원에 개시된 양태 또는 구현예를 벗어나지 않는 적당한 균등물을 사용하여 이루어질 수 있다는 것은 당업자에게 명백할 것이다. 이제 본 개시 내용을 상세히 설명하였지만, 하기의 실시예를 참고하여 보다 명확하게 이해될 것이며, 이는 단지 본 발명의 일부 양태 및 구현예를 예시하기 위한 것이며, 본 발명의 범주를 제한하는 것으로 간주되어서는 안된다. 본원에 인용된 모든 저널 참고문헌, 미국 특허 및 공보의 개시내용은 본원에 참고 문헌으로 포함된다.

본 발명은 다음의 비-제한적인 실시예에 의해 예시되는, 다수의 양태를 갖는다.

실시예 1

인간 디스트로핀 유전자 표적화

CRISPR/Cas9-기반 유전자 편집 시스템은 인간 디스트로핀 유전자의 엑손 51을 표적화하고 결실하는데 사용되었다. *에스. 피오게네스*(*S. pyogenes*) Cas9보다 약 1kb 작은 *에스. 아우레우스*(*S. aureus*) Cas9(SaCas9)는 CRISPR/Cas9-기반 유전자 편집 시스템을 전달하기 위해 아테노-연관 바이러스(AAV)와 함께 사용되었다. *에스. 아우레우스*(*S. aureus*) Cas9 분자를 코딩하는 코돈 최적화 핵산 서열은 SEQ ID NO: 43 또는 SEQ ID NO: 44에 개시되어있다. 도 3은 SaCas9 및 2개의 바이러스 벡터상의 2개의 gRNA를 근육 조직에 AAV-기반의 생체내 동시 전달하는 것의 개략도를 도시한다. 각각의 벡터는 CMV 및 hU6 프로모터(PT366-179(SEQ ID NO: 39) 및 PT366-183(SEQ ID NO: 40))에 의해 유도된 SaCas9 및 하나의 gRNA의 복제본을 가졌다.

인간 디스트로핀 유전자인 JCR89(엑손 51의 표적 상류)와 JCR91(엑손 51의 표적 하류)을 표적화하는 각 gRNA의 활성은 인간 계놈에 대하여 설계되었으며, HEK293T 세포(정상 버전의 디스트로핀 유전자) 및 DMD 환자 근모세포 세포주(DMD 8036 및 DMD 6594, 이는 각각 돌연변이 형태의 디스트로핀 유전자를 가짐)의 측량 검정에 의해 측정되었다(도 1 참조). 측량 검정은 CRISPR/Cas9-기반 유전자 편집 시스템의 indel을 나타내는 계놈 DNA의 미스매치를 탐지한다. JCR89의 경우, 모 밴드 크기는 555nt이고, 사용된 프라이머는 순방향 프라이머-aagttacttgcaggcatga(SEQ ID NO: 91); 및 역방향 프라이머-gaaaaacttctgccaactttatca(SEQ ID NO: 92)였다. 예상 절단 밴드 크기는 134 nt 및 421 nt이다. JCR91의 경우, 모 밴드 크기는 632nt이고, 사용된 프라이머는: 순방향 프라이머-tgcaaataacaaaagtagccataca(SEQ ID NO: 93); 및 역방향 프라이머-tctttagaaaggcttgaaagctg(SEQ ID NO: 94)였다. 예상 절단 밴드 크기는 210 nt 및 422 nt이다.

HEK293T 세포 및 DMD 근모세포(DMD 8036 및 DMD 6594)를 SaCas9 및 gRNA JCR89 및 JCR91(SEQ ID NO: 37 및 SEQ ID NO: 38)에 의해 처리하였다. 계놈 DNA는 순방향 프라이머-cttcactgctggccagttta(SEQ ID NO: 95); 및 역방향 프라이머-tctttagaaaggcttgaaagctg(SEQ ID NO: 94)로 증폭시켰다. 예상 모 밴드 크기는 1646nt이고, 예상되는 "완전한" 결실 밴드는 766nt였다(gRNA 절단 부위 사이의 실제 결실 크기는 indel의 발생으로 인해 766nt에서 다양함). 도 2A는 HEK293T 세포 및 DMD 근모세포의 계놈 DNA에서 엑손 51의 결실을 나타낸다. 도 2B는 DMD 근모세포로부터의 cDNA에서 엑손 51의 결실을 보여준다. "RT 없음"은 역전사효소가 첨가되지 않은 음성 대조군이다.

엑손 51을 결실시키기 위해 인간 DMD 유전자(hDMD/*mdx* 마우스)를 운반하는 트랜스제닉 마우스에 CRISPR/Cas9-기반 유전자 편집 시스템을 주사했다. SaCas9 및 gRNA를 운반하는 바이러스 벡터의 국소 AAV8 전달을 경골 전방(TA) 근육에 적용하였다. 표 1을 참조한다. 3마리의 마우스에게 AAV8을 주사하고: 1마리의 마우스에게 두 TA("HH")에 고용량의 AAV8을 주사하고, 1마리의 마우스에게 두 TA("LL")에 저용량의 AAV8을 주사하고, 1마리의 마우스에게 왼쪽 TA에 저용량을 주사하고, 오른쪽 TA("LH")에 고용량을 주사하였다. 투여량을 표 1의 2열에 기재하였다. 분석을 위한 조직을 채취하기 위해 치료 8주 후에 마우스를 희생시켰다("주 PT"). 중첩된 PCR은 HH 마우스의 양쪽 팔다리에서, LL 마우스의 오른쪽 TA에서 및 LH 마우스의 오른쪽 팔다리에서, 엑손 51의 결실을 나타냈다.

[0241] [표 1]

실험	AAV8 투여량	전달	주 PT	증접된 PCR - gDNA Δ51
HH, LL, LH	HH: 6.6E11 LL: 1E11	전경골근내 (TA)	8	HH: 둘다, LL: R, LH: R

[0242]

[0243] 마우스 TA 근육으로부터 채취된 게놈 DNA는 순방향 프라이머: cttcactgctggccagtta(SEQ ID NO: 95) 및 역방향 프라이머: tctttagaaggcttgaaagctg(SEQ ID NO: 94)를 사용하여 제1 PCR 반응에서 증폭시켰다. 1-3 μ L의 PCR 생성물을 순방향 프라이머-aagttacttgtccaggcatga(SEQ ID NO: 91); 및 역방향 프라이머-ttgaacatggcattgcataaA(SEQ ID NO: 96)를 사용하는 제2 PCR 반응(2X gDNA PCR)에 사용하였다. 이 제2 PCR은 1089nt의 예상 모 밴드와 323nt의 예상 결실 밴드를 가졌다(gRNA 절단 부위 사이의 실제 결실 크기는 indel의 발생으로 인해 323nt에서 다양함). 도 4는 제2 PCR 결과를 도시한다. "L" 레인은 왼쪽 TA 근육에 대한 결과를 나타내며, 대조군으로 사용되어 식염수를 받았다. "R" 레인은 동일한 양으로 미리 혼합된 2개의 바이러스 벡터가 주입된 오른쪽 TA 근육의 결과를 보여준다. CRISPR/Cas9-기반 유전자 편집 시스템은 또한 전신성 AAV8 전달을 통해 hDMD/mdx 마우스의 꼬리 정맥 내로 주사되었다(도 5 참조). 마우스 간(도 5-왼쪽 패널) 및 심장(도 5-오른쪽 패널)으로부터 채취된 게놈 DNA도 도 4와 동일한 프로토콜을 사용하여 증폭시켰다. 예상되는 약 300개의 뉴클레오타이드의 밴드는 엑손 51의 결실을 나타냈다.

[0244] 인간 및 붉은털 원숭이 디스트로핀 유전자 서열 또는 인간 디스트로핀 유전자 서열을 표적화하는 추가의 gRNA가 생성되고 선택되었다(도 6 및 표 2 참조). 표 2는 gRNA의 일반 표적, 인식되는 게놈 가닥, gRNA 서열 및 gRNA와 관련된 PAM 서열을 열거한다. 게놈 플러스 가닥에 표시된 gRNA의 표적 게놈 서열을 표 3에 나타내었다. 배양된 인간 세포에서 상기 gRNA를 시험하여 결실을 생성하는 최적의 활성 및 gRNA 조합을 발견하였다. 선택된 gRNA는 또한 인간 게놈에서 예측된 특이성에 기초하여 우선 순위를 정하고(도 8), 19-23개의 뉴클레오타이드에서 변화하는 최적 표적 서열 길이에 대해 스크리닝하였다(도 10 및 11). 도 6은 인간과 붉은털 원숭이 게놈간에 보존된 표 2에 열거된 다양한 gRNA 표적을 보여준다. 각 gRNA의 위치는 인간 디스트로핀 유전자의 엑손 51과 관련되어 표시된다.

[0245] [표 2]

gRNA의 목록

gRNA	일반 표적	가닥	gRNA 서열 (5'-3')	SEQ ID NO:	PAM 서열	SEQ ID NO:
JCR89	인간 DMD 엑손 51 (상류)	플러스	AAAGATATATAATGTCATGAAT	1	AAGAGT	53
JCR91	인간 DMD 엑손 51 (하류)	플러스	GCAGAATCAAATATAATAGTCT	2	GGAAT	54
JCR94	인간 DMD 엑손 52 (상류)	마이너스	AACAAATATCCCTTAGTATC	41	AGG	55
JCR99	인간 DMD 엑손 52 (하류)	마이너스	AATGTATTTCTTCTATTCAA	42	TGG	56
JCR159	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	AACAATAAGTCAAATTTAATTG	3	AAGAGT	53
JCR160	인간 및 붉은털 DMD 엑손 51 (하류)	마이너스	GAAGTGGTGGGAAATGGTCTAG	4	GAGAGT	57
JCR167	인간 및 붉은털 DMD 엑손 51 (하류)	마이너스	TCCTTTGGTAAATAAAAGTCCT	5	GGGAGT	58

[0246]

JCR166	인간 및 붉은털 DMD 엑손 51 (하류)	마이너스	TAGGAATCAAATGGACTTGGAT	6	TTGAAT	59
JCR168	인간 및 붉은털 DMD 엑손 51 (상류)	플러스	TAATTCTTCTAGAAAGAGCCT	7	CAGAGT	60
JCR170	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	CTCTTGCATCTTGACATGTCC	8	TGGAGT	61
JCR171	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	ACTTAGAGGTCTTACATACA	9	ATGAGT	62
JCR156	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	TCAGAGGTGAGTGGTGAGGGGA	10	AGGAAT	63
JCR157	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	ACACACAGCTGGGTTATCAGAG	11	GAGAGT	57
JCR176	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	CACAGCTGGGTTATCAGAG	12	GAGAGT	57
JCR177	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	ACACAGCTGGGTTATCAGAG	13	GAGAGT	57
JCR178	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	CACACAGCTGGGTTATCAGAG	14	GAGAGT	57
JCR179	인간 및 붉은털 DMD 엑손 51 (상류)	마이너스	AACACACAGCTGGGTTATCAGAG	15	GAGAGT	57
JCR180	인간 및 붉은털 DMD 엑손 51 (하류)	마이너스	CTGGTGGGAAATGGTCTAG	16	GAGAGT	57
JCR181	인간 및 붉은털 DMD 엑손 51 (하류)	마이너스	ACTGGTGGGAAATGGTCTAG	17	GAGAGT	57
JCR182	인간 및 붉은털 DMD 엑손 51 (하류)	마이너스	AACTGGTGGGAAATGGTCTAG	18	GAGAGT	57
JCR183	인간 및 붉은털 DMD 엑손 51 (하류)	마이너스	AGAACTGGTGGGAAATGGTCTAG	19	GAGAGT	57

[0247]

[0248] [표 3]

gRNA의 표적 서열

gRNA 명칭	표적 서열 (플러스 가닥 5'-3')	SEQ ID NO:	코멘트
JCR89	AAAGATATATAATGTCATGAAT	64	
JCR91	GCAGAATCAAATATAATAGTCT	65	
JCR159	CAATTAATTTGACTTATTGTT	66	엑손 51의 101 상류
JCR160	CTAGACCATTTCACACAGTTC	67	엑손 51의 78 하류
JCR167	AGGACTTTTATTTACCAAAGGA	68	엑손 51의 1534 하류
JCR166	ATCCAAGTCCATTGATTCTTA	69	엑손 51의 1266 하류
JCR168	TAATCTTTCTAGAAAGAGCCT	70	엑손 51의 1824 상류
JCR170	GGACATGTGCAAGATGCAAGAG	71	엑손 51의 2851 상류
JCR171	TGTATGTAGAAGACCTCTAAGT	72	엑손 51의 2947 상류
JCR156	TCCCCTCACCCTCACCTCTGA	73	엑손 51의 1300 상류
JCR157	CTCTGATAACCCAGCTGTGTGT	74	엑손 51의 1284 상류
JCR176	CTCTGATAACCCAGCTGTGTG	75	JCR157 - 19 뉴클레오타이드
JCR177	CTCTGATAACCCAGCTGTGTG	76	JCR157 - 20 뉴클레오타이드
JCR178	CTCTGATAACCCAGCTGTGTG	77	JCR157 - 21 뉴클레오타이드
JCR179	CTCTGATAACCCAGCTGTGTGTT	78	JCR157 - 23 뉴클레오타이드
JCR180	CTAGACCATTTCACACAG	79	JCR160 - 19 뉴클레오타이드
JCR181	CTAGACCATTTCACACAGT	80	JCR160 - 20 뉴클레오타이드
JCR182	CTAGACCATTTCACACAGTT	81	JCR160 - 21 뉴클레오타이드
JCR183	CTAGACCATTTCACACAGTTCT	82	JCR160 - 23 뉴클레오타이드

[0249]

[0250] 인간 HEK293T 세포를 표 2에 기재된 개별 후보 gRNA로 형질감염시켰다. 후보 gRNA의 활성은 측량 검정에 의해 측정하였다(도 7 참조). JCR160의 경우, 모 밴드 크기는 483nt이고, 사용된 프라이머는: 순방향 프라이머-cgggcttggacagaacttac(SEQ ID NO: 97); 역방향 프라이머-ctgcgtagtgcacaaacaaa(SEQ ID NO: 98)였다. 예상 절단 밴드 크기는 192 nt 및 291 nt이다. JCR157의 경우, 모 밴드 크기는 631nt이고, 사용된 프라이머는: 순방향 프라이머-gagatgtcttttgcagctttcc(SEQ ID NO: 99); 역방향 프라이머-gggaccttggtaaagccaca(SEQ ID NO: 100)였다. 예상 절단 밴드 크기는 147 nt 및 484 nt였다.

[0251]

후보 gRNA의 특이성은 CasOFFinder 프로그램을 사용하여 예측하였다(Bae et al. (2014) *Bioinformatics* 30: 1473-1475; 도 8 참조). 후보 gRNAs는 측량 검정에 의해 측정된 표적-이외 활성, 표적-상 활성 및 엑손으로부터의 거리에 기초하여 평가되고 선택되었다. gRNA JCR157 및 JCR160은 예측된 표적-이외 결합이 낮았으며, 추가 시험에 사용되었다.

[0252]

HEK293T 세포를 형질감염시키고, DMD 6594 세포를 gRNA JCR157을 함유하는 변형된 pD0240 플라스미드, gRNA JCR160을 함유하는 변형된 pD0240 플라스미드 및 SaCas9를 함유하는 플라스미드(pD0242; SEQ ID NO: 83)로 전기천공하였다. 모 밴드는 2451nt이고, 결실 밴드는 약 840 내지 850nt가 될 것으로 예측되었다. 도 9는 순방향 프라이머-tgccttttcaatcattgtttcg(SEQ ID NO: 101) 및 역방향 프라이머-agaaggcaaattggcacaga (SEQ ID NO: 102)를 사용하여 게놈 DNA(약 850개 뉴클레오타이드)의 PCR에 의해 측정된 엑손 51의 결실을 나타낸다. gRNA 절단 부위 사이에서 생성된 결실은 약 1611nt이다.

[0253]

도 10은 도 7의 JCR157에 사용된 프라이머 및 PCR 조건을 사용하여 HEK293T 세포에서 측량 검정에 의해 측정된 gRNA JCR157(19, 20, 21, 22 및 23개 뉴클레오타이드)의 다양한 표적 길이의 활성을 나타낸다. 도 11은 순방향 프라이머-cgggcttggacagaacttac(SEQ ID NO: 97) 및 역방향 프라이머-ctgcgtagtgcacaaacaaa(SEQ ID NO: 98)를 사용하여 HEK293T 세포에서 측량 검정에 의해 측정된 gRNA JCR160(19, 20, 21, 22 및 23개 뉴클레오타이드)의 다양한 표적 길이의 활성을 나타낸다. 모 밴드 크기는 483nt로 예측되고, 예상 절단 밴드 크기는 209nt 및 274nt이었다.

[0254]

gRNA JCR157 및 JCR160(21, 22 또는 23개 뉴클레오타이드)의 다양한 표적 길이의 조합을 도 9에서 사용된 조건

을 사용하여 HEK293T 세포에서 사용하였다. 도 12는 게놈 DNA의 PCR을 나타낸다. 각각 23개의 뉴클레오타이드 표적을 갖는 JCR157 및 JCR160의 조합은 거의 50% 결실을 가졌다.

[0255] HEK293T 세포("293s") 및 DMD6594 세포("DMD6594s")에서 표적-상 뉴클레아제 활성을 입증하기 위해 23 nt의 표적 서열을 사용하여 SaCas9로 엑손 51(상류 JCR179 및 하류 JCR183)에 인접한 각각의 gRNA를 수행하였다(도 13 참조). JCR179의 경우, 모 밴드 크기는 594nt이고, 사용된 프라이머는: 순방향 프라이머-tgcctttcaatcattgtttcg(SEQ ID NO: 101); 및 역방향 프라이머-aaggcccaaaatgtgaaat(SEQ ID NO: 103)였다. 예상 절단 밴드 크기는 594 nt 및 130 nt이다. JCR183의 경우, 모 밴드 크기는 731nt이고, 사용된 프라이머는: 순방향 프라이머-gagtttggctcaaattgttactctt(SEQ ID NO: 104); 역방향 프라이머-ctgcgtagtgcacaaacaaa(SEQ ID NO: 98)였다. 예상 절단 밴드 크기는 440 nt 및 291 nt이었다. 도 13은 측량 검정에 의한 시험관내 표적-상 뉴클레아제 활성을 나타낸다.

[0256] 인간 HEK293T 세포를 형질감염시키고, DMD 근모세포(DMD 6594)를 SaCas9 및 gRNA JCR179 및 JCR183을 함유하는 플라스미드(JCR157 및 JCR160의 23 nt 표적; SEQ ID NO: 37 및 SEQ ID NO: 38)로 전기천공하였다. DMD 6594 세포는 이미 엑손 48-50이 부족한 불멸화된 DMD 환자 근모세포이다. 모 밴드는 2451nt이고, 결실 밴드는 약 823nt로 예측된다. 도 14는 순방향 프라이머-tgcctttcaatcattgtttcg(SEQ ID NO: 101) 및 역방향 프라이머-agaaggcaaatggcacaga(SEQ ID NO: 102)를 사용하여 게놈 DNA의 PCR에 의해 측정된 인간 HEK293T 세포(왼쪽 패널) 및 DMD 6594s 세포(오른쪽 패널)에서의 게놈 DNA내 엑손 51의 시험관내 결실을 나타낸다. gRNA 절단 부위 사이에서 생성된 결실은 약 1628nt이었다. 상부 패널은 HEK293T 세포 및 DMD 6594 세포에서의 상류 및 하류 gRNA의 표적 유전자의 도식을 나타내며, 보라색은 정상적으로 가공된 엑손을 나타내고, 노란색은 돌연변이체 엑손을 나타낸다. 중간 패널은 게놈 결실 영역에 걸친 PCR의 결과를 나타내며, 별표는 결실을 나타낸다. 하단 패널은 게놈 DNA의 드롭렛 디지털 PCR을 나타낸다. HEK293T 세포에서, gRNA 및 SaCas9는 16%의 결실을 가졌으며, DMD 6594 세포는 약 10%의 편집을 가졌다.

[0257] 게놈 DNA의 변화가 RNA로 전사되었는지를 결정하기 위해, SaCas9 및 두 gRNA를 함께 형질감염시키고 7일 동안 분화시킨 DMD 근모세포로부터 RNA를 채취하였다(도 15 참조). RNA를 cDNA로 역전사시키고 cDNA를 당 분야에 공지된 표준 방법을 사용하여 PCR 증폭시켰다. 도 15에서, 하부 왼쪽 패널은 순방향 프라이머-tggcggcggttttcattat(SEQ ID NO: 105) 및 역방향 프라이머-TTCGATCCGTAATGATTGTCTAGCC(SEQ ID NO: 106)를 사용하여 엑손 44에서 엑손 52로의 PCR 증폭을 나타낸다. 모 밴드는 948nt인 것으로 예측되고, 결실 밴드는 약 715nt로 예측된다. 도 15는 SaCas9 및 두 gRNA로 처리된 세포에서만 결실 밴드를 보여준다. 하부 오른쪽 패널은 cDNA의 약 14% 편집을 나타내는 ddPCR을 보여준다. DMD 환자 근모세포의 cDNA에서의 시험관내 엑손 47 내지 52 접합을 시퀀싱하였다(도 16 참조). 도 16에서, 처리하지 않은 세포(대조군 세포 Δ48-50)로부터의 밴드의 서열은 엑손 47 내지 51이 예상대로 결합되었음을 나타내었지만, 처리된 세포(Δ48-50 + Δ51)의 결실 밴드는 엑손 47 내지 52의 접합을 나타냈다. 따라서 엑손 51과 게놈 DNA 수준을 목표로 하는 개시된 시스템의 뚜렷한 부족이 전사를 통해 수행되고 있었다.

[0258] 실시예 2

[0259] Δ52/*mdx* 마우스의 생성

[0260] 도 17은 Δ52/*mdx* 마우스의 디자인을 보여준다. hDMD/*mdx* 마우스는 Leiden University에서 얻었고, 엑손 52가 제거되고, 결실로 인해 해독 틀-변경 이동 및 DMD 유전자형이 나타나는 DMD 관련 모델을 생성하도록 조작되었다. hDMD/*mdx* 마우스는 마우스 디스트로핀이 발현되지 않도록 *mdx* 백그라운드에서 5번 염색체에 있는 전장 야생형 인간 디스트로핀 유전자를 포함한다.

[0261] DMD에 대한 관련 모델을 생성하기 위해 SpCas9 CRISPR/Cas9 편집 시스템과 gRNA를 사용하여 인간 디스트로핀 유전자의 엑손 52를 표적화 및 결실시켰다. 엑손 52의 상류 및 하류를 표적화하는 다양한 gRNA를 검사하고 측량 검정을 사용하여 검증하였다(도 18 참조). 상류 gRNA의 경우, 순방향 프라이머-ctccggaatgtctccatttg(SEQ ID NO: 87) 및 역방향 프라이머-TTGTGTGTCCTATGCTTGT(SEQ ID NO: 107)를 사용하였으며, 모 밴드 크기는 402 nt였다. JCR94(AACAAATATCCCTTAGTATC(SEQ ID NO: 41))의 경우, 예상 절단 크기는 243nt 및 159nt이었다. 하류 gRNA의 경우, 순방향 프라이머 CAACGCTGAAGAACCTGAT(SEQ ID NO: 108) 및 역방향 프라이머-atgaggagagactggcatc(SEQ ID NO: 88)를 사용하였으며, 모 밴드 크기는 509nt이었다. JCR99(AATGATTCTTCTATTCAA(SEQ ID NO: 42))의 경우, 예상 절단 크기는 346 nt 및 163 nt였다. 한 쌍의 gRNA를 순방향 프라이머-ctccggaatgtctccatttg(SEQ ID NO: 87) 및 역방향 프라이머-atgaggagagactggcatc(SEQ ID NO: 88)를 사용하여 JCR94 및 JCR99를 포함하는 HEK293T 세포의 게놈 DNA에서 엑손 51의 결실을 검출함으로써

시험하고 확인하였다(도 19 참조). 모 밴드 크기는 718nt이고, 결실 밴드는 392nt이었으며, gRNA 사이의 결실은 326nt이었다. JCR94와 JCR99 쌍은 게놈 편집 시스템에서 사용하여, $\Delta 52/mdx$ 마우스를 생성하였다. 구체적으로, 마우스는 JCR94 gRNA, JCR99 gRNA 및 SaCas9 mRNA를 마우스 배아에 주사함으로써 생성되었다.

[0262] 도 20은 BAC 재조합 서비스를 포함하는 DNA 미세주입 프로토콜을 나타낸다. 1일째: 임신한 암말을 혈청 생식선 자극 호르몬으로 복강내 처리하여 배란을 유도하였다. 3일째: 임신한 암말을 인간 융모막성 생식선자극 호르몬으로 복강내 처리하였다. 제1 라운드에서는 471개의 배아가 생산되었지만 5개의 플러그만 시각화되었다. 150개의 수정란이 사용되었다(14명의 암컷은 과배란을 동반했다). 전뇌 주입은 Cas9 50ng 미만, 각 20ng 미만으로 수행하였다. 도 21은 트랜스제닉 마우스를 생성하기 위한 마우스 번식 프로토콜을 나타낸다.

[0263] 파운더 마우스는 다음의 유전자형 분석 프로토콜을 사용하여 유전자형을 분석하였다. 게놈 DNA(gDNA)는 DNEasy Blood 및 Tissue kit(Qiagen)를 사용하여 마우스 꼬리 부분(tail snip)에서 추출하였다. 각 새끼를 유전자형 분석하기 위해 AccuPrime HiFi Taq 키트를 사용하여 gDNA를 다음과 같이 증폭시켰다: i. 100 ng gDNA; ii. 2.5 μ L AccuPrime 완충액 II; iii. 0.1 μ L AccuPrime HiFi Taq; iv. 1 μ L JRH261(ctccggaatgtctccatttg(SEQ ID NO: 87))(10 μ M); v. 1 μ L JRH264(atgaggagagagactggcatc(SEQ ID NO: 88))(10 μ M); 및 vi. 총 25 μ L의 물. 반응은 다음과 같이 써모사이클러에서 수행되었다: i. 4분간 95도; ii. 30초동안 95도; iii. 30초 동안 52도; iv. 1:00분간 68도; v. 주기 단계 ii-iv 35회; 및 영원히 4도. PCR 반응은 겔상에서 분리되었다(도 22). 예상된 밴드 크기는 결실이 존재하지 않는 경우(즉, 엑손 52가 여전히 존재하는 경우) 718nt이고, 엑손 52의 결실이 있는 경우 약 392nt이다. 도 22에 도시된 바와 같이, 파운더 마우스 7, 63, 및 76은 엑손 52 결실을 가졌다.

[0264] 증폭된 밴드를 JRH264 프라이머(도 23 참조)를 사용하여 시퀀싱하여 표적화된 영역의 재결합된 말단을 시퀀싱하였다. 도 23은 시퀀싱된 영역을 도시하고, 볼드체, 밑줄친 글씨 및 보통 글자는 천연 서열을 나타내고, 이탤릭체 문자는 삽입 또는 결실을 나타낸다. 예상된 순서("텔타 52")에서, 굵은 글씨는 밑줄이 그어진 문자에 결합된다. 파운더 마우스에서, 이 영역에 삽입(이탤릭체로된 문자) 및 결실(하이픈)이 있었다.

[0265] 수컷 파운더 마우스를 mdx/mdx 암컷과 교배시켜 키메라를 번식시켰다(도 24). mdx/mdx 암컷과 함께 파운더 수컷 76 또는 파운더 수컷 63으로부터 생성된 리터를 스크리닝하고, 도 22(도 25 및 도 26 각각)에서 사용된 조건을 사용하여, 엑손 52 결실에 대해 유전자형을 분석하였다. 엑손 52가 결실되면, 예상된 밴드 크기는 약 392nt이었다. 엑손 52가 존재한다면, 예상 밴드 크기는 약 718nt이었다. (파운더 수컷 63 + mdx/mdx 암컷 번식 쌍으로부터의) 새끼 54497 및 54498은 엑손 52 결실을 가지고, 시퀀싱되었다(도 27). 새끼 54497과 54498은 392 bp 시퀀싱 읽기에서 서로 92.86%의 동일성을 나타내었고, indels은 동일했다.

[0266] 건강한 hDMD/ mdx 마우스와 $\Delta 52/mdx$ 마우스에서 디스트로핀의 발현을 형광 면역조직화학을 이용하여 비교하였다. 도 28에 도시된 바와 같이, $\Delta 52/mdx$ 마우스 새끼 54497 및 54498에는 디스트로핀 단백질이 결핍되어 있었다. 심장 염색의 경우 라미닌 프로브 노출은 100ms 였고, 디스트로핀 노출은 900ms였다. TA 근육 샘플의 경우, 라미닌과 디스트로핀 프로브에 대한 노출은 2.0초였다. $\Delta 52/mdx$ 마우스가 디스트로핀 단백질이 결핍되었음을 보여주는 도 29를 참조한다. 심장 및 TA 근육 모두에서, 디스트로핀 발현은 $\Delta 52/mdx$ 마우스에서 소실되었다. TA 근육에는 몇 가지 자발적인 복귀돌연변이체 섬유(랜덤 스플라이스 이벤트, 또는 체세포 돌연변이)가 있었지만, 이는 mdx 마우스 모델과 일치한다. 웨스턴 블롯팅은 $\Delta 52/mdx$ 마우스가 DMD 유전자형과 일치하는 디스트로핀 단백질이 부족한 반면, 건강한 hDMD/ mdx 마우스는 디스트로핀을 발현함을 나타냈다. 도 30을 참조한다.

[0267] $\Delta 52/mdx$ 마우스는 오픈 필드 테스트의 처음 5분 후에 mdx 마우스와 비슷한 수준의 활성을 보였다. 오픈 필드 테스트에서, 마우스는 30분 동안 오픈 필드 경기장(20 x 20 x 30 cm)을 자유롭게 탐험하게 하였다. 동물의 활동 및 위치는 융합 활성(Fusion Activity) 소프트웨어(버전 5.3, Omnitech, Columbus, OH)를 실행하는 컴퓨터에 인터페이스된 적외선 다이오드(x, y 및 z 축)를 사용하여 자동으로 모니터링되었다. 도 31은 이동 및 탐색에 의해 지시된 mdx 마우스 및 hDMD/ mdx 마우스와 비교된 $\Delta 52/mdx$ 마우스의 전반적인 활성을 나타낸다. 이동한 거리와 직립 수직 활동은 각각 왼쪽 패널과 오른쪽 패널에 표시된다.

[0268] **실시예 3**

[0269] **엑손 51 제거에 의한 디스트로핀 복원**

[0270] 엑손 51의 제거는 $\Delta 52/mdx$ 마우스에서 베커형 근이영양증(BMD)-유사 유전자형을 생성할 수 있으며, 이론적으로 디스트로핀 발현을 회복시킨다. $\Delta 52/mdx$ 마우스를 사용하여 개시된 CRISPR/Cas9-기반 유전자 편집 시스템을 사용한 엑손 51의 제거에 의한 디스트로핀 발현의 회복을 입증하였다. 도 32는 SaCas9 및 gRNA를 사용하여, 인트론 영역의 엑손 51의 상류 및 하류의 gRNA를 표적화하여 엑손 51을 건너뛰는 교정 전략을 도시한다.

- [0271] 표준: gRNA JCR179(상류) 또는 JCR183(하류)(각각 SEQ ID NO: 37 및 SEQ ID NO: 38) 및 SaCas9를 함유하는 플라스미드를 DMD 환자 근모세포 내로 전기천공시켰다. 분화된 세포로부터 단백질을 채취하고, 디스트로핀 항체로 웨스턴 블롯을 사용하여 분석하였다. 도 33은 모든 3개의 성분(즉, gRNA 및 SaCas9 모두)으로 처리된 세포가 디스트로핀 발현을 나타냈기 때문에 게놈 DNA가 편집되어 디스트로핀 단백질을 복원할 수 있음을 나타낸다.
- [0272] 그런 다음 시스템을 $\Delta 52/mdx$ 마우스로 검사했다. 도 34는 gRNA 및 SaCas9 시스템을 사용하여 $\Delta 52/mdx$ 마우스를 치료하기 위한 실험 설계를 나타내며, 실험 설계에 사용된 2개의 바이러스 벡터의 개략도를 포함한다. AAV8 제조합 바이러스 구축물은 벡터 PT366-179(SEQ ID NO: 39) 및 PT366-183(SEQ ID NO: 40), 및 바이러스 입자를 생산하는 당 분야에 공지된 방법을 사용하여 생성하였다. 상기 바이러스 벡터(AAV8)는 2개의 바이러스 입자로서 생체내에서 동시 전달되었다. 각각의 바이러스성 입자는 SaCas9 및, gRNAs 중 하나를 함유하였다(도 3 참조). $\Delta 52/mdx$ 마우스를 AAV8 제조합 바이러스 구축물의 5E11로 처리하였다. 바이러스는 오른쪽 TA 근육에 근육내 주사하고, 왼쪽 TA 근육은 반대측성 대조군으로서 PBS를 주사하였다. 치료 후, 좌우 TA 근육을 모두 제거하고, 각각의 섹션을 게놈 DNA 분석을 위해 취했다. 도 35에 도시된 바와 같이, PCR은 왼쪽 젤상의 처리된 오른쪽 TA 근육에서 관심의 영역 및 결실 밴드에 걸쳐 수행되어, 유전자 편집의 일정 수준을 나타낸다. 결실 밴드가 시퀀싱되었고, 우성 생성물은 각 gRNA의 PAM으로부터 3개의 염기쌍의 예상된 결찰이었다. 도 35는 오른쪽 TA 근육에서 생체내 엑손 51 결실을 나타낸다.
- [0273] 유사하게, 치료된 오른쪽 TA 및 대조군 왼쪽 TA 근육의 절편을 분석하여 편집이 RNA를 통해 진행되었는지 여부를 결정하였다. PCR은 cDNA에서 엑손 50 내지 53에 걸쳐 수행되었다. 도 36의 오른쪽 겔에 도시된 바와 같이, 처리된 샘플 중 2개에는 엑손 51 및 52가 없다. 마우스가 이미 엑손 52가 없고 CRISPR/Cas9-기반의 유전자 편집 시스템이 엑손 51을 제거했다면 결실 밴드가 시퀀싱되었고, 우성 생성물은 엑손 50 내지 53의 결찰이었다(오른쪽 하단 서열 크로마토그램 참조). 도 36은 오른쪽 TA 근육에서 생체내 엑손 51 결실을 나타낸다.
- [0274] 도 37은 $\Delta 52/mdx$ 마우스의 대조군 PBS 주사한 왼쪽 TA 근육에 디스트로핀이 거의 존재하지 않는다는 것을 나타내는 대표적인 형광 면역조직화학 염색을 나타낸다. 대조군 왼쪽 TA에서 녹색으로 염색하는 어느 정도의 디스트로핀 염색은 복원 섬유 또는 죽은 세포에 기인할 수 있는데, 때로는 녹색으로 얼룩지게된다. 오른쪽 사진과 같이, 처리된 오른쪽 TA 근육에서 녹색 디스트로핀 염색이 명백하게 증가한다. 도 37은 처리된 TA 근육에서의 생체내 디스트로핀 단백질 복원을 나타낸다.
- [0275] 3마리의 시험 마우스로부터 좌우 TA 근육으로부터 단백질을 추출하고 웨스턴 블랏 분석을 수행하였다. 도 38은 처리된 TA 근육에서 생체내 디스트로핀 단백질 복원을 나타낸다. 3개의 오른쪽 TA 근육 모두 다양한 수준의 디스트로핀 단백질 발현을 나타낸 반면, 대조군 왼쪽 TA 근육에서는 단백질 발현이 관찰되지 않았다. 마우스 1의 오른쪽 TA에서 단백질 발현이 가장 강한 반면, 마우스 2와 3은 희미하지만 그럼에도 불구하고 밴드를 나타냈다. $\Delta 52/mdx$ 마우스에서 디스트로핀 단백질 발현을 어느 정도 복원하기 위해, 개시된 CRISPR/Cas9-기반 유전자 편집 시스템이 생체내에서 연구되었다.
- [0276] 마우스 생리학 테스트. 처리된 마우스(모든 수컷 hDMD- $\Delta 52(het)/mdx(hemi)$ 마우스)를 SaCas9- 및 gRNA-함유 AAV8(n = 10) 또는 AAV9(n = 10) 제조합 바이러스 구축물로 처리하고, 처리하지않은 마우스(hDMD- $\Delta 52/mdx$)(n = 10)와 비교하였다. 벡터 PT366-179(SEQ ID NO: 39) 및 PT366-183(SEQ ID NO: 40)을 사용하고, 당 분야에 공지된 방법을 사용하여 AAV 제조합 바이러스 구축물을 생성시켰다. 처리된 마우스는 6주령과 8주령 사이에 꼬리 정맥 내로 주사된 바이러스 200 μ L를 가졌다. 그 마우스를 8주 후에 테스트하였다.
- [0277] 오픈 필드 거리 테스트. 30분 동안 마우스를 오픈 필드 아레나를 자유롭게 탐험하도록 하였다. 융합 활성 소프트웨어를 실행하는 컴퓨터에 인터페이스된 적외선 다이오드를 사용하여 동물의 활동 및 위치를 자동으로 모니터링했다. 데이터는 연속적으로 수집되어 5분 간격으로 분류된다. 도 39는 치료되지 않은 16주령 마우스와 비교하여 8주령에 처리된 16주령 마우스에서 이동된 총 거리에 대한 모든 시간 지점의 평균을 나타낸다. 16주 후에 총 양육 자세에 대한 모든 시간 지점의 평균을 도 40에 나타낸다. 통계는: 일방향 ANOVA, 각 컬럼 평균을 처리되지 않은 평균과 비교됨, Dunnett 사후 검정(평균 \pm SEM). AAV8 및 AAV9 처리된 마우스는 처리하지않은 연령의 매칭된 마우스보다 통계적으로 유의하게 더 거리 이동한 것을 나타냈다(통계적으로 유의함). 처리된 모든 마우스는 처리되지 않은 연령의 매칭된 대조군과 비교하여 양육 자세의 통계적으로 유의한 증가된 양을 나타낸다.
- [0278] 그립 강도. 16주간 처리되지않은 및 처리된 마우스의 그립 강도를 시험하였다. 마우스는 그립 강도를 테스트하기 위해 전후방 피트에 대해 각각 3-5번의 시도를 받았다. 평균 시험은 도 41에 나타내었다. 그립 강도는 그램 중량(gram force)으로 보고된다. 도 41에 도시된 바와 같이, AAV9 처리된 마우스는 처리되지 않은 연령의 매칭된 마우스와 비교하여, 앞발에서 통계적으로 유의미한 증가된 그립 강도를 나타내었다. 통계는 다음과 같았다 :

양방향 ANOVA, Tukey의 사후 검증.

