(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0113940
(43) 공개일자 2022년08월17일

(51) 국제특허분류(Int. Cl.)
C12N 15/63 (2006.01) *A61K 48/00* (2006.01)
C12N 15/115 (2010.01) *C12N 15/86* (2006.01)
(52) CPC특허분류
C12N 15/63 (2013.01)
A61K 48/005 (2013.01)
(21) 출원번호 10-2022-7019191
(22) 출원일자(국제) 2020년09월30일
심사청구일자 없음
(85) 번역문제출일자 2022년06월07일
(86) 국제출원번호 PCT/US2020/053643
(87) 국제공개번호 WO 2021/096605
국제공개일자 2021년05월20일
(30) 우선권주장
62/933,714 2019년11월11일 미국(US)
PCT/US2020/025430 2020년03월27일 미국(US)

(71) 출원인
솔크 인스티튜트 포 바이올로지칼 스터디즈
미국 캘리포니아 92037-1002 라졸라 노쓰 토리 파인즈 로드10010
(72) 발명자
바흐만 루카스 크리스토프
미국 캘리포니아 92037-1002 라호야 노쓰 토리 파인즈 로드10010
파프 새뮤얼 로렌스
미국 캘리포니아 92037-1002 라호야 노쓰 토리 파인즈 로드10010
(74) 대리인
리앤목특허법인

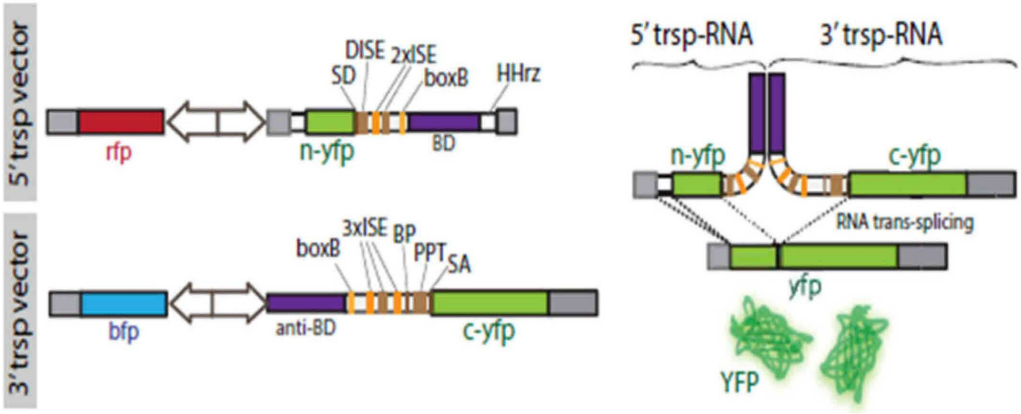
전체 청구항 수 : 총 50 항

(54) 발명의 명칭 **RNA 분자의 고-효율 재조합을 위한 조성물 및 방법**

(57) 요 약

RNA 분자를 이용하는 방법을 포함한, RNA 분자의 재구성을 위한 조성물 및 시스템이 제공된다. 예를 들면, 그러한 분자들은 2개 이상의 바이러스 벡터 (예를 들면, AAV)를 통해 단백질 코딩 서열을 전달하기 위해 이용되어, 세포에서 전장 단백질의 재구성을 가져올 수 있다. 그러한 방법은 치료 단백질, 예를 들면, 유전 질환 또는 암을 치료하는 치료 단백질을 전달하기 위해 이용될 수 있다.

대 표 도 - 도1a



(52) CPC특허분류
C12N 15/115 (2013.01)
C12N 15/86 (2013.01)
C12N 2310/16 (2013.01)
C12N 2750/14143 (2013.01)
C12N 2840/445 (2013.01)

명 세 서

청구범위

청구항 1

표적 단백질의 발현용 조성물로서, (a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너(splice donor); 및 (iii) 제 1 이합체화 도메인;을 포함하는 것인 제1 RNA 분자; 및 (b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; (ii) 분지점(branch point) 서열; (iii) 폴리피리미딘 트랙트(polypyrimidine tract); (iv) 스플라이스 억셉터(splice acceptor); 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제2 RNA 분자를 포함하는 것인 조성물.

청구항 2

청구항 1에 있어서, 상기 제1 이합체화 도메인과 제2 이합체화 도메인은 직접 결합, 간접 결합, 또는 이들의 조합에 의해 결합하는 것인 조성물.

청구항 3

청구항 2에 있어서, 직접 결합 또는 간접 결합은 염기쌍 형성 상호작용, 비-규범적(non-canonical) 염기쌍 형성 상호작용, 비-염기쌍 형성 상호작용(non-base pairing interaction), 또는 이들의 조합을 포함하는 것인 조성물.

청구항 4

청구항 2 또는 3에 있어서, 직접 결합은 키싱 루프(kissing loop) 또는 저다양성(hypodiverse) 영역간 염기쌍 형성 상호작용을 포함하는 것인 조성물.

청구항 5

청구항 2 또는 3에 있어서, 직접 결합은 압타머 영역간 규범적 염기쌍 형성 상호작용, 비-규범적 염기쌍 형성 상호작용, 비-염기쌍 형성 상호작용, 또는 이들의 조합을 포함하는 것인 조성물.

청구항 6

청구항 2 또는 3에 있어서, 간접 결합은 핵산 브릿지(nucleic acid bridge)를 통한 염기쌍 형성 상호작용을 포함하는 것인 조성물.

청구항 7

청구항 2에 있어서, 간접 결합은 압타머와 압타머 표적간, 또는 2개의 압타머간 비-염기쌍 형성 상호작용을 포함하는 것인 조성물.

청구항 8

청구항 1 내지 7 중 어느 한 항에 있어서, 상기 제1 이합체화 도메인 또는 제2 이합체화 도메인은 크립틱 스플라이스 억셉터(cryptic splice acceptor)를 포함하지 않는 것인 조성물.

청구항 9

청구항 1 내지 8 중 어느 한 항에 있어서, 상기 이합체화 도메인은 직접적으로 또는 간접적으로 결합하는 압타머 서열 이합체화 도메인인 것인 조성물.

청구항 10

청구항 1 내지 9 중 어느 한 항에 있어서, 상기 이합체화 도메인은 키싱 루프 상호작용 도메인인 것인 조성물.

청구항 11

청구항 1 내지 10 중 어느 한 항에 있어서, 상기 표적 단백질은 질병과 연관된 단백질, 또는 치료 단백질인 것인 조성물.

청구항 12

청구항 11에 있어서, 상기 질병은 단일유전자성(monogenic) 질병인 것인 조성물.

청구항 13

청구항 12에 있어서, 상기 치료 단백질은 독소인 것인 조성물.

청구항 14

청구항 11 내지 13 중 어느 한 항에 있어서, 상기 질병 및 상기 표적 단백질은 표 1에 열거된 것인 조성물.

청구항 15

청구항 1 내지 14 중 어느 한 항에 있어서, 상기 제1 RNA 분자는 상기 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 DISE(downstream intronic splice enhancer), 상기 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 ISE (intronic splice enhancer) 중 하나, 또는 둘 모두를 더 포함하고; 및/또는

상기 제2 RNA 분자는 상기 제2 이합체화 도메인의 3'이고 상기 분지점 서열의 5'인 ISE 및 상기 스플라이스 도너의 3'이고, 상기 이합체화 도메인의 5'인 DISE 중 하나, 또는 둘 모두를 더 포함하거나; 또는

이들의 조합인 것인 조성물.

청구항 16

청구항 1 내지 15 중 어느 한 항에 있어서,

상기 제1 RNA 분자는 상기 스플라이스 도너의 3'으로부터 임의의 위치에 배치된 자가-절단 RNA 서열 또는 RNA-절단 효소 표적 서열을 더 포함하여, 3'에 위치한 폴리아데닐화 테일을 절단하여 비-재조합 RNA 분자로부터의 단백질 단편 발현을 감소시키거나 억제하거나;

상기 제2 RNA 분자는 상기 분지점 서열의 5'으로부터 임의의 위치에 배치된 자가-절단 RNA 서열 또는 RNA-절단 효소 표적 서열을 더 포함하여, 5'에 위치한 RNA 캡을 절단하여 비-재조합 RNA 분자로부터의 단백질 절편 발현을 감소시키거나 억제하거나;

상기 제2 RNA 분자는 상기 스플라이스 억셉터의 3'에 있는 ORF(open reading frame) 대비 이동된 분지점 서열의 5'으로부터 임의의 위치에 배치된 개시 코돈을 더 포함하여, 비-재조합 RNA 분자로부터 표적 단백질 절편의 번역을 감소시키거나 또는 억제하거나;

상기 제1 RNA 분자는 상기 스플라이스 도너의 3'으로부터 임의의 위치에 마이크로 RNA 표적 부위를 더 포함하여, 비-연결(un-joined) RNA 단편이 핵 외부에서 마이크로 RNA 의존적 분해를 겪게 하거나;

상기 제2 RNA 분자는 상기 코딩 서열의 3'으로부터 임의의 위치에 마이크로 RNA 표적 부위를 더 포함하여, 비-연결 RNA 단편이 핵 외부에서 마이크로 RNA 의존적 분해를 겪게 하거나;

상기 제1 RNA 분자는 상기 스플라이스 도너의 5'에 있는 표적 단백질 ORF와 동일한 프레임 내에 위치하도록 상기 스플라이스 도너의 3'으로부터 임의의 위치에 데그론(degron) 단백질 분해 태그를 코딩하는 서열을 더 포함하여, 비-연결 단백질 단편이 분해를 위해 태깅되게 하거나;

상기 제2 RNA 분자는 상기 스플라이스 억셉터 부위의 3'에 있는 표적 단백질 ORF와 동일한 프레임 내에 위치하도록 상기 분지점 서열의 5'으로부터 임의의 위치에 개시 코돈 및 인-프레임(in-frame) 데그론 단백질 분해 태그를 더 포함하여, 비-연결 단백질 단편이 분해를 위해 태깅되게 하거나; 또는

이들의 조합인 것인 조성물.

청구항 17

(a) 청구항 1 내지 16 중 어느 한 항의 제1 RNA 분자를 코딩하는 제1 합성 DNA 분자로서, 상기 제1 합성 DNA 분자는 (i) 상기 제1 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제1 프로모터를 포함하는 것인 제1 합성 DNA 분자; 및 (b) 청구항 1 내지 16 중 어느 한 항에 따른 제2 RNA 분자를 코딩하는 제2 합성 DNA 분자로서, 상기 제2 합성 DNA 분자는 (i) 상기 제2 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제2 프로모터를 포함하는 것인 제2 합성 DNA 분자를 포함하는, 표적 단백질의 발현용 조성물.

청구항 18

청구항 17에 있어서, 각 프로모터는 독립적으로 선택되는 것인 조성물.

청구항 19

청구항 18 또는 19에 있어서,

상기 제1 프로모터와 상기 제2 프로모터는 동일한 프로모터이거나; 또는

상기 제1 프로모터와 상기 제2 프로모터는 상이한 프로모터인 것인 조성물.

청구항 20

청구항 17 내지 19 중 어느 한 항에 있어서, 상기 제1 프로모터 및 상기 제2 프로모터 각각은 항시적 프로모터; 조직-특이적 프로모터; 및 상기 표적 단백질에 내생적인(endogenous) 프로모터로부터 독립적으로 선택되는 것인 조성물.

청구항 21

청구항 17 내지 20 중 어느 한 항의 조성물을 포함하는, 표적 단백질의 발현용 시스템.

청구항 22

청구항 21에 있어서, 상기 시스템이 세포에 도입되는 경우, 상기 RNA 분자가 생성되고 적합한 순서로 재조합되어, 상기 표적 단백질의 전장 코딩 서열이 형성되는 것인 시스템.

청구항 23

청구항 21 또는 22에 있어서, 상기 제1 합성 및 제2 합성 RNA 분자 각각은 별개의 바이러스 벡터로부터 전사되는 것인 시스템.

청구항 24

청구항 21 내지 23 중 어느 한 항에 있어서, 상기 바이러스 벡터는 AAV인 것인 시스템.

청구항 25

청구항 21 내지 24 중 어느 한 항에 있어서, 상기 합성 DNA 분자 각각은: 약 2500 nt 내지 약 5000 nt, 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,500 nt 내지 약 4,750 nt, 약 2,500 nt 내지 약 5,000 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 4,750 nt, 약 2,750 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 4,750 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 4,750 nt, 약 3,250 nt 내지 약 5,000 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 4,750 nt, 약 3,500 nt 내지 약 5,000 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,750 nt, 약 3,750 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,750 nt, 약 4,000

nt 내지 약 5,000 nt, 약 4,250 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,750 nt, 약 4,250 nt 내지 약 5,000 nt, 약 4,500 nt 내지 약 4,750 nt, 약 4,500 nt 내지 약 5,000 nt, 약 4,750 nt 내지 약 5,000 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 약 4,500 nt, 약 4,750 nt, 및 약 5,000 nt로부터 독립적으로 선택된 크기를 갖는 것인 시스템.

청구항 26

청구항 21 내지 25 중 어느 한 항에 있어서, 상기 시스템의 합성 DNA 분자에 의해 코딩되는 표적 단백질의 N-말단 부분, 또는 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열은 각각: 약 2500 내지 4500 nt, 약 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,500 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000 nt, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 및 약 4,500 nt로부터 독립적으로 선택된 크기를 갖는 것인 시스템.

청구항 27

청구항 21 내지 26 중 어느 한 항에 있어서, 상기 시스템의 합성 DNA 분자에 의해 코딩되는 RNA 분자 중 하나 또는 둘 모두는 각각 약 2500 내지 4500 nt, 약 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,500 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000 nt, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 및 약 4,500 nt로부터 독립적으로 선택되는 크기를 갖는 것인 시스템.

청구항 28

청구항 21 내지 27 중 어느 한 항에 있어서, 상기 시스템은 청구항 17 내지 20 중 어느 한 항의 조성물을 포함하고,

상기 합성 DNA 분자는 약 5000 nt 내지 약 10,000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 9,500 nt, 약 5,000 nt 내지 약 10,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 8,500 nt, 약 5,500 nt 내지 약 9,000 nt, 약 5,500 nt 내지 약 9,500 nt, 약 5,500 nt 내지 약 10,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 9,500 nt, 약 6,000 nt 내지 약 10,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt,

약 6,500 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 8,500 nt, 약 6,500 nt 내지 약 9,000 nt, 약 6,500 nt 내지 약 9,500 nt, 약 6,500 nt 내지 약 10,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 9,500 nt, 약 7,000 nt 내지 약 10,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 9,500 nt, 약 7,500 nt 내지 약 10,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 9,500 nt, 약 8,000 nt 내지 약 10,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,500 nt, 약 8,500 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 9,500 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,500 nt 내지 약 10,000 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 약 9,000 nt, 약 9,500 nt, 및 약 10,000 nt로부터 선택된 전체 크기를 갖고;

전체 표적 단백질 코딩 서열은 약 2000 nt 내지 약 8000 nt, 약 2,000 nt 내지 약 3,000 nt, 약 2,000 nt 내지 약 3,500 nt, 약 2,000 nt 내지 약 4,000 nt, 약 2,000 nt 내지 약 4,500 nt, 약 2,000 nt 내지 약 5,000 nt, 약 2,000 nt 내지 약 5,500 nt, 약 2,000 nt 내지 약 6,000 nt, 약 2,000 nt 내지 약 6,500 nt, 약 2,000 nt 내지 약 7,000 nt, 약 2,000 nt 내지 약 7,500 nt, 약 2,000 nt 내지 약 8,000 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 5,500 nt, 약 3,000 nt 내지 약 6,000 nt, 약 3,000 nt 내지 약 6,500 nt, 약 3,000 nt 내지 약 7,000 nt, 약 3,000 nt 내지 약 7,500 nt, 약 3,000 nt 내지 약 8,000 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 5,000 nt, 약 3,500 nt 내지 약 5,500 nt, 약 3,500 nt 내지 약 6,000 nt, 약 3,500 nt 내지 약 6,500 nt, 약 3,500 nt 내지 약 7,000 nt, 약 3,500 nt 내지 약 7,500 nt, 약 3,500 nt 내지 약 8,000 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 5,500 nt, 약 4,000 nt 내지 약 6,000 nt, 약 4,000 nt 내지 약 6,500 nt, 약 4,000 nt 내지 약 7,000 nt, 약 4,000 nt 내지 약 7,500 nt, 약 4,000 nt 내지 약 8,000 nt, 약 4,500 nt 내지 약 5,000 nt, 약 4,500 nt 내지 약 5,500 nt, 약 4,500 nt 내지 약 6,000 nt, 약 4,500 nt 내지 약 6,500 nt, 약 4,500 nt 내지 약 7,000 nt, 약 4,500 nt 내지 약 7,500 nt, 약 4,500 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 또는 약 7,500 nt 내지 약 8,000 nt로부터 선택되고, 상기 전체 표적 단백질 코딩 서열은 약 2,000 nt, 약 3,000 nt, 약 3,500 nt, 약 4,000 nt, 약 4,500 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 및 약 8,000 nt이고; 및/또는

상기 2개의 합성 DNA 분자에 의해 코딩되는 RNA 분자의 합한 크기(summed size)는 약 5,000 nt 내지 약 9000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 8,500 nt, 약 5,500 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 8,500 nt, 약 6,500 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 및 약 9,000 nt로부터 선택되는 것인 시스템.

청구항 29

청구항 21 내지 28 중 어느 한 항에 있어서, 상기 제1 이합체화 도메인 및 상기 제2 이합체화 도메인은 각각 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어

도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 및

상기 시스템은 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 55%, 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95%, 또는 약 100%의 재조합 효율을 갖는 것인 시스템.

청구항 30

청구항 21 내지 29 중 어느 한 항에 있어서, 각각의 이합체화 도메인은 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 및

상기 시스템은 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 또는 약 100%의 재조합 효율을 갖는 것인 시스템.

청구항 31

청구항 21 내지 30 중 어느 한 항에 있어서, 상기 RNA 재조합 효율은 약 10% 내지 약 100%, 약 10% 내지 약 20%, 약 10% 내지 약 30%, 약 10% 내지 약 35%, 약 10% 내지 약 40%, 약 10% 내지 약 45%, 약 10% 내지 약 50%, 약 10% 내지 약 55%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 20% 내지 약 30%, 약 20% 내지 약 35%, 약 20% 내지 약 40%, 약 20% 내지 약 45%, 약 20% 내지 약 50%, 약 20% 내지 약 55%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 30% 내지 약 35%, 약 30% 내지 약 40%, 약 30% 내지 약 45%, 약 30% 내지 약 50%, 약 30% 내지 약 55%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 35% 내지 약 40%, 약 35% 내지 약 45%, 약 35% 내지 약 50%, 약 35% 내지 약 55%, 약 35% 내지 약 60%, 약 35% 내지 약 70%, 약 35% 내지 약 80%, 약 35% 내지 약 90%, 약 40% 내지 약 45%, 약 40% 내지 약 50%, 약 40% 내지 약 55%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 45% 내지 약 50%, 약 45% 내지 약 55%, 약 45% 내지 약 60%, 약 45% 내지 약 70%, 약 45% 내지 약 80%, 약 45% 내지 약 90%, 약 50% 내지 약 55%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 55% 내지 약 60%, 약 55% 내지 약 70%, 약 55% 내지 약 80%, 약 55% 내지 약 90%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 80% 내지 약 90%, 약 10%, 약 20%, 약 30%, 약 35%, 약 40%, 약 45%, 약 50%, 약 55%, 약 60%, 약 70%, 약 80%, 약 90%, 약 95%, 또는 약 100%인 것인 시스템.

청구항 32

청구항 21 내지 31 중 어느 한 항의 시스템을 포함하는 조성물.

청구항 33

청구항 32에 있어서, 상기 조성물은 각각 디스트로핀(dystrophin), 인자 8(factor 8), ABCA4, 또는 MYO7A의 적어도 일부를 코딩하는, 제1 RNA 분자, 제2 RNA 분자, 제3 RNA 분자, 및 선택적으로 제4 RNA 분자를 포함하는 것인 조성물.

청구항 34

청구항 21 내지 31 중 어느 한 항의 시스템, 또는 청구항 44 및 45 중 어느 한 항의 조성물을 포함하는 키트로서, 합성 제1 핵산 분자, 제2 핵산 분자, 제3 핵산 분자, 및 제4 핵산 분자는 별개의 용기에 담길 수 있고, 선택적으로, 약학적으로 허용가능한 담체와 같은 완충액을 더 포함하는 것인 키트.

청구항 35

세포에서 표적 단백질을 발현시키는 방법으로서,

청구항 21 내지 31 중 어느 한 항의 시스템, 또는 청구항 32 또는 33의 조성물을 세포 내로 도입시키는 단계, 및 상기 세포에서 제1 RNA 분자 및 제2 RNA 분자를 발현시키는 단계를 포함하고, 상기 표적 단백질이 상기 세포에서 생산되는 것인 방법.

청구항 36

청구항 35에 있어서, 상기 세포는 개체 내에 존재하고, 상기 도입시키는 단계는 상기 시스템의 치료 유효량을 상기 개체에게 투여하는 단계를 포함하는 것인 방법.

청구항 37

청구항 36에 있어서, 상기 방법은 상기 개체에서 상기 표적 단백질을 코딩하는 유전자 중 돌연변이에 의해 유발된 유전 질환을 치료하고, 상기 방법은 상기 개체에서 기능성 표적 단백질의 발현을 가져오는 것인 방법.

청구항 38

청구항 37에 있어서,

상기 유전 질환은 듀센 근이영양증 (Duchenne muscular dystrophy)이고, 상기 표적 단백질은 디스트로핀이거나;

상기 유전 질환은 A형 혈우병이고, 상기 표적 단백질은 F8이거나;

상기 유전 질환은 스타르가르트병(Stargardt disease)이고, 상기 표적 단백질은 ABCA4이거나; 또는

상기 유전 질환은 어셔 증후군(Usher syndrome)이고, 상기 표적 단백질은 MYO7A인 것인 방법.

청구항 39

청구항 21 내지 31 중 어느 한 항, 청구항 1 내지 16, 32, 및 33 중 어느 한 항, 또는 청구항 35 내지 38 중 어느 한 항에 있어서, 상기 제1 RNA 분자 및 상기 제2 RNA 분자 중 하나 또는 둘 모두는 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166 중 어느 하나로 제공되는 합성 인트론에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100% 서열 동일성을 포함하는 것인 시스템, 조성물, 또는 방법.

청구항 40

청구항 21 내지 31 및 39 중 어느 한 항, 청구항 1 내지 16, 32, 및 33 중 어느 한 항, 또는 청구항 35 내지 38 중 어느 한 항에 있어서, 상기 제1 RNA 분자 및 상기 제2 RNA 분자 중 하나 또는 둘 모두는 서열번호 20의 nt 3703 내지 3975, 서열번호 21의 nt 1 내지 228, 서열번호 22의 nt 3703 내지 3975, 서열번호 23의 nt 1 내지 225, 서열번호 24의 nt 3560 내지 3828, 및 서열번호 25의 nt 1-225로부터 선택된 합성 인트론을 포함하는 것인 시스템, 조성물, 또는 방법.

청구항 41

청구항 21 내지 31, 39, 및 40 중 어느 한 항, 청구항 1 내지 16, 32, 및 33 중 어느 한 항, 또는 청구항 35 내지 38 중 어느 한 항에 있어서, 상기 제1 RNA 분자 및 상기 제2 RNA 분자 중 하나 또는 둘 모두는 단백질 코딩 서열의 부분을 더 포함하는 것인 시스템, 조성물, 또는 방법.

청구항 42

청구항 21 내지 31 및 39 내지 41 중 어느 한 항, 청구항 1 내지 16, 32, 및 33 중 어느 한 항, 또는 청구항 35 내지 38 중 어느 한 항에 있어서, 상기 단백질 코딩 서열의 부분은 상기 단백질 코딩 서열의 N-말단 절반(N-terminal half), N-말단 부분(N-terminal portion), C-말단 절반, 또는 C-말단 부분을 포함하는 것인 시스템, 조성물, 또는 방법.

청구항 43

청구항 21 내지 31 및 39 내지 42 중 어느 한 항, 청구항 1 내지 16, 32, 및 33 중 어느 한 항, 또는 청구항 35 내지 38 중 어느 한 항에 있어서,

(a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로 (i) 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너; (ii-2) DISE, ISE, 또는 둘 모두; 및 (iii) 제1 이합체화 도메인을 포함하는 것인 제1 RNA 분자; 및

(b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로 (i) 상기 제1 이합체화 도메인에 결합하는, 제2 이합체화 도메인; (i-2) 하나 이상의 ISE 서열; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제2 RNA 분자를 포함하는 것인 시스템, 조성물, 또는 방법.

청구항 44

청구항 21 내지 31 및 39 내지 43 중 어느 한 항, 청구항 1 내지 16, 32, 및 33 중 어느 한 항, 또는 청구항 35 내지 38 중 어느 한 항에 있어서,

(a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너; (ii-2) DISE, ISE, 및 ISE; 및 (iii) 제1 이합체화 도메인을 포함하는 것인 제1 RNA 분자; 및

(b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 제1 이합체화 도메인에 결합하는, 제2 이합체화 도메인; (i-2) 3개의 ISE 서열; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제2 RNA 분자를 포함하는 것인 시스템, 조성물, 또는 방법.

청구항 45

청구항 1 내지 16 중 어느 한 항에 있어서, 상기 제1 RNA 분자 및 상기 제2 RNA 분자 중 하나 또는 둘 모두는 각각: 약 2500 내지 4500 nt, 약 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,500 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000 nt, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 및 약 4,500 nt로부터 독립적으로 선택된 크기를 갖는 것인 조성물.

청구항 46

청구항 1 내지 16 중 어느 한 항에 있어서,

전체 표적 단백질 코딩 서열 크기는 약 2000 nt 내지 약 8000 nt, 약 2,000 nt 내지 약 3,000 nt, 약 2,000 nt 내지 약 3,500 nt, 약 2,000 nt 내지 약 4,000 nt, 약 2,000 nt 내지 약 4,500 nt, 약 2,000 nt 내지 약 5,000 nt, 약 2,000 nt 내지 약 5,500 nt, 약 2,000 nt 내지 약 6,000 nt, 약 2,000 nt 내지 약 6,500 nt, 약 2,000 nt 내지 약 7,000 nt, 약 2,000 nt 내지 약 7,500 nt, 약 2,000 nt 내지 약 8,000 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 5,500 nt, 약 3,000 nt 내지 약 6,000 nt, 약 3,000 nt 내지 약 6,500 nt, 약 3,000 nt 내지 약 7,000 nt, 약 3,000 nt 내지 약 7,500 nt, 약 3,000 nt 내지 약 8,000 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 5,000 nt, 약 3,500 nt 내지 약 5,500 nt, 약 3,500 nt 내지 약 6,000 nt, 약 3,500 nt 내지 약 6,500 nt, 약 3,500 nt 내지 약 7,000 nt, 약 3,500 nt 내지 약 7,500 nt, 약 3,500 nt 내지 약 8,000 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 5,500 nt, 약 4,000 nt 내지 약 6,000 nt, 약 4,000 nt 내지 약 6,500 nt, 약 4,000 nt 내지 약 7,000 nt, 약 4,000 nt 내지 약 7,500 nt, 약 4,000 nt 내지 약 8,000 nt, 약 4,500 nt 내지 약 5,000 nt, 약 4,500 nt 내지 약 5,500 nt, 약 4,500 nt 내지 약 6,000 nt, 약 4,500 nt 내지 약 6,500 nt, 약 4,500 nt 내지 약 7,000 nt, 약 4,500 nt 내지 약 7,500 nt, 약 4,500 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt

내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 2,000 nt, 약 3,000 nt, 약 3,500 nt, 약 4,000 nt, 약 4,500 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 또는 약 8,000 nt이고; 및/또는

상기 2개의 RNA 분자의 합한 크기는 약 5,000 nt 내지 약 9000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 8,500 nt, 약 5,500 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 8,500 nt, 약 6,500 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 또는 약 9,000 nt인 것인 조성물.

청구항 47

청구항 1 내지 16 중 어느 한 항에 있어서, 상기 제1 이합체화 도메인 및 상기 제2 이합체화 도메인은 각각 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 상기 시스템은 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 55%, 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 또는 적어도 약 95%의 재조합 효율을 갖는 것인 조성물.

청구항 48

청구항 1 내지 16 중 어느 한 항에 있어서, 각각의 이합체화 도메인은 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 상기 시스템은 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 75%, 적어도 80%, 또는 적어도 90%의 재조합 효율을 갖는 것인 조성물.

청구항 49

청구항 1 내지 16 중 어느 한 항에 있어서, RNA 재조합 효율은 약 10% 내지 약 100%, 약 10% 내지 약 20%, 약 10% 내지 약 30%, 약 10% 내지 약 35%, 약 10% 내지 약 40%, 약 10% 내지 약 45%, 약 10% 내지 약 50%, 약 10% 내지 약 55%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 20% 내지 약 30%, 약 20% 내지 약 35%, 약 20% 내지 약 40%, 약 20% 내지 약 45%, 약 20% 내지 약 50%, 약 20% 내지 약 55%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 30% 내지 약 35%, 약 30% 내지 약 40%, 약 30% 내지 약 45%, 약 30% 내지 약 50%, 약 30% 내지 약 55%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 35% 내지 약 40%, 약 35% 내지 약 45%, 약 35% 내지 약 50%, 약 35% 내지 약 55%, 약 35% 내지 약 60%, 약 35% 내지 약 70%, 약 35% 내지 약 80%, 약 35% 내지 약 90%, 약 40% 내지 약 45%, 약 40% 내지 약 50%, 약 40% 내지 약 55%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 45% 내지 약 50%, 약 45% 내지 약 55%, 약 45% 내지 약 60%, 약 45% 내지 약 70%, 약 45% 내지 약 80%, 약 45% 내지 약 90%, 약 50% 내지 약 55%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 55% 내지 약

60%, 약 55% 내지 약 70%, 약 55% 내지 약 80%, 약 55% 내지 약 90%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 80% 내지 약 90%, 약 10%, 약 20%, 약 30%, 약 35%, 약 40%, 약 45%, 약 50%, 약 55%, 약 60%, 약 70%, 약 80%, 약 90%, 약 95%, 또는 약 100%인 것인 조성물.

청구항 50

청구항 1 내지 16 중 어느 한 항에 있어서,

(a) 상기 제1 RNA 분자 및 상기 제2 RNA 분자는 각각 약 2500 nt 내지 4500 nt이고;

(b) 전체 표적 단백질 코딩 서열 크기는 약 2000 nt 내지 약 8000 nt이며; 및/또는

(c) 상기 2개의 RNA 분자의 합한 크기는 약 5,000 nt 내지 약 9000 nt이며; 및 RNA 재조합 효율은 약 10% 내지 약 100%인 것인 조성물.

발명의 설명

기 술 분 야

[0001] **관련 출원에 대한 교차-참조**

[0002] 본 출원은 2019년 11월 11일에 출원된 미국 임시 출원 제62/933,714호에 기초한 우선권을 주장하는, 2020년 3월 27일에 출원된 PCT/US2020/025430의 일부 계속 출원이고, 이들 출원은 모두 전체로 참조에 의해 본 명세서에 포함된다.

[0003] 본 개시는 2개 이상의 RNA 분자의 재조합을 허용하여, 전장-단백질의 발현을 가능하게 하는 시스템, 키트, 조성물, 및 방법을 제공한다.

배 경 기 술

[0004] 유전자 치료는 기능상실(loss-of-function) 돌연변이에 의해 유발되는 유전 질환의 유망한 치료 방법이다. 대체 유전자들은 전형적으로 AAV와 같은 벡터를 이용하여 표적 세포에 재도입되고, 이는 상기 바이러스가 일반적으로 세포로의 진입에 안전하고 효율적이기 때문이다. 그러나, AAV의 경우에, 종래의 캡시드를 이용하여 약 5000개 이상의 뉴클레오티드를 캡슐화하기 어렵다. 큰 단백질을 코딩하는 유전자의 길이는 종종 AAV의 패키징 제한을 초과하므로, 많은 유전 질환들이 치료될 수 없는 상태이다. 이러한 한계를 극복하기 위한 전략이 과거에 모색되었으나, 비효율적인 것으로 판명되거나, 잠재적으로 유해한 절단된 단백질의 고수준 발현을 초래하거나, 또는 양자 모두였다. 질병을 치료하는 큰 단백질의 전달을 위한 안전한, 고 효율 전략이 요구된다.

발명의 내용

해결하려는 과제

[0005] **요약**

[0006] 표적 단백질의 발현을 위한 조성물이 본 명세서에 제공된다. 일 구체예(example)에서, 상기 조성물은 (a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너(splice donor); 및 (iii) 제 1 이합체화 도메인;을 포함하는 것인 제1 RNA 분자; 및 (b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; (ii) 분지점(branch point) 서열; (iii) 폴리피리미딘 트랙트(polypyrimidine tract); (iv) 스플라이스 억셉터(splice acceptor); 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제2 RNA 분자를 포함한다.

[0007] 일부 구체예에서, 상기 제1 이합체화 도메인과 제2 이합체화 도메인은 직접 결합, 간접 결합, 또는 이들의 조합에 의해 결합한다.

[0008] 일부 구체예에서, 상기 이합체화 도메인은 키싱 루프(kissing loop) 도메인 또는 저다양성(hypodiverse) 도메인이다.

[0009] 일부 구체예에서, 상기 제1 및/또는 제2 RNA 분자는 적어도 하나의 스플라이스 인핸서(splice enhancer)를 포함

한다.

[0010] 또한, (a) 청구항 1 내지 16 중 어느 한 항의 제1 RNA 분자를 코딩하는 제1 합성 DNA 분자로서, 상기 제1 합성 DNA 분자는 (i) 상기 제1 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제1 프로모터를 포함하는 것인 제1 합성 DNA 분자; 및 (b) 청구항 1 내지 16 중 어느 한 항에 따른 제2 RNA 분자를 코딩하는 제2 합성 DNA 분자로서, 상기 제2 합성 DNA 분자는 (i) 상기 제2 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제2 프로모터를 포함하는 것인 제2 합성 DNA 분자를 포함하는, 표적 단백질의 발현용 조성물이 제공된다.

[0011] 또한, 전술된 조성물을 포함하는, 표적 단백질의 발현용 시스템이 제공된다.

[0012] 또한, 세포에서 단백질을 발현하기 위해 본 명세서에서 개시되는 시스템 또는 상기 시스템에 의해 코딩된 RNA를 이용하는 방법이 제공된다. 그러한 방법은 상기 시스템을 세포 내로 도입시키는 단계, 및 동일한 세포에서 제1 합성 RNA 분자 및 제2 합성 RNA 분자를 발현시키는 단계를 포함할 수 있다. 일부 구체예에서, 상기 세포는 개체 내에 존재하고, 상기 방법은 상기 개체에서 상기 표적 단백질을 코딩하는 유전자 중 돌연변이에 의해 유발된 유전 질환을 치료한다. 일부 구체예에서, 상기 유전 질환은 듀센 근이영양증 (Duchenne muscular dystrophy), A형 혈우병(Hemophilia A), 스타르가르트병(Stargardt disease), 또는 어셔 증후군(Usher syndrome)이다.

[0013] 본 개시의 전술된 목적 및 특징과 기타 목적 및 특징이 첨부된 도면을 참조하여 기재된 하기 상세한 설명으로부터 보다 명확해질 것이다.

과제의 해결 수단

[0014] **상세한 설명**

[0015] 달리 기재되지 않으면, 기술적 용어는 통상적인 용법에 따라 사용된다. 분자 생물학에서 일반적인 용어의 정의는 Benjamin Lewin, *Genes VII*, published by Oxford University Press, 1999; Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, Blckwell Science Ltd. 출판, 1994; 및 Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, VCH Publishers, Inc. 출판, 1995; 및 기타 유사한 참고문헌에서 찾을 수 있다.

[0016] 본 명세서에서 사용된, 단수 형태("a," "an," 및 "the")는, 명확하게 달리 표시되지 않으면, 단수 및 복수 모두를 의미한다. 본 명세서에서 사용된, "포함하는 (comprises)"은 "함유하는 (includes)"을 의미한다. 따라서, "핵산 분자를 포함하는(comprising a nucleic acid molecule)"은 다른 요소들을 배제하지 않으면서, "핵산 분자를 포함하는(including a nucleic acid molecule)"을 의미한다. 또한, 핵산에 대해 주어진 모든 염기 크기는 대략적이고, 달리 표시되지 않으면, 설명적 목적으로 제공된다. 본 명세서에 기재된 것들과 유사하거나 또는 동등한 다수의 방법 및 재료가 사용될 수 있으나, 특정한 적합한 방법 및 재료가 하기에 기재된다. 상충의 경우에, 용어의 설명을 포함한, 본 명세서가 지배할 것이다. 또한, 재료, 방법, 및 실시예는 예시하는 것에 불과하고, 한정하는 것으로 의도되지 않는다. 특허 출원 및 특허, 및 GenBank 수탁 번호를 포함한 모든 참조문헌은 그 전체가 참조에 의해 본 명세서에 포함된다.

[0017] 본 개시의 다양한 구체예의 검토를 촉진하기 위해, 특정한 용어들의 하기 설명이 제공된다:

[0018] **투여(administration)**: 개체에게 본 명세서에서 제공된 치료 핵산 분자, 또는 기타 치료제와 같은 작용제(agent)를 임의의 효과적인 경로에 의해 제공하는 것이다. 전형적인 투여 경로는 주사 (예를 들면, 피하 주사, 근육내 주사, 피내 주사, 복막내 주사, 척수강내 주사, 종양내 주사, 골내 주사, 및 정맥내 주사), 경피, 비강내, 및 흡입 경로를 포함하나, 그에 한정되지 않는다. 투여는 전신적 또는 국소적일 수 있다.

[0019] **압타머(aptamer)**: 높은 친화도 및 특이성으로 특정한 표적 물질 또는 분자에 결합하는 핵산 분자(예를 들면, DNA 또는 RNA). 압타머는 개시된 핵산 분자에서 이합체화 도메인으로 사용될 수 있다. 일 구체예에서, 2개의 압타머가, 예를 들면, 표준 염기쌍 형성(basepairing), 비-규범적 염기쌍 상호작용, 또는 이들의 조합에 의해 상호 간에 결합하여 이합체화를 매개할 수 있다. 일 구체예에서, 압타머는 상기 압타머에 의해 인식되는 하나 이상의 표적의 존재시에만 RNA 이합체화(및 후속 재조합)를 가능하게 한다. 압타머는 SELEX(systematic evolution of ligands by exponential enrichment)로 불리는 조합 선택(combinatorial selection)를 통해 수득될 수 있다 (예를 들면, Ellington *et al.*, *Nature* 1990, 346, 818-822; Tuerk and Gold *Science* 1990, 249, 505-510; Liu *et al.*, *Chem. Rev.* 2009, 109, 1948-1998; Shamah *et al.*, *Acc. Chem. Res.* 2008, 41, 130-138; Famulok, *et al.*, *Chem. Rev.* 2007, 107, 3715-3743; Manimala *et al.*, *Recent Dev. Nucleic Acids Res.* 2004, 1, 207-231; Famulok *et al.*, *Acc. Chem. Res.* 2000, 33, 591-599; Hesselberth, *et al.*, *Rev. Mol.*

Biotech. 2000, 74, 15-25; Wilson *et al., Annu. Rev. Biochem*. 1999, 68, 611-647; Morris *et al., Proc. Natl. Acad. Sci. U.S.A*. 1998, 95, 2902-2907 참조). 그러한 과정에서, 목적 표적 분자에 결합할 수 있는 DNA 또는 RNA 분자가 선택, 증폭 및 돌연변이의 반복적 단계를 통해 10^{14}-10^{15} 개의 상이한 서열로 구성된 핵산 라이브러리로부터 선택된다. 압타머의 그들의 표적에 대한 친화도는 항체의 친화도에 필적할 수 있고, 해리 상수가 피코몰 범위까지 낮을 수 있다 (Morris *et al., Proc. Natl. Acad. Sci. U.S.A*. 1998, 95, 2902-2907; Green *et al., Biochemistry* 1996, *35*, 14413-14424).

[0020] 아데노신과 같은 소 유기분자로부터 트롬빈과 같은 단백질, 및 심지어 바이러스 및 세포에 이르는 다양한 범위의 표적에 특이적인 압타머가 밝혀졌다 (Liu *et al., Chem. Rev*. 2009, *109*, 1948-1998; Lee *et al., Nucleic Acids Res*. 2004, *32*, D95-D100; Navani and Li, *Curr. Opin. Chem. Biol*. 2006, *10*, 272-281; Song *et al., TrAC, Trends Anal. Chem*. 2008, *27*, 108-117). 예를 들면, Zn(II) (Ciesiolka *et al*., *RNA* 1: 538-550, 1995) 및 Ni(II) (Hofmann *et al*., *RNA*, 3:1289-1300, 1997)와 같은 금속 이온; 아데노신 트리포스페이트 (ATP) (Huizenga and Szostak, *Biochemistry*, 34:656-665, 1995); 및 구아닌 (Kiga *et al*., *Nucleic Acids Res*., 26:1755-60, 1998)과 같은 뉴클레오티드; NAD (Kiga *et al*., *Nucleic Acids Res*., 26:1755-60, 1998) 및 플라빈 (Lauhon and Szostak, *J. Am. Chem. Soc*., 117:1246-57, 1995)과 같은 보조-인자; 비오마이신 (viomycin) (Wallis *et al*., *Chem. Biol*. 4: 357-366, 1997) 및 스트렙토마이신 (Wallace and Schroeder, *RNA* 4:112-123, 1998)과 같은 항생제; HIV 역전사효소 (Chaloin *et al*., *Nucleic Acids Res*., 30:4001-8, 2002) 및 C형 간염 바이러스 RNA-의존성 RNA 폴리머라아제 (Biroccio *et al*., *J. Virol*. 76:3688-96, 2002)와 같은 단백질; 콜레라 전독소(cholera whole toxin) 및 스태필로콕코스 장내 독소 B(staphylococcal enterotoxin B) (Bruno and Kiel, *BioTechniques*, 32: pp. 178-180 and 182-183, 2002)와 같은 독소; 및 bacterial spores such as the anthrax (Bruno and Kiel, *Biosensors & Bioelectronics*, 14:457-464, 1999)와 같은 박테리아 포자를 인식하는 압타머가 이용가능하다.

[0021] **결합(binding)**: 2개의 물질 또는 분자간 회합(association), 예를 들면, 하나의 핵산 분자의 또 다른 핵산 분자 (또는 그 자신)로의 혼성화, 예를 들면, 2개의 이합체화 도메인간 혼성화, 또는 압타머의 그의 표적으로의 결합. 올리고뉴클레오티드 분자와 표적 핵산 간에 결합의 검출을 가능하게 할 정도로 충분한 개수의 상보적 염기쌍이 존재하는 경우, 상기 올리고뉴클레오티드 분자는 또 다른 핵산 분자에 결합하거나 또는 안정적으로 결합한다. 일부 구체예에서, 핵산 분자간 결합이 직접적으로 일어날 수 있다. 일부 구체예에서, 핵산 분자간 결합이 간접적으로, 예를 들면, 중간 분자(intermediate molecule)를 통해 일어날 수 있다. 직접 결합 또는 간접 결합은 표준 염기쌍 형성에 의해, 비-규범적 염기쌍 형성 상호작용에 의해, 비-염기쌍 형성 상호작용(non-base pairing interaction)에 의해, 또는 이들의 조합에 의해 일어날 수 있다. 비-규범적 염기쌍 형성 상호작용은 후그스틴(Hoogsteen) 염기쌍 및 워블(wobble) 염기쌍을 포함하나, 이에 한정되지 않는, 당업자에게 공지된 안정화의 임의의 수단에 의해 일어날 수 있다. 비-염기쌍 상호작용은 중간 분자를 통한 결합을 포함할 수 있다. 일부 구체예에서, 직접 결합은 키싱 루프 이합체화 도메인간에 일어난다. 일부 구체예에서, 직접 결합은 저다양성(hypodiverse) 이합체화 도메인간에 일어난다. 일부 구체예에서, 직접 결합은 압타머 영역간에 일어난다. 일부 구체예에서, 압타머 영역간 직접 결합은 비-규범적 염기쌍 상호작용을 포함한다. 일부 구체예에서, 압타머 영역간 직접 결합은 표준 염기쌍 형성 및 비-규범적 염기쌍 형성을 포함한다. 일부 구체예에서, 간접 결합은 핵산 브릿지(nucleic acid bridge)를 통해 일어난다. 일부 구체예에서 상기 핵산 브릿지는 mRNA이다. 핵산 브릿지의 비한정적 예가 도 7b에 도시된다. 일부 구체예에서, 간접 결합은 압타머 분자를 통해 일어난다. 압타머 분자를 통한 간접 결합의 비한정적 예가 도 7a에 도시된다. 일부 구현예에서, 압타머 분자를 통한 간접 결합은 압타머 분자와 결합 영역간 비-염기쌍 상호작용을 포함한다. 일부 구현예에서, 압타머 분자를 통한 간접 결합은 압타머 분자와 결합 영역간 비-염기쌍 상호작용, 및 결합 영역간 염기쌍 형성 상호작용을 포함한다.

[0022] **C-말단 부분(C-terminal portion)**: 단백질의 C-말단 잔기에서 또는 그 부근에서 시작되는 아미노산의 연속된 구간(stretch)을 포함하는 단백질 서열의 영역. 단백질의 C-말단 부분은 아미노산의 연속된 구간 (예를 들면, 다수의 아미노산 잔기)에 의해 정의될 수 있다.

[0023] **암(cancer)**: 비정상적 또는 비제어된 세포 증식을 특징으로 하는 악성 종양. 암과 종종 연관되는 기타 특징은 전이, 인접한 세포의 정상적 기능의 방해, 비정상적 수준의 사이토카인 또는 기타 분비성 산물의 분비, 및 염증성 또는 면역 반응의 억제 또는 악화, 주변 또는 원위 조직 또는 기관, 예를 들면, 림프절의 침윤 등을 포함한다. "전이성 질환(metastatic disease)"은 최초 종양 부위를 떠나 신체의 다른 부위로, 예를 들면, 혈류 또는 림프계를 통해, 이동한 암 세포를 의미한다.

[0024] **상보성(complementarity)**: 전통적인 Watson-Crick 염기쌍 형성 또는 기타 비-전통적 타입에 의해 또 다른 핵산 서열과 수소 결합을 형성하는 핵산의 능력. 퍼센트 상보성(percent complementarity)은 핵산 분자 중 제2 핵산 서열과 수소 결합 (예를 들면, Watson-Crick 염기쌍 형성)을 형성할 수 있는 잔기들의 비율을 나타낸다(예를 들면, 10개 중 5개, 6개, 7개, 8개, 9개, 10개는 각각 50%, 60%, 70%, 80%, 90%, 및 100% 상보성임). "완벽하게 상보적인(perfectly complementary)"은 핵산 서열의 모든 연속된 잔기들이 제2 핵산 서열 중 연속된 잔기들의 동일한 개수와 수소 결합을 형성할 것이라는 것을 의미한다. 본 명세서에서 사용된, "실질적으로 상보적인(substantially complementary)"은 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 35, 40, 45, 50개, 또는 그 이상의 뉴클레오티드의 영역에 대해 적어도 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, 또는 100%인 상보성의 정도를 의미하거나, 또는 엄격한(stringent) 조건 하에서 혼성화하는 2개의 핵산을 의미한다. 따라서, 일부 구체예에서, 제1 이합체화 도메인과 제2 이합체화 도메인은 상호 간에 완벽한 상보성(예를 들면, 100%)을 갖는다. 다른 구체예에서, 제1 이합체화 도메인과 제2 이합체화 도메인은 상호 간에 실질적 상보성(예를 들면, 적어도 80%)을 갖는다.

[0025] **접촉(contact)**: 고체 또는 액체 형태를 포함한, 직접적인 물리적 회합인 배치. 접촉은 인 비트로 또는 엑스 비보로, 예를 들면, 시약을 시료(예를 들면, 세포를 포함하는 시료)에 첨가하는 것에 의해, 또는 개체에게 투여하는 것에 의해 인 비보로 일어날 수 있다.

[0026] **하향조절 또는 넉다운된(downregulated or knocked down)**: 표적 핵산 또는 단백질과 같은 분자의 발현과 관련하여 사용되는 경우, 표적 RNA 또는 단백질의 생산에서 감소를 초래하나, 일부 구체예에서, 표적 RNA 산물 또는 표적 RNA 기능의 완전한 제거를 초래하지 않는 임의의 과정을 의미한다. 일 구체예에서, 하향조절 또는 넉다운은 검출가능한 표적 핵산/단백질 발현 또는 활성의 완전한 제거를 초래하지 않는다. 일부 구체예에서, 표적 핵산의 하향조절 또는 넉다운은 표적 RNA의 번역을 감소시키고 따라서 상응하는 단백질의 존재를 감소시킬 수 있는 과정을 포함한다. 본 명세서에서 개시되는 시스템은 목적 표적 핵산/단백질을 하향조절하기 위해 사용될 수 있다.

[0027] 하향조절 또는 넉다운은 표적 핵산/단백질의 임의의 검출가능한 감소를 포함한다. 특정한 구체예에서, 세포 또는 세포 불포함 시스템에서 검출가능한 표적 핵산/단백질은 대조군(상응하는 미처리 세포 또는 시료에서 검출되는 표적 핵산/단백질의 양) 대비 적어도 10%, 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 또는 적어도 99% (예를 들면, 40% 내지 90%, 40% 내지 80% 또는 50% 내지 95%) 감소한다. 일 구체예에서, 대조군은 정상 세포 (예를 들면, 본 명세서에서 제공되는 RNA 재조합을 위한 핵산 분자를 포함하지 않는 비-재조합 세포) 중 발현의 상대적 양이다.

[0028] **유효량(effective amount)**: 유용한 또는 원하는 결과를 달성하기에 충분한 작용제(agent) (예를 들면, 각각 치료 단백질, 예를 들면, 디스트로핀의 상이한 부분을 코딩하는, 복수 개의 벡터를 제공하는 시스템)의 양. 유효량은 또한 유용한 또는 원하는 결과를 달성하기에 충분한 생성된, 정확하게 연결된 RNA 또는 단백질의 양을 의미할 수 있다.

[0029] 유효량 (치료 유효량(therapeutically effective amount)으로도 지칭됨)은 치료대상 개체 및 질병 상태, 상기 개체의 체중 및 연령, 상기 질병 상태의 중증도, 투여 방식 등 중 하나 이상에 따라 변할 수 있고, 당업자에 의해 결정될 수 있다. 유용한 치료 효과는 진단적 결정의 가능화; 질병, 증상, 장애, 또는 병리적 질환의 개선; 질병, 증상, 장애, 또는 병리적 질환의 발병의 감소 또는 예방; 및 일반적으로, 질병, 증상, 장애, 또는 병리적 질환에 대한 반작용(counteract)을 포함할 수 있다.

[0030] 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 질병, 예를 들면, 유전 질환 또는 암을 치료하기에 충분하다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은(treated) 환자의 생존 기간을 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 적어도 500%, 또는 적어도 600%씩 증가시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 환자의 생존 기간을 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 6 개월, 적어도 9 개월, 적어도 1 년, 적어도 1.5 년, 적어도 2 년, 적어도 2.5 년, 적어도 3 년, 적어도 4 년, 적어도 5 년, 적어도 10 년, 적어도 12 년, 적어도 15 년, 또는 적어도 20 년씩 증가시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 핵산 분자의 "유효량"은 치료받

은 환자 (예를 들면, DMD 환자)의 이동성(mobility)을 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 적어도 500%, 또는 적어도 600% 증가시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 환자 (예를 들면, DMD 환자)의 이동성을 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 적어도 500%, 또는 적어도 600% 증가시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 환자 (예를 들면, DMD 환자)의 인지 능력(cognitive ability)을 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 적어도 500%, 또는 적어도 600% 증가시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 환자 (예를 들면, DMD 환자)의 호흡 기능을 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 적어도 500%, 또는 적어도 600% 증가시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 환자 (예를 들면, 혈우병 환자)의 혈액 응고를 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 적어도 500%, 또는 적어도 600% 증가시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 환자 (예를 들면, 어셔병 또는 스타르가르트병 환자)의 시력을 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 적어도 500%, 또는 적어도 600% 증가시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 환자 (예를 들면, 어셔병 환자)의 청력을 (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 예를 들면, 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 99%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 적어도 500%, 또는 적어도 600% 증가시키기에 충분한 양이다.

[0031] 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 DMD 환자의 종아리 근육 크기를 예를 들면, (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 또는 적어도 95% 감소시키기에 충분한 양이다. 일 구현예에서, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 "유효량"은 치료받은 DMD 환자의 심근병증(cardiomyopathy) 근육 크기를 예를 들면, (본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여 대비) 적어도 10%, 적어도 20%, 적어도 25%, 적어도 50%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 또는 적어도 95% 감소시키기에 충분한 양이다. 일부 구체예에서, 이러한 효과들의 조합이 달성된다.

[0032] **증가 또는 감소(increase or decrease)**: 대조값(control value)(예를 들면, 본 명세서에서 제공되는 2개 이상의 합성 핵산 분자의 미투여와 같은 치료제 불포함을 나타내는 값)으로부터 양의 통계적으로 유의한 양 또는 음(negative)의 변화. 증가는 양의 변화, 예를 들면, 대조값 대비 적어도 50%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400% 또는 적어도 500% 증가이다. 감소는 음의 변화, 예를 들면, 대조값 대비 적어도 20%, 적어도 25%, 적어도 50%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 적어도 100% 감소이다. 일부 구체예에서, 상기 감소는 100% 미만, 예를 들면, 90% 이하, 95% 이하, 또는 99% 이하의 감소이다.

[0033] **혼성화(hybridization)**: 핵산의 혼성화는 2개의 핵산 분자가 상당한 양의 상호간 수소 결합을 갖는 경우 일어난다. 혼성화의 엄격성(stringency)는 핵산 주위의 환경 조건, 혼성화 방법의 속성, 및 사용되는 핵산의 조성 및 길이에 따라 다를 수 있다. 특정한 정도의 엄격성을 달성하기 위해 요구되는 혼성화 조건에 대한 계산이 Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 2001); 및 Tijssen, *Laboratory Techniques in Biochemistry and Molecular Biology*—

Hybridization with Nucleic Acid Probes Part I, Chapter 2 (Elsevier, New York, 1993)에서 검토된다. T_m은 주어진 핵산 가닥의 50%가 그의 상보적 가닥에 혼성화되는 것인 온도이다.