- [0279] cDNA PCR. 마우스의 심장 조직은 RNEasy Plus Universal 미니 키트(Qiagen)를 사용하여 처리했다. 생성된 RNA를 SuperScript VILO cDNA 합성 키트를 사용하여 cDNA로 역전사시켰다. AccuPrime DNA 폴리머라제 및 엑손 48 (순방향 프라이머: gtttccagagctttacctgagaa(SEQ ID NO: 89)) 및 엑손 54(역방향 프라이머: CTTTATGAATGCTTCTCCAAG(SEQ ID NO: 90))의 프라이머를 사용하여 1 μ L의 cDNA를 PCR 증폭시켰다. 예상된 밴드 크기는 결실이 존재하지 않는 경우(즉, 엑손 52가 여전히 존재하는 경우) 997nt이고, 엑손 52의 결실이 있는 경우 약 764nt이다. 도 42는 36 내지 50시간된 AAV9("JA10(P2 AAV9)")를 안면정맥을 통해 P2 마우스에 주사하고, 및 꼬리 정맥을 통해 AAV8("JA11(TV AAV8)")의 3.3-7.7E12 또는 AAV9("JA12(TV AAV9)")의 4.3-7.5E12를 성체 마우스에 주사한 것을 나타낸다. 도 42에 도시된 바와 같이, P2 마우스(48 내지 54시간된 마우스) AAV9 처리된 마우스 및 AAV8 및 AAV9 성체 처리된 마우스에서 다양한 정도의 편집이 발생하였으며, 이는 엑손 53에 결합하는 tttctgtgattttcttttgattg(SEQ ID NO: 109) 프라이머를 사용하여 결실 밴드를 시퀀싱함으로써 추가로 확인되었다. 도 43은 Ja10 마우스 1의 결실 밴드의 서열에서 엑손 51 및 52의 결실을 나타내는 대표적인 크로마토그램을 나타낸다.
- [0280] 상기 상세한 설명 및 첨부되는 실시예는 단지 예시적인 것이며, 첨부되는 청구범위 및 이의 균등부에 의해서만 정의되는 본 발명의 범위에 대한 제한으로 간주되어서는 안 됨이 이해된다.
- [0281] 개시된 구현예에 대한 다양한 변화 및 변경은 당업자에게 자명할 것이다. 비제한적으로 본 발명의 화학적 구조, 치환체, 유도체, 중간체, 합성, 조성물, 제형물, 또는 이용 방법에 관한 것들을 포함하는 이러한 변화 및 변경은 이의 정신 및 범위에서 벗어나지 않고 수행될 수 있다.
- [0282] 완전성의 이유로, 본 발명의 다양한 양태를 다음의 번호를 매긴 항으로 나타낸다:
- [0283] 항 1. SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 42에 개시된 뉴클레오타이드 서열 또는 이의 상보체를 포함하는 표적화 도메인을 포함하는 가이드 RNA(gRNA).
- [0284] 항 2. 제1 gRNA 및 제2 gRNA를 포함하는 DNA 표적화 조성물로서, 상기 제1 gRNA 분자 및 제2 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 42에 개시된 뉴클레오타이드 서열 또는 이의 상보체를 포함하는 표적화 도메인을 포함하며, 상기 제1 gRNA 분자 및 제2 gRNA 분자는 상이한 표적화 도메인을 포함하는 DNA 표적화 조성물.
- [0285] 항 3. 제2항에 있어서, 상기 제1 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, 또는 SEQ ID NO: 41이며, 상기 제2 gRNA 분자가 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, 또는 SEQ ID NO: 42인, DNA 표적화 조성물.
- [0286] 항 4. 제2항 또는 제3항에 있어서, 상기 제1 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14 및 SEQ ID NO: 15로 구성된 그룹으로부터 선택되고, 상기 제2 gRNA 분자가 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, 및 SEQ ID NO: 19로 구성된 그룹으로부터 선택되는, DNA 표적화 조성물.
- [0287] 항 5. 제2항 내지 제4항 중 어느 한 항에 있어서, 상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가: (i) SEQ ID NO: 1에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ii) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iii) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iv) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (v) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는

표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vi) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (viii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ix) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (x) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (xi) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및 (xii) SEQ ID NO: 41에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 42에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자로 구성된 그룹으로부터 선택되는, DNA 표적화 조성물.

- [0288] 항 6. 제2항 내지 제5항 중 어느 한 항에 있어서, 클러스터링된, 규칙적으로 산재된, 짧은 회문 반복(Clustered Regularly Interspaced Short Palindromic Repeats) 관련(associated)(Cas) 단백질을 추가로 포함하는, DNA 표적화 조성물.
- [0289] 항 7. 제6항에 있어서, 상기 Cas 단백질이 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이스 인접 모티프(PAM)를 인식하는 Cas9 분자를 포함하는, DNA 표적화 조성물.
- [0290] 항 8. 제6항 또는 제7항에 있어서, 상기 Cas 단백질이 SEQ ID NO: 45의 아미노산 서열을 갖는 *스타필로코커스 아우레우스*(*Staphylococcus aureus*) Cas9 분자를 포함하는, DNA 표적화 조성물.
- [0291] 항 9. 제2항 내지 제8항 중 어느 한 항에 있어서, 상기 DNA 표적화 조성물이 SEQ ID NO: 83의 뉴클레오타이드 서열, SEQ ID NO: 84의 뉴클레오타이드 서열, SEQ ID NO: 37의 뉴클레오타이드 서열, 및/또는 SEQ ID NO: 38의 뉴클레오타이드 서열을 포함하는, DNA 표적화 조성물.
- [0292] 항 10. 제1항의 gRNA 분자 또는 제2항 내지 제9항 중 어느 한 항의 DNA 표적화 조성물을 포함하는 단리된 폴리뉴클레오타이드.
- [0293] 항 11. 제1항의 gRNA, 제2항 내지 제9항 중 어느 한 항의 DNA 표적화 조성물 또는 제10항의 단리된 폴리뉴클레오타이드를 포함하는 벡터.
- [0294] 항 12. 제6항 내지 제9항 중 어느 한 항의 DNA 표적화 조성물을 포함하는 벡터.
- [0295] 항 13. (a) 제1 가이드 RNA(gRNA) 분자, (b) 제2 gRNA 분자, 및 (c) NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이스 인접 모티프(PAM)를 인식하는 적어도 하나의 Cas9 분자를 코딩하는 벡터로서, 상기 제1 gRNA 분자 및 제2 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, SEQ ID NO: 41, SEQ ID NO: 42에 개시된 뉴클레오타이드 서열 또는 이의 상보체를 포함하는 표적화 도메인을 포함하고, 상기 제1 gRNA 분자 및 제2 gRNA 분자가 상이한 표적화 도메인을 포함하는, 벡터.
- [0296] 항 14. 제13항에 있어서, 상기 벡터는 인간 DMD 유전자의 엑손 51 측면의 제1 및 제2 인트론에 제1 및 제2 이중가닥 절단부를 형성하도록 구성되는, 벡터.
- [0297] 항 15. 제13항 또는 제14항에 있어서, 상기 제1 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, 또는 SEQ ID NO: 41이고, 상기 제2 gRNA 분자가 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, 또는 SEQ ID NO: 42인, 벡터.
- [0298] 항 16. 제13항 내지 제15항 중 어느 한 항에 있어서, 상기 제1 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14 및 SEQ ID NO: 15로 구성된 그룹으로부터 선택되고, 상기 제2 gRNA

분자는 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18 및 SEQ ID NO: 19로 구성된 그룹으로부터 선택되는, 벡터.

- [0299] 항 17. 제13항 내지 제16항 중 어느 한 항에 있어서, 상기 제1 gRNA 분자 및 제2 gRNA 분자가: (i) SEQ ID NO: 1에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ii) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iii) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iv) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (v) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vi) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (viii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ix) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (x) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및 (xi) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자로 구성된 그룹으로부터 선택되는, 벡터.
- [0300] 항 18. 제11항 내지 제17항 중 어느 한 항에 있어서, 상기 벡터가 바이러스 벡터인, 벡터.
- [0301] 항 19. 제18항에 있어서, 상기 벡터가 아데노-연관 바이러스(Adeno-associated virus, AAV) 벡터인, 벡터.
- [0302] 항 20. 제19항에 있어서, 상기 AAV 벡터가 AAV8 벡터 또는 AAV9 벡터인, 벡터.
- [0303] 항 21. 제11항 내지 제20항 중 어느 한 항에 있어서, 상기 벡터는 제1 gRNA 분자, 제2 gRNA 분자 및/또는 Cas9 분자를 코딩하는 뉴클레오타이드 서열에 작동가능하게 연결된 조직-특이적 프로모터를 포함하는, 벡터.
- [0304] 항 22. 제21항에 있어서, 상기 조직-특이적 프로모터가 근육 특이적 프로모터인, 벡터.
- [0305] 항 23. 제1항의 gRNA, 제2항 내지 제9항 중 어느 한 항의 DNA 표적화 조성물, 제10항의 단리된 폴리뉴클레오타이드 또는 제11항 내지 제22항 중 어느 한 항의 벡터를 포함하는 세포.
- [0306] 항 24. 제1항의 gRNA, 제2항 내지 제9항 중 어느 한 항의 DNA 표적화 시스템, 제10항의 단리된 폴리뉴클레오타이드, 제11항 내지 제22항 중 어느 한 항의 벡터, 또는 제23항의 세포 및 선택적으로 사용 지침서를 포함하는 키트.
- [0307] 항 25. 제1항의 gRNA, 제2항 내지 제9항 중 어느 한 항의 DNA 표적화 시스템, 제10항의 단리된 폴리뉴클레오타이드 또는 제11항 내지 제22항 중 어느 한 항의 벡터를 세포에 투여하는 단계를 포함하는, 세포내 돌연변이체 디스트로핀 유전자를 교정하는 방법.
- [0308] 항 26. 대상체의 돌연변이체 디스트로핀 유전자를 게놈 편집하는 방법으로서, 상기 방법은 제1항의 gRNA, 제2항 내지 제9항 중 어느 한 항의 DNA 표적화 시스템, 제10항의 단리된 폴리뉴클레오타이드, 제11항 내지 제22항 중 어느 한 항의 벡터 또는 제23항의 세포를 포함하는 게놈 편집 조성물을 대상체에 투여하는 단계를 포함하는, 방법.
- [0309] 항 27. 제26항에 있어서, 상기 게놈 편집 조성물이 대상체에게 근육 내, 정맥 내 또는 이들의 조합으로 투여되는, 방법.
- [0310] 항 28. 제25항 내지 제27항 중 어느 한 항에 있어서, 돌연변이체 디스트로핀 유전자를 교정하는 단계는 뉴클레

아제-매개된 비-상동성 말단 결합을 포함하는, 방법.

- [0311] 항 29. 돌연변이체 디스트로핀 유전자를 갖는, 치료를 필요로 하는 대상체를 치료하는 방법으로서, 상기 방법은 제1항의 gRNA, 제2항 내지 제9항 중 어느 한 항의 DNA 표적화 시스템, 제10항의 단리된 폴리뉴클레오타이드, 제11항 내지 제22항 중 어느 한 항의 벡터 또는 제23항의 세포를 대상체에 투여하는 단계를 포함하는, 방법.
- [0312] 항 30. 대상체에서 돌연변이체 디스트로핀 유전자를 게놈 편집하기 위한, 변형된 아데노-연관 바이러스 벡터로서, 제1항의 gRNA를 코딩하는 제1 폴리뉴클레오타이드 서열, 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25) 중 어느 하나의 프로토스페이서 인접 모티프(PAM)를 인식하는 Cas9 분자를 코딩하는 제2 폴리뉴클레오타이드 서열을 포함하는, 변형된 아데노-연관 바이러스 벡터.
- [0313] 항 31. 제30항에 있어서, 상기 변형된 아데노-연관 바이러스 벡터는 SEQ ID NO: 39 또는 SEQ ID NO: 40에 개시된 뉴클레오타이드 서열을 포함하는, 변형된 아데노-연관 바이러스 벡터.
- [0314] 항 32. 엑손 51을 포함하는 디스트로핀 유전자의 절편을 결실시키기 위한 조성물로서, 상기 조성물은: (a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및 (b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터를 포함하며, 상기 제1 및 제2 gRNA 분자 각각은 19 내지 24개의 뉴클레오타이드 길이의 표적화 도메인을 가지며, 제1 벡터 및 제2 벡터는 인간 DMD 유전자의 엑손 51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 디스트로핀 유전자의 절편을 결실시키도록 구성되는, 조성물.
- [0315] 항 33. 제32항에 있어서, 상기 절편은 약 50 염기쌍 내지 약 2,000 염기쌍의 길이를 갖는, 조성물.
- [0316] 항 34. 제33항에 있어서, 상기 절편은 약 118 염기쌍, 약 233 염기쌍, 약 326 염기쌍, 약 766 염기쌍, 약 805 염기쌍 또는 약 1611 염기쌍의 길이를 갖는, 조성물.
- [0317] 항 35. 제32항 내지 제34항 중 어느 한 항에 있어서, 상기 제1 Cas9 분자와 상기 제2 Cas9 분자는 동일한, 조성물.
- [0318] 항 36. 제35항에 있어서, 상기 제1 Cas9 분자 및 상기 제2 Cas9 분자는 *스타필로코커스 아우레우스* (*Staphylococcus aureus*) Cas9 분자인 조성물.
- [0319] 항 37. 제36항에 있어서, 상기 제1 Cas9 분자 및 상기 제2 Cas9 분자는 돌연변이체 *스타필로코커스 아우레우스* (*Staphylococcus aureus*) Cas9 분자인 조성물.
- [0320] 항 38. 제32항 내지 제34항 중 어느 한 항에 있어서, 상기 제1 Cas9 분자와 상기 제2 Cas9 분자가 상이한, 조성물.
- [0321] 항 39. 제38항에 있어서, 상기 제1 Cas9 분자 또는 상기 제2 Cas9 분자가 *스타필로코커스 아우레우스* (*Staphylococcus aureus*) Cas9 분자인 조성물.
- [0322] 항 40. 제32항 내지 제39항 중 어느 한 항에 있어서, 상기 제1 Cas9 분자 및/또는 제2 Cas9 분자는 SEQ ID NO: 45의 아미노산 서열을 갖는 SaCas9 분자를 포함하는, 조성물.
- [0323] 항 41. 제32항 내지 제40항 중 어느 한 항에 있어서, 상기 제1 벡터 및/또는 제2 벡터는 바이러스 벡터인, 조성물.
- [0324] 항 42. 제41항에 있어서, 상기 제1 벡터 및/또는 상기 제2 벡터는 아데노-연관 바이러스(Adeno-associated virus, AAV) 벡터인, 조성물.
- [0325] 항 43. 제42항에 있어서, 상기 AAV 벡터는 AAV8 벡터 또는 AAV9 벡터인, 조성물.
- [0326] 항 44. 제32항 내지 제43항 중 어느 한 항에 있어서, 상기 디스트로핀 유전자가 인간 디스트로핀 유전자인, 조성물.
- [0327] 항 45. 제32항 내지 제44항 중 어느 한 항에 있어서, 상기 제1 gRNA 분자 및 제2 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15,

SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19에 개시된 뉴클레오타이드 서열 또는 이의 상보체를 포함하는 표적화 도메인을 포함하며, 상기 제1 gRNA 분자 및 제2 gRNA 분자는 상이한 표적화 도메인을 포함하는, 조성물.

[0328] 항 46. 제32항 내지 제45항 중 어느 한 항에 있어서, 상기 제1 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 3, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, 또는 SEQ ID NO: 15이며, 상기 제2 gRNA 분자가 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, 또는 SEQ ID NO: 19인, 조성물.

[0329] 항 47. 제32항 내지 제46항 중 어느 한 항에 있어서, 상기 제1 gRNA 분자가 SEQ ID NO: 1, SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14 및 SEQ ID NO: 15로 구성된 그룹으로부터 선택되고, 상기 제2 gRNA 분자가 SEQ ID NO: 2, SEQ ID NO: 4, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, 및 SEQ ID NO: 19로 구성된 그룹으로부터 선택되는, 조성물.

[0330] 항 48. 제32항 내지 제47항 중 어느 한 항에 있어서, 상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가: (i) SEQ ID NO: 1에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ii) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iii) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iv) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (v) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vi) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (viii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ix) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (x) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및 (xi) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자로 구성된 그룹으로부터 선택되는, 조성물.

[0331] 항 49. 제32항 내지 제48항 중 어느 한 항에 있어서, 상기 제1 벡터는 SEQ ID NO: 39에 개시된 뉴클레오타이드 서열을 포함하고, 상기 제2 벡터는 SEQ ID NO: 40에 개시된 뉴클레오타이드 서열을 포함하는, 조성물.

[0332] 항 50. 약제에 사용하기 위한, 제32항 내지 제49항 중 어느 한 항의 조성물.

[0333] 항 51. 뒤시엔느 근이영양증(Duchenne Muscular Dystrophy)의 치료에 사용하기 위한, 제32항 내지 제50항 중 어느 한 항의 조성물.

[0334] 항 52. 제32항 내지 제51항 중 어느 한 항의 조성물을 포함하는 세포.

[0335] 항 53. 세포내에 돌연변이체 디스트로핀 유전자를 교정하는 방법으로서, 상기 방법은: (a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및 (b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터를 세포에 투여하는 단계를 포함하며, 상기 제1 gRNA 및 제2 gRNA 분자 각각은 19 내지 24개의 뉴클레오타이드 길이의 표적화 도메인을 가지며, 상기 벡터는 인간 디스트로핀 유전자의 엑손

51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 디스트로핀 유전자의 절편을 결실시키도록 구성되는, 방법.

- [0336] 항 54. 제53항에 있어서, 상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가: (i) SEQ ID NO: 1에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ii) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iii) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iv) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (v) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vi) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (viii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ix) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (x) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및 (xi) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자로 구성된 그룹으로부터 선택되는, 방법.
- [0337] 항 55. 제53항 또는 제54항에 있어서, 상기 돌연변이체 디스트로핀 유전자는 조기 중지 코돈, 손상된 해독 틀, 비정상 스플라이스 수신체 부위 또는 비정상 스플라이스 공여체 부위를 포함하는, 방법.
- [0338] 항 56. 제53항 내지 제55항 중 어느 한 항에 있어서, 상기 돌연변이체 디스트로핀 유전자가 조기 중지 코돈 및 절단된 유전자 생성물을 유발하는 틀 이동 돌연변이를 포함하는, 방법.
- [0339] 항 57. 제53항 내지 제55항 중 어느 한 항에 있어서, 상기 돌연변이체 디스트로핀 유전자가 해독 틀을 손상시키는 하나 이상의 엑손의 결실을 포함하는, 방법.
- [0340] 항 58. 제53항 내지 제57항 중 어느 한 항에 있어서, 상기 돌연변이체 디스트로핀 유전자의 교정이 조기 중지 코돈의 결실, 손상된 해독 틀의 교정, 또는 스플라이스 수신체 부위의 손상 또는 스플라이스 공여체 서열의 손상에 의한 스플라이싱의 조절을 포함하는, 방법.
- [0341] 항 59. 제53항 내지 제58항 중 어느 한 항에 있어서, 상기 돌연변이체 디스트로핀 유전자의 교정은 엑손 51의 결실을 포함하는, 방법.
- [0342] 항 60. 제53항 내지 제59항 중 어느 한 항에 있어서, 상기 돌연변이체 디스트로핀 유전자의 교정이 상동성-유도된 보수를 포함하는, 방법.
- [0343] 항 61. 제60항에 있어서, 세포에 공여체 DNA를 투여하는 단계를 추가로 포함하는, 방법.
- [0344] 항 62. 제53항 내지 제61항 중 어느 한 항에 있어서, 상기 돌연변이체 디스트로핀 유전자의 교정이 뉴클레아제 매개된 비-상동성 말단 결합을 포함하는, 방법.
- [0345] 항 63. 제53항 내지 제62항 중 어느 한 항에 있어서, 상기 세포가 근모세포 세포인, 방법.
- [0346] 항 64. 제53항 내지 제63항 중 어느 한 항에 있어서, 상기 세포가 뒤시엔느 근이영양증을 앓고있는 대상체로부터의 세포인, 방법.
- [0347] 항 65. 제53항 내지 제64항 중 어느 한 항에 있어서, 상기 세포가 뒤시엔느 근이영양증을 앓고있는 인간 대상체로부터의 근모세포인, 방법.

- [0348] 항 66. 제53항 내지 제65항 중 어느 한 항에 있어서, 상기 제1 gRNA 분자 및 제2 gRNA 분자는: (i) SEQ ID NO: 1에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ii) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및 (iii) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자로 구성된 그룹으로부터 선택되는, 방법.
- [0349] 항 67. 돌연변이체 디스트로핀 유전자를 갖는, 필요로 하는 대상체를 치료하는 방법으로서, 상기 방법은: (a) 제1 가이드 RNA(gRNA) 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제1 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제1 벡터, 및 (b) 제2 gRNA 분자를 코딩하는 폴리뉴클레오타이드 서열 및 NNGRRT(SEQ ID NO: 24) 또는 NNGRRV(SEQ ID NO: 25)의 프로토스페이서 인접 모티프(PAM)를 인식하는 제2 Cas9 분자를 코딩하는 폴리뉴클레오타이드 서열을 포함하는 제2 벡터를 대상체에 투여하는 단계를 포함하며, 제1 및 제2 gRNA 분자 각각은 19 내지 24개의 뉴클레오타이드 길이의 표적화 도메인을 가지며, 제1 벡터 및 제2 벡터는 인간 디스트로핀 유전자의 엑손 51 측면의 제1 및 제2 인트론에 각각 제1 및 제2 이중 가닥 절단부를 형성하여, 엑손 51을 포함하는 디스트로핀 유전자의 절편을 결실시키도록 구성되는, 방법.
- [0350] 항 68. 제67항에 있어서, 상기 제1 gRNA 분자 및 상기 제2 gRNA 분자가: (i) SEQ ID NO: 1에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 2에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ii) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iii) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (iv) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (v) SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vi) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (vii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (viii) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 4에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (ix) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 19에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; (x) SEQ ID NO: 14에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 15에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자; 및 (xi) SEQ ID NO: 11에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 18에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자로 구성된 그룹으로부터 선택되는, 방법.
- [0351] 항 69. 제68항에 있어서, 상기 대상체는 뒤시엔느 근이영양증을 앓고 있는, 방법.
- [0352] 항 70. 제67항 내지 제69항 중 어느 한 항에 있어서, 상기 제1 벡터 및 제2 벡터가 대상체의 근육에 투여되는, 방법.
- [0353] 항 71. 제70항에 있어서, 상기 근육이 골격근 또는 심장 근육인, 방법.
- [0354] 항 72. 제71항에 있어서, 상기 골격근은 경골 전방 근육(tibialis anterior muscle)인, 방법.
- [0355] 항 73. 제67항 내지 제72항 중 어느 한 항에 있어서, 상기 제1 벡터 및 제2 벡터가 대상체에게 근육 내, 정맥 내 또는 이들의 조합으로 투여되는, 방법.
- [0356] 항 74. 엑손 52 결실(Δ 52)을 갖는 인간 디스트로핀 유전자(hDMD)를 갖는 트랜스제닉 설치류 배아를 생성하는 방법으로서, 상기 방법은 제1항의 gRNA, 제2항 내지 제9항 중 어느 한 항의 DNA 표적화 시스템, 제10항의 단리

된 폴리뉴클레오타이드, 제11항 내지 제22항 중 어느 한 항의 벡터, 제30항 또는 제31항의 변형된 아데노-연관 바이러스 벡터 또는 제32항 내지 제51항 중 어느 한 항의 조성물을 설치류 배아에 투여하여 인간 디스트로핀 유전자의 엑손 52가 결실된 단계, 및 인간 디스트로핀 유전자의 엑손 52가 결실된 트랜스제닉 설치류 배아를 선별하는 단계를 포함하며, 상기 설치류 배아는 정상적인 인간 디스트로핀 유전자를 포함하는, 방법.

[0357] 항 75. 제74항에 있어서, 상기 설치류 배아가 마우스 배아인 방법.

[0358] 항 76. 제74항 또는 제75항에 있어서, 상기 트랜스제닉 설치류 배아가 이중 접합체 hDMD 또는 이중 접합체 hDMD-Δ52인, 방법.

[0359] 항 77. 제74항 내지 제76항 중 어느 한 항에 있어서, SEQ ID NO: 41에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제1 gRNA 분자, 및 SEQ ID NO: 42에 개시된 뉴클레오타이드 서열을 포함하는 표적화 도메인을 포함하는 제2 gRNA 분자가 설치류 배아에 투여되어, 인간 디스트로핀 유전자의 엑손 52를 결실시키는, 방법.

[0360] 항 78. 제74항 내지 제77항 중 어느 한 항에 있어서, SEQ ID NO: 27에 개시된 아미노산 서열을 포함하는 Cas 단백질을 설치류 배아에 투여하는 단계를 추가로 포함하는, 방법.

[0361] 항 79. 제74항 내지 제78항 중 어느 한 항의 방법에 의해 생성된 트랜스제닉 설치류 배아.

[0362] 항 80. 제79항의 트랜스제닉 설치류 배아로부터 생성된 트랜스제닉 설치류.

[0363] 별첨

[0364]

JCR179(SEQ ID NO: 37)를 갖는 pD0240(gRNA 볼드체)

ctaaattgtaagcgttaatatattttgtaaaattcgcgttaattttgttaaatcagctcatttttaaccaatagg
ccgaaatcggcaaaatccctataaatcaaaagaatagaccgagatagggttgagtggttccagtttggaacaag
agttccactattaaagaacgtggactccaacgtcaaaaggcgaaaaaccgtctatcagggcgatggccactacgtga
accatcacctaatcaagtttttggggtcaggtgcccgtaaagcactaaatcggaacccctaaaggagcccccgat
ttagagcttgacggggaagccggcgaaacgtggcgagaaaggaaaggaaagcgaaggagcggcgctaggggcg
ctggcaagtgtagcggctcacgctgcgcgttaaccaccaaccccgccgcttaatgcgccgtacaggcgctgccca
ttcgccattcaggctgcgcaactgttgggaaggcgatcgggtcgggcctcttcgtattacgccagctggcgaaag
ggggatgtgctgcaaggcgatlaaagttgggtaacgcagggtttccagtcacgacgttgtaaacgacggccagt
gagcgcgctgaatcagctcactataggcgaaattgggtaccaagcttgccattttccatgaticctcatatttg
catacagatacaaggctgttagagagataattggaattaattgactgtaaacacaaagataattagtacaaaatac
gtgacgtagaagaataaatttctgggtagtttgcagttttaaattatgttttaaattggactatcatatgctta
ccgttaacttgaaagtatttcgatttcttggtttatatatctgttgaaaggacgaaacacc**AACACACAGCTGGGT**
TATCAGAGgttttagtactctggaacagaatctactaaacaaggcaaatgcggtgtttatctcgtcaactgtt
ggcgagattttttgcggcccccgcggtggagctccagctttgttcccttagtgagggttaattgcgcgttgg
cgtaatcatggctatagctgtttctgtgtgaaattgttatccgctcacaattccacacatacagccggaagc
ataaagtgtaaagcctggggtgcctaagtgagtgagctaacacataaattgcgttgcgctcactgcccgtttcca
gtcgggaaacctgtgcgacgtgcattaatgaatcggcaacgcggggagagcggtttgcgtattggcgct
ctccgcttctcgtcactgactcgtcgcgtcgttcggtcggcgagcggtatcagctcactcaaaaggcg
gtataacgggtatccacagaatcagggtgatacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaa
ccgtaaaaaggccggtgtgctggcgttttccataggtccgccccctgacgagcatcacaaaaatcgacgtcaa
gtcagaggtggcgaaacccgacaggactataagatacaggcggtttcccttggaagctccctcgtgcgtctcct
gttccgacctgcgcttaccggatacctgtcgcctttctcccttgggaagcgtgagcgtttctcatagctcacg
ctgtaggatctcagttcgtgtaggtcgttcctcaagctgggctgtgtgcacgaacccccgttcagcccagc
gctgcgccttaccggtaactatcgttctgagtcacacccggttaagacacgacttatgccactggcagcagccact
ggtaacaggattagcagagcgaggtatgtaggcggtgctacaggttcttgaagtgggtgacctactcggtacac
tagaaggacagatttgatctgcgctcgtgaagccagttaccttcggaaaaagagttggtagctctgtatccg
gcaaaaaaacaccgctgtagcggtgtttttgttgcaggcagatiacgcgcagaaaaaaggatctcaa
gaagatcctttgatctttctacggggtgacgctcagtggaacgaaactcacgttaaggattttggtcatgag
attatcaaaaaggatcttcacctaagatccttttaaaataaaatgaagtttaaatcaatctaaagtatatatgagt

[0365]

aaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgtatctgtctatttcgttcacata
gttgccgtactccccgtcgtgtagataactacgatacgggagggttaccatctgccccagtgctgcaatgatacc
gcgagaccacgctcaccggctccagatttatcagcaataaacagccagccggaaggccgagcgcagaaagtggtc
ctgcaactttatccgctccatccagcttattattgttccgggaagctagagtaagtgttcgccagttatagt
ttgcgcaacgttgttccattgctacaggcatcgtggtgtcacgctgctgttgggtatggctcaticagctccgg
ttccaaacgatcaaggcgagttacatgatccccatgttgcgcaaaaagcgggttagctccttcggtcctccgatcg
ttgtcagaagtaagtggccgaggttatcatcatggttatggcagcactgcataattctcttactgtcatgccaa
tccgtaagatgctttctgtgactgggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgaccgagttg
ctcttgcggcgctcaatacgggataataccgcgccacatagcagaactttaaaagtctcatattggaaaacgtt
cttcggggcgaaaaactcgaaggatctaccgctgtgagatccagttcagatgaacccactgctgacccaactga
tcttcagatcttttacttaccagcgtttctgggtgagcaaaaacaggaaaggcaaaatgccgcaaaaagggaat
aaggcgacacggaaattgtgaatactcactcttcttcaataattattgaagcatttatcagggttattgtc
tcatgagcggatacatattgaaatgtatttagaaaaataacaaatagggttccgcgcacattccccgaaaagtgcacac

[0366]

[0367]

JCR183(SEQ ID NO: 38)을 갖는 pD0240(gRNA 볼드체)

ctaaattgtaagcgttaatatittgttaaattcgcgttaaattttgttaaatcagctcatttttaaccaatagg
ccgaatcggcctaaatccctataaatcaaaagaatagaccgagatagggtgagtggttccagtttggaacaag
agtccactatataagaacgtggactccaacgtcaaggcgcaaaaccgtctatcaggcgagggccactacgtga
accatcacctaatacaagttttggggtcaggtggttaagcactaaatcggaaccctaaaggagccccgat
ttagagcttgacggggaagccggcgaaacgtggcgagaaagggaagggaagcgaaggagggcgctaggcg
ctggcaagtgtagcgtgacgtgcgtgaaccacacccgccgcttaatgcgcgctacaggcgctgccca
ttcgccattcaggctgcgcaactgttgggaaggcgatcgggtcggtcctcttgcctattaccagctggcgaaag
ggggatgtgctgaaggcgatlaagttgggtaacgccagggtttccagtcacgacgttgtaaacgacggccagt
gagcgcgctgaatagactcactataggcgcaattgggtaccaagcttgcctattcccatgattcctcatattg
catatagatacaaggcgtgttagagagataattggaattaattgactgtaaacacaaagataattgacaaaatc
gtgacgtagaagaataaatttctgggtagtttgcagttttaaattatgtttaaattgactatcatatgctta
ccgtaactgaaagtatttcgatttctgctttatatactgtggaaggacgaacacc**AGAACTGGTGGGAAA**
TGGTCTAGgttttagtactctggaaacagaatctactaaaaaaggcctggtttatctcgtcaactgtt
ggcgagattttttgcggcccccgcggtgagctccagcttttgccttttagtgagggttaattgcgctgtg
cgtaatcatggctatagctgtttctgtgtgaattgttatccgctcacaattccacacacatacagccggaagc
ataaagtgtaaagcctggggtgcctaagtgagtgagtaactcacattaattgcgttcgctcactgccgctttcca

[0368]

gtcgggaaacctgtcgtccagctgcattaatgaatcgccaacgcggggagagcggttgcgtattggcgct
ctccgcttctcgtcactgactgcgtgcgtcggtcttgcggctgcggcgagcggtatcagctcactcaaggcg
gtaatcagggtatccagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaa
ccgtaaaaaggccggtgtcgtggttttccatagctccgccccctgacgagcatcacaataatgacgctcaa
gtcagagggtggcgaaccggacaggactataaagataccaggcgtttccctgggaagctccctcgtgcgtctcct
gttccgacctgcgctaccggatacctgtccgcttctccttccgggaagcgtggtcttctcatagctcacg
ctgtaggatctcagttcgtgtgaggtcgttcgtccaagctgggtgtgtgcacgaacccccgttcagcccgacc
gctgcgcttatccggtaactatcgtcttgagtcacccggtaagacagactatcgtccactggcagcagccact
ggtaacaggttagcagagcaggtatgtaggcggtgtcagaggttctgaagtggtggcctaactacggctacac
tagaaggacagattttggtatctgcgtctgtgaagccagttaccttccgaaaagagttggtagctcttgatccg
gcaaaaaaccaccgctgtgtagcgggtgtttttgtttgcaagcagcagattacgcgcaaaaaaaggatctcaa
gaagatcctttgatctttctacggggtctgacgtcagtggaacgaaaactcagtttaagggttttggtagctg
attatcaaaaaggatcttcactagatccttttaattaaaatgaagttttaaatcaatctaagtatatagat
aaacttggtctgacagttaccaatgcttaacagtgaggcacctatctcagcgatctgtctatttctgtcatccata
gttgcctgactccccgctgtgtagataactacgatacgggagggttaccatctgccccagtgctgcaatgatacc
gagagaccacgctcaccgggtccagattatcagcaataaacagccagccggaaggccgagcagagaagtggtc
ctgcaactttatccgctcctcagctatattaattgttgcgggaagctagagtaagtagttcggcagtaatagt
ttgcgcaacgttgttccattgctacaggcatcgtgtgtgcagctcgtgttggtaggttcattcagctccgg
ttcccaacgatcaaggcgagttatgatcccatgttggcaaaaaagggttagctccttcggtcctccgatcg
ttgtcagaagtaagttggccgaggttatcacatcagttatggcagcactgcataattctcttactgtcatgcca
tccgtaagatgctttctgtgactggtgagtaactcaacagtcattctgagaatagtgatgcggcgaccgagttg
ctcttgcggcggtgaataacgggataatccgcgccacatagcagaactttaaaaagtctcatcattggaacggt
cttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaaccactcgtgcaccaactga
tcttcagcatcttttacttaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaat
aaggcgacacggaaatgtgaatactcactcttcttcttcaatatattgaagcatttatcagggttatgtc
tcatgacgggatacatattgaattatgaaaaataaacaatatggggttcgcgcacatttccccgaaaagtgc ccac

[0369]

[0370]

JCR179 PT366AAV 179(SEQ ID NO: 39)를 갖는 PT366-생체내 연구에 사용된 AAV 플라스미드(gRNA 볼드체; SaCas9는 대문자; NLS는 소문자, 볼드체, 및 밑줄표기됨)

cctgcaggcagctgcgcgctcgtcgtcactgaggccgcccggcgctggcgacctttggtcgcgccgctcagt
gagcgcgcgagcgcgcagagaggagtgcccaactccatcactaggggttcctgcggcctctagactcgaggcgttg

[0371]

acattgattattgactagittattaatagtaatcaattacggggctcattggtcatagcccatataggagtccgcg
ttacataacttacggtaaatggccgcctggctgacccccaacgacccccgccattgacgtcaataatgacgtat
gttcccatagtaacccaatagggaactttccattgacgtcaatgggtggagtatttacggtaaacgccacttggc
agtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggtaaatggccgcctggcattatg
cccagfacatgaccttatgggaactttcctacttggcagtagatctactattagtcacgtattaccatggtgatg
cggttttggcagtagatcaatgggcgtggatagcggtttgactacggggatttccaagtctccacccattgacgt
caatgggaggtttgttggcaccaaaatcaacgggactttccaaatgtcgtacaactcgcgccattgacgcaaa
tgggcggttaggcgtgtacgtgggaggtctatataagcagagctctctgctaactaccggtgccaccatggcccca
aagaagaagcggaaggtcggtatccacggagtcacagccAAGCGGAACACTACATCTGGGCCTGGACAT
CGGCATCACCAGCGTGGGTACGGCATCATCGACTACGAGACACGGGACGTGATCG
ATGCCGGCTGTCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCG
GAGCAAGAGAGGGCCAGAAGGCTGAAGCGGCGGAGGCGCATAGAATCCAGAGA
GTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGG
CATCAACCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAA
GAGTTCTCTGCCCCCTGCTGCACCTGGCCAAGAGAAAGGCGTGCACAACGTGAAC
GAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACAAAGAGCAGATCAGCCGGA
ACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACCTGCAGCTGGAACGGCTGAA
GAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTG
AAAGAAGCCAAACAGCTGTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGA
GCTTCATCGACACCTACATCGACCTGTGGAAACCCGGCGGACCTACTATGAGGGA
CCTGGCGAGGGCAGCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCT
GATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACA
ACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGAC
GAGAACGAGAAGCTGGAATATTACGAGAAGTTCAGATCATCGAGAACGTGTTCAA
GCAGAAGAAGAAGCCACCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAA
GAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCGAGTTCACCAACCTG
AAGGTGTACCACGACATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACGC
CGAGCTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACA
TCCAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAG
ATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAA
CCTGATCCTGGACGAGCTGTGGCACCAACGACAACCAGATCGCTATCTCAACCG

[0372]

GCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAAAGAGATCCCCACCA
 CCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCA
 TCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCAACGACATCATTATC
 GAGCTGGCCCGCGAGAAGAAGTCCAAGGACGCCCAGAAAATGATCAACGAGATGC
 AGAAGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGG
 CAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAA
 GGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACAACCC
 CTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCTTCGACAACAGCTT
 CAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACC
 CCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAA
 GCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAG
 TATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAA
 CCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGA
 GCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTC
 ACCAGCTTCTCGCGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAA
 GCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTTCATCTCAAAGAGTG
 GAAGAACTGGACAAGGCCAAAAAAGTGATGGAACCAGATGTTTCGAGGAAAAG
 CAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCAT
 CACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACC
 GGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGG
 AAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAA
 GGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGT
 ACCACCACGACCCCCAGACCTACCAGAACTGAAGCTGATTATGGAACAGTACGGC
 GACGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACCTACCTGACCAA
 GTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACA
 AACTGAACGCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTC
 GTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGTGTAC
 AAGTTCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAATACTACGAAGT
 GAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCAGCAACCAGGCC
 GAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGCTGTAT
 AGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGA

[0373]

드체, 및 밀줄표기됨)

cctgcaggcagctcgcgctcgcctcactgaggccgcccgggctcgggagacctttggtcgcggcgctcagtg
gagcgagcgagcgcgagagaggagtgcccaactccatcactagggttctcgcggcctctagactcgagcggttg
acattgatttactagttattaatagtaataacgggtcattgtcatagcccatatattgaggttccg
ttacataacttacggttaattggccgctggctgacccccaacgacccccccattgacgtcaataatgacgtat
gttcccatagtaaccgaataggactttccattgacgtcaatgggtggagttattcggtaaaactgccacttggc
agtacatcaagtgtatcatatgccaaagtacgccccctattgacgtcaatgacggttaattggccgctggcattatg
ccagttacatgaccttaaggactttctacttggcagttacatctactgattagtcacgtattacattggtgatg
cgggttttggcagttacatcaatggcggtgtagcgggttggactcacgggatttccaagtctccacccattgacgt
caatgggagtttgggttggcaccacaaatcaacgggactttccaaatgtcgttaacaactccgccccattgacgcaa
tgggcggtaggcgtgtacgggtggaggtctatataagcagagctctctggctaactaccggtgccaccattggcccca

[0377]

aagaagaagcgggaaggctcggtatccacggagtgccagcagccAAGCGGAACATACATCTGGGCTGGACAT
CGGCATCACCAGCGTGGGCTACGGCATCATCGACTACGAGACACGGGACGTGATCG
ATGCCGGCTGTGCGGTGTCTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCG
GAGCAAGAGAGCGCCAGAAGGCTGAAGCGCGGAGGCGGCATAGAATCCAGAGA
GTGAAGAAGCTGCTGTTCTGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGG
CATCAACCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAA
GAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAA
CGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGG
AACAGCAAGGCCCTGGAAGAGAAAATACGTGGCCGAACCTGCAGCTGGAACGGCTGA
AGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGT
GAAAGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAG
AGCTTATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGG
ACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAGAATGGTACGAGATGC
TGATGGGCCACTGCACCTACTTCCCCGAGGAACCTGCGGAGCGTGAAGTACGCCTAC
AACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGA
CGAGAACGAGAAGCTGGAATATTACGAGAAGTCCAGATCATCGAGAACGTGTTCA
AGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCTCGTGAACGA
AGAGGATATTAAGGGTACAGAGTGACCAGCACCGGCAAGCCGAGTTCAACCAACC
TGAAGGTGTACCACGACATCAAGGACATTACCGCCCGAAAGAGATTATTGAGAAC
GCCGAGCTGCTGGATCAGATTGCCAAGATCTTGACCATCTACCAGAGCAGCGAGGA
CATCCAGGAAGAAGTACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGC
AGATCTCTAATCTGAAGGGTATACCGGCACCCACAACCTGAGCCTGAAGGCCATC
AACCTGATCCTGGACGAGCTGTGGCACCAACGACAACCAGATCGCTATCTTCAA
CCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCAGCAGAAAGAGATCCCCA
CCACCTGGTGACGACTTCATCTGAGCCCCGTGCTGAAGAGAAGCTTCATCCAGA
GCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCAACGACATCATT
ATCGAGCTGGCCCCGAGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGAT
GCAGAAGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGACCAACC
GGCAAAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGG
AAGGCAAGTGCCTGTACGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACAAC
CCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCTTCGACAACAGC

[0378]

TTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGA
 CCCCATTCCAGTACCTGAGCAGCAGCAGCAAGATCAGCTACGAAACCTTCAAG
 AAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAG
 AGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATC
 AACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCG
 GAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCT
 TCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTAC
 AAGCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGA
 GTGGAAGAACTGGACAAGGCCAAAAAGTGATGGAACCAGATGTTTCGAGGAA
 AAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCT
 TCATCACCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGC
 CACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCAC
 CCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACG
 ACAAGGACAATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCT
 GATGTACCACCACGACCCCGAGCCTACCAGAACTGAAGCTGATTATGGAACAGT
 ACGGCGACGAGAAGAATCCCTGTACAAGTACTACGAGGAAACCGGGAACCTACCTG
 ACCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGG
 CAACAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACA
 AGGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCG
 TGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAAGTACTAC
 GAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCAGCAACC
 AGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGATCAACGGCGAGC
 TGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATG
 ATCGACATCACCTACCGGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCAG
 GATCATTAAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGACA
 TTCTGGGCAACCTGTATGAAGTGAATCTAAGAAGCACCCCTCAGATCATCAAAAAG
 GGCaaaaggccggccacgaaaaaggccggccaggca

aaaaagaaaaagggatcctaccatacagatgttcagattacgcttaccatacagatgttcagattacgcttacc
 atacgatgttcagattacgcttaagaattctagagctcgtgatcagctcgtgcttcttagttgccagcc
 atctgttgttgccttccccctgccttcttgacctggaaggtgccactcccactgtcctttctaataaatg
 aggaattgcacgcattgtctgagtaggtgtgcatctattctgggggtgggtggggcaggacagcaagggggag

[0379]

gattgggagagaatagcaggcatgctggggaggtaccgagggcctatttccatgattcctcatattgcatata
cgatacaaggctgttagagagataaattggaattatttgaactgtaaacacaaagatattagtacaaatacgtgacg
tagaaagtaataatttctgggtagttgtagttttaaattatgttttaaattggaatcatatgcttaccgtaa
cttgaaagtatttcgatttctggctttatatacttgggaaaggacgaaacaccg**AGAACTGGTGGGAAATGGTC**
TAGgttttagtactctgaaacagaatctactaaaacaggcaaatgccgtttatctcgtaactgttggcgga
gattttgctggccgaggaacccctagtgatggagttggccactccctctctgctgctcgtcactgaggcc
ggcgaccaaaggctgcccagcggggcttggccggcgccctcagtgagcgaagcagcgcagctgcctgca
ggggcgccgtgatgctgtatttctcttaccgcatctgtgctgtatttccaccgcatacgtcaaaacacatagta
cgcgccctgtacggcgcatgaagcgcggggtgtgtgtgttacgcgcagcgtgaccgtacacttgccagcgccc
tagcgcccgtccttctgttcttcttcttcttctgcccagcttgcgggcttccccgtcaagctctaaatcgg
gggctccctttaggggttccgatttagtcttaccggcaccctgaccccaaaaactgttgggtgatgttcacg
tagtgggccaatcgccctgatagacggttttgcctttagcgttggagtcacgttctttaatgttgactcttgt
tccaaactggaacaacactcaaccctatctcgggctattctttagattataagggttggcgttctggcctat
tggttaaaaaagagctgatttaacaaaattacgcgaattttaacaaaattataacgtttacaattttatggtg
cactctcagtacaatctgctctgatgccgcatagtttaagccagccccgacccgccaacacccgctgacgcgcct
gacgggcttctgtctcccgcatccgcttacagacaagctgtgaccgtctccggagctgcatgtgtcagaggtt
tcaccgtcatcaccgaaacgcgcgagacgaaggccctctgtgatacgcctattttataggttaatgcatgataat
aatggttcttagacgtcaggtggcacttttggggaaatgtgcgcggaacccctatttgttattttctaaatac
attcaaatatgtaccgtcatgagacaataaccctgataaatgcttcaataattgaaaaaggagagatgagt
attcaacatttccgtgctcccttattcccttttggcgccatttggccttctgttttgcaccagaaacgct
gggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtggttacatcgaactggatctcaacagcggtgaaga
tccttgagaggttttgcggcggaagacgtttccaatgatgacacttttaagttctgtatgtggcgcggtatta
tccgtattgacggggcaagagcaactcgtgcggccatatactattctcagaatgacttgggtgagtactacc
agtcacagaaaagcatcttaccggatggcatgacagtaagagaattatgagtgctgccataacatgagtataaca
ctgcggccaacttacttctgacaacgatcgaggaccgaaggagctaaccgtttttgcaacatgggggatcat
gtaactcgcctgtgctgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgt
agcaatggcaacaacgttgcgcaactattactggcgaactacttactctagcttcccgcaacaattatagact
ggatggagggcgataaagttgcaggaaccacttctgcgctcggccctccggctggtgttattgtctgataaatct
ggagccggtgagcgtggaagcccggtatcatfagcagcactggggccagatgtaagccctccctatcgtatgtagt
ctacacgacggggagtcaggcaactatggatgaacgaataagacagatcgtgagataggtgcctactgattaagc
attggttaactgtcagaccaagtttactcatatatacttagattgattaaaacttattttaataaaaggatc

taggtgaagatcccttttgataatctcatgacaaaatcccttaacgtgagtttctgttccactgagcgtcagacc
cgtagaaaaagatcaaggatcttcttgagatccttttttctgcgctaatctgctgttgcacacaaaaaacac
cgctaccagcgggtgttgttggccgatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcg
cagataccaaatactgtccttctagttagccgtatgtagccaccacttcaagaactctgtagcaccgcctacata
cctcgtctgtctaactctgttaccagtgctgtcgcagtgccgataagtgctgtcttaccgggtgactcaagac
gatagttaccggataagcgcgagcggctgacgggggttcgtgcacagcccagcttggagcgaacgacc
tacaccgaactgagatacctacagcgtgagctatgagaaagccacgcttccgaaggagaaaggcgagaggtta
tccggtaagcggcgaggttcggaacgagagcgacgagggagcttccagggggaacgcctgtatctttatagtc
ctgtcgggttccacactctgactgagcgtcgtattttgtgatgctcgtcagggggcgagcctatggaaaaac
gccagcaacgcggcccttttaccggttctggtccttttgccttctgtccttctgctacatgt

pD0242(시험관내 연구를 위한 모든 JCR89/91 프로젝트 및 JCR157/160 프로젝트에 사용되는 SaCas9; SaCas9는 대문자)(SEQ ID NO: 83)

ctaaattgtaagcggttaattttgttaaaattcgcgttaaattttgttaaatcagctcatttttaaccaatagg
 ccgaatatcgcaaaatccctataaatcaaaagaatagaccgagatagggttgagtggttccagtttgaacaag
 agtccactattaaagacgtggactccaacgtcaaagggcgaaaaaccgtctatcaggcgcatggcccactactgta
 accatcacccatcaagtttttgggtcgaagtgccgtaaagcactaaatcggaaccctaaaggagcccccgat
 tttagagcttgacggggaagccggcgaaacgtggcgagaaaggaaagggaagaaaggagcggcgctaggggcg
 ctggcaagtgtagcggtcacgtgcgcgttaaccaccacaccccgcgcttaatgcccgtactacggcgcgctcca
 ttccgattcaggtgcgcaactgttgggaaggcgatcggtgcgggcctcttcgtattacgccagctggcgaaag
 ggggatgtctgaaggcgtaagtgtggtaacccagggtttccagtcacgacgttgtaaacgacggccagt
 gagcgcgctgaatacagactactataggcggaattgggtacCtttaattctagtactatgcaTgcgttgacattgat
 tattgactagtattataagtaaatcaattacgggtcattagttcatagcccatatagggagtccggttacataa
 ctacggtataatggcccctggctgacgcccacaccccccgcattgacgtcaataatgacgtatgttcccat
 agtacgccaatagggaatttccattgacgtcaatgggtgagttattacggttaactgccacttggcagtacac
 aagtgatcatatgccaagtagccccctattgacgtcaatgacggtaaatggcccgctggcattatgccagtag
 atgaccttagggactttcctacttggcagtagcatctacgtattatgctcatgattaccatggtgatgctgttttg
 gcagtagacatcaatggcggtggtagcggtttactcacggggatttccaagtctccacccattgacgtcaatggga
 gttgttttggcaccaaaatcaacgggactttccaaatgtcgtacaacactccgccattgacgcaaatggcggt
 aggcgtgtacgtgggaggtctatataagcagagctctctgtaactaccggtgccaccATGAAAAGGAACTACAT
 TCTGGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTATGAAA

[0383]

CAAGGGACGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAAC
 AATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCTGAAACGACGGAGAAGGC
 ACAGAATCCAGAGGGTGAAGAACTGCTGTTTCGATTACAACCTGCTGACCGACCAT
 TCTGAGCTGAGTGGAATTAATCCTTATGAAGCCAGGGTGAAGGCCTGAGTCAGAA
 GCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGCTAAGCGCCGAGGAGT
 GCATAACGTCAATGAGGTGGAAGAGGACACCGCAACGAGCTGTCTACAAAGGAAC
 AGATCTCACGCAATAGCAAAGCTCTGGAAGAGAAGTATGTCGACAGCTGCAGCTG
 GAACGGCTGAAGAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTCAAGACAA
 GCGACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAGTGCAGAAGGCTTACCACCAG
 CTGGATCAGAGCTTCATCGATACTTATATCGACCTGCTGGAGACTCGGAGAACCTAC
 ATGAGGGACCAGGAGAAGGGAGCCCCCTTCGGATGGAAAGACATCAAGGAATGGTA
 CGAGATGCTGATGGGACATTGACACCTATTTCCAGAAGAGCTGAGAAGCGTCAAGT
 ACGCTTATAACGCAGATCTGTACAACGCCCTGAATGACCTGAACAACTGGTCATCA
 CCAGGGATGAAAACGAGAAACTGGAATACTATGAGAAGTTCCAGATCATCGAAAAC
 GTGTTTAAGCAGAAGAAAAAGCCTACACTGAAACAGATTGCTAAGGAGATCCTGGT
 CAACGAAGAGGACATCAAGGGCTACCGGGTGACAAGCACTGGAACACAGAGTTC
 ACCAATCTGAAAGTGATCACGATATTAAGGACATCACAGCACGGAAGAATCATT
 GAGAACGCCGAAGTCTGGATCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCC
 GAGGACATCCAGGAAGAGCTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGA
 TCGAACAGATTAGTAATCTGAAGGGGTACACCGGAACACACAACCTGTCCCTGAAA
 GCTATCAATCTGATTCTGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATC
 TTTAACCGGCTGAAGCTGGTCCCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGAT
 CCCAACCACTGGTGGACGATTTTATTCTGTACCCGTGGTCAAGCGGAGCTTCAT
 CCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAATGATA
 TCATTATCGAGCTGGCTAGGGAGAAGAAGCAAGGACGCACAGAAGATGATCAAT
 GAGATGCAGAAACGAAACCGGCAGACCAATGAACGCATTGAAGAGATTATCCGAAC
 TACCGGGAAGAGAACGCAAAGTACCTGATTGAAAAAATCAAGCTGCACGATATGC
 AGGAGGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCCTGGAGGACCTGCTGAAC
 AATCCATTCAACTACGAGGTCGATCATATTATCCCCAGAAGCGTGTCTTCGACAAT
 TCCTTTAAACAAGGTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAGGGCAATAG
 GACTCCTTTCCAGTACCTGTCTAGTTTCAAGATCTCTTACGAAACCTTTAAA

[0384]

AAGCACATTCTGAATCTGGCCAAAGGAAAGGGCCGCATCAGCAAGACCAAAAAGG
 AGTACCTGCTGGAAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTA
 ACCGGAATCTGGTGGACACAAGATACGCTACTCGCGGCTGATGAATCTGCTGCGAT
 CCTATTTCCGGGTGAACAATCTGGATGTGAAAGTCAAGTCCATCAACGGCGGGTTCA
 CATCTTTTCTGAGGCGCAAATGGAAGTTAAAAAGGAGCGCAACAAAGGTACAAG
 CACCATGCCGAAGATGCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGG
 AAAAAGCTGGACAAAGCCAAGAAAGTGATGGAGAACCAGATGTTTCAAGAGAAGC
 AGGCCGAATCTATGCCGAAATCGAGACAGAACAGGAGTACAAGGAGATTTTCATC
 ACTCCTCACCAGATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCTACCGG
 GTGGATAAAAAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACAAGAAA
 AGACGATAAGGGGAATACCCTGATTGTGAAACAATCTGAACGGACTGTACGACAAAG
 ATAATGACAAGCTGAAAAAGCTGATCAACAAAAGTCCCAGAGAAGCTGCTGATGTAC
 CACCATGATCCTCAGACATATCAGAACTGAAGCTGATTATGGAGCAGTACGGCGA
 CGAGAAGAACCCTGTATAAGTACTATGAAGAGACTGGGAACCTACCTGACCAAGT
 ATAGCAAAAAGGATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAG
 CTGAATGCCCATCTGGACATCACAGACGATTACCCTAACAGTCGCAACAAGGTGGT
 CAAGCTGTCACTGAAGCCATACAGATTGATGTCTATCTGGACAACGGCGTGTATAA
 ATTTGTGACTGTCAAGAATCTGGATGTATCAAAAAGGAGAACTACTATGAAGTGA
 ATAGCAAGTGCTACGAAGAGGCTAAAAAGCTGAAAAAGATTAGCAACCAGGCAGA
 GTTCATCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAACTGTATAG
 GGTCATCGGGGTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTGACAT
 CACTTACCGAGAGTATCTGAAAAACATGAATGATAAGCGCCCCCTCGAATTATCAA
 AACAAATTGCCTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAA
 ACCTGTATGAGGTGAAGAGCAAAAAGCACCTCAGATTATCAAAAAGGGCagcggaggc
 aagcgt cctgctgctactaagaaagctggtcaagctaagaaaaagaaagatcctaccatagctgttccagattacgctta
 agaattcctagagctcgctgatcagcctcagctgtgccttctagttgccagccatctgttgttgccttccccgt
 gcccttctgaccctggaaggtgccactcccactgtcctttcctaataaatgaggaattgcatcgcatgtgtctga
 gtagggtgctattctattctgggggtgggggtggggcaggacagcaagggggaggattgggaagagaatagcaggcat
 gctggggaggtagcgccgcCCcggtggagctccagctttgttccctttagtgagggttaattgcgcgttggcg
 taatcatggtcatagctgttctgtgtgaaattgtatccgctcacaattccacacatacgagccggaagcat
 aaagtgtaagcctgggtgcctaagtgagtgagctaactcacattaattgcgttgcgtcactgcccgtttccagt

[0385]

cgggaaacctgtcgtgccagctgcatfaatgaatcgccaacgcgcggggagagcggttgcgtattggcgctct
 tccgttctcgtcactgactcgtcgtcgtcgttcggctgcgcgagcggtatcagctcactcaaaggcggt
 aatacgggtatccacagaatcaggggataacgcaggaaagaacatgtagcaaaaggccagcaaaaggccagggaacc
 gtaaaaaggccggttgcgtggcgttttccataggtccgccccctgacgagcatcacaaaaatcgacgtcaagt
 cagagggtggcgaaccggacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgtctcctgt
 tccgacctgcccgttaccggataccgtcgccttctccctcgggaagcgtggcgtttctcatagctcagct
 gtaggtatctcagttcgtgttaggtcgtcgtccaagctggcgtgtgtgcacgaacccccgttcagcccaccgc
 tgcgcttatccggtactatcgtttagtccaacccggtaagacacgactatcggcagcagccactgg
 taacaggattagcagagcgaggatgttagcggtgctacagagttctgaagtgggtgcctaactacggctacacta
 gaaggacagtatttggatctgcgtctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggc
 aaacaaccaccgctgtagcggtgtttttgttgaagcagcagattacgcgcagaaaaaggatctcaaga
 agatcctttgatctttctacgggctgacgctcagtggaacgaaactcaggttaagggttttggatgagat
 tatcaaaaaggatcttcacgtatctttaaataaaaaagagtttaaatcaatctaaagtatatatgagtaa
 acttggctgcaggttaccatgttaacatcagtgaggcacctatctcagcgtatctgtctatttctcatcatagt
 tgcctgactccccgtcgtgtagataactacgatacgggagggttaccatctgccccagtgctgcaatgataccgc
 gagaccacgctcaccggctcagattatcagcaataaccagccagccggaaaggccgagcgagaaagtggtcct
 gcaactttatccgctccatccagcttattattgttgcgggaagctagagtaagtagttcgcagtaatagttt
 gcgcaacgtgttgcattgctacaggcatcgtggtgtcgcgtcgtgttggatggcttattcagctccggtt
 cccaacgatcaaggcgagttacatgatccccatgttgcgcaaaaagcggttagctccttcggtcctccgacgtt
 gtcagaagtaagttggccgagtggtatcactcatggttatggcagcactgcataattctcttactgcatgccatc
 cgtaaagatgttttctgactggtgagtaactcaacgaagtcattctgagaatagtgtagcggcgaccgagttgct
 cttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtctcatcattggaaaacgttct
 tcggggcgaaaactctcaagatcttaccgctgttgagatccagttcgtatgaaccactcgtgcaccaactgatc
 ttcagcatcttttacttaccagcggttttgggtgagcaaaaacagggaaggcaaaatgccgcaaaaagggaataa
 gggcgacacggaaatgttgaatactatactcttcttttcaatatttgaagcatttatcagggtattgtctc
 atgagcggatacatatttgaatgtatttagaaaaataacaataagggggtccgcgcacatttccccgaaaagtgcc ac

[0386]

[0387] pJRH1(모든 JCR179/183 프로젝트를 위해 사용되는 SaCas9, SaCas9는 대문자; NLS는 소문자, 볼드체, 및 밑줄표 기함)(SEQ ID NO: 84)

ctaaattgtaagcgtaataatttgttaaaatcgcgtaaaatttgttaaatcagctcatttttaaccaatagg
 ccgaaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtggtgttccagtttggacaag

[0388]

agtccactattaaagaacgtggactccaacgtcaaaaggcgcaaaaaccgtctatcaggcgatggccactacgtga
 accatcacccaatcaagtttttggggcgcaggtgccgtaaaagcactaaatcggaaccctaaaggagccccgat
 ttagagcttgacggggaagccggcgacgtggcgagaaagggaagaaagcgaaggagcggcgctaggggcg
 ctggcaagtgtagcgggtcacgctgcgctaaccaccacaccccccgcgttaatgcgccgtacaggcgctcca
 ttgccattcagctgcgcaactgttgggaaggcgatcggtcggcgctcttcgtattacgccagctggcgaaag
 ggggatgtgtcgaaggcgatfaagttgggtaacccaagggtttccagtcacgacgttgaaaacgacggccagt
 gagcgcgctgaatacgaactactataggcggaattgggtaccttaattctagtactatgcatgcttgacattgat
 tattgactagtattattaagtaatacaattacggggcattagttcatagcccatatggagttccgcttacataa
 cttagcgtaaatggcccgcctggctgaccgccaacgacccccccattgacgtcaataatgacgtattgtccat
 agtaacgccaatagggaactttccattgacgtcaatgggtgagttattacgttaaaactgccacttggcagtaac
 aagttgtatcatatccaagtcagccccctattgacgtcaatgacggtaaatggccgctggcattatgccagtac
 atgaccttatgggactttcctacttggcagtcacatctacgtattagtcgtattacatggtgatgctgttttg
 gcagtcacatcaatggcgctggatagcgggttgactacggggatttccaagtccacccattgacgtcaatggga
 gttgttttggcaccaaaatcaacgggactttccaaatgtcgtacaactccgccattgacgcaaatggcggt
 agggcgtgtacggtggagctctatataagcagagctctctggtactaccgggtgccaccatggcccaagaagaa
 gcggaaggctcggtatccacggagtcacagaccAAGCGGAACACTACCTGGGCCTGGACATCGGCA
 TCACCAGCGTGGGTACGGCATCATCGACTACGAGACACGGGACGTGATCGATGCC
 GGCGTGC GGCTGTTCAAAGAGGCCAACGTGGAACAACGAGGGCAGGCGGAGCA
 AGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAA
 GAAGCTGCTGTTTCTGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCA
 ACCCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTT
 CTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGCGTGCACAACGTGAACGAGG
 TGGAAGAGGACACCGCAACGAGCTGTCCACCAAGAGCAGATCAGCCGGAACAG
 CAAGGCCCTGGAAGAGAAATACGTGGCCGAACCTGCAGCTGGAACGGCTGAAGAAA
 GACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAG
 AAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTC
 ATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGACCTGG
 CGAGGGCAGCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGG
 GCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCC
 GACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACGAGAA
 CGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTGTTCAAGCAGA

[0389]

AGAAGAAGCCACCCCTGAAGCAGATCGCCAAAGAAATCCTCGTGAACGAAGAGGA
TATTAAGGGCTACAGAGTGACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGG
TGTACCACGACATCAAGGACATTACCGCCCGAAAGAGATTATTGAGAACGCCGAG
CTGCTGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATCCA
GGAAGAAGTGAACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCT
CTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTG
ATCCTGGACGAGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCT
GAAGCTGGTGCCCAAGAAGGTGGACCTGTCCAGCAGAAAGAGATCCCCACCACCC
TGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGTTCATCCAGAGCATCA
AAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCAACGACATCATTATCGAG
CTGGCCCCGAGAGAAGTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGA
AGCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAA
AGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGGC
AAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACAACCCCTTC
AACTATGAGGTGGACCACATCATCCCCAGAACGCTGTCTTCGACAACAGCTTCAAC
AACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCAT
TCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAAGAAGCAC
ATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCAAGAAAGAGTATCT
GCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGA
ACCTGGTGGATACCAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTAC
TTCAGAGTGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTACCAG
CTTCTGCGGCGGAAGTGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACC
ACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAG
AAACTGGACAAGGCCAAAAAAGTGATGGAAAACAGATGTTTCGAGGAAAAGCAGG
CCGAGAGCATGCCCAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACC
CCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGT
GGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGG
ACGACAAGGGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGAC
AATGACAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCA
CCACGACCCCCAGACCTACCAGAACTGAAGCTGATTATGGAACAGTACGGCGACG
AGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACCTTGACCAAGTAC

[0390]

TCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAACT
GAACGCCCCTCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGTGA
AGCTGTCCCTGAAGCCCTACAGATTGACGTGTACCTGGACAATGGCGTGTACAAGT
TCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAACTACTACGAAGTGAAT
AGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTT
TATCGCCTCCTTCTACAAACAGTCTGATCAAGATCAACGGCGAGCTGTATAGAGT
GATCGGCGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCA
CCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAG
ACAATCGCCTCCAAGACCCAGAGCATTAAAGAAGTACAGCACAGACATTCTGGGCAA
CCTGTATGAAGTGAATCTAAGAAGCACCTCAGATCATCAAAAAGGGCaaaaggccgg
cgccacgaaaaaggccggccaggcaaaaaagaa

aaaggcatctaccatacatgttccagattacgcttaagaattctagagctcgtgatcagcctcgaactgtgcc
ttctagtccagccatctgtgtttgccctcccccgctccttccctggaagggtgccactccactgtcc
tttctaataaaatgaggaaatgcatcgcatgtctgagtaggtgtcatttattctgggggtgggggtgggcag
gacagcaagggggaggattgggaagagaatagcaggcatgtggggaggtagcgccgcccgcggtggagctccagc
ttttgttcctttagttaggggttaattgcgcgttggcgtaatcatgtgtcatagctgttctctgtgaaattgtta
tccgctcacaattccacacacatacagagccggaagcataaagttaagcctgggggtgcctaagtgtgagctaac
tcacattaatgtcggtgcgctcactgcccgcttccagtcgggaacacctgtcgtccagctgcatiaatgaatcgcc
caacgcgcggggagagggcggtttgcgtattggcgctcttccgcttccgctcactgactcgtcgtcgtcgtcgt
tcggctcggcgagcggtatcagctcactcaaggcggtataacaggttatccacagaatcaggggataacgcaggaa
agaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccggttgcgtcgtgtttccataggtc
cgcccccttgacgagcatcaaaaaatcgacgctcaagtcagaggtgcgaaccgcagagactataaagatacca
ggcggttccccctgggaagctccctcgtgcgctctcgttccgacctgcgcgttaccggatacctgtccgcttcc
tccttcgggaagcgtggcgcttctcatagctcacgctgtaggtatctcagttcgtgtaggtcgttcgctcaag
ctgggctgtgtgcacgaacccccgttcagcccagccgctgcgccttatccggtactatcgtctgagtcacacc
ggtaagacacacttatgccactggcagcagccatggttaacaggaatagcagagcgaggtatgtaggcggtgcta
cagagttctgaagtgtggtcctaactacgctacactagaagacagatattggtatctgcgctcgtcgtgaagcca
gttaccttggaaaaagagttggtagctcttgatccggcaacaacaccgctggtagcggtgtttttgtttg
caagcagcagattacgcgcagaaaaaaggatctcaagaagatccttgatctttctacggggtcgtcagctcag
ggaaacgaaactcaggttaagggtttgtcatgagattatcaaaaggatctcacctagatccttttaataaa
aaatgaagttttaatacaatcaaaagtatatgagtaaacctgtgtcagagttacaaatgcttaacagtgaaggc

acctatctcagcatctgtctatttctcatcatagttgcctgactccccgtcgtgtagataactacgatacggg
aggggttaccatctgccccagtgctgcaatgataccgcgagaccacgctcaccggctccagattatcagcaata
aaccagccagccggaaggccgagcgcagaagtgtcctgcaactttatccgctccatccagctatattgttg
ccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgtgttgccattgctacagggcatcgtggtgt
cacgctcgtcgtttgtatggttcatcagctccgggtccaacgatcaaggcgagttacatgatccccatgttg
tgcaaaaaagcggttagctccttcggtcctccgacgtgtgcagaagtaagtggccgaggtgtatcactcatggt
tatggcagcactgcataattcttctgtcatgcatccgtaagatgttttctgtgactggtgagtactcaacca
agtattctgagaatagtgtatgcggcgaccgagttgctcttgcggcggtcaatacgggataataccgcgccacat
agcagaactttaaaagtgtcatcattggaaaacgttcttggggcgaaaactctcaaggatcttaccgctgttgag
atccagtcgatgtaaccactcgtgcaccaactgatcttcagcatcttttcttaccagcggttctgggtgag
caaaaacaggaaggcaaaatgccgcaaaaaagggaataaggcgacacggaatgtgaatactcatactcttctt
ttcaatatattgaagcatttatcagggttattgtctcatgagcggatacatattgaatgtattagaaaaataa
acaaatagggttccgcgcacatttccccgaaaagtgcac

PT366에서 NLS 서열 (SEQ ID NO: 85)

AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG

pD0203 - SpCas9에 gRNA를 클로닝하는 일반 백본; JCR94 및 JCR99를 세포내에서 시험하기 위해 볼드체 부위에
넣은 다음, hMDM-delta52/mdx 마우스를 제조하기 위한 세포로부터 mRNA를 제조함 (SEQ ID NO: 86).

ctaaattgtaagcgttaattttgttaaaatcgcgttaaattttgttaaatcagctcatttttaaccaatagg
ccgaaatcggcaaaatcccttataatcaaaagaatagaccgagatagggttgagtggttccagtttggaacaag
agtccactatfaaaagacgtgagctcaacgtcaaggcgcaaaaccgtctatcagggcgatggccactacgtga
accatcacctaatacagtttttggggcagagtgccgtaaagcactaaatcggaaccctaaaggagccccgat
ttagagcttgacggggaagccggcgaacgtggcgagaaagggaagggaagcgaaggagcgggcgctagggcg
ctggcaagtgtagcgggtcacgtgcgctgaaccaccacacccgccgcttaatgcccgtacagggcgctccca
ttgccattcaggtcgtgcaactgttgggaaggcgatcgtgcggtcctcttctgctattacgccagctggcgaag
ggggatgtgctgcaaggcgattaaagtgggtaacccagggtttccagtcacgacgttgaacacgacggccagt
gagcgcgctgaatacgaactactataggcgcaattgggtaccgagggcctatttccatgattcctcattatgca
tatacgatacaggtctgttagagagataattggaattatgactgtaaacacaaagatattagtaaaaaacgt
gacgtagaagtaataatttctgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttctc

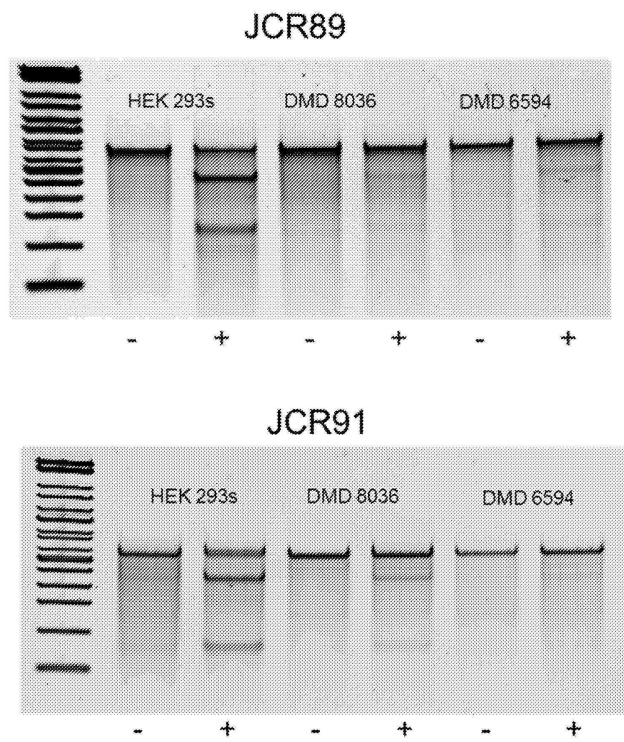
gtaactgaaagtatttctgatttctggcttatatcttgggaaaggacgaaacaccGGGTCTTCGAGAAGACC
Tgttttagagctagaaatagcaagtaaaataaggctagtcggttatcaacttgaaaaagtggcaccgagtcgggtgc
ttttttccgggtggagctcagcttttcttcccttagtgagggttaattgcgctgtggcgtatcatggtcat
agctgttctgtgtgaaattgtatccgctcacaaftccacacacatacagccgggaagcataaagtgaagcc
tggggctccaatagtgagtaactacattaatgctgtgctcactgcccgttccagtcgggaaacctgtc
gtgccagctgcatatgaatcggccaacgcgcggggagaggcgggttgcgtattgggcgtcttccgttctcgc
tactgactcgtcgtcgtcgttctggctgcggcgagcgggtatcagctcactcaaggcggtatacgggtatcc
acagaatcaggggataacgcaggaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgc
gttgcgtggctttttccataggctccgccccctgacgagcatcacaaaaatgacgctcaagtcagaggtggcgaa
accgcagacgactataaagataaccaggcgtttccctggaagctccctgctgctcctgttccgacctgccc
cttaccggatacctgtccgcttctccctcgggaaagcgtggcgcttctcatagctcagctgttaggtatctcag
ttcgggtgtaggtcgttcccaagctgggtgtgtgcacgaacccccgttcagcccagccgtgcgcttatccg
gtaactatcgtctgagtcaccccgtaagacacgactatcggcactggcagcagccactggtaacaggttagc
agagcgaggatgataggcggtctacagagttctgaagtgtggcctaactacggctacactagaaggacagtatt
tggtatctgcgtcgtcgtgaagccagttacctcgaaaaagagttggtagctcttgatccggcaaaacaccaccg
ctgtagcgggtgtgttttttgaagcagcagattacgcgcagaaaaaaggatctcaagaagatctttagatc
ttttctacgggtgtgacgtcagtggaacgaaactcacgttaagggaatttggctcatgagattatcaaaaggat
cttcacgtagatcttttaaaataaaatgaagtttaaatcaatctaaagtatatatgagtaaaactgtgtgaca
gttaccaatgctiaatcagtgaggcacctatctcagcgtatctgtatttcttcatccatagttgctgactcccc
gtcgtgtagataaactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagaccacgctc
accggctccagatttatcagcaataaaccagccagccggaaaggccgagcgcagaagtgttcttcaactttatccg
cctccatccagctatattgttggccgggaagctagagtaagtagttcgcaggttaatggttgcgcaacgtgtt
gccattgtacagcgtcgtgtgtcagctcgtgttgggtatggcttattcagctccggttcccaacgatcaag
gaggtatcagatccccatgttgcgaagaaagcggttagctccttgggtcctccgacgttgcagaagtaagt
tggccgcaggttatcactatggttatggcagcactgcataattcttactgtcatgccatccgtaagatgctt
tctgtgactggtgagtagtcaaccaagtcattctgagaatagtgatgcggcgaccgagttgctcttggccgctc
aatacgggataataccgcgccacatagcagaactttaaaagtgcctcattggaaaacgttcttggggcgaaaaac
tctcaaggatcttaccgctgttgagatccagttcgtatgaaccactcgtgcaccaactgatcttcagcatcttt
actttcaccagcgttctgggtgagcaaaaaaggaaggcaaaatccgcaaaaaagggaataaaggcgacacggaa
atgtgaatactactacttctcttttcaatatttgaagcatttatcagggtattgtctatgagcggataca
tatttgaatgtatttagaaaaataaacaatagggggtccgcacatttccccgaaaagtggccac

[0396]

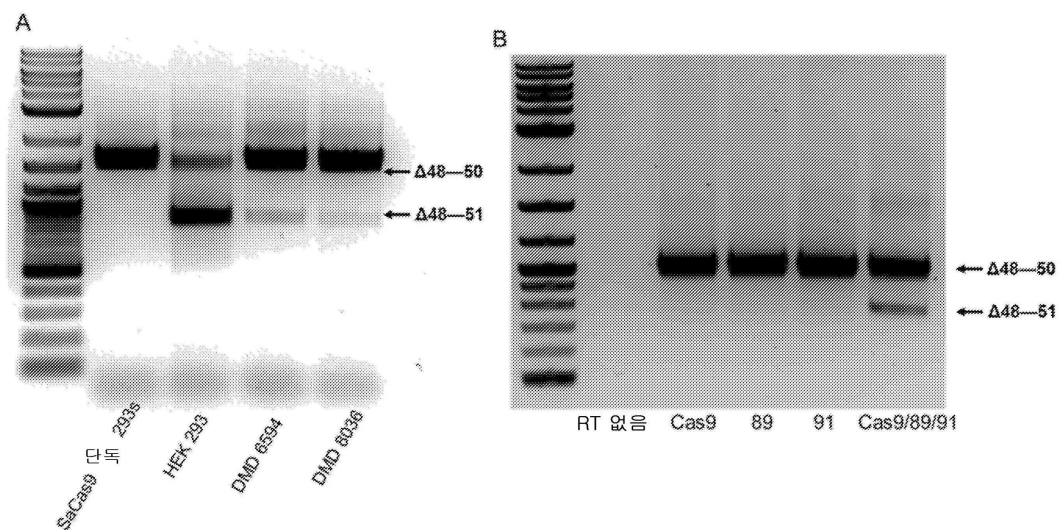
[0397]