[0034] **단리된(isolated):** "단리된(isolated)" 생물학적 성분(예를 들면, 핵산 분자 또는 단백질)은 상기 성분이 존재한 개체의 세포 또는 조직, 예를 들면, 기타 세포 (예를 들면, RBC), 염색체 및 염색체외 DNA 및 RNA, 및 단백질 중 다른 성분들로부터 실질적으로 분리되거나, 떨어져서 생산되거나 또는 정제된 것이다. "단리된" 핵산 및 단백질은 표준 정제 방법에 의해 정제된 핵산 및 단백질을 포함한다. 이 용어는 또한 숙주 세포에서 재조합 발현에 의해 제조된 핵산 및 단백질과 화학적으로 합성된 핵산 및 단백질을 포함한다.

[0035] **키싱 루프/키싱 스템 루프(kissing loop/kissing stem loop):** 2개의 헤어핀 루프(hairpin loop) 간 염기들이 쌍 상호작용(pair interaction)을 형성하는 경우의 형성되는 RNA 구조. 이러한 분자간 "키싱 상호작용(kissing interactions)"은 하나의 헤어핀 루프 중 쌍을 이루지 않은 뉴클레오티드가 또 따른 헤어핀 루프 중 쌍을 이루지 않은 뉴클레오티드와 염기쌍을 형성하여 안정한 상호작용 복합체를 형성하는 경우 일어난다. 예를 위해, 도 9a를 참조한다.

[0036] **N-말단 부분(N-terminal portion):** 단백질의 N-말단 잔기에서 시작되는 아미노산의 연속된 구간을 포함하는 단백질 서열의 영역. 단백질의 N-말단 부분은 아미노산의 연속된 구간 (예를 들면, 다수의 아미노산 잔기)에 의해 정의될 수 있다.

[0037] **비-천연, 합성, 또는 조작된(non-naturally occurring, synthetic, or engineered):** 본 명세서에서 호환적으로 사용되는 용어이고 사람에 의한 작업(hand of a human)의 개입을 나타낸다. 핵산 분자 또는 폴리펩티드를 지칭하는 경우, 이 용어들은 핵산 분자 또는 폴리펩티드가 그들이 자연적으로 회합되고 발견되는 적어도 하나의 다른 성분들이 실질적으로 없다는 것을 나타낸다. 또한, 이 용어들은 핵산 분자 또는 폴리펩티드가 자연에서 발견되지 않는 서열을 갖는다는 것을 나타낼 수 있다.

[0038] **핵산 분자(nucleic acid molecule):** 천연 뉴클레오티드/리보뉴클레오티드 및/또는 천연 뉴클레오티드와 유사한 방식으로 핵산 분자에 혼성화되는 천연 뉴클레오티드/리보뉴클레오티드의 유사체를 포함할 수 있는, 데옥시리보뉴클레오티드 폴리머(DNA) 또는 리보뉴클레오티드 폴리머(RNA). 핵산 분자는 단일가닥 (ss) DNA 또는 RNA 분자, 또는 이중가닥 (ds) 핵산 분자일 수 있다. 본 명세서에서 사용된 RNA 또는 mRNA는 프리-mRNA 분자 또는 성숙 RNA 전사물을 의미할 수 있다. 프리-mRNA 분자는 가공에 의해 제거되는 서열, 예를 들면, 본 명세서에 기재된 이합체화 도메인의 결합 후 스플라이싱에 의해 제거되는 인트론을 포함한다. 본 명세서에 기재된 핵산 분자는 DNA 상의 프로모터로부터 RNA가 전사되는 것인 DNA 분자, 예를 들면, DNA 발현 벡터의 상황(context) 중 DNA일 수 있다.

[0039] **작동가능하게 연결된(operably linked):** 제1 핵산 서열이 제2 핵산 서열과 기능적 관계로 배치된 경우, 제1 핵산 서열은 제2 핵산 서열에 작동가능하게 연결된다. 예를 들면, 프로모터가 핵산 서열의 발현에 영향을 미치는 경우, 예를 들면, 상기 프로모터가 스플라이싱되면 단백질 (예를 들면, DMD, 인자 8, 인자 9, 또는 ABCA4 코딩 서열의 일부)의 발현을 초래할 수 있는, 프리-mRNA의 전사에 영향을 미치는 경우, 프로모터 서열은 핵산 서열에 작동가능하게 연결된다.

[0040] **약학적으로 허용가능한 담체(pharmaceutically acceptable carriers):** 본 발명에서 유용한 약학적으로 허용가능한 담체는 통상적이다. *Remington's Pharmaceutical Sciences*, by E. W. Martin, Mack Publishing Co., Easton, PA, 15th Edition (1975)은 치료제, 예를 들면, 본 명세서에 개시된 핵산 분자의 약학적 전달을 위해 적합한 조성물 및 제형을 기술한다.

[0041] 일반적으로, 담체의 속성은 채택되는 특정한 투여 모드에 따라 결정될 것이다. 예를 들면, 비경구 제형은 일반적으로 약학적 및 생리적으로 허용가능한 유체, 예를 들면, 물, 생리적 식염수, 평형 염 용액(balanced salt solutions), 덱스트로오스 수용액, 글리세롤, 등을 비히클로 포함하는 주사액을 포함한다. 생물학적-중성 담체(biologically-neutral carriers) 외에, 투여될 약학적 조성물은 미량의 무독성 보조 물질, 예를 들면, 습윤제 또는 유화제, 보존제, 및 pH 완충제 등, 예를 들면, 소듐 아세테이트 또는 소르비탄 모노라우레이트를 포함할 수 있다.

[0042] **폴리펩티드, 펩티드, 및 단백질(polypeptide, peptide and protein):** 임의의 길이의 아미노산의 폴리머를 의미한다. 폴리머는 선형 또는 분지형일 수 있고, 변형된 아미노산을 포함할 수 있고, 비-아미노산(non-amino acid)에 의해 개입될 수 있다(interrupted). 이 용어들은 또한 변형된, 예를 들면, 디술피드 결합 형성, 글리코실화, 지질화(lipidation), 아세틸화, 인산화, 또는 기타 조작, 예를 들면, 표지화(labeling) 성분과의 접합에 의

해 변형된 아미노산 폴리머를 포함한다. 본 명세서에서 사용된, 용어 "아미노산"은, 글리신 및 D- 또는 L- 광학 이성질체, 및 아미노산 유사체 및 펩티도미메틱(peptidomimetic)을 포함한, 천연 및/또는 비천연, 또는 합성 아미노산을 포함한다. 일 구체예에서, 단백질은 질병, 예를 들면, 유전 질환과 연관된 단백질이다(예를 들면, 표 1 참조). 일 구체예에서, 단백질은 치료 단백질, 예를 들면, 암과 같은 질환의 치료에서 사용되는 단백질이다. 일 구체예에서, 단백질은 길이가 적어도 50 aa, 적어도 100 aa, 적어도 500 aa, 적어도 1000 aa, 적어도 1500 aa, 예를 들면, 적어도 2000 aa, 적어도 2500 aa, 적어도 3000 aa, 또는 적어도 5000 aa이다.

[0043] **폴리피리미딘 트랙트(polypyrimidine tract)**: 전사-후 변형의 과정 동안 RNA 스플라이싱을 수행하도록 특화된 단백질 복합체인 스플라이세오솜(splicesome)의 조립을 촉진하는 프리-mRNA(pre-messenger RNA)의 영역. 이 트랙트는 주로 우라실과 같은 피리미딘 뉴클레오티드일 수 있고, 일부 구체예에서, 스플라이싱되는 인트론의 3' 말단 전 약 5-40 bp에 위치한, 15-20 bp 길이이다.

[0044] **프로모터/인핸서(promoter/enhancer)**: 핵산 서열의 전사를 지시하는 핵산 조절 서열의 어레이. 프로모터는 전사의 개시 부위에 인접한 필요한 핵산 서열을 포함하고, 예를 들면, 폴리머라아제 II 타입 프로모터의 경우, TATA 요소를 포함한다. 프로모터는 또한 선택적으로 전사의 개시 부위로부터 수천 bp에 위치할 수 있는 원위(distal) 인핸서 또는 리프레서(repressor) 요소를 포함한다. 일부 구체예에서 프로모터 서열 + 그의 상응하는 코딩 서열은 AAV의 수용력(capacity)보다 더 크다. 일부 구체예에서, 표적 단백질의 프로모터 서열은 적어도 3500 nt, 적어도 4000 nt, 적어도 5000 nt, 또는 심지어 적어도 6000 nt이다.

[0045] "항시적 프로모터(constitutive promoter)"는 지속적으로 활성이고, 외래 신호 또는 분자에 의한 조절을 받지 않는 프로모터이다. 대조적으로, "유도성 프로모터(inducible promoter)"의 활성은 외래 신호 또는 분자 (예를 들면, 전사 인자)에 의해 조절된다. 항시적 프로모터 및 유도성 프로모터가 모두 본 명세서에서 제공되는 방법 및 시스템에서 이용될 수 있다 (예를 들면, Bitter *et al.*, *Methods in Enzymology* 153:516-544, 1987 참조). 조직-특이적 프로모터가, 예를 들면, 원하는 조직 또는 목적 세포, 예를 들면, 근육, 뉴런, 뼈, 피부, 혈액, 특정한 기관 (예를 들면, 간, 췌장), 또는 특정한 세포 종류 (예를 들면, 림프구)에서 주로 발현을 지시하기 위해, 본 명세서에서 제공되는 방법 및 시스템에서 이용될 수 있다. 일부 구체예에서, 본 명세서에서 사용되는 프로모터는 발현되는 표적 단백질에 내생적이다. 일부 구체예에서, 본 명세서에서 사용되는 프로모터는 발현되는 표적 단백질에 외생적이다

[0046] 또한, 프로모터-의존적 유전자 발현이 세포-타입 특이적, 조직-특이적으로 조절가능하게 하거나, 또는 외부 신호 또는 물질(agent)에 의해 유도될 수 있게 하기에 충분한 프로모터 요소가 포함된다; 그러한 요소들은 유전자의 5' 또는 3' 영역에 위치할 수 있다. 재조합 DNA 또는 합성 기법에 의해 생성된 프로모터도 핵산 서열의 전사를 위해 이용될 수 있다.

[0047] 본 명세서에서 제공되는 방법 및 시스템에서 이용될 수 있는 예시적 프로모터는 SV40 프로모터, CMV(cytomegalovirus) 프로모터 (선택적으로, CMV 인핸서를 가짐), pol III 프로모터 (예를 들면, U6 및 H1 프로모터), pol II 프로모터 (예를 들면, RSV(retroviral Rous sarcoma virus) LTR 프로모터 (선택적으로, RSV 인핸서를 가짐), DHFR(dihydrofolate reductase 프로모터), β-액틴 프로모터, PGK(phosphoglycerol kinase) 프로모터, 및 EF1α 프로모터를 포함하나, 이에 한정되지 않는다.

[0048] **재조합체(recombinant)**: 재조합체 핵산 분자 또는 단백질 서열은 자연발생적이 아닌 서열을 갖거나 또는 2개의 달리(otherwise) 분리된 서열의 세그먼트의 인위적 조합에 의해 제조된 서열을 갖는 것이다(예를 들면, 디스트로핀 코딩 서열의 일부, 예를 들면, 코딩 서열의 1/3, 또는 2/3를 포함하는 바이러스 벡터). 이러한 인위적 재조합은 예를 들면, 화학적 합성, 또는 핵산의 단리된 세그먼트의 인위적 조작, 예를 들면, 유전 공학 기법에 의해 달성될 수 있다. 유사하게, 재조합 세포 또는 형질전환(transgenic) 세포는 재조합 핵산 분자를 포함하는 세포이다.

[0049] **서열 동일성(sequence identity)**: 아미노산 (또는 뉴클레오티드) 서열간 유사성은, 달리 서열 동일성으로도 지칭되는, 서열간 유사성의 측면에서 표현된다. 서열 동일성은 종종 퍼센트 동일성 (또는 유사성 또는 상동성)의 측면에서 측정된다; 그 퍼센트가 높을수록, 2개의 서열은 더 유사하다.

[0050] 비교를 위해 서열을 정렬하는 방법이 알려져 있다. 다양한 프로그램 및 정렬 알고리즘이 하기에 기재된다: Smith and Waterman, *Adv. Appl. Math.* 2:482, 1981; Needleman and Wunsch, *J. Mol. Biol.* 48:443, 1970; Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85:2444, 1988; Higgins and Sharp, *Gene* 73:237, 1988; Higgins and Sharp, *CABIOS* 5:151, 1989; Corpet *et al.*, *Nucleic Acids Research* 16:10881, 1988; 및

Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85:2444, 1988. Altschul *et al.*, *Nature Genet.* 6:119, 1994는 서열 정렬 방법 및 상동성 계산의 상세한 고려사항을 제시한다.

[0051] NCBI Basic Local Alignment Search Tool (BLAST) (Altschul *et al.*, *J. Mol. Biol.* 215:403, 1990)이 서열 분석 프로그램 blastp, blastn, blastx, tblastn 및 tblastx과 관련하여 사용하기 위해 National Center for Biotechnology Information (NCBI, Bethesda, MD)을 포함한 여러 출처 및 인터넷에서 이용가능하다. 이 프로그램을 이용하여 서열 동일성을 결정하는 방법의 설명이 인터넷 상에서 NCBI 웹사이트에서 이용가능하다.

[0052] 원형의(native) 단백질 또는 코딩 서열 (예를 들면, DMD, 인자 8, 인자 9, 또는 ABCA4 서열)의 변이체는 일반적으로 디폴트 파라미터로 설정된 NCBI Blast 2.0, gapped blastp를 이용하여 아미노산 서열과의 전장 정렬에 대해 카운트된 적어도 약 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 서열 동일성의 소유를 특징으로 한다. 약 30개 아미노산보다 더 긴 아미노산 서열의 비교를 위해, 디폴트 파라미터(11의 gap existence cost, 및 1의 per residue gap cost)로 설정된 디폴트 BLOSUM62 매트릭스를 이용한 Blast 2 서열 기능이 채택된다. 짧은 펩티드 (약 30개 미만의 아미노산)를 정렬시키는 경우, 정렬은 디폴트 파라미터(open gap 9, extension gap 1 페널티)로 설정된 PAM30 매트릭스를 채택하는 Blast 2 서열 기능을 이용하여 수행되어야 한다. 기준 서열에 대한 더 큰 유사성을 갖는 단백질은 이 방법에 의해 평가된 증가하는 퍼센트 동일성, 예를 들면, 적어도 95%, 적어도 98%, 또는 적어도 99% 서열 동일성을 보일 것이다. 전체 서열보다 짧은 서열이 서열 동일성을 위해 비교되는 경우, 동족체(homolog) 및 변이체는 일반적으로 10-20개 아미노산의 짧은 창(window)에 대해 적어도 80% 서열 동일성을 갖고, 기준 서열에 대한 유사성에 따라, 적어도 85% 또는 적어도 90% 또는 적어도 95%의 서열 동일성을 가질 수 있다. 그러한 짧은 창에 대해 서열 동일성을 결정하는 방법이 인터넷의 NCBI 웹사이트에서 이용가능하다. 이러한 서열 동일성 범위는 단지 지침(guidance)을 위해서 제공된다; 제공된 범위에서 벗어난 강력하게 유의한 동족체를 수득하는 것이 가능하다.

[0053] 개시된 핵산 서열의 변이체 (예를 들면, 합성 인트론 서열 및 코딩 서열)는 일반적으로, 디폴트 파라미터로 설정된 gapped blastn인 NCBI Blast 2.0을 이용하여 핵산 서열과의 전장 정렬에 대해 카운트된 적어도 약 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 서열 동일성을 특징으로 한다. 당업자는 이러한 서열 동일성 범위가 단지 지침을 위해서 제공되고; 제공된 범위에서 벗어난 기능성 서열을 수득하는 것이 가능하다는 것을 이해할 것이다.

[0054] **개체(subject)**: 포유동물, 예를 들면, 인간. 포유동물은 마우스, 유인원, 인간, 가축(farm animal), 스포츠 동물(sport animal), 및 펫을 포함하나, 이에 한정되지 않는다. 일 구현예에서, 상기 개체는 비-인간 포유동물(non-human mammalian) 개체, 예를 들면, 원숭이, 또는 기타 비-인간 영장류, 마우스, 랫트, 토끼, 돼지, 염소, 양, 돌고래, 개, 고양이, 말, 또는 소이다. 일부 구체예에서, 상기 개체는 실험용(laboratory) 동물/개체, 예를 들면, 마우스, 토끼, 또는 랫트이다. 일부 구체예에서, 본 명세서에서 개시된 방법을 이용하여 치료되는 개체는 인간이다.

[0055] 일부 구체예에서, 상기 개체는 본 명세서에서 개시된 방법을 이용하여 치료될 수 있는 유전 질환, 예를 들면, 표 1에 열거된 유전 질환을 갖는다. 일부 구체예에서, 본 명세서에서 개시된 방법을 이용하여 치료되는 개체는 유전 질환을 갖는 인간 개체이다. 일부 구체예에서, 본 명세서에서 개시된 방법을 이용하여 치료되는 개체는 암을 갖는 인간 개체이다.

[0056] **치료제(therapeutic agent)**: 개체로의 투여시 어떤 유용한 효과를 부여하는 하나 이상의 분자 또는 화합물을 의미한다. 본 명세서에서 제공된 개시되는 합성 핵산 분자 및 시스템은 치료제이다. 유용한 치료 효과는 진단적 결정의 가능화; 질병, 증상, 장애, 또는 병리적 질환의 개선; 질병, 증상, 장애, 또는 병리적 질환의 발병의 감소 또는 예방; 및 일반적으로, 질병, 증상, 장애, 또는 병리적 질환에 대한 반작용을 포함할 수 있다.

[0057] **형질도입된, 형질전환된, 및 형질감염된(transduced, transformed and transfected)**: 바이러스 또는 벡터가 세포 내로 핵산 분자를 전달할 때, 바이러스 또는 벡터는 세포를 "형질도입(transduce)" 시킨다. 핵산 분자가 세포 게놈으로의 결합에 의해 또는 에피좀 발현에 의해 세포에 의해 안정적으로 발현되는 경우, 세포는 세포 내로 형질도입된 핵산에 의해 "형질전환(transforrmed)"되거나 또는 "형질감염(transfected)"된다.

[0058] 이러한 용어들은 바이러스 벡터에 의한 형질감염, 플라스미드 벡터에 의한 형질전환, 및 전기천공, 리포펙션(lipofection), 입자 건 가속(particle gun acceleration) 및 당해 분야의 기타 방법에 의한 노출(naked) DNA의 도입을 포함한, 핵산 분자가 그러한 세포 내로 도입될 수 있게 하는 모든 기법을 포함한다. 일부 구체예에서, 상기 방법은 화학적 방법 (예를 들면, 인산칼슘 형질감염), 물리적 방법 (예를 들면, 전기천공, 미

세주입(microinjection), 입자 충격(particle bombardment), 융합 (예를 들면, liposomes), 수용체-매개 내포(endocytosis) (예를 들면, DNA-단백질 복합체, 바이러스 외피/캡시드 단백질 복합체(viral envelope/capsid-DNA complexes)) 및 재조합 바이러스와 같은 바이러스에 의한 생물학적 감염 (Wolff, J. A., ed, *Gene Therapeutics*, Birkhauser, Boston, USA, 1994)이다. 핵산 분자를 세포 내로 도입하는 방법들이 알려져 있다 (예를 들면, 미국특허 제6,110,743호 참조). 이러한 방법들은 세포를 개시된 핵산 분자로 형질도입시키기 위해 이용될 수 있다.

[0059] **전이 유전자(transgene):** 외생 유전자, 예를 들면, AAV와 같은 벡터에 의해 공급되는 외생 유전자. 일 구체예에서, 전이 유전자는 표적 유전자의 일부, 예를 들면, 프로모터 서열에 작동가능하게 연결된, 표적 단백질의 약 1/3, 1/2, 또는 2/3를 코딩한다. 일 구체예에서, 전이 유전자는 디스트로핀 코딩 서열의 일부, 예를 들면, 프로모터 서열에 작동가능하게 연결된, 디스트로핀 코딩 서열(또는 기타 치료제 코딩 서열, 예를 들면, 표 1에 열거된 단백질을 코딩하는 서열)의 약 1/3, 1/2, 또는 2/3를 포함한다.

[0060] **치료하는, 치료, 및 치료법(treating, treatment, and therapy):** 증상의 완화, 경감, 감소 또는 증상을 환자가 견딜수 있게 하기, 퇴화 속도의 둔화 또는 약화, 퇴화의 최후(final point)에 덜 약화되게 하기, 개체의 신체적 또는 정신적 웰빙의 개선, 또는 생존의 기간을 연장하기를 포함한, 손상, 질환 또는 질병의 경감 또는 개선의 성공 또는 성공의 징후. 치료는 신체 검사, 혈액 및 기타 임상적 테스트의 결과 등을 포함한, 객관적 또는 주관적 파라미터에 의해 평가될 수 있다. 일부 구체예에서, 본 명세서에서 개시되는 방법에 의한 치료는 유전 질환과 연관된 증상의 갯수 또는 중증도의 감소, 예를 들면, 유전 질환을 갖는 치료 환자의 생존 기간의 증가를 가져온다.

[0061] 일부 구체예에서, 본 명세서에 개시된 방법에 의한 치료는 DMD 또는 기타 유전 질환과 연관된 증상의 갯수 또는 중증도의 감소, 예를 들면, 생존율 증가, 이동성(mobility)(예를 들면, 걷기, 등반)의 증가, 인지 능력의 개선, 종아리 근육 크기의 감소, 심근병증(cardiomyopathy)의 경감, 시력 개선, 청력 개선, 혈액 응고 개선, 또는 호흡 기능의 개선을 가져온다. 일부 구체예에서, 이러한 효과들의 조합이 달성된다.

[0062] **종양, 신생물, 악성종양 또는 암(tumor, neoplasia, malignancy or cancer):** 신생물은 과도한 세포 분열로부터 유발된 조직 또는 세포의 비정상적 증식이다. 신생물성 증식은 종양을 생성한다. 개체에서 종양의 양이 종양의 갯수, 부피 또는 중량에 의해 측정될 수 있는 "**종양 부하(tumor burden)**"이다. 전이되지 않는 종양은 "**양성(benign)**"으로 지칭된다. 주변 조직으로 침윤하고 및/또는 전이될 수 있는 종양은 "**악성(malignant)**"으로 지칭된다. "**비-암성 조직(non-cancerous tissue)**"은 악성 신생물이 형성된 것인 동일한 기관으로부터 유래되나, 신생물의 특징적인 병리를 갖지 않는 조직이다. 일반적으로, 비-암성 조직은 조직학적으로 정상으로 보인다. "**정상 조직(normal tissue)**"은 암 또는 또다른 질환 또는 질병에 의해 영향을 받지 않은 기관으로부터 유래된 조직이다. "암-불포함(cancer-free)" 개체는 해당 기관의 암으로 진단받은 적이 없고 검출가능한 암을 갖지 않는다.

[0063] 본 명세서에서 개시되는 방법 및 시스템에 의해 치료될 수 있는 예시적 종양, 예를 들면, 암은 고형 종양, 예를 들면, 유방암종(예를 들면, 소엽 및 유관 암종), 육종, 폐의 암종 (예를 들면, 비소세포폐암, 대세포암, 편평세포 암종(squamous carcinoma), 및 선암종), 폐의 중피종, 대장 선암종, 위암종, 전립선암종(prostatic adenocarcinoma), 난소암종 (예를 들면, 장액 낭선암종(serous cystadenocarcinoma) 및 점액성 낭선암종), 난소 배아세포종(ovarian germ cell tumors), 고환암종 및 배아세포종, 췌장 선암종, 담관 선암종(biliary adenocarcinoma), 간세포 암종(hepatocellular carcinoma), 방광 암종(예를 들면, 이행세포 암종(transitional cell carcinoma), 선암종, 및 편평세포 암종 포함), 신세포 선암종(renal cell adenocarcinoma), 자궁내막 암종(예를 들면, 선암종 및 혼합 뮬러관 종양(mixed Mullerian tumors) (암육종) 포함), 자궁경관점막, 외자궁경부, 및 질의 암종 (예를 들면, 각각의 선암종 및 편평세포 암종), 피부의 종양 (예를 들면, 편평세포 암종, 기저세포 암종, 악성 흑색종, 피부 부속기 종양(skin appendage tumors), 카포시 육종, 피부 림프종, 피부 부속기 종양(skin adnexal tumors) 및 다양한 종류의 육종 및 메르겔 세포 암종), 식도 암종, 비인두 및 구강인두의 암종(이들의 편평세포 암종 및 선암종 포함), 침샘 암종, 뇌 및 중추신경계 종양 (예를 들면, 아교, 신경원, 및 수막 기원의 종양(tumors of glial, neuronal, and meningeal origin) 포함), 말초신경의 종양, 연조직 육종, 및 골 및 연골의 육종, 및 림프 종양 (B-세포 및 T-세포 악성 림프종 포함)을 포함한다. 일 구체예에서, 상기 종양은 선암종이다.

[0064] 상기 방법 및 시스템은 또한 림프 종양, 백혈구 종양, 또는 기타 종류의 백혈병과 같은 액체 종양을 치료하기 위해 이용될 수 있다. 특정한 구체예에서, 치료되는 종양은 혈액의 종양, 예를 들면, 백혈병 (예를 들면, 급성

림프구성 백혈병 (ALL), 만성 림프구성 백혈병 (CLL), 급성 골수성 백혈병 (AML), 만성 골수성 백혈병 (CML), 유모 세포 백혈병 (HCL), T-세포 전림프구성 백혈병 (T-PLL), 거대 과립 림프구성 백혈병(large granular lymphocytic leukemia), 및 성인 T-세포 백혈병 (adult T-cell leukemia)), 림프종 (예를 들면, 호지킨 림프종 및 비-호지킨 림프종), 및 골수종이다.

[0065] **상향조절된(upregulated):** 표적 핵산/단백질과 같은 분자의 발현과 관련하여 사용되는 경우, 상기 표적 핵산/단백질의 생산의 증가를 가져오는 임의의 과정을 의미한다. 일부 구체예에서, 표적 RNA의 상향조절 또는 활성화는 표적 RNA의 번역을 증가시키고, 따라서 상응하는 단백질의 존재를 증가시킬 수 있는 과정을 포함한다.

[0066] 상향조절은 표적 핵산/단백질의 검출가능한 증가를 포함한다. 일부 구체예에서, 세포 또는 무세포(cell free) 시스템에서 검출가능한 표적 핵산/단백질 발현은 대조군(본 명세서에서 제공되는 핵산으로 처리되지 않은 상응하는 시료에서 검출되는 표적 핵산/단백질의 양) 대비 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 100%, 적어도 200%, 적어도 400%, 또는 적어도 500% 증가한다. 일 구체예에서, 대조군은 정상 세포(예를 들면, 본 명세서에서 제공되는 시스템을 포함하지 않는, 비-재조합 세포)에서의 발현의 상대적 양이다.

[0067] **충분한 조건 하에서(under conditions sufficient for):** 원하는 활성을 가능하게 하는 임의의 환경을 기술하기 위해 사용되는 구절. 일 구체예에서, 원하는 활성은 질병을 치료하기 위해 필요한 단백질의 증가된 발현 또는 활성이다. 일 구체예에서, 원하는 활성은 인 비보에서, 예를 들면, 본 명세서에서 개시된 방법 및 시스템을 이용하여, DMD (또는 표 1에 열거된 기타 유전 질환)와 같은 유전 질환의 치료 또는 진행의 둔화이다.

[0068] **벡터(vector):** 숙주 세포에서 벡터가 복제 및/또는 통합하는 능력을 방해하지 않으면서, 외래 핵산 분자가 도입될 수 있는 핵산 분자. 벡터는 단일가닥, 이중가닥, 또는 부분적 이중가닥인 핵산 분자; 하나 이상의 유리 말단(free end)을 갖거나 또는 유리 말단을 갖지 않는(예를 들면, 고리형(circular)) 핵산 분자; DNA, RNA, 또는 둘 모두를 포함하는 핵산 분자; 및 폴리뉴클레오티드의 기타 종류를 포함하나, 그에 한정되지 않는다.

[0069] 벡터는 숙주 세포에서 복제될 수 있게 하는 핵산 서열, 예를 들면, 복제 원점을 포함할 수 있다. 벡터는 또한 하나 이상의 선택 마커 유전자 및 기타 유전 요소를 포함할 수 있다. 통합 벡터(integrating vector)는 그 자신을 숙주 핵산 내로 통합시킬 수 있다. 발현 벡터는 삽입된 유전자 또는 유전자들의 전사 및 번역을 가능하게 하는 필요한 조절 서열을 포함하는 벡터이다.

[0070] 벡터의 한 종류는 "플라스미드(plasmid)"이고, 플라스미드는, 예를 들면, 표준 분자 클로닝 기법에 의해, 추가적인 DNA 세그먼트가 삽입될 수 있는 고리형 이중가닥 DNA 루프를 의미한다. 또 다른 종류의 벡터는 바이러스 유래 DNA 또는 RNA 서열이 바이러스로의 패키징을 위해 벡터에 존재하는 것인 바이러스 벡터이다 (예를 들면, 레트로바이러스, 복제 결함(replication defective) 레트로바이러스, 아데노바이러스, 복제 결함 아데노바이러스, 및 아데노-연관 바이러스(AAV)). 바이러스 벡터는 또한 숙주 세포 내로의 형질감염을 위해 바이러스에 의해 운반되는 폴리뉴클레오티드를 포함한다. 일부 구현예에서, 상기 벡터는 렌티바이러스 (예를 들면, 통합-결함(integration-deficient) 렌티바이러스 벡터) 또는 아데노-연관 바이러스 (AAV) 벡터이다.

[0071] 일부 구현예에서, 상기 벡터는 AAV, 예를 들면, AAV 혈청형 AAV9 또는 AAVrh.10이다. 일부 구현예에서, 상기 벡터는 예를 들면, 정맥내 투여 후에, 혈뇌 장벽을 통과할 수 있는 벡터이다. AAVrh.10(adeno-associated virus serotype rh.10) 벡터는 부분적으로 혈뇌 장벽을 통과하여, 높은 수준의 전이 유전자 발현 및 전파를 제공한다.

[0072] **II. 여러 구현예의 개요**

[0073] 유전 질환을 앓는 환자를 치료하기 위한 하나의 접근방식은 유전자 대체 요법(일반적으로 유전자 요법으로 지칭됨)이다. 그러한 접근방식에서, 결함 유전자가 예를 들면, 바이러스 벡터를 통해 전달된, 그의 온전한 버전에 의해 대체되어, 수개월 내지 수년까지 지속되는 발현을 달성한다. 아데노 연관 바이러스 (AAV)가 임상 유전자 대체 요법을 위해 사용되고 있으나, 그들은 제한된 패키징 능력 (예를 들면, 약 5 kb 미만)을 갖는다. 따라서, 약 5 kb 크기 한계를 초과하는 유전자의 유전자 대체를 달성하기 위해 이러한 패키징 한계를 극복하는 전략이 요구된다. 예를 들면, 일부 프로모터 단독, 코딩 서열 단독, 또는 합쳐진 프로모터 + 코딩 서열이 AAV의 약 5 kb 크기 한계를 초과한다. 따라서, 그러한 프로모터 및 코딩 서열에 의해 코딩되는 단백질이 본 명세서에서 개시되는 시스템을 이용하여 발현될 수 있다.

[0074] AAV의 카고 한계를 극복하기 위한 이전의 방법들은 질병을 치료하기 위해 충분한 개수의 세포에서 적절한 수준의 표적 단백질을 생산하기 위해 요구되는 효율을 달성하는 것으로 보이지 않는다. 예를 들면, 디스트로핀은 약

11kb이므로, AAV 패키징 한계와 양립되기 위해 최소 3개의 단편으로 전달되어야 한다.

[0075] RNA 단편 중 하나 또는 둘 모두에 대해 자연 발생적 인트론 서열을 이용한 2개의 RNA 분자의 스플라이싱 매개 재조합은 비효율적이다. 첫째, 이러한 천연 인트론 서열은 자연발생적 인트론으로부터 유래되고 모든 4개의 RNA 뉴클레오티드의 혼합으로 구성된다. 그러한 서열들은 분자간 상호작용을 위해 이용가능하기 보다, 강력한 분자내 염기쌍을 형성하는 것에 의해 트랜스-상호작용을 방해할 수 있는 구조로 폴딩되는 경향이 있다. 둘째, 고등 진핵생물에서 인트론이 아닌 엑손이 엑손 정의를 주도하므로, 이러한 자연발생적 인트론 서열은 스플라이세오솜 성분들을 강력하게 이끌도록 진화하지 않았다. 이전 전략들의 이러한 2가지 한계가 자연적으로 발견되지 않는 합성 인트론 서열을 설계하는 것에 의해 본 발명에서 해소된다. 이러한 합성 서열은 한편으로는 스플라이세오솜을 강력하고 끌어당기고 스플라이세오솜 동원을 촉진하면서, 2개의 RNA 단편들이 모이는 것을 방해하는 이차 구조(및 일부 구체예에서, 삼차 구조와 같은 기타 구조)를 최소화하는 요소들을 포함한다.

[0076] 본 발명자들은 다수의 일련의 단편들로부터 큰 유전자의 코딩 서열을 효율적으로 재구성하기 위해 이용될 수 있는 신규한 핵산 기반 요소를 개발했다. 본 명세서에 개시된 방법 및 시스템은 이전의 방법들과 다르다. 본 명세서에서 개시되는 고효율 합성 인트론은 비-공유결합에 의해 연결된 RNA (프리-mRNA) 간 RNA 스플라이싱 반응을 효율적으로 이끄는 RNA 요소들 (또는 이러한 요소들을 코딩하는 DNA)의 최적 배열을 이용한다. 상기 방법/시스템은 유전 질환을 치료하기 위한 단백질의 치료 수준에 보다 더 근접하는 높은 수준의 기능성 단백질을 생성하기 때문에 트랜스-스플라이싱을 이용하는 이전의 시도 대비 중대한 발전이다. 혁신은 내재적으로, 상보적 가닥을 갖는 (또한 내재적으로 낮은 시스-결합력(cis-binding capacity)을 갖는) 제2 RNA와의 트랜스-상호작용을 방해하는 강력한 시스-결합 상호작용을 형성할 수 없는 비-천연 RNA 도메인을 선택하는 것에 기반한다. 이러한 최적화된 이합체화 도메인 및/또는 합성 인트론은 RNA 스플라이싱을 촉진하는 최적화된 모티프(스플라이스 도너, 스플라이스 억셉터, 스플라이스 인핸서, 및 스플라이스 분지점 서열 포함)와 조합하여 사용되는 비-천연 서열(예를 들면, 인간 세포에서 발견되지 않는 서열 및/또는 또 다른 생물계에서 발견되지 않는 서열)을 포함한다. 합성 핵산은 비-천연 핵산 서열, 예를 들면, 인간 세포에서 발견되지 않고 및/또는 또 다른 생물학적 시스템에서 발견되지 않는 서열일 수 있다. 효율적인 스플라이싱을 매개하는 적합한 RNA 모티프의 상황에서 RNA 가닥의 트랜스-이합체화를 최적화하는 것에 의해, 2개 또는 3개의 상이한 RNA가 동일한 세포에서 정확하게 및 효율적으로 공유결합에 의해 연결되어, 인 비보 및 인 비트로에서 높은 수준의 기능성 단백질을 생산할 수 있다는 것이 최초로 본 발명에서 입증된다. 궁극적으로 성숙 전사물로부터 DNA 재조합 부위를 제거하기 위한 시스 RNA 스플라이싱으로 이어지는, DNA 재조합을 통한 DNA 수준에서의 비효율적 조합을 제공하는 "하이브리드(hybrid)" 접근 방식과 달리, 본 명세서에서 개시되는 방법/시스템은 비-기능성 및/또는 유해한 산물들을 코딩하는 재조합 산물을 생성할 더 낮은 위험으로, 2개의 단백질 코딩 RNA 단편들이 프리-mRNA 수준에서 함께 연결되는 것인 보다 효율적인 반응을 촉진한다.

[0077] 데이터는 효율적인 합성 RNA-이합체화 및 재조합 도메인(sRdR 도메인, RNA 말단-연결 (REJ) 도메인으로도 지칭됨)을 이용하는 것에 의해, 목적 유전자가 동일한 세포에서 발현된 2개 또는 3개의 별개의 유전자 단편으로부터 효율적으로 재구성될 수 있다는 것을 보여준다. 이러한 결과는 본 명세서에서 개시되는 방법 및 시스템이 각각 듀센 근이영양증, A형 혈우병, 또는 스타르가르트병을 치료하기 위해, AAV를 이용하여 디스트로핀 또는 혈액 응고 인자 VIII, 또는 Abca4 (ATP binding cassette subfamily A member 4)와 같은 큰 유전자를 재구성하는 능력을 보여준다. 이러한 관찰에 근거하여, 큰 단백질의 발현으로부터 유익을 얻는 질환과 같은 기타 유전 질환(예를 들면, 표 1에 열거된 질환들 참조)이 유사하게 치료될 수 있다. 기타 적용은 연구 및 생물공학 적용을 포함한다.

[0078] 다수의 AAV로부터 단편화된 유전자의 재구성을 위한 기존 전략들의 일부 한계를 해소하기 위해, 표적 세포에서 2개 이상의 개별적인 합성 RNA 분자들을 연속적으로 정렬시키고 재조합하는 시스템이 본 명세서에서 제공된다. 각각의 개별적인 합성 RNA 분자는 RNA 스플라이싱을 위해 필요한 이합체화 도메인 및 요소들을 포함한, 합성 뉴트론 서열을 포함하고, 이는 이합체화 도메인이 상호 간에 정확한 순서로 결합시, 개별적인 단편들의 효율적 RNA 재조합을 매개한다. 일 구체예에서, 2개의 단편으로부터 코딩 서열의 재구성은 제1 합성 인트론 (A)을 N-말단 코딩 단편의 3' 말단에 부가하고, 상보적 제2 합성 도메인 (A')을 C-말단 코딩 단편의 5' 말단에 부가하는 것에 의해 달성된다. 2개의 RNA가 세포의 고유한 RNA 스플라이싱 기구(즉, 스플라이세오솜 기구(spliceosome machinery))에 의해 재조합된다. 합성 인트론 도메인은 2개의 기능성 요소를 포함한다: (1) 재조합될 2개의 절반간 염기쌍 형성을 매개하는 이합체화 도메인 및 (2) 2개의 RNA 분자의 효율적 재구성을 매개하기 위해 효율적으로 스플라이싱 기구를 동원하도록 최적화된 도메인. 일부 구체예에서, 합성 인트론은 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148,

155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166으로 제공된 임의의 합성 인트론에 대해 적어도 50% 적어도 60%, 적어도 70%, 적어도 75%, 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100% 서열 동일성을 갖는 서열을 포함한다 (예를 들면, 도 10a 내지 10z 참조). 일부 구체예에서, 합성 인트론은 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166으로 제공된 임의의 합성 인트론에 대해 적어도 50% 적어도 60%, 적어도 70%, 적어도 75%, 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100% 서열 동일성을 가지나, 제공된 프로모터 서열은 갖지 않는 서열에 의해 코딩되는 RNA 분자이다. 당업자는 서열번호 1, 2, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166으로 제공된 분자가 단백질 코딩 부분(예를 들면,도 6a의 114 및 164)을 또 다른 목적 단백질 코딩 서열로 치환하기 위해 변형될 수 있다는 것을 이해할 것이다(예를 들면, 서열번호 1, 2, 22 또는 23의 YFP 코딩 서열이 치료 단백질 코딩 서열로 치환될 수 있음). 따라서, 본 명세서에서, 서열번호 1, 2, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166으로 제공된 임의의 합성 인트론 부분에 대한 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100% 서열 동일성을 갖는 합성 인트론 분자가 또한 제공된다 (예를 들면, 서열번호 22의 nt 3703-3975 및 서열번호 23의 nt 1-225). 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166으로 제공된 임의의 합성 인트론에 적어도 50% 적어도 60%, 적어도 70%, 적어도 75%, 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100% 서열 동일성을 가지나, 제공된 프로모터 서열은 갖지 않는 서열에 의해 코딩된 합성 인트론 RNA 분자가 또한 제공된다.

[0079] 예시적인 이합체화 도메인은 그들의 내부 2차/3차 구조를 최소화/최적화하도록 생물정보학을 이용하여(bioinformatically) 선택하였다. 테스트된 이합체화 도메인은 분자내 어닐링을 방지하기 위해 낮은 다양성 뉴클레오티드 서열(low diversity nucleotide sequences)의 긴 구역을 포함했다. 분자내 어닐링을 방지하는 것에 의해, 이러한 이합체화 도메인은 열린 구성(open configuration)으로 존재하고, 따라서, 상응하는 상보적 이합체화 도메인 서열과의 쌍 형성을 위해 이용가능하다. 합성 인트론 도메인은 스플라이싱 기구의 효율적 동원을 일으키는 ISE 요소(intronic splice enhancing elements)를 포함한다.

[0080] 본 명세서에서 개시되는 RNA 분자들은 적어도, RNA의 효율적인 스플라이싱 및 재조합을 가능하게 하도록 상보적 이합체화 도메인에 결합하기 위해 이용가능한 개방되고 이용가능한 단일가닥 영역을 갖는다. 일부 구체예에서, 이는 결합 도메인에 대해 퓨린만 또는 피리미딘만을 이용하는 것에 의해 달성된다. 퓨린 (및 마찬가지로 피리미딘)의 그 스스로와 쌍을 형성할 수 없는 불능 때문에, RNA의 이러한 구역들은 개방된 예측 구조를 갖는다.

[0081] RNA 분자들은 세포에서 단일가닥으로 존재한다. 단일가닥이므로, 그들은 내재적으로 스스로 혼성화되고, 그에 의해 강한 2차 및 3차 구조를 형성하는 경향이 있다. 가장 안정한 염기쌍은 G와 C, A와 U, 및 G와 U 워블 쌍(wobble pair)일 것이다. 열역학적으로, 2개의 염기의 쌍 형성이 개방 구성에 비해 선호된다. 효율적인 합성 핵산 분자를 설계하기 위해, 상호 간에 상보성을 갖는 2개의 이합체화 도메인은 상기 이합체화 도메인이 분자간 염기쌍 형성을 위해 이용가능하도록 개방 구성으로 존재한다. 합성 핵산 분자의 다른 부분들 간 분자내 염기쌍 형성을 방지하기 위해, 비양립성(incompatible) 염기들을 포함하는 비-다양성(non-diverse) 서열의 긴 구역이 포함될 수 있다. 예를 들면, 피리미딘 (즉, C 및 T) 또는 퓨린 (즉, A 및 G)의 긴 구역이 합성 핵산 분자에 존재할 수 있다. 피리디민은 다른 피리미딘과 규범적 염기쌍을 형성할 수 없고, 퓨린은 다른 퓨린과 규범적 염기쌍을 형성할 수 없다. 그러한 퓨린 또는 피리미딘의 구역은 수 bp 내지 수백 bp 범위일 수 있다. 이러한 구역들은 분자내 결합을 형성할 수 없으므로, 그들은 상보적 단편과의 분자간 염기쌍 형성을 위해 이용가능하다. 예를 들면, 합성 핵산 분자 A와 A'은 A는 피리미딘 구역 (예를 들면, 5'-CCUU(...)CCUU-3')을 포함하고, A'은 상보적인 퓨린 서열 (예를 들면, 5'-AAGG(...)AAGG-3')을 포함하도록 구성될 수 있다.

[0082] 본 명세서에서 개시된 합성 핵산 분자 (예를 들면, RNA 또는 상기 RNA를 코딩하는 DNA)는 게놈에서 정확하지 않은 부위로의 오프-타겟(off-target) 결합을 최소화하도록 설계된다. 오프 타겟 결합은 핵산 분자의 서열을 변형시키는 것에 의해 감소될 수 있다.

[0083] 동일한 설계 원칙, 즉, 개방된 합성 핵산 구조를 달성하기 위한 RNA 염기의 저다양성 구간(hypodiverse stretch)의 이용이 이합체화 도메인에서 단일 염기의 구간의 이용, 예를 들면, 연속된 C와 염기쌍을 형성할 연속된 G, 및 연속된 U와 염기쌍을 형성할 연속된 A의 이용까지 확장될 수 있다.

[0084] 2개 이상의 합성 핵산 분자의 재조합을 증가시키기 위해, 하기 방법이 이용될 수 있다. RNA 스플라이싱은 인트론의 5' 말단(스플라이스 도너 부위) 및 인트론의 3' 말단(그와 연관된 분지점 서열 및 피리미딘 트랙트를 갖는, 스플라이스 억셉터 부위)으로의 스플라이세오솜 성분들의 동원에 의존적이다. 상이한 RNP(ribonucleoprotein)가 단백질 연관 snRNA (small nuclear RNA)와 인트론 서열의 염기쌍 형성을 통해 인트론으로 동원된다. RNA 이합체화 및 재조합 도메인에 완벽한 일치 컨센서스 서열(perfect match consensus sequences)을 배치하는 것에 의해, 스플라이세오솜 성분들의 동원이 촉진되고, 그에 의해 스플라이세오솜 매개 재조합의 효율을 증가시킬 수 있다. 이전에 규명된 ISE 서열이 ISE로 지칭되는 추가적인 스플라이싱 촉진 인자들을 동원할 수 있다.

[0085] 일부 구체예에서, RNA 스플라이싱 서열을 위해 천연 RNA 서열을 이용하는 대신에, 컨센서스 서열이 이용된다. 예를 들면, 스플라이스 도너, 스플라이스 억셉터, 스플라이스 인핸서 및 스플라이스 분지점 서열을 포함한, 스플라이싱에 관련된 서열을 위해 컨센서스 서열이 이용될 수 있다. 이러한 합성 핵산 분자를 이용하여, 2개 (또는 그 이상의) RNA 분자들이 세포에서 엑스 비보, 인 비트로, 또는 인 비보로 함께 연속적으로 연결될 수 있다. 코딩된 합성 인트론 도메인 외부에, 합성 핵산 분자는 프로모터 및 코딩 서열을 포함할 수 있다. 예를 들면, 2개의 합성 핵산 분자가 단일 유전자의 2개의 절반을 가질 수 있다. YFP(yellow fluorescent protein)의 2개의 절반을 재구성하는 것에 의해 인 비보 및 인 비트로에서 이를 테스트하였고, 효율적인 것으로 확인되었다 (도 3a-3d 참조).

[0086] 합성 핵산 분자의 모듈성(modular nature)이 최적화된 상보적 이합체화 도메인의 조합 세트를 이용하여 다수의 RNA 단편의 연속 재조합(즉, >2)의 효율을 테스트하는 것을 가능하게 했다 (도 4a-4b). 3원 분할 YFP(three-way split yellow fluorescent protein)도 형질감염된 세포의 >80%에서 효율적으로 재구성되고, 고수준으로 발현되었다.

[0087] 이러한 결과는 너무 길어서 AAV와 같은 단일 유전자 요법 벡터 내에 맞출 수 없는 프로모터 및/또는 코딩 서열을 갖는 질병 유발 유전자 (또는 치료 단백질)의 발현의 경우와 같이, 단일 RNA 분자가 적어도 3개의 상이한 핵산 분자로부터 재구성될 수 있다는 것을 보여준다.

[0088] 일부 구체예에서, 본 발명의 조성물, 시스템, 키트, 및 방법의 합성 핵산 분자, 예를 들면, 합성 DNA 분자는 역전사효소에 의한 RNA 바이러스 게놈의 전사에 의해 생산된다.

[0089] 본 명세서에서 개시된 시스템은 개별적인 단편들 간의 효율적인 RNA 재조합을 가능하게 한다. 일부 구체예에서, 본 개시의 조성물, 시스템 또는 방법을 이용하여 달성되는 재구성 (즉, 스플라이싱 또는 재조합) 효율이 당업자에게 공지된 임의의 적합한 방법을 이용하여 결정된다. 일부 구체예에서, 재구성 효율은 대조군 RNA 대비 정확하게 연결된 RNA의 수준(meausre), 또는 대조군 단백질 대비 전장 단백질 또는 단백질 활성의 수준으로 표시된다. 일부 구체예에서, 대조군 RNA는 비연결(unjoined) RNA이고, 재구성 효율은 비연결 RNA 대비 연결된 RNA의 수준에 의해 표시된다. 이 측정은 연결부(junction) RNA와 비연결 3' RNA 종을 검출하고 비교하는 것에 의해 이루어질 수 있다(예를 들면, 연결부 RNA: 3' RNA). 일부 구체예에서, 2개 초과의 RNA가 연결되는 경우, 모든 연결부에서의 연결이 평가된다. 일부 구체예에서, 재구성 효율은 단백질 단편 또는 비활성 단백질 대비 전장 단백질 또는 활성 단백질의 수준에 의해 표시된다.

[0090] 일부 구체예에서, 재구성, 재조합 또는 스플라이싱 효율 (상이한 RNA 분자에 존재하는 2개 이상의 상이한 코딩 서열의 정확한 연결 및/또는 원하는 전장 단백질의 생산의 수준)은 약 10% 내지 약 100%이다. 일부 구체예에서, 재구성 효율은 약 10% 내지 약 15%, 약 10% 내지 약 20%, 약 10% 내지 약 25%, 약 10% 내지 약 30%, 약 10% 내지 약 40%, 약 10% 내지 약 50%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 10% 내지 약 100%, 약 15% 내지 약 20%, 약 15% 내지 약 25%, 약 15% 내지 약 30%, 약 15% 내지 약 40%, 약 15% 내지 약 50%, 약 15% 내지 약 60%, 약 15% 내지 약 70%, 약 15% 내지 약 80%, 약 15% 내지 약 90%, 약 15% 내지 약 100%, 약 20% 내지 약 25%, 약 20% 내지 약 30%, 약 20% 내지 약 40%, 약 20% 내지 약 50%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 20% 내지 약 100%, 약 25% 내지 약 30%, 약 25% 내지 약 40%, 약 25% 내지 약 50%, 약 25% 내지 약 60%, 약 25% 내지 약 70%, 약 25% 내지 약 80%, 약 25% 내지 약 90%, 약 25% 내지 약 100%, 약 30% 내지 약 40%, 약 30% 내지 약 50%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 30% 내지 약 100%, 약 40% 내지 약 50%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 40% 내지 약 100%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 50% 내지 약 100%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 60% 내지 약

100%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 70% 내지 약 100%, 약 80% 내지 약 90%, 약 80% 내지 약 100%, 또는 약 90% 내지 약 100%이다. 일부 구체예에서, 재구성 효율은 약 10%, 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다. 일부 구체예에서, 재구성 효율은 적어도 약 10%, 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 또는 약 90%이다. 일부 구체예에서, 재구성 효율은 최대(at most) 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다.

[0091] 일부 구체예에서, 재구성, 재조합 또는 스플라이싱 효율 (이 구체예에서, 상이한 RNA 분자에 존재하는 2개의 상이한 코딩 서열의 정확한 연결 및/또는 원하는 전장 단백질의 생산의 수준이고, 상기 2개의 상이한 코딩 서열은 약 3200 nt 내지 9000 nt, 예를 들면, 약 4000 to 9000 nt, 약 4400 to 9000 nt, 약 3200 내지 4000 nt, 약 3200 내지 3600 nt, 예를 들면, 약 4500 nt, 약 4000 nt, 약 3800 nt, 약 3600 nt, 또는 약 3200 nt의 전사물을 코딩함)은 약 10% 내지 약 100%이다. 일부 구체예에서, 2-파트 시스템(two-part system)을 이용한 재구성 효율은 약 10% 내지 약 15%, 약 10% 내지 약 20%, 약 10% 내지 약 25%, 약 10% 내지 약 30%, 약 10% 내지 약 40%, 약 10% 내지 약 50%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 10% 내지 약 100%, 약 15% 내지 약 20%, 약 15% 내지 약 25%, 약 15% 내지 약 30%, 약 15% 내지 약 40%, 약 15% 내지 약 50%, 약 15% 내지 약 60%, 약 15% 내지 약 70%, 약 15% 내지 약 80%, 약 15% 내지 약 90%, 약 15% 내지 약 100%, 약 20% 내지 약 25%, 약 20% 내지 약 30%, 약 20% 내지 약 40%, 약 20% 내지 약 50%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 20% 내지 약 100%, 약 25% 내지 약 30%, 약 25% 내지 약 40%, 약 25% 내지 약 50%, 약 25% 내지 약 60%, 약 25% 내지 약 70%, 약 25% 내지 약 80%, 약 25% 내지 약 90%, 약 25% 내지 약 100%, 약 30% 내지 약 40%, 약 30% 내지 약 50%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 30% 내지 약 100%, 약 40% 내지 약 50%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 40% 내지 약 100%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 50% 내지 약 100%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 60% 내지 약 100%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 70% 내지 약 100%, 약 80% 내지 약 90%, 약 80% 내지 약 100%, 또는 약 90% 내지 약 100%이다. 일부 구체예에서, 재구성 효율은 약 10%, 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다. 일부 구체예에서, 재구성 효율은 적어도 약 10%, 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 또는 약 90%이다. 일부 구체예에서, 재구성 효율은 최대 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다.

[0092] 일부 구체예에서, 재구성, 재조합 또는 스플라이싱 효율 (이 구체예에서, 상이한 RNA 분자에 존재하는 2개의 상이한 코딩 서열의 정확한 연결 및/또는 원하는 전장 단백질의 생산의 수준이고, 상기 2개의 상이한 코딩 서열은 약 4000 nt의 전사물을 코딩함)은 약 40% 내지 약 60%, 예를 들면, 약 40% 내지 약 50%, 약 42% 내지 약 47%, 예를 들면, 약 45%이다.