도면

도면1



도면2



도면3

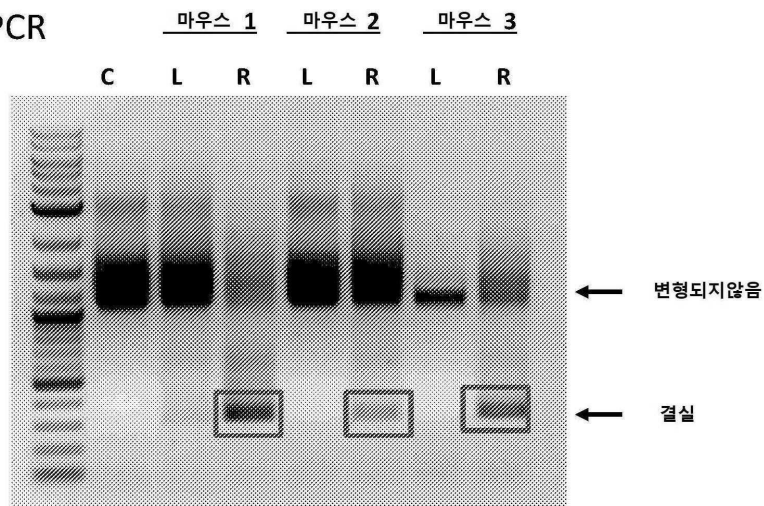
전달 셋업



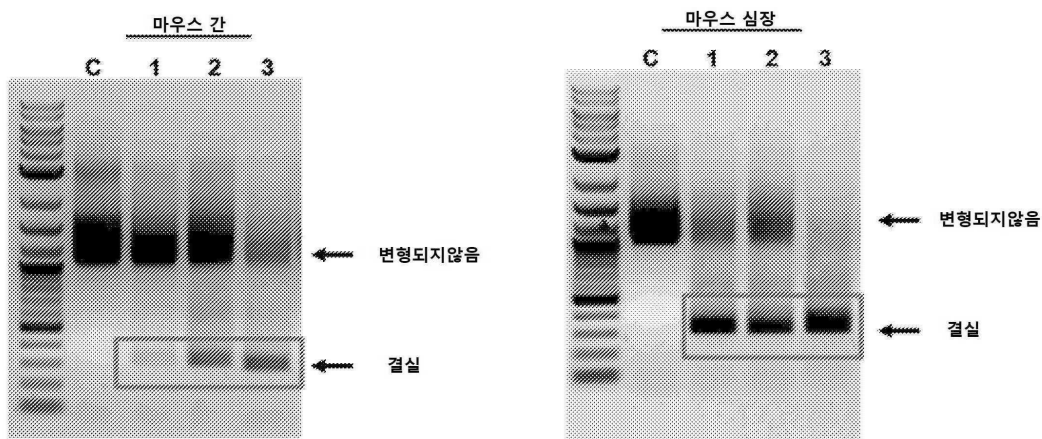
동시 전달 2 바이러스 w/JCR89, 91

도면4

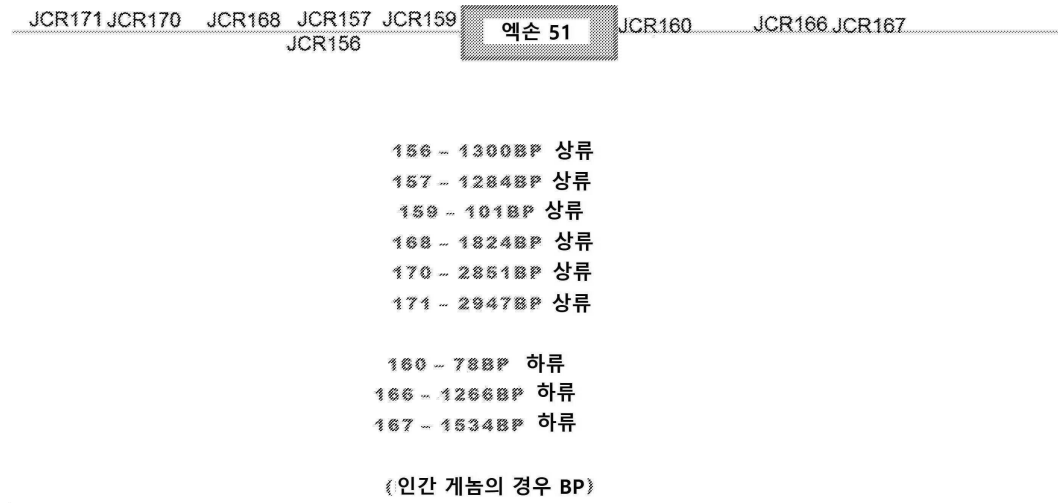
2x gDNA PCR



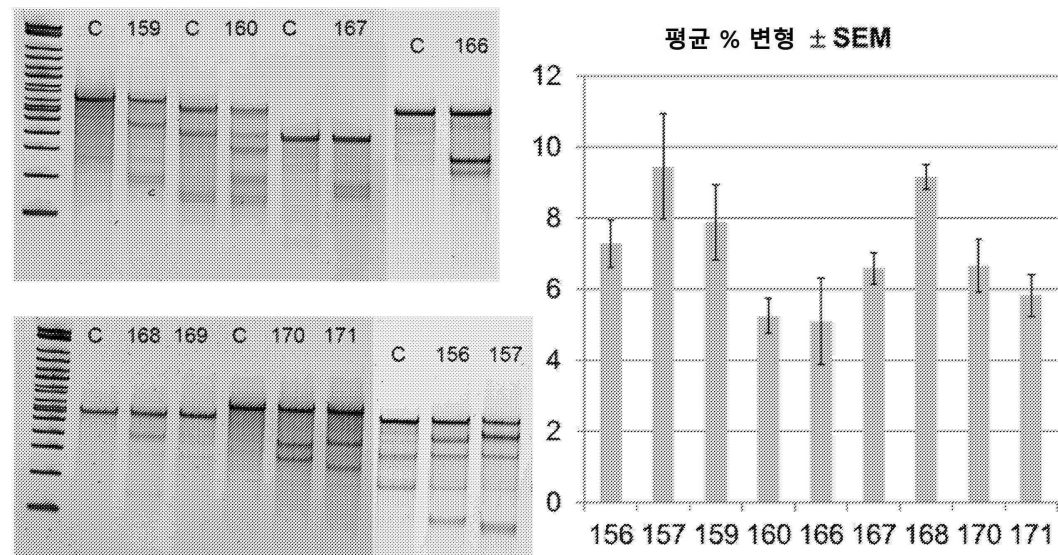
도면5



도면6



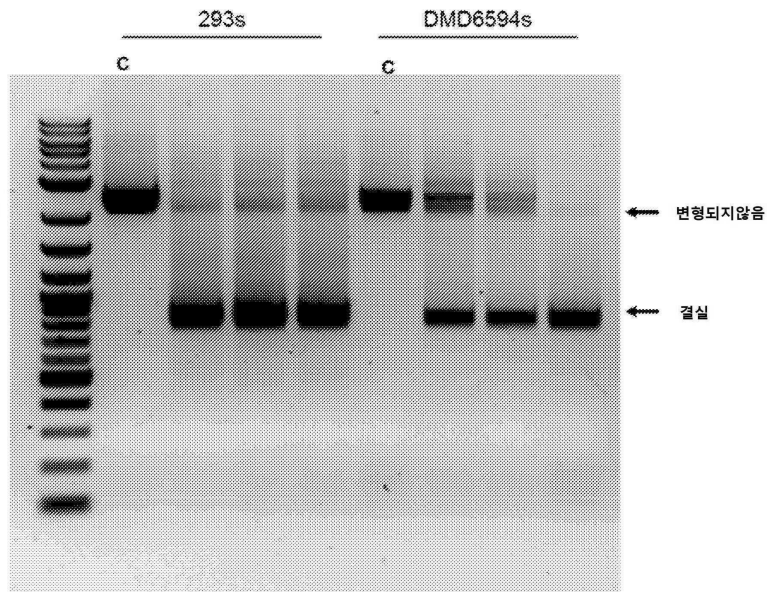
도면7



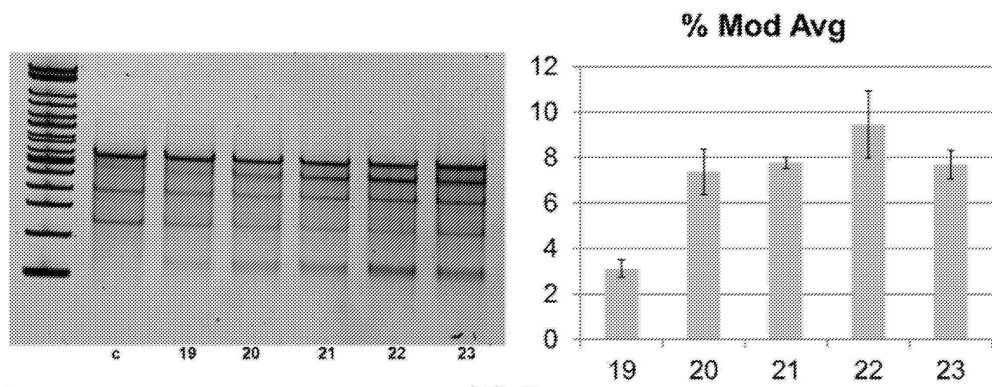
도면8

gRNA	어디에	얼마나 멀리(nt)	% mod ± SEM	인간 탈표적 (미스매치: 게놈내 히트)	붉은털 탈표적
JCR156	상류	1300	7.28% ± 0.66	4 미스매치: 23 3 미스매치: 6 2 미스매치: 0	4 미스매치: 19 3 미스매치: 7 2 미스매치: 1 <2: 0
JCR157	상류	1284	9.45% ± 1.48	*인간 게놈내에서 프로토스페이를 발견하지 못함* 4 미스매치: 6 <4 미스매치: 0	4 미스매치: 5 <4 미스매치: 0
JCR159	상류	101	7.88% ± 1.07	4 미스매치: 16 <4 미스매치: 0	4 미스매치: 15 3 미스매치: 1 <3 미스매치: 0
JCR168	상류	1824	9.17% ± 0.35	4 미스매치: 21 3 미스매치: 1 <3 미스매치: 0	4 미스매치: 13 <4 미스매치: 0
JCR170	상류	2851	6.66% ± 0.74	4 미스매치: 1 3 미스매치: 1 <3 미스매치: 0	4 미스매치: 3 <4 미스매치: 0
JCR171	상류	2947	5.82% ± 0.59	4 미스매치: 5 3 미스매치: 1 <3 미스매치: 0	4 미스매치: 2 3 미스매치: 1 2 미스매치: 1 <2 미스매치: 0
JCR160	하류	78	5.25% ± 0.50	4 미스매치: 2 <4 미스매치: 0	4 미스매치: 4 <4 미스매치: 0
JCR166	하류	1266	5.09% ± 1.21	4 미스매치: 25 3 미스매치: 5 <3: 0	4 미스매치: 17 3 미스매치: 5 <3 미스매치: 0
JCR167	하류	1534	6.59% ± 0.44	4 미스매치: 4 3 미스매치: 2 <3 미스매치: 0	4 미스매치: 10 <4 미스매치: 0

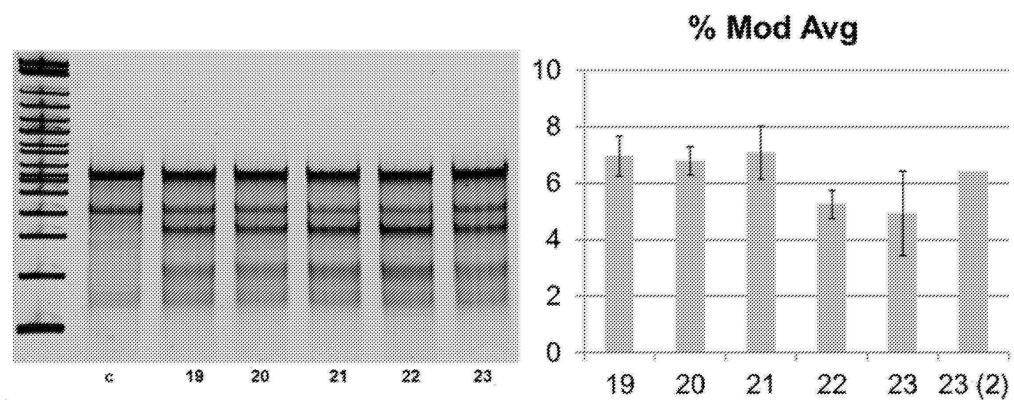
도면9



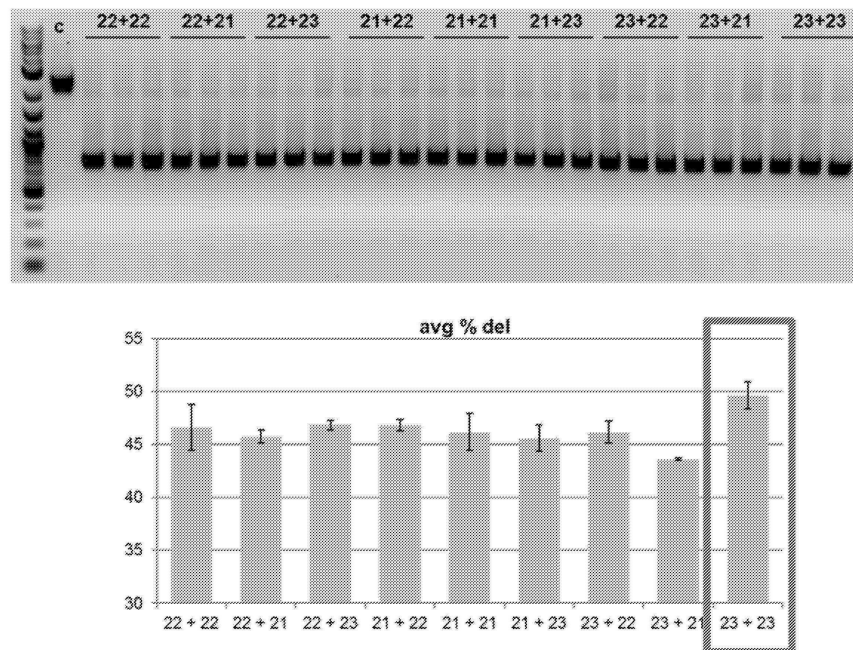
도면10



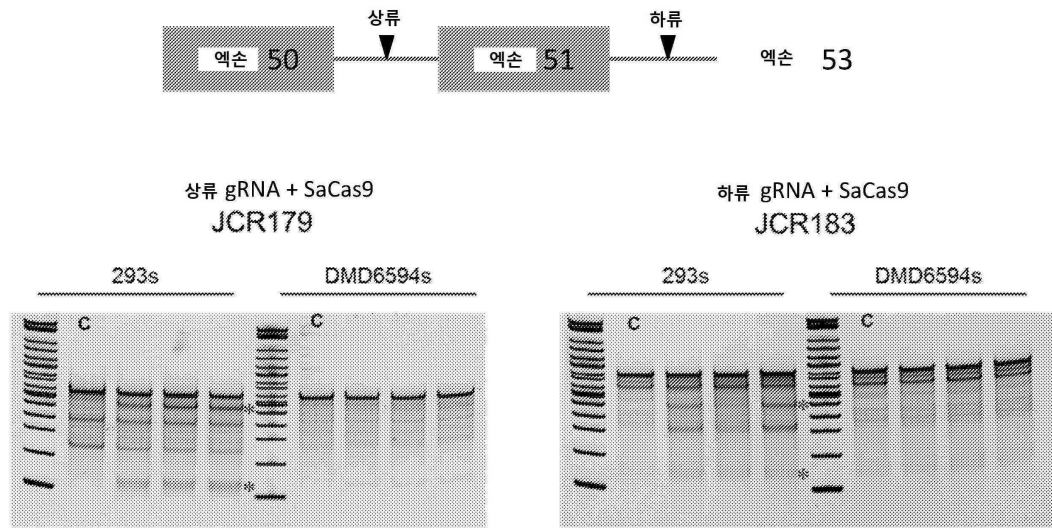
도면11



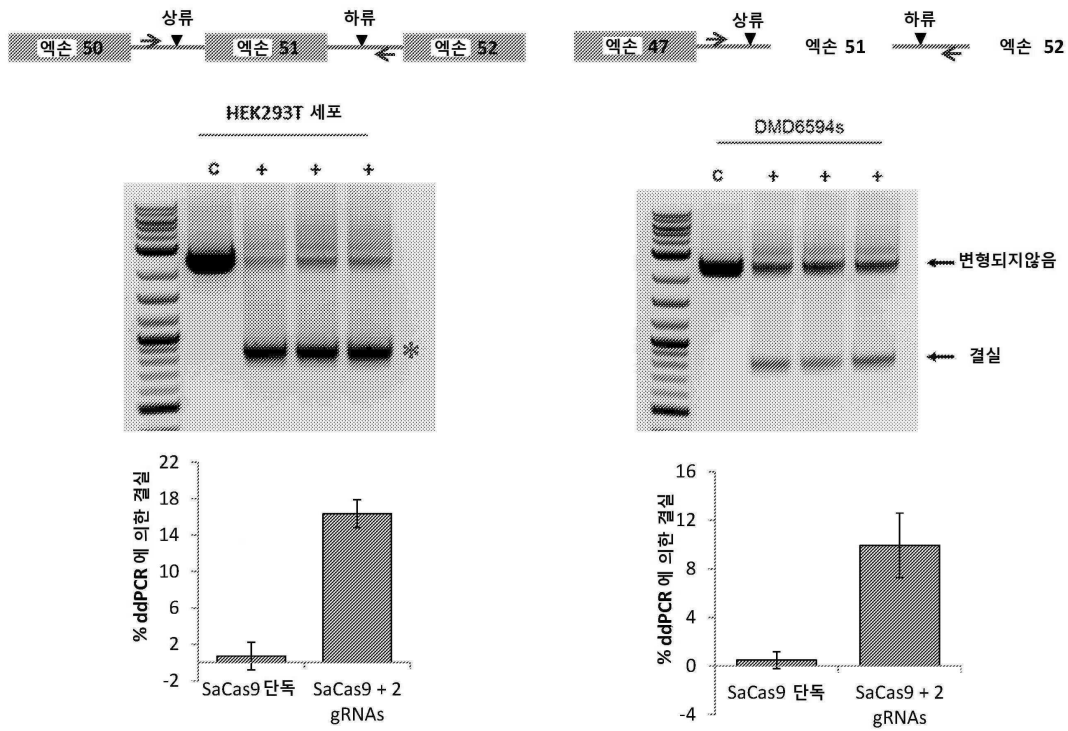
도면12



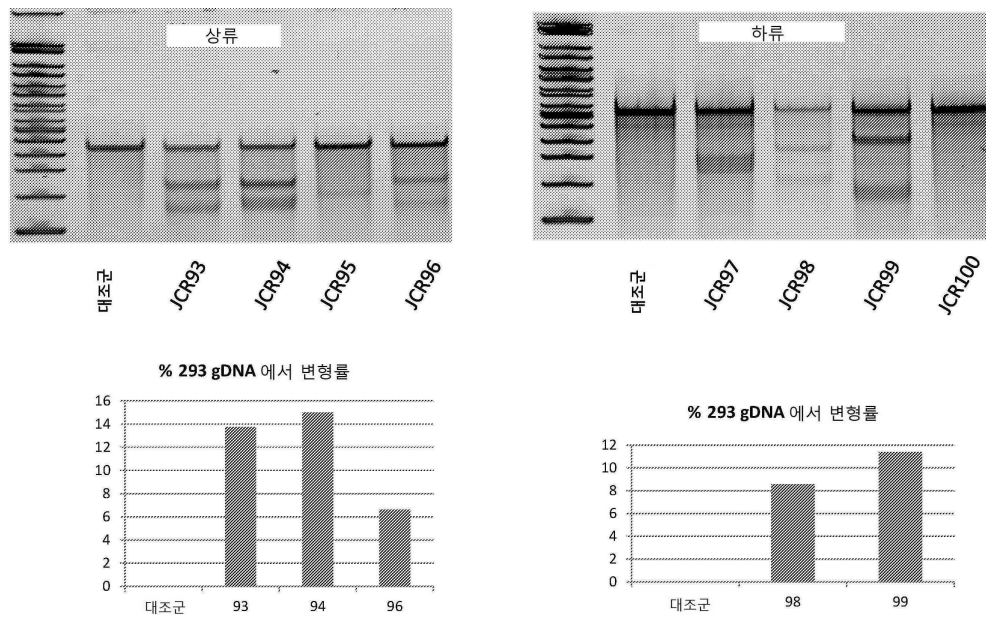
도면13



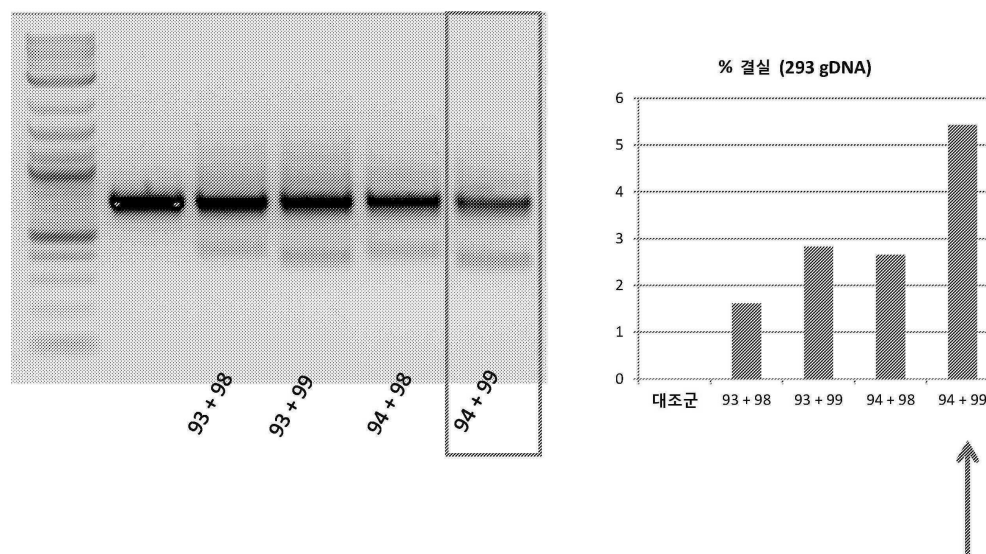
도면14



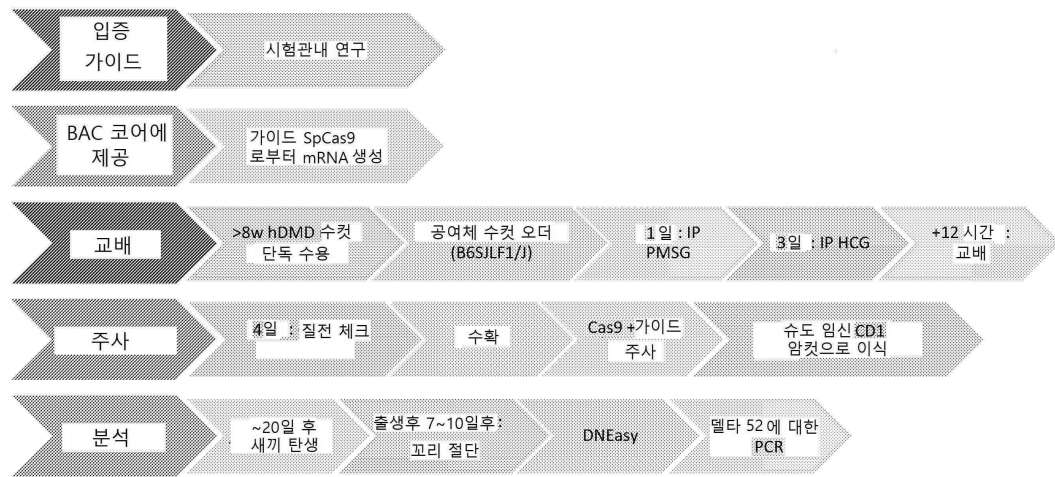
도면18



도면19



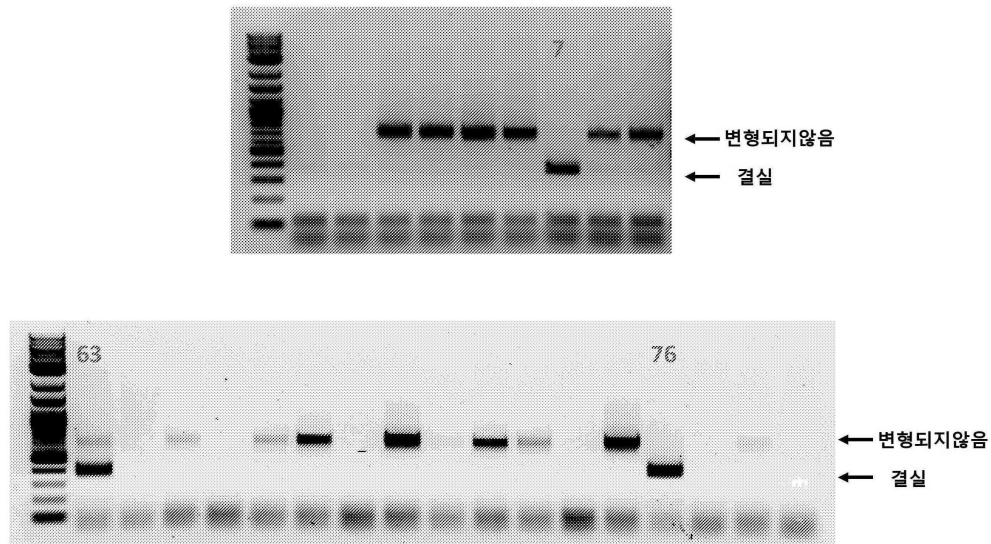
도면20



도면21



도면22



도면23

보라색 볼드체 : JCR94; 녹색 밑줄: JCR99; 회색 이탤릭체: 삼입/결실

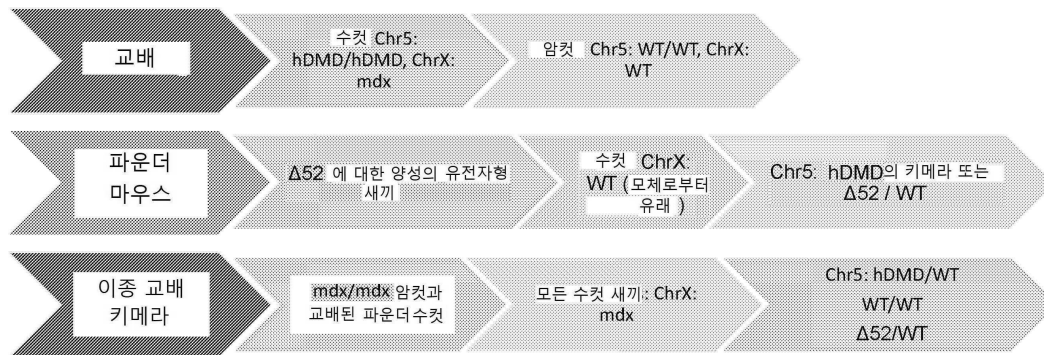
델타 52: AACGCTGAAGAACCCT**GAT**AATAGAAGAAATACATTTTTAAATCAATTCAGG (SEQ ID NO: 46)

파운더 76: AAGAACCCTGATA*-TTATCTTAGTAGATTA* - *ATAGAAGAAATACATT*TTTAAA (SEQ ID NO: 47)

파운더 63: GCTGAAGAACCCTGA-----AAAAATACATTTTTTTATCAATTCAGG (SEQ ID NO: 48)

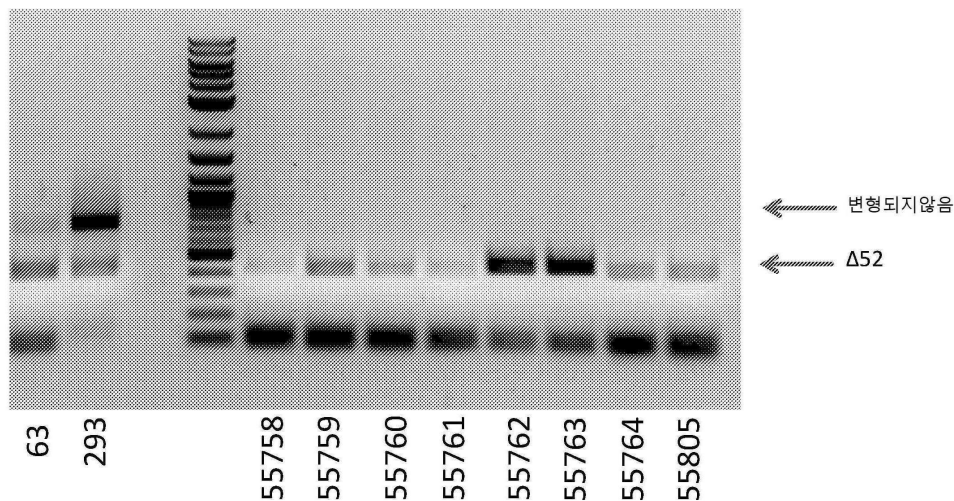
파운더 7: GAAT-----*{19x}*-----**GAT**-TTTCTTGTAGAAGAAATAACAATT—AAATC (SEQ ID NO: 49)

도면24

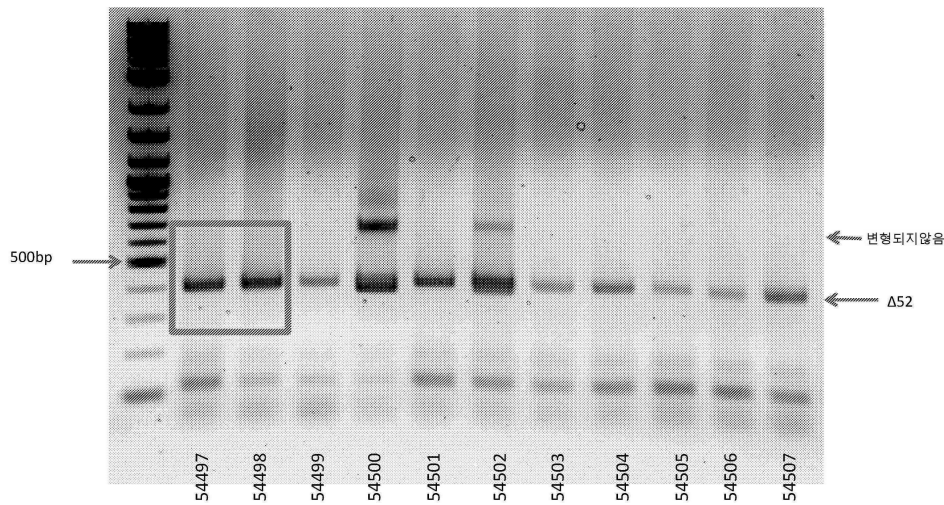


도면25

파운더 수컷 76 + mdx/mdx



도면26



도면27

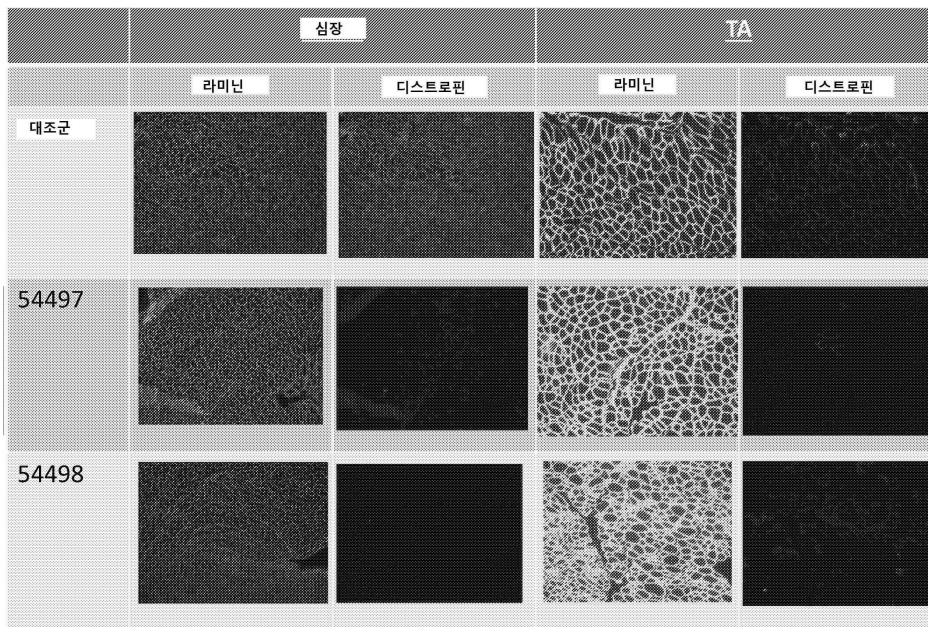
보라색 볼드체 : JCR94; 녹색 밑줄: JCR99; 회색 이탤릭체: 삼입/결실

Delta 52: AACGCTGAAGAACCCTGATAATAGAAGAAATACATTTTAAATCAATTCAGG (SEQ ID NO: 50)

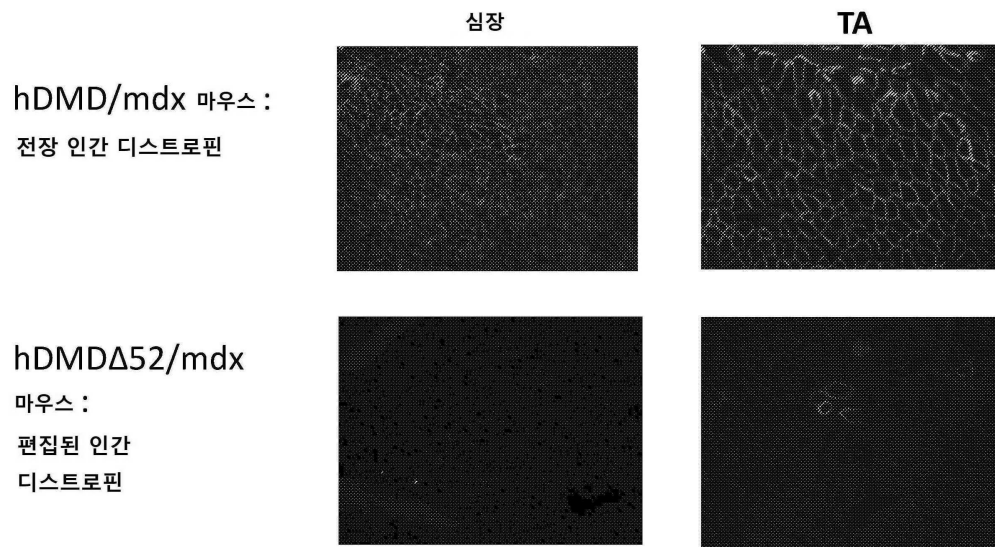
54497: AACGCTGAAGAACCCTGATATTATCTTAGTAGATTAAATAGAAGAAATACATTTTAAAT (SEQ ID NO: 51)

54498: AACGCTGAAGAACCCTGATATTATCTTAGTAGATTAAATAGAAGAAATACATTTTAAAT (SEQ ID NO: 52)

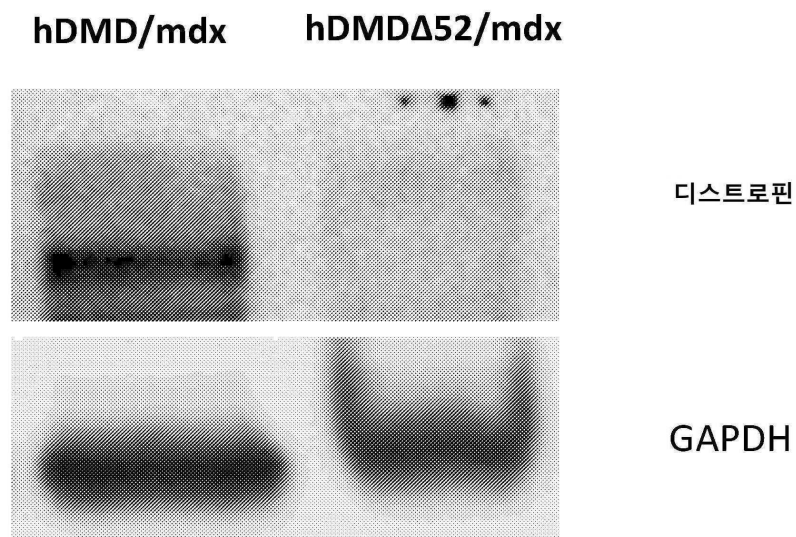
도면28



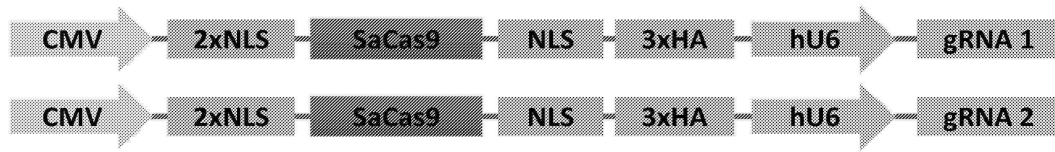
도면29



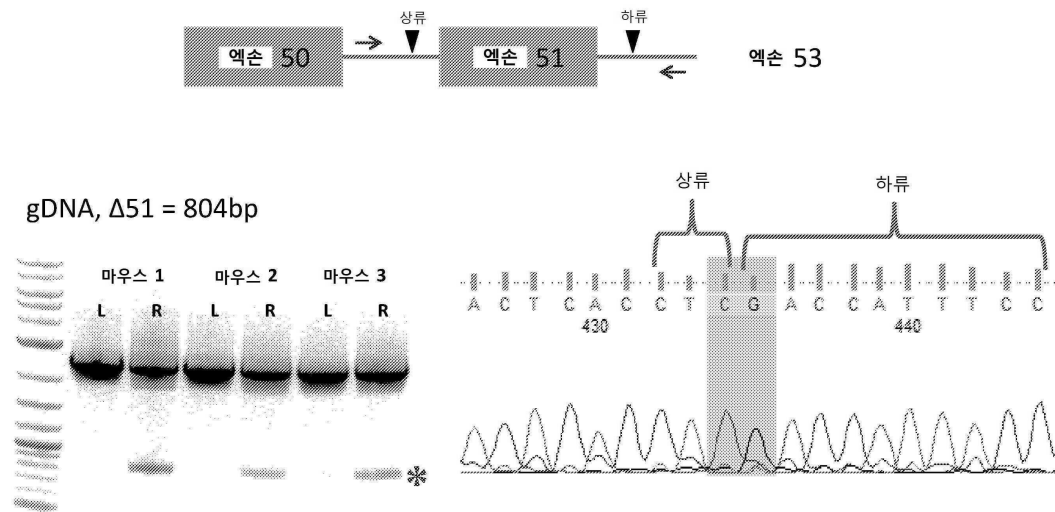
도면30



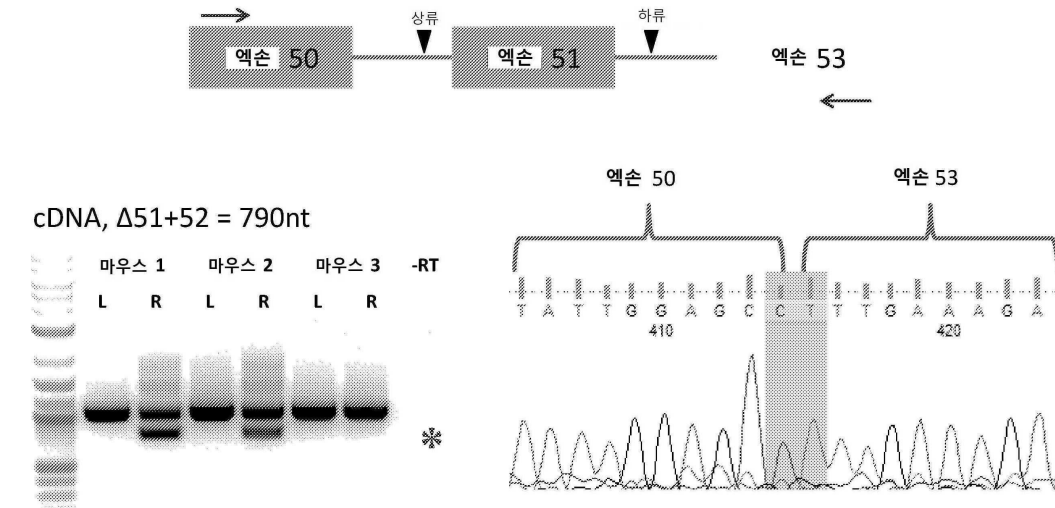
도면34



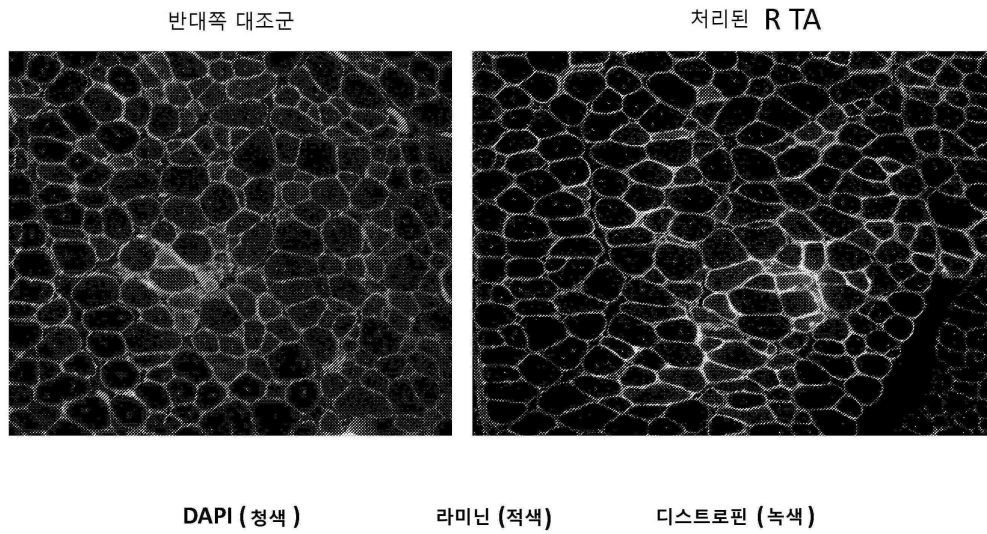
도면35



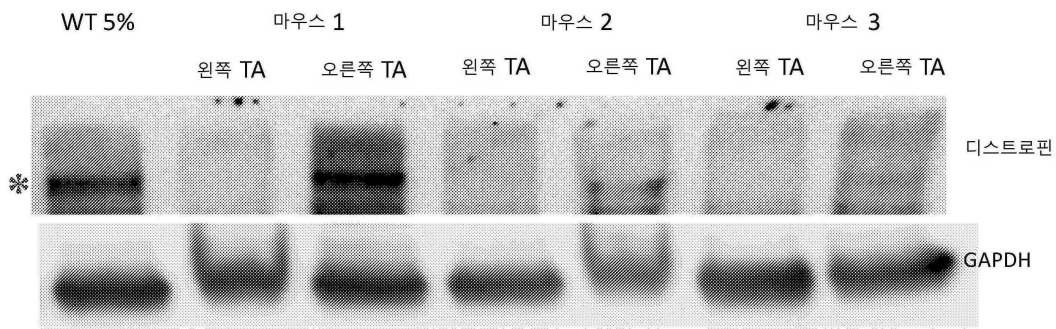
도면36



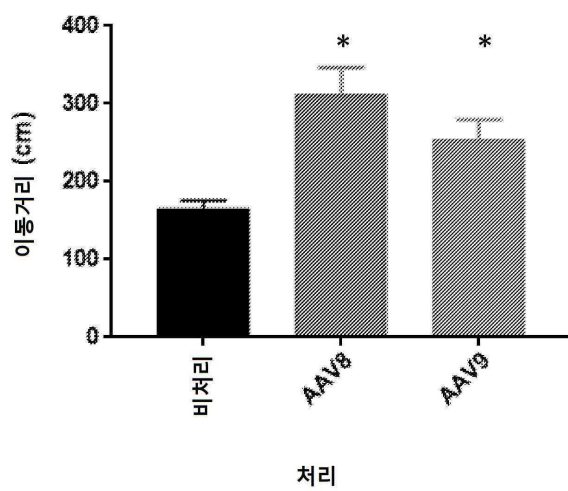
도면37



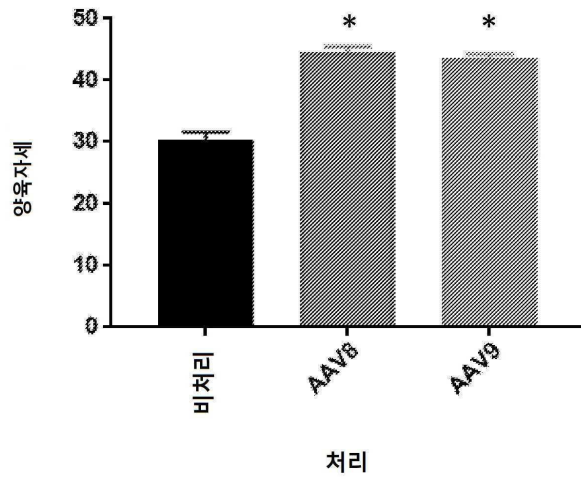
도면38



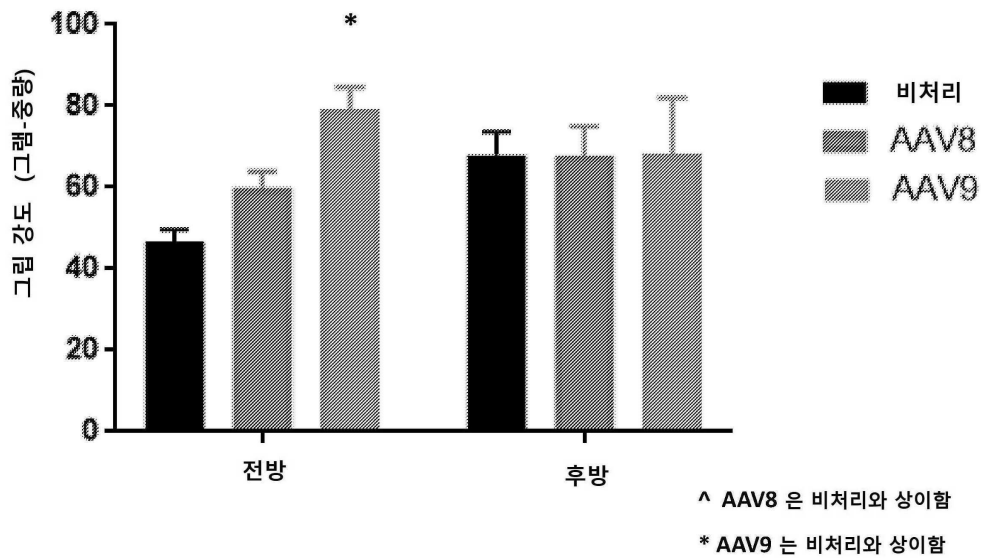
도면39



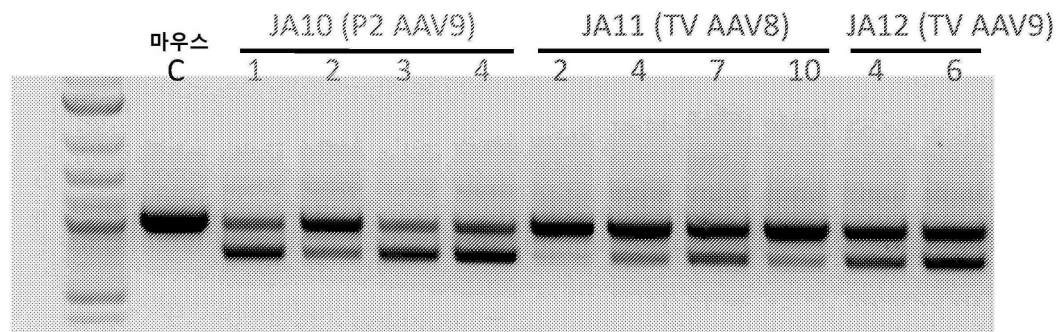
도면40



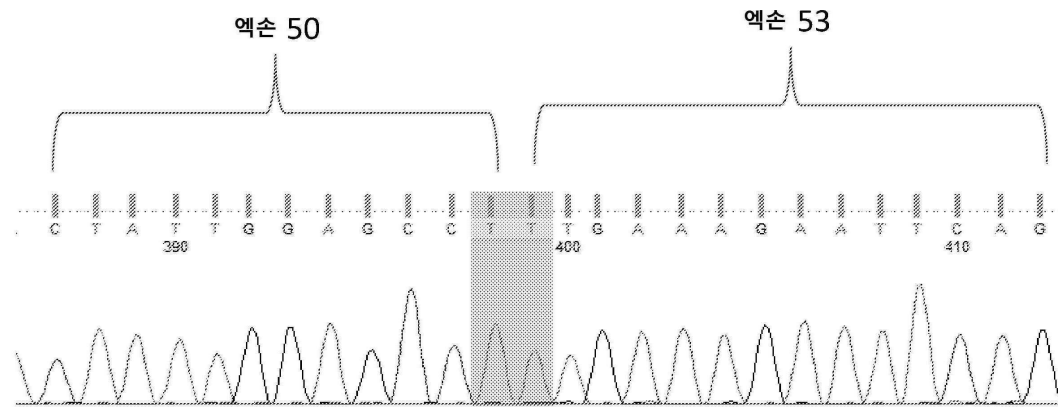
도면41



도면42



도면43



서열 목록

SEQUENCE LISTING

<110> DUKE UNIVERSITY

<120> THERAPEUTIC TARGETS FOR THE CORRECTION OF THE HUMAN DYSTROPHIN
GENE BY GENE EDITING AND METHODS OF USE

<130> 028193-9251-US02

<150> 62/260,712

<151> 2015-11-30

<150> 62/330,336

<151> 2016-05-02

<150> PCT/US2016/064285

<151> 2016-11-30

<160> 109

<170> PatentIn version 3.5

<210> 1

<211> 22

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 1

aaagatatat aatgtcatga at

22

<210> 2

<211> 22

<212> DNA

<213> Artificial sequence
 <220><223> Synthetic
 <400> 2
 gcagaatcaa atataatagt ct 22
 <210> 3
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 3
 caattaaatt tgacttattg tt 22
 <210> 4
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 4
 ctagaccatt tcccaccagt tc 22
 <210> 5
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 5
 aggactttta tttaccaaag ga 22
 <210> 6
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 6
 atccaagtcc atttgattcc ta 22
 <210> 7
 <211> 22

<212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 7
 taattctttc tagaaagagc ct 22

<210> 8
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 8
 ggacatgtgc aagatgcaag ag 22

<210> 9
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 9
 tgtatgtaga agacctctaa gt 22

<210> 10
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 10
 tccccacc actcacctct ga 22

<210> 11
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 11
 ctctgataac ccagctgtgt gt 22

<210> 12
 <211> 19
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 12
 ctctgataac ccagctgtg 19
 <210> 13
 <211> 20
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 13
 ctctgataac ccagctgtgt 20

 <210> 14
 <211> 21
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 14
 ctctgataac ccagctgtgt g 21
 <210> 15
 <211> 23
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 15
 ctctgataac ccagctgtgt gtt 23
 <210> 16
 <211> 19
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 16

ctagaccatt tcccaccag	19
<210> 17	
<211> 20	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<400> 17	
ctagaccatt tcccaccagt	20
<210> 18	
<211> 21	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<400> 18	
ctagaccatt tcccaccagt t	21
<210> 19	
<211> 23	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<400> 19	
ctagaccatt tcccaccagt tct	23
<210> 20	
<211> 7	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<220><221> misc_feature	
<222> (1)..(2)	
<223> n is a, c, g, or t	
<220><221> w	
<222> (7)..(7)	
<223> a or t	

<400> 20

nnagaaw

7

<210> 21

<211> 4

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(1)

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

<220><221> r

<222> (4)..(4)

<223> a or g

<400> 21

naar

4

<210> 22

<211> 5

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(2)

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

<220><221> r

<222> (4)..(5)

<223> a or g

<400> 22

nngrr

5

<210> 23

<211> 6

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(2)

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

<220><221> r

<222> (4)..(5)

<223> a or g

<220><221> misc_feature

<222> (6)..(6)

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

<400> 23

nngrrn

6

<210> 24

<211> 6

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(2)

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

<220><221> r

<222> (4)..(5)

<223> a or g

<400> 24

nngrrt

6

<210> 25

<211> 6

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(2)

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

<220><221> r

<222> (4)..(5)

<223> a or g

<400> 25

nngrrv 6

<210> 26

<211> 4107

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 26

atggataaaa agtacagcat cgggctggac atcgggtacaa actcagtggg gtgggccgtg 60

attacggacg agtacaaggt accctccaaa aaatttaaag tgctgggtaa cacggacaga 120

cactctataa agaaaaatct tattggagcc ttgctgttcg actcaggcga gacagccgaa 180

gccacaaggt tgaagcggac cgccaggagg cgggtatacca ggagaaagaa ccgcatatgc 240

tacctgcaag aaatcttcag taacgagatg gcaaaggttg acgatagctt ttccatcgcc 300

ctggaagaat cttttcttgt tgaggaagac aagaagcacg aacggcaccc catctttggc 360

aatattgtcg acgaagtggc atatcacgaa aagtaccga ctatctacca cctcaggaag 420

aagctggtgg actctaccga taaggcggac ctcagactta tttatttggc actcgcccac 480

atgattaaat ttagaggaca tttcttgatc gagggcgacc tgaaccggga caacagtgac 540

gtcgataagc tgttcatcca acttgtgcag acctacaatc aactgttcga agaaaaccct 600

ataaatgctt caggagtcca cgctaaagca atcctgtccg cgcgcccttc aaaatctaga 660

agacttgaga atctgattgc tcagttgccc ggggaaaaga aaaatggatt gtttggaac 720

ctgatcgccc tcagtctcgg actgaccca aatttcaaaa gtaacttcga cctggccgaa 780

gacgctaagc tccagctgtc caaggacaca tacgatgacg acctcgacaa tctgctggcc 840

cagattgggg atcagtacgc cgatctcttt ttggcagcaa agaacctgtc cgacgccatc 900

ctgttgagcg atatcttgag agtgaacacc gaaattacta aagcaccct tagcgcatct 960

atgatcaagc ggtacgacga gcatcatcag gatctgacct tgctgaaggc tcttgtgagg 1020

caacagctcc ccgaaaaata caaggaaatc ttctttgacc agagcaaaaa cggtacgct 1080

ggctatatag atggtggggc cagtcaggag gaattctata aattcatcaa gccattctc 1140

gagaaaatgg acggcacaga ggagttgtgt gtcaaaacta acaggaggga cctgctgcgg 1200

aagcagcgga cctttgacaa cgggtctatc ccccaccaga ttcactgtgg cgaactgcac 1260

gcaatcctga ggaggcagga ggatttttat ctttttctta aagataaccg cgagaaaata 1320

gaaaagattc ttacattcag gatcccgtag tacgtgggac ctctcgcccg gggcaattca	1380
cggtttgcct ggatgacaag gaagtcagag gagactatta caccttggaa cttcgaagaa	1440
gtggtggaca aggggtgcac tgcccagctt ttcacgcagc ggatgacaaa ttttgacaag	1500
aacctcccta atgagaaggt gctgccccaa cattctctgc tctacgagta ctttacgctc	1560
tacaatgaac tgactaaagt caagtacgtc accgagggaa tgaggaagcc ggcattcctt	1620
agtggagaac agaagaagc gattgtagac ctgttgttca agaccaacag gaaggtgact	1680
gtgaagcaac ttaaagaaga ctactttaag aagatcgaat gttttgacag tgtggaaatt	1740
tcaggggttg aagaccgctt caatgcgtca ttggggactt accatgatct tctcaagatc	1800
ataaaggaca aagacttctt ggacaacgaa gaaaatgagg atattctcga agacatcgctc	1860
ctcacctga cctgtttcga agacagggaa atgatagaag agcgcttgaa aacctatgcc	1920
cacctcttcg acgataaagt tatgaagcag ctgaagcgca ggagatacac aggatgggga	1980
agattgtcaa ggaagctgat caatggaatt agggataaac agatlggcaa gaccatactg	2040
gatttcctca aatctgatgg cttcgccaat aggaacttca tgcaactgat tcacgatgac	2100
tctcttacct tcaaggagga cattcaaaaag gctcaggtga gcgggcaggg agactccctt	2160
catgaacaca tcgcgaattt ggcaggttcc cccgctatta aaaagggcat cttcaaact	2220
gtcaaggtgg tggatgaatt ggtcaaggta atgggcagac ataagccaga aaatatgtg	2280
atcgagatgg cccgcgaaaa ccagaccaca cagaagggcc agaaaaatag tagagagcgg	2340
atgaagagga tcgaggagg catcaaagag ctgggatctc agattctcaa agaacacccc	2400
gtagaaaaca cacagctgca gaacgaaaaa ttgtacttgt actatctgca gaacggcaga	2460
gacatgtacg tcgaccaaga acttgatatt aatagactgt ccgactatga cgtagacat	2520
atcgtcccc agtccttctt gaaggacgac tccattgata acaaagtctt gacaagaagc	2580
gacaagaaca ggggtaaaag tgataatgtg cctagcgagg aggtggtgaa aaaaatgaag	2640
aactactggc gacagctgct taatgcaaag ctattacac aacggaagtt cgataatctg	2700
acgaaagcag agagaggtgg cttgtctgag ttggacaagg cagggtttat taagcggcag	2760
ctggtggaaa ctaggcagat cacaagcac gtggcgcaga ttttgacag ccgatgaac	2820
acaaaatcg acgaaaatga taaactgata cgagaggtca aagttatcac gctgaaaagc	2880
aagctggtgt ccgattttcg gaaagacttc cagttctaca aagttcgca gattaataac	2940
taccatcatg ctacgatgc gtacctgaac gctgttgcg ggaccgcctt gataaagaag	3000
tacccaaagc tggaatccga gtctgtatc ggggattaca aagtgtacga tgtgaggaaa	3060

atgatagcca agtccgagca ggagattgga aaggccacag ctaagtactt cttttattct 3120
aacatcatga atttttttaa gacggaaatt accctggcca acggagagat cagaaagcgg 3180
ccccctatag agacaaatgg tgaaacaggt gaaatcgtct gggataaggg cagggatttc 3240
gctactgtga ggaaggtgct gagtatgcca caggtaaata tcgtgaaaaa aaccgaagta 3300
cagaccggag gattttccaa ggaaagcatt ttgcctaaaa gaaactcaga caagtcacac 3360
gcccgcgaaga aagattggga ccctaagaaa tacgggggat ttgactcacc caccgtagcc 3420
tattctgtgc tgggtggtagc taaggtggaa aaaggaaagt ctaagaagct gaagtcctgt 3480

aaggaactct tgggaatcac tatcatggaa agatcatcct ttgaaaagaa ccctatcgat 3540
ttcctggagg ctaagggtta caaggaggtc aagaaagacc tcatcattaa actgccaaaa 3600
tactctctct tcgagctgga aaatggcagg aagagaatgt tggccagcgc cggagagctg 3660
caaaagggaa acgagcttgc tctgcctcc aaatatgtta atttctcta tctcgttcc 3720
cactatgaaa agctgaaagg gtctcccgaa gataacgagc agaagcagct gtctgtcgaa 3780
cagcacaagc actatctgga tgaaataatc gaacaaataa gcgagttcag caaaagggtt 3840
atcctggcgg atgctaattt ggacaaagta ctgtctgctt ataacaagca ccgggataag 3900

cctattaggg aacaagccga gaatataatt cacctcttta cactcacgaa tctcggagcc 3960
cccgccgect tcaataactt tgatacgact atcgaccgga aacggtatac cagtacaaaa 4020
gaggtcctcg atgccaccct cateccaccag tcaattactg gcctgtacga aacacggatc 4080
gacctctctc aactgggcgg cgactag 4107

<210> 27

<211> 1368

<212> PRT

<213> Artificial sequence

<220><223> Synthetic

<400> 27

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

1 5 10 15

Gly Trp Ala Val Ile Thr Asp Glu Tyr Lys Val Pro Ser Lys Lys Phe

20 25 30

Lys Val Leu Gly Asn Thr Asp Arg His Ser Ile Lys Lys Asn Leu Ile

35 40 45

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

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys
65 70 75 80

Tyr Leu Gln Glu Ile Phe Ser Asn Glu Met Ala Lys Val Asp Asp Ser
85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys
100 105 110

His Glu Arg His Pro Ile Phe Gly Asn Ile Val Asp Glu Val Ala Tyr
115 120 125

His Glu Lys Tyr Pro Thr Ile Tyr His Leu Arg Lys Lys Leu Val Asp
130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His
145 150 155 160

Met Ile Lys Phe Arg Gly His Phe Leu Ile Glu Gly Asp Leu Asn Pro
165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr
180 185 190

Asn Gln Leu Phe Glu Glu Asn Pro Ile Asn Ala Ser Gly Val Asp Ala
195 200 205

Lys Ala Ile Leu Ser Ala Arg Leu Ser Lys Ser Arg Arg Leu Glu Asn
210 215 220

Leu Ile Ala Gln Leu Pro Gly Glu Lys Lys Asn Gly Leu Phe Gly Asn
225 230 235 240

Leu Ile Ala Leu Ser Leu Gly Leu Thr Pro Asn Phe Lys Ser Asn Phe
245 250 255

Asp Leu Ala Glu Asp Ala Lys Leu Gln Leu Ser Lys Asp Thr Tyr Asp
260 265 270

Asp Asp Leu Asp Asn Leu Leu Ala Gln Ile Gly Asp Gln Tyr Ala Asp
275 280 285

Leu Phe Leu Ala Ala Lys Asn Leu Ser Asp Ala Ile Leu Leu Ser Asp
290 295 300

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

305 310 315 320
 Met Ile Lys Arg Tyr Asp Glu His His Gln Asp Leu Thr Leu Leu Lys
 325 330 335

 Ala Leu Val Arg Gln Gln Leu Pro Glu Lys Tyr Lys Glu Ile Phe Phe
 340 345 350
 Asp Gln Ser Lys Asn Gly Tyr Ala Gly Tyr Ile Asp Gly Gly Ala Ser
 355 360 365
 Gln Glu Glu Phe Tyr Lys Phe Ile Lys Pro Ile Leu Glu Lys Met Asp
 370 375 380
 Gly Thr Glu Glu Leu Leu Val Lys Leu Asn Arg Glu Asp Leu Leu Arg
 385 390 395 400

 Lys Gln Arg Thr Phe Asp Asn Gly Ser Ile Pro His Gln Ile His Leu
 405 410 415
 Gly Glu Leu His Ala Ile Leu Arg Arg Gln Glu Asp Phe Tyr Pro Phe
 420 425 430
 Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460

 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560

Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
565 570 575

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

Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
595 600 605

Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
610 615 620

Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
625 630 635 640

His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
645 650 655

Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
660 665 670

Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe
675 680 685

Ala Asn Arg Asn Phe Met Gln Leu Ile His Asp Asp Ser Leu Thr Phe
690 695 700

Lys Glu Asp Ile Gln Lys Ala Gln Val Ser Gly Gln Gly Asp Ser Leu
705 710 715 720

His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
725 730 735

Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
740 745 750

Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
755 760 765

Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
770 775 780

Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
785 790 795 800

Val Glu Asn Thr Gln Leu Gln Asn Glu Lys Leu Tyr Leu Tyr Tyr Leu

805 810 815
Gln Asn Gly Arg Asp Met Tyr Val Asp Gln Glu Leu Asp Ile Asn Arg

820 825 830
Leu Ser Asp Tyr Asp Val Asp His Ile Val Pro Gln Ser Phe Leu Lys

835 840 845

Asp Asp Ser Ile Asp Asn Lys Val Leu Thr Arg Ser Asp Lys Asn Arg
850 855 860

Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
865 870 875 880

Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
885 890 895

Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
900 905 910

Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
915 920 925

Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
980 985 990

Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
1010 1015 1020

Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
1040 1045 1050

Asn Gly	Glu Ile Arg Lys Arg	Pro Leu Ile Glu Thr	Asn Gly Glu
1055	1060	1065	
Thr Gly	Glu Ile Val Trp Asp	Lys Gly Arg Asp Phe	Ala Thr Val
1070	1075	1080	
Arg Lys	Val Leu Ser Met Pro	Gln Val Asn Ile Val	Lys Lys Thr
1085	1090	1095	
Glu Val	Gln Thr Gly Gly Phe	Ser Lys Glu Ser Ile	Leu Pro Lys
1100	1105	1110	
Arg Asn	Ser Asp Lys Leu Ile	Ala Arg Lys Lys Asp	Trp Asp Pro
1115	1120	1125	
Lys Lys	Tyr Gly Gly Phe Asp	Ser Pro Thr Val Ala	Tyr Ser Val
1130	1135	1140	
Leu Val	Val Ala Lys Val Glu	Lys Gly Lys Ser Lys	Lys Leu Lys
1145	1150	1155	
Ser Val	Lys Glu Leu Leu Gly	Ile Thr Ile Met Glu	Arg Ser Ser
1160	1165	1170	
Phe Glu	Lys Asn Pro Ile Asp	Phe Leu Glu Ala Lys	Gly Tyr Lys
1175	1180	1185	
Glu Val	Lys Lys Asp Leu Ile	Ile Lys Leu Pro Lys	Tyr Ser Leu
1190	1195	1200	
Phe Glu	Leu Glu Asn Gly Arg	Lys Arg Met Leu Ala	Ser Ala Gly
1205	1210	1215	
Glu Leu	Gln Lys Gly Asn Glu	Leu Ala Leu Pro Ser	Lys Tyr Val
1220	1225	1230	
Asn Phe	Leu Tyr Leu Ala Ser	His Tyr Glu Lys Leu	Lys Gly Ser
1235	1240	1245	
Pro Glu	Asp Asn Glu Gln Lys	Gln Leu Phe Val Glu	Gln His Lys
1250	1255	1260	
His Tyr	Leu Asp Glu Ile Ile	Glu Gln Ile Ser Glu	Phe Ser Lys
1265	1270	1275	
Arg Val	Ile Leu Ala Asp Ala	Asn Leu Asp Lys Val	Leu Ser Ala

1280 1285 1290
Tyr Asn Lys His Arg Asp Lys Pro Ile Arg Glu Gln Ala Glu Asn
1295 1300 1305
Ile Ile His Leu Phe Thr Leu Thr Asn Leu Gly Ala Pro Ala Ala
1310 1315 1320
Phe Lys Tyr Phe Asp Thr Thr Ile Asp Arg Lys Arg Tyr Thr Ser
1325 1330 1335

Thr Lys Glu Val Leu Asp Ala Thr Leu Ile His Gln Ser Ile Thr
1340 1345 1350
Gly Leu Tyr Glu Thr Arg Ile Asp Leu Ser Gln Leu Gly Gly Asp
1355 1360 1365

<210> 28

<211> 3158

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 28

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt 60
attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac 120

gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgccctgaa acgacggaga 180
aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgaccat 240
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg 300
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg agtgcataac 360
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc 420
aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa 480
gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc 540

aagcagctgc tgaagtgc gaagccttac caccagctgg atcagagctt catcgatact 600
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga agggagcccc 660
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctathtt 720
ccagaagagc tgagaagcgt caagtacgt tataacgcag atctgtacaa cgccctgaat 780
gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag 840
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct 900

aaggagatcc tggtaacga agaggacatc aagggctacc gggtgacaag cactggaaaa 960

ccagagttca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa 1020

atcattgaga acgccgaact gctggatcag attgctaaga tctgactat ctaccagagc 1080

tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgacca ggaagagatc 1140

gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc 1200

aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg 1260

ctgaagctgg tccccaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg 1320

gtggacgatt tcattctgtc acccgtggtc aagcggagct tcattccagag catcaaagtg 1380

atcaacgcca tcattcaaga gtacggcctg cccaatgata tcattatcga gctggctagg 1440

gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag 1500

accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg 1560

attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtattc tctggaggcc 1620

tccccctgga ggacctgctg aacaatccat tcaactacga ggtcgatcat attatcccca 1680

gaagcgtgtc cttcgacaat tcctttaaca acaaggtgct ggtcaagcag gaagagaact 1740

ctaaaaaggg caataggact cctttccagc acctgtctag ttcagattcc aagatctctt 1800

acgaaacctt taaaaagcac attctgaatc tggccaaagg aaagggccgc atcagcaaga 1860

ccaaaaagga gtacctgctg gaagagcggg acatcaacag attctccgtc cagaaggatt 1920

ttattaaccg gaatctggtg gacacaagat acgctactcg cggcctgatg aatctgctgc 1980

gacctatatt ccgggtgaac aatctggatg tgaaagtcaa gtccatcaac ggcgggttca 2040

catcttttct gaggcgcaaa tggaagttaa aaaaggagcg caacaaaggg tacaagcacc 2100

atgccgaaga tgctctgatt atcgcaaatg ccgacttcat ctttaaggag tggaaaaagc 2160

tggacaaagc caagaaagtg atggagaacc agatgttcga agagaagcag gccgaatcta 2220

tgcccgaat cgagacagaa caggagtaca aggagatatt catcactcct caccagatca 2280

agcatatcaa ggatttcaag gactacaagt actctaccg ggtggataaa aagccaaca 2340

gagagctgat caatgacacc ctgtatagta caagaaaaga cgataagggg aataacctga 2400

ttgtgaacaa tctgaacgga ctgtacgaca aagataatga caagctgaaa aagctgatca 2460

acaaaagtcc cgagaagctg ctgatgtacc accatgatcc tcagacatat cagaaactga 2520

agctgattat ggagcagtac ggcgacgaga agaaccact gtataagtac tatgaagaga 2580

ctgggaacta cctgaccaag tatagcaaaa aggataatgg ccccgatgc aagaagatca 2640

agtactatgg gaacaagctg aatgcccatc tggacatcac agacgattac cctaacagtc	2700
gcaacaaggt ggtcaagctg tcaactgaagc catacagatt cgatgtctat ctggacaacg	2760
gcgtgtataa atttgtgact gtcaagaatc tggatgtcat caaaaaggag aactactatg	2820
aagtgaatag caagtgtctac gaagaggctc aaaagctgaa aaagattagc aaccaggcag	2880
agttcatcgc ctctttttac aacaacgacc tgattaagat caatggcgaa ctgtataggg	2940
tcatcggggg gaacaatgat ctgctgaacc gcattgaagt gaatatgatt gacatcactt	3000
accgagagta tctggaaaac atgaatgata agcgccccc tcgaattatc aaaacaattg	3060
cctctaagac tcagagtatc aaaaagtact caaccgacat tctgggaaac ctgtatgagg	3120
tgaagagcaa aaagcacctc cagattatca aaaagggc	3158
<210> 29	
<211> 3159	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<400> 29	
atgaagcggg actacatcct gggcctggac atcgccatca ccagcgtggg ctacggcatc	60
atcgactacg agacacggga cgtgatcgat gccggcgtgc ggctgttcaa agaggccaac	120
gtggaaaaca acgagggcag gcggagcaag agaggcgcca gaaggctgaa gcggcggagg	180
cggcatagaa tccagagagt gaagaagctg ctgttcgact acaacctgct gaccgaccac	240
agcgagctga gcggcatcaa cccctacgag gccagagtga agggcctgag ccagaagctg	300
agcgaggaag agttctctgc cgccctgctg cacctggcca agagaagagg cgtgcacaac	360
gtgaacgagg tggaagagga caccggcaac gagctgtcca ccaaagagca gatcagccgg	420
aacagcaagg ccctggaaga gaaatactg gccgaactgc agctggaacg gctgaagaaa	480
gacggcgaag tgcggggcag catcaacaga ttcaagacca gcgactacgt gaaagaagcc	540
aaacagctgc tgaaggtgca gaagccctac caccagctgg accagagctt catcgacacc	600
tacatcgacc tgctggaaac ccggcggacc tactatgagg gacctggcga gggcagcccc	660
ttcggctgga aggacatcaa agaatggtac gagatgctga tgggccactg cacctacttc	720
cccaggaac tgcggagcgt gaagtacgcc tacaacccg acctgtacaa cgccctgaac	780
gacctgaaca atctcgtgat caccaggac gagaacgaga agctggaata ttacgagaag	840
ttccagatca tcgagaacct gttcaagcag aagaagaagc ccacctgaa gcagatcgcc	900
aaagaaatcc tcgtgaacga agaggatatt aagggtaca gagtgaccag caccggcaag	960
cccaggttca ccaacctgaa ggtgtaccac gacatcaagg acattaccgc ccggaagag	1020

attattgaga acgccgagct gctggatcag attgccaaga tcctgacat ctaccagagc 1080

agcgaggaca tccaggaaga actgaccaat ctgaactccg agctgacca ggaagagatc 1140

gagcagatct ctaatctgaa gggctatacc ggcacccaca acctgagcct gaaggccatc 1200

aacctgatcc tggacgagct gtggcacacc aacgacaacc agatcgctat cttcaaccgg 1260

ctgaagctgg tgcccaagaa ggtggacctg tcccagcaga aagagatccc caccaccctg 1320

gtggacgact tcattctgag ccccgctgtg aagagaagct tcattccagag catcaaagtg 1380

atcaacgcca tcattcaagaa gtacggcctg cccaacgaca tcattatcga gctggcccg 1440

gagaagaact ccaaggacgc ccagaaaatg atcaacgaga tgcagaagcg gaaccggcag 1500

accaacgagc ggatcgagga aatcatccgg accaccggca aagagaacgc caagtacctg 1560

atcgagaaga tcaagctgca cgacatgcag gaaggcaagt gcctgtacag cctggaagcc 1620

atccctctgg aagatctgct gaacaacccc ttcaactatg aggtggacca catcatcccc 1680

agaagcgtgt ccttcgacaa cagcttcaac aacaagggtgc tcgtgaagca ggaagaaaac 1740

agcaagaagg gcaaccggac cccattccag tacctgagca gcagcgacag caagatcagc 1800

tacgaaacct tcaagaagca catctgaat ctggccaagg gcaagggcag aatcagcaag 1860

accaagaag agtatctgct ggaagaacgg gacatcaaca ggttctccgt gcagaaagac 1920

ttcatcaacc ggaacctggt ggataccaga tacgccacca gaggcctgat gaacctgctg 1980

cggagctact tcagagtga caacctggac gtgaaagtga agtccatcaa tggcggttc 2040

accagctttc tcggcggaag gtggaagttt aagaaagagc ggaacaaggg gtacaagcac 2100

cacgccgagg acgccctgat cattgccaac gccgatttca tttcaaaga gtggaagaaa 2160

ctggacaagg ccaaaaaagt gatgaaaaac cagatgttcg aggaaaagca ggccgagagc 2220

atgcccgaga tcgaaaccga gcaggagtac aaagagatct tcattacccc ccaccagatc 2280

aagcacatta aggacttcaa ggactacaag tacagccacc ggggtggacaa gaagcctaat 2340

agagagctga ttaacgacac cctgtactcc acccggaagg acgacaaggg caacaccctg 2400

atcgtgaaca atctgaacgg cctgtacgac aaggacaatg acaagctgaa aaagctgatc 2460

aacaagagcc ccgaaaagct gctgatgtac caccacgacc ccagacctt ccagaaactg 2520

aagctgatta tggaacagta cggcgacgag aagaatcccc tgtacaagta ctacaggaa 2580

accgggaact acctgacca gtactccaaa aaggacaacg gccccgtgat caagaagatt 2640

aagtattacg gcaacaaact gaacgccccat ctggacatca ccgacgacta cccaacagc 2700

agaaacaagg tcgtgaagct gtcctgaag ccctacagat tcgacgtgta cctggacaat 2760

ggcgtgtaca agttcgtgac cgtgaagaat ctggatgtga tcaaaaaaga aaactactac 2820
gaagtgaata gcaagtgcta tgaggaagct aagaagctga agaagatcag caaccaggcc 2880
gagtttatcg cctccttcta caacaacgat ctgatcaaga tcaacggcga gctgtataga 2940
gtgatcggcg tgaacaacga cctgctgaac cggatcgaag tgaacatgat cgacatcacc 3000
taccgcgagt acctggaaaa catgaacgac aagaggcccc ccaggatcat taagacaatc 3060
gcctccaaga ccagagcat taagaagtac agcacagaca ttctgggcaa cctgtatgaa 3120
gtgaaatcta agaagcacc ctagatcatc aaaaagggc 3159