[0093] 일부 구체예에서, 재구성, 재조합 또는 스플라이싱 효율 (이 구체예에서, 상이한 RNA 분자에 존재하는 2개의 상이한 코딩 서열의 정확한 연결 및/또는 원하는 전장 단백질의 생산의 수준이고, 상기 2개의 상이한 코딩 서열은 약 3800 nt의 전사물을 코딩함)은 약 40% 내지 약 60%, 예를 들면, 약 40% 내지 약 50%, 약 42% 내지 약 47%, 예를 들면, 약 45%이다.

[0094] 일부 구체예에서, 재구성, 재조합 또는 스플라이싱 효율 (이 구체예에서, 상이한 RNA 분자에 존재하는 2개의 상이한 코딩 서열의 정확한 연결 및/또는 원하는 전장 단백질의 생산의 수준이고, 상기 2개의 상이한 코딩 서열은 약 3600 nt의 전사물을 코딩함)은 약 25% 내지 약 50%, 예를 들면, 약 30% 내지 약 40%, 예를 들면, 약 35%이다.

[0095] 일부 구체예에서, 재구성, 재조합 또는 스플라이싱 효율 (이 구체예에서, 상이한 RNA 분자에 존재하는 2개의 상이한 코딩 서열의 정확한 연결 및/또는 원하는 전장 단백질의 생산의 수준이고, 상기 2개의 상이한 코딩 서열은 약 3200 nt의 전사물을 코딩함)은 약 25% 내지 약 50%, 예를 들면, 약 30% 내지 약 40%, 예를 들면, 약 35%이다.

[0096] 일부 구체예에서, 재구성, 재조합 또는 스플라이싱 효율 (이 구체예에서, 상이한 RNA 분자에 존재하는 3개의 상이한 코딩 서열의 정확한 연결 및/또는 원하는 전장 단백질의 생산의 수준이고, 상기 3개의 상이한 코딩 서열은 약 3200 nt 내지 약 13,500 nt, 예를 들면, 약 4000 nt 내지 약 5,000 nt, 약 4000 nt 내지 약 13,500 nt, 약

6000 nt 내지 약 12,000 nt, 약 6000 nt 내지 약 10,000 nt, 또는 약 8000 nt 내지 약 12,000 nt, 예를 들면, 최대 약 13,500 nt의 전사물을 코딩함)은 약 10% 내지 약 100%이다. 일부 구체예에서, 3-파트 시스템을 이용한 재구성 효율은 약 10% 내지 약 15%, 약 10% 내지 약 20%, 약 10% 내지 약 25%, 약 10% 내지 약 30%, 약 10% 내지 약 40%, 약 10% 내지 약 50%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 10% 내지 약 100%, 약 15% 내지 약 20%, 약 15% 내지 약 25%, 약 15% 내지 약 30%, 약 15% 내지 약 40%, 약 15% 내지 약 50%, 약 15% 내지 약 60%, 약 15% 내지 약 70%, 약 15% 내지 약 80%, 약 15% 내지 약 90%, 약 15% 내지 약 100%, 약 20% 내지 약 25%, 약 20% 내지 약 30%, 약 20% 내지 약 40%, 약 20% 내지 약 50%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 20% 내지 약 100%, 약 25% 내지 약 30%, 약 25% 내지 약 40%, 약 25% 내지 약 50%, 약 25% 내지 약 60%, 약 25% 내지 약 70%, 약 25% 내지 약 80%, 약 25% 내지 약 90%, 약 25% 내지 약 100%, 약 30% 내지 약 40%, 약 30% 내지 약 50%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 30% 내지 약 100%, 약 40% 내지 약 50%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 40% 내지 약 100%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 50% 내지 약 100%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 60% 내지 약 100%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 70% 내지 약 100%, 약 80% 내지 약 90%, 약 80% 내지 약 100%, 또는 약 90% 내지 약 100%이다. 일부 구체예에서, 재구성 효율은 약 10%, 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다. 일부 구체예에서, 재구성 효율은 적어도 약 10%, 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 또는 약 90%이다. 일부 구체예에서, 재구성 효율은 최대 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다.

[0097] 일부 구체예에서, 재구성, 재조합 또는 스플라이싱 효율 (이 구체예에서, 상이한 RNA 분자에 존재하는 4개의 상이한 코딩 서열의 정확한 연결 및/또는 원하는 전장 단백질의 생산의 수준이고, 상기 4개의 상이한 코딩 서열은 약 3200 nt 내지 약 18,000 nt, 예를 들면, 약 4000 nt 내지 약 18,000 nt, 약 4000 nt 내지 약 5,000 nt, 약 10,000 nt 내지 약 18,000 nt, 약 15,000 nt 내지 약 18,000nt, 또는 약 12,000 nt 내지 약 15,000 nt, 예를 들면, 최대 약 18,000 nt의 전사물을 코딩함)은 약 10% 내지 약 100%이다. 일부 구체예에서, 4-파트 시스템을 이용한 재구성 효율은 약 10% 내지 약 15%, 약 10% 내지 약 20%, 약 10% 내지 약 25%, 약 10% 내지 약 30%, 약 10% 내지 약 40%, 약 10% 내지 약 50%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 10% 내지 약 100%, 약 15% 내지 약 20%, 약 15% 내지 약 25%, 약 15% 내지 약 30%, 약 15% 내지 약 40%, 약 15% 내지 약 50%, 약 15% 내지 약 60%, 약 15% 내지 약 70%, 약 15% 내지 약 80%, 약 15% 내지 약 90%, 약 15% 내지 약 100%, 약 20% 내지 약 25%, 약 20% 내지 약 30%, 약 20% 내지 약 40%, 약 20% 내지 약 50%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 20% 내지 약 100%, 약 25% 내지 약 30%, 약 25% 내지 약 40%, 약 25% 내지 약 50%, 약 25% 내지 약 60%, 약 25% 내지 약 70%, 약 25% 내지 약 80%, 약 25% 내지 약 90%, 약 25% 내지 약 100%, 약 30% 내지 약 40%, 약 30% 내지 약 50%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 30% 내지 약 100%, 약 40% 내지 약 50%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 40% 내지 약 100%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 50% 내지 약 100%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 60% 내지 약 100%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 70% 내지 약 100%, 약 80% 내지 약 90%, 약 80% 내지 약 100%, 또는 약 90% 내지 약 100%이다. 일부 구체예에서, 재구성 효율은 약 10%, 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다. 일부 구체예에서, 재구성 효율은 적어도 약 10%, 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 또는 약 90%이다. 일부 구체예에서, 재구성 효율은 최대 약 15%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다. 일부 구체예에서, 본 개시의 조성물, 시스템, 또는 방법은 당업자에게 공지된 적절한 방법을 이용하여 RNA 또는 단백질 생산 수준을 결정하는 것에 의해 평가된다. 일부 구체예에서, RNA 생산 수준은 대조군 RNA 대비 정확하게 연결된 RNA의 수준, 또는 대조군 대비 전장 단백질의 수준에 의해 표시된다. 일부 구체예에서, 상기 대조군 RNA는 상응하는 돌연변이 RNA 또는 내생 RNA이다. 예를 들면, 정확하게 연결된 RNA의 생산 수준을 결정하기 위해, 형질감염된 세포에서 생산된 돌연변이 또는 내생 RNA의 양 대비 연결된 RNA의 양의 비율이 비형질감염 세포에서의 동일한 비율과 비교된다. 일부 구체예에서, 대조군 RNA의 양 또는 대조군 단백질의 양 또는 활성 대비 정확하게 연결된 RNA의 양, 전장 단백질의 양, 또는 단백질 활성의 비율이 비교된다.

[0098] 일부 구체예에서, 달성되는 RNA 생산 수준은 5% 내지 100%이다. 일부 구체예에서, 달성되는 RNA 생산 수준은 약 5% 내지 약 100%이다. 일부 구체예에서, 달성되는 RNA 생산 수준은 약 5% 내지 약 10%, 약 5% 내지 약 20%, 약 5% 내지 약 25%, 약 5% 내지 약 30%, 약 5% 내지 약 40%, 약 5% 내지 약 50%, 약 5% 내지 약 60%, 약 5% 내지 약 70%, 약 5% 내지 약 80%, 약 5% 내지 약 90%, 약 5% 내지 약 100%, 약 10% 내지 약 20%, 약 10% 내지 약 25%, 약 10% 내지 약 30%, 약 10% 내지 약 40%, 약 10% 내지 약 50%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 10% 내지 약 100%, 약 20% 내지 약 25%, 약 20% 내지 약 30%, 약 20% 내지 약 40%, 약 20% 내지 약 50%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 20% 내지 약 100%, 약 25% 내지 약 30%, 약 25% 내지 약 40%, 약 25% 내지 약 50%, 약 25% 내지 약 60%, 약 25% 내지 약 70%, 약 25% 내지 약 80%, 약 25% 내지 약 90%, 약 25% 내지 약 100%, 약 30% 내지 약 40%, 약 30% 내지 약 50%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 30% 내지 약 100%, 약 40% 내지 약 50%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 40% 내지 약 100%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 50% 내지 약 100%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 60% 내지 약 100%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 70% 내지 약 100%, 약 80% 내지 약 90%, 약 80% 내지 약 100%, 또는 약 90% 내지 약 100%이다. 일부 구체예에서, 달성되는 RNA 생산 수준은 약 5%, 약 10%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다. 일부 구체예에서, 달성되는 RNA 생산 수준은 적어도 약 5%, 약 10%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 또는 약 90%이다. 일부 구체예에서, 달성되는 RNA 생산 수준은 최대 약 10%, 약 20%, 약 25%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 80%, 약 90%, 또는 약 100%이다.

[0099] 일부 구체예에서, 단백질 생산 수준은 대조군 단백질의 양 또는 활성 대비 전장 단백질의 양 또는 단백질 활성의 수준에 의해 표시된다. 일부 구체예에서, 대조군 단백질은 상응하는 돌연변이 단백질 또는 내생 단백질이다. 예를 들면, 형질감염된 세포에서 생산된 돌연변이 또는 내생 단백질의 양에 대한 전장 단백질의 양 또는 단백질의 활성의 비가 비형질감염 세포에서의 동일한 비율과 비교된다. 일부 구체예에서, 대조군 단백질은 예를 들면, 대조군 전장 단백질을 발현하도록 조작된 세포(상기 세포는 본 발명의 구조체로 형질감염되지 않음), 또는 대조군 전장 단백질을 발현하는 정상 개체로부터의 비-형질감염 세포에서 생산된 전장 단백질이고, 단백질 생산 수준은 형질감염된 세포에서 상기 단백질의 양 또는 활성을 측정하고, 이를 대조군 단백질의 양 또는 활성과 비교하는 것에 의해 결정된다. 일부 구체예에서, 대조군 단백질은 상기 구조체로 형질감염되거나 또는 형질감염되지 않은 세포에서 생산된 단백질의 돌연변이 형태이고, 단백질 생산 수준을 결정하기 위해 전장 단백질의 양 또는 단백질 활성이 대조군 단백질의 양 또는 활성과 비교된다. 일부 구체예에서, 단백질 생산 수준을 결정하기 위해 전장 단백질의 양 또는 단백질 활성이 내생(endogenous), 또는 하우스키핑(housekeeping) 단백질의 양 또는 단백질 활성과 비교된다.

[0100] 일부 구체예에서, 달성되는 단백질 생산 수준은 약 1% 내지 약 100%이다. 일부 구체예에서, 달성되는 단백질 생산 수준은 약 10% 내지 약 100%이다. 일부 구체예에서, 달성되는 단백질 생산 수준은 약 10% 내지 약 20%, 약 10% 내지 약 30%, 약 10% 내지 약 40%, 약 10% 내지 약 50%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 75%, 약 10% 내지 약 80%, 약 10% 내지 약 85%, 약 10% 내지 약 90%, 약 10% 내지 약 100%, 약 20% 내지 약 30%, 약 20% 내지 약 40%, 약 20% 내지 약 50%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 75%, 약 20% 내지 약 80%, 약 20% 내지 약 85%, 약 20% 내지 약 90%, 약 20% 내지 약 100%, 약 30% 내지 약 40%, 약 30% 내지 약 50%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 75%, 약 30% 내지 약 80%, 약 30% 내지 약 85%, 약 30% 내지 약 90%, 약 30% 내지 약 100%, 약 40% 내지 약 50%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 75%, 약 40% 내지 약 80%, 약 40% 내지 약 85%, 약 40% 내지 약 90%, 약 40% 내지 약 100%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 75%, 약 50% 내지 약 80%, 약 50% 내지 약 85%, 약 50% 내지 약 90%, 약 50% 내지 약 100%, 약 60% 내지 약 70%, 약 60% 내지 약 75%, 약 60% 내지 약 80%, 약 60% 내지 약 85%, 약 60% 내지 약 90%, 약 60% 내지 약 100%, 약 70% 내지 약 75%, 약 70% 내지 약 80%, 약 70% 내지 약 85%, 약 70% 내지 약 90%, 약 70% 내지 약 100%, 약 75% 내지 약 80%, 약 75% 내지 약 85%, 약 75% 내지 약 90%, 약 75% 내지 약 100%, 약 80% 내지 약 85%, 약 80% 내지 약 90%, 약 80% 내지 약 100%, 약 85% 내지 약 90%, 약 85% 내지 약 100%, 또는 약 90% 내지 약 100%이다. 일부 구체예에서, 달성되는 단백질 생산 수준은 약 10%, 약 20%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 또는 약 100%이다. 일부 구체예에서, 달성되는 단백질 생산 수준은 적어도 약 10%, 약 20%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 75%, 약 80%, 약 85%, 또는 약 90%이다. 일부 구체예에서, 달성되는 단백질 생산 수준은 최대 약 20%, 약 30%, 약 40%, 약 50%, 약 60%, 약 70%, 약 75%, 약 80%, 약

85%, 약 90%, 또는 약 100%이다.

[0101] 일부 구체예에서, 달성되는 단백질 활성 수준은 약 50% 내지 약 100%이다. 일부 구체예에서, 달성되는 단백질 활성 수준은 약 50% 내지 약 100%이다. 일부 구체예에서, 달성되는 단백질 활성 수준은 약 50% 내지 약 55%, 약 50% 내지 약 60%, 약 50% 내지 약 65%, 약 50% 내지 약 70%, 약 50% 내지 약 75%, 약 50% 내지 약 80%, 약 50% 내지 약 85%, 약 50% 내지 약 90%, 약 50% 내지 약 95%, 약 50% 내지 약 100%, 약 55% 내지 약 60%, 약 55% 내지 약 65%, 약 55% 내지 약 70%, 약 55% 내지 약 75%, 약 55% 내지 약 80%, 약 55% 내지 약 85%, 약 55% 내지 약 90%, 약 55% 내지 약 95%, 약 55% 내지 약 100%, 약 60% 내지 약 65%, 약 60% 내지 약 70%, 약 60% 내지 약 75%, 약 60% 내지 약 80%, 약 60% 내지 약 85%, 약 60% 내지 약 90%, 약 60% 내지 약 95%, 약 60% 내지 약 100%, 약 65% 내지 약 70%, 약 65% 내지 약 75%, 약 65% 내지 약 80%, 약 65% 내지 약 85%, 약 65% 내지 약 90%, 약 65% 내지 약 95%, 약 65% 내지 약 100%, 약 70% 내지 약 75%, 약 70% 내지 약 80%, 약 70% 내지 약 85%, 약 70% 내지 약 90%, 약 70% 내지 약 95%, 약 70% 내지 약 100%, 약 75% 내지 약 80%, 약 75% 내지 약 85%, 약 75% 내지 약 90%, 약 75% 내지 약 95%, 약 75% 내지 약 100%, 약 80% 내지 약 85%, 약 80% 내지 약 90%, 약 80% 내지 약 95%, 약 80% 내지 약 100%, 약 85% 내지 약 90%, 약 85% 내지 약 95%, 약 85% 내지 약 100%, 약 90% 내지 약 95%, 약 90% 내지 약 100%, 또는 약 95% 내지 약 100%이다. 일부 구체예에서, 달성되는 단백질 활성 수준은 약 50%, 약 55%, 약 60%, 약 65%, 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 약 95%, 또는 약 100%이다. 일부 구체예에서, 달성되는 단백질 활성 수준은 적어도 약 50%, 약 55%, 약 60%, 약 65%, 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 또는 약 95%이다. 일부 구체예에서, 달성되는 단백질 활성 수준은 최대 약 55%, 약 60%, 약 65%, 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 약 95%, 또는 약 100%이다.

[0102] 일부 구체예에서, 세포에서 생산되는 정확하게 연결된 RNA 또는 전장 단백질의 양은 특정한 질병 또는 질환과 관련하여 당업자에 의해 이해되는 바와 같이, 개체에서 질병 또는 질환을 개선 또는 치유하기에 충분하다. 일부 구체예에서, 세포에서 생산되는 정확하게 연결된 RNA 또는 전장 단백질의 양은 유효량이다. 일부 구체예에서, 이 양은 정상 세포에서 생산되는 상기 RNA 또는 단백질의 양의 약 50% 내지 100%에 해당한다. 일부 구체예에서, 이 양은 정상 세포에서 생산되는 상기 RNA 또는 단백질의 양의 약 40% 내지 100%에 해당한다. 일부 구체예에서, 이 양은 정상 세포에서 생산되는 상기 RNA 또는 단백질의 양의 약 40% 내지 약 45%, 약 40% 내지 약 50%, 약 40% 내지 약 55%, 약 40% 내지 약 60%, 약 40% 내지 약 65%, 약 40% 내지 약 70%, 약 40% 내지 약 75%, 약 40% 내지 약 80%, 약 40% 내지 약 85%, 약 40% 내지 약 90%, 약 40% 내지 약 100%, 약 45% 내지 약 50%, 약 45% 내지 약 55%, 약 45% 내지 약 60%, 약 45% 내지 약 65%, 약 45% 내지 약 70%, 약 45% 내지 약 75%, 약 45% 내지 약 80%, 약 45% 내지 약 85%, 약 45% 내지 약 90%, 약 45% 내지 약 100%, 약 50% 내지 약 55%, 약 50% 내지 약 60%, 약 50% 내지 약 65%, 약 50% 내지 약 70%, 약 50% 내지 약 75%, 약 50% 내지 약 80%, 약 50% 내지 약 85%, 약 50% 내지 약 90%, 약 50% 내지 약 100%, 약 55% 내지 약 60%, 약 55% 내지 약 65%, 약 55% 내지 약 70%, 약 55% 내지 약 75%, 약 55% 내지 약 80%, 약 55% 내지 약 85%, 약 55% 내지 약 90%, 약 55% 내지 약 100%, 약 60% 내지 약 65%, 약 60% 내지 약 70%, 약 60% 내지 약 75%, 약 60% 내지 약 80%, 약 60% 내지 약 85%, 약 60% 내지 약 90%, 약 60% 내지 약 100%, 약 65% 내지 약 70%, 약 65% 내지 약 75%, 약 65% 내지 약 80%, 약 65% 내지 약 85%, 약 65% 내지 약 90%, 약 65% 내지 약 100%, 약 70% 내지 약 75%, 약 70% 내지 약 80%, 약 70% 내지 약 85%, 약 70% 내지 약 90%, 약 70% 내지 약 100%, 약 75% 내지 약 80%, 약 75% 내지 약 85%, 약 75% 내지 약 90%, 약 75% 내지 약 100%, 약 80% 내지 약 85%, 약 80% 내지 약 90%, 약 80% 내지 약 100%, 약 85% 내지 약 90%, 약 85% 내지 약 100%, 또는 약 90% 내지 약 100%에 해당한다. 일부 구체예에서, 이 양은 정상 세포에서 생산되는 상기 RNA 또는 단백질의 양의 약 40%, 약 45%, 약 50%, 약 55%, 약 60%, 약 65%, 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 또는 약 100%에 해당한다. 일부 구체예에서, 이 양은 정상 세포에서 생산되는 상기 RNA 또는 단백질의 양의 약 적어도 약 40%, 약 45%, 약 50%, 약 55%, 약 60%, 약 65%, 약 70%, 약 75%, 약 80%, 약 85%, 또는 약 90%에 해당한다. 일부 구체예에서, 이 양은 정상 세포에서 생산되는 상기 RNA 또는 단백질의 양의 최대 약 45%, 약 50%, 약 55%, 약 60%, 약 65%, 약 70%, 약 75%, 약 80%, 약 85%, 약 90%, 또는 약 100%에 해당한다.

[0103] 재조합 효율 또는 생산 수준을 결정하기 위해 이용되는 RNA 또는 단백질의 측정은 당업자에게 공지된 적절한 방법에 의해 이루어질 수 있다. 일부 구체예에서, 재조합 효율 또는 생산 수준은 발현된 기능성 단백질의 양을, 예를 들면, 웨스턴 블롯팅에 의해 측정하는 것에 의해 결정된다. 일부 구체예에서, 재조합 효율 또는 생산 수준은 RNA 전사물을, 예를 들면, 2-프로브 기반 정량적 실시간 PCR을 이용하여 측정하는 것에 의해 결정된다. 예를 들면, 제1 어세이는 3' 엑손 코딩 서열에 완전히 포함된 서열에 걸친다(표지된 3' 프로브). 제2 어세이는 5' 엑손 코딩 서열과 3' 엑손 코딩 서열간 연결부에 걸친다 (표지된 연결부 프로브). 재구성 효율은 (연결부 프로브 카운트)/(3' 프로브 카운트)의 비로 계산될 수 있다. "재구성 효율", "재조합 효율" 및 "스플라이싱 효율"이 본

명세서에서 호환적으로 사용된다.

[0104] 일부 구체예에서, 이합체화 도메인은 약 20 내지 약 1000 nt, 또는 약 50 내지 약 160 nt, 또는 약 50 내지 약 500 nt, 또는 약 50 to 1000 nt이고, 재구성 효율은 정확하게 연결된 RNA 또는 전장 단백질의 유효량의 생산을 초래한다. 일부 구체예에서, 이합체화 도메인은 약 50 내지 약 160 nt이고, 재구성 효율은 정확하게 연결된 RNA 또는 전장 단백질의 유효량의 생산을 초래한다.

[0105] 다수의 RNA 분자간 효율적인 재조합을 달성하는 것은 전이 유전자의 AAV로의 패키징 및 전달을 가능하게 하고, 이는 단일 AAV의 패키징 한계를 초과한다. AAV 패키징 한계가 큰 유전자의 부재/결함에 의해 유발되는 질환에 대한 유전자 요법 방식에 대한 주요한 장애를 나타낸다. 이 시스템의 하나의 적용은 제한된 패키징 능력을 갖는 바이러스 벡터를 이용한 큰 질병-유발 유전자의 발현이다. 질병 및 유전자는 (질병 (유전자, OMIM 유전자 식별자)): 1) 듀센 근이영양증(Duchenne muscular dystrophy) 및 벡터 근이영양증(Becker muscular dystrophy) (디스트로핀, OMIM:300377); 2) 디스페를린병(Dysferlinopathies) (디스페를린(Dysferlin), OMIM:603009); 3) 낭포성 섬유증 (CFTR, OMIM:602421); 4) 어셔 증후군 1B(Usher's Syndrome 1B) (미오신 VIIA, OMIM:276903); 5) 스타르가르트병 1(Stargardt disease 1) (ABCA4, OMIM:601691); 6) A형 혈우병(Hemophilia A) (응고인자 VIII, OMIM:300841); 7) 폰 빌리브란트병(Von Willebrand disease) (폰 빌리브란트 인자, OMIM:613160); 8) 마르판 증후군(Marfan Syndrome) (피브릴린(Fibrillin) 1, OMIM:134797); 9) 폰 레클링하우젠병(Von Recklinghausen disease)(신경섬유병증(neurofibromatosis)-1, OMIM:162200), 및 난청 (OTOF, OMIM: 603681)을 포함하나, 그에 한정되지 않는다. 기타가 표 1에 제공된다. 또한, Cas9 단백질(예를 들면, 실시예 20-23에 예시된 것들)이, 예를 들면, 게놈 점 돌연변이를 치료하기 위해 또는 유전자를 활성화시키거나 또는 과발현시키기 위해, 본 명세서에서 제공되는 개시된 시스템을 이용하여 발현될 수 있다. 전이 유전자의 전달은 본 명세서에서 제공되는 방식을 이용하여 다수의 단편들로 분할하는 것에 의해 달성될 수 있다.

[0106] 본 명세서에 개시된 방법 및 시스템의 추가적 적용은 표적화된 유전자 발현을 위한 교차점 유전자 전달(intersectional gene delivery)을 포함한다. 하나는 단편화된 유전자를 코딩하는 2개의 바이러스의 차등적 감염/발현 패턴을 이용할 수 있다. 2개의 바이러스가 각각 단독으로 발현하는 것의 교차점(intersection)을 나타내는 세포들의 중복된 집단에서 재구성된 단백질이 발현될 것이다. 그러한 적용의 예는: (1) 2갈래로 분지되는 이중 투사 뉴런(bifurcating dual projection neurons)을 표지하기 위한 2개 (또는 그 이상)의 투사 표적(projection target)으로부터 역행으로 수송되는 바이러스 벡터를 이용한 단백질의 2개의 절반 (또는 3개의 1/3 부분, 또는 기타 부분)의 전달, (2) AUB 집단을 특이적으로 태깅/조작하기 위한, 집단 A에서 활성인 프로모터의 제어 하에 있는 하나의 단편 및 집단 B에서 활성인 프로모터로부터의 제2 단편의 전달, (3) AUB 집단을 특이적으로 태깅/조작하기 위한, 집단 A에 대한 친화성(tropism)을 갖는 바이러스 벡터에 의한 단백질의 제1 절반 및 집단 B에 대한 친화성을 갖는 바이러스 벡터에 의한 제2 절반의 전달, 또는 이러한 방식들의 조합을 포함할 수 있다.

[0107] 일 구체예에서, 이합체화 도메인은 압타머 서열, 예를 들면, (a) 압타머에 의해 인식되는 작은 분자 트리거(small molecular trigger), 또는 (b) 2개의 절반에 결합하고 따라서, 이합체화를 촉진하는, 세포에 존재하는 단백질의 존재 하에 이합체화를 촉진하는 압타머 서열이다.

[0108] 일부 구현예에서, 말단-연결(end-joining)을 위해 필요한 RNA-RNA 상호작용은 다른 뉴클레오티드에 의해 양성적으로 또는 음성적으로 조절될 수 있고, 예를 들면, (a) 2개의 절반에 대한 상동성을 갖는 안티센스 올리고뉴클레오티드 서열(ssDNA 유발 이합체화 (ssDNA triggered dimerization))로서, 그러한 구체예에서, 두 절반 모두에 상보적 서열을 갖는 안티센스 올리고뉴클레오티드가 2개의 분자를 서로 가교시켜서(bridge), 2개의 분자의 스플라이세오솜 매개 재조합을 촉진하거나, (b) 2개의 연결되는 RNA 중 하나에 대해 상동성을 갖는 안티센스 올리고뉴클레오티드가 2개의 분자의 RNA-이합체화를 막고, 유전자 발현에 대한 오프-스위치(off-switch)로 작용할 수 있거나, 또는 (c) 2개의 절반에 대한 상동성을 갖는 내생적 세포 RNA (RNA 유발 이합체화)로서, 그러한 구체예에서, 두 절반 모두에 상보적 서열을 갖는 세포 RNA (예를 들면, mRNA 또는 레트로요소(retroelement))가 2개의 분자를 서로 가교시켜서, 2개의 분자의 스플라이세오솜 매개 재조합을 촉진한다.

[0109] 이러한 분자, 단백질, 또는 RNA 매개 상호작용이 제어가능한/미세조정된(fine tuned) 유전자 발현 수준을 가능하게 한다: 결합 도메인과 상호작용하는 분자 (예를 들면, 안티센스 올리고뉴클레오티드, 소분자, 내생적 세포 RNA)에서의 적정(titrating)을 통해, 프로모터 활성과 독립적으로 발현 수준을 조정하기 위해, 2개의 절반간 이합체화 효율이 조절될 수 있다. 좁은 범위의 단백질 발현 수준이 요구되는 경우, 그러한 분할(installment)이 이용될 수 있다.

[0110] **III. 시스템**

[0111] 이합체화 서열을 포함하는 합성 인트론을 이용하여, 2개 이상의 RNA 분자, 예를 들면, 적어도 2개, 적어도 3개, 적어도 4개, 또는 적어도 5개의 상이한 RNA 분자들 (예를 들면, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 상이한 RNA 분자)을 재조합하기 위해 이용될 수 있는 시스템이 본 명세서에서 제공된다. 단백질 수준에서 단편화 및 2개의 단편의 재구성과 달리, 본 명세서에서 개시되는 접근방식은 적절한 분할점(split point)을 찾기 위한 광범위한 단백질 공학을 요구하지 않는다. RNA 수준에서의 재구성은 단백질의 2개의 단편의 무결절성(seamless) 연결을 가능하게 한다. 본 명세서에서 개시되는 방법 및 시스템은 큰 유전자 (및 상응하는 단백질), 예를 들면, 약 4.5 kb보다 큰 유전자, 적어도 5 kb, 적어도 5.5 kb, 적어도 6 kb, 적어도 kb, 적어도 8 kb, 적어도 8 kb, 적어도 10 kb, 적어도 13.5 kb, 또는 적어도 18 kb인 유전자가, 각각 별개의 벡터, 예를 들면, 다수의 AAV를 통해 세포 또는 개체 내로 도입될 수 있는 것인 2개 이상의 단편 또는 부분으로 나누어질 수 있게 한다. 일 구체예에서, 상기 시스템은 2개의 RNA 분자를 재조합하기 위한 2개의 부분을 포함하고, 예를 들면, 표적 단백질은 적어도 약 4500 nt 내지 약 9000 nt, 예를 들면, 4000 nt 내지 5000 nt에 의해 코딩된다. 일 구체예에서, 상기 시스템은 3개의 RNA 분자를 재조합하기 위한 3개의 부분을 포함하고, 예를 들면, 표적 단백질은 최대 약 13,500 nt, 예를 들면, 약 4500 nt 내지 약 13,500 nt 또는 4000 nt 내지 5000 nt에 의해 코딩된다. 일 구체예에서, 상기 시스템은 4개의 RNA 분자를 재조합하기 위한 4개의 부분을 포함하고, 예를 들면, 표적 단백질은 최대 약 18,000 nt, 예를 들면, 약 4500 nt 내지 약 18,000 nt, 또는 4000 nt 내지 5000 nt에 의해 코딩된다. 이는 벡터에서 이용가능한 한정된 공간을 극복하는데 기여한다. 일부 구체예에서, 내생적 프로모터 길이가 그의 상응하는 유전자가 AAV에서 발현되는 능력을 제한한다. 일부 구체예에서, 코딩 서열 길이가 그의 AAV에 발현되는 능력을 제한한다. 일부 구체예에서, 내생적 프로모터 길이 및 그의 코딩 서열 길이가 AAV에서 함께 발현되는 능력을 제한한다. 본 명세서에서 개시되는 시스템은 AAV에서 이전에 발현시키기 어려웠던 긴 서열을 발현시키기 위해 이용될 수 있다.

[0112] 일부 구체예에서, 재구성될 표적 단백질은 질병, 예를 들면, 단일 유전자성 질환, 열성 유전 질환, 큰 유전자 (예를 들면, 약 4500 nt 초과, 예를 들면, 적어도 5 kb, 적어도 5.5 kb, 적어도 6 kb, 적어도 kb, 적어도 8 kb, 적어도 8 kb, 적어도 10 kb, 적어도 13.5 kb, 또는 적어도 18 kb의 유전자) 및/또는 AAV의 수용능을 초과하는 유전자 (예를 들면, 5000 nt 초과) (예를 들면, 프로모터 + 코딩 서열) 중 돌연변이에 의해 유발되는 질환과 연관된 단백질이다. 그러한 질환의 예는 A형 혈우병 (F8 유전자, 7kb 코딩 영역 중 돌연변이에 의해 유발됨), B형 혈우병 (F9 유전자 중 돌연변이에 의해 유발됨), 듀센 근이영양증 (디스트로핀 유전자, 11 kb 코딩 영역 중 돌연변이에 의해 유발됨), 겸상세포 빈혈 (약 3.5 kb의 프로모터를 갖는, 헤모글로빈의 베타 글로빈 도메인 중 돌연변이에 의해 유발됨), 스타르가르트병 (ABCA4 유전자, 6.9 kb 코딩 영역 중 돌연변이에 의해 유발됨), 어셔 증후군 (청력 소실 및 시력 손상을 초래하는, MYO7A, 7 kb 코딩 영역 중 돌연변이에 의해 유발됨)을 포함하나, 이에 한정되지 않는다.

[0113] 일 구체예에서, 재구성되는 표적 단백질은 질병, 예를 들면, 암, 예를 들면, 유방, 폐, 전립선, 신장, 뇌, 뼈, 난소, 자궁, 피부, 또는 결장의 암을 치료할 수 있는 단백질이다. 일 구체예에서, 재구성되는 표적 단백질은 독소, 예를 들면, AB 독소, 예를 들면, 디프테리아 독소 A 또는 슈도모나스 외독소 A, 또는 수용체 결합 활성이 결여된 형태(예를 들면, 디프테리아 독소 DAB389, DAB486, DT388, DT390, 또는 슈도모나스 외독소 A PE38 또는 PE40)일 수 있다.

[0114] 일부 구체예에서, 표적 단백질을 코딩하고 본 명세서에서 개시된 방법 및 시스템에서 사용되는 RNA 서열은 표적 개체 또는 세포에서의 발현을 위해 코돈 최적화되고, 예를 들면, 인간, 개, 돼지, 고양이, 마우스, 또는 랫트 세포에서의 발현을 위해 코돈 최적화된다. 따라서, 일부 구체예에서, 상기 RNA 코딩 서열은 선호되는 코돈을 포함한다 (예를 들면, 낮은 이용율을 갖는 드문 코돈은 포함하지 않음). 코돈 최적화는 표적 개체 또는 세포에서 풍부한 tRNA 수준을 확인하는 것에 의해 수행될 수 있다. 일부 구체예에서, 단백질을 코딩하는 RNA 서열은 RNA 재조합 반응을 최대화하기 위해 크립틱 스플라이스 도너 및 억셉터 부위에 대해 탈-농축된다(de-enriched).

[0115] 일부 구체예에서, 단백질은 2개의 부분, 예를 들면, 약 2개의 동등한 절반 (또는 기타 비율, 예를 들면, 약 1/3을 발현하는 부분 A 및 약 2/3를 발현하는 부분 B, 또는 약 1/4을 발현하는 부분 A 및 약 3/4를 발현하는 부분 B, 등)으로 나뉠 수 있다. 그러나, 각 부분이 동일한 개수의 뉴클레오티드이어야 하는 것 (또는 동일한 개수의 아미노산을 코딩함)이 요구되지 않는다. 그러한 구체예에서, 상기 방법은 2개의 합성 핵산 분자 (예를 들면, RNA 또는 그러한 RNA를 코딩하는 DNA), 즉, 단백질의 N-말단 부분에 대한 코딩 서열을 포함하는 하나의 분자와 상기 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 또 다른 하나의 분자를 이용할 수 있다. 이러한 기반에 근거하여, 당업자는 단백질을 2개의 단편 또는 부분으로 나누는 것 외에, 목적 단백질이 2개 초과의 단편,

예를 들면, 3개의 단편으로 분리 또는 분할될 수 있다는 것을 이해할 것이다. 3개의 RNA 분자의 인트론 서열의 설계 원칙은 2개의 RNA 분자에 대한 것과 유사하나, 대신에, 2개의 연결부 중 하나에 대한 이합체화 도메인의 상이한 쌍이 이용된다. 따라서, 예를 들면, N-말단 단백질 코딩 서열이 특이적 결합 도메인을 갖는 인트론 서열(예를 들면, 제1 이합체화 서열)로 이어지고, 중간 코딩 서열은 상기 제1 이합체화 서열에 상보적인 서열을 갖는 인트론 서열 (제2 이합체화 서열)을 포함한다. 중간 코딩 단편은 또 다른 이합체화 서열 (제2 이합체화 서열과 다른, 제3 이합체화 서열)을 갖는 또 다른 인트론 단편으로 이어진다. 제3 단편은 상기 단백질의 C-말단 코딩 서열을 포함하고, 상기 제3 이합체화 서열에 상보적인 이합체화 서열(제4 이합체화 서열)을 갖는 인트론 영역을 포함한다. 1개 초과의 중간 부분의 이용에서, 2개의 중간 부분은 개별적인 부분들을 구별하는 것으로 이해되는 방식으로, 중간 부분(middle portion) 및 제1 중간 부분, 또는 제1 중간 부분 및 제2 중간 부분, 또는 제1 중간 부분, 제2 중간 부분, 및 제3 중간 부분 등으로 지칭될 수 있다.

[0116] 일 구체예에서, 원하는 단백질이 N-말단 부분 및 C-말단 부분 (예를 들면, 거의 반으로, 또는 불균등한 비율로, 예를 들면, 1/3 및 2/3 또는 1/4 및 3/4으로 나뉨)으로 나뉘고, 이들은 본 명세서에서 개시된 시스템 및 방법을 이용하여 재구성될 수 있다. 도 6a를 참조하면, 그러한 구체예에서, 상기 시스템은 적어도 2개의 합성 핵산 분자(110, 150)를 포함한다. 각 핵산 분자 (110, 150)는 DNA 또는 RNA로 구성될 수 있다 (RNA인 경우, 프로모터 (112, 152)가 없다). 일부 구체예에서, 분자 (110, 150) 각각은 길이가 약 적어도 100개 뉴클레오티드/리보뉴클레오티드 (nt), 예를 들면, 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 적어도 8000 nt, 적어도 10,000 nt, 예를 들면, 200 내지 10,000 nt, 200 내지 8000 nt, 500 내지 5000 nt, 또는 200 내지 1000 nt이다. 분자 (110, 150)은 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다.

[0117] 분자(110)은 스플라이스 도너 (116)를 포함하므로, 상기 시스템의 5'-위치(5'-located) 분자이다. 분자 (110)가 DNA인 것인 구현예에서, 상기 분자는 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 프로모터 (112)를 포함하고, 상기 RNA 분자는 5'에서 3'으로: 표적 단백질의 N-말단 부분에 대한 코딩 서열 (114)로서, 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열 (114)은 상기 표적 단백질 코딩 서열의 3'-말단에 스플라이스 연결부, SD (116), 선택적인 DISE (118), 선택적인 ISE (120), 이합체화 도메인 (122), 및 선택적인 폴리아데닐화 서열 (124)을 포함한다. 임의의 프로모터 (112) (또는 인핸서), 예를 들면, RNA 폴리머라아제 II를 이용하는 프로모터, 예를 들면, 항시적 또는 유도성 프로모터가 사용될 수 있다. 일부 구체예에서, 프로모터 (112)는 조직-특이적 프로모터, 예를 들면, 근육 조직 (예를 들면, 골격근 또는 심장근), 시각 조직 (예를 들면, 망막 조직), 내이 조직, 간 조직, 췌장 조직, 폐 조직, 피부 조직, 골 조직, 또는 신장 조직에서 항시적으로 활성인 프로모터이다. 일부 구체예에서, 프로모터 (112)는 세포-특이적 프로모터, 예를 들면, 암 세포 또는 정상 세포에서 항시적으로 활성인 프로모터이다. 일부 구체예에서, 프로모터 (112)는 발현되는 표적 단백질의 내생적 프로모터이고, 일부 구체예에서, 길다 (예를 들면, 적어도 2500 nt, 적어도 3000 nt, 적어도 4000 nt, 적어도 5000 nt, 또는 적어도 7500 nt). 일부 구체예에서, 프로모터 (112)는 길이가 적어도 약 50 nt(nucleotides), 예를 들면, 적어도 100, 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 적어도 8000 nt, 적어도 9000 nt, 또는 적어도 10,000 nt, 예를 들면, 50 내지 10,000 nt, 100 내지 5000 nt, 500 내지 5000 nt, 또는 50 내지 1000 nt 길이이다. 일부 구체예에서, 분자 (110)은 DNA이고, 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 또는 적어도 8000 nt, 예를 들면, 200 내지 10,000 nt, 200 내지 8000 nt, 500 내지 5000 nt, 또는 200 내지 1000 nt 길이이다. 도 6f에 도시된 바와 같이, 분자 (110)가 RNA, 예를 들면, DNA의 RNA로의 전사 후 RNA인 것인 구현예에서, 분자 (110)는 프로모터 (112)를 포함하지 않고, 분자 (114)는 표적 단백질의 N-말단 부분에 대한 코딩 서열에 의해 코딩되는 RNA이다. 일부 구체예에서, 분자 (110)는 RNA이고, 프로모터 (112)를 포함하지 않고, 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 또는 적어도 8000 nt, 예를 들면, 200 내지 10,000 nt, 200 내지 8000 nt, 500 내지 5000 nt, 또는 200 내지 1000 nt 길이이다. (프로모터 (112)를 갖거나 또는 갖지 않는) 분자 (110)는 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다.

[0118] N-말단 코딩 서열 (또는 그에 의해 코딩된 RNA 서열) (114)의 3' 말단 근처의 스플라이스 연결부는 분자 (110, 150)가 도입되는 표적 세포 또는 개체에서 발견되는 컨센서스 서열과 조화(match)될 수 있다. 인간에서, 스플라이스 연결부 서열은 U2-의존성 인트론에 대한 5' 스플라이스 부위의 위치 -1 및 -2에 있는 AG (아데닌-구아닌) 또는 UG (우라실-구아닌), 또는 U12-의존성 인트론에 대한 AG, UG, CU (시토신-우라실), 또는 UU이다. 따라서, 일부 구체예에서, 스플라이스 연결부는 길이가 2 nt이고, N-말단 코딩 부분 (114)의 3' 말단은 AG, UG, CU 또는

UU이다. 일부 구체예에서, 표적 단백질의 부분을 코딩하는 DNA 분자는 다수의 스플라이스 연결부의 부분을 코딩하는 서열들을, 예를 들면, 표적 단백질의 N-말단 부분을 코딩하는 DNA 분자의 3' 말단, 및 표적 단백질의 C-말단 부분을 코딩하는 DNA 분자의 5' 말단에 포함한다.

[0119] 분자 (110)의 나머지 3'-말단 부분은 인트론 (130)이다. 일부 구체예에서, 인트론 서열 (130)은 길이가 약 적어도 10 nt, 예를 들면, 적어도 20 nt, 적어도 50 nt, 적어도 100 nt, 적어도 250 nt, 적어도 250 nt, 적어도 300 nt, 적어도 400 nt, 또는 적어도 500 nt이고, 예를 들면, 20 내지 500, 20 내지 250, 20 내지 100, 50 내지 100, 또는 50 내지 200 nt 길이이다. N-말단 코딩 서열 (또는 그에 의해 코딩된 RNA) (114)의 바로 다음은 스플라이스 도너 (SD) (116) (예를 들면, SD 컨센서스 서열, 예를 들면, SD 인간 컨센서스 서열)이다. 따라서, 인트론 서열 (130)의 SD (116)은 N-말단 코딩 서열 (114)의 3'이다. SD (116)는 스플라이세오솜 성분이 RNA 분자에 결합하기 위한 인식 서열을 형성한다. SD (116)의 서열은 분자 (110, 150)가 도입되는 표적 세포 또는 개체에서 발견되는 SD 컨센서스 서열일 수 있다. 일부 구체예에서, SD (116)는 길이가 적어도 2 nt, 예를 들면, 적어도 5 nt, 또는 적어도 10 nt, 예를 들면, 2 내지 10, 2 내지 8, 2 내지 5, 또는 5 내지 10 nt이다. SD (116)는 U2 또는 U12 의존성 스플라이싱 기구를 동원하기 위해 이용될 수 있다. 일 구체예에서, U2 의존성 스플라이싱이 인간 세포에서 이용되고, SD (116) 서열은 GUAAGUAUU를 포함하거나, 또는 GUAAGUAUU이다. 일 구체예에서, U12 의존성 스플라이싱이 인간 세포에서 이용되고, SD (116) 서열은 AUAUCCUUUUUA (서열번호 137) 또는 GUAUCCUUUUUA (서열번호 138)를 포함하거나, 또는 AUAUCCUUUUUA (서열번호 137) 또는 GUAUCCUUUUUA (서열번호 138)이다. 명세서 전체에서, RNA 서열은 뉴클레오티드 A,G,T 및 C를 이용하여 기재될 수 있고, DNA 서열은 뉴클레오티드 A,G,U 및 C를 이용하여 기재될 수 있는 것으로 이해된다.

[0120] 인트론 서열 (130)은 선택적으로 DISE (downstream intronic splice enhancer) (118) 및 ISE (intronic splice enhancer) (120)로 지칭되는 스플라이싱 인핸서 서열의 세트 중 하나 또는 둘 모두를 포함하고, 이들은 스플라이세오솜의 작용을 촉진한다(예를 들면, 활성을 증가시킨다). 일부 구체예에서, 인트론 서열 (130)은 적어도 2개의 스플라이싱 인핸서 서열, 예를 들면, 적어도 3개, 적어도 4개, 또는 적어도 5개의 스플라이싱 인핸서 서열을 포함한다. 예시적인 스플라이싱 인핸서 서열은 DISE (118) 및 ISE (120)를 포함한다. 일부 구체예에서, 인트론 서열 (130) 중 하나 이상의 스플라이싱 인핸서 서열 (118, 120)의 포함은 스플라이싱 효율을 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 75%, 적어도 80%, 적어도 90% 또는 적어도 95% 증가시킨다. 이용될 수 있는 예시적인 스플라이싱 인핸서 서열이 서열번호 26-136, 151, 및 152, 및 GGGTTT, GGTGGT, TTTGGG, GAGGGG, GGTATT, GTAACG, GGGGGTAGG, GGAGGGTTT, GGGTGGTGT TTCAT, CCATTT, TTTTAAA, TGCAT, TGCATG, TGTGTT, CTAAC, TCTCT, TCTGT, TCTTT, TGCATG, CTAAC, CTGCT, TAACC, AGCTT, TTCATTA, GTTAG, TTTTGC, ACTAAT, ATGTTT, CTCTG, GGG, GGG(N)2-4GGG, TGGG, YCAY, UGCAUG, 또는 3x($G_{3-6}N_{1-7}$)로 제공된다. 일부 구체예에서, DISE (118)가 존재하는 경우, 길이가 적어도 3 nt, 적어도 4 nt, 적어도 5 nt, 적어도 10 nt, 적어도 25 nt, 적어도 50 nt, 적어도 75 nt, 또는 적어도 100 nt, 예를 들면, 3 내지 10, 3 내지 11, 4 내지 11, 5 내지 11, 10 내지 50, 5 내지 100, 10 내지 25, 10 내지 20, 또는 20 내지 75 nt일 수 있고, DISE (118)의 서열은 CUCUUUCUUUTCCAUGGGUUGGCU (서열번호 134), TGCATG, CTAAC, CTGCT, TAACC, AGCTT, TTCATTA, GTTAG, TTTTGC, ACTAAT, ATGTTT 또는 CTCTG이거나, 또는 이를 포함할 수 있다. 일부 구체예에서, ISE (120)가 존재하는 경우, 길이가 약 적어도 3 nt, 적어도 4 nt,적어도 5 nt, 적어도 10 nt, 예를 들면, 적어도 20 nt, 적어도 25 nt, 적어도 30 nt, 적어도 40 nt, 또는 적어도 50 nt, 예를 들면, 3 내지 10, 3 내지 11, 4 내지 11, 5 내지 11, 10 내지 50, 20 내지 25, 10 내지 25, 10 내지 20, 또는 20 내지 40 nt일 수 있다. 일 구체예에서, ISE (120)의 서열은 GGCUGAGGGAAGGACUGUCCUGGG (서열번호 135), GGGUUAUGGGACC (서열번호 136), TTCAT, CCATTT, TTTTAAA, TGCAT, TGCATG, TGTGTT, CTAAC, TCTCT, TCTGT, 또는 TCTTT이거나 또는 이를 포함할 수 있다. 일부 구체예에서, 인트론 서열 (130)은 적어도 2개, 적어도 3개, 또는 적어도 4개의 ISE (120)를 포함한다. 일부 구체예에서, ISE (120)는 서열번호 173, 174, 175, 176, 177, 178, 179, 180, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 199, 200, 201, 202, 또는 203에 대해 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 1개 이상의 서열, 예를 들면, 적어도 2개, 적어도 3개의 그러한 서열, 예를 들면, 1, 2, 3, 4 또는 5개의 그러한 서열이거나 또는 이를 포함한다. 일부 구체예에서, DISE (118)는 서열번호 173, 174, 175, 176, 177, 178, 179, 180, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 199, 200, 201, 202, 또는 203에 대해 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 1개 이상의 서열, 예를 들면, 적어도 2개, 또는 적어도 3개의 그러한 서열, 예를 들면, 1, 2, 3, 4 또는 5개의 그러한 서열이거나 또는 이를 포함한다.

[0121] SD (116) (및 존재하는 경우, 인핸서 서열 (118, 120))는 3'으로 조합될 N-말단 코딩 서열 (또는 그에 의해 코딩되는 RNA) (114)과 C-말단 코딩 서열 (154)을 모으기 위해 이용되는 이합체화 도메인 (122)으로 이어진다. 분자 (110)의 인트론 서열 (130) 부분은 선택적으로 3'-말단에 폴리아데닐화 부위 (124)를 포함할 수 있고, 이는 그 단편의 전사를 종료시킨다. 일부 구체예에서, 폴리아데닐화 서열 (124)은 적어도 15개의 A, 예를 들면, 15 내지 30개 또는 15 내지 20개의 A의 폴리A 서열이다.

[0122] 일부 구체예에서, 제1 이합체화 도메인 (122) (및 분자 (150)의 제2 이합체화 도메인 (154))은 복수의 쌍을 이루지 않은(unpaired) 뉴클레오티드(즉, 분자 (110) 자체의 구조 중 쌍을 이루지 않은 뉴클레오티드)를 포함한다. 이합체화 도메인에 쌍을 이루지 않은 뉴클레오티드를 갖는 것은 5' (또는 제1) 이합체화 도메인 (122)과 3' (또는 제2) 이합체화 도메인 (154)이 염기쌍 형성을 통해 상호작용할 수 있게 한다. 이러한 상호작용을 통해, 분자 (110 및 150)는 근접하게 유지되어 스플라이세오솜이 N-말단 코딩 영역 (또는 그에 의해 코딩되는 RNA) (114)과 C 말단 코딩 영역 (또는 그에 의해 코딩되는 RNA) (164)을 연결시키는 것에 의해 두 분자들을 재조합하도록 촉발한다.

[0123] 일 구체예에서, 이합체화 도메인 (122) (및 154)은 "저다양성(hypodiverse) 서열"을 포함하고, 상기 서열은 제한된 다양성의 뉴클레오티드를 포함하고 따라서 각 분자 (110, 150)의 이차 구조에서 스스로와 스템 루프(stem loop)를 형성할 가능성이 없다. 그러한 저다양성 이합체화 도메인 (122) (및 154)은 단백질의 N- 및 C-말단을 코딩하는 DNA (또는 그에 의해 코딩되는 RNA)(114, 164)의 서열과 독립적으로, 상대적으로 개방된 구성(configuration)일 수 있다. 이는 제1 이합체화 도메인 (122)의 뉴클레오티드가 분자 (150)의 상응하는 제2 이합체화 도메인 (154)와 염기쌍 형성을 위해 이용될 수 있게 하여, N-말단 코딩 서열 (또는 그에 의해 코딩되는 RNA) (114)과 C-말단 코딩 서열 (또는 그에 의해 코딩되는 RNA) (164)의 후속 연결을 가능하게 한다. 일부 구체예에서, 제1 및 제2 이합체화 도메인 (122, 154)은 스템(stem)을 형성할 수 있는 서열이 점재된 저다양성 서열을 포함하여, 슈도노트(pseudoknot) 형성의 부재시 개방되어 염기쌍 형성에 이용될 수 있는 국소 RNA 루프를 초래한다 (도 6b). 예시적인 저다양성 서열은 반복된 일련(repeated series)의 U (예를 들면, 30 내지 500개의 U), 반복된 일련의 A(예를 들면, 30 내지 500개의 A). 반복된 일련의 G(예를 들면, 30 내지 500개의 G), 반복된 일련의 C(예를 들면, 30 내지 500개의 C), A와 G만 포함하는 혼합 서열(예를 들면, 30 내지 500개의 A 및 G, 예를 들면, 반복될 수 있는 AAAGAAGGAA(...) (서열번호 149)), C와 U만 포함하는 혼합 서열(예를 들면, 30 내지 500개의 C 및 U, 예를 들면, 반복될 수 있는 CUUUCUUUUCUU(...) (서열번호 150))을 포함한다. 기타 예시적인 저다양성 서열은 저다양성 서열에 의해 플랭킹된(flanked) 헬릭스를 형성하는 상보적 서열을 포함한다.

[0124] 일부 구체예에서, 제1 및 제2 이합체화 도메인 (122, 154)은 퓨린만 포함하거나 또는 피리미딘만 포함한다. 일 구체예에서, 제1 이합체화 도메인 (122)은 퓨린만 포함하고, 제2 이합체화 도메인 (154)은 피리미딘만 포함한다. 또 다른 구체예에서, 제1 이합체화 도메인 (122)은 피리미딘만 포함하고, 제2 이합체화 도메인 (154)은 퓨린만 포함한다. 퓨린 (및 마찬가지로 피리미딘)의 그 스스로와 쌍을 형성할 수 없는 불능 때문에, RNA의 이러한 구간들은 개방된 예측 구조를 갖는다.

[0125] 일부 구체예에서, 제1 및 제2 이합체화 도메인 (122, 154)은 RNA 재조합과 경쟁할 수 있는 크립틱 스플라이스 억셉터, 예를 들면, 스플라이스 도너 컨센서스 서열 NNNAGGUNNNN (서열번호 151) 또는 NNNUGGUNNNN (서열번호 152) (서열 중 N은 임의의 뉴클레오티드를 의미함)와 유사한 서열을 포함하지 않는다. 일부 구체예에서, 제1 이합체화 도메인 (122)은 1000 nt 이하, 예를 들면, 750 nt 이하, 500 nt 이하, 예를 들면, 6 내지 1000 nt, 10 내지 1000 nt, 20 내지 1000 nt, 30 내지 1000 nt, 30 내지 750 nt, 30 내지 500 nt, 50 내지 500 nt, 50 내지 100 nt, 또는 100 내지 250 nt이다. 일부 구체예에서, 제1 이합체화 도메인 (122)은 50 nt보다 길고, 예를 들면, 적어도 51 nt, 적어도 100 nt, 적어도 150 nt, 적어도 161 nt, or 적어도 170 nt, 예를 들면, 51 내지 159 nt, 51 내지 150 nt, 51 내지 120 nt, 51 내지 100 nt, 또는 51 내지 70 nt이다. 일부 구체예에서, 제1 이합체화 도메인 (122)은 160 nt 초과이고, 예를 들면, 적어도 161 nt, 적어도 170 nt, 적어도 180 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 적어도 600 nt, 적어도 700 nt, 적어도 800 nt, 적어도 900 nt, 또는 적어도 1000 nt, 예를 들면, 161 내지 100 nt, 161 내지 500 nt, 161 내지 300 nt, 161 내지 200 nt, 또는 161 내지 170 nt이다. 일부 구체예에서, 제1 이합체화 도메인 (122)은 50 nt 미만, 예를 들면, 6 내지 49 nt, 6 내지 45 nt, 6 내지 40 nt, 6 내지 30 nt, 6 내지 20 nt, 또는 6 내지 10 nt이다.

[0126] 일부 구체예에서, 이합체화 도메인은 20 내지 160 nt, 50-500 nt, 또는 500-1000 nt이다. 일부 구체예에서, 이합체화 도메인은 약 20 nt 내지 약 160 nt이다. 일부 구체예에서, 이합체화 도메인은 약 20 nt 내지 약 40 nt, 약 20 nt 내지 약 50 nt, 약 20 nt 내지 약 70 nt, 약 20 nt 내지 약 90 nt, 약 20 nt 내지 약 100 nt, 약 20 nt 내지 약 110 nt, 약 20 nt 내지 약 120 nt, 약 20 nt 내지 약 130 nt, 약 20 nt 내지 약 140 nt, 약 20 nt

내지 약 150 nt, 약 20 nt 내지 약 160 nt, 약 40 nt 내지 약 50 nt, 약 40 nt 내지 약 70 nt, 약 40 nt 내지 약 90 nt, 약 40 nt 내지 약 100 nt, 약 40 nt 내지 약 110 nt, 약 40 nt 내지 약 120 nt, 약 40 nt 내지 약 130 nt, 약 40 nt 내지 약 140 nt, 약 40 nt 내지 약 150 nt, 약 40 nt 내지 약 160 nt, 약 50 nt 내지 약 70 nt, 약 50 nt 내지 약 90 nt, 약 50 nt 내지 약 100 nt, 약 50 nt 내지 약 110 nt, 약 50 nt 내지 약 120 nt, 약 50 nt 내지 약 130 nt, 약 50 nt 내지 약 140 nt, 약 50 nt 내지 약 150 nt, 약 50 nt 내지 약 160 nt, 약 70 nt 내지 약 90 nt, 약 70 nt 내지 약 100 nt, 약 70 nt 내지 약 110 nt, 약 70 nt 내지 약 120 nt, 약 70 nt 내지 약 130 nt, 약 70 nt 내지 약 140 nt, 약 70 nt 내지 약 150 nt, 약 70 nt 내지 약 160 nt, 약 90 nt 내지 약 100 nt, 약 90 nt 내지 약 110 nt, 약 90 nt 내지 약 120 nt, 약 90 nt 내지 약 130 nt, 약 90 nt 내지 약 140 nt, 약 90 nt 내지 약 150 nt, 약 90 nt 내지 약 160 nt, 약 100 nt 내지 약 110 nt, 약 100 nt 내지 약 120 nt, 약 100 nt 내지 약 130 nt, 약 100 nt 내지 약 140 nt, 약 100 nt 내지 약 150 nt, 약 100 nt 내지 약 160 nt, 약 110 nt 내지 약 120 nt, 약 110 nt 내지 약 130 nt, 약 110 nt 내지 약 140 nt, 약 110 nt 내지 약 150 nt, 약 110 nt 내지 약 160 nt, 약 120 nt 내지 약 130 nt, 약 120 nt 내지 약 140 nt, 약 120 nt 내지 약 150 nt, 약 120 nt 내지 약 160 nt, 약 130 nt 내지 약 140 nt, 약 130 nt 내지 약 150 nt, 약 130 nt 내지 약 160 nt, 약 140 nt 내지 약 150 nt, 약 140 nt 내지 약 160 nt, 또는 약 150 nt 내지 약 160 nt이다. 일부 구체예에서, 이합체화 도메인은 약 20 nt, 약 40 nt, 약 50 nt, 약 70 nt, 약 90 nt, 약 100 nt, 약 110 nt, 약 120 nt, 약 130 nt, 약 140 nt, 약 150 nt, 또는 약 160 nt이다. 일부 구체예에서, 이합체화 도메인은 적어도 약 20 nt, 약 40 nt, 약 50 nt, 약 70 nt, 약 90 nt, 약 100 nt, 약 110 nt, 약 120 nt, 약 130 nt, 약 140 nt, 또는 약 150 nt이다. 일부 구체예에서, 이합체화 도메인은 최대 약 40 nt, 약 50 nt, 약 70 nt, 약 90 nt, 약 100 nt, 약 110 nt, 약 120 nt, 약 130 nt, 약 140 nt, 약 150 nt, 또는 약 160 nt이다.

[0127] 일부 구체예에서, 이합체화 도메인은 약 50 nt 내지 약 500 nt이다. 일부 구체예에서, 이합체화 도메인은 약 50 nt 내지 약 100 nt, 약 50 nt 내지 약 150 nt, 약 50 nt 내지 약 200 nt, 약 50 nt 내지 약 250 nt, 약 50 nt 내지 약 300 nt, 약 50 nt 내지 약 350 nt, 약 50 nt 내지 약 400 nt, 약 50 nt 내지 약 500 nt, 약 100 nt 내지 약 150 nt, 약 100 nt 내지 약 200 nt, 약 100 nt 내지 약 250 nt, 약 100 nt 내지 약 300 nt, 약 100 nt 내지 약 350 nt, 약 100 nt 내지 약 400 nt, 약 100 nt 내지 약 500 nt, 약 150 nt 내지 약 200 nt, 약 150 nt 내지 약 250 nt, 약 150 nt 내지 약 300 nt, 약 150 nt 내지 약 350 nt, 약 150 nt 내지 약 400 nt, 약 150 nt 내지 약 500 nt, 약 200 nt 내지 약 250 nt, 약 200 nt 내지 약 300 nt, 약 200 nt 내지 약 350 nt, 약 200 nt 내지 약 400 nt, 약 200 nt 내지 약 500 nt, 약 250 nt 내지 약 300 nt, 약 250 nt 내지 약 350 nt, 약 250 nt 내지 약 400 nt, 약 250 nt 내지 약 500 nt, 약 300 nt 내지 약 350 nt, 약 300 nt 내지 약 400 nt, 약 300 nt 내지 약 500 nt, 약 350 nt 내지 약 400 nt, 약 350 nt 내지 약 500 nt, 또는 약 400 nt 내지 약 500 nt이다. 일부 구체예에서, 이합체화 도메인은 약 50 nt, 약 100 nt, 약 150 nt, 약 200 nt, 약 250 nt, 약 300 nt, 약 350 nt, 약 400 nt, 또는 약 500 nt이다. 일부 구체예에서, 이합체화 도메인은 적어도 약 50 nt, 약 100 nt, 약 150 nt, 약 200 nt, 약 250 nt, 약 300 nt, 약 350 nt, 또는 약 400 nt이다. 일부 구체예에서, 이합체화 도메인은 최대 약 100 nt, 약 150 nt, 약 200 nt, 약 250 nt, 약 300 nt, 약 350 nt, 약 400 nt, 또는 약 500 nt이다.

[0128] 일부 구체예에서, 제1 및 제2 이합체화 도메인 (122 및 154)의 서열은 인 실리코 구조 예측 스크리닝(in silico structure prediction screening)(예를 들면, RNA 폴딩 구조 예측이 가능한 이합체화 도메인 서열의 라이브러리를 스크리닝하기 위해 이용된다; 이합체화 도메인 및 상응하는 안티-이합체화 도메인 모두에서 높은 비율의 쌍을 이루지 않은 뉴클레오티드를 갖는 서열이 선택된다), 저다양성 뉴클레오티드 설계 (예를 들면, 이합체화 도메인은 저다양성 서열, 예를 들면, U 단독, A 단독, C 단독, G 단독, R (G 및 A) 단독, 또는 Y (U 및 C)의 반복 서열의 구간을 포함하도록 설계되고, 상기 서열은 그 자체 위로 폴딩될 수 없다), 또는 경험적 스크리닝 (empirical screening) (예를 들면, 이합체화 도메인 및 상응하는 안티-이합체화 도메인의 라이브러리를 합성하고 최대 재조합 효율에 대해 스크리닝한다)에 의해 결정된다.

[0129] 일부 구체예에서, 제1 및 제2 이합체화 도메인 (122, 154)의 서열은 그들의 대응 부분(counter parts)과 강력한 키싱 루프 상호작용을 형성할 수 있는 상보적 RNA 헤어핀 구조(스템 루프(stem loop)로도 지칭됨)를 포함하도록 설계된다. 일부 구체예에서, 3개 이상의 이합체화 도메인, 예를 들면, 4개 이상, 또는 5개 이상의 이합체화 도메인, 예를 들면, 3, 4, 5, 6, 7, 8, 9 또는 10개의 이합체화 도메인이 이 코딩 서열의 3개 이상의 부분을 연결시키기 위해 이용되는 경우, 키싱 루프가 이용된다 (예를 들면, 도 6e). 키싱 루프의 각 헤어핀 루프 (또는 스템 루프)는 비-상보적 서열의 영역 (예를 들면, 루프를 형성함)에 의해 분리된 적어도 2개의 상보적 서열 (예를 들면, 스템을 형성함)로 구성된다. 일부 구체예에서, 이합체화 도메인은 1개 이상 (예를 들면, 적어도 2, 적어도 3, 적어도 4, 또는 적어도 5개, 예를 들면, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18,

19, 또는 20개)의 루프로 구성될 수 있다. 다수의 루프를 갖는 일부 구체예에서, 상기 루프들 중 전부 또는 일부가 반복될 수 있다. 다수의 루프를 갖는 일부 구체예에서, 상기 루프들 중 전부 또는 일부가 다를 수 있다. 일부 구체예에서, 각각의 상보적 서열은 약 4 내지 100 nt이고, 약 3 내지 20 nt의 루프에 의해 분리된다. 2개의 상보적 서열간 염기쌍 형성은 헬릭스 (또는 스템), 예를 들면, 적어도 4 bp, 적어도 5 bp, 적어도 10 bp, 적어도 20 bp, 적어도 30 bp, 적어도 40 bp, 적어도 50 bp, 적어도 75 bp, 적어도 90 bp, 또는 적어도 100 bp, 예를 들면, 4 내지 100 bp, 5 내지 75 bp, 또는 10 내지 50 bp의 헬릭스 (또는 스템)을 초래한다. 일부 구체예에서, 루프 부분은 적어도 3 nt, 적어도 5 nt, 적어도 10 nt, 적어도 15 nt, 또는 적어도 20 nt, 예를 들면, 3 내지 20 nt, 5 내지 15 nt 또는 5 내지 10 nt이고, 상기 루프는 염기쌍을 형성하지 않는다. 2개의 헤어핀 루프간 상보적 서열은 염기쌍 형성, 및 키싱 루프/키싱 스템 루프 상호작용의 생성을 초래한다. 일부 구체예에서, 2개의 헤어핀 루프간 상보적 서열은 제1 루프의 적어도 3개의 뉴클레오티드와 제2 루프의 적어도 3개의 뉴클레오티드, 예를 들면, 제1 루프의 적어도 4, 적어도 5, 적어도 6, 적어도 7, 적어도 8, 적어도 9, 적어도 10, 적어도 11, 적어도 12, 적어도 13, 적어도 14, 적어도 15, 적어도 16, 적어도 17, 적어도 19, 또는 적어도 20 nt (예를 들면, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20 nt)와 제2 루프의 적어도 4, 적어도 5, 적어도 6, 적어도 7, 적어도 8, 적어도 9, 적어도 10, 적어도 11, 적어도 12, 적어도 13, 적어도 14, 적어도 15, 적어도 16, 적어도 17, 적어도 19, 또는 적어도 20 nt (예를 들면, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20 nt) 간에 일어난다. 일부 구체예에서, 2개의 헤어핀 루프간 상보적 서열은 전체 루프 서열의 적어도 15%, 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 또는 100%에서 일어난다.

[0130] 일부 경우에, 키싱 루프의 스템은 2개의 RNA 분자 간에 트랜스로 염기쌍을 형성하도록 선택된다. 그러한 구체예에서, 하나의 분자에 있는 하나의 헤어핀 루프와 제2 분자에 있는 또 다른 헤어핀 루프의 키싱 루프 상호작용을 형성한 후에, 초기 헤어핀 루프의 개별적인 스템 (또는 헬릭스) 영역은 가닥 치환/침윤(strand replacement/invasion) 및 연장된 듀플렉스 형성(extended duplex formation)을 통해 2개의 RNA 분자 간에 트랜스로 염기쌍을 형성할 수 있다. 일부 구체예에서, 초기 루프 서열 내에서, 최대 85%의 뉴클레오티드가 연장된 듀플렉스 형성 후에 쌍을 이루지 않은 상태로 유지될 수 있다(예를 들면, nt의 약 15%가 2개의 루프 간에 염기쌍을 형성한다). 일부 구체예에서, 키싱 루프는 HIV-1 DIS 루프에 기반하고 (서열번호 139 및 140, 도 17a), 상보적 서열의 6개 뉴클레오티드의 5' 쪽에 2개의 A 뉴클레오티드, 및 3' 쪽에 이어진 1개의 A 뉴클레오티드를 포함한다 (예를 들면, AANNNNNA, 상기 서열 중 N은 A, U, G, 또는 C일 수 있음). 일부 구체예에서, 키싱 루프는 HIV-2 키싱 루프 이합체화 도메인에 기반하고 (서열번호 141 및 142, 도 17b), 상보적 서열의 6개 뉴클레오티드의 5' 쪽에 G 및 A 뉴클레오티드, 및 3' 쪽에 이어진 3개의 A 뉴클레오티드를 포함한다 (예를 들면, GANNNNNNAAA (서열번호 153), 상기 서열 중 N은 A, U, G, 또는 C일 수 있음).

[0131] 하나의 구성에서, 초기 스템에 미스매치(mismatch)의 내포에 의해 연장된 듀플렉스 형성이 선호되어, 연장된 듀플렉스 중 더 높은 비율의 매칭을 초래한다. 따라서, 일부 구체예에서, 헤어핀 루프의 헬릭스 또는 스템 영역은 초기에 쌍을 이루지 않는 염기쌍을 최대 30% 포함한다 (예를 들면, 30% 이하, 20% 이하, 15% 이하, 10% 이하, 5% 이하, 또는 1% 이하, 예를 들면, 1 내지 30%, 5 내지 30%, 10 내지 30%, 또는 25 내지 30%의 염기쌍이 처음에 염기쌍을 형성하지 않는다). 이러한 비-염기쌍 형성의 영역이 벌지(bulge), 미스매치, 또는 내부 루프(internal loop)를 형성할 수 있다.

[0132] 2개의 헤어핀 루프의 상호작용 (키싱 루프 상호작용) 외에, 다른 형태의 루프 상호작용이 제1 및 제2 이합체화 도메인 (122, 154)에 대해 이용될 수 있다. 일 구체예에서, 루프는 벌지이고, 염기쌍 형성 헬릭스의 하나의 가닥이 스템 구조로부터 돌출되는 1개 이상의 뉴클레오티드를 포함한다. 예시적인 벌지는 적어도 1 nt, 적어도 2 nt, 적어도 3 nt, 적어도 4 nt, 적어도 5 nt, 적어도 10 nt 또는 적어도 20 nt, 예를 들면, 1 내지 20 nt, 1 내지 15 nt, 1 내지 10 nt, 또는 5 내지 10 nt, 또는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20 nt이다. 일 구체예에서, 루프는 내부 루프이고, 예를 들면, 헬릭스 중 1개 이상의 뉴클레오티드가 미스매칭되어, 미스매치의 위치에서 내부 루프에 의해 중단된 헬릭스를 초래한다. 일부 구체예에서, 헬릭스는 각각의 가닥 상에서 적어도 1 nt (예를 들면, 각 가닥 상에서 적어도 2 nt, 적어도 3 nt, 적어도 4 nt, 적어도 5 nt, 적어도 10 nt 또는 적어도 20 nt, 예를 들면, 1 내지 20 nt, 1 내지 15 nt, 1 내지 10 nt, 또는 5 내지 10 nt, 또는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 또는 20 nt)인 내부 루프의 양쪽에 적어도 4 nt (예를 들면, 적어도 5 nt, 적어도 10 nt, 적어도 20 nt, 적어도 30 nt, 적어도 40 nt, 적어도 50 nt, 적어도 75 nt, 적어도 90 nt, 또는 적어도 100 nt, 예를 들면, 4 내지 100 nt, 5 내지 75 nt, 또는 10 내지 50 nt, 예를 들면, 4 내지 100 nt)이다. 일 구체예에서, 루프는 다분지(multi-branched) 루프이고, 3개의 헬릭스 또는 스템이 삼각형을 형성하고, 하나 이상의 쌍을 이루지 않은 뉴클레오티

드가 상기 3개의 헬릭스를 연결한다. 일부 구체예에서, 헬릭스 각각은 적어도 4 bp (예를 들면, 적어도 5 bp, 적어도 10 bp, 적어도 20 bp, 적어도 30 bp, 적어도 40 bp, 적어도 50 bp, 적어도 75 bp, 적어도 90bp, 또는 적어도 100 bp, 예를 들면, 4 내지 100 bp, 5 내지 75 bp, 또는 10 내지 50 bp)이고, 상기 삼각형을 형성하는 쌍을 이루지 않은 뉴클레오티드는 적어도 3 nt (예를 들면, 적어도 4 nt, 적어도 5 nt, 적어도 10 nt, 적어도 20, 적어도 15, 적어도 30, 적어도 40, 적어도 50, 또는 적어도 60 nt, 예를 들면, 3 내지 60 nt, 3 내지 30 nt, 3 내지 25 nt, 또는 5 내지 20 nt, 예를 들면, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 2, 25, 30, 35, 40, 45, 50, 55 또는 60 nt)이다. 키싱 상호작용은 이러한 종류의 루프 중 2개 간에(예를 들면, 각각 1개 이상의 헬릭스를 포함하는 것인 2개 이상의 도메인 간에) 일어날 수 있다. 일부 구체예에서, 1개의 이합체화 도메인 (예를 들면, 제1 이합체화 도메인 (122)) 내의 헬릭스는 다른 결합 도메인 (예를 들면, 제2 이합체화 도메인 (154))에 직접적인 대응물을 가져서 초기 루프 키싱 상호작용 후에 연장된 듀플렉스 형성을 가능하게 한다. 일부 구체예에서, 루프를 생성하는 헬릭스를 포함하는 이합체화 도메인은 2개 이상의 이합체화 도메인 (예를 들면, 도 6a의 122, 154) 간 상호작용시 단일 키싱 스템 루프를 형성한다. 일부 구체예에서, 헬릭스를 포함하는 이합체화 도메인은 2개 이상의 이합체화 도메인 (예를 들면, 도 6a의 122, 154) 간 상호작용시 단일 키싱 루프 상호작용을 위한 복수의 루프를 형성한다. 일부 구체예에서, 1개 이상의 이합체화 도메인 (예를 들면, 도 6a의 122)은 벌지, 단일 염기 벌지, 미스매치, 또는 내부 루프, 또는 G-U 워블 쌍의 포함에 의해 불안정해진 헬릭스를 포함하나, 나머지 결합 도메인(예를 들면, 도 6a의 154)과 매칭되어, 초기 키싱/염기쌍 형성 후 연장된 듀플렉스 형성에 알맞다. 일부 구체예에서, 1개 이상의 이합체화 도메인 (예를 들면, 도 6a의 122)은 불안정화된 헬릭스를 포함하고, 이는 안정화되면 (예를 들면, 테오필린 스위치 키싱 루프 (theophylline switch kissing loop)), 루프-루프 상호작용 (예를 들면, 키싱/쌍 형성)을 통해 제2 이합체화 도메인 (예를 들면, 도 6a의 122)과 상호작용할 수 있는 루프를 노출시킨다.

[0133] 일부 구체예에서, 이러한 스템 루프는 길이가 적어도 10 nt, 예를 들면, 적어도 20 nt, 적어도 25 nt, 적어도 50 nt, 적어도 75 nt, 또는 적어도 100 nt, 예를 들면, 길이가 10 내지 50, 20 내지 25, 10 내지 100, 10 내지 20, 또는 20 내지 40 nt를 포함한다. 각각의 이합체화 도메인은 적어도 1개의 개별적인 스템 루프, 예를 들면, 적어도 2개, 적어도 5개, 적어도 10개, 적어도 15개, 또는 적어도 20개, 예를 들면, 1 내지 20개, 2 내지 5개 또는 1 내지 10개의 개별적인 스템 루프를 포함할 수 있다.

[0134] 일부 구체예에서, 코딩 서열의 3 내지 10개의 부분들이 2 내지 9개의 키싱 루프에 의해 연결되고, 예를 들면, 3개의 부분이 2개의 키싱 루프에 의해 연결되고, 4개의 부분이 3개의 키싱 루프에 의해 연결되고, 등이며, 상기 2개 내지 9개의 키싱 루프 각각은 상이하다. 일부 구체예에서, 키싱 루프는 다수의 스템 루프, 예를 들면, 2 내지 20개의 스템 루프를 포함한다. 일부 구체예에서, 상기 키싱 루프 중 다수의 스템 루프 각각은 동일하다. 일부 구체예에서, 상기 키싱 루프 중 다수의 스템 루프 각각은 상이하다. 일부 구체예에서, 이합체화 도메인은 1 내지 20개의 스템 루프를 포함한다. 일부 구체예에서, 이합체화 도메인은 1개의 스템 루프 내지 20개의 스템 루프를 포함한다. 일부 구체예에서, 이합체화 도메인은 1개의 스템 루프 내지 2개의 스템 루프, 1개의 스템 루프 내지 3개의 스템 루프, 1개의 스템 루프 내지 4개의 스템 루프, 1개의 스템 루프 내지 5개의 스템 루프, 1개의 스템 루프 내지 6개의 스템 루프, 1개의 스템 루프 내지 7개의 스템 루프, 1개의 스템 루프 내지 8개의 스템 루프, 1개의 스템 루프 내지 290개의 스템 루프, 1개의 스템 루프 내지 10개의 스템 루프, 1개의 스템 루프 내지 15개의 스템 루프, 1개의 스템 루프 내지 20개의 스템 루프, 2개의 스템 루프 내지 3개의 스템 루프, 2개의 스템 루프 내지 4개의 스템 루프, 2개의 스템 루프 내지 5개의 스템 루프, 2개의 스템 루프 내지 6개의 스템 루프, 2개의 스템 루프 내지 7개의 스템 루프, 2개의 스템 루프 내지 8개의 스템 루프, 2개의 스템 루프 내지 9개의 스템 루프, 2개의 스템 루프 내지 10개의 스템 루프, 2개의 스템 루프 내지 15개의 스템 루프, 2개의 스템 루프 내지 20개의 스템 루프, 3개의 스템 루프 내지 4개의 스템 루프, 3개의 스템 루프 내지 5개의 스템 루프, 3개의 스템 루프 내지 6개의 스템 루프, 3개의 스템 루프 내지 7개의 스템 루프, 3개의 스템 루프 내지 8개의 스템 루프, 3개의 스템 루프 내지 9개의 스템 루프, 3개의 스템 루프 내지 10개의 스템 루프, 3개의 스템 루프 내지 15개의 스템 루프, 3개의 스템 루프 내지 20개의 스템 루프, 4개의 스템 루프 내지 5개의 스템 루프, 4개의 스템 루프 내지 6개의 스템 루프, 4개의 스템 루프 내지 7개의 스템 루프, 4개의 스템 루프 내지 8개의 스템 루프, 4개의 스템 루프 내지 9개의 스템 루프, 4개의 스템 루프 내지 10개의 스템 루프, 4개의 스템 루프 내지 15개의 스템 루프, 4개의 스템 루프 내지 20개의 스템 루프, 5개의 스템 루프 내지 6개의 스템 루프, 5개의 스템 루프 내지 7개의 스템 루프, 5개의 스템 루프 내지 8개의 스템 루프, 5개의 스템 루프 내지 9개의 스템 루프, 5개의 스템 루프 내지 10개의 스템 루프, 5개의 스템 루프 내지 15개의 스템 루프, 5개의 스템 루프 내지 20개의 스템 루프, 6개의 스템 루프 내지 7개의 스템 루프, 6개의 스템 루프 내지 8개의 스템 루프, 6개의 스템 루프 내지 9개의 스템 루프, 6개의 스템 루프 내지 10개의 스템 루프, 6개의 스템 루프 내지 15개의 스템 루프,

6개의 스템 루프 내지 20개의 스템 루프, 7개의 스템 루프 내지 8개의 스템 루프, 7개의 스템 루프 내지 9개의 스템 루프, 7개의 스템 루프 내지 10개의 스템 루프, 7개의 스템 루프 내지 15개의 스템 루프, 7개의 스템 루프 내지 20개의 스템 루프, 8개의 스템 루프 내지 9개의 스템 루프, 8개의 스템 루프 내지 10개의 스템 루프, 8개의 스템 루프 내지 15개의 스템 루프, 8개의 스템 루프 내지 20개의 스템 루프, 9개의 스템 루프 내지 10개의 스템 루프, 9개의 스템 루프 내지 15개의 스템 루프, 9개의 스템 루프 내지 20개의 스템 루프, 10개의 스템 루프 내지 15개의 스템 루프, 10개의 스템 루프 내지 20개의 스템 루프, 또는 15개의 스템 루프 내지 20개의 스템 루프를 포함한다. 일부 구체예에서, 이합체화 도메인은 1개의 스템 루프, 2개의 스템 루프, 3개의 스템 루프, 4개의 스템 루프, 5개의 스템 루프, 6개의 스템 루프, 7개의 스템 루프, 8개의 스템 루프, 9개의 스템 루프, 10개의 스템 루프, 15개의 스템 루프, 또는 20개의 스템 루프를 포함한다. 일부 구체예에서, 이합체화 도메인은 적어도 1개의 스템 루프, 2개의 스템 루프, 3개의 스템 루프, 4개의 스템 루프, 5개의 스템 루프, 6개의 스템 루프, 7개의 스템 루프, 8개의 스템 루프, 9개의 스템 루프, 10개의 스템 루프, 또는 15개의 스템 루프를 포함한다. 일부 구체예에서, 이합체화 도메인은 최대 2개의 스템 루프, 3개의 스템 루프, 4개의 스템 루프, 5개의 스템 루프, 6개의 스템 루프, 7개의 스템 루프, 8개의 스템 루프, 9개의 스템 루프, 10개의 스템 루프, 15개의 스템 루프, 또는 20개의 스템 루프를 포함한다.

[0135] 2개 이상의 이합체화 도메인(예를 들면, 도 6a의 122, 154)이 코딩 서열의 재조합이 일어나기에 충분하게 상호 간에 결합하거나 상호작용할 수 있게 하는 다른 메카니즘이 이용될 수 있다. 일부 구체예에서, 2개 이상의 이합체화 도메인(예를 들면, 도 6a의 122, 154)은 상호 간에, 예를 들면, 비-염기쌍 형성 상호작용을 통해 상호작용할 수 있거나, 공통 분자(예를 들면, 단백질, ATP, 금속 이온, 보조인자, 또는 합성 리간드)에 결합할 수 있는 핵산 압타머 (예를 들면, RNA 압타머)이다. 일부 구체예에서, 2개 이상의 이합체화 도메인(예를 들면, 도 6a의 122, 154)은 상호 간에 혼성화하지 않으나, 동일한 브릿지 핵산 분자에 모두 혼성화될 수 있다. 일부 구체예에서, 그러한 브릿지 핵산 분자는 세포, 조직 또는 개체에 외인적으로 제공될 수 있다. 일부 구체예에서, 그러한 브릿지 핵산 분자는 세포 내에 DNA 또는 RNA 서열, 예를 들면, 전사물 또는 게놈 유전자좌(genomic locus)일 수 있다. 일부 구체예에서, 2개 이상의 이합체화 도메인(예를 들면, 도 6a의 122, 154)은 예를 들면, 비-염기쌍 형성 상호작용을 통해, 상호 간에 상호작용할 수 있는 서열이다.

[0136] 분자 (150)는 3'-위치 분자이고, SA (splice acceptor) (162) 및 제2 이합체화 도메인 (154)을 포함한다. 분자 (150)가 DNA인 것인 구현예에서, 분자는 인트론 서열 (170)로 이어지는 제2 프로모터 (152)를 포함한다. 프로모터 (152)는 인트론 서열 (170)에 작동가능하게 연결될 수 있다. 임의의 프로모터 (152), 예를 들면, 항시적 또는 유도성 프로모터가 이용될 수 있다. 일부 구체예에서, 프로모터 (152)는 조직-특이적 프로모터, 예를 들면, 근육 조직 (예를 들면, 골격근 또는 심장근), 눈 조직 (예를 들면, 망막 조직), 내이 조직, 간 조직, 췌장 조직, 폐 조직, 피부 조직, 골 조직, 또는 신장 조직에서 항시적으로 활성인 프로모터이다. 일부 구체예에서, 프로모터 (112)는 세포-특이적 프로모터, 예를 들면, 암 세포 또는 정상 세포에서 항시적으로 활성인 프로모터이다. 일부 구체예에서, 프로모터 (112)는 발현되는 표적 단백질의 내생적 프로모터이고, 일부 구체예에서, 길다 (예를 들면, 적어도 2500 nt, 적어도 3000 nt, 적어도 4000 nt, 적어도 5000 nt, 또는 적어도 7500 nt). 일부 구체예에서, 프로모터 (112)는 길이가 적어도 약 50 nt, 예를 들면, 적어도 100, 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 적어도 8000 nt, 적어도 9000 nt, 또는 적어도 10,000 nt, 예를 들면, 50 내지 10,000 nt, 100 내지 5000 nt, 500 내지 5000 nt, 또는 50 내지 1000 nt 길이이다. 일부 구체예에서, 프로모터 (112) 및 프로모터 (152)는 동일한 프로모터이다. 다른 구체예에서, 프로모터 (112) 및 프로모터 (152)는 상이한 프로모터이다. 일부 구체예에서, 분자 (150)은 DNA이고, 길이가 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 또는 적어도 8000 nt, 예를 들면, 200 내지 10,000 nt, 200 내지 8000 nt, 500 내지 5000 nt, 또는 200 내지 1000 nt이다. 도 6f에 도시된 바와 같이, 분자 (150)가 RNA인 것인 구현예에서, 예를 들면, DNA가 RNA로 발현된 후, 분자 (150)는 더 이상 프로모터 (152)를 포함하지 않고, 164는 표적 단백질의 C-말단 부분에 대한 코딩 서열에 의해 코딩되는 RNA이다. 일부 구체예에서, 분자 (150)는 RNA이고, 프로모터 (152)를 포함하지 않고, 길이가 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 또는 적어도 8000 nt, 예를 들면, 200 내지 10,000 nt, 200 내지 8000 nt, 500 내지 5000 nt, 또는 200 내지 1000 nt이다. (프로모터 (152)를 갖거나 또는 갖지 않는) 분자 (150)는 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다.

[0137] 인트론 서열 (170)은 제2 이합체화 도메인 (154), 선택적 ISE (156), 분지점 (158), 폴리피리미딘 트랙트 (160), 및 뒤이은 스플라이스 억셉터 서열 (162)을 포함한다. 일부 구체예에서, 인트론 서열 (130)은 길이가 약

적어도 10 nt, 예를 들면, 적어도 20 nt, 적어도 30 nt, 적어도 50 nt, 적어도 100 nt, 적어도 250 nt, 적어도 250 nt, 적어도 300 nt, 적어도 400 nt, 또는 적어도 500 nt, 예를 들면, 길이가 20 내지 500, 20 내지 250, 20 내지 100, 50 내지 100, 30 내지 500, 또는 50 내지 200 nt이다.

[0138] 제2 이합체화 도메인 (154)은 분자 (110)의 제1 이합체화 도메인 (122)의 역 상보체(reverse complement)인 서열을 갖는다. 따라서, 앞서 검토된 제1 이합체화 도메인 (122)의 동일한 설계 특징 및 고려사항이 제2 이합체화 도메인 (154)에도 적용된다. 예를 들면, 일부 구체예에서, 제2 이합체화 도메인 (154)은 제1 이합체화 도메인 (122)과 키싱 루프 상호작용을 형성할 수 있다. 일부 구체예에서, 제2 이합체화 도메인 (154)은 RNA 재조합과 경쟁할 수 있는 크립틱 스플라이스 억셉터 (예를 들면, NNNAGGUNNN; 서열번호 143)를 포함하지 않는다. 일부 구체예에서, 이합체화 도메인 (154)은 저다양성 서열을 갖는다. 일부 구체예에서, 제2 이합체화 도메인 (154)은 1000 nt 이하, 예를 들면, 750 nt 이하, 또는 500 nt 초과, 예를 들면, 30 내지 1000 nt, 30 내지 750 nt, 30 내지 500 nt, 50 내지 500 nt, 50 내지 100 nt, 또는 100 내지 250 nt이다. 일부 구체예에서, 제2 이합체화 도메인 (154)은 50 nt보다 크고, 예를 들면, 적어도 51 nt, 적어도 100 nt, 적어도 150 nt, 적어도 161 nt, 또는 적어도 170 nt, 예를 들면, 51 내지 159 nt, 51 내지 150 nt, 51 내지 120 nt, 51 내지 100 nt, 또는 51 내지 70 nt이다. 일부 구체예에서, 제2 이합체화 도메인 (154)은 160 nt보다 크고, 예를 들면, 적어도 161 nt, 적어도 170 nt, 적어도 180 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 적어도 600 nt, 적어도 700 nt, 적어도 800 nt, 적어도 900 nt, 또는 적어도 1000 nt, 예를 들면, 161 내지 100 nt, 161 내지 500 nt, 161 내지 300 nt, 161 내지 200 nt, 또는 161 내지 170 nt이다. 일부 구체예에서, 제2 이합체화 도메인 (154)은 50 nt 미만, 예를 들면, 6 내지 49 nt, 6 내지 45 nt, 6 내지 40 nt, 6 내지 30 nt, 6 내지 20 nt, 또는 6 내지 10 nt이다.

[0139] 제2 이합체화 도메인 (154)의 3'-쪽은 선택적 ISE (156), 분지점 (158) (예를 들면, 분지점 컨센서스 서열), 폴리피리미딘 트랙트 (160), 및 스플라이스 억셉터 서열 (162)로 이어진다. 분자 (110)의 ISE (120) 및 DISE (118) 처럼, ISE (156)는 스플라이세오솜이 재조합 반응을 촉매하도록 촉진한다. 일부 구체예에서, 인트론 서열 (150)은 적어도 2개의 ISE (156), 예를 들면, 적어도 3개, 적어도 4개, 또는 적어도 5개의 ISE (156)를 포함한다. 예시적 스플라이싱 인핸서 서열은 ISE (156)를 포함한다. 일부 구체예에서, 인트론 서열 (150) 중 1개 이상의 스플라이싱 인핸서 서열 (156)의 포함은 재조합 또는 스플라이싱 효율을 적어도 10%, 적어도 20%, 적어도 30%, 적어도 40%, 또는 적어도 50% 증가시킨다. 이용될 수 있는 예시적 스플라이싱 인핸서 서열이 서열번호 26-136, 151, 및 152, GGGTTT, GGTGGT, TTTGGG, GAGGGG, GGTATT, GTAACG, GGGGGTAGG, GGAGGGTTT, GGGTGGTGT TTCAT, CCATTT, TTTTAAA, TGCAT, TGCATG, TGTGTT, CTAAC, TCTCT, TCTGT, TCTTT, TGCATG, CTAAC, CTGCT, TAACC, AGCTT, TTCATTA, GTTAG, TTTTGC, ACTAAT, ATGTTT, CTCTG, GGG, GGG(N)2-4GGG, TGGG, YCAY, UGCAUG, 또는 3x($G_{3-6}N_{1-7}$)로 제공된다. 일부 구체예에서, ISE (156)가 존재하는 경우, 길이가 약 적어도 3 nt, 적어도 4 nt, 적어도 5 nt, 적어도 10 nt, 예를 들면, 적어도 20 nt, 적어도 25 nt, 적어도 30 nt, 적어도 40 nt, 또는 적어도 50 nt, 예를 들면, 길이가 3 내지 10, 3 내지 11, 4 내지 11, 5 내지 11, 10 내지 50, 20 내지 25, 10 내지 25, 10 내지 20, 또는 20 내지 40 nt일 수 있다. 일 구체예에서, ISE (156)의 서열은 GGCUGAGGGAAGGACUGUCCUGGG (서열번호 135), GGGUUAUGGGACC (서열번호 136), TTCAT, CCATTT, TTTTAAA, TGCAT, TGCATG, TGTGTT, CTAAC, TCTCT, TCTGT, 또는 TCTTT이거나, 이를 포함한다. 일부 구체예에서, ISE (120) 및 ISE (156)은 동일한 서열이다. 다른 구체예에서, ISE (120) 및 ISE (156)은 상이한 서열이다

[0140] 제2 이합체화 도메인 (154) (및 존재하는 경우, ISE (156))에 대한 3'은 분지점 서열 (158) (예를 들면, 분지점 컨센서스 서열), 폴리피리미딘 트랙트 (160), 및 뒤이은 스플라이스 억셉터 서열 (162) (예를 들면, 스플라이스 억셉터 컨센서스 서열)이다. 분지점 (158)의 서열은 표적 세포 또는 개체의 종의 컨센서스 서열에 기반한다. 예를 들면, 인간 스플라이싱의 경우, 컨센서스 서열은 YUNAY를 포함하거나, YUNAY일 수 있다. 따라서, 사용되는 서열은 U2-의존성 인트론의 경우, CUAAC일 수 있거나, 또는 U12-의존성 인트론의 경우, UUUUCCUUAACU (서열번호 144)일 수 있다.

[0141] 폴리피리미딘 트랙트 (160)는 C, U, 또는 C 및 U 뉴클레오티드, 예를 들면, n+y는 10개의 뉴클레오티드이거나 그보다 큰 것인 CnUy를 포함할 수 있고, 3'-스플라이스 연결부로부터 -3 내지 -22에 뉴클레오티드를 포함할 수 있다. 일부 구체예에서, 폴리피리미딘 트랙트 (160)는 적어도 80%의 Y 뉴클레오티드 (즉, U, C, 또는 U 및 C)를 포함한다. 일부 구체예에서, 폴리피리미딘 트랙트 (160)는 폴리C 또는 폴리U 서열이다. 일부 구체예에서, 폴리피리미딘 트랙트 (160)는 적어도 15개의 U, 예를 들면, 15 내지 30개, 또는 15개 내지 20개의 U의 폴리U 서열이다. 분지점 (158) 및 폴리피리미딘 트랙트 (160)는 필수적인 스플라이싱 성분이다. SA (162)의 서열은 표적 세포 또는 개체의 종의 컨센서스 서열에 기반할 수 있다. 예를 들면, 인간에서, SA 서열은 U2-의존성 인트론의 경

우 3'-스플라이스 부위로부터 -1 및 -2 위치에 있는 AG일 수 있고, U12-의존성 인트론의 경우, AC 또는 AG일 수 있다. 따라서, 일부 구체예에서, SA (162)는 AG 또는 AC와 같이 길이가 2 nt일 수 있다.

[0142] SA (162) 바로 다음이 그의 5' 말단에 스플라이스 연결부를 갖는 표적 단백질 (164)의 C-말단 부분을 코딩하는 DNA 서열을 포함하는 엑손 서열이다. 표적 단백질의 C-말단 부분을 코딩하는 DNA 서열 (164)의 5' 말단에 있는 스플라이스 연결부는 분자 (110, 150)가 도입되는 표적 세포 또는 개체에서 발견되는 컨센서스 서열에 매칭될 수 있다. 일부 구체예에서, 스플라이스 연결부는 U12-의존성 인트론의 경우, 3' 스플라이스 부위의 +1 및 +2 위치에 있는 GA 또는 GU, 또는 U12-의존성 인트론의 경우, AU일 수 있다. 따라서, 일부 구체예에서, 스플라이스 연결부는 길이가 2 nt이고, C-말단 코딩 부분 (164)의 5' 말단은 GA, GU, 또는 AU이다.

[0143] 분자 (150)의 인트론 부분 (170) 다음의 엑손 서열은 표적 단백질의 제2 코딩 부분 (예를 들면, 절반(half)), 예를 들면, C 말단 단편 (164) 및 선택적 폴리아데닐화 서열 (166)을 포함한다. 따라서, 분자 (150)는 표적 단백질의 C-말단 부분을 코딩하는 서열 (164)을 포함한다. 분자 (150)의 3'-말단은 선택적으로 폴리아데닐화 서열 (166)을 포함하고, 이는 스플라이세오솜의 조립을 촉진한다. 일부 구체예에서, 폴리아데닐화 서열 (166)은 적어도 15개의 A, 예를 들면, 15 내지 30개 또는 15 내지 20개의 A의 폴리A 서열이다. 일부 구체예에서, 폴리아데닐화 서열 (166) 및 폴리아데닐화 서열 (124)는 동일한 서열이다. 다른 구체예에서, 폴리아데닐화 서열 (166) 및 폴리아데닐화 서열 (124)는 상이한 서열이다.

[0144] 일부 구체예에서, N-말단 코딩 영역 (114) 및/또는 C-말단 코딩 영역 (164)은 원형의 코딩 서열이다. 예를 들면, 상기 코딩 서열은 본 명세서에서 개시된 시스템이 도입되는 세포 또는 개체에서 발견된다 (예를 들면, 인간 세포 또는 개체에 도입되는 경우, 인간 코딩 서열). 일부 구체예에서, N-말단 코딩 영역 (114) 및/또는 C-말단 코딩 영역 (164)은 원형의 코딩 서열 대비, 예를 들면, tRNA 이용률을 최대화하거나 또는 크립틱 스플라이스 부위에 대해 탈-농축시키기 위해 (예를 들면,부정확한 스플라이싱을 감소시키거나 방지하고, 정확한 연결부 형성을 촉진하기 위해), 코돈 최적화된다. 일부 구체예에서, N-말단 코딩 영역 (114) 및/또는 C-말단 코딩 영역 (164)의 부분은 원형의 코딩 서열 대비 코돈 최적화되고, 예를 들면, 각 연결부에 인접한 약 200 nt (예를 들면, 114의 3'-말단, 및 164의 5'-말단)가 코돈 최적화되거나, 또는 (SR 단백질에 결합하는) ESE(exonic splice enhancer) 부위를 포함하도록 변경될 수 있다. 예를 들면, 상기 코딩 서열은 본 명세서에서 개시된 시스템이 도입되는 세포 또는 개체에서 발견되지 않는 것일 수 있다 (예를 들면, 마우스 세포 또는 개체에 도입되는 경우, 인간 코딩 서열).

[0145] 일부 구체예에서, N-말단 코딩 영역 (114) 및/또는 C-말단 코딩 영역 (164)은 속성상 천연 또는 합성이고, 스플라이스 도너 및 억셉터 부위를 포함하는 인트론을 포함한다. 예를 들면, 발현될 코딩 서열 내부에 임베딩된 인트론은 N-말단 코딩 영역 (114)의 내부에, 서열 (116)의 상류 (예를 들면, 상류 약 200 nt)에 포함될 수 있거나, 발현될 코딩 서열 내부에 임베딩된 인트론은 C-말단 코딩 영역 (164)의 내부에, 서열 (162)의 하류 (예를 들면, 하류 약 200 nt)에 포함될 수 있거나, 또는 둘 모두일 수 있다. 그러한 인트론의 포함은 스플라이싱 기구의 트랜스-스플라이싱 인트론 도너 및 억셉터로의 부착을 촉진하기 위해 이용될 수 있다. 일부 구체예에서, 그러한 (촉진성-) 인트론은 110 및 150이 발현되는 숙주로부터 유래될 수 있다. 일부 구체예에서, s그러한 (촉진성-) 인트론은 기타 개체로부터 유래되거나, 바이러스 기원이거나 또는 합성 기원일 수 있다.

[0146] 일부 구체예에서, 분자 (150)를 안정화시키는 서열의 포함(예를 들면, 도 6a에서 150의 3' 비번역 영역 중 164와 166 사이에 배치됨)은 재조합 산물의 발현 효율을 적어도 25%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 또는 적어도 75%, 예를 들면, 25 내지 95%, 25 내지 75%, 25 내지 60%, 25 내지 50%, 40 내지 95%, 40 내지 60%, 또는 50 내지 60% 증가시킬 수 있다. 일부 구체예에서, WPRE(woodchuck post-transcriptional regulatory element) 또는 그의 절단물(예를 들면, WPRE3)이 재조합 산물 발현 효율을 증가시키기 위한 안정화 요소로서 3'-UTR에 포함된다. 일부 구체예에서, WPRE 서열은 GenBank accession no. J04514의 nt 1093 내지 1684, 또는 WPRE3의 247 bp 서열에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 또는 100% 서열 동일성을 갖는다.

[0147] 도 6c에 도시된 바와 같이, 분자 (110)의 제1 이합체화 도메인 (122)과 분자 (150)의 제2 이합체화 도메인 (154) 간의 상호작용 및 혼성화 (염기쌍 형성)가 스플라이세오솜 성분이 N-말단 코딩 서열 (114)과 C-말단 코딩 서열 (164)을 재조합할 수 있게 한다. 구체적으로, 두 부분 간 무결절성 연결부로서 N-말단 단백질 코딩 서열 (114)의 3' 말단이 C-말단 단백질 서열 (164)의 5' 말단에 융합된다.

[0148] 도 6d는 표적 단백질이 3개의 부분, N-말단, 중간, 및 C-말단 부분으로 나뉜 것인 시스템의 개략도를 보여준다 (각각의 부분은 크기가 유사하거나 다를 수 있음). 당업자는 단백질이 임의의 갯수의 원하는 세그먼트 또는 부

분으로 나뉠 수 있고, 본 명세서에서 제공되는 정보를 이용하여, 적절한 갯수의 분자가 설계될 수 있다는 것을 이해할 것이다. 그러한 구체예에서, 상기 시스템은 적어도 3개의 합성 핵산 분자(110, 200, 및 150)를 포함하고, 분자 (110)는 상기 단백질의 N-말단 부분을 코딩하는 분자 (114)를 포함하고, 분자 (200)는 상기 단백질의 중간 부분을 코딩하는 분자 (216)를 포함하며, 분자 (150)는 상기 단백질의 C-말단 부분을 코딩하는 분자 (164)를 포함한다. 각각의 핵산 분자 (110, 200, 150)는 DNA로 구성될 수 있고, 전사 후에, 프로모터 (112, 202, 152)가 없는 RNA일 수 있다. 일부 구체예에서, 각각의 110, 200, 150 (프로모터 (112, 202, 152) 포함 또는 불포함)은 적어도 약 100 nt(nucleotides/ribonucleotides) 길이, 예를 들면, 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 또는 적어도 8000 nt 길이, 예를 들면, 200 내지 10,000 nt, 200 내지 8000 nt, 500 내지 5000 nt, 또는 200 내지 1000 nt 길이이다. 분자 (110, 150, 200) (프로모터 (112, 202, 152) 포함 또는 불포함)는 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다. 2개 (또는 그 이상)의 직교(orthogonal) 이합체화 도메인을 이용하는 것에 더해, 2개의 인트론 중 하나는 U2-타입 인트론일 수 있고, 제2 인트론은 U12-타입 인트론일 수 있다. U2 및 U12 의존성 인트론의 스플라이스 도너 및 억셉터는 두 타입의 인트론간 컨센서스 인식 서열이 다르기 때문에 최소 교차 반응성을 보인다. 두 전략 (즉, 직교 이합체화 도메인, 및 U2 vs. U12 타입 인트론) 모두 (예를 들면, 제1 단편이 마지막 단편에 직접 연결되는 것을 방지하고, 중간 단편이 그 자체로 고리화되는 것을 방지하기 위해) 정확한 순서로 3개의 단편의 재조합을 촉진한다.

[0149] 도 6d의 분자 (110)는 도 1a에 대해 전술된 동일한 특징들을 포함하고, 즉, RNA 분자를 코딩하는 서열에 작동가능하게 연결된 프로모터 (112)로서, 상기 RNA 분자는 5'에서 3'으로: 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열 (114)로서, 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열 (114)은 상기 표적 단백질 코딩 서열의 3'-말단의 스플라이스 연결부, SD (116), 선택적인 DISE (118), 선택적인 ISE (120), 이합체화 도메인 (122), 및 선택적인 폴리아데닐화 서열 (124)을 포함하나, 제1 이합체화 도메인 (122)은 분자 (200)의 제3 이합체화 도메인에 역 상보성을 갖는 것인 프로모터를 포함한다. 도 6f에 도시된 바와 같이, 분자 (110)가 RNA, 예를 들면, DNA의 RNA로의 전사 후 RNA인 것인 구체예에서, 분자 (110)는 프로모터 (112)를 포함하지 않고, 114는 표적 단백질의 N-말단 부분에 대한 코딩 서열에 의해 코딩되는 RNA이다. (프로모터 (112)를 갖거나 또는 갖지 않는) 분자 (110)는 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다.

[0150] 도 6d의 분자 (150)는 도 1a에 대해 전술된 동일한 특징들을 포함하고, 즉, RNA 분자를 코딩하는 서열에 작동가능하게 연결된 프로모터 (152)로서, 상기 RNA 분자는 5'에서 3'으로: 제2 이합체화 도메인 (154), 선택적 ISE (156), 분지점 서열 (158), 폴리피리미딘 트랙트 (160), 스플라이스 억셉터 서열 (162); 및 표적 단백질의 C-말단 부분에 대한 코딩 서열 (164)를 포함하고, 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열은 상기 표적 단백질 코딩 서열의 5'-말단의 스플라이스 연결부, 및 선택적인 폴리아데닐화 서열 (166)을 포함하는 것인 프로모터를 포함한다. 상기 제2 이합체화 도메인 (154)은 분자 (200)의 제4 이합체화 도메인 (226)에 대한 역 상보성을 갖는다. (프로모터 (152)를 갖거나 또는 갖지 않는) 분자 (150)는 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다.

[0151] 분자 (200)는 각각 분자 (110) 및 분자 (150)의 이합체화 도메인 (122, 154)에 대해 역 상보성을 갖는 이합체화 도메인을 제공하는 것에 의해 N- 및 C-말단 코딩 영역 (114, 164)의 연결을 가능하게 한다. 분자 (200)는 2개의 인트론 서열 (230, 240)을 포함한, 분자 (110) 및 분자 (150)으로부터의 특징들을 포함한다. 구체적으로, 분자 (200)이 DNA인 것인 구현예에서, 분자 (220)는 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 프로모터 (210)(프로모터 (112) 및/또는 프로모터 (152)와 동일하거나 또는 다른 수 있음)를 포함하고, 상기 RNA 분자는 5'에서 3'으로: 제3 이합체화 도메인 (204) (도 6d에서 분자 (110)의 제1 이합체화 도메인 (122)에 대한 역 상보체임), 선택적인 ISE (206), 분지점 (208), 폴리피리미딘 트랙트 (210), SA (212), 표적 단백질의 중간 부분에 대한 코딩 서열 (216)을 포함하고, 상기 표적 단백질의 중간 부분에 대한 코딩 서열 (216)은 상기 표적 단백질 코딩 서열의 5'-말단의 스플라이스 연결부 및 상기 표적 단백질 코딩 서열의 3'-말단의 스플라이스 연결부, SD (220), 선택적인 DISE (222), 선택적인 ISE (224), 제4 이합체화 도메인 (226) (도 6d에서 분자 (150)의 제4 이합체화 도메인 (154)에 대한 역 상보체임), 및 선택적인 폴리아데닐화 서열 (228)을 포함한다. 일부 구체예에서, 분자 (220)는 DNA이고, 길이가 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 또는 적어도 8000 nt, 예를 들면, 200 내지 10,000 nt, 200 내지 8000 nt, 500 내지 5000 nt, 또는 200 내지 1000 nt이다. 분자 (200)가 RNA, 예를 들면, DNA의 RNA로의 발현 후 RNA인 것인 구현예에서, 분자 (200)는 더 이상 프로모터 (202)를 포함하지 않고, 216은 표적 단백질의 중간 부분에 대한 코딩 서열에 의해 코딩되는 RNA이다. 일부 구체예에서, 분자 (200)는 RNA이고, 프로모터 (202)를 포함하지 않고, 길이가 적어도 200, 적어도 300, 적어도 500, 적어도 1000, 적어도 2000, 적

어도 3000, 적어도 4000, 적어도 5000, 적어도 6000, 적어도 7000, 또는 적어도 8000 nt, 예를 들면, 200 내지 10,000 nt, 200 내지 8000 nt, 500 내지 5000 nt, 또는 200 내지 1000 nt이다. (프로모터 (202)를 갖거나 또는 갖지 않는) 분자 (200)는 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다.

[0152] 도 6e에 도시된 바와 같이, 분자 (110)의 제1 이합체화 도메인 (122)과 분자 (200)의 제3 이합체화 도메인 (204) 간의 상호작용 및 혼성화 (염기쌍 형성), 및 분자 (200)의 제4 이합체화 도메인 (226)과 분자 (150)의 제2 이합체화 도메인 (154) 간의 상호작용 및 혼성화 (염기쌍 형성)가 스플라이세오솜 성분이 N-말단 코딩 서열 (114), 중간 코딩 서열 (216), 및 C-말단 코딩 서열 (164)을 재조합할 수 있게 한다. 구체적으로, 세 부분 간 무결절성 연결부로서, N-말단 단백질 코딩 서열 (114)의 3' 말단이 중간 단백질 서열 (216)의 5' 말단에 융합되고, 중간 단백질 서열 (216)의 3' 말단이 C-말단 단백질 서열 (164)의 5' 말단에 융합된다.