<210> 30

<211> 3159

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 30

atgaagcgca actacatcct cggactggac atcggcatta cctccgtggg atacggcatc 60
atcgattacg aaactagga tgtgatcgac gctggagtca ggctgttcaa agaggcgaac 120
gtggagaaca acgaggggcg gcgctcaaag aggggggccc gccggctgaa gcgccccgc 180
agacatagaa tccagcgctg gaagaagctg ctgttcgact acaaccttct gaccgaccac 240
tccgaacttt ccggcatcaa cccatatgag gctagagtga agggattgtc ccaaaagctg 300
tccgaggaag agttctccgc cgcgttgctc cacctcgcca agcgcagggg agtgcacaat 360

gtgaacgaag tggaagaaga taccggaaac gagctgtcca ccaaggagca gatcagccgg 420
aactccaagg ccctggaaga gaaatcgtg gcggaactgc aactggagcg gctgaagaaa 480
gacggagaag tgcgcggctc gatcaaccgc ttcaagacct cggactacgt gaaggaggcc 540
aagcagctcc tgaaagtga aaaggcctat caccaacttg accagtcctt tatcgatacc 600
tacatcgatc tgctcgagac tcggcggact tactacgagg gtccagggga gggctcccca 660
tttggttga aggatattaa ggagtgttac gaaatgctga tgggacactg cacatacttc 720
cctgaggagc tgcggagcgt gaaatcgc tacaacgcag acctgtacaa cgcgctgaac 780

gacctgaaca atctcgtgat caccggggac gagaacgaaa agctcgagta ttacgaaaag 840
ttccagatta ttgagaacgt gttcaaacag aagaagaagc cgacactgaa gcagattgcc 900
aaggaaatcc tcgtgaacga agaggacatc aagggtatc gagtgacctc aacgggaaag 960
ccggagtcca ccaatctgaa ggtctaccac gacatcaaag acattaccgc ccggaaggag 1020
atcattgaga acgcggagct gttggaccag attgcgaaga ttctgacat ctaccaatcc 1080
tccgaggata ttcaggaaga actcaccaac ctcaacagcg aactgacca ggaggagata 1140

gagcaaatct ccaacctgaa gggctacacc ggaactcata acctgagcct gaaggccatc	1200
aacttgatcc tggacgagct gtggcacacc aacgataacc agatcgctat tttcaatcgg	1260
ctgaagctgg tccccaagaa agtggacctc tcacaacaaa aggagatccc tactaccctt	1320
gtggacgatt tcattctgtc ccccggtgtc aagagaagct tcatacagtc aatcaaagtg	1380
atcaatgcc a ttatcaagaa atacggtctg cccaacgaca ttatcattga gctcgcccg	1440
gagaagaact cgaaggacgc ccagaagatg attaacgaaa tgcagaagag gaaccgacag	1500
actaacgaac ggatcgaaga aatcatccgg accaccggga aggaaaacgc gaagtacctg	1560
atcgaaaaga tcaagctcca tgacatgcag gaaggaaagt gtctgtactc gctggaggcc	1620
attccgctgg aggacttgt gaacaaccct tttaactacg aagtggatca tatcattccg	1680
aggagcgtgt cattcgacaa ttccttcaac aacaaggtcc tcgtgaagca ggaggaaaac	1740
tcgaagaagg gaaaccgcac gccgttccag tacctgagca gcagcgactc caagatttcc	1800
tacgaaacct tcaagaagca catcctcaac ctggcaaagg ggaagggctg catctccaag	1860
accaagaagg aatatctgtt ggaagaaaga gacatcaaca gattctccgt gcaaaaggac	1920
ttcatcaacc gcaacctcgt ggatactaga tacgtactc ggggtctgat gaacctcctg	1980
agaagctact ttagagttaa caatctggac gtgaaggtca agtcgattaa cggaggtttc	2040
acctccttcc tgcggcgcaa gtggaagttc aagaaggaac ggaacaaggg ctacaagcac	2100
cacgccgagg acgcccgtat cattgccaac gccgacttca tcttcaaaga atggaagaaa	2160
cttgacaagg ctaagaaggt catgaaaaac cagatgttcg aagaaaagca ggccgagttt	2220
atgcctgaaa tcgagactga acaggagtac aaggaaatct ttattacgcc acaccagatc	2280
aaacacatca aggatttcaa ggattacaag tactcacatc gcgtggacaa aaagccgaac	2340
agggaactga tcaacgacac cctctactcc acccggaagg atgacaaagg gaatacctc	2400
atcgtcaaca accttaacgg cctgtacgac aaggacaacg ataagctgaa gaagtcatt	2460
aacaagtcgc ccgaaaagtt gctgatgtac caccacgacc ctcagactta ccagaagctc	2520
aagctgatca tggagcagta tggggacgag aaaaaccgt tgtacaagta ctacgaagaa	2580
actgggaatt atctgactaa gtactccaag aaagataacg gccccgtgat taagaagatt	2640
aagtactacg gcaacaagct gaacgcccac ctggacatca ccgatgacta ccctaattcc	2700
cgcaacaagg tcgtcaagct gagcctcaag cctaccggt ttgatgtgta ccttgacaat	2760
ggagtgtaca agttcgtgac tgtgaagaac cttgacgtga tcaagaagga gaactactac	2820
gaagtcaact ccaagtgtca cgaggaagca aagaagttga agaagatctc gaaccaggcc	2880

gagttcattg cctccttcta taacaacgac ctgattaaga tcaacggcga actgtaccgc 2940
 gtcattggcg tgaacaacga tctcctgaac cgcacgaag tgaacatgat cgacatcact 3000
 taccgggaat acctggagaa tatgaacgac aagcggccgc cccggatcat taagactatc 3060
 gcctcaaaga ccagtcgat caagaagtac agcaccgaca tcctgggcaa cctgtacgag 3120
 gtcaaatcga agaagcacc ccagatcatc aagaaggga 3159
 <210> 31
 <211> 3255
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 31
 atggcccaa agaagaagcg gaaggtcggg atccacggag tcccagcagc caagcggaac 60
 tacatcctgg gcctggacat cggcatcacc agcgtgggct acggcatcat cgactacgag 120
 acacgggacg tgatcgatgc cggcgtgcgg ctgttcaaag aggccaacgt ggaaaacaac 180
 gagggcaggc ggagcaagag aggcgccaga aggctgaagc ggcgaggcgc gcatagaatc 240
 cagagagtga agaagctgct gtctgactac aacctgctga ccgaccacag cgagctgagc 300
 ggcatcaacc cctacgagc cagagtgaag ggctgagcc agaagctgag cgaggaagag 360
 ttctctgccg cctgctgca cctggccaag agaagaggcg tgcacaacgt gaacgaggtg 420
 gaagaggaca ccggcaacga gctgtccacc agagagcaga tcagccggaa cagcaaggcc 480
 ctggaagaga aatcgtggc cgaactgcag ctggaacggc tgaagaaaga cggcgaagtg 540
 cggggcagca tcaacagatt caagaccagc gactacgtga aagaagcaa acagctgctg 600
 aaggtgcaga aggcctacca ccagctggac cagagcttca tcgacaccta catcgacctg 660
 ctggaaaccc ggcgaccta ctatgaggga cctggcgagg gcagcccctt cggctggaag 720
 gacatcaaag aatggtacga gatgctgatg ggccactgca cctacttccc cgaggaactg 780
 cggagcgtga agtacgcta caacccgac ctgtacaacg ccctgaacga cctgaacaat 840
 ctctgatca ccagggacga gaacgagaag ctggaatatt acgagaagtt ccagatcatc 900
 gagaacgtgt tcaagcagaa gaagaagccc accctgaagc agatcgcaa agaaatcctc 960
 gtgaacgaag aggatattaa gggctacaga gtgaccagca ccggcaagcc cgagttcacc 1020
 aacctgaagg tgiaccagca catcaaggac attaccgccc ggaaagagat tattgagaac 1080
 gccgagctgc tggatcagat tgccaagatc ctgaccatct accagagcag cgaggacatc 1140
 caggaagaac tgaccaatct gaactccgag ctgaccaggg aagagatcga gcagatctct 1200
 aatctgaagg gctataccgg caccacaac ctgagcctga aggccatcaa cctgatcctg 1260

gacgagctgt ggcacaccaa cgacaaccag atcgctatct tcaaccggct gaagctgggtg	1320
cccaagaagg tggacctgtc ccagcagaaa gagatcccca ccacctggt ggacgacttc	1380
atcctgagcc ccgtcgtgaa gagaagcttc atccagagca tcaaagtgat caacgccatc	1440
atcaagaagt acggcctgcc caacgacatc attatcgagc tggcccgcga gaagaactcc	1500
aaggacgccc agaaaatgat caacgagatg cagaagcgga accggcagac caacgagcgg	1560
atcgaggaaa tcatccggac caccggcaaa gagaacgcca agtacctgat cgagaagatc	1620
aagctgcacg acatgcagga aggcaagtgc ctgtacagcc tggaagccat ccctctggaa	1680
gatctgctga acaaccctt caactatgag gtggaccaca tcatccccag aagcgtgtcc	1740
ttcgacaaca gcttcaaca caagtgctc gtgaagcagg aagaaaacag caagaagggc	1800
aaccggacct cattccagta cctgagcagc agcgacagca agatcagcta cgaaaccttc	1860
aagaagcaca tctgaatct ggccaagggc aaggcgagaa tcagcaagac caagaaagag	1920
tatctgctgg aagaacggga catcaacagg ttctccgtgc agaaagactt catcaaccgg	1980
aacctggtgg ataccagata cgccaccaga ggctgatga acctgctgcg gagctacttc	2040
agagtgaaca acctggacgt gaaagtgaag tccatcaatg gcggtttcac cagctttctg	2100
cggcggaagt ggaagttaa gaaagagcgg aacaaggggt acaagcacca cgccgaggac	2160
gcctgatca ttgccaacgc cgatttcac ttcaaagagt ggaagaaact ggacaaggcc	2220
aaaaaagtga tggaaaacca gatgttcgag gaaaggcagg ccgagagcat gcccagatc	2280
gaaaccgagc aggagtacaa agagatcttc atcaccccc accagatcaa gcacattaag	2340
gacttcaagg actacaagta cagccaccgg gtggacaaga agcctaatag agagctgatt	2400
aacgacacc tgiactccac ccggaaggac gacaaggga acacctgat cgtgaacaat	2460
ctgaacggcc tgtacgaca ggacaatgac aagctgaaaa agctgatcaa caagagcccc	2520
gaaaagctgc tgatgtacca ccacgacccc cagacctacc agaaactgaa gctgattatg	2580
gaacagtacg gcgacgagaa gaatccccctg tacaagtact acgaggaaac cgggaactac	2640
ctgaccaagt actccaaaa ggacaacggc cccgtgatca agaagattaa gtattacggc	2700
aacaaactga acgcccattt ggacatcacc gacgactacc ccaacagcag aaacaaggtc	2760
gtgaagctgt cctgaagcc ctacagattc gacgtgtacc tggacaatgg cgtgtacaag	2820
ttcgtgaccg tgaagaatct ggatgtgatc aaaaaagaaa actactacga agtgaatagc	2880
aagtgtatg aggaagctaa gaagctgaag aagatcagca accaggccga gtttatcgcc	2940
tccttttaca acaacgatct gatcaagatc aacggcgagc tgtatagagt gatcggcgtg	3000

aacaacgacc tgctgaaccg gatcgaagtg aacatgatcg acatcaccta ccgcgagtac 3060
ctggaaaaca tgaacgacaa gaggccccc aggatcatta agacaatcgc ctccaagacc 3120
cagagcatta agaagtacag cacagacatt ctgggcaacc tgtatgaagt gaaatctaag 3180
aagcacctc agatcatcaa aaaggcaca aggccggcgg ccacgaaaaa ggccggccag 3240
gcaaaaaaga aaaag 3255
<210> 32
<211> 3242
<212> DNA
<213> Artificial sequence
<220><223> Synthetic
<400> 32
accggtgcc ccatgtacc atacgatgtt ccagattacg cttcgccgaa gaaaaagcgc 60

aaggtcgaag cgtccatgaa aaggaactac attctggggc tggacatcgg gattacaagc 120
gtgggtatg ggattattga ctatgaaaca agggacgtga tcgacgcagg cgtcagactg 180
ttcaaggagg ccaacgtgga aaacaatgag ggacggagaa gcaagagggg agccaggcgc 240
ctgaaacgac ggagaaggca cagaatccag agggtgaaga aactgctgtt cgattacaac 300
ctgctgaccg accattctga gctgagtgga attaatcctt atgaagccag ggtgaaaggc 360
ctgagtcaga agctgtcaga ggaagagttt tccgcagctc tgctgcacct ggctaagcgc 420
cgaggagtgc ataacgtcaa tgagggtgaa gaggacaccg gcaacgagct gtctacaaag 480

gaacagatct cagcaatag caaagctctg gaagagaagt atgtcgaga gctgcagctg 540
gaacggctga agaaagatgg cgaggtgaga gggtaatta ataggttcaa gacaagcgac 600
tacgtcaaag aagccaagca gctgctgaaa gtgcagaagg cttaccacca gctggatcag 660
agcttcatcg atacttatat cgacctgctg gagactcgga gaacctacta tgagggacca 720
ggagaaggga gccccttcgg atggaagac atcaaggaat ggtacgagat gctgatggga 780
cattgcacct atttccaga agagctgaga agcgtcaagt acgttataa cgcagatctt 840
acaacgccct gaatgacctg aacaacctgg tcatcaccag ggatgaaaac gagaaactgg 900

aatactatga gaagttccag atcatcgaac acgtgtttta gcagaagaaa aagcctacac 960
tgaaacagat tgctaaggag atccttggtca acgaagagga catcaagggc taccgggtga 1020
caagcactgg aaaaccagag ttcaccaatc tgaaagtgtg tcacgatatt aaggacatca 1080
cagcagcgaa agaaatcatt gagaacgccg aactgctgga tcagattgct aagatcctga 1140
ctatctacca gagctccgag gacatccagg aagagctgac taacctgaac agcgagctga 1200
cccaggaaga gatcgaacag attagtaatc tgaaggggta caccggaaca cacaacctgt 1260

ccctgaaagc tatcaatctg attctggatg agctgtggca taaaaacgac aatcagattg	1320
caatctttta cgggtgaag ctggccccaa aaaaggtgga cctgagtcag cagaaagaga	1380
tcccaaccac actggtggac gatttcattc tgcacccgt ggtcaagcgg agcttcatcc	1440
agagcatcaa agtgcataac gccatcatca agaagtacgg cctgccaat gatatcatta	1500
tcgagctggc tagggagaag aacagcaagg acgcacagaa gatgatcaat gagatgcaga	1560
aacgaaaccg gcagaccaat gaacgcattg aagagattat ccgaactacc gggaaagaga	1620
acgcaaagta cctgattgaa aaaatcaagc tgcacgatat gcaggaggga aagtgtctgt	1680
attctctgga ggccatcccc ctggaggacc tgcgaacaa tccattcaac tacgaggtcg	1740
atcatattat cccagaaagc gtgtccttcg acaattcctt taacaacaag gtgctggtca	1800
agcaggaaga gaactctaaa aagggaata ggactccttt ccagtacctg tctagttcag	1860
attccaagat ctcttacgaa acctttaaaa agcacattct gaatctggcc aaaggaaagg	1920
gccgcatcag caagacaaa aaggagtacc tgctggaaga gcgggacatc aacagattct	1980
ccgtccagaa ggattttatt aaccggaatc tgggtggacac aagatacgt actcgggcc	2040
tgatgaatct gctgcgatcc tatttcggg tgaacaatct ggatgtgaaa gtcaagtcca	2100
tcaacggcgg gttcacatct tttctgaggc gcaaatggaa gtttaaaaag gagcgcaaca	2160
aagggtacaa gcaccatgcc gaagatgctc tgattatcgc aaatgccgac ttcattctta	2220
aggagtggaa aaagctggac aaagccaaga aagtgatgga gaaccagatg ttcgaagaga	2280
agcaggccga atctatgccc gaaatcgaga cagaacagga gtacaaggag attttcatca	2340
ctctcacca gatcaagcat atcaaggatt tcaaggacta caagtactct caccgggtgg	2400
ataaaaagcc caacagagag ctgatcaatg acaccctgta tagtacaaga aaagacgata	2460
aggggaatac cctgattgtg aacaatctga acggactgta cgacaaagat aatgacaagc	2520
tgaaaaagct gatcaacaaa agtcccgaga agctgctgat gtaccacat gatcctcaga	2580
catatcagaa actgaagctg attatggagc agtacggcga cgagaagaac cactgtata	2640
agtactatga agagactggg aactacctga ccaagtatag caaaaaggat aatggccccg	2700
tgatcaagaa gatcaaglac tatgggaaca agctgaatgc ccatctggac atcacagacg	2760
attaccctaa cagtcgcaac aagggtgtca agctgtcact gaagccatc agattcgatg	2820
tctatctgga caacggcgtg tataaatttg tgaactgtca gaatctggat gtcataaaaa	2880
aggagaacta ctatgaagtg aatagcaagt gctacgaaga ggctaaaaag ctgaaaaaga	2940
ttagcaacca ggcagagttc atgcctcct tttacaacaa cgacctgatt aagatcaatg	3000

gcgaactgta tagggtcatc ggggtgaaca atgatctgct gaaccgcatt gaagtgaata 3060
 tgattgacat cacttaccga gagtatctgg aaaacatgaa tgataagcgc cccctcgaa 3120
 ttatcaaaac aattgcctct aagactcaga gtatcaaaaa gtactcaacc gacattctgg 3180
 gaaacctgta tgaggagaag agcaaaaagc accctcagat tatcaaaaag ggctaagaat 3240
 tc 3242

<210> 33

<211> 1053

<212> PRT

<213> Artificial sequence

<220><223> Synthetic

<400> 33

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

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

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

130 135 140
 Leu Glu Glu Lys Tyr Val Ala Glu Leu Gln Leu Glu Arg Leu Lys Lys
 145 150 155 160
 Asp Gly Glu Val Arg Gly Ser Ile Asn Arg Phe Lys Thr Ser Asp Tyr

165 170 175
 Val Lys Glu Ala Lys Gln Leu Leu Lys Val Gln Lys Ala Tyr His Gln
 180 185 190
 Leu Asp Gln Ser Phe Ile Asp Thr Tyr Ile Asp Leu Leu Glu Thr Arg

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

 260 265 270
 Glu Lys Leu Glu Tyr Tyr Glu Lys Phe Gln Ile Ile Glu Asn Val Phe
 275 280 285
 Lys Gln Lys Lys Lys Pro Thr Leu Lys Gln Ile Ala Lys Glu Ile Leu
 290 295 300
 Val Asn Glu Glu Asp Ile Lys Gly Tyr Arg Val Thr Ser Thr Gly Lys
 305 310 315 320
 Pro Glu Phe Thr Asn Leu Lys Val Tyr His Asp Ile Lys Asp Ile Thr

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

 385 390 395 400
 Asn Leu Ile Leu Asp Glu Leu Trp His Thr Asn Asp Asn Gln Ile Ala
 405 410 415

Ile Phe Asn Arg Leu Lys Leu Val Pro Lys Lys Val Asp Leu Ser Gln
420 425 430
Gln Lys Glu Ile Pro Thr Thr Leu Val Asp Asp Phe Ile Leu Ser Pro
435 440 445
Val Val Lys Arg Ser Phe Ile Gln Ser Ile Lys Val Ile Asn Ala Ile
450 455 460
Ile Lys Lys Tyr Gly Leu Pro Asn Asp Ile Ile Ile Glu Leu Ala Arg
465 470 475 480
Glu Lys Asn Ser Lys Asp Ala Gln Lys Met Ile Asn Glu Met Gln Lys
485 490 495
Arg Asn Arg Gln Thr Asn Glu Arg Ile Glu Glu Ile Ile Arg Thr Thr
500 505 510
Gly Lys Glu Asn Ala Lys Tyr Leu Ile Glu Lys Ile Lys Leu His Asp
515 520 525
Met Gln Glu Gly Lys Cys Leu Tyr Ser Leu Glu Ala Ile Pro Leu Glu
530 535 540
Asp Leu Leu Asn Asn Pro Phe Asn Tyr Glu Val Asp His Ile Ile Pro
545 550 555 560
Arg Ser Val Ser Phe Asp Asn Ser Phe Asn Asn Lys Val Leu Val Lys
565 570 575
Gln Glu Glu Asn Ser Lys Lys Gly Asn Arg Thr Pro Phe Gln Tyr Leu
580 585 590
Ser Ser Ser Asp Ser Lys Ile Ser Tyr Glu Thr Phe Lys Lys His Ile
595 600 605
Leu Asn Leu Ala Lys Gly Lys Gly Arg Ile Ser Lys Thr Lys Lys Glu
610 615 620
Tyr Leu Leu Glu Glu Arg Asp Ile Asn Arg Phe Ser Val Gln Lys Asp
625 630 635 640
Phe Ile Asn Arg Asn Leu Val Asp Thr Arg Tyr Ala Thr Arg Gly Leu
645 650 655
Met Asn Leu Leu Arg Ser Tyr Phe Arg Val Asn Asn Leu Asp Val Lys

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

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

 770 775 780
 Asn Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu
 785 790 795 800
 Ile Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu
 805 810 815
 Lys Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His
 820 825 830
 Asp Pro Gln Thr Tyr Gln Lys Leu Lys Leu Ile Met Glu Gln Tyr Gly

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

 900 905 910

Arg Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val
915 920 925
Lys Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser
930 935 940
Lys Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala
945 950 955 960
Glu Phe Ile Ala Ser Phe Tyr Asn Asn Asp Leu Ile Lys Ile Asn Gly
965 970 975
Glu Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile
980 985 990
Glu Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met
995 1000 1005
Asn Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys
1010 1015 1020
Thr Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu
1025 1030 1035
Tyr Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly
1040 1045 1050

<210> 34

<211> 3159

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 34

atgaaaagga actacattct ggggctggcc atcgggatta caagcgtggg gtatgggatt	60
attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac	120
gtggaaaaca atgagggacg gagaagcaag aggggagcca ggcgcctgaa acgacggaga	180
aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgaccat	240
tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg	300
tcagaggaag agttttccgc agctctgctg cacctggcta agcgcaggagg agtgcataac	360
gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc	420
aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa	480

gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc	540
aagcagctgc tgaaagtga gaaggcttac caccagctgg atcagagctt catcgatact	600
tatatcgacc tgctggagac tccgagaacc tactatgagg gaccaggaga agggagcccc	660
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatfff	720
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat	780
gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag	840
ttccagatca tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct	900
aaggagatcc tggtaacga agaggacatc aagggtacc ggggtgacaag cactggaaaa	960
ccagagtcca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa	1020
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc	1080
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgacca ggaagagatc	1140
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc	1200
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg	1260
ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg	1320
gtggacgatt tcattctgtc acccgtgggtc aagcggagct tcattccagag catcaaagtg	1380
atcaacgcca tcattcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg	1440
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag	1500
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg	1560
attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtattc tctggaggcc	1620
atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc	1680
agaagcgtgt ccttcgacaa ttcttttaac aacaaggtgc tggtaacga ggaagagaac	1740
tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct	1800
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg catcagcaag	1860
acaaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat	1920
tttattaacc ggaatctggt ggacacaaga tacgtactc gcggcctgat gaatctgctg	1980
cgatcctatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc	2040
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac	2100
catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga gtggaaaaag	2160
ctggacaaaag ccaagaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct	2220

atgcccgaag tcgagacaga acaggagtag aaggagattt tcatcactcc tcaccagatc 2280

aagcatatca aggtatttcaa ggactacaag tactctcacc gggtggataa aaagcccaac 2340

agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg 2400

attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac 2460

aacaaaagtc ccgagaagct gctgatgtac caccatgac ctcagacata tcagaaactg 2520

aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag 2580

actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagatc 2640

aagtactatg ggaacaagct gaatgcccac ctggacatca cagacgatta ccctaacagt 2700

cgcaacaagg tggtaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac 2760

ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat 2820

gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca 2880

gagttcatcg cctccttita caacaacgac ctgattaaga tcaatggcga actgtatagg 2940

gtcatcgggg tgaacaatga tctgtgaac cgcattgaag tgaatatgat tgacatcact 3000

taccgagagt atctggaaa catgaatgat aagcgcccc ctcgaattat caaaacaatt 3060

gcctctaaga ctcagaglat caaaaagtag tcaaccgaca ttctgggaaa cctgtatgag 3120

gtgaagagca aaaagcacc tcagattatc aaaaagggc 3159

<210> 35

<211> 3159

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 35

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg gtatgggatt 60

attgactatg aaacaaggga cgtgatcgac gcaggcgtca gactgttcaa ggaggccaac 120

gtggaaaaca atgagggacg gagaagcaag aggggagcca ggccgctgaa acgacggaga 180

aggcacagaa tccagagggt gaagaaactg ctgttcgatt acaacctgct gaccgacat 240

tctgagctga gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg 300

tcagaggaag agttttccgc agctctgctg cacctggcta agcggcagg agtgcataac 360

gtcaatgagg tggaagagga caccggcaac gagctgtcta caaaggaaca gatctcacgc 420

aatagcaaag ctctggaaga gaagtatgtc gcagagctgc agctggaacg gctgaagaaa 480

gatggcgagg tgagagggtc aattaatagg ttcaagacaa gcgactacgt caaagaagcc 540

aagcagctgc tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact	600
tatatcgacc tgctggagac tcgagaaacc tactatgagg gaccaggaga agggagcccc	660
ttcggatgga aagacatcaa ggaatggtac gagatgctga tgggacattg cacctatitt	720
ccagaagagc tgagaagcgt caagtacgct tataacgcag atctgtacaa cgccctgaat	780
gacctgaaca acctggtcat caccagggat gaaaacgaga aactggaata ctatgagaag	840
ttccagatca tcgaaaacgt gttaaagcag aaaaaaagc ctacactgaa acagattgct	900
aaggagatcc tggtaacga agaggacatc aagggtacc gggtgacaag cactggaaaa	960
ccagagtcca ccaatctgaa agtgtatcac gatattaagg acatcacagc acggaaagaa	1020
atcattgaga acgccgaact gctggatcag attgctaaga tcctgactat ctaccagagc	1080
tccgaggaca tccaggaaga gctgactaac ctgaacagcg agctgacca ggaagagatc	1140
gaacagatta gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc	1200
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat ctttaaccgg	1260
ctgaagctgg tccccaaaaa ggtggacctg agtcagcaga aagagatccc aaccacactg	1320
gtggacgatt tcattctgtc acccgtggtc aagcggagct tcattccagag catcaaagtg	1380
atcaacgcca tcatcaagaa gtacggcctg cccaatgata tcattatcga gctggctagg	1440
gagaagaaca gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag	1500
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc aaagtacctg	1560
attgaaaaaa tcaagctgca cgatatgcag gagggaaagt gtctgtattc tctggaggcc	1620
atccccctgg aggacctgct gaacaatcca ttcaactacg aggtcgatca tattatcccc	1680
agaagcgtgt ccttcgacaa ttcttttaac aacaaggtgc tggtaagca ggaagaggcc	1740
tctaaaaagg gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct	1800
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggcgg catcagcaag	1860
acaaaaagg agtacctgct ggaagagcgg gacatcaaca gattctccgt ccagaaggat	1920
tttattaacc ggaatctggt ggacacaaga tacgtactc gcggcctgat gaatctgctg	1980
cgatcttatt tccgggtgaa caatctggat gtgaaagtca agtccatcaa cggcgggttc	2040
acatcttttc tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac	2100
catgccgaag atgtctgat tatcgaaat gccgacttca tctttaagga gtggaaaaag	2160
ctggacaaag ccaagaaagt gatggagaac cagatgttcg aagagaagca ggccgaatct	2220
atgcccgaaa tcgagacaga acaggagtac aaggagattt tcattactcc tcaccagatc	2280
aagcatatca aggtattcaa ggactacaag tactctcacc ggggtgataa aaagcccaac	2340

agagagctga tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg 2400

attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa aaagctgac 2460

aacaaaagtc ccgagaagct gctgatgtac caccatgac ctcagacata tcagaaactg 2520

aagctgatta tggagcagta cggcgacgag aagaaccac tgtataagta ctatgaagag 2580

actgggaact acctgaccaa gtatagcaaa aaggataatg gccccgtgat caagaagac 2640

aagtactatg ggaacaagct gaatgcccac ctggacatca cagacgatta ccctaacagt 2700

cgcaacaagg tggccaagct gtcactgaag ccatacagat tcgatgtcta tctggacaac 2760

ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca tcaaaaagga gaactactat 2820

gaagtgaata gcaagtgcta cgaagaggct aaaaagctga aaaagattag caaccaggca 2880

gagttcatcg cctcctttta caacaacgac ctgattaaga tcaatggcga actgtatagg 2940

gtcatcgagg tgaacaatga tctgtgaac cgcattgaag tgaatatgat tgacatcact 3000

taccgagagt atctggaaaa catgaatgat aagcgcccc ctcgaattat caaaacaatt 3060

gcctctaaga ctcagagtat caaaaagtac tcaaccgaca ttctgggaaa cctgtatgag 3120

gtgaagagca aaaagcaccc tcagattatc aaaaagggc 3159

<210> 36

<211> 5

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<220><221> misc_feature

<222> (1)..(1)

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

<220><221> misc_feature

<222> (4)..(4)

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

<400> 36

nggng 5

<210> 37

<211> 3238

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 37

ctaaattgta agcgtaata ttttgtaaa attcggtta aattttgtt aaatcagctc	60
atttttaac caataggccg aaatcgcaa aatcccttat aaatcaaaag aatagaccga	120
gatagggttg agtgtgttc cagtttgaa caagagtcca ctattaaaga acgtggactc	180
caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc	240
ctaatcaagt tttttgggt cgaggtgccg taaagcacta aatcggaacc ctaaaggag	300
ccccgattt agagcttgac ggggaaagcc ggcaacgtg gcgagaaagg aagggaagaa	360
agcgaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgtgc gcgtaaccac	420
cacaccgcc gcgcttaatg cgccgtaca gggcgctcc cattcgccat tcaggtgcg	480
caactgttg gaagggcgat cgggtcgggc ctcttcgcta ttacgccagc tggcgaaagg	540
gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttccagt cagcagttg	600
taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtacaa	660
gcttgctat ttccatgat tccttcataat ttgcatatac gatacaagc tgttagagag	720
ataattggaa ttaatttgac tgtaaacaca aagatattag taaaaatac gtgacgtaga	780
aagtaataat ttcttgggt gtttcagtt ttaaaattat gttttaaat ggactatcat	840
atgcttaccg taacttgaag gtatttcgat ttcttggctt tatatatctt gtggaaagga	900
cgaaacacca acacacagct gggttatcag aggttttagt actctggaaa cagaatctac	960
taaaacaagg caaatgccg tgtttatctc gtcaacttgt tggcgagatt ttttgcggc	1020
cgcccggtt ggagctccag cttttgttcc ctttagtgag ggttaattgc gcgcttggcg	1080
taatcatggt catagctgtt tcctgtgtga aattgttacc cgctcacaat tccacacaac	1140
atagagccg gaagcataaa gtgtaaagcc tgggtgcct aatgagtgag ctaactcaca	1200
ttaattgcgt tgcgtcact gcccgtttc cagtcgggaa acctgtcgtg ccagctgcat	1260
taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc ttccgttcc	1320
tcgtcactg actcgtcgcg ctcggtcgtt cggtgcggc gagcggtatc agctcactca	1380
aaggcggtaa tacggttacc cacagaatca ggggataacg caggaaagaa catgtgagca	1440
aaaggccagc aaaaggccag gaaccgtaaa aaggccgcgt tgctggcggt tttccatagg	1500
ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa gtcagagtg gcgaaacccg	1560
acaggactat aaagatacca ggcgtttccc cctggaagct ccttcgtgcg ctctcctgtt	1620
ccgaccctgc cgcttaccgg atacctgtcc gcctttctcc cttcggaag cgtggcgctt	1680
tctcatagct cagctgttag gtatctcagt tcggtgtagg tcgttcgctc caagctgggc	1740

tgtgtgcacg aacccccctg tcagcccgac cgtgcgccct tatccggtaa ctatcgtctt 1800

 gagtccaacc cggtaagaca cgacttatcg ccactggcag cagccactgg taacaggatt 1860
 agcagagcga ggtatgtagg cggctgtaca gatttcttga agtggtggcc taactacggc 1920
 tacactagaa ggacagtatt tggatatctgc gctctgtga agccagttac cttcgaaaa 1980
 agagtggta gctcttgatc cggcaaaaa accaccgtg gtagcggtgg tttttttgtt 2040
 tgcaagcagc agattacgag cagaaaaaaa ggatctcaag aagatccttt gatcttttct 2100
 acggggtctg acgctcagtg gaacgaaaac tcacgttaag ggatttttgt catgagatta 2160
 tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagtttta atcaatctaa 2220

 agtatatatg agtaaaactg gtctgacagt taccaatgct taatcagtga ggcacctatc 2280
 tcagcgatct gtctatttcg ttcattcata gttgcctgac tccccgtcgt gtagataact 2340
 acgatacggg agggcttacc atctggcccc agtgctgcaa tgataccgag agaccacgc 2400
 tcaccggctc cagatttatt agcaataaac cagccagccg gaagggccga gcgcagaagt 2460
 ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttccggga agctagagta 2520
 agtagttcgc cagttaatag ttgctgcaac gttgttgcca ttgctacagg catcgtggtg 2580
 tcacgctcgt cgtttggtat ggcttcattc agctccggtt cccaacgac aaggcgagtt 2640

 acatgatecc ccatgttgtg caaaaaagcg gttagctcct tcggctcctc gatcgttgtc 2700
 agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt 2760
 actgtcatgc catccgtaag atgcttttct gtgactggtg agtactcaac caagtcattc 2820
 tgagaatagt gtatgcggcg accgagttgc tcttgcccg cgtcaatacg ggataatacc 2880
 gcgccacata gcagaacttt aaaagtgtc atcattggaa aacgttcttc ggggcgaaaa 2940
 ctctcaagga tcttaccgt gttgagatcc agttcgatgt aaccactcg tgcaccaac 3000
 tgatcttcag catcttttac ttccaccagc gtttctgggt gagcaaaaac aggaaggcaa 3060

 aatgccgcaa aaaagggaat aaggcgaca cggaaatgtt gaatactcat actcttcctt 3120
 tttcaatatt attgaagcat ttatcagggt tatgtctca tgagcgata catatttgaa 3180
 tgtatttaga aaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgccac 3238