[0153] 대안적인 이합체화 도메인이 도 7a-7b 및 9a에 도시된다. 즉, 상호 간에 혼성화되는 이합체화 도메인을 이용하는 것에 대한 대안(예를 들면, 112 내지 204, 226 내지 154, 도 6d, 6e)으로, 일 구체예에서, 압타머 서열이 이용된다. 도 7a에 도시된 바와 같이, 합성 핵산 분자 (500, 600) 모두에서, 이합체화 도메인 대신에 압타머 서열 (512, 602)이 이용되고, 상기 압타머들은 그들의 표적(예를 들면, 아데노신, 도파민, 또는 카페인)과의 상호작용을 통해 모인다. 그러한 구체예에서, 각 분자 (500, 600)의 압타머 서열 (512, 602)은 동일한 서열일 수 있거나 또는 상이한 서열일 수 있다. 도 7a의 분자 (500)는 도 6a의 분자 (110)에 대해 전술된 것과 동일한 특징들을 포함하고, 즉, DNA가 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 프로모터 (152)로서, 상기 RNA 분자는 5'에서 3'으로: 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열 (502)로서, 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열 (502)은 상기 표적 단백질 코딩 서열의 3'-말단의 스플라이스 연결부, SD (506), 선택적인 DISE (508), 선택적인 ISE (510), 제1 이합체화 도메인 대신인 제1 압타머 (512), 및 선택적인 폴리아데닐화 서열 (124)을 포함하는 것인 프로모터를 포함한다. 분자 (500)가 RNA인, 예를 들면, 상기 DNA 분자로부터 전사된 RNA인 것인 구현예에서, 분자 (500)는 프로모터를 포함하지 않는다 (예를 들면, 도 7a에 도시됨). 마찬가지로, 도 7a의 분자 (600)는 도 6a의 분자 (150)에 대해 전술된 것과 동일한 특징을 포함하고, 즉, DNA가 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 프로모터를 포함하고, 상기 RNA 분자는 5'에서 3'으로: 제2 이합체화 도메인 (154) 대신인 제2 압타머 (602), 선택적인 ISE (604), 분지점 (606), 폴리피리미딘 트랙트 (608), SA (610), 그의 5'-말단에 스플라이스 연결부를 갖는 표적 단백질의 C-말단 부분을 코딩하는 DNA, 및 선택적인 폴리아데닐화 서열 (616)을 포함한다. 분자 (600)가 RNA인, 예를 들면, 상기 DNA 분자로부터 전사된 RNA인 것인 구현예에서, 분자 (500)는 프로모터를 포함하지 않는다 (예를 들면, 도 7a에 도시됨). 2개의 압타머 (512, 602)의 상호간 또는 분자 (700)와의 상호작용이 스플라이세오솜 성분이 N-말단 코딩 서열 (502)과 C-말단 코딩 서열 (614)을 재조합할 수 있게 한다. 구체적으로, 두 부분 간 무결절성 연결부로서, N-말단 단백질 코딩 서열 (502)의 3' 말단이 C-말단 단백질 서열 (614)의 5' 말단에 융합된다. 분자 (500 및 600)는 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다.

[0154] 일부 구체예에서, 압타머 서열 (512, 602)은 동일한 표적 (700)을 인식할 수 있거나 (예를 들면, 특이적으로 결합하거나)(도 7a), 또는 상이한 표적을 인식할 수 있다 (합성 분자는 또한 본 명세서에서 제공되는 시스템과 함께 투여되고, 상기 시스템은 각각의 압타머에 의해 특이적으로 인식되는 각 분자, 또는 상기 압타머에 의해 인식되는 분자의 부분, 예를 들면, 카페인/도파민 하이브리드 분자를 포함한다). 압타머에 의해 인식되는 예시적인 표적은 세포 단백질, 소분자, 외래 단백질, 또는 RNA 분자를 포함한다.

[0155] 도 7b는 도 7a에 유사한 구체예를 보여준다. 이합체화 도메인(도 7a의 512, 602)은 RNA 분자를 인식한다. 도 7b에 도시된 구체예에서, 각 도메인은 표적 세포 (표적 단백질 발현이 바람직한 것인 세포)에서만 발현되는 mRNA 분자의 상이한 부분, 예를 들면, 암-특이적 전사물을 인식한다. 그러한 구체예에서, RNA에 포함된 코딩 서열 (도 7a의 502, 614)은 이합체화 도메인에 의해 인식되는 특이적 RNA 분자의 존재시에만 재조합된다. 이때, 표적 단백질은 암 세포에서만 발현되고, 정상 세포에서는 발현되지 않을 것이다. 그러한 시스템은 암 세포에서 표적 단백질 (예를 들면, 암에 대한 치료 단백질, 예를 들면, 독소, 또는 세포독성 효소, 예를 들면, 간시클로비르를 동반한 티미딘 키나아제; 따라서, 일부 구체예에서, 표적 단백질은 독소 또는 티미딘 키나아제임) 발현의 조절을 가능하게 하여, 정상의 비-암 세포에서 표적 단백질 발현의 원치않는 효과를 감소시킨다.

[0156] 도 7c는 예시적인 "오프-스위치(off-switch)" 구체예를 제공한다. 여기에서, 합성 핵산 분자 (800, 900)의 이합체화 도메인 (812, 902)(상호 간에 역 상보체임)의 혼성화/결합이 상기 결합/혼성화와 경쟁하는 안티-결합 도메인 올리고뉴클레오티드 (예를 들면, RNA 또는 DNA) (1000) (2개의 상이한 안티-결합 도메인 올리고뉴클레오티드 (1000)일 수 있고, 하나는 812의 역 상보체이고, 하나는 912의 역 상보체임)를 제공하는 것에 의해 감소될 수 있다. 따라서, 안티-결합 도메인 올리고뉴클레오티드 (1000)는 각각 N- 및 C-말단 코딩 부분 (802 및 914)에 의

해 코딩되는 단백질의 재구성에 대한 "오프-스위치"로 작용할 수 있다. 도 7c의 분자 (800)는 RNA 분자인(따라서, 프로모터가 없음) 도 6a의 분자 (110)에 대해 전술된 동일한 특징들을 포함하고, 상기 RNA 분자는 5'에서 3'으로: 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열 (802)로서, 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열 (802)은 상기 표적 단백질 코딩 서열의 3'-말단의 스플라이스 연결부, SD (806), 선택적인 DISE (808), 선택적인 ISE (810), 이합체화 도메인 (812), 및 선택적인 폴리아데닐화 서열 (814)을 포함한다. 마찬가지로, 도 7c의 분자 (900)는 RNA 분자인 (따라서, 프로모터가 없음) 도 6a의 분자 (150)에 대해 전술된 동일한 특징들을 포함하고, 상기 RNA 분자는 5'에서 3'으로: 안티-이합체화 도메인 (902), 선택적인 ISE (904), 분지점 (906), 폴리피리미딘 트랙트 (908), SA (910), 표적 단백질의 C-말단 부분을 코딩하는 RNA (914), 및 선택적인 폴리아데닐화 서열 (916)을 포함한다. 2개의 이합체화 도메인 (812, 902)은 안티-결합 도메인 올리고뉴클레오티드 (1000)의 존재시 상호 간에 상호작용/혼성화할 수 없고, 따라서, N-말단 코딩 서열 (802)과 C-말단 코딩 서열 (914)의 재조합을 방지하거나 감소시킬 수 있다. 그러한 적용이 상기 시스템에 의해 코딩되는 단백질의 발현을 감소시키거나 또는 제거하기 위해 이용될 수 있다. 분자 (800 및 900)는 천연 및/또는 비-천연 뉴클레오티드 또는 리보뉴클레오티드를 포함할 수 있다.

[0157] 도 9a는 이합체화를 위해 역 상보적 서열 혼성화 대신에 키싱 루프 상호작용을 이용하는 예시적인 이합체화 도메인을 제공한다. 2개의 RNA 헤어핀의 루프 중 염기가 2개의 RNA 분자간 상호작용하는 쌍을 형성할 때 키싱 루프 상호작용이 형성된다. n-yfp로 표시된, 좌측에 있는 분자가 스플라이스 도너 부위, DISE 요소, 및 2개의 ISE 요소를 포함하는 합성 인트론에 연결된, yfp의 n-말단 부분을 코딩하는 RNA 분자를 나타낸다. 이 분자의 이합체화 도메인은 각각 스템 (RNA가 그 자체에 혼성화됨) 및 루프 (RNA가 그 자체에 혼성화되지 않음)로 구성된, 3개의 RNA 헤어핀 루프를 포함한다. 이 구체예에서, 이합체화 도메인은 3개의 스템 및 루프 요소들 (헤어핀 루프로도 지칭됨)을 포함하고, 삼중 키싱 루프 이합체화 도메인(trimodal kissing loop dimerization domain)으로 지칭된다. c-yfp로 표시된, 우측에 있는 분자는 yfp의 c-말단 부분을 코딩하는 RNA 분자를 나타낸다. 5'에서 3'으로, 이 분자는 3개의 헤어핀 루프의 세트를 포함하는 삼중 키싱 루프 이합체화 도메인으로 구성된다. 루프 부분은 상보적 n-yfp 분자 상에 있는 상응하는 루프와 키싱 루프 상호작용을 형성할 수 있다. 삼중 키싱 루프 이합체화 도메인은 3개의 ISE 서열, 분지점 서열, 폴리피리미딘 트랙트, 및 스플라이스 억셉터 부위를 포함하는 합성 인트론 서열로 이어진다. 상기 합성 인트론 서열은 c-말단 yfp 코딩 서열로 이어지고, 이는 폴리아데닐화 신호를 포함하는 3' 비번역 영역으로 이어진다. 도면의 상단에, 키싱 루프 상호작용의 대표적인 3차원 표현이 도시된다. 이 표현은 헤어핀 루프의 꼬인 형태가 어떻게 루프 잔기들을 외부로 노출시켜 그들이 키싱 루프 상호작용에 이용될 수 있게 하는 지를 보여준다.

[0158] 2개의 분자의 회합시, 스플라이세오솜이 트랜스-스플라이싱 반응을 매개하여, n-말단 및 c-말단 ypf 코딩 서열의 연결을 초래하고, 전장 형광 단백질의 발현을 가능하게 한다.

[0159] 도 6a-7c 및 9a는 시스템이 2개의 합성 핵산 분자를 이용하는 것인 구체예(즉, 표적 단백질 코딩 서열이 2개의 합성 핵산 분자 간에 분할됨)를 도시하나, 당업자는 그러한 구현예가 본 명세서의 교시를 이용하여, 2개 초과의 핵산 분자, 예를 들면, 3개, 4개, 5개, 6개, 7개, 8개, 9개, 또는 10개의 합성 핵산 분자에도 유사하게 이용될 수 있다는 것을 이해할 것이다.

[0160] 일부 구체예에서, 상기 시스템은 비-조립(un-assembled)/비-재조합(un-recombined) 단편들의 발현을 억제하는 핵산 분자를 포함한다. 그러한 구체예에서, 전장 코딩 서열의 2개 이상의 부분 (예를 들면, 각각 도 6a의 110의 114, 150의 164)이 재조합되지 않으면, 상기 핵산 분자가 전장 단백질로 재조합되지 않은 전장 코딩 서열의 각 부분의 발현을 억제할 것이다. 예를 들면, 그러한 억제성(suppressive) 핵산 분자는 핵 외부에서, RNA를 불안정화시키거나, 번역을 방지하거나, 이동된(shifted) 개시 코돈으로부터의 번역을 촉진하거나, 또는 번역되면, 단백질 활성을 억제하거나 또는 그를 분해를 위해 표시하는(flag) 단백질 데그론(degron) 또는 불안정화 도메인을 포함한다.

[0161] 일 구체예에서, 비-재조합 RNA 분자의 불안정화는 자가-절단 RNA 서열 (예를 들면, Hammerhead 리보자임 또는 HDV 리보자임)을 합성 인트론 내에, 예를 들면, 도 6a 또는 6f의 인트론 서열 (130) 내 임의의 위치에 포함시키는 것에 의해 달성된다. 일 구체예에서, RNA 분자의 절단은 RNA 안정화 폴리 A 테일의 소실을 초래하고, 이는 도 6a 또는 6F의 ORF (114)로부터 비-재조합 단백질의 발현을 억제할 수 있다. 일 구체예에서, 자가-절단 RNA 서열이 도 6a 또는 6f의 인트론 서열 (170) 내의 임의의 위치에 포함되어, 5' 말단 CAP를 절단시키고, 이는, 일 구체예에서 도 6a 또는 6f의 코딩 서열 (164)의 일부 또는 전체를 포함하는 ORF의 감소된 발현을 초래할 수 있다. 일 구체예에서, 자가-절단 RNA 서열이 RNA 절단 효소 표적 부위, 예를 들면, Csy4 표적 부위로 치환된다.

[0162] 일부 구체예에서, 억제성 핵산 분자는 도 6a 또는 6f의 ORF 서열 (164) 대비 -1, -2, +1, 또는 +2 뉴클레오티드가 이동된 ORF의 번역을 지시하는, 도 6a 또는 6f의 인트론 서열 (170) 내의 임의의 위치에 개시 코돈 (ATG) 또는 코작 강화 개시 코돈(Kozak enhanced start codon) (GCCGCCACCATG (서열번호 154) 또는 GCCACCATG 또는 ACCATG)을 포함한다. 일 구체예에서, 비-조립 단편 발현은 도 6a 또는 6f의 억제될 ORF 서열 (164)로부터의 번역을 벗어나게 하기 위해 이 데코이(decoy) 개시 코돈을 이용하는 것에 의해 감소되거나 억제된다.

[0163] 일부 구체예에서, 억제성 핵산 분자는 도 6a 또는 6f의 인트론 서열 (130) 내의 임의의 위치 및/또는 도 6a 또는 6f의 인트론 서열 (170) 내의 임의의 위치에 하나 이상의 마이크로 RNA 표적 부위를 포함한다. 특정한 분자 (예를 들면, 도 6a 또는 6f의 110 또는 150)가 핵으로부터 외부로 수송되는 경우, 상기 분자는 핵으로부터 수송된 비-연결 RNA를 분해/억제시키는 것에 의해 의도되지 않은 비-연결 단편 발현을 억제할 수 있는 마이크로 RNA/sh(small hairpin) RNA 의존성 분해를 겪게 된다. 일 구체예에서, 그러한 마이크로 RNA 표적 서열은 도 6a 또는 6f의 분자 (110 및 150)가 도입되는 세포, 또는 조직, 또는 동물에서 발현되는 것으로 알려진 마이크로 RNA에 상보적일 수 있다. 일 구체예에서, 이 마이크로 RNA 표적 서열은 상기 세포, 또는 조직. 또는 동물에 도입되는 서열에 상보적이다. 일 구체예에서, 그러한 마이크로 RNA는 shRNA의 형태로 RNA 폴리머라아제 III 의존성 프로모터로부터 발현될 수 있다. 일 구체예에서, 그러한 microRNA는 RNA 폴리머라아제 II 의존성 프로모터로부터 발현되고 마이크로 RNA 가공 루프 (예를 들면, mir30 스캐폴드)에 임베딩될 수 있다.

[0164] 일부 구체예에서, ORF (예를 들면, 도 6의 114)로부터의 비-재조합 단백질 산물의 불안정화는 도 6a 또는 6f의 인트론 서열 (130)에서 종결 코돈 출현의 결실 및 도 6a 또는 6f의 인트론 서열 (130) 내의 임의의 위치에 배치되고, 도 6a 또는 6f의 서열 (114)로부터 벗어나 연장된 ORF와 동일한 프레임 내에 있는 단백질을 분해를 위해 표시할 수 있는 프레임 내 단백질 신호(in frame protein signal)(예를 들면, 데그론 서열)를 코딩하는 RNA 서열의 추가적인 포함에 의해 달성될 수 있다. 일 구체예에서 데그론 서열은 PEST 서열의 서열 또는 CL1 데그론 서열의 서열일 수 있다. 이용되는 데그론 서열은 프로테아솜(proteasome)-의존성, 프로테아솜-비의존성, 유비퀴틴-의존성, 또는 유비퀴틴-비의존성 경로를 이용할 수 있다. 일 구체예에서, 비-재조합 단백질 불안정화는 수개의 동일한 또는 상이한 데그론 서열의 포함에 의해 강화된다.

[0165] 일부 구체예에서, 도 6a의 ORF 서열 (164)로부터의 비-재조합 단백질 산물의 불안정화는 도 6에서 서열 (164) 내의 ORF와 프레임 내인 도 6a의 인트론 서열 (170) 내 임의의 위치에 있는 데그론 서열로 이어지는 개시 코돈 (ATG)의 도입에 의해 달성된다. 이 구체예에서, 데그론 서열은 분해를 위해 표시되는 것에 의해 억제될 비-재조합 단백질 단편에 N-말단 연결될 것이다.

[0166] **IV. 조성물 및 키트**

[0167] 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자를 포함하는 조성물 및 키트로서, 상기 합성 핵산 분자는 재조합되면 전장-단백질을 코딩하는 것인 조성물 및 키트가 제공된다. 일부 구체예에서, 상기 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자는 DNA이다. 일부 구체예에서, 상기 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자는 RNA이고, 프로모터 서열을 포함하지 않는다. 일 구체예에서, 상기 조성물 또는 키트는 2개의 본 명세서에서 제공되는 합성 핵산 분자를 포함하고, 상기 2개의 합성 핵산 분자 각각은 표적 단백질, 예를 들면, 표 1에 열거된 것 (또는 치료 단백질, 예를 들면, 독소 또는 티미딘 키나아제)의 상이한 부분(즉, N-말단 및 C-말단 부분, 상기 2개의 합성 핵산 분자간 재조합이 일어나는 경우, 전체 코딩 서열이 생성됨)을 코딩한다. 일 구체예에서, 상기 조성물 또는 키트는 3개의 본 명세서에서 제공되는 합성 핵산 분자를 포함하고, 상기 3개의 합성 핵산 분자 각각은 표적 단백질, 예를 들면, 표 1에 열거된 것 (또는 치료 단백질, 예를 들면, 독소 또는 티미딘 키나아제)의 상이한 부분(즉, N-말단, 중간, 및 C-말단 부분, 상기 3개의 합성 핵산 분자간 재조합이 일어나는 경우, 전체 코딩 서열이 생성됨)을 코딩한다. 일 구체예에서, 상기 조성물 또는 키트는 4개 이상의 본 명세서에서 제공되는 합성 핵산 분자를 포함하고, 상기 4개의 합성 핵산 분자 각각은 표적 단백질, 예를 들면, 표 1에 열거된 것 (또는 치료 단백질, 예를 들면, 독소 또는 티미딘 키나아제)의 상이한 부분(즉, N-말단, 제1 중간, 제2 중간 (및 선택적으로 추가적인 중간), 및 C-말단 부분, 상기 4개 이상의 합성 핵산 분자간 재조합이 일어나는 경우, 전체 코딩 서열이 생성됨)을 코딩한다. 일 구체예에서, 상기 조성물 또는 키트는 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자의 2개 이상의 세트를 포함하고, 상기 합성 핵산 분자의 세트 각각은 상이한 표적 단백질, 예를 들면, 표 1에 열거된 것 (또는 치료 단백질, 예를 들면, 독소 또는 티미딘 키나아제) 중 2개 이상을 코딩한다.

[0168] 일 구체예에서, 상기 조성물 또는 키트 중 각각의 합성 핵산 분자는 벡터, 예를 들면, AAV, 또는 기타 유전자 요법 벡터의 일부이다. 일 구체예에서, 상기 조성물 또는 키트는 세포, 예를 들면, 박테리아 세포 또는 진핵 세

포를 포함하고, 2개 이상의 개시된 합성 핵산 분자를 포함하며, 상기 합성 핵산 분자는 재조합되면 전장 표적 단백질을 코딩한다.

[0169] 그러한 조성물은 약제학적으로 허용가능한 담체 (예를 들면, 염수(saline), 물, 글리세롤, DMSO, 또는 PBS)를 포함할 수 있다. 일부 구체예에서, 상기 조성물은 액체, 동결건조된 분말, 또는 동결보존물(cryopreserved)이다.

[0170] 일부 구체예에서, 상기 키트는 세포 타입 특이적 흡수를 유도하고/엔도좀 탈출(endosome escape)을 증가시키고/혈뇌 장벽 통과 등을 가능하게 하기 위해 전달 시스템 (예를 들면, 리포좀, 입자, 엑소좀, 또는 미세소포(microvesicle))을 포함한다. 일부 구체예에서, 상기 키트는 세포 배양 또는 증식 배지, 예를 들면, 박테리아, 식물, 곤충, 또는 포유동물 세포의 증식에 적합한 배지를 더 포함한다. 일부 구체예에서, 키트의 그러한 부분들이 별개의 용기에 존재한다. 예시적인 용기는 플라스틱 또는 유리 바이알 또는 튜브를 포함한다.

[0171] 일부 구체예에서, 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자 각각은 별개의 용기에 존재한다. 일부 구체예에서, 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자의 2개 이상의 세트 각각은 별개의 용기에 존재한다.

[0172] **V. 치료 방법**

[0173] 본 명세서에서 개시된 방법 및 시스템은, 예를 들면, 단백질이 너무 커서 치료 바이러스 (예를 들면, AAV)에 의해 발현될 수 없거나 또는 완전한 유전자 서열 (예를 들면, 내생적 프로모터 + 코딩 서열)이 너무 커서 치료 바이러스 (예를 들면, AAV)에 의해 발현될 수 없는 경우, 목적 단백질을 발현하기 위해 이용될 수 있다. 그러한 경우에, 표적 단백질의 코딩 서열은 본 명세서에서 개시되는 시스템을 이용하여 2개 이상의 부분으로 나뉠 수 있고, 정확한 순서로 재조합되어, 상기 단백질이 원하는 때 및 위치에서 발현될 수 있게 한다.

[0174] 치료되는 개체는 임의의 포유동물, 예를 들면, 단일유전자성 질환(monogenetic disorder), 예를 들면, 표 1에 열거된 질환을 갖는 포유동물일 수 있다. 일 구체예에서, 상기 개체는 암을 가질 수 있다. 따라서, 인간, 고양이, 돼지, 랫트, 마우스, 소, 염소, 및 개가 본 명세서에서 개시되는 방법을 이용하여 치료될 수 있다. 일부 구체예에서, 상기 개체는 6개월령 미만의 인간 유아이다. 일부 구체예에서, 상기 개체는 1세 미만의 인간 유아이다. 일부 구체예에서, 상기 개체는 인간 청소년이다. 일부 구체예에서, 상기 개체는 18세 이상의 성인이다. 일부 구체예에서, 상기 개체는 여성이다. 일부 구체예에서, 상기 개체는 남성이다.

[0175] 개체를 치료하기 위해 이용되는 상기 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자는 치료되는 개체에 매칭될 수 있다. 따라서, 예를 들면, 상기 치료되는 개체가 개인 경우, 표적 단백질에 대한 개 코딩 서열이 이용될 수 있고, 인트론 서열이 개 세포에서의 발현을 위해 최적화될 수 있고, 상기 치료되는 개체가 인간인 경우, 표적 단백질에 대한 인간 코딩 서열이 이용될 수 있고, 인트론 서열은 인간 세포에서의 발현을 위해 최적화될 수 있다.

[0176] 상기 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자는 벡터, 예를 들면, AAV (adeno-associated vector), 예를 들면, AAV 혈청형 rh.10의 일부로 투여될 수 있다. 일부 구체예에서, 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자 중 하나를 포함하는 벡터 (예를 들면, AAV)는 전신적으로, 예를 들면, 정맥내로 투여된다. 따라서, 코딩 서열이 2개의 본 명세서에서 제공되는 합성 핵산 분자 간에 나뉘는 경우, 2개의 AAV가 투여되고, 각 AAV는 2개의 본 명세서에서 제공되는 합성 핵산 분자 중 하나를 포함한다.

[0177] 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자의 치료 유효량이 예를 들면, AAV로 투여된다. 일부 구체예에서, 상기 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자는 바이러스 벡터 (예를 들면, AAV)의 일부로서 개체당 적어도 $1x10^{11}$ gc(genome copies), 적어도 $1x10^{12}$ gc, 적어도 $2x10^{12}$ gc, 적어도 $1x10^{13}$ gc, 적어도 $2x10^{13}$ gc/개체, or 적어도 $1x10^{14}$ gc/개체, 예를 들면, $2x10^{11}$ gc/개체, $2x10^{12}$ gc/개체, $2x10^{13}$ gc, 또는 $2x10^{14}$ gc/개체의 용량으로 투여된다. 일부 구체예에서, 상기 2개 이상의 본 명세서에서 제공되는 합성 핵산 분자는 바이러스 벡터 (예를 들면, AAV)의 일부로서 적어도 $1x10^{11}$ gc/kg, 적어도 $5x10^{11}$ gc/kg, 적어도 $1x10^{12}$ gc/kg, 적어도 $5x10^{12}$ gc/kg, 적어도 $1x10^{13}$ gc/kg, 또는 적어도 $4x10^{13}$ gc/kg, 예를 들면, $4x10^{11}$ gc/kg, $4x10^{12}$ gc/kg, 또는 $4x10^{13}$ gc/kg의 용량으로 투여된다.

[0178] 유해한 증상, 예를 들면, 혈액 중 AAV-캡시드 특이적 T 세포가 발생하는 경우, 코르티코스테로이드가 투여될 수

있다 (예를 들면, Nathwani *et al.*, *N Engl J Med.* 365(25):2357-65, 2011 참조).

[0179] 본 명세서에서 개시되는 방법으로 치료될 수 있는 질환은 혈액의 유전 질환 (예를 들면, 겸상 세포 질환, 원발성 면역결핍성 질환), HIV (예를 들면, HIV-1), 및 혈액암(hematologic malignancies) 또는 암을 포함한다. 원발성 면역결핍성 질환 및 그의 상응하는 돌연변이의 예는 Al-Herz *et al.* (*Frontiers in Immunology*, volume 5, article 162, April 22, 2014, 참조에 의해 그 전체로 본 명세서에 포함됨)에 열거된 것들을 포함한다. 혈액암 또는 암은 혈액, 골수, 및 림프절에 영향을 미치는 종양이다. 예는 백혈병 (예를 들면, 급성 림프구성 백혈병, 만성 림프구성 백혈병, 급성 골수성 백혈병, 만성 골수성 백혈병, 급성 단핵구성 백혈병), 림프종 (예를 들면, 호지킨 림프종 및 비-호지킨 림프종), 및 골수종을 포함한다. 일부 구체예에서, 상기 질환은 단일기원(monogenetic) 질환이다. 표 1은 개시된 시스템 및 방법에 의해 표적화될 수 있는 예시적 질병 및 유전자의 목록을 제공한다. 추가적인 예가 rarediseases.info.nih.gov/diseases/diseases-by-category/5/congenital-and-genetic-diseases (목록이 참조에 의해 본 명세서에 포함됨)에서 제공된다. 단백질의 결핍 (예를 들면, 열성 돌연변이) 또는 단백질의 부족(insufficiency)에 의해 유발된 유전 질환은 본 명세서에서 개시되는 시스템 및 방법으로부터 유익을 얻을 수 있다. 유전자의 코딩 영역이 상대적으로 작은 경우, 본 명세서에서 개시되는 시스템 및 방법이 조절 서열, 예를 들면, 유전자 발현을 적합한 수준으로 적합한 세포 타입으로 유도하기 위해 조직 특이적 프로모터 또는 특이적 비-코딩 RNA 세그먼트를 부가하기에 유용할 수 있다.

[0180] **표 1: 예시적 질환 및 상응하는 돌연변이**

표 1

[0181]

질병	**유전자**	**돌연변이**
혈액 세포 질환		
겸상 적혈구 빈혈증	헤모글로빈의 β-글로빈 사슬	점 돌연변이를 생성하는 SNP (A to T) (6번 아미노산에서 Glu->Val)
혈우병	응고 인자 I 내지 XIII 중 하나	
A형 혈우병	응고 인자 VIII	큰 결실, 삽입, 역위, 및 점 돌연변이
B형 혈우병	응고 인자 IX	
알파-탈라세미아	HBA1 또는 HBA2	염색체 16 p 중 돌연변이 또는 결실
베타-탈라세미아	HBB	염색체 11 중 돌연변이
델타-탈라세미아	HBD	돌연변이
폰 빌레브란트병(von Willebrand Disease)	폰 빌리브란트 인자(von Willebrand factor)	돌연변이 또는 결실
악성 빈혈	MTHFR	
판코니 빈혈(Fanconi anemia)	FANCA, FANCC, FANCD2, FANCG, FANCJ	FANCA: c.3788_3790del (p.Phe1263del); c.1115_1118delTTGG (p.Val372fs); Exon 12-17del; Exon 12-31del; c.295C>T (p.Gln99X) FANCC: c.711+4A>T (처음에 IVS4+4A>T로 보고됨); c.67delG (처음에 322delG로 보고됨) FANCD2: c.1948-16T>G FANCG; c.313G>T (p.Glu105X); c.1077-2A>G; c.1480+1G>C; c.307+1G>C; c.1794_1803del (p.Trp599fs); c.637_643del (p.Tyr213fs) FANCJ: c.2392C>T (p.Arg798X)

혈소판 감소성 자반증(Thrombocytopenic purpura)	ADAMTS13	미스센스 및 넌센스 돌연변이
혈전성향증(thrombophilia)	인자 V Leiden 프로트롬빈	F5 유전자 중 1691번 돌연변이 프로트롬빈 G20210A
원발성 면역결핍 질환		
T-B+ SCID	IL-2RG, JAK3, IL-2, -4,-7,-9, -15 및 -21에 대한 수용체의 감마 사슬 중 결함	
T-B- SCID	RAG1, RAG2	
WHIM 증후군	CXCR4	이형접합 돌연변이 (예를 들면, 카르복시-말단 중 돌연변이); 카르복시-말단 절단 (예를 들면, 10-19개 잔기)
기타 원발성 면역결핍 (PID) 증후군		
IL-7 수용체 중증 복합 면역 결핍 장애(severe combined immune deficiency: SCID)	IL7 수용체	
아데노신 탈아미나아제 결핍(ADA) SCID	ADA	
퓨린 뉴클레오시드 포스포릴라아제 (PNP) 결핍	PNP	
위스콧-알드리치 증후군 (WAS)	WAS	300개 이상의 돌연변이가 식별됨
만성 육아종증 (CGD)	CYBA, CYBB, NCF1, NCF2, 또는 NCF4	
백혈구 부착 결핍 (LAD)	베타-2 인테그린	
HIV	C-C 케모카인 수용체 타입 5 (CCR5), MSRB1 HIV LTR(long terminal repeats) CSCR4 P17 PSIP1	CCR5 중 32 bp의 결실
듀센 근육 이영양증(Duchenne muscular dystrophy)	CCR5DMD	
글리코겐 저장성 질병 타입 IA(Glycogen storage disease type IA)	G6Pase	
망막 이영양증(Retinal Dystrophy)	CEP290ABCA4	C2991+1655A>G 5196+1216C>A; 5196+1056A>G; 5196+1159G>A; 5196+1137G>A; 938-619A>G; 4539+2064C>T
XMEN(X-linked immunodeficiency with magnesium defect, Epstein-Barr virus infection, and neoplasia)	MAGT1	
단일유전자성 질환(MonoGenetic Disorders)		
이염성 백질이영양증 (MLD)	ARSA(arylsulfatase A)	
부신 백질이영양증 (ALD)	ABCD1	
뮤코다당증 (MPS) 장애 헌터(Hunter) 증후군 헐러(Hurler) 증후군 샤이에(Scheie) 증후군 산필립포(Sanfilippo) 증후군 A, B, C, 및 D 모르퀴오(Morquio) 증후군 A 모르퀴오 증후군 B 마르토-라미(Maroteaux-Lamy) 증후군 슬라이(Sly) 증후군 나토위츠(Natowicz) 증후군	 IDS IDUA IDUA SGSH, NAGLU, HGSNAT, GNS GALNS GLB1 ARSB GUSB HYAL1	
알파 만노시드증(Alpha manosidosis)	MAN2B1	

니만 피크병(Nieman Pick disease) 타입 A, B, 및 C	SMPD1, NPC1, NPC2	
낭포성 섬유증	CFTR(cystic fibrosis transmembrane conductance regulator)	△F508
다낭성 신장 질환	PKD-1, PDK-2, PDK-3	
테이 삭스병(Tay Sachs Disease)	HEXA	1278insTATC
고셔병(Gaucher disease)	GBA	
헌팅턴병(Huntington's disease)	HTT	CAG repeat
신경섬유종증 타입 1 및 2	NF-1 및 NF2	NF1 중 CGA->UGA->Arg1306Term
가족성 고콜레스테롤혈증	APOB, LDLR, LDLRAP1, 및 PCSK9	
암		
만성 골수성 백혈병 (CML)	BCR-ABLASXL1	융합
급성 골수성 백혈병 (AML)	염색체 11q23 또는 t(9;11)	전좌
골육종	RUNX2	
대장암	EPHA1	
위암, 흑색종	PD-1	
전립선암	안드로겐 수용체	
자궁경부암	E6, E7	
교모세포종	CD	
신경질환		
알츠하이머병	NGF	
이염성 백질이영양증 (MLD)	ARSA	
다발성 경화증	MBP	
위스콧-알드리치 증후군	WASP	
X-연관 부신백질이영양증(X-linked adrenoleukodystrophy)	ABCD1	
AACD 결핍	AADC	
바텐병(Batten disease)	CLN2	
카나반병(Canavan disease)	ASPA	
거대 축삭 신경병증	GAN	
레버 유전성 시신경병증 (Leber's hereditary optic neuropathy)	MT-ND4	
MPS IIIA	SGSH, SUMF1	
파킨슨병	GAD, NTRN, TH, AADC, CH1, GDNF, AADC	
폼페병	GAA	
척수 근이영양증 타입 1	SMN	

[0182] 본 명세서에서 개시되는 방법 및 시스템을 이용하는 것이 표 1에 열거된 질환, 또는 기타 공지된 유전 질환을 치료하기 위해 이용될 수 있다. 본 명세서에서 개시되는 방법은 또한 기타 질환, 예를 들면, 암 세포에서 독소 또는 티미딘 키나아제와 같은 치료 단백질의 발현으로부터 유의을 얻을 수 있는 암을 치료하기 위해 이용될 수 있다. 개체에 전장 티미딘 키나아제를 발현하는, 본 명세서에서 제공되는 2개 이상의 합성 분자가 투여되는 경우, 상기 개체에 간시클로비르도 투여된다. 치료는 질환의 모든 특징의 100% 제거를 요구하지 않으나, 감소일 수 있다. 특정한 구체예가 하기에 제공되나, 이 교시에 근거하여, 당업자는 기타 질환의 증상이 유사하게 영향을 받을 수 있다는 것을 이해할 것이다. 예를 들면, 본 명세서에서 개시되는 방법은 개체에 의해 발현되지 않거나 또는 감소된 발현을 갖는 단백질의 발현을 증가시키거나, 또는 개체에 의해 바람직하지 않게 발현되거나 또는 감소된 발현을 갖는 단백질의 발현을 감소시키기 위해 이용될 수 있다. 예를 들면, 본 명세서에서 개시되는 방법은 유전 질환의 원치않는 효과를 치료 또는 경감시키기 위해 이용될 수 있다.

[0183] 예를 들면, 본 명세서에서 개시되는 방법 및 시스템은 헤모글로빈의 전장 야생형 β-글로빈 사슬을 발현시키는 것에 의해 겸상 세포 질환의 원치않는 효과를 치료하거나 또는 경감시킬 수 있다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체에서 겸상 세포 질환의 증상 (예를 들면, 혈액 중 겸상 세포의 존재, 통증, 허혈, 괴사, 빈혈, 혈관폐쇄 발증(vaso-occlusive crisis), 골수무형성 발증(aplastic crisis), 비장 격리 발증(splenic sequestration crisis), 및 용혈 발증(haemolytic crisis) 중 하나 이상)을, 예를 들면, (치료 핵산

분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시킨다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체에서 겸상 세포의 개수를 감소시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 적어도 90%, 또는 적어도 95% 감소시킨다.

[0184] 예를 들면, 본 명세서에서 개시되는 방법 및 시스템은 전장 야생형 인자 V Leiden 또는 프로트롬빈 유전자를 발현시키는 것에 의해 혈전성향증(thrombophilia)의 원치않는 효과를 치료하거나 또는 경감시킬 수 있다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체에서 혈전성향증의 증상 (예를 들면, 심부정맥 혈전증과 같은 혈전증, 폐 색전증, 정맥 혈전색전증, 종창, 흉통, 심계항진 중 하나 이상)을 감소시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시킨다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체에서 응고 인자의 활성을 감소시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 적어도 90%, 또는 적어도 95% 감소시킨다.

[0185] 예를 들면, 본 명세서에서 개시되는 방법 및 시스템은 전장 야생형 CD40 리간드 유전자를 발현시키는 것에 의해 CD40 리간드 결핍증의 원치않는 효과를 치료하거나 또는 경감시킬 수 있다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체에서 CD4 리간드 결핍증의 증상 (예를 들면, 증가된 혈청 IgM, 기타 면역 글로불린의 낮은 혈정 수준, 기회 감염, 자가면역, 및 악성 종양 중 하나 이상)을 감소시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시킨다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체에서 CD40 리간드의 양 또는 활성을 증가시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 적어도 90%, 적어도 100%, 적어도 200% 또는 적어도 500% 증가시킨다.

[0186] 예를 들면, 본 명세서에서 개시되는 방법은 유전적 결함으로부터 유발된 원발성 면역결핍증의 원치않는 효과를 치료하거나 또는 경감시키기 위해 이용될 수 있다. 예를 들면, 상기 개시되는 방법 및 시스템 (예를 들면, AAV를 이용하여, 개체에서 결실되거나 또는 결함인 기능성 단백질을 발현시키기 위해 2개 이상의 합성 핵산 분자를 이용할 수 있음)은 원발성 면역결핍성 질환의 원치않는 효과를 치료하거나 또는 경감시킬 수 있다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체에서 원발성 면역결핍성 질환의 증상 (예를 들면, 박테리아 감염, 균류 감염, 바이러스 감염, 기생충 감염, 림프샘 종창, 비장 확대, 상처, 및 체중 감소 중 하나 이상)을 감소시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시킨다. 일 구체예에서, 본 명세서에서 개시되는 방법은 원발성 면역결핍성 질환을 갖는 피투여 개체에서 면역 세포(예를 들면, CD8 세포와 같은 T 세포)의 개수를 증가시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비), 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 적어도 90%, 적어도 95%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 또는 적어도 500% 증가시킨다. 일 구체예에서, 본 명세서에서 개시되는 방법은 원발성 면역결핍성 질환을 갖는 피투여 개체에서 정해진 기간 동안 (예를 들면, 1년에 걸쳐) 감염(예를 들면, 박테리아 감염, 바이러스 감염, 균류 감염, 또는 이들의 조합)의 개수를 감소시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 적어도 90%, 또는 적어도 95% 감소시킨다.

[0187] 예를 들면, 본 명세서에서 개시되는 방법은 단일기원(단일유전자성) 질환의 원치않는 효과를 치료하거나 또는 경감시키기 위해 이용될 수 있다. 예를 들면, 본 명세서에서 개시되는 방법(예를 들면, AAV를 이용하여, 개체에서 결실되거나 또는 결함인 기능성 단백질을 발현시키기 위해 2개 이상의 합성 핵산 분자를 이용할 수 있음)은 단일기원 질환의 원치않는 효과를 치료하거나 또는 경감시킬 수 있다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체에서 단일기원 질환의 증상을 경감시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 경감시킨다. 일 구체예에서, 본 명세서에서 개시되는 방법은 단일기원 질환을 갖는 피투여 개체에 의해 정상적으로 발현되지 않는 정상 단백질의 양을 증가시키고, 예를 들면, (치료 핵산 분자의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 적어도 90%, 적어도 95%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 또는 적어도 500% 증가시킨다.

[0188] 예를 들면, 본 명세서에서 개시되는 방법은 피투여 개체에서 혈액암의 원치않는 효과를 치료하거나 또는 경감시키기 위해 이용될 수 있다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체 (예를 들면, 백혈병을 갖는 개체)에서 비정상 백혈구 세포(예를 들면, B 세포)의 개수를 감소시키고, 예를 들면, (개시된 요법의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시킨다. 일 구체예

에서, 본 명세서에서 개시되는 요법의 투여는 림프종의 원치않는 효과를 치료하거나 또는 경감시키기 위해, 예를 들면, 림프종의 크기, 림프종의 부피, 림프종의 증식 속도, 림프종의 전이를 감소시키기 위해, 예를 들면, (개시된 요법의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시키기 위해 이용될 수 있다. 일 구체예에서, 본 명세서에서 개시되는 요법의 투여는 다발성 골수종의 원치않는 효과를 치료하거나 또는 경감시키기 위해, 예를 들면, 피투여 개체에서 비정상 형질 세포의 개수를 예를 들면, (개시된 요법의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시키기 위해 이용될 수 있다.

[0189] 예를 들면, 본 명세서에서 개시되는 방법은 피투여 개체에서 유전적 결함으로부터 유발되는 것과 같은, 악성 종양의 원치않는 효과를 치료하거나 또는 경감시키기 위해 이용될 수 있다. 일 구체예에서, 본 명세서에서 개시되는 방법은 피투여 개체(예를 들면, 본 명세서에 열거된 암을 가진 개체)에서 암세포의 개수, 종양의 크기, 종양의 부피, 또는 전이의 개수를, 예를 들면, (개시된 요법의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시키기 위해 이용될 수 있다. 일 구체예에서, 본 명세서에서 개시되는 요법의 투여는 림프종의 원치않는 효과를 치료하거나 또는 경감시키기 위해, 예를 들면, 종양의 크기, 종양의 부피, 암의 증식 속도, 암의 전이를 예를 들면, (개시된 요법의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 또는 적어도 90% 감소시키기 위해 이용될 수 있다. 예를 들면, 본 명세서에서 개시되는 방법은 피투여 개체에서 유전적 결함으로부터 유발되는 신경 질환의 원치않는 효과를 치료하거나 또는 경감시키기 위해 이용될 수 있다. 일 구체예에서, 상기 개시되는 방법은 피투여 개체 (예를 들면, 전술된 신경 질환을 갖는 개체)에서 신경 기능을, 예를 들면, (개시된 요법의 투여가 없는 경우 대비) 적어도 10%, 적어도 20%, 적어도 50%, 적어도 70%, 적어도 90%, 적어도 100%, 적어도 200%, 적어도 300%, 적어도 400%, 또는 적어도 500% 증가시킨다.

[0190] **듀센 근이영양증 (DMD)의 치료**

[0191] 듀센 근이영양증 (DMD, MIM:310200)은 진행성 근육 약화 및 퇴행을 특징으로하는 치명적인 유전성 질환이다. 질병이 진행되면서, 퇴행성 근섬유가 지방 및 섬유증 조직에 의해 대체된다. DMD는 유전자 디스트로핀 (MIM:300377)의 결함에서 기원한다. 디스트로핀 유전자는 22k bp의 영역에 걸치고, 돌연변이되는 경향이 있다. 따라서, DMD는 일부 경우에 질병-유발 돌연변이의 가족력이 없는 환자에서도 산발적으로 나타날 수 있다. DMD는 디스트로핀병증(dystrophinopathies)으로 알려진 4개의 질병 중 하나이다. 이 그룹에 속하는 나머지 3개의 질병은 벡커 근이영양증 (BMD, DMD의 마일드형); DMD와 BMD 사이의 중간 임상 양태(intermediate clinical presentation); 및 임상적 골격 질환, 또는 자발적(voluntary), 근육 질환이 거의 또는 전혀 없는 DMD-연관 확장성 심근병증 (심장 질환)이다. 따라서, 일부 구체예에서, DMD, BMD, DMD와 BMD 사이의 중간 임상 양태; 또는 임상적 골격 질환, 또는 자발적, 근육 질환이 거의 또는 전혀 없는 DMD-연관 확장성 심근병증 (심장 질환)을 갖는 환자가 본 명세서에서 개시되는 시스템 및 방법으로 치료된다.

[0192] 본 명세서에서 개시되는 방법 및 시스템은 디스트로핀을 발현시키는 것에 의해 DMD의 단일기원 원인을 치료하기 위해 이용될 수 있다. 디스트로핀은 긴 코딩 영역, 예를 들면, 디스트로핀 (dystrophin)을 갖는다. 단일 AAV로부터 디스트로핀을 발현시키는 현재의 방법은 디스트로핀의 단축/절단 버전(마이크로-디스트로핀 및 미니-디스트로핀)을 이용한다. 이러한 절단 디스트로핀 전달 요법 중 다수가 I/II상 임상 시험(NCT03362502, NCT00428935, NCT03368742, NCT03375164)에서 테스트되고 있다. 이러한 디스트로핀의 절단 버전이 DMD에서 디스트로핀 결함의 최악의 결과들을 개선시킬 수 있으나, 절단 버전들은 전장 단백질의 로드(rod) 및 힌지(hinge) 영역에서 주요한 도메인이 결실되므로, 전장 디스트로핀과 비교시 완전한 기능을 가질 것으로 예상되지 않는다. 본 명세서에서 개시되는 방법 및 시스템은, 다수의 AAV 바이러스가 높은 MOI (multiplicity of infection, 즉, 높은 역가)로 도입되는 경우, 동일한 세포를 효율적으로 감염시킬 수 있기 때문에, "다중(multiplexed)" AAV 조합물을 이용하는 것에 의해 전이 유전자 페이로드(transgenic payload)의 크기 제한을 완화시킨다.

[0193] 따라서, 일부 구체예에서, 각각 본 명세서에서 개시되는 합성 분자의 세트 중 하나를 포함하는, 2개 이상의 AAV를 포함하는 조성물이 DMD 개체에게 치료 유효량으로, 예를 들면, 재조합되면 전장 디스트로핀 코딩 서열을 초래하는, (각각 상이한 AAV에) 2개, 3개, 4개, 또는 5개의 상이한 합성 RNA 분자의 세트로서 (예를 들면, i.v.) 투여된다.

[0194] **VI. 예시적 구현예**

[0195] 1. 표적 단백질의 발현용 시스템으로서, (a) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제1 프로모터를 포함하는 제1 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 상기 표적 단백질의 N-말단 부분에 대한 코

딩 서열; 스플라이스 도너; 및 제 1 이합체화 도메인을 포함하는 것인 제1 합성 핵산 분자; 및 (b) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제2 프로모터를 포함하는 제2 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; 분지점 서열; 폴리피리미딘 트랙트; 스플라이스 억셉터; 및 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제2 합성 핵산 분자를 포함하는 것인 시스템.

[0196] 2. 표적 단백질의 발현용 시스템으로서, (a) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제1 프로모터를 포함하는 제1 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; 스플라이스 도너; 및 제 1 이합체화 도메인을 포함하는 것인 제1 합성 핵산 분자; 및 (b) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제2 프로모터를 포함하는 제2 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; 분지점 서열; 폴리피리미딘 트랙트; 스플라이스 억셉터; 및 상기 표적 단백질의 중간 부분(middle portion)에 대한 코딩 서열; 제2 스플라이스 도너; 및 제3 이합체화 도메인을 포함하는 것인 제2 합성 핵산 분자; 및 (c) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제3 프로모터를 포함하는 제3 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 제4 이합체화 도메인으로서, 상기 제4 이합체화 도메인은 상기 제3 이합체화 도메인에 결합하는 것인 제4 이합체화 도메인; 분지점 서열; 폴리피리미딘 트랙트; 스플라이스 억셉터; 및 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제3 합성 핵산 분자를 포함하는 것인 시스템.

[0197] 3. 표적 단백질의 발현용 시스템으로서, (a) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제1 프로모터를 포함하는 제1 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; 스플라이스 도너; 및 제 1 이합체화 도메인을 포함하는 것인 제1 합성 핵산 분자; (b) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제2 프로모터를 포함하는 제2 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; 분지점 서열; 폴리피리미딘 트랙트; 스플라이스 억셉터; 및 상기 표적 단백질의 중간 부분에 대한 코딩 서열; 제2 스플라이스 도너; 및 제3 이합체화 도메인을 포함하는 것인 제2 합성 핵산 분자; 및 (c) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제3 프로모터를 포함하는 제3 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 제4 이합체화 도메인으로서, 상기 제4 이합체화 도메인은 상기 제3 이합체화 도메인에 결합하는 것인 제4 이합체화 도메인; 분지점 서열; 폴리피리미딘 트랙트; 스플라이스 억셉터; 및 상기 표적 단백질의 제1 중간 부분에 대한 코딩 서열; 제2 스플라이스 도너; 및 제5 이합체화 도메인을 포함하는 것인 제3 합성 핵산 분자; 및 (d) RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제4 프로모터를 포함하는 제4 합성 핵산 분자로서, 상기 RNA 분자는 5'에서 3'으로: 제6 이합체화 도메인으로서, 상기 제6 이합체화 도메인은 상기 제5 이합체화 도메인에 결합하는 것인 제6 이합체화 도메인; 분지점 서열; 폴리피리미딘 트랙트; 스플라이스 억셉터; 및 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제4 합성 핵산 분자를 포함하는 것인 시스템.

[0198] 4. 구현예 1 내지 3 중 어느 하나에 있어서, 각 프로모터는 독립적으로 선택되는 것인 시스템.

[0199] 5. 구현예 1 내지 4 중 어느 하나에 있어서,

[0200] 상기 제1 및 제2 프로모터는 동일한 프로모터이거나;

[0201] 상기 제1 및 제2 프로모터는 상이한 프로모터이거나;

[0202] 상기 제1, 제2, 및 제3 프로모터는 동일한 프로모터이거나;

[0203] 상기 제1, 제2, 및 제3 프로모터는 상이한 프로모터이거나;

[0204] 상기 제1, 제2, 제3, 및 제4 프로모터는 동일한 프로모터이거나; 또는

[0205] 상기 제1, 제2, 제3, 및 제4 프로모터는 상이한 프로모터인 것인 시스템.

[0206] 6. 구현예 1 내지 5 중 어느 하나에 있어서, 상기 제1, 제2, 제3, 및 제4 프로모터 각각은 항시적 프로모터; 조직-특이적 프로모터; 및 상기 표적 단백질에 내생적인(endogenous) 프로모터로부터 독립적으로 선택되는 것인 시스템.

[0207] 7. 구현예 1 내지 6 중 어느 하나에 있어서, 상기 제1 이합체 도메인과 상기 제2 이합체 도메인, 상기 제3 이합체 도메인과 상기 제4 이합체 도메인, 및/또는 상기 제5 이합체 도메인과 상기 제6 이합체 도메인은 직접 결합,

간접 결합, 또는 이들의 조합에 의해 결합하는 것인 시스템.

[0208] 8. 구현예 7에 있어서, 직접 결합 또는 간접 결합은 염기쌍 형성 상호작용, 비-규범적(non-canonical) 염기쌍 형성 상호작용, 비-염기쌍 형성 상호작용(non-base pairing interaction), 또는 이들의 조합을 포함하는 것인 시스템.

[0209] 9. 구현예 7 또는 8에 있어서, 직접 결합은 키싱 루프(kissing loop) 또는 저다양성(hypodiverse) 영역간 염기쌍 형성 상호작용을 포함하는 것인 시스템.

[0210] 10. 구현예 7 또는 8에 있어서, 직접 결합은 압타머 영역간 규범적 염기쌍 형성 상호작용, 비-규범적 염기쌍 형성 상호작용, 비-염기쌍 형성 상호작용, 또는 이들의 조합을 포함하는 것인 시스템.

[0211] 11. 구현예 7 또는 8에 있어서, 간접 결합은 핵산 브릿지(nucleic acid bridge)를 통한 염기쌍 형성 상호작용을 포함하는 것인 시스템.

[0212] 12. 구현예 7 또는 8에 있어서, 간접 결합은 압타머와 압타머 표적간, 또는 2개의 압타머간 비-염기쌍 형성 상호작용을 포함하는 것인 시스템.

[0213] 13. 구현예 1 내지 12 중 어느 하나에 있어서, 상기 제1, 제2, 제3, 제4, 제5, 및/또는 제6 이합체화 도메인은 크립틱 스플라이스 억셉터(cryptic splice acceptor)를 포함하지 않는 것인 시스템.

[0214] 14. 구현예 1 내지 13 중 어느 하나에 있어서, 적어도 1쌍의 직접적으로 또는 간접적으로 결합하는 압타머 서열 이합체화 도메인을 포함하는 것인 시스템.

[0215] 15. 구현예 1 내지 14 중 어느 하나에 있어서, 적어도 1쌍의 키싱 루프 상호작용 이합체화 도메인을 포함하는 것인 시스템.

[0216] 16. 구현예 1 내지 15 중 어느 하나에 있어서, 상기 표적 단백질은 질병과 연관된 단백질, 또는 치료 단백질인 것 시스템.

[0217] 17. 구현예 16에 있어서, 상기 질병은 단일유전자성(monogenic) 질병인 것인 시스템.

[0218] 18. 구현예 17에 있어서, 상기 치료 단백질은 독소인 것인 시스템.

[0219] 19. 구현예 16 내지 18 중 어느 하나에 있어서, 상기 질병 및 상기 표적 단백질은 표 1에 열거된 것인 시스템.

[0220] 20. 구현예 1 내지 19중 어느 하나에 있어서, 상기 제1, 제2, 제3, 및/또는 제4 합성 핵산 분자는 상기 제1, 제2, 제3, 및/또는 제4 합성 핵산 분자의 3'-말단에 폴리아데닐화 서열을 더 포함하는 것인 시스템.

[0221] 21. 구현예 1, 또는 4 내지 20 중 어느 하나에 있어서,

[0222] 상기 제1 합성 핵산 분자는 상기 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'(3' to the splice donor and 5' to the first dimerization domain)인 DISE(downstream intronic splice enhancer), 상기 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 ISE (intronic splice enhancer) 중 하나, 또는 둘 모두를 더 포함하고; 및/또는

[0223] 상기 제2 합성 핵산 분자는 상기 제2 이합체화 도메인의 3'이고 상기 분지점 서열의 5'인 ISE 및 상기 스플라이스 도너의 3'이고, 상기 이합체화 도메인의 5'인 DISE 중 하나, 또는 둘 모두를 더 포함하거나; 또는

[0224] 이들의 조합인 것인 시스템.

[0225] 22. 구현예 2, 또는 4 내지 20 중 어느 하나에 있어서,

[0226] 상기 제1 합성 핵산 분자는 상기 제1 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 DISE(downstream intronic splice enhancer), 상기 제1 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 ISE (intronic splice enhancer), 또는 DISE 및 ISE를 더 포함하고;

[0227] 상기 제2 합성 핵산 분자는 상기 제2 이합체화 도메인의 3'이고 상기 분지점 서열의 5'인 ISE, 상기 제2 스플라이스 도너의 3'이고, 상기 제2 이합체화 도메인의 5'인 DISE, 상기 제2 스플라이스 도너의 3'이고 상기 제3 이합체화 도메인의 5'인 ISE, 또는 이들의 조합을 더 포함하고; 및/또는

[0228] 상기 제3 합성 핵산 분자는 상기 제4 이합체화 도메인의 3'이고 상기 제2 분지점 서열의 5'인 ISE를 더 포함하고; 및

[0229] 이들의 조합인 것인 시스템.

[0230] 23. 구현예 3 내지 20 중 어느 하나에 있어서,

[0231] 상기 제1 합성 핵산 분자는 상기 제1 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 DISE, 상기 제1 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 ISE, 또는 DISE 및 ISE 모두를 더 포함하고;

[0232] 상기 제2 합성 핵산 분자는 상기 제2 이합체화 도메인의 3'이고 상기 제1 분지점 서열의 5'인 ISE, 상기 제2 스플라이스 도너의 3'이고, 상기 제2 이합체화 도메인의 5'인 DISE, 상기 제2 스플라이스 도너의 3'이고 상기 제3 이합체화 도메인의 5'인 ISE, 또는 이들의 조합을 더 포함하고;

[0233] 상기 제3 합성 핵산 분자는 상기 제4 이합체화 도메인의 3'이고 상기 제2 분지점 서열의 5'인 ISE를 더 포함하고; 및/또는

[0234] 상기 제4 합성 핵산 분자는 상기 제5 이합체화 도메인의 3'이고 상기 제3 분지점 서열의 5'인 ISE, 상기 제3 스플라이스 도너의 3'이고, 상기 제5 이합체화 도메인의 5'인 DISE, 상기 제3 스플라이스 도너의 3'이고 상기 제6 이합체화 도메인의 5'인 ISE, 또는 이들의 조합을 더 포함하고, 및

[0235] 이들의 조합인 것인 시스템.

[0236] 24. 구현예 1 내지 23 중 어느 하나에 있어서, 상기 시스템이 세포에 도입되는 경우, 상기 RNA 분자가 생성되고 적합한 순서로 재조합되어, 상기 표적 단백질의 전장 코딩 서열이 형성되는 것인 시스템.

[0237] 25. 구현예 1 내지 24 중 어느 하나에 있어서, 상기 제1, 제2, 제3, 및 제4 합성 핵산 분자 각각은 별개의 바이러스 벡터의 부분인 것인 시스템.

[0238] 26. 구현예 25에 있어서, 상기 바이러스 벡터는 AAV인 것인 시스템.

[0239] 27. 구현예 1 내지 26 중 어느 하나에 있어서,

[0240] 상기 제1 및/또는 제3 합성 핵산 분자는 상기 스플라이스 도너의 3'으로부터 임의의 위치에 배치된(positioned anywhere 3' to the splice donor) 자가-절단 RNA 서열 또는 RNA-절단 효소 표적 서열을 더 포함하여, 3'에 위치한 폴리아데닐화 테일을 절단하여 비-재조합(non-recombined) RNA 분자로부터의 단백질 단편 발현을 감소시키거나 억제하거나;

[0241] 상기 제2 및/또는 제4 합성 핵산 분자는 상기 분지점 서열의 5'으로부터 임의의 위치에 배치된 자가-절단 RNA 서열 또는 RNA-절단 효소 표적 서열을 더 포함하여, 5'에 위치한 RNA 캡을 절단하여 비-재조합 RNA 분자로부터의 단백질 절편 발현을 감소시키거나 억제하거나;

[0242] 상기 제2 및/또는 제4 합성 핵산 분자는 상기 스플라이스 억셉터의 3'에 있는 ORF(open reading frame) 대비 이동된 분지점 서열의 5'으로부터 임의의 위치에 배치된 개시 코돈을 더 포함하여, 비-재조합 RNA 분자로부터 표적 단백질 절편의 번역을 감소시키거나 또는 억제하거나;

[0243] 상기 제1 및/또는 제3 합성 핵산 분자는 상기 스플라이스 도너의 3'으로부터 임의의 위치에 마이크로 RNA 표적 부위를 더 포함하여, 비-연결(un-joined) RNA 단편이 핵 외부에서 마이크로 RNA 의존적 분해를 겪게 하거나;

[0244] 상기 제2 및/또는 제4 합성 핵산 분자는 상기 코딩 서열의 3'으로부터 임의의 위치에 마이크로 RNA 표적 부위를 더 포함하여, 비-연결 RNA 단편이 핵 외부에서 마이크로 RNA 의존적 분해를 겪게 하거나;

[0245] 상기 제1 및/또는 제3 합성 핵산 분자는 상기 스플라이스 도너의 5'에 있는 표적 단백질 ORF와 동일한 프레임 내에 위치하도록(in frame with the target protein open reading frame 5' to the splice donor) 상기 스플라이스 도너의 3'으로부터 임의의 위치에 데그론(degron) 단백질 분해 태그를 코딩하는 서열을 더 포함하여, 비-연결 단백질 단편이 분해를 위해 태깅되게 하거나;

[0246] 상기 제2 및/또는 제4 합성 핵산 분자는 상기 스플라이스 억셉터 부위의 3'에 있는 표적 단백질 ORF와 동일한 프레임 내에 위치하도록 상기 분지점 서열의 5'으로부터 임의의 위치에 개시 코돈 및 인-프레임(in-frame) 데그론 단백질 분해 태그를 더 포함하여, 비-연결 단백질 단편이 분해를 위해 태깅되게 하거나; 또는

[0247] 이들의 조합인 것인 시스템.

[0248] 28. 구현예 1 내지 27 중 어느 하나에 있어서, 상기 시스템의 1개, 2개, 3개, 또는 4개의 합성 핵산 분자는 각각 약 2500 nt 내지 약 5000 nt, 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내

지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,500 nt 내지 약 4,750 nt, 약 2,500 nt 내지 약 5,000 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 4,750 nt, 약 2,750 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 4,750 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 4,750 nt, 약 3,250 nt 내지 약 5,000 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 4,750 nt, 약 3,500 nt 내지 약 5,000 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,750 nt, 약 3,750 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,750 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,250 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,750 nt, 약 4,250 nt 내지 약 5,000 nt, 약 4,500 nt 내지 약 4,750 nt, 약 4,500 nt 내지 약 5,000 nt, 약 4,750 nt 내지 약 5,000 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000 nt, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 약 4,500 nt, 약 4,750 nt, 및 약 5,000 nt로부터 독립적으로 선택된 크기를 갖는 것인 시스템.

[0249] 29. 구현예 1 내지 28 중 어느 하나에 있어서, 상기 시스템의 합성 핵산 분자에 의해 코딩되는 표적 단백질의 N-말단 부분, 상기 표적 단백질의 중간 부분, 또는 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열은 각각: 약 1000 nt 내지 약 4000 nt, 약 1,000 nt 내지 약 1,500 nt, 약 1,000 nt 내지 약 2,000 nt, 약 1,000 nt 내지 약 2,500 nt, 약 1,000 nt 내지 약 3,000 nt, 약 1,000 nt 내지 약 3,500 nt, 약 1,000 nt 내지 약 4,000 nt, 약 1,500 nt 내지 약 2,000 nt, 약 1,500 nt 내지 약 2,500 nt, 약 1,500 nt 내지 약 3,000 nt, 약 1,500 nt 내지 약 3,500 nt, 약 1,500 nt 내지 약 4,000 nt, 약 2,000 nt 내지 약 2,500 nt, 약 2,000 nt 내지 약 3,000 nt, 약 2,000 nt 내지 약 3,500 nt, 약 2,000 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,000 nt, 약 1,000 nt, 약 1,500 nt, 약 2,000 nt, 약 2,500 nt, 약 3,000 nt, 약 3,500 nt, 및 약 4,000 nt로부터 독립적으로 선택된 크기를 갖는 것인 시스템.

[0250] 30. 구현예 1 내지 29 중 어느 하나에 있어서, 상기 시스템의 1개, 2개, 3개, 또는 4개의 합성 핵산 분자에 의해 코딩되는 1개, 3개, 3개, 또는 4개의 RNA 분자는 각각: 약 2500 내지 4500 nt, 약 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,500 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000 nt, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 및 약 4,500 nt로부터 독립적으로 선택되는 크기를 갖는 것인 시스템.

[0251] 31. 구현예 1 및 4 내지 30 중 어느 하나에 있어서

[0252] 상기 합성 핵산 분자는 약 5000 nt 내지 약 10,000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 9,500 nt, 약 5,000 nt 내지 약 10,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 8,500 nt, 약 5,500 nt 내지 약 9,000 nt, 약 5,500 nt 내지 약 9,500 nt, 약 5,500 nt 내지 약

10,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 9,500 nt, 약 6,000 nt 내지 약 10,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 8,500 nt, 약 6,500 nt 내지 약 9,000 nt, 약 6,500 nt 내지 약 9,500 nt, 약 6,500 nt 내지 약 10,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 9,500 nt, 약 7,000 nt 내지 약 10,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 9,500 nt, 약 7,500 nt 내지 약 10,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 9,500 nt, 약 8,000 nt 내지 약 10,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,500 nt, 약 8,500 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 9,500 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,500 nt 내지 약 10,000 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 약 9,000 nt, 약 9,500 nt, 및 약 10,000 nt로부터 선택된 전체 크기를 갖고;

[0253] 전체 표적 단백질 코딩 서열은 약 2000 nt 내지 약 8000 nt, 약 2,000 nt 내지 약 3,000 nt, 약 2,000 nt 내지 약 3,500 nt, 약 2,000 nt 내지 약 4,000 nt, 약 2,000 nt 내지 약 4,500 nt, 약 2,000 nt 내지 약 5,000 nt, 약 2,000 nt 내지 약 5,500 nt, 약 2,000 nt 내지 약 6,000 nt, 약 2,000 nt 내지 약 6,500 nt, 약 2,000 nt 내지 약 7,000 nt, 약 2,000 nt 내지 약 7,500 nt, 약 2,000 nt 내지 약 8,000 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 5,500 nt, 약 3,000 nt 내지 약 6,000 nt, 약 3,000 nt 내지 약 6,500 nt, 약 3,000 nt 내지 약 7,000 nt, 약 3,000 nt 내지 약 7,500 nt, 약 3,000 nt 내지 약 8,000 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 5,000 nt, 약 3,500 nt 내지 약 5,500 nt, 약 3,500 nt 내지 약 6,000 nt, 약 3,500 nt 내지 약 6,500 nt, 약 3,500 nt 내지 약 7,000 nt, 약 3,500 nt 내지 약 7,500 nt, 약 3,500 nt 내지 약 8,000 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 5,500 nt, 약 4,000 nt 내지 약 6,000 nt, 약 4,000 nt 내지 약 6,500 nt, 약 4,000 nt 내지 약 7,000 nt, 약 4,000 nt 내지 약 7,500 nt, 약 4,000 nt 내지 약 8,000 nt, 약 4,500 nt 내지 약 5,000 nt, 약 4,500 nt 내지 약 5,500 nt, 약 4,500 nt 내지 약 6,000 nt, 약 4,500 nt 내지 약 6,500 nt, 약 4,500 nt 내지 약 7,000 nt, 약 4,500 nt 내지 약 7,500 nt, 약 4,500 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 또는 약 7,500 nt 내지 약 8,000 nt로부터 선택되고, 상기 전체 표적 단백질 코딩 서열은 약 2,000 nt, 약 3,000 nt, 약 3,500 nt, 약 4,000 nt, 약 4,500 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 및 약 8,000 nt이고; 및/또는

[0254] 상기 2개의 합성 핵산 분자에 의해 코딩되는 RNA 분자는 약 5,000 nt 내지 약 9000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 8,500 nt, 약 5,500 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 8,500 nt, 약 6,500 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, or 약 8,500 nt 내지 약 9,000 nt로부터 선택된 전체 크기를 갖고, 상기 2개의 합성 핵산 분자에 의해 코딩되는 RNA 분자는 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 및 약 9,000 nt의 전체 크기를 갖는 것인 시스템.

[0255] 32. 구현예 2 및 4 내지 30 중 어느 하나에 있어서,

[0256] 상기 합성 핵산 분자는 약 7500 nt 내지 약 15,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,500 nt, 약 7,500 nt 내지 약 10,000 nt, 약 7,500 nt 내지 약 10,500 nt, 약 7,500 nt 내지 약 11,000 nt, 약 7,500 nt 내지 약 11,500 nt, 약 7,500 nt 내지 약 12,000 nt, 약 7,500 nt 내지 약 12,500 nt, 약 7,500 nt 내지 약 13,000 nt, 약 7,500 nt 내지 약 14,000 nt, 약 7,500 nt 내지 약 15,000 nt, 약 8,500 nt 내지 약 9,500 nt, 약 8,500 nt 내지 약 10,000 nt, 약 8,500 nt 내지 약 10,500 nt, 약 8,500 nt 내지 약 11,000 nt, 약 8,500 nt 내지 약 11,500 nt, 약 8,500 nt 내지 약 12,000 nt, 약 8,500 nt 내지 약 12,500 nt, 약 8,500 nt 내지 약 13,000 nt, 약 8,500 nt 내지 약 14,000 nt, 약 8,500 nt 내지 약 15,000 nt, 약 9,500 nt 내지 약 10,000 nt, 약 9,500 nt 내지 약 10,500 nt, 약 9,500 nt 내지 약 11,000 nt, 약 9,500 nt 내지 약 11,500 nt, 약 9,500 nt 내지 약 12,000 nt, 약 9,500 nt 내지 약 12,500 nt, 약 9,500 nt 내지 약 13,000 nt, 약 9,500 nt 내지 약 14,000 nt, 약 9,500 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 10,500 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 11,500 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 12,500 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,500 nt 내지 약 11,000 nt, 약 10,500 nt 내지 약 11,500 nt, 약 10,500 nt 내지 약 12,000 nt, 약 10,500 nt 내지 약 12,500 nt, 약 10,500 nt 내지 약 13,000 nt, 약 10,500 nt 내지 약 14,000 nt, 약 10,500 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 11,500 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 12,500 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,500 nt 내지 약 12,000 nt, 약 11,500 nt 내지 약 12,500 nt, 약 11,500 nt 내지 약 13,000 nt, 약 11,500 nt 내지 약 14,000 nt, 약 11,500 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 12,500 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,500 nt 내지 약 13,000 nt, 약 12,500 nt 내지 약 14,000 nt, 약 12,500 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 또는 약 14,000 nt 내지 약 15,000 nt로부터 선택된 전체 크기를 갖고, 상기 합성 핵산 분자는 약 7,500 nt, 약 8,500 nt, 약 9,500 nt, 약 10,000 nt, 약 10,500 nt, 약 11,000 nt, 약 11,500 nt, 약 12,000 nt, 약 12,500 nt, 약 13,000 nt, 약 14,000 nt, 및 약 15,000 nt의 전체 크기를 갖고;

[0257] 전체 표적 단백질 코딩 서열은 약 3000 nt 내지 약 12,000 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 6,000 nt, 약 3,000 nt 내지 약 7,000 nt, 약 3,000 nt 내지 약 7,500 nt, 약 3,000 nt 내지 약 8,000 nt, 약 3,000 nt 내지 약 8,500 nt, 약 3,000 nt 내지 약 9,000 nt, 약 3,000 nt 내지 약 1,000 nt, 약 3,000 nt 내지 약 11,000 nt, 약 3,000 nt 내지 약 12,000 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 6,000 nt, 약 4,000 nt 내지 약 7,000 nt, 약 4,000 nt 내지 약 7,500 nt, 약 4,000 nt 내지 약 8,000 nt, 약 4,000 nt 내지 약 8,500 nt, 약 4,000 nt 내지 약 9,000 nt, 약 4,000 nt 내지 약 1,000 nt, 약 4,000 nt 내지 약 11,000 nt, 약 4,000 nt 내지 약 12,000 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 1,000 nt, 약 5,000 nt 내지 약 11,000 nt, 약 5,000 nt 내지 약 12,000 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 1,000 nt, 약 6,000 nt 내지 약 11,000 nt, 약 6,000 nt 내지 약 12,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 1,000 nt, 약 7,000 nt 내지 약 11,000 nt, 약 7,000 nt 내지 약 12,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 1,000 nt, 약 7,500 nt 내지 약 11,000 nt, 약 7,500 nt 내지 약 12,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 1,000 nt, 약 8,000 nt 내지 약 11,000 nt, 약 8,000 nt 내지 약 12,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 1,000 nt, 약 8,500 nt 내지 약 11,000 nt, 약 8,500 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 1,000 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 12,000 nt, 약 1,000 nt 내지 약 11,000 nt, 약 1,000 nt 내지 약 12,000 nt, 또는 약 11,000 nt 내지 약 12,000 nt로부터 선택되며, 상기 전체 표적 단백질 코딩 서열은 약 3,000 nt, 약 4,000 nt, 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 약 9,000 nt, 약 1,000 nt, 약 11,000 nt, 및 약 12,000 nt이고; 및/또는

[0258] 상기 3개의 합성 핵산 분자에 의해 코딩되는 RNA는 약 7500 nt 내지 약 13,500 nt, 약 7,500 nt 내지 약 8,500

nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 9,500 nt, 약 7,500 nt 내지 약 10,000 nt, 약 7,500 nt 내지 약 10,500 nt, 약 7,500 nt 내지 약 11,000 nt, 약 7,500 nt 내지 약 11,500 nt, 약 7,500 nt 내지 약 12,000 nt, 약 7,500 nt 내지 약 12,500 nt, 약 7,500 nt 내지 약 13,000 nt, 약 7,500 nt 내지 약 13,500 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,500 nt, 약 8,500 nt 내지 약 10,000 nt, 약 8,500 nt 내지 약 10,500 nt, 약 8,500 nt 내지 약 11,000 nt, 약 8,500 nt 내지 약 11,500 nt, 약 8,500 nt 내지 약 12,000 nt, 약 8,500 nt 내지 약 12,500 nt, 약 8,500 nt 내지 약 13,000 nt, 약 8,500 nt 내지 약 13,500 nt, 약 9,000 nt 내지 약 9,500 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 10,500 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 11,500 nt, 약 9,000 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 12,500 nt, 약 9,000 nt 내지 약 13,000 nt, 약 9,000 nt 내지 약 13,500 nt, 약 9,500 nt 내지 약 10,000 nt, 약 9,500 nt 내지 약 10,500 nt, 약 9,500 nt 내지 약 11,000 nt, 약 9,500 nt 내지 약 11,500 nt, 약 9,500 nt 내지 약 12,000 nt, 약 9,500 nt 내지 약 12,500 nt, 약 9,500 nt 내지 약 13,000 nt, 약 9,500 nt 내지 약 13,500 nt, 약 10,000 nt 내지 약 10,500 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 11,500 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 12,500 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 13,500 nt, 약 10,500 nt 내지 약 11,000 nt, 약 10,500 nt 내지 약 11,500 nt, 약 10,500 nt 내지 약 12,000 nt, 약 10,500 nt 내지 약 12,500 nt, 약 10,500 nt 내지 약 13,000 nt, 약 10,500 nt 내지 약 13,500 nt, 약 11,000 nt 내지 약 11,500 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 12,500 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 13,500 nt, 약 11,500 nt 내지 약 12,000 nt, 약 11,500 nt 내지 약 12,500 nt, 약 11,500 nt 내지 약 13,000 nt, 약 11,500 nt 내지 약 13,500 nt, 약 12,000 nt 내지 약 12,500 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 13,500 nt, 약 12,500 nt 내지 약 13,000 nt, 약 12,500 nt 내지 약 13,500 nt, 또는 약 13,000 nt 내지 약 13,500 nt로부터 선택된 전체 크기를 갖고, 상기 2개의 합성 핵산 분자에 의해 코딩되는 RNA는 약 7,500 nt, 약 8,500 nt, 약 9,000 nt, 약 9,500 nt, 약 10,000 nt, 약 10,500 nt, 약 11,000 nt, 약 11,500 nt, 약 12,000 nt, 약 12,500 nt, 약 13,000 nt, 및 약 13,500 nt의 전체 크기를 갖는 것인 시스템.

[0259] 33. 구현예 1 내지 23 중 어느 하나에 있어서,

[0260] 상기 합성 핵산 분자는 약 10,000 nt 내지 약 20,000 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 16,000 nt, 약 10,000 nt 내지 약 17,000 nt, 약 10,000 nt 내지 약 18,000 nt, 약 10,000 nt 내지 약 19,000 nt, 약 10,000 nt 내지 약 20,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 16,000 nt, 약 11,000 nt 내지 약 17,000 nt, 약 11,000 nt 내지 약 18,000 nt, 약 11,000 nt 내지 약 19,000 nt, 약 11,000 nt 내지 약 20,000 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 16,000 nt, 약 12,000 nt 내지 약 17,000 nt, 약 12,000 nt 내지 약 18,000 nt, 약 12,000 nt 내지 약 19,000 nt, 약 12,000 nt 내지 약 20,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 16,000 nt, 약 13,000 nt 내지 약 17,000 nt, 약 13,000 nt 내지 약 18,000 nt, 약 13,000 nt 내지 약 19,000 nt, 약 13,000 nt 내지 약 20,000 nt, 약 14,000 nt 내지 약 15,000 nt, 약 14,000 nt 내지 약 16,000 nt, 약 14,000 nt 내지 약 17,000 nt, 약 14,000 nt 내지 약 18,000 nt, 약 14,000 nt 내지 약 19,000 nt, 약 14,000 nt 내지 약 20,000 nt, 약 15,000 nt 내지 약 16,000 nt, 약 15,000 nt 내지 약 17,000 nt, 약 15,000 nt 내지 약 18,000 nt, 약 15,000 nt 내지 약 19,000 nt, 약 15,000 nt 내지 약 20,000 nt, 약 16,000 nt 내지 약 17,000 nt, 약 16,000 nt 내지 약 18,000 nt, 약 16,000 nt 내지 약 19,000 nt, 약 16,000 nt 내지 약 20,000 nt, 약 17,000 nt 내지 약 18,000 nt, 약 17,000 nt 내지 약 19,000 nt, 약 17,000 nt 내지 약 20,000 nt, 약 18,000 nt 내지 약 19,000 nt, 약 18,000 nt 내지 약 20,000 nt, 또는 약 19,000 nt 내지 약 20,000 nt로부터 선택된 전체 크기를 갖고, 상기 합성 핵산 분자는 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 약 15,000 nt, 약 16,000 nt, 약 17,000 nt, 약 18,000 nt, 약 19,000 nt, 및 약 20,000 nt의 전체 크기를 갖고;

[0261] 전체 표적 단백질 코딩 서열은 약 4000 nt 내지 약 16,000 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 10,000 nt, 약 5,000 nt 내지 약 11,000 nt, 약 5,000 nt 내지 약 12,000 nt, 약 5,000 nt 내지 약 13,000 nt, 약 5,000 nt 내지 약 14,000 nt, 약 5,000 nt 내지 약 15,000 nt, 약 5,000 nt 내지 약 16,000 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 10,000 nt, 약 6,000 nt 내지 약 11,000 nt, 약 6,000 nt 내지 약 12,000 nt, 약 6,000 nt 내지 약 13,000 nt,

약 6,000 nt 내지 약 14,000 nt, 약 6,000 nt 내지 약 15,000 nt, 약 6,000 nt 내지 약 16,000 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 10,000 nt, 약 7,000 nt 내지 약 11,000 nt, 약 7,000 nt 내지 약 12,000 nt, 약 7,000 nt 내지 약 13,000 nt, 약 7,000 nt 내지 약 14,000 nt, 약 7,000 nt 내지 약 15,000 nt, 약 7,000 nt 내지 약 16,000 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 10,000 nt, 약 8,000 nt 내지 약 11,000 nt, 약 8,000 nt 내지 약 12,000 nt, 약 8,000 nt 내지 약 13,000 nt, 약 8,000 nt 내지 약 14,000 nt, 약 8,000 nt 내지 약 15,000 nt, 약 8,000 nt 내지 약 16,000 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 13,000 nt, 약 9,000 nt 내지 약 14,000 nt, 약 9,000 nt 내지 약 15,000 nt, 약 9,000 nt 내지 약 16,000 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 16,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 16,000 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 16,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 16,000 nt, 약 14,000 nt 내지 약 15,000 nt, 약 14,000 nt 내지 약 16,000 nt, 또는 약 15,000 nt 내지 약 16,000 nt로부터 선택되고, 상기 전체 표적 단백질 코딩 서열은 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 8,000 nt, 약 9,000 nt, 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 약 15,000 nt, 또는 약 16,000 nt이고, 상기 전체 표적 단백질 코딩 서열은 적어도 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 8,000 nt, 약 9,000 nt, 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 및 약 15,000 nt이고; 및/또는

[0262] 상기 2개의 합성 핵산 분자에 의해 코딩되는 RNA는 약 10,000 nt 내지 약 18,000 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 16,000 nt, 약 10,000 nt 내지 약 17,000 nt, 약 10,000 nt 내지 약 18,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 16,000 nt, 약 11,000 nt 내지 약 17,000 nt, 약 11,000 nt 내지 약 18,000 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 16,000 nt, 약 12,000 nt 내지 약 17,000 nt, 약 12,000 nt 내지 약 18,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 16,000 nt, 약 13,000 nt 내지 약 17,000 nt, 약 13,000 nt 내지 약 18,000 nt, 약 14,000 nt 내지 약 15,000 nt, 약 14,000 nt 내지 약 16,000 nt, 약 14,000 nt 내지 약 17,000 nt, 약 14,000 nt 내지 약 18,000 nt, 약 15,000 nt 내지 약 16,000 nt, 약 15,000 nt 내지 약 17,000 nt, 약 15,000 nt 내지 약 18,000 nt, 약 16,000 nt 내지 약 17,000 nt, 약 16,000 nt 내지 약 18,000 nt, 또는 약 17,000 nt 내지 약 18,000 nt로부터 선택된 전체 크기를 갖고, 상기 2개의 합성 핵산 분자에 의해 코딩되는 RNA 분자는 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 약 15,000 nt, 약 16,000 nt, 약 17,000 nt, 및 약 18,000 nt의 전체 크기를 갖는 것인 시스템.

[0263] 34. 구현예 1 내지 33 중 어느 하나에 있어서, RNA 재조합 효율은 약 10% 내지 약 95%, 약 10% 내지 약 20%, 약 10% 내지 약 30%, 약 10% 내지 약 35%, 약 10% 내지 약 40%, 약 10% 내지 약 45%, 약 10% 내지 약 50%, 약 10% 내지 약 55%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 20% 내지 약 30%, 약 20% 내지 약 35%, 약 20% 내지 약 40%, 약 20% 내지 약 45%, 약 20% 내지 약 50%, 약 20% 내지 약 55%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 30% 내지 약 35%, 약 30% 내지 약 40%, 약 30% 내지 약 45%, 약 30% 내지 약 50%, 약 30% 내지 약 55%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 35% 내지 약 40%, 약 35% 내지 약 45%, 약 35% 내지 약 50%, 약 35% 내지 약 55%, 약 35% 내지 약 60%, 약 35% 내지 약 70%, 약 35% 내지 약 80%, 약 35% 내지 약 90%, 약 40% 내지 약 45%, 약 40% 내지 약 50%, 약 40% 내지 약 55%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 45% 내지 약 50%, 약 45% 내지 약 55%, 약 45% 내지 약 60%, 약 45% 내지 약 70%, 약 45% 내지 약 80%, 약 45% 내지 약 90%, 약 50% 내지 약 55%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 55% 내지 약 60%, 약 55% 내지 약 70%, 약 55% 내지 약 80%, 약 55% 내지 약 90%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 80% 내지 약 90%, 약 10%, 약 20%, 약 30%, 약 35%, 약 40%, 약 45%, 약 50%, 약 55%, 약 60%, 약 70%, 약 80%, 또는 약 90%, 또는 약 95%인 것인 시스템.

[0264] 35. 구현예 1 내지 34 중 어느 하나에 있어서, 상기 제1 이합체 도메인과 상기 제2 이합체 도메인, 상기 제3 이합체 도메인과 상기 제3 이합체 도메인, 및/또는 상기 제5 이합체 도메인과 상기 제6 이합체 도메인은 각각 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 to 1000 nt, 50 to 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 상기 시스템은 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 55%, 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 또는 적어도 약 95%의 재조합 효율을 갖는 것인 시스템.

[0265] 36. 구현예 1 내지 35 중 어느 하나에 있어서, 각각의 이합체화 도메인은 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 및 상기 시스템은 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 75%, 적어도 80%, 또는 적어도 90%의 재조합 효율을 갖는 것인 시스템.

[0266] 37. 구현예 1 내지 36 중 어느 하나의 시스템을 포함하는 조성물.

[0267] 38. 구현예 1 내지 37 중 어느 하나에 기재된 RNA 분자를 포함하는 조성물.

[0268] 39. 구현예 1 내지 37 중 어느 하나에 기재된 RNA 분자를 1개, 2개, 3개, 또는 4개 포함하는 조성물.

[0269] 40. 구현예 37 내지 39 중 어느 하나에 있어서, 상기 조성물은 제1, 제2, 제3, 및 선택적으로 제4 합성 핵산 분자 또는 RNA 분자를 포함하고, 각각은 디스트로핀(dystrophin), 인자 8 (factor 8), ABCA4, 또는 MYO7A의 적어도 일부를 코딩하는 것인 조성물.

[0270] 41. 구현예 1 내지 36 중 어느 하나에 기재된 RNA 분자.

[0271] 42. 구현예 1 내지 41 중 어느 하나의 시스템, 또는 구현예 37 내지 40 중 어느 하나의 조성물을 포함하는 키트로서, 상기 제1, 제2, 제3, 및 제4 합성 핵산 분자는 별개의 용기에 담길 수 있고, 선택적으로 약학적으로 허용 가능한 담체와 같은 완충액을 더 포함하는 것인 키트.

[0272] 43. 세포에서 표적 단백질을 발현시키는 방법으로서,

[0273] 구현예 1 내지 36 중 어느 하나의 시스템, 또는 구현예 35 내지 37 중 어느 하나의 조성물을 세포 내로 도입시키는 단계, 및 상기 세포에서 제1 및 제2 합성 RNA 분자, 제1, 제2, 및 제3 합성 RNA 분자, 또는 제1, 제2, 제3, 및 제4 합성 RNA 분자를 발현시키는 단계를 포함하고, 상기 표적 단백질이 상기 세포에서 생산되는 것인 방법.

[0274] 44. 구현예 43에 있어서, 상기 세포는 개체 내에 존재하고, 상기 도입시키는 단계는 상기 시스템의 치료 유효량을 상기 개체에게 투여하는 단계를 포함하는 것인 방법.

[0275] 45. 구현예 44에 있어서, 상기 방법은 상기 개체에서 상기 표적 단백질을 코딩하는 유전자 중 돌연변이에 의해 유발된 유전 질환을 치료하고, 상기 방법은 상기 개체에서 기능성 표적 단백질의 발현을 가져오는 것인 방법.

[0276] 46. 구현예 45에 있어서,

[0277] 상기 유전 질환은 듀센 근이영양증 (Duchenne muscular dystrophy)이고, 상기 표적 단백질은 디스트로핀(dystrophin)이거나;

[0278] 상기 유전 질환은 A형 혈우병이고, 상기 표적 단백질은 F8이거나;

[0279] 상기 유전 질환은 스타르가르트병(Stargardt disease)이고, 상기 표적 단백질은 ABCA4이거나; 또는

[0280] 상기 유전 질환은 어셔 증후군(Usher syndrome)이고, 상기 표적 단백질은 MYO7A인 것인 방법.

[0281] 47. 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166 중 어느 하나로 제공되는 합성 인트론에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100% 서열 동일성을 포함하는 합성 핵산 분자.

[0282] 48. 구현예 47에 있어서, 상기 합성 인트론은 서열번호 20의 nt 3703 내지 3975, 서열번호 21의 nt 1 내지

228, 서열번호 22의 nt 3703 내지 3975, 서열번호 23의 nt 1 내지 225, 서열번호 24의 nt 3560 내지 3828, 및 서열번호 25의 nt 1-225인 것인 합성 핵산 분자.

[0283] 49. 구현예 47 또는 48에 있어서, 단백질 코딩 서열의 부분을 더 포함하는 것인 합성 핵산 분자.

[0284] 50. 구현예 48에 있어서, 상기 단백질 코딩 서열의 부분은 상기 단백질 코딩 서열의 N-말단 절반(N-terminal half), N-말단 1/3(N-terminal third), 중간 부분, C-말단 절반, 또는 C-말단 1/3을 포함하는 것인 합성 핵산 분자.

[0285] 51. 구현예 1 내지 36 중 어느 하나, 또는 구현예 37 내지 40 중 어느 하나에 있어서, 적어도 하나의 합성 핵산 분자는 구현예 47 내지 50 중 어느 하나에 기재된 핵산 분자를 포함하는 합성 인트론을 포함하는 것인 시스템 또는 조성물.

[0286] 52. 전술된 구현예에 있어서, 상기 합성 핵산은 역전사 효소에 의해 RNA 바이러스 게놈의 전사에 의해 생산된 DNA인 것인 조성물, 시스템, 방법, 또는 키트.

[0287] **VII. 추가적인 예시적 구체예**

[0288] 1. 표적 단백질의 발현용 조성물로서, (a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너; 및 (iii) 제 1 이합체화 도메인;을 포함하는 것인 제1 RNA 분자; 및 (b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제2 RNA 분자를 포함하는 것인 조성물.

[0289] 2. 표적 단백질의 발현용 조성물로서, (a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너; 및 (iii) 제 1 이합체화 도메인;을 포함하는 것인 제1 RNA 분자; (b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 중간 부분에 대한 코딩 서열; (vi) 제2 스플라이스 도너; 및 (vii) 제3 이합체화 도메인을 포함하는 것인 제2 RNA 분자; 및 (c) 제3 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로: (i) 제4 이합체화 도메인으로서, 상기 제4 이합체화 도메인은 상기 제3 이합체화 도메인에 결합하는 것인 제4 이합체화 도메인; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제3 RNA 분자를 포함하는 것인 조성물.

[0290] 3. 표적 단백질의 발현용 조성물로서, (a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너; 및 (iii) 제 1 이합체화 도메인;을 포함하는 것인 제1 RNA 분자; (b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 중간 부분에 대한 코딩 서열; (vi) 제2 스플라이스 도너; 및 (vii) 제3 이합체화 도메인을 포함하는 것인 제2 RNA 분자; 및 (c) 제3 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로: (i) 제4 이합체화 도메인으로서, 상기 제4 이합체화 도메인은 상기 제3 이합체화 도메인에 결합하는 것인 제4 이합체화 도메인; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; (v) 상기 표적 단백질의 제1 중간 부분(a first middle portion)에 대한 코딩 서열; (vi) 제2 스플라이스 도너; 및 (vii) 제5 이합체화 도메인을 포함하는 것인 제3 RNA 분자; 및 (d) 제4 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로: (i) 제6 이합체화 도메인으로서, 상기 제6 이합체화 도메인은 상기 제5 이합체화 도메인에 결합하는 것인 제6 이합체화 도메인; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제4 RNA 분자를 포함하는 것인 조성물.

[0291] 4. 구현예 1 내지 3 중 하나에 있어서, 상기 제1 및 제2 이합체화 도메인, 상기 제3 및 제4 이합체화 도메인, 및/또는 상기 제5 및 제6 이합체화 도메인은 직접 결합, 간접 결합, 또는 이들의 조합에 의해 결합하는 것인 조성물.

[0292] 5. 구현예 4에 있어서, 직접 결합 또는 간접 결합은 염기쌍 형성 상호작용, 비-규범적 염기쌍 형성 상호작용,

비-염기쌍 형성 상호작용, 또는 이들의 조합을 포함하는 것인 조성물.

[0293] 6. 구현예 4 또는 5에 있어서, 직접 결합은 키싱 루프 또는 저다양성 영역간 염기쌍 형성 상호작용을 포함하는 것인 조성물.

[0294] 7. 구현예 4 또는 5에 있어서, 직접 결합은 압타머 영역간 규범적 염기쌍 형성 상호작용, 비-규범적 염기쌍 형성 상호작용, 비-염기쌍 형성 상호작용, 또는 이들의 조합을 포함하는 것인 조성물.

[0295] 8. 구현예 4 또는 5에 있어서, 간접 결합은 핵산 브릿지를 통한 염기쌍 형성 상호작용을 포함하는 것인 조성물.

[0296] 9. 구현예 4 또는 5에 있어서, 간접 결합은 압타머와 압타머 표적 물질 (aptamer target agent)간, 또는 2개의 압타머간 비-염기쌍 형성 상호작용을 포함하는 것인 조성물.

[0297] 10. 구현예 1 내지 9 중 어느 하나에 있어서, 상기 제1, 제2, 제3, 제4, 제5, 및/또는 제6 이합체화 도메인은 크립틱 스플라이스 억셉터를 포함하지 않는 것인 조성물.

[0298] 11. 구현예 1 내지 10 중 어느 하나에 있어서, 적어도 한쌍의 직접적으로 또는 간접적으로 결합하는 압타머 서열 이합체화 도메인을 포함하는 것인 조성물.

[0299] 12. 구현예 1 내지 11 중 어느 하나에 있어서, 적어도 한쌍의 키싱 루프 상호작용 도메인을 포함하는 것인 조성물.

[0300] 13. 구현예 1 내지 12 중 어느 하나에 있어서, 상기 표적 단백질은 질병과 연관된 단백질, 또는 치료 단백질인 것인 조성물.

[0301] 14. 구현예 13에 있어서, 상기 질병은 단일유전자성(monogenic) 질병인 것인 조성물.

[0302] 15. 구현예 14에 있어서, 상기 치료 단백질은 독소인 것인 조성물.

[0303] 16. 구현예 13 내지 15 중 어느 하나에 있어서, 상기 질병 및 상기 표적 단백질은 표 1에 열거된 것인 조성물.

[0304] 17. 구현예 1 내지 16 중 어느 하나에 있어서, 상기 제1, 제2, 제3, 및/또는 제4 RNA 분자는 상기 제1, 제2, 제3, 또는 제4 RNA 분자의 3'-말단에 폴리 A 테일을 더 포함하는 것인 조성물.

[0305] 18. 구현예 1 또는 4 내지 17 중 어느 하나에 있어서,

[0306] 상기 제1 RNA 분자는 상기 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 DISE(downstream intronic splice enhancer), 상기 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 ISE (intronic splice enhancer), 또는 둘 모두를 더 포함하고; 및/또는

[0307] 상기 제2 RNA 분자는 상기 제2 이합체화 도메인의 3'이고 상기 분지점 서열의 5'인 ISE 및 상기 스플라이스 도너의 3'이고, 상기 이합체화 도메인의 5'인 DISE 중 하나, 또는 둘 모두를 더 포함하거나; 또는

[0308] 이들의 조합인 것인 조성물.

[0309] 19. 구현예 2 또는 4 내지 17 중 어느 하나에 있어서,

[0310] 상기 제1 RNA 분자는 상기 제1 스플라이스 도너의 3'이고 상기 제1 이합체화 도메인의 5'인 DISE, 상기 제1 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 ISE, 또는 둘 모두를 더 포함하고;

[0311] 상기 제2 RNA 분자는 상기 제2 이합체화 도메인의 3'이고 상기 제1 분지점 서열의 5'인 ISE, 상기 제2 스플라이스 도너의 3'이고, 상기 제2 이합체화 도메인의 5'인 DISE, 상기 제2 스플라이스 도너의 3'이고, 상기 제3 이합체화 도메인의 5'인 ISE, 또는 이들의 조합을 더 포함하고; 및/또는

[0312] 상기 제3 RNA 분자는 상기 제4 이합체화 도메인의 3'이고 상기 제2 분지점 서열의 5'인 ISE를 더 포함하거나, 또는

[0313] 이들의 조합인 것인 조성물.

[0314] 20. 구현예 3 내지 17 중 어느 하나에 있어서,

[0315] 상기 제1 RNA 분자는 상기 제1 스플라이스 도너의 3'이고 상기 제1 이합체화 도메인의 5'인 DISE, 상기 제1 스플라이스 도너의 3' 및 상기 제1 이합체화 도메인의 5'인 ISE 중 하나, 또는 둘 모두를 더 포함하고;

[0316] 상기 제2 RNA 분자는 상기 제2 이합체화 도메인의 3'이고 상기 제1 분지점 서열의 5'인 ISE, 상기 제2 스플라이

스 도너의 3'이고, 상기 제2 이합체화 도메인의 5'인 DISE, 상기 제2 스플라이스 도너의 3'이고, 상기 제3 이합체화 도메인의 5'인 ISE, 또는 이들의 조합을 더 포함하고; 및/또는

[0317] 상기 제3 RNA 분자는 상기 제4 이합체화 도메인의 3'이고 상기 제2 분지점 서열의 5'인 ISE를 더 포함하고; 및/또는

[0318] 상기 제4 RNA 분자는 상기 제5 이합체화 도메인의 3'이고 상기 제3 분지점 서열의 5'인 ISE, 상기 제3 스플라이스 도너의 3'이고, 상기 제5 이합체화 도메인의 5'인 DISE, 상기 제3 스플라이스 도너의 3'이고, 상기 제6 이합체화 도메인의 5'인 ISE, 또는 이들의 조합을 더 포함하거나;

[0319] 또는 이들의 조합인 것인 조성물.

[0320] 24. 구현예 1 내지 23 중 어느 하나에 있어서,

[0321] 상기 제1 및/또는 제3 RNA 분자는 상기 스플라이스 도너의 3'으로부터 임의의 위치에 배치된 자가-절단 RNA 서열 또는 RNA-절단 효소 표적 서열을 더 포함하여, 3'에 위치한 폴리아데닐화 테일을 절단하여 비-재조합 RNA 분자로부터의 단백질 단편 발현을 감소시키거나 억제하거나;

[0322] 상기 제2 및/또는 제4 RNA 분자는 상기 분지점 서열의 5'으로부터 임의의 위치에 배치된 자가-절단 RNA 서열 또는 RNA-절단 효소 표적 서열을 더 포함하여, 5'에 위치한 RNA 캡을 절단하여 비-재조합 RNA 분자로부터의 단백질 절편 발현을 감소시키거나 억제하거나;

[0323] 상기 제2 및/또는 제4 RNA 분자는 상기 스플라이스 억셉터의 3'에 있는 ORF(open reading frame) 대비 이동된 분지점 서열의 5'으로부터 임의의 위치에 배치된 개시 코돈을 더 포함하여, 비-재조합 RNA 분자로부터 표적 단백질 절편의 번역을 감소시키거나 또는 억제하거나;

[0324] 상기 제1 및/또는 제3 RNA 분자는 상기 스플라이스 도너의 3'으로부터 임의의 위치에 마이크로 RNA 표적 부위를 더 포함하여, 비-연결(un-joined) RNA 단편이 핵 외부에서 마이크로 RNA 의존적 분해를 겪게 하거나;

[0325] 상기 제2 및/또는 제4 RNA 분자는 상기 코딩 서열의 3'으로부터 임의의 위치에 마이크로 RNA 표적 부위를 더 포함하여, 비-연결 RNA 단편이 핵 외부에서 마이크로 RNA 의존적 분해를 겪게 하거나;

[0326] 상기 제1 및/또는 제3 RNA 분자는 상기 스플라이스 도너 부위의 5'에 있는 표적 단백질 ORF와 동일한 프레임 내에 위치하도록 상기 스플라이스 도너의 3'으로부터 임의의 위치에 데그론(degron) 단백질 분해 태그를 코딩하는 서열을 더 포함하여, 비-연결 단백질 단편이 분해를 위해 태깅되게 하거나;

[0327] 상기 제2 및/또는 제4 RNA 분자는 상기 스플라이스 억셉터 부위의 3'에 있는 표적 단백질 ORF와 동일한 프레임 내에 위치하도록 상기 분지점 서열의 5'으로부터 임의의 위치에 개시 코돈 및 인-프레임(in-frame) 데그론 단백질 분해 태그를 더 포함하여, 비-연결 단백질 단편이 분해를 위해 태깅되게 하거나; 또는

[0328] 이들의 조합인 것인 조성물.

[0329] 25. 표적 단백질의 발현용 조성물로서: (a) 구현예 1 및 4 내지 24 중 어느 하나의 제1 RNA 분자를 코딩하는 제1 합성 DNA 분자로서, 상기 제1 합성 DNA 분자는 (i) 상기 제1 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제1 프로모터를 포함하는 것인 제1 합성 DNA 분자; 및 (b) 구현예 1 및 4 내지 24 중 어느 하나의 제2 RNA 분자를 코딩하는 제2 합성 DNA 분자로서, 상기 제2 합성 DNA 분자는 (i) 상기 제2 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제2 프로모터를 포함하는 것인 제2 합성 DNA 분자를 포함하는 것인 조성물.

[0330] 26. 표적 단백질의 발현용 조성물로서: (a) 구현예 2 및 4 내지 24 중 어느 하나의 제1 RNA 분자를 코딩하는 제1 합성 DNA 분자로서, 상기 제1 합성 DNA 분자는 (i) 상기 제1 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제1 프로모터를 포함하는 것인 제1 합성 DNA 분자; (b) 구현예 2 및 4 내지 24 중 어느 하나의 제2 RNA 분자를 코딩하는 제2 합성 DNA 분자로서, 상기 제2 합성 DNA 분자는 (i) 상기 제2 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제2 프로모터를 포함하는 것인 제2 합성 DNA 분자; 및 (c) 구현예 2 및 4 내지 24 중 어느 하나의 제3 RNA 분자를 코딩하는 제3 합성 DNA 분자로서, 상기 제3 합성 DNA 분자는 (i) 상기 제3 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제3 프로모터를 포함하는 것인 제3 합성 DNA 분자를 포함하는 것인 조성물.

[0331] 27. 표적 단백질의 발현용 조성물로서: (a) 구현예 3 및 4 내지 24 중 어느 하나의 제1 RNA 분자를 코딩하는 제1 합성 DNA 분자로서, 상기 제1 합성 DNA 분자는 (i) 상기 제1 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제1 프로모터를 포함하는 것인 제1 합성 DNA 분자; (b) 구현예 3 및 4 내지 24 중 어느 하나의 제2 RNA 분자

를 코딩하는 제2 합성 DNA 분자로서, 상기 제2 합성 DNA 분자는 (i) 상기 제2 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제2 프로모터를 포함하는 것인 제2 합성 DNA 분자; (c) 구현예 3 및 4 내지 24 중 어느 하나의 제3 RNA 분자를 코딩하는 제3 합성 DNA 분자로서, 상기 제3 합성 DNA 분자는 (i) 상기 제3 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제3 프로모터를 포함하는 것인 제3 합성 DNA 분자; 및 (d) 구현예 3 및 4 내지 24 중 어느 하나의 제4 RNA 분자를 코딩하는 제4 합성 DNA 분자로서, 상기 제4 합성 DNA 분자는 (i) 상기 제4 RNA 분자를 코딩하는 서열에 작동가능하게 연결된 제4 프로모터를 포함하는 것인 제4 합성 DNA 분자를 포함하는 것인 조성물.

[0332] 28. 구현예 25 내지 27 중 어느 하나에 있어서, 각 프로모터는 독립적으로 선택되는 것인 조성물.

[0333] 29. 구현예 25 내지 28 중 어느 하나에 있어서,

[0334] 상기 제1 및 제2 프로모터는 동일한 프로모터이거나;

[0335] 상기 제1 및 제2 프로모터는 상이한 프로모터이거나;

[0336] 상기 제1, 제2, 및 제3 프로모터는 동일한 프로모터이거나;

[0337] 상기 제1, 제2, 및 제3 프로모터는 상이한 프로모터이거나;

[0338] 상기 제1, 제2, 제3, 및 제4 프로모터는 동일한 프로모터이거나; 또는

[0339] 상기 제1, 제2, 제3, 및 제4 프로모터는 상이한 프로모터인 것인 조성물.

[0340] 30. 구현예 25 내지 29 중 어느 하나에 있어서, 상기 제1, 제2, 제3, 및 제4 프로모터 각각은 항시적 프로모터; 조직-특이적 프로모터; 및 상기 표적 단백질에 내생적인 프로모터로부터 독립적으로 선택되는 것인 조성물.

[0341] 31. 구현예 25 내지 30 중 어느 하나의 조성물을 포함하는, 표적 단백질의 발현을 위한 시스템.

[0342] 32. 구현예 31에 있어서, 상기 시스템이 세포에 도입되는 경우, 상기 RNA 분자가 생성되고 적합한 순서로 재조합되어, 상기 표적 단백질의 전장 코딩 서열이 형성되는 것인 시스템.

[0343] 33. 구현예 31 또는 32에 있어서, 상기 제1 및 제2 RNA 분자 (2-파트 시스템(2-part system)의 경우) 각각, 상기 제1, 제2, 및 제3 RNA 분자 (3-파트 시스템의 경우) 각각, 또는 상기 제1, 제2, 제3, 및 제4 RNA 분자 (4-파트 시스템의 경우) 각각은 별개의 바이러스 벡터로부터 전사되는 것인 시스템.

[0344] 34. 구현예 31 내지 33 중 어느 하나에 있어서, 상기 바이러스 벡터는 AAV인 것인 시스템.

[0345] 35. 구현예 31 내지 34 중 어느 하나에 있어서, 상기 시스템의 제1, 제2, 제3, 또는 제4 합성 DNA 분자는 각각: 약 2500 nt 내지 약 5000 nt, 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,500 nt 내지 약 4,750 nt, 약 2,500 nt 내지 약 5,000 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 4,750 nt, 약 2,750 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 4,750 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 4,750 nt, 약 3,250 nt 내지 약 5,000 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 4,750 nt, 약 3,500 nt 내지 약 5,000 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,750 nt, 약 3,750 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,750 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,250 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,750 nt, 약 4,250 nt 내지 약 5,000 nt, 약 4,500 nt 내지 약 4,750 nt, 약 4,500 nt 내지 약 5,000 nt, 약 4,750 nt 내지 약 5,000 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000 nt, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 약 4,500 nt, 약 4,750 nt, 및 약 5,000 nt로부터 독립적으로 선택된 크기를 갖는 것인 시스템.

[0346] 36. 구현예 31 내지 35 중 어느 하나에 있어서, 상기 시스템의 합성 DNA 분자에 의해 코딩되는 표적 단백질의 N-말단 부분(2-파트, 3-파트, 또는 4-파트 시스템의 경우), 상기 표적 단백질의 중간 부분(3-파트 시스템의 경우), 상기 표적 단백질의 제1 중간 부분(4-파트 시스템의 경우), 또는 상기 표적 단백질의 C-말단 부분(2-파트, 3-파트, 또는 4-파트 시스템의 경우)에 대한 코딩 서열은 각각: 약 1,000 nt 내지 약 4,500 nt으로부터 독립적으로 선택된 크기를 갖고, 각각: 약 1,000 nt 내지 약 1,500 nt, 약 1,000 nt 내지 약 2,000 nt, 약 1,000 nt 내지 약 2,500 nt, 약 1,000 nt 내지 약 3,000 nt, 약 1,000 nt 내지 약 3,500 nt, 약 1,000 nt 내지 약 4,000 nt, 약 1,000 nt 내지 약 4,500 nt, 약 1,500 nt 내지 약 2,000 nt, 약 1,500 nt 내지 약 2,500 nt, 약 1,500 nt 내지 약 3,000 nt, 약 1,500 nt 내지 약 3,500 nt, 약 1,500 nt 내지 약 4,000 nt, 약 1,500 nt 내지 약 4,500 nt, 약 2,000 nt 내지 약 2,500 nt, 약 2,000 nt 내지 약 3,000 nt, 약 2,000 nt 내지 약 3,500 nt, 약 2,000 nt 내지 약 4,000 nt, 약 2,000 nt 내지 약 4,500 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,500 nt, 약 1,000 nt, 약 1,500 nt, 약 2,000 nt, 약 2,500 nt, 약 3,000 nt, 약 3,500 nt, 약 4,000 nt, 또는 약 4,500 nt로부터 독립적으로 선택된 크기를 갖는 것인 시스템.

[0347] 37. 구현예 31 내지 36 중 어느 하나에 있어서, 상기 시스템의 1개, 2개, 3개, 또는 4개의 합성 핵산 분자에 의해 코딩되는 1개, 2개, 3개, 또는 4개의 RNA 분자는 각각: 약 2500 내지 4500 nt, 약 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,500 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000 nt, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 및 약 4,500 nt로부터 독립적으로 선택된 크기를 갖는 것인 시스템.