 <210> 38
 <211> 3238
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 38

ctaaattgta agcgtttaata ttttgttaaa attcgcgta aatttttgtt aaatcagctc	60
attttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga	120
gatagggttg agtgtttgtc cagtttgtaa caagagtcca ctattaaaga acgtggactc	180
caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc	240
ctaatacaagt tttttgggtt cgaggtgccg taaagcacta aatcggaacc ctaaaggag	300
cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa	360
agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgctgc gcgtaaccac	420
cacacccgcc gcgttaatg gcgcgtaca gggcgctcc cattgccat tcaggtcgcg	480
caactgttgg gaagggcgat cgggtcgggc ctcttcgcta ttacgccagc tggcgaaagg	540
gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cacgacgttg	600
taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtaccaa	660
gcttgcttat ttcccatgat tcttcataat ttgcatatac gatacaaggc tgttagagag	720
ataattggaa ttaatttgac tgtaaacaca aagatattag tacaaaatac gtgacgtaga	780
aagtaataat ttcttgggta gtttgcagtt ttaaaattat gttttaaaat ggactatcat	840
atgcttaccg taacttgaat gtatttcgat ttcttggctt tataatatctt gtggaaagga	900
cgaacacca gaactggtgg gaaatggtct aggttttagt actctggaaa cagaatctac	960
taaaacaagg caaatgccg tgtttatctc gtcaacttgt tggcgagatt tttttcggc	1020
cgcccgcggt ggagctccag cttttgttcc ctttagtgag ggttaattgc gcgcttggcg	1080
taatcatggt catagctgtt tcctgtgtga aattgttata cgctcacaat tccacacaac	1140
atacgagccg gaagcataaa gtgtaaagcc tggggtgcct aatgagttag ctaactcaca	1200
ttaattgcgt tcgctcact gcccgcttcc cagtcgggaa acctgtcgtg ccagctgcat	1260
taatgaatcg gccaacgcgc ggggagaggc ggtttgcgta ttgggcgctc ttccgcttcc	1320
tcgctcactg actcgtcgcg ctcggtcgtt cggctgcggc gagcggtatc agctcactca	1380
aaggcggtaa tacggttatc cacagaatca ggggataacg caggaaagaa catgtgagca	1440
aaaggccagc aaaaggccag gaaccgtaaa aaggcccgct tgctggcggt tttccatagg	1500
ctccgcccc ctgacgagca tcacaaaaat cgacgtcaa gtcagagggt gcgaaaccgc	1560
acaggactat aaagatacca ggcgtttccc cctggaagct cctcgtgcg ctctcctgtt	1620
ccgacctgc cgcttaccgg atacctgtcc gcctttctcc cttcggaag cgtggcgctt	1680
tctcatagct cacgtgttag gtatctcagt tcggttagg tcgttcgctc caagctgggc	1740
tgtgtgcacg aacccccgt tcagcccgac cgtgcgcct tatccggtaa ctatcgtctt	1800

gagtccaacc cggtaaagaca cgacttatcg ccactggcag cagccactgg taacaggatt 1860

agcagagcga ggtatgtagg cgggtctaca gatttcttga agtgggtggc taactacggc 1920

tacactagaa ggacagtatt tggatatctgc gctctgctga agccagttac cttcggaaaa 1980

agagttggta gctcttgatc cggcaaacaa accaccgtg gtagcgggtg tttttttgtt 2040

tgcaagcagc agattacgag cagaaaaaaa ggatctcaag aagatccttt gatcttttct 2100

acgggggtctg acgtcagtg gaacgaaaac tcacgttaag ggatttttgt catgagatta 2160

tcaaaaagga tcttcaccta gatcctttta aattaaaaat gaagttttaa atcaatctaa 2220

agtatatatg agtaaaactg gtctgacagt taccaatgct taatcagtga ggcacctatc 2280

tcagcgatct gtctatttcg ttcacccata gttgcctgac tccccgtcgt gtagataact 2340

acgatacggg agggccttacc atctggcccc agtgcctgcaa tgataccgag agaccacgc 2400

tcaccggctc cagattttatc agcaataaac cagccagccg gaagggccga gcgcagaagt 2460

ggtcctgcaa ctttatccgc ctccatccag tctattaatt gttgccggga agctagagta 2520

agtagttcgc cagttaatag tttagcgaac gttgttgcca ttgctacagg catcgtggtg 2580

tcacgctcgt cgtttgggat ggcttcattc agctccggtt cccaacgac aaggcgagtt 2640

acatgatccc ccatgttggt caaaaaagcg gttagctcct tcggtcctcc gatcgttgct 2700

agaagtaagt tggccgcagt gttatcactc atggttatgg cagcactgca taattctctt 2760

actgtcatgc catccgtaag atgcttttct gtgactgggtg agtactcaac caagtcattc 2820

tgagaatagt gtatcggcg accgagttgc tcttgcccgg cgtcaatacg ggataatacc 2880

gcgccacata gcagaacttt aaaagtgtc atcattggaa aacgttcttc ggggcgaaaa 2940

ctctcaagga tcttaccgt gttgagatcc agttcgatgt aaccactcg tgcaccaac 3000

tgatcttcag catcttttac tttcaccagc gtttctgggt gagcaaaaac aggaaggcaa 3060

aatgcgcaa aaaagggaat aaggcgaca cggaaatgtt gaatactcat actcttcctt 3120

tttcaatatt attgaagcat ttatcagggt tattgtctca tgagcggata catatttgaa 3180

tgtatttaga aaaataaaca aataggggtt ccgcgcacat ttccccgaaa agtgccac 3238

<210> 39

<211> 7450

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 39

cctgcaggca gctgcgcgt cgtcgtctca ctgaggccgc ccgggcgtcg ggcgaccttt 60

ggtcgccccg cctcagtgag cgagcgagcg cgagagagg gaggggccaa ctccatcact	120
aggggttcct gggcctcta gactcgaggc gttgacattg attattgact agttattaat	180
agtaatcaat tacgggggtca ttagttcata gcccataat ggagttccgc gttacataac	240
ttacggtaaa tggcccgctt ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa	300
tgacgtatgt tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt	360
atttacggta aactgcccac ttggcagtag atcaagtgt tcatatgcca agtacgcccc	420
ctattgacgt caatgacggt aaatggcccc cctggcatta tgcccagtag atgaccttat	480
gggactttcc tacttggcag tacatctacg tattagtcac cgctattacc atggtgatgc	540
ggttttggca gtacatcaat gggcgtggat agcggtttga ctacgggga ttccaagtc	600
tccaccccat tgacgtcaat gggagtttgt ttggcacca aaatcaacgg gactttccaa	660
aatgtcgtaa caactccgcc ccattgacgc aaatgggagg taggcgtgt cgggtggagg	720
tctatataag cagagctctc tggctaacta ccggtgccac catggcccca aagaagaagc	780
ggaaggtcgg tatccacgga gtcccagcag ccaagcggaa ctacatcctg ggcctggaca	840
tcggcatcac cagcgtgggc tacggcatca tcgactacga gacacgggac gtgatcgatg	900
ccggcgtgcg gctgttcaaa gaggccaacg tggaaaacaa cgagggcagg cggagcaaga	960
gaggcgccag aaggctgaag cggcggaggc ggcatagaat ccagagagt aagaagctgc	1020
tgttcgacta caacctgctg accgaccaca gcgagctgag cggcatcaac ccctacgagg	1080
ccagagtga gggcctgagc cagaagctga gcgaggaaga gttctctgcc gccctgctgc	1140
acctggccaa gagaagaggc gtgcacaacg tgaacgaggt ggaagaggac accggcaacg	1200
agctgtccac caaagagcag atcagccgga acagcaaggc cctggaagag aaatacgtgg	1260
ccgaactgca gctggaacgg ctgaagaaag acggcgaagt gcggggcagc atcaacagat	1320
tcaagaccag cgactacgtg aaagaagcca aacagctgct gaaggtgcag aaggcctacc	1380
accagctgga ccagagcttc atcgacacct acatcgacct gctggaaacc cggcggacct	1440
actatgaggg acctggcgag ggcagcccct tcggctggaa ggacatcaaa gaatgtacg	1500
agatgctgat gggccactgc acctacttcc ccgaggaact gcggagcgtg aagtacgcct	1560
acaacgccga cctgtacaac gccctgaacg acctgaacaa tctcgtgatc accagggacg	1620
agaacgagaa gctggaatat tacgagaagt tccagatcat cgagaacgtg ttcaagcaga	1680
agaagaagcc caccctgaag cagatcgcca aagaaatcct cgtgaacgaa gaggatatta	1740
agggctacag agtgaccagc accggcaagc ccgagttcac caacctgaag gtgtaccacg	1800
acatcaagga cattaccgcc cggaagagaa ttattgagaa cgccgagctg ctggatcaga	1860

ttgccaagat cctgaccatc taccagagca gcgaggacat ccaggaagaa ctgaccaatc	1920
tgaactccga gctgaccag gaagagatcg agcagatctc taatctgaag ggctataaccg	1980
gcacccacaa cctgagcctg aaggccatca acctgatcct ggacgagctg tggcacacca	2040
acgacaacca gatcgctatc ttcaaccggc tgaagctggg gcccaagaag gtggacctgt	2100
cccagcagaa agagatcccc accaccctgg tggacgactt catcctgagc cccgtcgtga	2160
agagaagctt catccagagc atcaaagtga tcaacgccat catcaagaag tacggcctgc	2220
ccaacgacat cattatcgag ctggcccgcg agaagaactc caaggacgcc cagaaaatga	2280
tcaacgagat gcagaagcgg aaccggcaga ccaacgagcg gatcgaggaa atcatccgga	2340
ccaccggcaa agagaacgcc aagtacctga tcgagaagat caagctgcac gacatgcagg	2400
aaggcaagtg cctgtacagc ctggaagcca tcctctgga agatctgctg aacaaccct	2460
tcaactatga ggtggaccac atcatcccca gaagcgtgtc cttcgacaac agcttcaaca	2520
acaagtgct cgtgaagcag gaagaaaaca gcaagaaggg caaccggacc ccattccagt	2580
acctgagcag cagcgacagc aagatcagct acgaaacctt caagaagcac atcctgaatc	2640
tggccaaggg caagggcaga atcagcaaga ccaagaaaga gtatctgctg gaagaacggg	2700
acatcaacag gtctccgtg cagaaagact tcatcaaccg gaacctgggtg gataccagat	2760
acgccaccag aggctgatg aacctgctgc ggagctactt cagagtgaac aacctggacg	2820
tgaaagtga gtccatcaat ggcggcttca ccagctttct gcggcggaag tggaaagtta	2880
agaaagagcg gaacaagggg tacaagcacc acgccgagga cgccctgatc attgccaacg	2940
ccgatttcat cttcaaagag tggaagaaac tggacaaggc caaaaaagtg atggaaaacc	3000
agatgttca ggaagcag gccgagagca tgcccagat cgaaccgag caggagtaca	3060
aagagatctt catcacccc caccagatca agcacattaa ggacttcaag gactacaagt	3120
acagccaccg ggtggacaag aagcctaata gagagctgat taacgacacc ctgtactcca	3180
cccggaagga cgacaagggc aacaccctga tcgtgaacaa tctgaacggc ctgtacgaca	3240
aggacaatga caagctgaaa aagctgatca acaagagccc cgaaaagtct ctgatgtacc	3300
accacgaccc ccagacctac cagaaactga agctgattat ggaacagtac ggcgacgaga	3360
agaatcccct gtacaagtac tacgaggaag ccgggaacta cctgaccaag tactccaaaa	3420
aggacaacgg ccccgatgac aagaagatta agtattacgg caacaaactg aacgcccac	3480
tggacatcac cgacgactac cccaacagca gaaacaaggt cgtgaagctg tccctgaagc	3540
cctacagatt cgacgtgtac ctggacaatg gcgtgtacaa gttcgtgacc gtgaagaatc	3600

tggatgtgat caaaaaagaa aactactacg aagtgaatag caagtgtat gaggaagcta	3660
agaagctgaa gaagatcagc aaccaggccg agtttatcgc ctctttctac aacaacgac	3720
tgatcaagat caacggcgag ctgtatagag tgatcggcgt gaacaacgac ctgctgaacc	3780
ggatcgaagt gaacatgac gacatcacct accgcgagta cctggaaaac atgaacgaca	3840
agaggccccc caggatcatt aagacaatcg cctccaagac ccagagcatt aagaagtaca	3900
gcacagacat tctgggcaac ctgtatgaag tgaatctaa gaagcaccct cagatcatca	3960
aaaagggcaa aaggccggcg gccacgaaaa aggccggcca ggcaaaaaag aaaaagggat	4020
cctaccata cgatgttcca gattacgtt acccatacga tgttccagat tacgttacc	4080
catacgtgt tccagattac gcttaagaat tcctagagct cgctgatcag cctcgactgt	4140
gccttctagt tgccagccat ctgttgtttg ccctccccc gtgccttcct tgaccctgga	4200
aggtgccact cccactgtcc ttctctaata aaatgaggaa attgcatcgc attgtctgag	4260
taggtgtcat tctattctgg ggggtggggt ggggcaggac agcaaggggg aggattggga	4320
agagaatagc aggcattgtg gggaggtacc gagggcctat ttcccatgat tccttcatat	4380
ttgcatatac gatacaaggc tgttagagag ataattggaa ttaatttgac tgtaaacaca	4440
aagatattag taaaaatagc gtgacgtaga aagtaataat ttcttgggta gtttgagtt	4500
ttaaaattat gttttaaagt ggactatcat atgcttaccg taacttgaaa gtatttcgat	4560
ttcttggctt tatatatctt gtggaaagga cgaacacccg aacacacagc tgggttatca	4620
gaggttttag tactctggaa acagaatcta ctaaaacaag gcaaaatgcc gtgtttatct	4680
cgtaacttg ttggcgagat ttttgcggcc gcaggaaccc ctatgatgg agttggccac	4740
tcctctctg cgcgctcgt cgctcactga ggccgggcga ccaaaggctc cccgacgccc	4800
gggctttgcc cgggcggcct cagtgagcga gcgagcgcgc agctgcctgc aggggcgcct	4860
gatgcggtat ttctctctta cgcactctgt cggtatttca caccgcatac gtcaaagcaa	4920
ccatagtacg cgcctgtag cggcgcatca agcgcggcgg gtgtggtggt tacgcgcagc	4980
gtgaccgcta cacttgccag cgccttagcg cccgtctctt tcgtttctt ccttctctt	5040
ctgccacgt tcgccgctt tccccgtcaa gctctaaatc gggggctccc tttagggttc	5100
cgatttagtg ctttacggca cctcgacccc aaaaaacttg atttgggtga tggttcacgt	5160
agtgggcat cgcctgata gacggttttt cgccttttga cgttggagtc cacgttcttt	5220
aatagtggac tctgttcca aactggaaca acactcaacc ctatctcggg ctattctttt	5280
gatttataag ggattttgcc gatttcggcc tatttggtta aaaatgagct gatttaacaa	5340
aaatttaacg cgaattttta caaaatatta acgtttacaa ttttatggtg cactctcagt	5400
acaatctgct ctgatgccg atagttaagc cagccccgac acccgccaac acccgctgac	5460

gcgccttgac gggcttgtct gctcccggca tccgcttaca gacaagctgt gaccgtctcc 5520

gggagctgca tgtgtcagag gttttcaccg tcatcaccga aacgcgcgag acgaaagggc 5580

ctcgtgatac gcctatTTTT ataggttaat gtcatgataa taatggtttc ttagacgtca 5640

ggtggcactt ttcggggaaa tgtgcgcgga acccctatTT gtttattttt ctaaatacat 5700

tcaaatatgt atccgctcat gagacaataa ccttgataaa tgcttcaata atattgaaaa 5760

aggaagagta tgagtattca acatttcctg gtcgccctta ttcctttttt tgcggcattt 5820

tgcttctctg tttttgctca cccagaaacg ctggtgaaag taaaagatgc tgaagatcag 5880

ttgggtgcac gagtgggtta catcgaactg gatctcaaca gcggtaaagat ccttgagagt 5940

tttcgccccg aagaacgttt tccaatgatg agcactttta aagtctctgt atgtggcgcg 6000

gtattatccc gtattgacgc cgggcaagag caactcggtc gccgcataca ctattctcag 6060

aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta 6120

agagaattat gcagtgtgc cataacatg agtgataaca ctgcggccaa cttacttctg 6180

acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta 6240

actcgccttg atcgttggga accggagctg aatgaagcca taccaaacga cgagcgtgac 6300

accacgatgc ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt 6360

actctagctt cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca 6420

cttctgcgtc cggcccttcc ggctggctgg tttattgtcg ataaatctgg agccggtgag 6480

cgtggaagcc gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta 6540

gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgtgag 6600

atagggtcct cactgattaa gcatitggtta ctgtcagacc aagtttactc atataactt 6660

tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat cctttttgat 6720

aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta 6780

gaaaagatca aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa 6840

acaaaaaac caccgtacc agcgttgggt tgtttgccgg atcaagagct accaactctt 6900

tttccgaagg taactggctt cagcagagcg cagatacaca atactgtcct tctagtgtag 6960

ccgtagttag gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta 7020

atcctgttac cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca 7080

agacgatagt taccgataa ggcgagcgg tcgggctgaa cgggggggtc gtgcacacag 7140

cccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa 7200

agcgccacgc ttcccgaagg gagaaggcg gacaggtatc cggtaagcgg cagggtcgga 7260

acaggagagc gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtctgtc 7320

gggtttcggc accctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc 7380

ctatggaaaa acgccagcaa cgcggccttt ttacggttcc tggccttttg ctggcctttt 7440

gctcacatgt 7450

<210> 40

<211> 7450

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 40

cctgcaggca gctgcgcgct cgctcgtca ctgaggccgc ccgggcgtcg ggcgaccttt 60

ggtcgcccgg cctcagttag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120

aggggttcct gcggcctcta gactcgaggc gttgacattg attattgact agttattaat 180

agtaatcaat tacgggggtca ttagttcata gcccatatat ggagttccgc gttacataac 240

ttacggtaaa tggcccgctt ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa 300

tgacgtatgt tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt 360

atttacggta aactgccac ttggcagtag atcaagtgtg tcatatgcca agtaccccc 420

ctattgacgt caatgacggt aaatggcccg cctggcatta tgcccagtag atgaccttat 480

gggactttcc tacttggcag tacatctacg tattagtcac cgctattacc atggtgatgc 540

ggttttggca gtacatcaat gggcgtggat agcggtttga ctcacgggga ttccaagtc 600

tccaccccat tgacgtcaat gggagtttgt tttggcacca aaatcaacgg gactttccaa 660

aatgtcgtaa caactccgcc ccattgacgc aaatgggcgg taggcgtgta cgggtggagg 720

tctatataag cagagctctc tggctaacta ccggtgccac catggcccca aagaagaagc 780

ggaaggtcgg tatccacgga gtcccagcag ccaagcggaa ctacatcctg ggcctggaca 840

tcggcatcac cagcgtgggc tacggcatca tcgactacga gacacgggac gtgatcgatg 900

ccggcgtgcg gctgttcaaa gaggccaacg tggaaaacaa cgagggcagg cggagcaaga 960

gaggcgccag aaggctgaag cggcggaggc ggcatagaat ccagagagtg aagaagctgc 1020

tgttcgacta caacctgctg accgaccaca gcgagctgag cggcatcaac ccctacgagg 1080

ccagagtga gggcctgagc cagaagctga gcgaggaaga gttctctgcc gccctgctgc 1140

acctggccaa gagaagaggc gtgcacaacg tgaacgaggt ggaagaggac accggcaacg 1200

agctgtccac caaagagcag atcagccgga acagcaaggc cctggaagag aaatacgtgg 1260

ccgaactgca gctggaacgg ctgaagaaag acggcgaagt gcggggcagc atcaacagat 1320

tcaagaccag cgactacgtg aaagaagcca aacagctgct gaaggtgcag aaggcctacc	1380
accagctgga ccagagcttc atcgacacct acatcgacct gctggaaacc cggcggacct	1440
actatgaggg acctggcgag ggcagcccct tcggctggaa ggacatcaaa gaatggtacg	1500
agatgctgat gggccactgc acctacttcc ccgaggaact gcggagcgtg aagtacgcct	1560
acaacgccga cctgtacaac gccctgaacg acctgaacaa tctcgtgac accagggacg	1620
agaacgagaa gctggaatat tacgagaagt tccagatcat cgagaacgtg ttcaagcaga	1680
agaagaagcc caccctgaag cagatcgcca aagaaatcct cgtgaacgaa gaggatatta	1740
agggctacag agtgaccagc accggcaagc ccgagttcac caacctgaag gtgtaccacg	1800
acatcaagga cattaccgcc cggaaagaga ttattgagaa cgccgagctg ctggatcaga	1860
ttgccaagat cctgaccatc taccagagca gcgaggacat ccaggaagaa ctgaccaatc	1920
tgaactccga gctgaccag gaagagatcg agcagatctc taatctgaag ggctataccg	1980
gcaccacaa cctgagcctg aaggccatca acctgatcct ggacgagctg tggcacacca	2040
acgacaacca gatcgtatc ttcaaccggc tgaagctggt gcccaagaag gtggacctgt	2100
cccagcagaa agagatcccc accaccctgg tggacgactt catcctgagc cccgtcgtga	2160
agagaagctt catccagagc atcaaagtga tcaacgcat catcaagaag tacggcctgc	2220
ccaacgacat cattatcgag ctggcccgcg agaagaactc caaggacgcc cagaaaatga	2280
tcaacgagat gcagaagcgg aaccggcaga ccaacgagcg gatcgaggaa atcatccgga	2340
ccaccggcaa agagaacgcc aagtacctga tcgagaagat caagctgcac gacatgcagg	2400
aaggcaagtg cctgtacagc ctggaagcca tccctctgga agatctgctg aacaaccct	2460
tcaactatga ggtggaccac atcatccca gaagcgtgtc cttcgacaac agcttcaaca	2520
acaagtgct cgtgaagcag gaagaaaaca gcaagaaggg caaccggacc ccattccagt	2580
acctgagcag cagcgacagc aagatcagct acgaaacctt caagaagcac atcctgaatc	2640
tggccaaggg caagggcaga atcagcaaga ccaagaaaga gtatctgctg gaagaacggg	2700
acatcaacag gttctccgtg cagaaagact tcatcaaccg gaacctggtg gataccagat	2760
acgccaccag aggcctgatg aacctgctgc ggagctactt cagagtgaac aacctggacg	2820
tgaaagtga gtccatcaat ggcggttca ccagctttct gcggcggaag tggaagtta	2880
agaaagagcg gaacaagggg tacaagcacc acgccagga cgccctgac attgccaacg	2940
ccgatttcat cticaaagag tggaagaaac tggacaaggc caaaaaagt atggaaaacc	3000
agatgttcga ggaaaagcag gccgagagca tgcccagat cgaaaccgag caggagtaca	3060
aagagatctt catcaccccc caccagatca agcacattaa ggacttcaag gactacaagt	3120

acagccaccg ggtggacaag aagcctaata gagagctgat taacgacacc ctgtactcca 3180

cccgaagga cgacaagggc aacaccctga tctgaacaa tctgaacggc ctgtacgaca 3240

aggacaatga caagctgaaa aagctgatca acaagagccc cgaaaagctg ctgatgtacc 3300

accacgaccc ccagacctac cagaaactga agctgattat ggaacagtac ggcgacgaga 3360

agaatcccct gtacaagtac tacgaggaaa ccgggaacta cctgaccaag tactccaaaa 3420

aggacaacgg ccccgatgac aagaagatta agtattacgg caacaaactg aacgcccac 3480

tggacatcac cgacgactac cccaacagca gaaacaaggt cgtgaagctg tcctgaagc 3540

cctacagatt cgacgtgtac ctggacaatg gcgtgtacaa gttcgtgacc gtgaagaatc 3600

tggatgtgat caaaaaagaa aactactacg aagtgaatag caagtgctat gaggaagcta 3660

agaagctgaa gaagatcagc aaccaggccg agtttatcgc ctctttctac aacaacgac 3720

tgatcaagat caacggcgag ctgtatagag tgatcggcgt gaacaacgac ctgctgaacc 3780

ggatcgaagt gaacatgac gacatcacct accgcgagta cctggaaaac atgaacgaca 3840

agaggcccc caggatcatt aagacaatcg cctccaagac ccagagcatt aagaagtaca 3900

gcacagacat tctgggcaac ctgtatgaag tgaatctaa gaagcaccct cagatcatca 3960

aaaagggcaa aaggccggcg gccacgaaa aggccggcca ggcaaaaaag aaaaagggat 4020

cctaccata cgatgttcca gattacgctt acccatacga tgttcagat tacgcttacc 4080

catacgtatg tccagattac gcttaagaat tcttagagct cgctgatcag cctcgactgt 4140

gccttctagt tgccagccat ctgttgtttg ccctccccc gtgccttcct tgaccctgga 4200

aggtgccact cccactgtcc ttctctaata aaatgaggaa attgcatcgc attgtctgag 4260

taggtgtcat tctattctgg ggggtggggt ggggcaggac agcaaggggg aggattggga 4320

agagaatagc aggcattgct gggaggtacc gagggcctat ttcccatgat tccttcatat 4380

ttgcatatac gatacaagcg tgtagagag ataattggaa ttaatttgac tgtaaacaca 4440

aagatattag taaaaatac gtgacgtaga aagtaataat ttcttgggta gtttgagtt 4500

ttaaaattat gttttaaaat ggactatcat atgcttaccg taacttgaaa gtatttcgat 4560

ttcttggctt tatatatctt gtggaaagga cgaaacaccg agaactggtg ggaaatggtc 4620

taggttttag tactctggaa acagaatcta ctaaaacaag gcaaatgcc gtgtttatct 4680

cgtaacttg ttggcgagat ttttcgggcc gcaggaaccc ctatgatgg agttggccac 4740

tcctctctg cgcgctcgt cgctcactga ggccgggcga ccaaggtcg cccgacgccc 4800

gggctttgcc cgggcggcct cagtgagcga gcgagcgcgc agctgcctgc aggggcgcct 4860

gatgcggtat tttctcctta cgcatctgtg cggatattca caccgcatac gtcaaagcaa	4920
ccatagtagc cgccctgtag cggcgcatta agcgcggcgg gtgtggtggt tacgcgcagc	4980
gtgaccgcta cacttgccag cgccctagcg cccgctcctt tcgctttctt cctttccttt	5040
ctcgccacgt tcgcccgttt tccccgtcaa gctctaaatc gggggctccc tttagggttc	5100
cgatttagtg ctttacggca cctcgacccc aaaaaacttg atttgggtga tggttcacgt	5160
agtgggcat cgccctgata gacggttttt cgcccttga cgttggagtc cacgttcttt	5220
aatagtggac tcttgttcca aactggaaca aactcaacc ctatctcggg ctattctttt	5280
gatttataag ggattttgcc gatttcggcc tatitggttaa aaaatgagct gatttaacaa	5340
aaatttaacg cgaattttta caaaatattt acgtttacaa ttttatgggtg cactctcagt	5400
acaatctgct ctgatgccgc atagttaagc cagccccgac acccgccaac acccgctgac	5460
gcgcctgac gggtttgtct gctcccgga tccgcttaca gacaagctgt gaccgtctcc	5520
gggagctgca tgtgtcagag gttttcaccg tcatcaccga aacgcgcgag acgaaagggc	5580
ctcgtgatac gcctattttt ataggttaat gtcatgataa taatggtttc ttagacgtca	5640
ggtggcactt ttcggggaaa tgtgcgcgga acccctatit gtttattttt ctaaatacat	5700
tcaaatatgt atccgctcat gagacaataa ccttgataaa tgcttcaata atattgaaaa	5760
aggaagagta ttagtattca acatttcctg gtgcacctta ttcctttttt tgcggcattt	5820
tgctttcctg tttttgtca cccagaaacg ctggtgaaag taaaagatgc tgaagatcag	5880
ttgggtgcac gagtgggtta catcgaactg gatctcaaca gcggtlaagat ccttgagagt	5940
tttcgccccg aagaacgttt tccaatgatg agcactttta aagtctctgt atgtggcgcg	6000
gtattatccc gtattgacgc cgggcaagag caactcggtc gccgcataca ctattctcag	6060
aatgacttgg ttgagtactc accagtcaca gaaaagcatc ttacggatgg catgacagta	6120
agagaattat gcagtgtgc cataacatg agtgataaca ctgcggccaa ctacttctg	6180
acaacgatcg gaggaccgaa ggagctaacc gcttttttgc acaacatggg ggatcatgta	6240
actcgcttg atcgttggga accggagctg aatgaagcca taccaaagca cgagcgtgac	6300
accacgatgc ctgtagcaat ggcaacaacg ttgcgcaaac tattaactgg cgaactactt	6360
actctagctt cccggcaaca attaatagac tggatggagg cggataaagt tgcaggacca	6420
cttctgcgt cgcccttcc ggctggctgg tttattgtg ataaatctgg agccggtgag	6480
cgtggaagcc gcggtatcat tgcagcactg gggccagatg gtaagccctc ccgtatcgta	6540
gttatctaca cgacggggag tcaggcaact atggatgaac gaaatagaca gatcgtgag	6600
ataggtagct cactgattaa gcattggtaa ctgtcagacc aagtttactc atatatactt	6660
tagattgatt taaaacttca tttttaattt aaaaggatct aggtgaagat cttttttgat	6720

aatctcatga ccaaaatccc ttaacgtgag ttttcgttcc actgagcgtc agaccccgta 6780
 gaaaagatca aaggatcttc ttgagatcct ttttttctgc gcgtaatctg ctgcttgcaa 6840
 acaaaaaaac caccgctacc agcgggtggtt tgtttgccgg atcaagagct accaactctt 6900
 tttccgaagg taactggctt cagcagagcg cagataccaa atactgtcct tctagttag 6960

ccgtagttag gccaccactt caagaactct gtagcaccgc ctacatacct cgctctgcta 7020
 atcctgttac cagtggctgc tgccagtggc gataagtcgt gtcttaccgg gttggactca 7080
 agacgatagt taccggataa ggcgacgcgg tcgggctgaa cgggggggttc gtgcacacag 7140
 cccagcttgg agcgaacgac ctacaccgaa ctgagatacc tacagcgtga gctatgagaa 7200
 agcgccacgc ttcccgaagg gagaaaggcg gacaggatc cggttaagcgg cagggtcgga 7260
 acaggagagc gcacgaggga gcttccaggg ggaaacgcct ggtatcttta tagtcctgtc 7320
 gggtttcgcc acctctgact tgagcgtcga tttttgtgat gctcgtcagg ggggcggagc 7380

ctatggaaaa acgccagcaa cgcgcccttt ttacgggtcc tggccttttg ctggcctttt 7440
 gctcacatgt 7450

<210> 41

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 41

aacaaatata ccttagtata 20

<210> 42

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 42

aatgtatttc ttctattcaa 20

<

210> 43

<211> 3255

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 43

atggcccaaa agaagaagcg gaaggtcggg atccacggag tcccagcagc caagcggaac	60
tacatcctgg gcctggacat cggcatcacc agcgtgggct acggcatcat cgactacgag	120
acacgggacg tgatcgatgc cggcgtgcgg ctgttcaaag aggccaacgt ggaaaacaac	180
gagggcaggc ggagcaagag aggcgccaga aggctgaagc ggcgaggcg gcatagaatc	240
cagagagtga agaagctgct gttcgactac aacctgctga ccgaccacag cgagctgagc	300
ggcatcaacc cctacgagcg cagagtgaag ggctgagcc agaagctgag cgaggaagag	360
ttctctgccg cctgctgca cctggccaag agaagaggcg tgcacaacgt gaacgaggtg	420
gaagaggaca cggcaacga gctgtccacc aaagagcaga tcagccggaa cagcaaggcc	480
ctggaagaga aatcgtggc cgaactgcag ctggaacggc tgaagaaaga cggcgaagtg	540
cggggcagca tcaacagatt caagaccagc gactacgtga aagaagccaa acagctgctg	600
aaggtgcaga aggcctacca ccagctggac cagagcttca tcgacaccta catcgacctg	660
ctggaaacc ggcgaccta ctatgaggga cctggcgagg gcagcccctt cggctggaag	720
gacatcaaag aatggtacga gatgctgatg ggccactgca cctacttccc cgaggaactg	780
cggagcgtga agtacgccta caacccgac ctgtacaacg cctgaacga cctgaacaat	840
ctcgtgatca ccagggacga gaacgagaag ctggaatatt acgagaagtt ccagatcatc	900
gagaactgtg tcaagcaga gaagaagccc accctgaagc agatcgccaa agaaatcctc	960
gtgaacgaag aggatattaa gggctacaga gtgaccagca ccggcaagcc cgagttcacc	1020
aacctgaagg tgtaccagca catcaaggac attaccgccc ggaaagagat tattgagaac	1080
gccgagctgc tggatcagat tgccaagatc ctgaccatct accagagcag cgaggacatc	1140
caggaagaac tgaccaatct gaactccgag ctgaccagg aagagatcga gcagatctct	1200
aatctgaagg gctataccgg caccacaac ctgagcctga aggccatcaa cctgatcctg	1260
gacgagctgt ggacaccaa cgacaaccag atcgtatct tcaaccggct gaagctgggtg	1320
ccaagaagg tggacctgtc ccagcagaaa gagatcccca ccaccctggg ggacgacttc	1380
atcctgagcc ccgtcgtgaa gagaagcttc atccagagca tcaaagtgat caacgccatc	1440
atcaagaagt acggcctgcc caacgacatc attatcgagc tggcccgca gaagaactcc	1500
aaggacgccc agaaaatgat caacgagatg cagaagcgga accggcagac caacgagcgg	1560
atcgaggaaa tcatccggac caccggcaaa gagaacgcca agtacctgat cgagaagatc	1620
aagctgcacg acatgcagga aggcaagtgc ctgtacagcc tggaagccat ccctctggaa	1680
gatctgctga acaaccctt caactatgag gtggaccaca tcatccccag aagcgtgtcc	1740
ttcgacaaca gcttcaaca caagtgctc gtgaagcagg aagaaaacag caagaagggc	1800

aaccggaccc cattccagta cctgagcagc agcgacagca agatcagcta cgaaaccttc	1860
aagaagcaca tcctgaatct ggccaagggc aagggcagaa tcagcaagac caagaaagag	1920
tatctgctgg aagaacggga catcaacagg ttctccgtgc agaaagactt catcaaccgg	1980
aacctggtgg ataccagata cgccaccaga ggcctgatga acctgctgcg gagctacttc	2040
agagtgaaca acctggacgt gaaagtgaag tccatcaatg gcggcttcac cagctttctg	2100
cggcggaagt ggaagtttaa gaaagagcgg aacaaggggt acaagcacca cgccgaggac	2160
gccctgatca ttgccaacgc cgatttcac ttcaaagagt ggaagaaact ggacaaggcc	2220
aaaaaagtga tggaaaacca gatgttcgag gaaaagcagg ccgagagcat gcccgagatc	2280
gaaaccgagc aggagtacaa agagatcttc atcaccccc accagatcaa gcacattaag	2340
gacttcaagg actacaagta cagccaccgg gtggacaaga agcctaatag agagctgatt	2400
aacgacaccc tgtactccac ccggaaggac gacaagggca acacctgat cgtgaacaat	2460
ctgaacggcc tgiacgacaa ggacaatgac aagctgaaaa agctgatcaa caagagcccc	2520
gaaaagctgc tgatgtacca ccacgacccc cagacctacc agaaactgaa gctgattatg	2580
gaacagtacg gcgacgagaa gaatccccctg tacaagtact acgaggaaac cgggaactac	2640
ctgaccaagt actccaaaa ggacaacggc cccgtgatca agaagattaa gtattacggc	2700
aacaaactga acgcccactt ggacatcacc gacgactacc ccaacagcag aaacaaggtc	2760
gtgaagctgt cctgaagcc ctacagattc gacgtgtacc tggacaatgg cgtgtacaag	2820
ttcgtgaccg tgaagaatct ggatgtgatc aaaaaagaaa actactacga agtgaatagc	2880
aagtgtatg aggaagctaa gaagctgaag aagatcagca accaggccga gtttatcgcc	2940
tccttctaca acaacgatct gatcaagatc aacggcgagc tgtatagagt gatcggcgtg	3000
aacaacgacc tgctgaaccg gatcgaagtg aacatgatcg acatcaccta ccgcgagtac	3060
ctggaaaaca tgaacgacaa gagggccccc aggatcatta agacaatcgc ctccaagacc	3120
cagagcatta agaagtacag cacagacatt ctgggcaacc tgtatgaagt gaaatctaag	3180
aagcacctc agatcatcaa aaagggcaaa aggccggcgg ccacgaaaaa ggccggccag	3240
gcaaaaaaga aaaag	3255

<210> 44

<211> 3156

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 44

aagcggaaact acatcctggg cctggacatc ggcatcacca gcgtgggcta cggcatcatc	60
gactacgaga cacgggagct gatcgatgcc ggctgctggc tgttcaaaga ggccaacgtg	120
gaaaacaacg agggcaggcg gagcaagaga ggccgagaa ggctgaagcg gcggaggcgg	180
catagaatcc agagagttaa gaagctgctg ttgactaca acctgctgac cgaccacagc	240
gagctgagcg gcatcaaccc ctacgaggcc agagtgaagg gcctgagcca gaagctgagc	300
gaggaagagt tctctgccgc cctgctgcac ctggccaaga gaagaggcgt gcacaacgtg	360
aacgaggtgg aagaggacac cggcaacgag ctgtccacca aagagcagat cagccggaac	420
agcaaggccc tggaagagaa atacgtggcc gaactgcagc tggaacggct gaagaaagac	480
ggcgaagtgc ggggcagcat caacagattc aagaccagcg actacgtgaa agaagccaaa	540
cagctgctga aggtgcagaa ggcctaccac cagctggacc agagcttcat cgacacctac	600
atcgacctgc tggaacccc gcggacctac tatgaggagc ctggcgaggg cagccccctc	660
ggctggaagg acatcaaaga atggtacgag atgctgatgg gccactgcac ctacttcccc	720
gaggaactgc ggagcgtgaa gtacgcctac aacgccgacc tgtacaacgc cctgaacgac	780
ctgaacaatc tcgtgatcac caggagcagc aacgagaagc tggaatatta cgagaagttc	840
cagatcatcg agaactgttt caagcagaag aagaagccca cctgaagca gatcgccaaa	900
gaaatcctcg tgaacgaaga ggatattaag ggctacagag tgaccagcac cggcaagccc	960
gagttcacca acctgaaggt gtaccagcac atcaaggaca ttaccgcccg gaaagagatt	1020
attgagaacg ccgagctgct ggatcagatt gccaaagatc tgaccatcta ccagagcagc	1080
gaggacatcc aggaagaact gaccaatctg aactccgagc tgaccagga agagatcgag	1140
cagatctcta atctgaaggg ctataccggc acccacaacc tgagcctgaa ggccatcaac	1200
ctgatcctgg acgagctgtg gcacaccaac gacaaccaga tcgtatctt caaccggctg	1260
aagctggtgc ccaagaaggt ggacctgtcc cagcagaaag agatccccac caccctggtg	1320
gacgacttca tctgagccc cgtcgtgaag agaagcttca tccagagcat caaagtgatc	1380
aacgccatca tcaagaagta cggcctgccc aacgacatca ttatcgagct ggcccccgag	1440
aagaactcca aggacgccc gaaaatgatc aacgagatgc agaagcggaa ccggcagacc	1500
aacgagcggg tcgaggaaat catccggacc accggcaaag agaacgcaa gtacctgatc	1560
gagaagatca agctgcacga catgcaggaa ggcaagtgcc tgtacagcct ggaagccatc	1620
cctctggaag atctgctgaa caacccttc aactatgagg tggaccacat catcccaga	1680
agcgtgtcct tcgacaacag cttcaacaac aaggtgctcg tgaagcagga agaaaacagc	1740
aagaagggca accggacccc attccagtac ctgagcagca gcgacagcaa gatcagctac	1800
gaaaccttca agaagcacat cctgaatctg gccaaaggca agggcagaat cagcaagacc	1860

aagaaagagt atctgctgga agaacgggac atcaacaggt tctccgtgca gaaagacttc 1920
atcaaccgga acctggtgga taccagatac gccaccagag gcctgatgaa cctgctgcgg 1980
agctacttca gagtgaacaa cctggacgtg aaagtgaagt ccatcaatgg cggttcacc 2040

agctttctgc ggcggaagtg gaagttaaag aaagagcgga acaaggggta caagcaccac 2100
gccgaggacg cctgatcat tgccaacgcc gatttcattc tcaaagagtg gaagaaactg 2160
gacaaggcca aaaaagtgat ggaaaaccag atgttcgagg aaaagcaggc cgagagcatg 2220
cccagatcg aaaccgagca ggagtacaaa gagatcttca tcaccccca ccagatcaag 2280
cacattaagg acttcaagga ctacaagtac agccaccggg tggacaagaa gcctaataga 2340
gagctgatta acgacacct gtactccacc cggaaggacg acaagggcaa caccctgatc 2400
gtgaacaatc tgaacggcct gtacgacaag gacaatgaca agctgaaaaa gctgatcaac 2460

aagagcccg aaaaagtgtg gatgtaccac cagaccccc agacctacca gaaactgaag 2520
ctgattatgg aacagtacgg cgacgagaag aatcccctgt acaagtacta cgaggaaacc 2580
gggaactacc tgaccaagta ctcaaaaag gacaacggcc ccgtgatcaa gaagattaag 2640
tattacggca aaaaactgaa cgcccatctg gacatcacgg acgactacce caacagcaga 2700
aacaaggctg tgaagctgtc cctgaagccc tacagattcg acgtgtacct ggacaatggc 2760
gtgtacaagt tcgtgaccgt gaagaatctg gatgtgatca aaaaagaaaa ctactacgaa 2820
gtgaatagca agtgctatga ggaagctaag aagctgaaga agatcagcaa ccaggccgag 2880

tttatgcct cttctacaa caacgatctg atcaagatca acggcgagct gtatagagtg 2940
atcgcggtga acaacgacct gctgaaccgg atcgaagtga acatgatcga catcacctac 3000
cgcgagtacc tgaaaaacat gaacgacaag agggcccca ggatcattaa gacaatcgcc 3060
tccaagacc agagcattaa gaagtacagc acagacattc tgggcaacct gtatgaagtg 3120
aaatctaaga agcacctca gatcatcaaa aagggc 3156