[0348] 38. 구현예 31 내지 37 중 어느 하나에 있어서, 상기 시스템은 구현예 25 및 28 내지 30 중 어느 하나의 조성물을 포함하고,

[0349] 상기 제1 및 제2 합성 DNA 분자는 약 5000 nt 내지 약 10,000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 9,500 nt, 약 5,000 nt 내지 약 10,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 8,500 nt, 약 5,500 nt 내지 약 9,000 nt, 약 5,500 nt 내지 약 9,500 nt, 약 5,500 nt 내지 약 10,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 9,500 nt, 약 6,000 nt 내지 약 10,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 8,500 nt, 약 6,500 nt 내지 약 9,000 nt, 약 6,500 nt 내지 약 9,500 nt, 약 6,500 nt 내지 약 10,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 9,500 nt, 약 7,000 nt 내지 약 10,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 9,500 nt, 약 7,500 nt 내지 약 10,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 9,500 nt, 약 8,000 nt 내지 약 10,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,500 nt, 약 8,500 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 9,500 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,500 nt 내지 약 10,000 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 약 9,000 nt, 약 9,500 nt, 및 약 10,000 nt로부터 선택된 전체 크기를 갖고;

[0350] 전체 단백질 코딩 서열 크기는 약 2000 nt 내지 약 8000 nt, 약 2,000 nt 내지 약 3,000 nt, 약 2,000 nt 내지 약 3,500 nt, 약 2,000 nt 내지 약 4,000 nt, 약 2,000 nt 내지 약 4,500 nt, 약 2,000 nt 내지 약 5,000 nt, 약 2,000 nt 내지 약 5,500 nt, 약 2,000 nt 내지 약 6,000 nt, 약 2,000 nt 내지 약 6,500 nt, 약 2,000 nt 내지 약 7,000 nt, 약 2,000 nt 내지 약 7,500 nt, 약 2,000 nt 내지 약 8,000 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 5,500 nt, 약 3,000 nt 내지 약 6,000 nt, 약 3,000 nt 내지 약 6,500 nt, 약 3,000 nt 내지 약 7,000 nt, 약 3,000 nt 내지 약 7,500 nt, 약 3,000 nt 내지 약 8,000 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 5,000 nt, 약 3,500 nt 내지 약 5,500 nt, 약 3,500 nt 내지 약 6,000 nt, 약 3,500 nt 내지 약 6,500 nt, 약 3,500 nt 내지 약 7,000 nt, 약 3,500 nt 내지 약 7,500 nt, 약 3,500 nt 내지 약 8,000 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 5,500 nt, 약 4,000 nt 내지 약 6,000 nt, 약 4,000 nt 내지 약 6,500 nt, 약 4,000 nt 내지 약 7,000 nt, 약 4,000 nt 내지 약 7,500 nt, 약 4,000 nt 내지 약 8,000 nt, 약 4,500 nt 내지 약 5,000 nt, 약 4,500 nt 내지 약 5,500 nt, 약 4,500 nt 내지 약 6,000 nt, 약 4,500 nt 내지 약 6,500 nt, 약 4,500 nt 내지 약 7,000 nt, 약 4,500 nt 내지 약 7,500 nt, 약 4,500 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 또는 약 7,500 nt 내지 약 8,000 nt로부터 선택되고, 상기 전체 표적 단백질 코딩 서열은 약 2,000 nt, 약 3,000 nt, 약 3,500 nt, 약 4,000 nt, 약 4,500 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 및 약 8,000 nt이고; 및/또는

[0351] 상기 2개의 합성 DNA 분자에 의해 코딩되는 RNA 분자의 합한 크기는 약 5,000 nt 내지 약 9000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 8,500 nt, 약 5,500 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 8,500 nt, 약 6,500 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 또는 약 9,000 nt인 것인 시스템.

[0352] 39. 구현예 31 내지 36 중 어느 하나에 있어서, 상기 시스템은 구현예 26 및 28 내지 30 중 어느 하나의 조성물을 포함하고,

[0353] 상기 제1, 제2, 및 제3 합성 DNA 분자는 약 7500 nt 내지 약 15,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,500 nt, 약 7,500 nt 내지 약 10,000 nt, 약 7,500 nt 내지 약 10,500 nt, 약 7,500 nt 내지 약 11,000 nt, 약 7,500 nt 내지 약 11,500 nt, 약 7,500 nt 내지 약 12,000 nt, 약 7,500 nt 내지 약 12,500 nt, 약 7,500 nt 내지 약 13,000 nt, 약 7,500 nt 내지 약 14,000 nt, 약 7,500 nt 내지 약 15,000 nt, 약 8,500 nt 내지 약 9,500 nt, 약 8,500 nt 내지 약 10,000 nt, 약 8,500 nt 내지 약 10,500 nt, 약 8,500 nt 내지 약 11,000 nt, 약 8,500 nt 내지 약 11,500 nt, 약 8,500 nt 내지 약 12,000 nt, 약 8,500 nt 내지 약 12,500 nt, 약 8,500 nt 내지 약 13,000 nt, 약 8,500 nt 내지 약 14,000 nt, 약 8,500 nt 내지 약 15,000 nt, 약 9,500 nt 내지 약 10,000 nt, 약 9,500 nt 내지 약 10,500 nt, 약 9,500 nt 내지 약 11,000 nt, 약 9,500 nt 내지 약 11,500 nt, 약 9,500 nt 내지 약 12,000 nt, 약 9,500 nt 내지 약 12,500 nt, 약 9,500 nt 내지 약 13,000 nt, 약 9,500 nt 내지 약 14,000 nt, 약 9,500 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 10,500 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 11,500 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 12,500 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000

nt 내지 약 15,000 nt, 약 10,500 nt 내지 약 11,000 nt, 약 10,500 nt 내지 약 11,500 nt, 약 10,500 nt 내지 약 12,000 nt, 약 10,500 nt 내지 약 12,500 nt, 약 10,500 nt 내지 약 13,000 nt, 약 10,500 nt 내지 약 14,000 nt, 약 10,500 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 11,500 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 12,500 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,500 nt 내지 약 12,000 nt, 약 11,500 nt 내지 약 12,500 nt, 약 11,500 nt 내지 약 13,000 nt, 약 11,500 nt 내지 약 14,000 nt, 약 11,500 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 12,500 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,500 nt 내지 약 13,000 nt, 약 12,500 nt 내지 약 14,000 nt, 약 12,500 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 또는 약 14,000 nt 내지 약 15,000 nt, 약 7,500 nt, 약 8,500 nt, 약 9,500 nt, 약 10,000 nt, 약 10,500 nt, 약 11,000 nt, 약 11,500 nt, 약 12,000 nt, 약 12,500 nt, 약 13,000 nt, 약 14,000 nt, 또는 약 15,000 nt의 전체 크기를 갖고;

[0354] 전체 표적 단백질 코딩 서열은 약 3000 nt 내지 약 12,000 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 6,000 nt, 약 3,000 nt 내지 약 7,000 nt, 약 3,000 nt 내지 약 7,500 nt, 약 3,000 nt 내지 약 8,000 nt, 약 3,000 nt 내지 약 8,500 nt, 약 3,000 nt 내지 약 9,000 nt, 약 3,000 nt 내지 약 1,000 nt, 약 3,000 nt 내지 약 11,000 nt, 약 3,000 nt 내지 약 12,000 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 6,000 nt, 약 4,000 nt 내지 약 7,000 nt, 약 4,000 nt 내지 약 7,500 nt, 약 4,000 nt 내지 약 8,000 nt, 약 4,000 nt 내지 약 8,500 nt, 약 4,000 nt 내지 약 9,000 nt, 약 4,000 nt 내지 약 1,000 nt, 약 4,000 nt 내지 약 11,000 nt, 약 4,000 nt 내지 약 12,000 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 1,000 nt, 약 5,000 nt 내지 약 11,000 nt, 약 5,000 nt 내지 약 12,000 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 1,000 nt, 약 6,000 nt 내지 약 11,000 nt, 약 6,000 nt 내지 약 12,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 1,000 nt, 약 7,000 nt 내지 약 11,000 nt, 약 7,000 nt 내지 약 12,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 1,000 nt, 약 7,500 nt 내지 약 11,000 nt, 약 7,500 nt 내지 약 12,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 1,000 nt, 약 8,000 nt 내지 약 11,000 nt, 약 8,000 nt 내지 약 12,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 1,000 nt, 약 8,500 nt 내지 약 11,000 nt, 약 8,500 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 1,000 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 12,000 nt, 약 1,000 nt 내지 약 11,000 nt, 약 1,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 3,000 nt, 약 4,000 nt, 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 약 9,000 nt, 약 1,000 nt, 약 11,000 nt, 또는 약 12,000 nt이고; 및/또는

[0355] 상기 3개의 합성 DNA 분자에 의해 코딩되는 RNA 분자의 합한 크기는 약 7500 nt 내지 약 13,500 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 9,500 nt, 약 7,500 nt 내지 약 10,000 nt, 약 7,500 nt 내지 약 10,500 nt, 약 7,500 nt 내지 약 11,000 nt, 약 7,500 nt 내지 약 11,500 nt, 약 7,500 nt 내지 약 12,000 nt, 약 7,500 nt 내지 약 12,500 nt, 약 7,500 nt 내지 약 13,000 nt, 약 7,500 nt 내지 약 13,500 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,500 nt, 약 8,500 nt 내지 약 10,000 nt, 약 8,500 nt 내지 약 10,500 nt, 약 8,500 nt 내지 약 11,000 nt, 약 8,500 nt 내지 약 11,500 nt, 약 8,500 nt 내지 약 12,000 nt, 약 8,500 nt 내지 약 12,500 nt, 약 8,500 nt 내지 약 13,000 nt, 약 8,500 nt 내지 약 13,500 nt, 약 9,000 nt 내지 약 9,500 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 10,500 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 11,500 nt, 약 9,000 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 12,500 nt, 약 9,000 nt 내지 약 13,000 nt, 약 9,000 nt 내지 약 13,500 nt, 약 9,500 nt 내지 약 10,000 nt, 약 9,500 nt 내지 약 10,500 nt, 약 9,500 nt 내지 약 11,000 nt, 약 9,500 nt 내지 약 11,500 nt, 약 9,500 nt 내지 약 12,000 nt, 약 9,500 nt 내지 약 12,500 nt, 약 9,500 nt 내지 약 13,000 nt, 약 9,500 nt 내지 약 13,500 nt, 약 10,000 nt 내지 약 10,500 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 11,500 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 12,500 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 13,500 nt, 약 10,500 nt 내지 약 11,000 nt, 약 10,500 nt 내지 약 11,500 nt, 약 10,500 nt 내지 약 12,000 nt, 약 10,500 nt 내지 약 12,500 nt, 약 10,500 nt 내지 약 13,000 nt, 약 10,500 nt 내지 약 13,500 nt, 약 11,000 nt 내지 약 11,500 nt, 약 11,000 nt 내지 약 12,000

nt, 약 11,000 nt 내지 약 12,500 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 13,500 nt, 약 11,500 nt 내지 약 12,000 nt, 약 11,500 nt 내지 약 12,500 nt, 약 11,500 nt 내지 약 13,000 nt, 약 11,500 nt 내지 약 13,500 nt, 약 12,000 nt 내지 약 12,500 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 13,500 nt, 약 12,500 nt 내지 약 13,000 nt, 약 12,500 nt 내지 약 13,500 nt, 약 13,000 nt 내지 약 13,500 nt 약 7,500 nt, 약 8,500 nt, 약 9,000 nt, 약 9,500 nt, 약 10,000 nt, 약 10,500 nt, 약 11,000 nt, 약 11,500 nt, 약 12,000 nt, 약 12,500 nt, 약 13,000 nt, 또는 약 13,500 nt인 것인 시스템.

[0356] 40. 구현예 31 내지 36 중 어느 하나에 있어서, 상기 시스템은 구현예 27 및 28 내지 30 중 어느 하나의 조성물을 포함하고,

[0357] 상기 제1, 제2, 제3, 및 제4 합성 DNA 분자는 약 10,000 nt 내지 약 20,000 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 16,000 nt, 약 10,000 nt 내지 약 17,000 nt, 약 10,000 nt 내지 약 18,000 nt, 약 10,000 nt 내지 약 19,000 nt, 약 10,000 nt 내지 약 20,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 16,000 nt, 약 11,000 nt 내지 약 17,000 nt, 약 11,000 nt 내지 약 18,000 nt, 약 11,000 nt 내지 약 19,000 nt, 약 11,000 nt 내지 약 20,000 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 16,000 nt, 약 12,000 nt 내지 약 17,000 nt, 약 12,000 nt 내지 약 18,000 nt, 약 12,000 nt 내지 약 19,000 nt, 약 12,000 nt 내지 약 20,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 16,000 nt, 약 13,000 nt 내지 약 17,000 nt, 약 13,000 nt 내지 약 18,000 nt, 약 13,000 nt 내지 약 19,000 nt, 약 13,000 nt 내지 약 20,000 nt, 약 14,000 nt 내지 약 15,000 nt, 약 14,000 nt 내지 약 16,000 nt, 약 14,000 nt 내지 약 17,000 nt, 약 14,000 nt 내지 약 18,000 nt, 약 14,000 nt 내지 약 19,000 nt, 약 14,000 nt 내지 약 20,000 nt, 약 15,000 nt 내지 약 16,000 nt, 약 15,000 nt 내지 약 17,000 nt, 약 15,000 nt 내지 약 18,000 nt, 약 15,000 nt 내지 약 19,000 nt, 약 15,000 nt 내지 약 20,000 nt, 약 16,000 nt 내지 약 17,000 nt, 약 16,000 nt 내지 약 18,000 nt, 약 16,000 nt 내지 약 19,000 nt, 약 16,000 nt 내지 약 20,000 nt, 약 17,000 nt 내지 약 18,000 nt, 약 17,000 nt 내지 약 19,000 nt, 약 17,000 nt 내지 약 20,000 nt, 약 18,000 nt 내지 약 19,000 nt, 약 18,000 nt 내지 약 20,000 nt, 약 19,000 nt 내지 약 20,000 nt 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 약 15,000 nt, 약 16,000 nt, 약 17,000 nt, 약 18,000 nt, 약 19,000 nt, 또는 약 20,000 nt의 전체 크기를 갖고;

[0358] 전체 표적 단백질 코딩 서열은 약 4000 nt 내지 약 16,000 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 10,000 nt, 약 5,000 nt 내지 약 11,000 nt, 약 5,000 nt 내지 약 12,000 nt, 약 5,000 nt 내지 약 13,000 nt, 약 5,000 nt 내지 약 14,000 nt, 약 5,000 nt 내지 약 15,000 nt, 약 5,000 nt 내지 약 16,000 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 10,000 nt, 약 6,000 nt 내지 약 11,000 nt, 약 6,000 nt 내지 약 12,000 nt, 약 6,000 nt 내지 약 13,000 nt, 약 6,000 nt 내지 약 14,000 nt, 약 6,000 nt 내지 약 15,000 nt, 약 6,000 nt 내지 약 16,000 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 10,000 nt, 약 7,000 nt 내지 약 11,000 nt, 약 7,000 nt 내지 약 12,000 nt, 약 7,000 nt 내지 약 13,000 nt, 약 7,000 nt 내지 약 14,000 nt, 약 7,000 nt 내지 약 15,000 nt, 약 7,000 nt 내지 약 16,000 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 10,000 nt, 약 8,000 nt 내지 약 11,000 nt, 약 8,000 nt 내지 약 12,000 nt, 약 8,000 nt 내지 약 13,000 nt, 약 8,000 nt 내지 약 14,000 nt, 약 8,000 nt 내지 약 15,000 nt, 약 8,000 nt 내지 약 16,000 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 13,000 nt, 약 9,000 nt 내지 약 14,000 nt, 약 9,000 nt 내지 약 15,000 nt, 약 9,000 nt 내지 약 16,000 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 16,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 16,000 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 16,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 16,000 nt, 약 14,000 nt 내지 약 15,000 nt, 약 14,000 nt 내지 약 16,000 nt, or 약 15,000 nt 내지 약 16,000 nt. the total target protein coding sequence is 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 8,000 nt, 약 9,000 nt, 약 10,000 nt, 약 11,000

nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 약 15,000 nt, 약 16,000 nt 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 8,000 nt, 약 9,000 nt, 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 또는 약 15,000 nt이고; 및/또는

[0359] 상기 4개의 합성 DNA 분자에 의해 코딩되는 RNA 분자의 합한 크기는 약 10,000 nt 내지 약 18,000 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 16,000 nt, 약 10,000 nt 내지 약 17,000 nt, 약 10,000 nt 내지 약 18,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 16,000 nt, 약 11,000 nt 내지 약 17,000 nt, 약 11,000 nt 내지 약 18,000 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 16,000 nt, 약 12,000 nt 내지 약 17,000 nt, 약 12,000 nt 내지 약 18,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 16,000 nt, 약 13,000 nt 내지 약 17,000 nt, 약 13,000 nt 내지 약 18,000 nt, 약 14,000 nt 내지 약 15,000 nt, 약 14,000 nt 내지 약 16,000 nt, 약 14,000 nt 내지 약 17,000 nt, 약 14,000 nt 내지 약 18,000 nt, 약 15,000 nt 내지 약 16,000 nt, 약 15,000 nt 내지 약 17,000 nt, 약 15,000 nt 내지 약 18,000 nt, 약 16,000 nt 내지 약 17,000 nt, 약 16,000 nt 내지 약 18,000 nt, 약 17,000 nt 내지 약 18,000 nt, 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 약 15,000 nt, 약 16,000 nt, 약 17,000 nt, 또는 약 18,000 nt인 것인 시스템.

[0360] 41. 구현예 31 내지 40 중 어느 하나에 있어서, 상기 제1 이합체화 도메인과 제2 이합체화 도메인, 상기 제3 이합체화 도메인과 제4 이합체화 도메인, 및/또는 상기 제5 이합체화 도메인과 제6 이합체화 도메인은 각각 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 상기 시스템은 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 55%, 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 적어도 약 95%, 또는 약 100%의 재조합 효율을 갖는 것인 시스템.

[0361] 42. 구현예 31 내지 41 중 어느 하나에 있어서, 각각의 이합체화 도메인은 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 상기 시스템은 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 90%, 또는 약 100%의 재조합 효율을 갖는 것인 시스템.

[0362] 43. 구현예 31 내지 42 중 어느 하나에 있어서, 상기 RNA 재조합 효율은 약 10% 내지 약 100%, 약 10% 내지 약 20%, 약 10% 내지 약 30%, 약 10% 내지 약 35%, 약 10% 내지 약 40%, 약 10% 내지 약 45%, 약 10% 내지 약 50%, 약 10% 내지 약 55%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 20% 내지 약 30%, 약 20% 내지 약 35%, 약 20% 내지 약 40%, 약 20% 내지 약 45%, 약 20% 내지 약 50%, 약 20% 내지 약 55%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 30% 내지 약 35%, 약 30% 내지 약 40%, 약 30% 내지 약 45%, 약 30% 내지 약 50%, 약 30% 내지 약 55%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 35% 내지 약 40%, 약 35% 내지 약 45%, 약 35% 내지 약 50%, 약 35% 내지 약 55%, 약 35% 내지 약 60%, 약 35% 내지 약 70%, 약 35% 내지 약 80%, 약 35% 내지 약 90%, 약 40% 내지 약 45%, 약 40% 내지 약 50%, 약 40% 내지 약 55%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 45% 내지 약 50%, 약 45% 내지 약 55%, 약 45% 내지 약 60%, 약 45% 내지 약 70%, 약 45% 내지 약 80%, 약 45% 내지 약 90%, 약 50% 내지 약 55%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 55% 내지 약 60%, 약 55% 내지 약 70%, 약 55% 내지 약 80%, 약 55% 내지 약 90%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 80% 내지 약 90%, 약 10%, 약 20%, 약 30%, 약 35%, 약 40%, 약 45%, 약 50%, 약 55%, 약 60%, 약 70%, 약 80%, 약 90%, 약 95%, 또는 약 100%인 것인 시스템.

[0363] 44. 구현예 31 내지 43 중 어느 하나의 시스템을 포함하는 조성물.

[0364] 45. 구현예 44에 있어서, 상기 조성물은 제1 RNA 분자, 제2 RNA 분자, 제3 RNA 분자, 및 선택적으로 제4 RNA 분

자를 포함하고, 각각은 디스트로핀, 인자 8, ABCA4, 또는 MYO7A의 적어도 일부를 코딩하는 것인 조성물.

[0365] 46. 구현예 31 내지 43 중 어느 하나의 시스템, 또는 구현예 44 및 45 중 어느 하나의 조성물을 포함하는 키트로서, 상기 제1, 제2, 제3, 및 제4 핵산 분자는 별개의 용기에 담길 수 있고, 선택적으로 약학적으로 허용가능한 담체와 같은 완충액을 더 포함하는 것인 키트.

[0366] 47. 세포에서 표적 단백질을 발현시키는 방법으로서,

[0367] 구현예 31 내지 43 중 어느 하나의 시스템, 또는 구현예 44 및 45 중 어느 하나의 조성물을 세포 내로 도입시키는 단계, 및 상기 세포에서 제1 및 제2 합성 RNA 분자, 제1, 제2, 및 제3 합성 RNA 분자, 또는 제1, 제2, 제3, 및 제4 RNA 분자를 발현시키는 단계를 포함하고, 상기 표적 단백질이 상기 세포에서 생산되는 것인 방법.

[0368] 48. 구현예 47에 있어서, 상기 세포는 개체 내에 존재하고, 상기 도입시키는 단계는 상기 시스템의 치료 유효량을 상기 개체에게 투여하는 단계를 포함하는 것인 방법.

[0369] 49. 구현예 48에 있어서, 상기 방법은 상기 개체에서 상기 표적 단백질을 코딩하는 유전자 중 돌연변이에 의해 유발된 유전 질환을 치료하고, 상기 방법은 상기 개체에서 기능성 표적 단백질의 발현을 가져오는 것인 방법.

[0370] 50. 구현예 49에 있어서,

[0371] 상기 유전 질환은 듀센 근이영양증이고, 상기 표적 단백질은 디스트로핀이거나;

[0372] 상기 유전 질환은 A형 혈우병이고, 상기 표적 단백질은 F8이거나;

[0373] 상기 유전 질환은 스타르가르트병이고, 상기 표적 단백질은 ABCA4이거나; 또는

[0374] 상기 유전 질환은 어셔 증후군이고, 상기 표적 단백질은 MYO7A인 것인 방법.

[0375] 51. 구현예 31 내지 43 중 어느 하나, 구현예 1 내지 24, 44 및 45 중 어느 하나, 구현예 46, 또는 구현예 47 내지 50 중 어느 하나에 있어서, 1개, 2개, 3개, 또는 4개의 RNA 분자는 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166 중 어느 하나로 제공되는 합성 인트론에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100% 서열 동일성을 포함하는 것인 시스템, 조성물, 키트, 또는 방법.

[0376] 52. 구현예 31 내지 43 중 어느 하나, 구현예 1 내지 24, 44, 및 45 중 어느 하나, 구현예 46, 또는 구현예 47 내지 50 중 어느 하나에 있어서, 1개, 2개, 3개, 또는 4개의 RNA 분자는 서열번호 20의 nt 3703 내지 3975, 서열번호 21의 nt 1 내지 228, 서열번호 22의 nt 3703 내지 3975, 서열번호 23의 nt 1 내지 225, 서열번호 24의 nt 3560 내지 3828, 및 서열번호 25의 nt 1-225로부터 선택되는 합성 인트론을 포함하는 것인 시스템, 조성물, 키트, 또는 방법.

[0377] 53. 구현예 31 내지 43, 51, 및 52 중 어느 하나, 구현예 1 내지 24, 44, 및 45 중 어느 하나, 또는 구현예 47 내지 50 중 어느 하나에 있어서, 상기 1개, 2개, 3개, 또는 4개의 RNA 분자는 단백질 코딩 서열의 부분을 더 포함하는 것인 시스템, 조성물, 키트, 또는 방법.

[0378] 54. 구현예 31 내지 43 중 어느 하나, 구현예 1 내지 24, 44, 및 45 중 어느 하나, 또는 구현예 47 내지 50 중 어느 하나에 있어서, 상기 단백질 코딩 서열의 부분은 상기 단백질 코딩 서열의 N-말단 절반, N-말단 1/3, 중간 부분, 제1 중간 부분, C-말단 절반, 또는 C-말단 1/3을 포함하는 것인 시스템, 조성물, 또는 방법.

[0379] 55. 구현예 31 내지 43 및 51 내지 54 중 어느 하나, 구현예 1 내지 24, 44, 및 45 중 어느 하나, 또는 구현예 47 내지 50 중 어느 하나에 있어서, (a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 표적 단백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너; (ii-2) DISE, ISE, 또는 둘 모두; 및 (iii) 제1 이합체화 도메인;을 포함하는 것인 제1 RNA 분자; 및 (b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; (i-2) 하나 이상의 ISE 서열; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제2 RNA 분자를 포함하는 것인 시스템, 조성물, 또는 방법.

[0380] 56. 구현예 31 내지 43 및 51 내지 55 중 어느 하나, 구현예 1 내지 24, 44, 및 45 중 어느 하나, 또는 구현예 47 내지 50 중 어느 하나에 있어서, (a) 제1 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 상기 표적 단

백질의 N-말단 부분에 대한 코딩 서열; (ii) 스플라이스 도너; (ii-2) DISE, ISE, 및 ISE; 및 (iii) 제 1 이합체화 도메인;을 포함하는 것인 제1 RNA 분자; 및 (b) 제2 RNA 분자로서, 상기 RNA 분자는 5'에서 3'으로, (i) 제2 이합체화 도메인으로서, 상기 제2 이합체화 도메인은 상기 제1 이합체화 도메인에 결합하는 것인 제2 이합체화 도메인; (i-2) 3개의 ISE 서열; (ii) 분지점 서열; (iii) 폴리피리미딘 트랙트; (iv) 스플라이스 억셉터; 및 (v) 상기 표적 단백질의 C-말단 부분에 대한 코딩 서열을 포함하는 것인 제2 RNA 분자를 포함하는 것인 시스템, 조성물, 또는 방법.

[0381] 57. 구현예 1 및 4 내지 24 중 어느 하나, 또는 구현예 2 및 4 내지 24 중 어느 하나, 또는 구현예 3 및 4 내지 24 중 어느 하나에 있어서, 상기 2개의 RNA 분자 중 1개 또는 2개, 또는 상기 3개의 RNA 분자 중 1개, 2개, 또는 3개, 또는 상기 4개의 RNA 분자 중 1개, 2개, 3개, 또는 4개는 각각: 약 2500 내지 4500 nt, 약 2,500 nt 내지 약 2,750 nt, 약 2,500 nt 내지 약 3,000 nt, 약 2,500 nt 내지 약 3,250 nt, 약 2,500 nt 내지 약 3,500 nt, 약 2,500 nt 내지 약 3,750 nt, 약 2,500 nt 내지 약 4,000 nt, 약 2,500 nt 내지 약 4,250 nt, 약 2,500 nt 내지 약 4,500 nt, 약 2,750 nt 내지 약 3,000 nt, 약 2,750 nt 내지 약 3,250 nt, 약 2,750 nt 내지 약 3,500 nt, 약 2,750 nt 내지 약 3,750 nt, 약 2,750 nt 내지 약 4,000 nt, 약 2,750 nt 내지 약 4,250 nt, 약 2,750 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 3,250 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 3,750 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,250 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,250 nt 내지 약 3,500 nt, 약 3,250 nt 내지 약 3,750 nt, 약 3,250 nt 내지 약 4,000 nt, 약 3,250 nt 내지 약 4,250 nt, 약 3,250 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 3,750 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,250 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,750 nt 내지 약 4,000 nt, 약 3,750 nt 내지 약 4,250 nt, 약 3,750 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 4,250 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,250 nt 내지 약 4,500 nt, 약 2,500 nt, 약 2,750 nt, 약 3,000 nt, 약 3,250 nt, 약 3,500 nt, 약 3,750 nt, 약 4,000 nt, 약 4,250 nt, 및 약 4,500 nt로부터 독립적으로 선택된 크기를 갖는 것인 조성물.

[0382] 58. 구현예 1 및 4 내지 24 중 어느 하나에 있어서,

[0383] 전체 표적 단백질 코딩 서열은 약 2000 nt 내지 약 8000 nt, 약 2,000 nt 내지 약 3,000 nt, 약 2,000 nt 내지 약 3,500 nt, 약 2,000 nt 내지 약 4,000 nt, 약 2,000 nt 내지 약 4,500 nt, 약 2,000 nt 내지 약 5,000 nt, 약 2,000 nt 내지 약 5,500 nt, 약 2,000 nt 내지 약 6,000 nt, 약 2,000 nt 내지 약 6,500 nt, 약 2,000 nt 내지 약 7,000 nt, 약 2,000 nt 내지 약 7,500 nt, 약 2,000 nt 내지 약 8,000 nt, 약 3,000 nt 내지 약 3,500 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 4,500 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 5,500 nt, 약 3,000 nt 내지 약 6,000 nt, 약 3,000 nt 내지 약 6,500 nt, 약 3,000 nt 내지 약 7,000 nt, 약 3,000 nt 내지 약 7,500 nt, 약 3,000 nt 내지 약 8,000 nt, 약 3,500 nt 내지 약 4,000 nt, 약 3,500 nt 내지 약 4,500 nt, 약 3,500 nt 내지 약 5,000 nt, 약 3,500 nt 내지 약 5,500 nt, 약 3,500 nt 내지 약 6,000 nt, 약 3,500 nt 내지 약 6,500 nt, 약 3,500 nt 내지 약 7,000 nt, 약 3,500 nt 내지 약 7,500 nt, 약 3,500 nt 내지 약 8,000 nt, 약 4,000 nt 내지 약 4,500 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 5,500 nt, 약 4,000 nt 내지 약 6,000 nt, 약 4,000 nt 내지 약 6,500 nt, 약 4,000 nt 내지 약 7,000 nt, 약 4,000 nt 내지 약 7,500 nt, 약 4,000 nt 내지 약 8,000 nt, 약 4,500 nt 내지 약 5,000 nt, 약 4,500 nt 내지 약 5,500 nt, 약 4,500 nt 내지 약 6,000 nt, 약 4,500 nt 내지 약 6,500 nt, 약 4,500 nt 내지 약 7,000 nt, 약 4,500 nt 내지 약 7,500 nt, 약 4,500 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지 약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 또는 약 7,500 nt 내지 약 8,000 nt이거나, 전체 표적 단백질 코딩 서열은 약 2,000 nt, 약 3,000 nt, 약 3,500 nt, 약 4,000 nt, 약 4,500 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 또는 약 8,000 nt이고; 및/또는

[0384] 상기 2개의 RNA 분자의 합한 크기는 약 5,000 nt 내지 약 9000 nt, 약 5,000 nt 내지 약 5,500 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 6,500 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,500 nt 내지 약 6,000 nt, 약 5,500 nt 내지 약 6,500 nt, 약 5,500 nt 내지 약 7,000 nt, 약 5,500 nt 내지

약 7,500 nt, 약 5,500 nt 내지 약 8,000 nt, 약 5,500 nt 내지 약 8,500 nt, 약 5,500 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 6,500 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,500 nt 내지 약 7,000 nt, 약 6,500 nt 내지 약 7,500 nt, 약 6,500 nt 내지 약 8,000 nt, 약 6,500 nt 내지 약 8,500 nt, 약 6,500 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 5,000 nt, 약 5,500 nt, 약 6,000 nt, 약 6,500 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 또는 약 9,000 nt인 것인 조성물.

[0385] 59. 구현예 2 및 4 내지 24 중 어느 하나에 있어서,

[0386] 전체 표적 단백질 코딩 서열 크기는 약 3000 nt 내지 약 12,000 nt, 약 3,000 nt 내지 약 4,000 nt, 약 3,000 nt 내지 약 5,000 nt, 약 3,000 nt 내지 약 6,000 nt, 약 3,000 nt 내지 약 7,000 nt, 약 3,000 nt 내지 약 7,500 nt, 약 3,000 nt 내지 약 8,000 nt, 약 3,000 nt 내지 약 8,500 nt, 약 3,000 nt 내지 약 9,000 nt, 약 3,000 nt 내지 약 1,000 nt, 약 3,000 nt 내지 약 11,000 nt, 약 3,000 nt 내지 약 12,000 nt, 약 4,000 nt 내지 약 5,000 nt, 약 4,000 nt 내지 약 6,000 nt, 약 4,000 nt 내지 약 7,000 nt, 약 4,000 nt 내지 약 7,500 nt, 약 4,000 nt 내지 약 8,000 nt, 약 4,000 nt 내지 약 8,500 nt, 약 4,000 nt 내지 약 9,000 nt, 약 4,000 nt 내지 약 1,000 nt, 약 4,000 nt 내지 약 11,000 nt, 약 4,000 nt 내지 약 12,000 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 7,500 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 8,500 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 1,000 nt, 약 5,000 nt 내지 약 11,000 nt, 약 5,000 nt 내지 약 12,000 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 7,500 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 8,500 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 1,000 nt, 약 6,000 nt 내지 약 11,000 nt, 약 6,000 nt 내지 약 12,000 nt, 약 7,000 nt 내지 약 7,500 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 8,500 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 1,000 nt, 약 7,000 nt 내지 약 11,000 nt, 약 7,000 nt 내지 약 12,000 nt, 약 7,500 nt 내지 약 8,000 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 1,000 nt, 약 7,500 nt 내지 약 11,000 nt, 약 7,500 nt 내지 약 12,000 nt, 약 8,000 nt 내지 약 8,500 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 1,000 nt, 약 8,000 nt 내지 약 11,000 nt, 약 8,000 nt 내지 약 12,000 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 1,000 nt, 약 8,500 nt 내지 약 11,000 nt, 약 8,500 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 1,000 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 12,000 nt, 약 1,000 nt 내지 약 11,000 nt, 약 1,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 3,000 nt, 약 4,000 nt, 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 7,500 nt, 약 8,000 nt, 약 8,500 nt, 약 9,000 nt, 약 1,000 nt, 약 11,000 nt, 또는 약 12,000 nt이고; 및/또는

[0387] 상기 3개의 RNA 분자의 합한 크기는 약 7500 nt 내지 약 13,500 nt, 약 7,500 nt 내지 약 8,500 nt, 약 7,500 nt 내지 약 9,000 nt, 약 7,500 nt 내지 약 9,500 nt, 약 7,500 nt 내지 약 10,000 nt, 약 7,500 nt 내지 약 10,500 nt, 약 7,500 nt 내지 약 11,000 nt, 약 7,500 nt 내지 약 11,500 nt, 약 7,500 nt 내지 약 12,000 nt, 약 7,500 nt 내지 약 12,500 nt, 약 7,500 nt 내지 약 13,000 nt, 약 7,500 nt 내지 약 13,500 nt, 약 8,500 nt 내지 약 9,000 nt, 약 8,500 nt 내지 약 9,500 nt, 약 8,500 nt 내지 약 10,000 nt, 약 8,500 nt 내지 약 10,500 nt, 약 8,500 nt 내지 약 11,000 nt, 약 8,500 nt 내지 약 11,500 nt, 약 8,500 nt 내지 약 12,000 nt, 약 8,500 nt 내지 약 12,500 nt, 약 8,500 nt 내지 약 13,000 nt, 약 8,500 nt 내지 약 13,500 nt, 약 9,000 nt 내지 약 9,500 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 10,500 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 11,500 nt, 약 9,000 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 12,500 nt, 약 9,000 nt 내지 약 13,000 nt, 약 9,000 nt 내지 약 13,500 nt, 약 9,500 nt 내지 약 10,000 nt, 약 9,500 nt 내지 약 10,500 nt, 약 9,500 nt 내지 약 11,000 nt, 약 9,500 nt 내지 약 11,500 nt, 약 9,500 nt 내지 약 12,000 nt, 약 9,500 nt 내지 약 12,500 nt, 약 9,500 nt 내지 약 13,000 nt, 약 9,500 nt 내지 약 13,500 nt, 약 10,000 nt 내지 약 10,500 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 11,500 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 12,500 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 13,500 nt, 약 10,500 nt 내지 약 11,000 nt, 약 10,500 nt 내지 약 11,500 nt, 약 10,500 nt 내지 약 12,000 nt, 약 10,500 nt 내지 약 12,500 nt, 약 10,500 nt 내지 약 13,000 nt, 약 10,500 nt 내지 약 13,500 nt, 약 11,000 nt 내지 약 11,500 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 12,500 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 13,500 nt, 약 11,500 nt 내지 약 12,000 nt, 약

11,500 nt 내지 약 12,500 nt, 약 11,500 nt 내지 약 13,000 nt, 약 11,500 nt 내지 약 13,500 nt, 약 12,000 nt 내지 약 12,500 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 13,500 nt, 약 12,500 nt 내지 약 13,000 nt, 약 12,500 nt 내지 약 13,500 nt, 약 13,000 nt 내지 약 13,500 nt 약 7,500 nt, 약 8,500 nt, 약 9,000 nt, 약 9,500 nt, 약 10,000 nt, 약 10,500 nt, 약 11,000 nt, 약 11,500 nt, 약 12,000 nt, 약 12,500 nt, 약 13,000 nt, 또는 약 13,500 nt인 것인 조성물.

[0388] 60. 구현예 3 및 4 내지 24 중 어느 하나에 있어서,

[0389] 전체 표적 단백질 코딩 서열 크기는 약 4000 nt 내지 약 16,000 nt, 약 5,000 nt 내지 약 6,000 nt, 약 5,000 nt 내지 약 7,000 nt, 약 5,000 nt 내지 약 8,000 nt, 약 5,000 nt 내지 약 9,000 nt, 약 5,000 nt 내지 약 10,000 nt, 약 5,000 nt 내지 약 11,000 nt, 약 5,000 nt 내지 약 12,000 nt, 약 5,000 nt 내지 약 13,000 nt, 약 5,000 nt 내지 약 14,000 nt, 약 5,000 nt 내지 약 15,000 nt, 약 5,000 nt 내지 약 16,000 nt, 약 6,000 nt 내지 약 7,000 nt, 약 6,000 nt 내지 약 8,000 nt, 약 6,000 nt 내지 약 9,000 nt, 약 6,000 nt 내지 약 10,000 nt, 약 6,000 nt 내지 약 11,000 nt, 약 6,000 nt 내지 약 12,000 nt, 약 6,000 nt 내지 약 13,000 nt, 약 6,000 nt 내지 약 14,000 nt, 약 6,000 nt 내지 약 15,000 nt, 약 6,000 nt 내지 약 16,000 nt, 약 7,000 nt 내지 약 8,000 nt, 약 7,000 nt 내지 약 9,000 nt, 약 7,000 nt 내지 약 10,000 nt, 약 7,000 nt 내지 약 11,000 nt, 약 7,000 nt 내지 약 12,000 nt, 약 7,000 nt 내지 약 13,000 nt, 약 7,000 nt 내지 약 14,000 nt, 약 7,000 nt 내지 약 15,000 nt, 약 7,000 nt 내지 약 16,000 nt, 약 8,000 nt 내지 약 9,000 nt, 약 8,000 nt 내지 약 10,000 nt, 약 8,000 nt 내지 약 11,000 nt, 약 8,000 nt 내지 약 12,000 nt, 약 8,000 nt 내지 약 13,000 nt, 약 8,000 nt 내지 약 14,000 nt, 약 8,000 nt 내지 약 15,000 nt, 약 8,000 nt 내지 약 16,000 nt, 약 9,000 nt 내지 약 10,000 nt, 약 9,000 nt 내지 약 11,000 nt, 약 9,000 nt 내지 약 12,000 nt, 약 9,000 nt 내지 약 13,000 nt, 약 9,000 nt 내지 약 14,000 nt, 약 9,000 nt 내지 약 15,000 nt, 약 9,000 nt 내지 약 16,000 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 16,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 16,000 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 16,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 16,000 nt, 약 14,000 nt 내지 약 15,000 nt, 약 14,000 nt 내지 약 16,000 nt, 또는 약 15,000 nt 내지 약 16,000 nt이거나, 전체 표적 단백질 코딩 서열은 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 8,000 nt, 약 9,000 nt, 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 약 15,000 nt, 약 16,000 nt 약 5,000 nt, 약 6,000 nt, 약 7,000 nt, 약 8,000 nt, 약 9,000 nt, 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 또는 약 15,000 nt이고; 및/또는

[0390] 상기 4개의 합성 DNA 분자에 의해 코딩되는 RNA 분자의 합한 크기는 약 10,000 nt 내지 약 18,000 nt, 약 10,000 nt 내지 약 11,000 nt, 약 10,000 nt 내지 약 12,000 nt, 약 10,000 nt 내지 약 13,000 nt, 약 10,000 nt 내지 약 14,000 nt, 약 10,000 nt 내지 약 15,000 nt, 약 10,000 nt 내지 약 16,000 nt, 약 10,000 nt 내지 약 17,000 nt, 약 10,000 nt 내지 약 18,000 nt, 약 11,000 nt 내지 약 12,000 nt, 약 11,000 nt 내지 약 13,000 nt, 약 11,000 nt 내지 약 14,000 nt, 약 11,000 nt 내지 약 15,000 nt, 약 11,000 nt 내지 약 16,000 nt, 약 11,000 nt 내지 약 17,000 nt, 약 11,000 nt 내지 약 18,000 nt, 약 12,000 nt 내지 약 13,000 nt, 약 12,000 nt 내지 약 14,000 nt, 약 12,000 nt 내지 약 15,000 nt, 약 12,000 nt 내지 약 16,000 nt, 약 12,000 nt 내지 약 17,000 nt, 약 12,000 nt 내지 약 18,000 nt, 약 13,000 nt 내지 약 14,000 nt, 약 13,000 nt 내지 약 15,000 nt, 약 13,000 nt 내지 약 16,000 nt, 약 13,000 nt 내지 약 17,000 nt, 약 13,000 nt 내지 약 18,000 nt, 약 14,000 nt 내지 약 15,000 nt, 약 14,000 nt 내지 약 16,000 nt, 약 14,000 nt 내지 약 17,000 nt, 약 14,000 nt 내지 약 18,000 nt, 약 15,000 nt 내지 약 16,000 nt, 약 15,000 nt 내지 약 17,000 nt, 약 15,000 nt 내지 약 18,000 nt, 약 16,000 nt 내지 약 17,000 nt, 약 16,000 nt 내지 약 18,000 nt, 약 17,000 nt 내지 약 18,000 nt, 약 10,000 nt, 약 11,000 nt, 약 12,000 nt, 약 13,000 nt, 약 14,000 nt, 약 15,000 nt, 약 16,000 nt, 약 17,000 nt, 또는 약 18,000 nt인 것인 조성물.

[0391] 61. 구현예 1 내지 24 및 57 내지 60 중 어느 하나에 있어서, 상기 제1 이합체화 도메인과 제2 이합체화 도메인, 상기 제3 이합체화 도메인과 제4 이합체화 도메인, 및/또는 상기 제5 이합체화 도메인과 제6 이합체화 도메인은 각각 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150,

200, 250, 300, 400, 또는 500 nt이고; 상기 시스템은 적어도 약 20%, 적어도 약 25%, 적어도 약 30%, 적어도 약 35%, 적어도 약 40%, 적어도 약 45%, 적어도 약 50%, 적어도 약 55%, 적어도 약 60%, 적어도 약 65%, 적어도 약 70%, 적어도 약 75%, 적어도 약 80%, 적어도 약 85%, 적어도 약 90%, 또는 적어도 약 95%의 재조합 효율을 갖는 것인 시스템.

[0392] 62. 구체예 1 내지 24 및 57 내지 61 중 어느 하나에 있어서, 각각의 이합체화 도메인은 1000 nt 이하, 예를 들면, 적어도 50 nt, 적어도 100 nt, 적어도 150 nt, 적어도 200 nt, 적어도 300 nt, 적어도 400 nt, 적어도 500 nt, 50 내지 1000 nt, 50 내지 500 nt, 50 내지 150 nt, 50, 100, 150, 200, 250, 300, 400, 또는 500 nt이고; 상기 시스템은 적어도 20%, 적어도 30%, 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 75%, 적어도 80%, 또는 적어도 90%의 재조합 효율을 갖는 것인 시스템.

[0393] 63. 구체예 1 내지 24 및 57 내지 62 중 어느 하나에 있어서, RNA 재조합 효율은 약 10% 내지 약 100%, 약 10% 내지 약 20%, 약 10% 내지 약 30%, 약 10% 내지 약 35%, 약 10% 내지 약 40%, 약 10% 내지 약 45%, 약 10% 내지 약 50%, 약 10% 내지 약 55%, 약 10% 내지 약 60%, 약 10% 내지 약 70%, 약 10% 내지 약 80%, 약 10% 내지 약 90%, 약 20% 내지 약 30%, 약 20% 내지 약 35%, 약 20% 내지 약 40%, 약 20% 내지 약 45%, 약 20% 내지 약 50%, 약 20% 내지 약 55%, 약 20% 내지 약 60%, 약 20% 내지 약 70%, 약 20% 내지 약 80%, 약 20% 내지 약 90%, 약 30% 내지 약 35%, 약 30% 내지 약 40%, 약 30% 내지 약 45%, 약 30% 내지 약 50%, 약 30% 내지 약 55%, 약 30% 내지 약 60%, 약 30% 내지 약 70%, 약 30% 내지 약 80%, 약 30% 내지 약 90%, 약 35% 내지 약 40%, 약 35% 내지 약 45%, 약 35% 내지 약 50%, 약 35% 내지 약 55%, 약 35% 내지 약 60%, 약 35% 내지 약 70%, 약 35% 내지 약 80%, 약 35% 내지 약 90%, 약 40% 내지 약 45%, 약 40% 내지 약 50%, 약 40% 내지 약 55%, 약 40% 내지 약 60%, 약 40% 내지 약 70%, 약 40% 내지 약 80%, 약 40% 내지 약 90%, 약 45% 내지 약 50%, 약 45% 내지 약 55%, 약 45% 내지 약 60%, 약 45% 내지 약 70%, 약 45% 내지 약 80%, 약 45% 내지 약 90%, 약 50% 내지 약 55%, 약 50% 내지 약 60%, 약 50% 내지 약 70%, 약 50% 내지 약 80%, 약 50% 내지 약 90%, 약 55% 내지 약 60%, 약 55% 내지 약 70%, 약 55% 내지 약 80%, 약 55% 내지 약 90%, 약 60% 내지 약 70%, 약 60% 내지 약 80%, 약 60% 내지 약 90%, 약 70% 내지 약 80%, 약 70% 내지 약 90%, 약 80% 내지 약 90%, 약 10%, 약 20%, 약 30%, 약 35%, 약 40%, 약 45%, 약 50%, 약 55%, 약 60%, 약 70%, 약 80%, 약 90%, 약 95%, 또는 약 100%인 것인 조성물.

[0394] 64. 구체예 25 내지 30 및 44 내지 45 중 어느 하나, 구체예 31 내지 43 중 어느 하나, 또는 구체예 47 내지 50 중 어느 하나에 있어서, 상기 합성 DNA는 역전사효소에 의한 RNA 바이러스 게놈의 전사에 의해 생성되는 것인 조성물, 시스템, 또는 방법.

도면의 간단한 설명

[0395] 특허 또는 출원 파일은 적어도 하나의 컬러 도면을 포함한다. 컬러 도면을 포함하는 본 특허 또는 특허 출원의 카피는 신청 및 필요한 수수료의 납부시 특허청에 의해 제공될 것이다.

도 1a는 벡터 설계 (좌측) 및 RNA 상호작용 및 스플라이싱 (우측)의 개략도이다. 좌측: 5' 트랜스-스플라이스 (trsp) DNA 벡터: 개방 화살표(open arrow)는 2개의 대향하는 프로모터이다. RFP 코딩 도메인 및 폴리 아데닐화 요소를 갖는 3' UTR이 SD (splice donor sequence), DISE (downstream intronic splicing enhancer), 및 2개의 ISE (intronic splicing enhancer) (2xISE), 결합 도메인 (BD, 이합체화 도메인으로도 지칭됨), 및 안정한 스템 루프(stem loop) BoxB 요소 (boxB), 자가-절단(self-cleaving) 해머헤드(hammerhead) 리보자임 (HHrz)으로 이어지고, 폴리아데닐화 요소를 포함하는 3' UTR로 종료되는, YFP의 N-말단 부분 (n-yfp)으로부터 반대편에 발현된다. n-yfp 세그먼트는 삽입된 작은 인트론 (n-yfp 내의 백색 세그먼트)을 갖는다. 3' trsp DNA 벡터: 개방 화살표는 2개의 대향하는 프로모터이다. BFP 코딩 도메인 및 폴리 아데닐화 요소를 갖는 3'UTR이, 상보적인 결합 도메인 (안티-BD, 이합체화 도메인으로도 지칭됨)으로부터 반대편에 발현되고, 상기 도메인은 3개의 ISE (intronic splicing enhancer sequence)(3xISE), 분지점 (BP), 폴리피리미딘 트랙트 (PPT), 스플라이스 억셉터 서열 (SA), YFP 코딩 서열의 c-말단 부분으로 이어지고, 폴리아데닐화 요소를 포함하는 3' UTR로 종료된다. 우측: YFP 단백질을 코딩하는 mRNA를 생성하기 위한 프리-mRNA(pre-mRNA) 상호작용 (5' trsp-RNA + 3' trsp-RNA) 및 트랜스-스플라이싱이 도시된다.

도 1b는 N-말단 발현 플라스미드 단독의 형질감염은 YFP 형광을 초래하지 않는다는 것을 도시한다.

도 1c는 C-말단 발현 플라스미드 단독의 형질감염은 YFP 형광을 초래하지 않는다는 것을 도시한다.

도 1d는 결합 도메인이 없는, N-말단 및 C-말단 단편의 발현이 낮은 수준의 YFP 유도를 보인다는 것을

도시한다.

도 1e는 루프 구성(looped configuration)의 합리적으로 설계된 이합체화/결합 도메인 (이중가닥 스템 구조를 형성하는 상보적 서열에 의해 중단된(interrupted) 피리미딘만으로 또는 퓨린만으로 구성된 저다양성 서열(hypodiverse sequence))을 도시한다.

도 1f는 "루프형(looped)" 이합체화 도메인 구성의 3D 표현을 도시한다.

도 1g는 C-말단 절반에 결합 도메인을 갖지 않는 음성 대조군을 도시한다.

도 1h는 N-말단 절반에 결합 도메인을 갖지 않는 음성 대조군을 도시한다.

도 1i는 N-말단 절반 및 C-말단 절반 모두에서 루프 구성의 매칭되는 결합 도메인이 세포의 90%에서 강한 YFP 유도를 보인다는 것을 도시한다.

도 1j-1n은 피리미딘만으로 (또는 대안적으로 퓨린만으로) 구성된 150개 뉴클레오티드로 이루어진 저다양성 서열을 포함하는 서열이 완전히 개방된 형태(fully open configuration)를 초래한, 결합 도메인의 구성에 대한 도 1e-1i에 도시된 데이터에 균등한 데이터를 도시한다.

도 1j는 상보적 염기쌍 형성을 위한 완전히 열린 구성을 초래한 150개 뉴클레오티드 저다양성 피리미딘 서열을 도시한다.

도 1k는 (1j)로부터의 150개 저다양성 피리미딘 서열의 3D 표현을 도시한다.

도 1l은 상보적 저다양성 결합 도메인이 결여된 C-말단-YFP 코딩 구조체에 의한 대조군 HEK293T 세포 형질감염을 도시한다. 소수의 형질감염된 세포가 YFP를 발현한다.

도 1m은 상보적 저다양성 결합 도메인이 결여된 N-말단-YFP 코딩 구조체에 의한 대조군 HEK293T 세포 형질감염을 도시한다. 소수의 형질감염된 세포가 YFP를 발현한다.

도 1n은 상보적 저다양성 이합체화 도메인을 갖는 N-말단-YFP 및 C-말단-YFP 코딩 구조체에 의한 HEK293T 세포 형질감염을 도시한다. 다수의 세포가 높은 수준으로 YFP를 발현한다.

도 1o는 도 1g에서 표시된 세포에 대한 대표적인 형광 이미지를 보여준다. 형질감염에 대한 양성 마커(RFP+BFP)가 발현되나, YFP 단백질은 효율적으로 재구성되지 않는다.

도 1p는 도 1l에서 표시된 세포에 대한 대표적인 형광 이미지를 보여준다. 형질감염에 대한 양성 마커(RFP+BFP)가 발현되고 RFP 및 BFP 이중 양성인 세포에서 YFP 단백질이 높은 수준으로 재구성된다.

도 1q는 도 1d, 도 1g 내지 1i, 및 도 1l 내지 1n에 표시된 조건의 비교이다. N: 결합 도메인 불포함(no binding domain), Loop: 루프형 저다양성 결합 도메인 구성(looped hypodiverse binding domain configuration), Lin: 선형 저다양성 구성(linear hypodiverse configuration).

도 2a는 벡터 설계의 개략도이다. YFP(yellow fluorescent protein)의 단백질 코딩 서열이 N-말단, 중간 단편(m-yfp) 및 C-말단 단편으로 분할된다. n 및 m 단편을 코딩하는 RNA의 연결부(junction)가 루프형 결합 도메인(BD1)에 의해 연결되고, m 단편과 c 단편 사이의 연결부가 루프형 결합 도메인 (BD2)에 의해 연결된다. 피리미딘 (Y) 및 퓨린 (R) 서열이 m-단편의 자가-고리화(self-circularization)를 방지하고 N-단편과 C-단편의 직접적인 재조합을 방지하도록 배열된다. N-말단 단편은 형질감염 대조군으로서 RFP (red fluorescent protein)와 공-발현되고, C-말단 단편은 형질감염 대조군으로서 BFP (blue fluorescent protein)와 공-발현된다. 프로모터 서열이 열린 화살표로 표시된다. 스플라이스 도너 (SD) 및 스플라이스 억셉터 (SA) 부위가 표시된다. 도 1a에서 SA의 상류(5') 및 SD의 하류(3')에서 이용된 요소와 유사한, 스플라이스 인핸서, 폴리피리미딘 트랙트 및 분지점을 포함한 ISE (intronic splicing element)가 포함된다.

도 2b는 형질감염된 세포의 80%에서 높은 수준의 YFP 발현을 효율적으로 재구성한, 플라스미드 I+II+III (도 2a 참조)의 인간 세포주 형질감염을 보여준다.

도 2c는 n 및 m 단편의 발현 (플라스미드 I+II, 도 2a 참조)의 대표적인 형광 이미지가 yfp 형광을 보이지 않는다 (음성 대조군)는 것을 보여준다.

도 2d는 m 및 c 단편의 발현 (플라스미드 II+III, 도 2a 참조)의 대표적인 형광 이미지가 yfp 형광을 보이지 않는다 (음성 대조군)는 것을 보여준다.

도 2e는 3개의 단편 모두의 공-형질감염 (플라스미드 I+II+III, 도 2a 참조)에 의해 강한 YFP 형광이 유도된다는 것을 보여주는 대표적인 형광 이미지를 보여준다.