<210> 45

<211> 1052

<212> PRT

<213> Artificial sequence

<220><223> Synthetic

<400> 45

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

1 5 10 15

Tyr Gly Ile Ile Asp Tyr Glu Thr Arg Asp Val Ile Asp Ala Gly Val

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

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

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

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

 260 265 270

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

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

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

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

 705 710 715 720
 Asp Lys Ala Lys Lys Val Met Glu Asn Gln Met Phe Glu Glu Lys Gln
 725 730 735
 Ala Glu Ser Met Pro Glu Ile Glu Thr Glu Gln Glu Tyr Lys Glu Ile
 740 745 750
 Phe Ile Thr Pro His Gln Ile Lys His Ile Lys Asp Phe Lys Asp Tyr
 755 760 765

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

770

775

780

Asp Thr Leu Tyr Ser Thr Arg Lys Asp Asp Lys Gly Asn Thr Leu Ile

785

790

795

800

Val Asn Asn Leu Asn Gly Leu Tyr Asp Lys Asp Asn Asp Lys Leu Lys

805

810

815

Lys Leu Ile Asn Lys Ser Pro Glu Lys Leu Leu Met Tyr His His Asp

820

825

830

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

835

840

845

Glu Lys Asn Pro Leu Tyr Lys Tyr Tyr Glu Glu Thr Gly Asn Tyr Leu

850

855

860

Thr Lys Tyr Ser Lys Lys Asp Asn Gly Pro Val Ile Lys Lys Ile Lys

865

870

875

880

Tyr Tyr Gly Asn Lys Leu Asn Ala His Leu Asp Ile Thr Asp Asp Tyr

885

890

895

Pro Asn Ser Arg Asn Lys Val Val Lys Leu Ser Leu Lys Pro Tyr Arg

900

905

910

Phe Asp Val Tyr Leu Asp Asn Gly Val Tyr Lys Phe Val Thr Val Lys

915

920

925

Asn Leu Asp Val Ile Lys Lys Glu Asn Tyr Tyr Glu Val Asn Ser Lys

930

935

940

Cys Tyr Glu Glu Ala Lys Lys Leu Lys Lys Ile Ser Asn Gln Ala Glu

945

950

955

960

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

965

970

975

Leu Tyr Arg Val Ile Gly Val Asn Asn Asp Leu Leu Asn Arg Ile Glu

980

985

990

Val Asn Met Ile Asp Ile Thr Tyr Arg Glu Tyr Leu Glu Asn Met Asn

995

1000

1005

Asp Lys Arg Pro Pro Arg Ile Ile Lys Thr Ile Ala Ser Lys Thr

1010	1015	1020
Gln Ser Ile Lys Lys Tyr Ser Thr Asp Ile Leu Gly Asn Leu Tyr		
1025	1030	1035
Glu Val Lys Ser Lys Lys His Pro Gln Ile Ile Lys Lys Gly		
1040	1045	1050
<210> 46		
<211> 52		
<212> DNA		
<213> Artificial sequence		
<220><223> Synthetic		
<400> 46		
aacgctgaag aaccctgata atagaagaaa tacattttta aatcaattca gg	52	
<210> 47		
<211> 51		
<212> DNA		
<213> Artificial sequence		
<220><223> Synthetic		
<400> 47		
aagaaccctg atattatctt agtagattaa tagaagaaat acatttttaa a	51	
<210> 48		
<211> 42		
<212> DNA		
<213> Artificial sequence		
<220><223> Synthetic		
<400> 48		
gctgaagaac cctgaaaaaa tacatttttt tatcaattca gg	42	
<210> 49		
<211> 36		
<212> DNA		
<213> Artificial sequence		
<220><223> Synthetic		
<400> 49		
gaatgatttt cttgtagaag aaataacaat taaatc	36	

<210> 50
 <211> 52
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 50
 aacgctgaag aaccctgata atagaagaaa tacatTTTTa aatcaattca gg 52

<210> 51
 <211> 59
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 51
 aacgctgaag aaccctgata ttatcttagt agattaatag aagaaataca tttttaaat 59

<210> 52
 <211> 59
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 52
 aacgctgaag aaccctgata ttatcttagt agattaatag aagaaataca tttttaaat 59

<210> 53
 <211> 6
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 53
 aagagt 6

<210> 54
 <211> 6
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic

<400> 54	
gggaat	6
<210> 55	
<211> 3	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<400> 55	
agg	3
<210> 56	
<211> 3	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<400> 56	
tgg	3
<210> 57	
<211> 6	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<400> 57	
gagagt	6
<210> 58	
<211> 6	
<212> DNA	
<213> Artificial sequence	
<220><223> Synthetic	
<400> 58	
gggagt	6
<210> 59	
<211> 6	
<212> DNA	
<213> Artificial sequence	

<220><223> Synthetic

<400> 59

ttgaat

6

<210> 60

<211> 6

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 60

cagagt

6

<210> 61

<211> 6

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 61

tggagt

6

<210> 62

<211> 6

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 62

atgagt

6

<210> 63

<211> 6

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 63

aggaat

6

<210> 64

<211> 22

<212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 64
 aaagatatat aatgtcatga at 22
 <210> 65
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 65
 gcagaatcaa atataatagt ct 22
 <210> 66
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 66
 caattaaatt tgacttattg tt 22
 <210> 67
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 67
 ctagaccatt tcccaccagt tc 22
 <210> 68
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 68
 aggactttta ttaccaaag ga 22

<210> 69
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 69
 atccaagtcc atttgattcc ta 22
 <210> 70
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 70
 taattctttc tagaaagagc ct 22
 <210> 71
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 71
 ggacatgtgc aagatgcaag ag 22

 <210> 72
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 72
 tgtatgtaga agacctctaa gt 22
 <210> 73
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 73

tccccctcacc actcacctct ga 22
 <210> 74
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 74
 ctctgataac ccagctgtgt gt 22

 <210> 75
 <211> 19
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 75
 ctctgataac ccagctgtg 19
 <210> 76
 <211> 20
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 76
 ctctgataac ccagctgtgt 20
 <210> 77
 <211> 21
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 77
 ctctgataac ccagctgtgt g 21

 <210> 78
 <211> 23
 <212> DNA
 <213> Artificial sequence

<220><223> Synthetic

<400> 78

ctctgataac ccagctgtgt gtt 23

<210> 79

<211> 19

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 79

ctagaccatt tcccaccag 19

<210> 80

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 80

ctagaccatt tcccaccagt 20

<210> 81

<211> 21

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 81

ctagaccatt tcccaccagt t 21

<210> 82

<211> 23

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 82

ctagaccatt tcccaccagt tct 23

<210> 83

<211> 7009

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 83

ctaaattgta agcggttaata ttttgtaaata attcgcggtta aatttttgtt aaatcagctc	60
attttttaac caataggccg aaatcggtcaa aatcccttat aaatcaaaag aatagaccga	120
gatagggttg agtgttgttc cagtttggaa caagagtcca ctattaaaga acgtggactc	180
caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc	240
ctaatacaagt tttttgggtt cgaggtgccc taaagcacta aatcggaacc ctaaaggag	300
cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa	360
agcgaaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgtgc gcgtaaccac	420
cacacccgcc gcgcttaatg cgccgtaca gggcgcgctc cattcgccat tcaggtcgcg	480
caactgttgg gaagggcgat cgggtcgggc ctcttcgcta ttacgccagc tggcgaaagg	540
gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttcccagt cagcagttg	600
taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtacctt	660
taattctagt actatgcatg cgttgacatt gattattgac tagttattaa tagtaatcaa	720
ttacggggtc attagttcat agcccatata tggagtccg cgttacataa cttacggtaa	780
atggcccgcc tggtcgaccg cccaacgacc ccgcgccatt gacgtcaata atgacgtatg	840
ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt	900
aaactgcccc cttggcagta catcaagtgt atcatatgcc aagtacgcc cctattgacg	960
tcaatgacgg taaatggccc gcctggcatt atgccagta catgacctta tgggactttc	1020
ctacttggca gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc	1080
agtacatcaa tgggcgtgga tagcggtttg actcacgggg atttccaagt ctccacccca	1140
ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgt	1200
acaactccgc ccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa	1260
gcagagctct ctggctaact accggtgcc ccatgaaaag gaactacatt ctggggctgg	1320
acatcgggat tacaagcgtg gggatatgga ttattgacta tgaacaagg gacgtgatcg	1380
acgcaggcgt cagactgttc aaggaggcca acgtggaaaa caatgaggga cggagaagca	1440
agaggggagc caggcgcttg aaacgacgga gaaggcacag aatccagagg gtgaagaaac	1500
tgctgttcga ttacaacctg ctgaccgacc attctgagct gagtgggaatt aatccttatg	1560
aagccagggt gaaaggcctg agtcagaagc tgcagagga agagttttcc gcagctctgc	1620
tgcacctggc taagcgccga ggagtgcata acgtcaatga ggtggaagag gacaccggca	1680

acgagctgtc tacaaaggaa cagatctcac gcaatagcaa agctctggaa gagaagtatg 1740

tcgcagagct gcagctggaa cggctgaaga aagatggcga ggtgagaggg tcaattaata 1800

ggttcaagac aagcgactac gtcaaagaag ccaagcagct gctgaaagtg cagaaggctt 1860

accaccagct ggatcagagc ttcatcgata cttatatcga cctgctggag actcggagaa 1920

cctactatga gggaccagga gaaggagacc ctttcggatg gaaagacatc aaggaatggt 1980

acgagatgct gatgggacat tgcacctatt ttccagaaga gctgagaagc gtcaagtacg 2040

cttataacgc agatctgtac aacgcctga atgacctgaa caacctggtc atcaccaggg 2100

atgaaaacga gaaactggaa tactatgaga agttccagat catcgaaaac gtgtttaagc 2160

agaagaaaaa gcctacactg aaacagattg ctaaggagat cctggccaac gaagaggaca 2220

tcaagggcta ccgggtgaca agcactggaa aaccagagtt caccaatctg aaagtgtatc 2280

acgatattaa ggacatcaca gcacggaaaag aaatcattga gaacgccgaa ctgctggatc 2340

agattgctaa gatcctgact atctaccaga gctccgagga catccaggaa gagctgacta 2400

acctgaacag cgagctgacc caggaagaga tcgaacagat tagtaatctg aagggttaca 2460

ccggaacaca caacctgtcc ctgaaagcta tcaatctgat tctggatgag ctgtggcata 2520

caaacgacaa tcagattgca atctttaacc ggctgaagct ggtcccaaaa aaggtggacc 2580

tgagtacga gaaagagatc ccaaccacac tgggtggacga tttcattctg tcaccctgg 2640

tcaagcggag ctcatccag agcatcaaag tgatcaacgc catcatcaag aagtacggcc 2700

tgcccaatga tatcattatc gagctggcta gggagaagaa cagcaaggac gcacagaaga 2760

tgatcaatga gatgcagaaa cgaaaccggc agaccaatga acgcattgaa gagattatcc 2820

gaactaccgg gaaagagaac gcaaagtacc tgattgaaaa aatcaagctg cacgatatgc 2880

aggagggaat gtgtctgtat tctctggagg ccattcccct ggaggacctg ctgaacaatc 2940

cattcaacta cgaggtcgat catattatcc ccagaagcgt gtctctcgac aattccttta 3000

acaacaaggt gctggccaag caggaagaga actctaaaaa gggcaatagg actcctttcc 3060

agtacctgtc tagttcagat tccaagatct cttacgaaac ctttaaaaag cacattctga 3120

atctggccaa aggaaagggc cgcatcagca agaccaaaaa ggaglacctg ctggaagagc 3180

gggacatcaa cagattctcc gtccagaagg attttattaa ccggaatctg gtggacacaa 3240

gatacgctac tcgggcctg atgaatctgc tgcgatccta tttccgggtg aacaatctgg 3300

atgtgaaagt caagtccatc aacggcgggt tcacatcttt tctgaggcgc aaatggaagt 3360

ttaaaaagga gcgcaacaaa gggtaacaag accatgccga agatgctctg attatcgcaa 3420

atgccgactt catctttaag gagtggaaaa agctggacaa agccaagaaa gtgatggaga 3480

accagatgtt cgaagagaag caggccgaat ctatgcccga aatcgagaca gaacaggagt 3540

acaaggagat tttcatcact cctcaccaga tcaagcatat caaggatttc aaggactaca 3600

agtactctca ccgggtggat aaaaagccca acagagagct gatcaatgac accctgtata 3660

gtacaagaaa agacgataag gggaataccc tgattgtgaa caatctgaac ggactgtacg 3720

acaaagataa tgacaagctg aaaaagctga tcaacaaaag tcccgagaag ctgctgatgt 3780

accacatga tctcagaca tatcagaaac tgaagctgat tatggagcag tacggcgacg 3840

agaagaaccc actgtataag tactatgaag agactgggaa ctacctgacc aagtatagca 3900

aaaaggataa tggccccgtg atcaagaaga tcaagtacta tgggaacaag ctgaatgccc 3960

atctggacat cacagacgat taccctaaca gtcgcaacaa ggtggccaag ctgtcactga 4020

agccatacag attcgatgtc tatctggaca acggcgtgta taaatttttg actgtcaaga 4080

atctggatgt catcaaaaag gagaactact atgaagtga tagcaagtgc tacgaagagg 4140

ctaaaaagct gaaaaagatt agcaaccagg cagagttcat cgcctccttt tacaacaacg 4200

acctgattaa gatcaatggc gaactgtata gggtcatcgg ggtgaacaat gatctgctga 4260

accgcattga agtgaatatg attgacatca cttaccgaga gtatctggaa aacatgaatg 4320

ataagcgccc cctcgaatt atcaaaacaa ttgcctctaa gactcagagt atcaaaaagt 4380

actcaaccga cattctggga aacctgtatg aggtgaagag caaaaagcac cctcagatta 4440

tcaaaaaggg cagcggaggc aagcgtcctg ctgctactaa gaaagctggt caagctaaga 4500

aaaagaaagg atcctaccca tacgatgttc cagattacgc ttaagaattc ctagagctcg 4560

ctgatcagcc tcgactgtgc cttctagttag ccagccatct gttgtttgcc cctccccgt 4620

gccttccttg acctggaag gtgccactcc cactgtcctt tcctaataaa atgaggaaat 4680

tgcatcgcat tgtctgagta ggtgtcattc tattctgggg ggtggggtgg ggcaggacag 4740

caagggggag gattgggaag agaatagcag gcatgctggg gaggtagcgg ccgcccgcgg 4800

tggagctcca gcttttgttc cctttagtga gggtttaattg cgcgcttgge gtaatcatgg 4860

tcatagctgt ttctgtgtg aaattgttat ccgctcaca ttecacacaa catacgagcc 4920

ggaagcataa agtgtaaagc ctgggggtgcc taatgagtga gctaactcac attaatgctg 4980

ttgcgtcac tgccgcttt ccagtcggga aacctgtcgt gccagctgca ttaatgaatc 5040

ggccaacgcg cggggagagg cggtttgcgt attgggcgct cttccgcttc ctgcctcact 5100

gactcgtgc gctcggctgt tcggctgcgg cgagcggtat cagctcactc aaaggcggtg 5160

atacggttat ccacagaatc aggggataac gcaggaaaga acatgtgagc aaaaggccag 5220

caaaaggcca ggaaccgtaa aaaggccgcg ttgctggcgt ttttccatag gctccgcccc 5280

cctgacgagc atcacaaaaa tcgacgctca agtcagaggt ggcgaaaccc gacaggacta 5340
 taaagatacc aggcgtttcc ccctggaagc tcctctgtgc gctctcctgt tccgaccctg 5400
 ccgcttaccg gatacctgtc cgcttttctc ccttcgggaa gcgtggcgct ttctcatagc 5460
 tcacgctgta ggatctctcag ttccgtgtag gtcgttcgct ccaagctggg ctgtgtgcac 5520

gaaccccccg ttcagcccca ccgtgcgcc ttatccgta actatcgtct tgagtccaac 5580
 ccgtaagac acgacttacc gccactggca gcagccactg gtaacaggat tagcagagcg 5640
 aggtatgtag gcggtgctac agagtctctg aagtgggtggc ctaactacgg ctacactaga 5700
 aggacagtat ttggtatctg cgctctgctg aagccagtta ccttcggaaa aagagttggt 5760
 agctcttgat ccggcaaaaa aaccaccgct ggtagcgggtg gtttttttgt ttgcaagcag 5820
 cagattacgc gcagaaaaaa aggatctcaa gaagatcctt tgatcttttc tacggggctct 5880
 gacgctcagt ggaacgaaaa ctacagttaa gggattttgg tcatgagatt atcaaaaagg 5940

atcttcacct agatcctttt aaattaaaaa tgaagtttta aatcaatcta aagtatatat 6000
 gagtaaacctt ggtctgacag ttaccaatgc ttaatcagtg aggcacctat ctacgcatc 6060
 tgtctatttc gttcatccat agttgcctga ctccccgtcg tgtagataac tacgatacgg 6120
 gagggcttac catctggccc cagtgtctga atgataccgc gagaccacg ctaccggct 6180
 ccagatttat cagcaataaa ccagccagcc ggaagggccg agcgcagaag tggctctgca 6240
 actttatccg cctccatcca gtctattaat tgttgccggg aagctagagt aagtagttcg 6300
 ccagtttaata gtttgccgaa cgttgttgcc attgctacag gcacgtggt gtacgctcg 6360

tcgtttggta tggtttcatt cagctccggt tcccaacgat caaggcgagt tacatgatec 6420
 cccatgttgt gcaaaaaagc ggtagctcc ttccgtctctc cgatcgttgt cagaagtaag 6480
 ttggccgcag tgttatcact catggttatg gcagcactgc ataattctct tactgtcatg 6540
 ccatccgtaa gatgcttttc tgtgactggt gactactcaa ccaagtcatt ctgagaatag 6600
 tgtatgcggc gaccgagttg ctcttgcccc gcgtcaatac gggataatac cgcgccacat 6660
 agcagaactt taaaagtgt catcattgga aaacgttctt cggggcgaaa actctcaagg 6720
 atcttaccgc tgttgagatc cagttcgatg taaccactc gtgcacccaa ctgatcttca 6780

gcacctttta ctttcaccag cgtttctggg tgagcaaaaa caggaaggca aaatgccgca 6840
 aaaaagggaa taaggcgac acggaaatgt tgaatactca tactcttctt ttttcaatat 6900
 tattgaagca ttatcagggt ttattgtctc atgagcggat acatatttga atgtatttag 6960
 aaaaataaac aaatagggtg tccgcgcaca tttccccgaa aagtgccac 7009

<210> 84

<211> 7048

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 84

ctaaattgta agcgttaata ttttgtaaa attcgcgtta aatitttgtt aaatcagctc	60
atittttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga	120
gatagggttg agtgttgttc cagtttggaa caagagtcca ctattaaaga acgtggactc	180
caacgtcaaa gggcgaaaaa ccgtctatca gggcgatggc ccactacgtg aaccatcacc	240
ctaatcaagt tttttgggtt cgaggtgccg taaagcacta aatcggaacc cttaaaggag	300
ccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa	360
agcgaagga gcgggcgcta gggcgctggc aagtgtagcg gtcacgtgc gcgtaaccac	420
cacaccgcc gcgcttaatg cgccgtaca gggcgctcc cattcgccat tcaggtgcg	480
caactgttgg gaaggcgat cggcgggc ctcttcgcta ttacccagc tggcgaaagg	540
gggatgtgct gcaaggcgat taagttgggt aagccaggg ttttccagt cagcagttg	600
taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtacctt	660
taattctagt actatgcatg cgttgacatt gattattgac tagttattaa tagtaatcaa	720
ttacggggtc attagtcat agccatata tggagtccg cgttacataa cttacggtaa	780
atggccgcc tggctgaccg cccaacgacc ccgcccatt gacgtcaata atgacgtatg	840
ttcccatagt aacgccaata gggactttcc attgacgtca atgggtggag tatttacggt	900
aaactgccc cttggcagta catcaagtgt atcatatgcc aagtacgcc cctattgacg	960
tcaatgacgg taaatggccc gcctggcatt atgccagta catgacctta tgggactttc	1020
ctacttgga gtacatctac gtattagtca tcgctattac catggtgatg cggttttggc	1080
agtacatcaa tgggcgtgga tagcggtttg actcacggg atttccaagt ctccaccca	1140
ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg ggactttcca aaatgtcgta	1200
acaactccgc ccattgacg caaatgggcg gtaggcgtgt acggtgggag gtctatataa	1260
gcagagctct ctggctaact accggtgcc ccatggcccc aaagaagaag cggaaggtcg	1320
gtatccacgg agtcccagca gccaaaggga actacatcct gggcctggac atcggcacga	1380
ccagcgtggg ctacggcatc atcgactacg agacacggga cgtgatcgat gccggcgtgc	1440
ggctgttcaa agaggccaac gtggaaaaca acgagggcag gcggagcaag agaggcgcca	1500
gaaggctgaa gcggcggagg cggcatagaa tccagagagt gaagaagctg ctgttcgact	1560
acaacctgct gaccgaccac agcgagctga gcggcatcaa cccctacgag gccagagtga	1620

agggcctgag ccagaagctg agcgaggaag agttctctgc cgccttgctg cacctggcca	1680
agagaagagg cgtgcacaac gtgaacgagg tggaagagga caccggcaac gagctgtcca	1740
ccaaagagca gatcagccgg aacagcaagg ccctggaaga gaaatacgtg gccgaactgc	1800
agctggaacg gctgaagaaa gacggcgaag tgcggggcag catcaacaga ttcaagacca	1860
gcgactacgt gaaagaagcc aaacagctgc tgaaggtgca gaaggcctac caccagctgg	1920
accagagctt catcgacacc tacatcgacc tgcctggaac ccggcggacc tactatgagg	1980
gacctggcga gggcagcccc ttcggctgga aggacatcaa agaattgtac gagatgctga	2040
tgggccactg cacctacttc cccgaggaac tgcggagcgt gaagtacgcc tacaacgccg	2100
acctgtacaa cgcctgaac gacctgaaca atctcgtgat caccagggac gagaacgaga	2160
agctggaata ttacgagaag ttccagatca tcgagaacgt gttcaagcag aagaagaagc	2220
ccacctgaa gcagatcgcc aaagaaatcc tcgtgaacga agaggatatt aagggtaca	2280
gagtgaccag caccggcaag ccgagttca ccaacctgaa ggtgtaccac gacatcaagg	2340
acattaccgc ccgaaagag attattgaga acgccgagct gctggatcag attgccaaga	2400
tcctgacat ctaccagagc agcgaggaca tccaggaaga actgaccaat ctgaactccg	2460
agctgaccca ggaagagatc gagcagatct ctaatctgaa gggctatacc ggcaccaca	2520
acctgagcct gaaggccatc aacctgatcc tggacgagct gtggcacacc aacgacaacc	2580
agatcgctat cttaaccgg ctgaagctgg tgccaagaa ggtggacctg tcccagcaga	2640
aagagatccc caccacctg gtggacgact tcatectgag cccctcgtg aagagaagct	2700
tcateccagag catcaaagt atcaacgcca tcataagaa gtacggcctg cccaacgaca	2760
tcattatcga gctggccgc gagaagaact ccaaggacgc ccagaaaatg atcaacgaga	2820
tgcagaagcg gaaccggcag accaacgagc ggatcgagga aatcatccgg accaccggca	2880
aagagaacgc caagtacctg atcgagaaga tcaagctgca cgacatgcag gaaggcaagt	2940
gcctgtacag cctggaagcc atccctctgg aagatctgct gaacaacccc ttcaactatg	3000
aggtggacca catcatcccc agaagcgtgt ccttcgacaa cagcttcaac aacaaggtgc	3060
tcgtgaagca ggaagaaaac agcaagaagg gcaaccggac cccattccag tacctgagca	3120
gcagcgacag caagatcagc tacgaaacct tcaagaagca catcctgaat ctggccaagg	3180
gcaagggcag aatcagcaag accaagaaaag agtatctgct ggaagaacgg gacatcaaca	3240
ggttctccgt gcagaaagac ttcatcaacc ggaacctggg ggataccaga tacgccacca	3300
gaggcctgat gaacctgctg cggagctact tcagagtga caacctggac gtgaaagtga	3360
agtcacatca tggcgcttc accagctttc tgcggcggaa gtggaagttt aagaaagagc	3420

ggaacaaggg gtacaagcac cacgccgagg acgccctgat cattgccaac gccgatttca	3480
tcttcaaaga gtggaagaaa ctggacaagg ccaaaaaagt gatggaaaac cagatgttcg	3540
aggaaaagca ggccgagagc atgcccagaga tcgaaaccga gcaggagtac aaagagatct	3600
tcatcacccc ccaccagatc aagcacatta aggacttcaa ggactacaag tacagccacc	3660
gggtggacaa gaagccta atagagctga ttaacgacac cctgtactcc acccggaagg	3720
acgacaaggg caacacctg atcgtgaaca atctgaacgg cctgtacgac aaggacaatg	3780
acaagctgaa aaagctgac aacaagagcc ccgaaaagct gctgatgtac caccacgacc	3840
cccagaccta ccagaaactg aagctgatta tggaacagta cggcgacgag aagaatcccc	3900
tgtacaagta ctacgaggaa accgggaact acctgaccaa gtactccaaa aaggacaacg	3960
gccccgtgat caagaagatt aagtattacg gcaacaaact gaacgcccac ctggacatca	4020
ccgacgacta cccaacagc agaaacaagg tcgtgaagct gtcctgaag ccctacagat	4080
tcgacgtgta cctggacaat ggcgtgtaca agttcgtgac cgtgaagaat ctggatgtga	4140
tcaaaaaaga aaactactac gaagtgaata gcaagtgcta tgaggaagct aagaagctga	4200
agaagatcag caaccaggcc gagtttatcg cctccttcta caacaacgat ctgatcaaga	4260
tcaacggcga gctgtataga gtgatcggcg tgaacaacga cctgctgaac cggatcgaag	4320
tgaacatgat cgacatcacc taccgcgagt acctggaaaa catgaacgac aagaggcccc	4380
ccaggatcat taagacaatc gcctccaaga ccagagcat taagaagtac agcacagaca	4440
ttctgggcaa cctgtatgaa gtgaaatcta agaagcacc tcagatcatc aaaaagggca	4500
aaaggccggc ggccacgaaa aaggccggcc aggcaaaaaa gaaaaaggga tcctacccat	4560
acgatgttcc agattacgt taagaattcc tagagctcgc tgatcagcct cgactgtgcc	4620
ttctagtgtc cagccatctg ttgtttgccc ctccccgtg ccttccttga ccctggaagg	4680
tgccactccc actgtccttt cctaataaaa tgaggaaatt gcatcgcat gtctgagtag	4740
gtgtcattct attctggggg gtgggtggg gcaggacagc aagggggagg attgggaaga	4800
gaatagcagg catgctgggg aggtagcggc cgcccgcggt ggagctccag cttttgttcc	4860
ctttagttag ggtaattgc gcgcttggcg taatcatggt catagctgtt tcctgtgtga	4920
aattgtatc cgctcacaat tccacacaac atacgagccg gaagcataaa gtgtaaagcc	4980
tgggtgtcct aatgagttag ctaactcaca ttaattgcgt tgcgtcact gcccgcttcc	5040
cagtcgggaa acctgtcgtg ccagctgcat taatgaatcg gccaacgcgc ggggagaggc	5100
ggtttgcgta ttgggcgctc ttccgcttcc tcgtcactg actcgtcgc ctcggtcgtt	5160

cggtgcggc gagcggatc agtcactca aaggcggtaa tacggttatc cacagaatca 5220
 ggggataacg caggaaagaa catgtgagca aaaggccagc aaaaggccag gaaccgtaaa 5280
 aaggccgcgt tgctggcggtt ttccatagc ctccgcccc ctgacgagca tcacaaaaat 5340
 cgacgtcaa gtcagagggtg gcgaaccgc acaggactat aaagatacca ggcgtttccc 5400
 cctggaagct cctcgtgcg ctctctgtt ccgacctgc cgcttaccgg atacctgtcc 5460
 gcctttctcc ctccgggaag cgtggcgctt tctcatagc cacgtgtag gtatctcagt 5520
 tcggtgtagg tcgttcgtc caagctgggc tgtgtgcacg aacccccgt tcagcccgac 5580

 cgctgcgcct tatccggtaa ctatcgtctt gagtccaacc cgtaagaca cgacttatcg 5640
 ccactggcag cagccactgg taacaggatt agcagagcga ggatgtagg cgggtctaca 5700
 gagtcttga agtgggtggc taactacggc tacactagaa ggacagtatt tggatctgc 5760
 gctctgtga agccagttac ctccgaaaa agagtggta gctcttgatc cggcaaaaa 5820
 accaccgtg gtagcgggtg ttttttgtt tgcaagcagc agattacgc cagaaaaaa 5880
 ggatctcaag aagatcttt gatctttct acggggtctg acgtcagtg gaacgaaac 5940
 tcacgttaag ggattttgg catgagatta tcaaaaagga tcttcaccta gatctttta 6000

 aattaaaaat gaagtttta atcaatctaa agtatatatg agtaaaactg gtctgacagt 6060
 taccaatgct taatcagtga ggcacctatc tcagcgatct gtctatttcg ttcattcata 6120
 gtgacctgac tccccgtgt gtagataact acgatacggg agggcttacc atctggcccc 6180
 agtgctgcaa tgataccgc agaccacgc tcaccgctc cagatttatc agcaataaac 6240
 cagccagccg gaaggccga gcgcagaagt ggtcctgcaa ctttatccgc ctccatccag 6300
 tctattaatt gttgccggga agctagagta agtagttgc cagttaatag ttgcgcaac 6360
 gttgttgcca ttgtacagg catcgtggtg tcacgctcgt cgtttggtat ggcttcattc 6420

 agctccggtt cccaacgac aaggcgagt acatgatccc ccatgttgtg caaaaaagcg 6480
 gttagctcct tcggtcctcc gatcgttgtc agaagtaagt tggccgcagt gttatcctc 6540
 atggttatgg cagcactgca taattctctt actgtcatgc catccgtaag atgctttct 6600
 gtgactggtg agtactcaac caagtcattc tgagaatagt gtatgcggcg accgagtgc 6660
 tcttgcccg cgtaaatcg ggataatac gcgccacata gcagaacttt aaaagtgtc 6720
 atcattggaa aacgttcttc ggggcgaaaa ctctcaagga tcttaccgt gttgagatcc 6780
 agttcgtatg aaccactcg tgaccaaac tgatcttcag catcttttac tttaccagc 6840

 gtttctgggt gagcaaaaac aggaaggcaa aatgccgcaa aaaagggaat aaggcgaca 6900
 cggaatgtt gaatactcat actcttctt tttcaatatt attgaagcat ttatcagggt 6960
 tattgtctca tgacgggata catattgaa tgtatttaga aaaataaaca aatagggtt 7020

ccgcgcacat ttccccgaaa agtgccac 7048

<210> 85

<211> 48

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 85

aaaaggccgg cgccacgaa aaaggccggc caggcaaaaa agaaaaag 48

<210> 86

<211> 3223

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 86

ctaaattgta agcgttaata ttttgtaaa attcgcgtta aatttttgtt aaatcagctc 60

attttttaac caataggccg aaatcggcaa aatcccttat aaatcaaaag aatagaccga 120

gatagggttg agtggtgttc cagtttggaa caagagtcca ctattaaaga acgtggactc 180

caacgtcaaa gggcgaaaaa ccgtctatca gggcgcgtgc ccactacgtg aaccatcacc 240

ctaatcaagt tttttgggtt cgaggtgccg taaagcacta aatcggaacc ctaaaggagag 300

cccccgattt agagcttgac ggggaaagcc ggcgaacgtg gcgagaaagg aagggaagaa 360

agcgaaagga gggggcgcta gggcgcgtgc aagtgtagcg gtcacgtgc gcgtaaccac 420

cacaccgcc gcgttaatg cgccgtaca gggcgcgtec cattcgccat tcaggtgcg 480

caactgttgg gaaggcgcat cggtcggggc ctcttcgcta ttacgccagc tggcgaaagg 540

gggatgtgct gcaaggcgat taagttgggt aacgccaggg ttttccagt cagcagttg 600

taaaacgacg gccagtgagc gcgcgtaata cgactcacta tagggcgaat tgggtaccga 660

gggcctattt cccatgattc cttcatattt gcatatacga tacaaggctg ttagagagat 720

aattggaatt aatttgactg taaacacaaa gatattagta caaaatacgt gacgtagaaa 780

glaataatth ctgggtagt ttgcagtttt aaaattatgt tttaaaatgg actatcatat 840

gcttatcgta acttgaaagt atttcgattt ctggcttta tatacttgt ggaaaggacg 900

aaacaccggg ttttcgagaa gacctgtttt agagctagaa atagcaagtt aaaataaggc 960

tagtccgtta tcaacttgaa aaagtggcac cgagtcggtg ctttttttcc gcggtggagc 1020

tccagctttt gttcccttta gtgagggtta attgcgcgct tggcgtaatc atggtcatag 1080

ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc	1140
ataaagtgtg aagcctgggg tgcctaata gtagagtaac tcacattaat tgcgttgcgc	1200
tcactgcccc ctttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcggccaa	1260
cgcgcgggga gaggcgggtt gcgtattggg cgctcttccg cttcctcgct cactgactcg	1320
ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaagcg gtaataacgg	1380
ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag	1440
gccaggaacc gtaaaaagcg cgcgttgctg gcgtttttcc ataggctccg cccccctgac	1500
gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga	1560
taccaggcgt tccccctgg aagctccctc gtgcgctctc ctgttccgac cctgccgctt	1620
accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgctttctca tagctcacgc	1680
tgtaggtatc tcagttcggg ttaggtcgtt cgctccaagc tgggctgtgt gcacgaacct	1740
cccgttcagc ccgaccgctg cgccttatcc ggtaactatc gtcttgagtc caaccggta	1800
agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggtat	1860
gtaggcggtg ctacagagtt ctggaagtgg tggcctaact acggtacac tagaaggaca	1920
gtatttggta tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct	1980
tgatccggca aacaaaccac cgctggtagc ggtgggtttt ttgtttgcaa gcagcagatt	2040
acgcgcagaa aaaaaggatc tcaagaagat cctttgatct tttctacggg gtctgacgct	2100
cagtggaaacg aaaactcacg ttaagggtt ttggtcatga gattatcaa aaggatcttc	2160
acctagatcc ttttaaatga aaaatgaagt tttaaatcaa tctaaagtat atatgagtaa	2220
acttggctctg acagttacca atgcttaatc agtgaggcac ctatctcagc gatctgtcta	2280
tttcgttcat ccatagtgc ctgactcccc gtctgttaga taactacgat acgggagggc	2340
ttaccatctg gccccagtgc tgcaatgata ccgcgagacc cacgtcacc ggctccagat	2400
ttatcagcaa taaaccagcc agccggaagg gccgagcgca gaagtgtcc tgcaacttta	2460
tccgcctcca tccagtctat taattgttgc cggaagcta gagtaagtag ttcgccagtt	2520
aatagtttgc gcaacgttgt tgccattgct acaggcatcg tgggtgcacg ctcgctgttt	2580
ggtagtgctt cattcagctc cggttcccaa cgatcaagcg gagttacatg atccccatg	2640
ttgtgcaaaa aagcggttag ctcttcgggt cctccgatcg ttgtcagaag taagtggcc	2700
gcagtgttat cactcatggt tatggcagca ctgcataatt ctcttactgt catgccatcc	2760
gtaagatgct tttctgtgac tggtagtac tcaaccaagt cattctgaga atagtgtatg	2820

cggcgaccga gttgctcttg cccggcgtca atacgggata ataccgcgcc acatagcaga 2880

actttaaaag tgctcatcat tggaaaacgt tcttcggggc gaaaactctc aaggatctta 2940

ccgctgttga gatccagttc gatgtaacc actcgtgcac ccaactgac ttcagcatct 3000

tttactttca ccagcgtttc tgggtgagca aaaacaggaa ggcaaaatgc cgcaaaaaag 3060

ggaataaggg cgacacggaa atgttgaata ctcatctct tctttttca atattattga 3120

agcatttate agggttattg tctcatgagc ggatacatat ttgaatgtat ttagaaaaat 3180

aaacaaatag gggttccgcg cacatttccc cgaaaagtgc cac 3223

<210> 87

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 87

ctccggaatg tctccatttg 20

<210> 88

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 88

atgagggaga gactggcatc 20

<210> 89

<211> 23

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 89

gtttccagag ctttacctga gaa 23

<210> 90

<211> 22

<212> DNA

<213>

Artificial sequence

<220><223> Synthetic

<400> 90

cttttatgaa tgcttctcca ag 22

<210> 91

<211> 21

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 91

aagttacttg tccagcatg a 21

<210> 92

<211> 25

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 92

gaaaaacttc tgccaacttt tatca 25

<210> 93

<211> 25

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 93

tgcaaataac aaaagtagcc ataca 25

<210> 94

<211> 23

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 94

tcttttagaaa ggcttgaaag ctg 23

<210> 95

<211> 20

<212> DNA

<213> Artificial sequence
 <220><223> Synthetic
 <400> 95
 cttcactgct ggccagttta 20

<210> 96
 <211> 21
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 96
 ttgaacatgg cattgcataa a 21

<210> 97
 <211> 20
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 97
 cgggcttgga cagaacttac 20

<210> 98
 <211> 20
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 98
 ctgcgtagt ccaaaacaaa 20

<210> 99
 <211> 22
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 99
 gagatgtctt ttgcagcttt cc 22

<210> 100

<211> 20
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 100
 gggaccttgg taaagccaca 20
 <210> 101
 <211> 21
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 101
 tgcctttcaa tcattgtttc g 21

 <210> 102
 <211> 20
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 102
 agaaggcaaa ttggcacaga 20
 <210> 103
 <211> 20
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 103
 aaggcccaa aatgtgaaat 20
 <210> 104
 <211> 25
 <212> DNA
 <213> Artificial sequence
 <220><223> Synthetic
 <400> 104

gagttttggct caaattgtta ctctt

25

<210> 105

<211> 18

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 105

tggcggcggtt ttcattat

18

<210> 106

<211> 26

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 106

ttcgatccgt aatgattgtt ctagcc

26

<210> 107

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 107

ttgtgtgtcc catgcttggt

20

<210> 108

<211> 20

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 108

caacgctgaa gaaccctgat

20

<210> 109

<211> 24

<212> DNA

<213> Artificial sequence

<220><223> Synthetic

<400> 109

tttctgtgat tttcttttgg attg

24