도 3a-3d는 신생 (P3) 마우스 새끼에서 전신 투여 후 2개의 AAV2/8s로부터 발현된 2개의 단편 (서열번호 1 및 2)으로부터의 YFP (yellow fluorescent protein)의 효율적인 재구성을 도시한다. 도 3a는 YFP의 n-말단 절반 단편을 코딩하는 AAV 1 및 c-말단 절반 단편을 코딩하는 AAV 2를 도시한다. AAV 1+AAV 2를 동일한 역가(titer)로 혼합하고 마우스에 정맥내로 주사했다. 주사 후 3주차에 조직 시료를 수집했다. 도 3b는 희생 시점에 미성년 마우스(juvenile mouse)의 간에서 YFP 형광을 보여준다 (그린). 주사되지 않았던 마우스의 간(uninjected liver)이 비교를 위해 표시된다 (대조군: YFP 미검출). DRAQ5 핵 염색이 상황(context)을 위해 마젠타색으로 표시된다. 도 3c는 희생 시점에 심장 근육에서 강한 YFP 형광을 q보여준다 (그린). 상부 패널은 상황을 위한 DARQ5 핵 염색(마젠타색)을 갖는 단면을 보여준다. 하부 패널은 상황을 위한 DARQ5 핵 염색(마젠타색)을 갖는 단면을 보여준다. YFP가 결여된, 주사되지 않았던 마우스의 심장이 대조군으로 표시된다. 도 3d는 희생 시점에 다리의 골격근에서 강한 YFP 형광을 보여준다. 주사되지 않았던 마우스의 다리가 비교를 위해 표시된다 (음성 대조군, YFP 미검출). 상부 패널은 마젠타색으로 레드 자가발광을 갖는 육안 관찰도(macroscopic view)이다. 하부 패널은 다리를 관통하는 단면의 현미경 이미지를 보여준다. 하부 패널은 상황을 위한 마제타색의 DRAQ5 핵 염색을 보여준다.

도 4a-4b는 신생 (P3) 마우스 새끼에서 3종의 AAV2/8의 근육내 주사 후 마우스 전경근(tibialis anterior muscle)에서 3개의 단편 (각각, 서열번호 145, 146 및 2)으로부터의 YFP (yellow fluorescent protein)의 효율적인 재구성을 도시한다. 도 4a는 개별적인 YFP의 N-말단 단편, M-단편, 및 C-말단 단편을 갖는 3개의 AAV 입자의 개략도를 보여준다. 도 4b는 3개의 바이러스 입자 모두가 주사된 마우스의 전경근의 종단면에서 강한 YFP 형광을 보여준다. DRAQ5 핵 염색이 상황을 위해 마젠타로 표시된다.

도 5a-5f는 성체 마우스 전경근에서 2개의 단편 및 3개의 단편으로부터의 YFP의 효율적인 재구성을 도시한다. 도 5a는 YFP 코딩 서열의 N-말단 절반 및 C-말단 절반이 합성 RNA-이합체화 및 재조합 도메인을 갖는다는 것을 도시한다. 도 5b는 이러한 2개의 단편을 발현하는 2종의 AAV 전달 플라스미드를 성체 마우스 TA(tibialis anterior) 근육 내로 경피로 전기천공시키고, 전기천공 후 5일차에 강한 형광이 검출되었다는 것을 보여준다. 도 5c는 반대쪽 비-주사(contralateral non-injected) TA에서 형광이 검출되지 않았다는 것을 보여준다. 도 5d는 n-말단, 중간, 및 c-말단 YFP 코딩 서열이 각 단편을 그의 인접한 단편에 연결시키는 합성 RNA-이합체화 및 재조합 도메인을 갖는다는 것을 보여준다. 도 5e는 이러한 3개의 단편을 발현하는 3종의 AAV 전달 플라스미드의 경피 전기천공을 도시한다. 강한 YFP 형광이 검출되어, 3개의 단편으로부터 YFP의 효율적인 재구성을 나타낸다. 도 5f는 반대쪽 비-주사 TA에서의 형광을 도시한다. 형광 채널(fluorescent channel)이 상황을 위해 그레이 스케일 사진 위에 오버레이된다.

도 6a는 표적 단백질이 2개의 부분으로 나뉘고, 각 부분이 상이한 핵산 분자에 의해 코딩되는 것인 2개의 핵산 분자 (110, 150)를 이용하는, 본 명세서에서 개시되는 RNA 재조합 방법을 위한 예시적 시스템을 제공하는 개략도이다. 일부 구체예에서, 상기 시스템의 핵산 분자 (110, 150)는 DNA이고, 프로모터 (112, 152)를 포함한다. 일부 구체예에서, 상기 시스템의 핵산 분자 (110, 150)는 RNA이고, 따라서, 프로모터 (112, 152)를 포함하지 않는다. 도면은 축척에 따라 도시되지 않음.

도 6b는 슈도노트(pseudoknot) 형성의 부재시 개방되어 염기쌍 형성에 이용될 수 있는 국소 RNA 루프를 초래하는, 스템(stem)을 형성할 수 있는 서열들이 간간이 배치된(interspersed) 저다양성 서열을 포함하는 예시적 이합체화 도메인(예를 들면, 도 6a의 122, 154)을 제공하는 개략도이다. 도면은 축척에 따라 도시되지 않음.

도 6c는 분자 (110)의 프리-mRNA 이합체화 도메인 (122)(도 6a)과 분자 (150)의 프리-mRNA 이합체화 도메인 (154)(도 6a) 간의 상호작용 및 혼성화 (염기쌍 형성)가 스플라이세오솜(spliceosome) 성분이 N-말단 코딩 서열 (114)과 C-말단 코딩 서열 (164)을 재조합할 수 있게 한다는 것을 보여주는 개략도이다. 이는 N-말단 단백질 코딩 서열 (114)의 3' 말단의 C-말단 단백질 서열 (164)의 5' 말단으로의 융합 및 N-말단 부분과 C-말단 부분 간에 무결절성 연결(seamless junction)을 초래한다. 도면은 축척에 따라 도시되지 않음.

도 6d는 표적 단백질이 3개의 부분 (N-말단, 중간, C-말단)으로 나뉘고, 각각의 부분이 상이한 핵산 분자에 의해 코딩되는 것인 3개의 핵산 분자 (110, 200, 150)를 이용한, 본 명세서에서 개시되는 RNA 재조합 방법을 위한 예시적 시스템을 제공하는 개략도이다. 전사 전에, 상기 시스템의 핵산 분자(110, 150, 200)는 DNA이고, 프로모터 (112, 152, 202)를 포함한다. 전사 후에, 상기 시스템의 핵산 분자 (110, 150, 200)는 RNA이고, 따라서, 프

로모터(112, 152, 202)가 없다. 도면은 축척에 따라 도시되지 않음.

도 6e는 분자 (110)의 이합체화 도메인 (122)(도 6d)과 분자 (200)의 이합체화 도메인 (204)(도 6d) 간, 및 분자 (200)의 이합체화 도메인 (204)(도 6d)과 분자 (150)의 이합체화 도메인 (154)(도 6d) 간의 상호작용 및 혼성화 (염기쌍 형성)가 스플라이세오솜 성분이 N-말단 코딩 서열 (114), 중간 단백질 코딩 서열 (216), 및 C-말단 코딩 서열 (164)을 재조합할 수 있게 한다는 것을 보여주는 개략도이다. 이는 N-말단 단백질 코딩 서열 (114)의 3' 말단의 중간 코딩 서열 (216)의 5' 말단으로의 융합, 및 중간 코딩 서열 (216)의 3' 말단의 C-말단 서열 (164)의 5' 말단으로의 융합 및 N-말단 부분, 중간 부분, 및 C-말단 부분 간에 무결절성 연결을 초래한다. 일부 구체예에서, 전사 후에, 표시된 요소들은 RNA이다. 도면은 축척에 따라 도시되지 않음.

도 6f는 표적 단백질이 2개의 부분으로 나뉘고, 각 부분이 상이한 핵산 분자에 의해 코딩되는 것인 2개의 핵산 분자 (110, 150)를 이용하는, 본 명세서에서 개시되는 RNA 재조합 방법을 위한 예시적 시스템을 제공하는 개략도이다. 이 구체예에서, DNA는 RNA로 전사되어, 상기 시스템의 핵산 분자 (110, 150)는 RNA이고, 따라서, DNA에 존재하는 프로모터 (112, 152)가 없다 (도 6a 참조). 도면은 축척에 따라 도시되지 않음.

도 7a는 도 6a에서와 같이, 2개의 핵산 분자 (500, 600)를 이용하나, 이합체화 도메인이 동일한 표적 단백질 (700)을 인식하는 압타머 (512, 602)인 것인 본 명세서에서 개시되는 RNA 재조합 방법을 위한 예시적 시스템을 제공하는 개략도이다. 일부 구체예에서, 예를 들면, 전사 후에, 표시된 요소들은 RNA이다. 도면은 축척에 따라 도시되지 않음.

도 7b는 도 7a와 관련하여, 동일한 표적 단백질을 인식하는 이합체화 도메인을 이용하는, 본 명세서에서 개시되는 RNA 재조합 방법을 위한 예시적 시스템을 제공하는 개략도이다. 여기에서, 이합체화 도메인에 의해 인식되는 표적은 (도 7a의 분자 (700), 예를 들면, 단백질 또는 소분자 대신에) 특이적 RNA 분자이다. 각 도메인은 표적 세포 (즉, 표적 단백질 발현이 요구되는 세포)에서만 발현되는 mRNA 분자의 상이한 부분, 예를 들면, 암-특이적 전사물을 인식한다. 일부 구체예에서, 예를 들면, 전사 후에, 표시된 요소들은 RNA이다. 도면은 축척에 따라 도시되지 않음.

도 7c는 도 6a 및 7a에서와 같이, 2개의 핵산 분자 (800, 900)를 이용하고, 이합체화 도메인이 상호 간에 상호작용하는 것을 방지하고, 따라서 N-말단 코딩 서열 (802)과 C-말단 코딩 서열 (914)의 재조합을 방지하거나 또는 감소시키는 올리고뉴클레오티드 (1000)에 혼성화되는 이합체화 도메인 (812, 902)을 보여주는, 본 명세서에서 개시되는 RNA 재조합 방법을 위한 예시적 시스템을 제공하는 개략도이다. 일부 구체예에서, 예를 들면, 전사 후에, 표시된 요소들은 RNA이다. 도면은 축척에 따라 도시되지 않음.

도 8은 3' 비번역(untranslated) 영역 중 WPRE3 서열의 존재 (w/) 또는 부재 (w/o) 하에 YFP 단백질 발현의 재구성을 비교하는 막대 그래프이다.

도 9a는 고 친화도 이합체화를 위한 키싱 루프 상호작용을 포함하는 이합체화 도메인 (예를 들면, 도 6a의 122, 154)의 이용을 위한 구체예를 제공하는 개략도이다. 본 명세서에서 제공되는 교시를 이용하여, 본 명세서에서 개시되는 코딩 부분들(예를 들면, YFP)이 다른 표적 단백질 코딩 서열로 대체될 수 있다는 것을 이해할 것이다. 도면은 축척에 따라 도시되지 않음.

도 9b는 저다양성 설계 원칙에 따른 선형 이합체화 도메인 또는 키싱 루프-루프 상호작용을 위해 설계된 구조화된(structured) 이합체화 도메인을 갖춘, 분할된(split) YFP의 두 절반 모두에 의해 형질감염된 HEK293T 세포에서의 RFP, BFP, 및 YFP 신호를 보여준다. 강한 황색 형광 신호(yellow fluorescent signal)가 효율적인 재구성을 나타낸다.

도 10a 내지 10z는 본 발명의 시스템 및 방법에서 사용될 수 있는 예시적 합성 핵산 분자이다. 일부 구체예에서, 합성 핵산 분자는 서열번호 1 (도 10a-10b), 2 (도 10c-10e), 7 (도 10e), 8 (도 10f), 9 (도 10g), 10 (도 10h), 11 (도 10i), 12 (도 10j), 13 (도 10k), 14 (도 10l), 15 (도 10m), 16 (도 10n), 17 (도 10o), 18 (도 10p), 19 (도 10q), 20 (도 10r-10u), 및 21 (도 10v-10z) 중 하나의 서열에 대해 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 가지나, 상이한 표적 단백질 코딩 서열을 갖는다. 따라서, 본 명세서에서 제공되는 시스템 또는 방법에서 사용되는 인트론 영역은 서열번호 1, 2, 3, 4, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 또는 21의 인트론 서열에 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 가질 수 있다. 예를 들면, 도 10a-d는 전장 YFP를 발현하기 위해 이용될 수 있는 예시적 (10a, 10b) 제1 (서열번호 1) 및 (10c, 10d) 제2 (서열번호 2) 합성 분자를 보여주고, 서열번호 3 및 4는 YFP 코딩 영역이 없는 상응하는 합성

인트론 부분을 제공한다. 일부 구체예에서, 합성 인트론 서열은 서열번호 3 또는 4에 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는다. 따라서, 본 명세서에서 제공되는 임의의 합성 분자의 코딩 서열 부분 (예를 들면, 서열번호 1의 nt 544 내지 1032 및 서열번호 2의 nt 905 내지 1141)은 또 다른 코딩 서열 부분에 의해 교체될 수 있다.

도 11은 상이한 길이의 랜덤 상보적(complimentary) 염기쌍 형성 결합 도메인 (50 bp, 100 bp, 150 bp, 200 bp, 300 bp, 400 bp, 및 500 bp)의 재구성 효율을 보여주는 막대 그래프이다. YFP 형광 강도 중앙값이 매칭되는 RFP 및 BFP 형질감염 수준을 갖는 세포들 간에 비교된다. 조건 당 n=3개 시료.

도 12a-12b는 스플라이스 인핸서의 합성 인트론으로의 내포가 재구성 효율을 증가시킨다는 것을 보여준다. 도 12a는 사용된 5'-N 및 3'-C-말단 구조체 (서열번호 1 및 2)의 개략도이다(약어는 도 1a를 참조한다). 도 12b는 서열번호 1 및 2, 또는 Δ에 의해 표시된 그의 다양한 절단의 세포로의 형질감염 후 결과적으로 수득된 YFP 형광을 보여주는 막대 그래프이다. 조건 당 n=3개 시료.

도 13a-13d는 2개의 절편 (서열번호 147 및 148)으로부터의 전장 flp 재조합효소 (Flpo)의 재구성에 의한 정중선-교차 피질 뉴런 추적(midline-crossing cortical neuron tracing)을 보여준다. (도 13a) flop를 재구성하기 위해 사용된 5'- 및 3'-서열의 개략적 표현 (도 12a의 구조체와 유사함) (도 13b) 각각 피질의 좌측 영역 및 우측 영역에 주사된, N-flop 및 C-flop 코딩 AAV 바이러스가 주사된 flp-리포터 마우스 라인의 개략적 표현. (도 13c 및 13d)는 뇌의 대측성 반구로 돌출하고, 따라서, N-flpo 및 C-flpo 바이러스 모두에 의해 감염된 피질 뉴런의 신경 세포체(neuronal cell body) 및 액손 표지(도 13c 및 13d)를 보여준다. Hoechst 염색 (핵)이 상황을 위해 표시된다.

도 14a-14d는 세포 배양 및 인 비보에서 마우스 일차 운동 피질(primary motor cortex) 중 거대 카고(oversized cargo)(즉, 긴 RNA에 의해 코딩된 단백질)의 발현을 보여준다. (도 14a) 긴 스터퍼(stuffer) 서열(비중단(uninterrupted) 개방 해독 프레임; 각각 서열번호 22 및 23)을 포함하는, YFP를 재구성하기 위해 이용된 5'- 및 3'-서열의 개략적 표현. (도 14b) HEK 293t 세포에서 거대 YFP 구조체의 재구성 효율의 정량적 실시간 PCR 분석. 조건 당 N=3. (도 14c) 일시적으로 형질감염된 HEK 293t 세포의 유동세포분석법에 의해 평가된 전장 거대 YFP 발현 및 분할-REJ 발현으로부터의 재구성 YFP 단백질 발현. 상이한 조건에서 동일한 형질감염 대조군 (블루 및 레드) 형광을 갖는 세포 집단 간에 황색 형광 강도 중앙값을 비교한다. Y-축은 황색 형광 강도 중앙값 [a.u.]을 나타낸다. 조건 당 N=3. (도 14d) 마우스 일차 운동 피질로의 주사의 개략도, 및 인 비보에서 긴 (2401 aa) YFP 단백질의 성공적인 재구성을 보여주는, 주사 후 10일차 뇌 조직의 이미지.

도 15a-15c는 N-말단 HA 태그 (N-말단 신호 펩티드를 치환함)를 갖는 전장 인간 응고 인자 VIII (FVIII) (2317 aa)의 효율적인 재구성을 보여준다. (도 15a) FVIII을 재구성하기 위해 이용된 5'- 및 3'-서열(각각, 서열번호 24 및 25)의 개략적 표현. (도 15b) 연결부(junction)의 PCR 증폭. (도 15c) FVIII의 발현을 보여주는 웨스턴 블롯. 래인 1-3: 전장 FVIII의 발현 (290kDa 밴드가 전장, 미가공(unprocessed) FVIII을 나타냄). 래인 4-6: 재구성된 FVIII의 발현 (290kDa 밴드가 성공적으로 재구성된 FVIII을 나타냄). 래인 7 및 8: N-말단 단독의 발현이 290 kDa에서 전장 FVIII 밴드의 부재를 보여준다. 모든 래인에 대해: 예상되는 단백질분해 가공 산물은 ~75kDa 내지 ~210kDa 범위에서 관찰된다. 마우스 항-HA 일차 항체를 이용하여 FVIII을 탐색한다. 모든 래인에 5 마이크로그램의 세정된(cleared) 세포 단백질 추출물을 로딩했다. GAPDH (토끼 항-GAPDH)를 로딩 대조군으로서 탐색한다.

도 16a-16f는 C-말단 FLAG-태그를 갖는 전장 인간 Abca4 (2300 aa)의 효율적 재구성을 보여준다. (도 16a) Abca4를 재구성하기 위해 이용된 5'- 및 3'-서열(각각, 서열번호 20 및 21), 및 연결부를 포함한 생거 시퀀싱 트레이스(Sanger sequencing trace across the junction)의 개략적 표현. (도 16b) 연결부의 PCR 증폭. (도 16c) 5'- 및 3'-단편의 재조합을 분석하기 위해 사용된 프로브의 개략적 표현. (도 16d) HEK 293t 세포에서 2일의 발현 후 재구성 효율의 PCR 정량. 조건당 N=2. (도 16e) Abca4의 발현을 보여주는 웨스턴 블롯. 래인 1-3: 전장 Abca4의 발현 (~260kDa 밴드가 전장 Abca4를 나타냄). 래인 4-6: 재구성된 Abca4의 발현 (260kDa 밴드가 성공적으로 재구성된 Abca4를 나타냄). 래인 7 및 8: 형질감염 불포함 대조군 (즉, HEK 293t 용해물 단독)은 신호의 부재를 보여준다. 마우스 항-FLAG 일차 항체를 이용하여 Abca4를 탐색한다. 모든 래인에 5 마이크로그램의 세정된 세포 단백질 추출물을 로딩했다. GAPDH (토끼 항-GAPDH)를 로딩 대조군으로서 탐색한다. (도 16f) 차등적 BFP 농도에 대해 정규화된 (도 16e)의 웨스턴 블롯의 정량. 전장 발현 대조군의 평균으로 정규화된 데이터가 표시된다.

도 17a 및 17b는 (도 17a) HIV-1 기반 키싱 루프 이합체화 도메인 (N-단편, 서열번호 139, C-단편 서열번호

140); 및 (도 17b) HIV-2 기반 키싱 루프 이합체화 도메인 (N-단편, 서열번호 141, C-단편 서열번호 142)을 제공한다.

도 18a-18c는 C-말단 FLAG-태그를 갖는 전장 마우스 Otof (2019 aa)의 효율적인 재구성을 보여준다. 사용된 5' 및 3' 분자의 DNA 서열이 서열번호 155 및 156으로 표시된다. (도 18a) Otof의 발현을 보여주는 웨스턴 블롯. 래인 1-3: 전장 Otof의 발현 (~250kDa 밴드가 전장 Otof를 나타냄). 래인 4-6: 재구성 Otof의 발현 (250k Da의 밴드는 성공적으로 재구성된 Otof를 나타냄). 래인 7: 형질감염 불포함 대조군 (즉, HEK 293t 용해물 단독)은 신호의 부재를 보여준다. 마우스 항-FLAG 일차 항체를 이용하여 Otof를 탐색한다. 모든 래인에 5 마이크로그램의 세정된 세포 단백질 추출물을 로딩했다. GAPDH (토끼 항-GAPDH)를 로딩 대조군으로서 탐색한다. (도 18b) 웨스턴 블롯의 원시 정량화(raw quantification) 및 (도 18c) 차등적 BFP 농도에 대해 정규화된 정량화. 전장 발현 대조군의 평균으로 정규화된 데이터가 표시된다.

도 19a-19c는 C-말단 FLAG-태그를 가진 전장 인간 Myo7a (2243 aa)의 효율적인 재구성을 보여준다. 사용된 5' 및 3' 분자의 DNA 서열이 서열번호 157 및 158로 표시된다. (도 19a) Myo7a의 발현을 보여주는 웨스턴 블롯. 래인 1-3: 전장 Myo7a의 발현 (~270kDa 밴드가 전장 AncBE4를 나타냄). 래인 4-6: 재구성된 Myo7a의 발현 (270kDa 밴드는 성공적으로 재구성된 AncBE4를 나타냄). 래인 7: 형질감염 불포함 대조군 (즉, HEK 293t 용해물 단독)은 신호의 부재를 보여준다. 마우스 항-FLAG 일차 항체를 이용하여 Myo7a를 탐색한다. 모든 래인에 5 마이크로그램의 세정된 세포 단백질 추출물을 로딩했다. GAPDH (토끼 항-GAPDH)를 로딩 대조군으로서 탐색한다. (도 19b) 웨스턴 블롯의 원시 정량화 및 (도 19c) 차등적 BFP 농도에 대해 정규화된 정량화. 전장 발현 대조군의 평균으로 정규화된 데이터가 표시된다.

도 20a-20d는 전장 DCas9-VPR (1951 aa)의 효율적인 재구성을 보여준다. 사용된 5' 및 3' 분자의 DNA 서열이 서열번호 159 및 160으로 표시된다. (도 20a) DCas9-VPR의 발현을 보여주는 웨스턴 블롯. 래인 1-3: 전장 DCas9-VPR의 발현 (~250kDa 밴드가 전장 DCas9-VPR을 나타냄). 래인 4-6: 재구성된 DCas9-VPR의 발현 (250kDa 밴드는 성공적으로 재구성된 AncBE4를 나타냄). 래인 7: 형질감염 불포함 대조군 (즉, HEK 293t 용해물 단독)은 신호의 부재를 보여준다. 마우스 항-Cas9 일차 항체를 이용하여 DCas9-VPR을 탐색한다. 모든 래인에 5 마이크로그램의 세정된 세포 단백질 추출물을 로딩했다. GAPDH (토끼 항-GAPDH)를 로딩 대조군으로서 탐색한다. (도 20b) 웨스턴 블롯의 원시 정량화 및 (도 20c) 차등적 BFP 농도에 대해 정규화된 정량화. 전장 발현 대조군의 평균으로 정규화된 데이터가 표시된다. (도 20d) HEK 293t 세포에서 YFP 발현 플라스미드의 전사적 활성화의 예. 전장 (상부 패널) 또는 이원 분할(two-way split) REJ-듀얼 dCas9-VPR (하부 패널)이 비-표적화 가이드 RNA (좌측 패널) 또는 UAS-표적화 가이드 RNA (우측 패널) 발현 플라스미드와 함께 일시적으로 형질감염된다. 모든 세포를 또한 YFP의 발현을 초래하는 dCas9-VPR이 최소 프로모터(minimal promoter)의 상류 영역에 표적화될 때까지 전사적으로 불활성인 UAS-YFP 플라스미드로 형질감염시킨다. RFP(Red fluorescent protein)가 dCas9-VPR의 N-말단 단편과 함께 발현되고, BFP(Blue fluoreent protein)는 각각 전장 dCas9-VPR 또는 dCas9-VPR의 C-말단 단편과 함께 발현된다. RFP 및 BFP가 형질감염 대조군으로 작용한다. UAS-표적화 가이드 RNA와 쌍을 이룬 전장 및 이원 분할 dCas-VPR의 발현시, YFP 단백질 발현이 관찰되어, 재구성된 전장 단백질의 기능성을 확인한다.

도 21a-21d는 전장 인간화 Prime Editor (2118 aa)의 효율적인 재구성을 보여준다. 사용된 5' 및 3' 분자의 DNA 서열이 서열번호 161 및 162로 표시된다. (도 21a) Prime Editor의 발현을 보여주는 웨스턴 블롯. 래인 1-3: 전장 Prime Editor의 발현 (~260kDa 밴드가 전장 Prime Editor를 나타냄). 래인 4-6: 재구성된 Prime Editor의 발현 (260kDa 밴드는 성공적으로 재구성된 Prime Editor를 나타냄). 래인 7: 형질감염 불포함 대조군 (즉, HEK 293t 용해물 단독)은 신호의 부재를 보여준다. 마우스 항-Cas9 일차 항체를 이용하여 Prime Editor를 탐색한다. 모든 래인에 5 마이크로그램의 세정된 세포 단백질 추출물을 로딩했다. GAPDH (토끼 항-GAPDH)를 로딩 대조군으로서 탐색한다. (도 21b) 웨스턴 블롯의 원시 정량화 및 (도 21c) 차등적 BFP 농도에 대해 정규화된 정량화. 전장 발현 대조군의 평균으로 정규화된 데이터가 표시된다. (도 21d)는 Prime Editor가 HEK293t 세포의 FANCF 및 VEGFA3 유전자좌에서 유도된 G의 T로의 전환(transversion)을 유도했다는 것을 보여준다. 상부 패널은 각각 FANCF 및 VEGFA3 유전자좌에 대한 서열 상황(context)을 보여준다. 회색 화살표는 pegRNA (prime editor guide RNA)에 의해 표적화되는 서열을 나타낸다. PAM (protospacer adjacent motif)은 회색 박스로 표시된다. T로의 전환을 위해 표적화되는 G가 서열 중에 하이라이트된다. 3가지 조건에서 Sanger 서열을 이용하여 게놈 유전자좌들을 시퀀싱한다. 상부 패널은 편집되지 않은 야생형 조건에 대한 대표적인 생거 트레이스(sanger trace)를 보여준다. 상부 패널로부터 2번째는 전장 발현 프라임 에디터(prime editor) 구조체를 나타내는 대표적인 생거 트레이스를 보여준다. 블랙 박스로 하이라이트된 영역은 세포의 일부에서 에디트 (edit)의 성공적인 포함

을 나타내는, 생거 서열 중 T 밴드의 출현을 보여준다. 최하단의 패널은 이원 분할 재구성 프라임 에디터에 의해 편집된 세포에 대한 대표적인 생거 트레이스를 보여준다. T 트레이스 (블랙 박스)의 출현은 2개의 단편으로부터 재구성되는 경우, 프라임 에디터의 기능성을 입증한다.

도 22a-22c는 전장 인간화 AncBE4 (Cytosine Base Editor) (1854 aa)의 효율적 재구성을 보여준다. 사용된 5' 및 3' 분자의 DNA 서열이 서열번호 163 및 164로 표시된다. (도 22a) AncBE4의 발현을 보여주는 웨스턴 블롯. 래인 1-3: 전장 AncBE4의 발현 (~230kDa 밴드가 전장 AncBE4를 나타냄). 래인 4-6: 재구성된 AncBE4의 발현 (230kDa 밴드는 성공적으로 재구성된 AncBE4를 나타냄). 래인 7: 형질감염 불포함 대조군 (즉, HEK 293t 용해물 단독)은 신호의 부재를 보여준다. 마우스 항-Cas9 일차 항체를 이용하여 AncBE4를 탐색한다. 모든 래인에 5 마이크로그램의 세정된 세포 단백질 추출물을 로딩했다. GAPDH (토끼 항-GAPDH)를 로딩 대조군으로서 탐색한다. (도 22b) 웨스턴 블롯의 원시 정량화. 전장 발현 대조군의 평균으로 정규화된 데이터가 표시된다. (도 22c)는 AncBE4가 HEK293t 세포의 EMX1 및 HEK 부위 3 유전자좌에서 유도된 C의 T로의 전이(transition) 돌연변이를 유도했다는 것을 보여준다. 상부 패널은 각각 EMX1 및 HEK 부위 3 유전자좌에 대한 서열 상황을 보여준다. 회색 화살표는 AncBE4 가이드 RNA (sgRNA)에 의해 표적화되는 서열을 나타낸다. PAM (protospacer adjacent motif)은 회색 박스로 표시된다. T로의 전이를 위해 표적화되는 C가 서열 중에 하이라이트된다. 3가지 조건에서 Sanger 서열을 이용하여 게놈 유전자좌들을 시퀀싱한다. 상부 패널은 편집되지 않은 야생형 조건에 대한 대표적인 생거 트레이스를 보여준다. 상부 패널로부터 2번째는 전장 발현 AncBE4 구조체를 나타내는 대표적인 생거 트레이스를 보여준다. 블랙 박스로 하이라이트된 영역은 세포의 일부에서 상기 에디트의 성공적인 포함을 나타내는, 생거 서열 중 T 밴드의 출현을 보여준다. 최하단의 패널은 이원 분할 재구성 AncBE4에 의해 편집된 세포에 대한 대표적인 생거 트레이스를 보여준다. T 트레이스 (블랙 박스)의 출현은 2개의 단편으로부터 재구성되는 경우, AncBE4의 기능성을 입증한다.

도 23a-23c는 전장 인간화 Abe8e (Adenine Base Editor) (1606 aa)의 효율적 재구성을 보여준다. 사용된 5' 및 3' 분자의 DNA 서열이 서열번호 165 및 166으로 표시된다. (도 23a) Abe8e의 발현을 보여주는 웨스턴 블롯. 래인 1-3: 전장 Abe8e의 발현 (~230kDa 밴드가 전장 AncBE4를 나타냄). 래인 4-6: 재구성된 Abe8e의 발현 (230kDa 밴드는 성공적으로 재구성된 AncBE4를 나타냄). 래인 7: 형질감염 불포함 대조군 (즉, HEK 293t 용해물 단독)은 신호의 부재를 보여준다. 마우스 항-Cas9 일차 항체를 이용하여 Abe8e를 탐색한다. 모든 래인에 5 마이크로그램의 세정된 세포 단백질 추출물을 로딩했다. GAPDH (토끼 항-GAPDH)를 로딩 대조군으로서 탐색한다. (도 23b) 웨스턴 블롯의 원시 정량화. 전장 발현 대조군의 평균으로 정규화된 데이터가 표시된다. (도 23c)는 Abe8e가 HEK293t 세포의 BCL11A 및 HGB1/2 유전자좌에서 유도된 A의 G로의 전이를 유도했다는 것을 보여준다. 상부 패널은 각각 BCL11A 및 HGB1/2 유전자좌에 대한 서열 상황을 보여준다. 회색 화살표는 Abe8e 가이드 RNA (sgRNA)에 의해 표적화되는 서열을 나타낸다. PAM (protospacer adjacent motif)은 회색 박스로 표시된다. G로의 전이를 위해 표적화되는 A가 서열 중에 하이라이트된다. 3가지 조건에서 Sanger 서열을 이용하여 게놈 유전자좌들을 시퀀싱한다. 상부 패널은 편집되지 않은 야생형 조건에 대한 대표적인 생거 트레이스를 보여준다. 상부 패널로부터 2번째는 전장 발현 Abe8e 구조체를 나타내는 대표적인 생거 트레이스를 보여준다. 블랙 박스로 하이라이트된 영역은 세포의 일부에서 상기 에디트의 성공적인 포함(incorporation)을 나타내는, 생거 서열 중 G 밴드의 출현을 보여준다. 최하단의 패널은 이원 분할 재구성 Abe8e에 의해 편집된 세포에 대한 대표적인 생거 트레이스를 보여준다. G 트레이스 (블랙 박스)의 출현은 2개의 단편으로부터 재구성되는 경우, Abe8e의 기능성을 입증한다.

도 24a-24c는 DISE(downstream intronic splicing enhancers) 및 ISE(intronic splicing enhancers) 및 억셉터 서열의 RNA 말단 연결(RNA end joining)의 효율에 대한 영향을 보여준다. (도 24a) 스크린 셋업의 개략도. 5' 단편은 인간 CMV 프로모터 및 인핸서를 이용하여 DNA 구조체로부터 전사되는 RNA 분자이다. 생성된 RNA 분자는 큰 카고 크기를 모사하기 위해 긴 스터퍼 개방 해독 프레임(ORF)를 포함한다. 이 스터퍼 서열은 2A 자가-절단 펩티드 서열로 종료되고 YFP의 5' 단편에 대한 코딩 영역 (n-yfp)으로 이어진다. yfp의 5' 단편은 SD(spice donor site)로 종료된다. 이 스플라이스 도너 부위는 RNA 말단 연결 모듈의 5' 인트론 부분으로 이어진다. DISE 및 ISE 서열의 RNA 말단 연결 반응 효율에 대한 영향을 결정하기 위한 목적으로, 5' 인트론 부분을 3개의 단편으로 세분한다: ds: 하류 세그먼트(downstream segment); m: 중간 인트론 세그먼트(mid intronic segment); dd: 도너 원위 세그먼트(donor distal segment). 5' 인트론 부분은 삼중(trimodal) 키싱 루프 DNA 이합체화 도메인으로 이어진다. 메시지(message)는 짧은 폴리 아데닐화 신호로 종료된다. 이 5' RNA 분자의 전체 길이는 큰 카고 재구성 시나리오를 모사하기 위해 ~4kb이다. 3' 단편은 인간 CMV 프로모터 및 인핸서를 이용하여 DNA 구조체로부터 전사되는 RNA 분자이다. 3' 단편은 5' 단편 코딩 RNA 분자 상에 있는 도메인에 상보적인 삼중 키싱 루프 RNA 이합체화 도메인으로 개시된다. 상기 이합체화 도메인은 RNA 말단 연결 모듈의 3' 인트론 부분으로 이어

진다. 이 3' 인트론 부분은 2개의 세그먼트로 세분된다: m: 중간-인트론 세그먼트(mid-intronic segment); ap: 억셉터 근위 세그먼트(acceptor proximal segment). 억셉터 근위 세그먼트는 스플라이세오솜 매개 RNA 연결 반응을 위해 필수적인 분지점 및 폴리피리미딘 트랙트의 변형을 포함한다. 스플라이스 억셉터 (SA) 부위는 3' yfp 코딩 서열로 이어지고, 이는 자가-절단 2A 서열로 이어지고, 뒤이어 긴 스터퍼 ORF로 이어진다. 메시지는 SV40 폴리 아데닐화 신호에 의해 종료된다. 3' RNA 분자의 전체 길이는 큰 카고 재구성 시나리오를 모사하기 위해 ~4kb이다. 2개의 RNA 분자 (5' 단편 및 3' 단편)의 회합은 삼중 키싱 루프 RNA 이합체화 도메인에 의해 매개되고, 스플라이세오솜과 RNA 말단 연결 반응의 동원은 인트론 세그먼트에 의해 매개된다. 성공적인 RNA 말단 연결은 yfp ORF의 재구성 및 뒤이은 YFP의 번역을 가져온다. (도 24b) 유동 세포측정법에 의해 측정된 YFP 발현 강도 중앙값이 다수의 인트론 구성에 대해 표시된다. 제1 그룹(막대 1 내지 9)에서, 잠재적인 하류 인트론 스플라이싱 인핸서(DISE) 서열의 선택이 막대 1 내지 8에서 표시된, 컨센서스 스플라이스 도너 부위 (DNA 구조체 중 GTAAGTATT 및 RNA 서열 중 GUAAGUAUU)와 쌍을 형성한다. 이들은 모두 4개의 염기의 동등한 부분으로 구성된 스크램블된(scrambled) 서열(ds9)로 이어지는 컨센서스 스플라이스 도너에 비교된다. 제2 그룹에서, 잠재적인 인트론 스플라이싱 인핸서의 선택, m1-m16이 스크램블된 서열 (m16)에 비교된다. 마지막 그룹핑에서, 잠재적인 강력한 분지점, 폴리피리미딘 트랙트, 및 스플라이스 억셉터의 선택이 비교된다. 기준 구조체(reference constructs)는 컨센서스 도너의 모든 비-가변 위치의 스크램블된 서열, 및 뒤이어 ds 위치의 스크램블된 서열 및 컨센서스 스플라이스 억셉터 서열 (전체 폴리피리미딘 트랙트는 각각 DNA 구조체 중 T 및 RNA 단편 중 U로 구성됨)로 구성된다. (도 24c) 사용된 상이한 DISE, ISE, 및 스플라이스 억셉터 요소의 목록.

서열 목록

첨부된 서열 목록에 열거된 핵산 및 아미노산 서열은 37 C.F.R. 1.822에 정의된 바와 같이, 뉴클레오티드 염기에 대한 표준 문자 약어, 및 아미노산에 대한 3문자 코드를 이용하여 표시된다. 각 핵산 서열의 하나의 가닥만이 표시되나, 상보적 가닥은 표시된 가닥에 대한 참조에 의해 포함된 것으로 이해된다. 서열 목록은 2020년 9월 30일에 작성된, 157 KB, ASCII 텍스트 파일로 제출되고, 이는 본 명세서에 참조에 의해 첨부된다. 첨부된 서열 목록에서:

서열번호 **1** 및 **2**는 각각 전장 YFP를 발현시키기 위해 사용된 N- 및 C-말단 서열이다. 서열번호 1, CMV 프로모터 nt 1 내지 543, YFP 코딩 서열 nt 544 내지 1032, 합성 인트론 nt 1033 내지 1436, 및 비번역 폴리 A 영역 nt 1437 내지 1491. 서열번호 2, CMV 프로모터 nt 1 내지 522, 합성 인트론 nt 523 내지 904, YFP 코딩 서열 nt 905 내지 1141이고, 및 nt 1142 내지 1302는 비번역 폴리 A 영역이다.

서열번호 **3** 및 **4**는 원하는 전장 단백질을 발현하기 위해 사용될 수 있는, 5'- 및 3'-인트론 서열이고, 상기 전장 단백질의 N-말단 부분은 서열번호 3의 nt 1에 첨가될 수 있고, 상기 전장 단백질의 C-말단 영역은 서열번호 4의 nt 382에 첨가될 수 있다.

서열번호 **5** 및 **6**은 각각 전장 YFP를 발현시키기 위해 사용된 N- 및 C-말단 코딩 서열이다.

서열번호 **7**은 예시적 합성 인트론 이합체화 도메인이다 (도 10e).

서열번호 **8**은 예시적 ISE(intronic splicing enhancer) 불포함 합성 인트론 이다 (도 10f).

서열번호 **9**는 예시적 ISE 불포함 합성 인트론이다 (도 10g).

서열번호 **10**은 예시적 ISE 불포함 합성 인트론이다 (도 10h).

서열번호 **11**은 예시적 결합 도메인 불포함 합성 인트론이다 (도 10i).

서열번호 **12**는 예시적 이합체화 도메인을 갖는 합성 인트론이다 (도 10j).

서열번호 **13**은 예시적 이합체화 도메인을 갖는 합성 인트론이다 (도 10k).

서열번호 **14**는 예시적 ISE 불포함 합성 인트론이다 (도 10l).

서열번호 **15**는 예시적 DISE만 갖는 합성 인트론이다 (도 10m).

서열번호 **16**는 예시적 HHrz 불포함 합성 인트론이다 (도 10n).

서열번호 **17**은 예시적 ISE 불포함 합성 인트론이다 (도 10o).

서열번호 **18**은 예시적 결합 도메인을 갖는 U12 의존적 인트론이다 (도 10p).

서열번호 19는 예시적 결합 도메인을 갖는 U12 의존적 인트론이다 (도 10q).

서열번호 20 및 21은 각각, 전장 Abca4를 초래하는, RNA (프리-mRNA)를 발현하기 위해 사용되는 N- 및 C-말단 DNA 서열이다. 서열번호 20에서, N-말단 Abca4 코딩 서열에 해당하는 서열은 nt 22 내지 3702이고, nt 3703 내지 3912는 합성 인트론이고, 3921 내지 3969는 비번역 폴리 A 영역이다. 서열번호 20은 또한 nt 3703-3711에 스플라이스 도너, nt 3714-3737에 랫트 FGFR2 DISE, nt 3747-3770에 cTNT ISE, nt 3782-3794에 M2 ISE, 및 nt 3801-3975에 키싱 루프 이합체화 도메인을 포함한다. 서열번호 21에서, nt 1 내지 228은 합성 인트론이고, nt 229 내지 3366은 C-말단 Abca4 코딩 영역이며, 3367 내지 3447은 FLAG 에피토프 태그이고, nt 3476 내지 3607은 비번역 폴리 A 영역 (신호)이다. 서열번호 21은 또한, nt 3-114에 키싱 루프 이합체화 도메인, nt 121-133에 M2 ISE, nt 140-163에 cTNT ISE, nt 175-187에 M2 ISE, nt 194-201에 분지점 모티프(Branch Point Motif), nt 207-226에 폴리피리미딘 트랙트, 및 nt 228에 스플라이스 억셉터를 포함한다.

서열번호 22 및 23은 긴 전장 YFP를 초래하는 RNA (프리-mRNA)를 발현하기 위해 사용되는, 각각 N-말단 및 C-말단 DNA 서열이고, 각각 스플라이스 인핸서를 포함한다. 서열번호 22에서, N-말단 YFP 코딩 영역은 nt 22 내지 3702이고, nt 3703 내지 3912는 합성 인트론이며, nt 3921 내지 3969는 비번역 폴리 A 영역이다. 서열번호 22도 nt 3703-3711의 스플라이스 도너, nt 3714-3737의 랫트 FGFR2 DISE, nt 3747-3770의 cTNT 인트론 스플라이싱 인핸서, nt 3782-3794의 M2 인트론 스플라이싱 인핸서, 및 3801-3975의 키싱 루프 이합체화 도메인을 포함한다. 서열번호 23에서, nt 1 내지 225는 합성 인트론이고, nt 226 내지 3747은 C-말단 YFP 코딩 영역이고, nt 3748 내지 3912는 비번역 폴리 A 영역이다. 서열번호 23은 nt 3-114의 키싱 루프 이합체화 도메인, nt 118-130의 M2 인트론 스플라이싱 인핸서, nt 137-160의 cTNT 인트론 스플라이싱 인핸서, nt 172-184의 의 M2 인트론 스플라이싱 인핸서, nt 191-198의 분지점 모티프(Branch Point Motif), nt 204-223의 폴리피리미딘 트랙트, 및 nt 225의 스플라이스 억셉터를 포함한다.

서열번호 24 및 25는 전장 인간 인자 VIII을 초래하는 RNA (프리-mRNA)를 발현하기 위해 사용되는, 각각 N-말단 및 C-말단 서열이다. 서열번호 24에서, N-말단 HA 에피토프 태그를 갖는 N-말단 FVIII 코딩 영역이 nt 22 내지 3561에 존재하고, nt 3562 내지 3771은 합성 인트론이며, nt 3780 내지 3828은 비번역 폴리 A 영역이다. 서열번호 24는 또한 nt 3562-3570의 스플라이스 도너, nt 3573-3596의 랫트 FGFR2 DISE, nt 3606-3629의 cTNT 인트론 스플라이싱 인핸서, nt 3641-3653의 M2 인트론 스플라이싱 인핸서, 및 nt 3660-3834의 키싱 루프 이합체화 도메인을 포함한다. 서열번호 25에서, nt 1 내지 225는 합성 인트론이고, nt 226 내지 3636은 C-말단 FVIII 코딩 영역이며, nt 3665 내지 3797은 비번역 폴리 A 영역이다. 서열번호 25은 또한 nt 3703-3711의 스플라이스 도너, nt 3714-3737의 랫트 FGFR2 DISE, nt 3747-3770의 cTNT 인트론 스플라이싱 인핸서, nt 3782-3794의 M2 인트론 스플라이싱 인핸서, 및 nt 3801-3975의 키싱 루프 이합체화 도메인을 포함한다.

서열번호 26-136은 본 명세서에 기재된 시스템에서 사용될 수 있는 예시적 스플라이싱 인핸서이다 (예를 들면, 도 6a의 118, 120, 156).

서열번호 137 및 138은 예시적 스플라이스 도너 서열이다.

서열번호 139 및 140은 각각 HIV-1 기반 키싱 루프 이합체화 도메인의 N-단편 및 C-단편이다.

서열번호 141 및 142는 각각 HIV-2 기반 키싱 루프 이합체화 도메인의 N-단편 및 C-단편이다.

서열번호 143은 예시적 크립틱 스플라이스 억셉터 서열이다.

서열번호 144는 예시적 분지점 컨센서스 서열이다.

서열번호 145 및 146은 서열번호 2 (C-말단 단편)와 함께 전장 YFP를 발현하기 위해 사용되는, 각각 N-서열 및 중간 서열이다. 서열번호 145에서, nt 1 내지 543은 CMV 프로모터 서열이고, nt 544 내지 849는 N-말단 YFP 코딩 영역이며, nt 850 내지 1305는 합성 인트론이다. 서열번호 146에서, nt 1 내지 522는 CMV 프로모터 서열이고, nt 523 내지 901은 합성 인트론이며, nt 902 내지 1084는 중간 YFP 코딩 영역이고, nt 1085 내지 1543은 비번역 폴리 A 영역이다.

서열번호 147 및 148은 각각, 전장 Flpo를 발현하기 위해 사용되는, 5' 및 3'-합성 서열이다. 서열번호 147에서, nt 1 내지 540은 CMV 프로모터 서열이고, nt 541 내지 1112는 N-말단 Flpo 코딩 영역이며, nt 1113 내지 1571은 합성 인트론이다. 서열번호 148에서, nt 1 내지 522는 CMV 프로모터 서열이고, nt 523 내지 904는 합성 인트론이며, nt 905 내지 1604는 C-말단 Flpo 코딩 영역이고, nt 1605 내지 1765는 비번역 폴리 A 영역이다.

서열번호 149 및 150은 예시적 저다양성 서열이다.

서열번호 151 및 152는 예시적 스플라이스 도너 컨센서스 서열이다.

서열번호 153은 예시적 HIV-2 키싱 루프 이합체화 도메인에 기반한 키싱 루프이다 (서열번호 141 및 142, 도 17b).

서열번호 154는 예시적 코작 강화 (Kozak enhanced) 개시 코돈이다.

서열번호 155 및 156은 인 비보에서 마우스 Otof 코딩 서열을 발현하기 위해 사용될 수 있는 예시적 구조체이다. 서열번호 155는 N-말단 Otof RNA를 생성하기 위해 사용된다. 서열번호 155는 nt 1-522의 인간 CMV 인핸서 및 프로모터, nt 523의 추정적 전사 개시 부위(putative transcription start site), 및 nt 4263-4311의 폴리 아데닐화 신호를 포함한다. 이는 하기와 같이 N-말단 Otof RNA 요소를 코딩한다: nt 523-546의 Kozak 서열을 포함하는 5' 비번역 영역; nt 547-4044의 5' Otoferlin 코딩 서열; nt 4045-4142의 5' 합성 인트론 서열; nt 4143-4254의 5' 삼중 키싱 루프 이합체화 도메인; 및 nt 4255-4262의 링커를 포함한다. 서열번호 155는 C-말단 Otof RNA를 생성하기 위해 사용된다. 서열번호 155는 nt 1-522의 인간 CMV 인핸서 및 프로모터, nt 523의 추정적 전사 개시 부위, 및 nt 3335-3467의 폴리 아데닐화 신호를 포함한다. 이는 하기와 같이 C-말단 Otof RNA 요소를 코딩한다: nt 525-636의 3' 삼중 키싱 루프 이합체화 도메인; nt 637-747의 3' 합성 인트론 서열; nt 748-3225의 3' Otoferlin 코딩 서열; nt 3226-3306의 C-말단 3xFlag 태그; 및 nt 3307-3334의 링커.

서열번호 157 및 158은 인 비보에서 인간 Myo7a(MYOSIN VIIA) 코딩 서열을 발현하기 위해 사용될 수 있는 예시적 구조체이다. 서열번호 157은 N-말단 Myo7a RNA를 생성하기 위해 사용된다. 서열번호 157은 nt 1-522의 인간 CMV 인핸서 및 프로모터, nt 523의 추정적 전사 개시 부위, 및 nt 4344-4392의 폴리 아데닐화 신호를 포함한다. 이는 하기와 같이 N-말단 Myo7A RNA 요소를 코딩한다: nt 523-543의 Kozak 서열을 포함한 5' 비번역 영역; nt 544-4125의 5' Myo7a 코딩 서열; nt 4126-4223의 5' 합성 인트론 서열; nt 4224-4335의 5' 삼중 키싱 루프 이합체화 도메인; 및 nt 4336-4343의 링커. 서열번호 158은 C-말단 Myo7a RNA를 생성하기 위해 사용된다. 서열번호 158은 nt 1-522의 인간 CMV 인핸서 및 프로모터, nt 523의 추정적 전사 개시 부위, 및 nt 3923-4055의 폴리 아데닐화 신호를 포함한다. 서열번호 158은 하기와 가티 C-말단 Myo7a RNA 요소를 코딩한다: nt 525-636의 3' 삼중 키싱 루프 이합체화 도메인; nt 637-747의 3' 합성 인트론 서열; nt 748-3813의 3' Myo7a 코딩 서열; nt 3814-3894의 C-말단 3xFlag 태그; 및 nt 3895-3922의 링커.

서열번호 159 및 160은 인 비보에서 전장 dCas9-VPR(enzymatically dead Cas9 fused to a VPR transcriptional activator domain) 코딩 서열을 발현하기 위해 사용될 수 있는 예시적 구조체이다. 서열번호 159는 N-말단 DCas9-VPR RNA을 생성하기 위해 사용된다. 서열번호 159는 nt 1-522의 인간 CMV 인핸서 및 프로모터, nt 523의 추정적 전사 개시 부위, 및 nt 4112-4161의 폴리 아데닐화 신호를 포함한다. 서열번호 159는 하기와 같이 N-말단 DCas9-VPR RNA 요소를 코딩한다: nt 523-543의 Kozak 서열을 포함한 5' 비번역 영역; nt 544-3894의 5' DCas9-VPR 코딩 서열; nt 3895-3992의 5' 합성 인트론 서열; nt 3993-4104의 5' 삼중 키싱 루프 이합체화 도메인; 및 nt 4105-4112의 링커. 서열번호 160은 C-말단 DCas9-VPR RNA를 생성하기 위해 사용된다. 서열번호 160은 nt 1-522의 인간 CMV 인핸서 및 프로모터, nt 523의 추정적 전사 개시 부위, 및 nt 3278-3410의 폴리 아데닐화 신호를 포함한다. 서열번호 160은 하기와 같이 C-말단 DCas9-VPR RNA 요소를 코딩한다: nt 525-636의 3' 삼중 키싱 루프 이합체화 도메인; nt 637-747의 3' 합성 인트론 서열; nt 748-3249의 3' DCas9-VPR 코딩 서열; 및 nt 3250-3277의 링커.

서열번호 161 및 162는 인 비보에서 전장 인간화 Cas9 Prime Editor (Prime Editor) 코딩 서열을 발현하기 위해 사용될 수 있는 예시적 구조체이다. 서열번호 161은 하기와 같이 N-말단 Prime Editor 서열을 코딩한다: nt 1-522의 인간 CMV 인핸서 및 프로모터; nt 523의 추정적 전사 개시 부위; nt 523-543의 Kozak 서열을 포함하는 5' 비번역 영역; nt 544-3894의 5' Prime Editor 코딩 서열; nt 3895-3992의 5' 합성 인트론 서열; nt 3993-4104의 5' 삼중 키싱 루프 이합체화 도메인; nt 4105-4112의 링커; nt 4112-4161의 폴리 아데닐화 신호. 서열번호 162는 하기와 같이 C-말단 Prime Editor 서열을 코딩한다: nt 1-522의 인간 CMV 인핸서 및 프로모터; nt 523의 추정적 전사 개시 부위; nt 525-636의 3' 삼중 키싱 루프 이합체화 도메인; nt 637-747의 3' 합성 인트론 서열; nt 748-3750의 3' Prime Editor 코딩 서열; nt 3751-3778의 링커; nt 3779-3911의 폴리 아데닐화 신호.

서열번호 163 및 164는 인 비보에서 전장 인간화 AncBE4(Cytosine Base Editor) 코딩 서열을 발현하기 위해 사용될 수 있는 예시적 구조체이다. 서열번호 163은 하기와 같이 N-말단 AncBE4 서열을 코딩한다: nt 1-522의 인간 CMV 인핸서 및 프로모터; nt 523의 추정적 전사 개시 부위; nt 523-540의 Kozak 서열을 포함하는 5' 비번역

영역; nt 541-2892의 5' AncBE4 코딩 서열; nt 2893-2990의 5' 합성 인트론 서열; nt 3103-3110의 5' 삼중 키싱 루프 이합체화 도메인; nt 3103-3110의 링커; nt 3111-3159의 폴리 아데닐화 신호. 서열번호 164는 하기와 같이 C-말단 AncBE4 서열을 코딩한다: nt 1-522의 인간 CMV 인핸서 및 프로모터; nt 523의 추정적 전사 개시 부위; nt 525-636의 3' 삼중 키싱 루프 이합체화 도메인; nt 637-747의 3' 합성 인트론 서열; nt 748-3957의 3' AncBE4 코딩 서열; nt 3958-3982의 링커; nt 3983-4115의 폴리 아데닐화 신호.

서열번호 **165** 및 **166**은 인 비보에서 전장 인간화 Abe8e(Adenine Base Editor) 코딩 서열을 발현하기 위해 사용될 수 있는 예시적 구조체이다. 서열번호 165는 하기와 같이 N-말단 Abe8e 서열을 코딩한다: nt 1-522의 인간 CMV 인핸서 및 프로모터; nt 523의 추정적 전사 개시 부위; nt 523-540의 Kozak 서열을 포함하는 5' 비번역 영역; nt 541-2706의 5' Abe8e 코딩 서열; nt 2707-2804의 5' 합성 인트론 서열; nt 2805-2916의 5' 삼중 키싱 루프 이합체화 도메인; nt 2917-2924의 링커; nt 2925-2973의 폴리 아데닐화 신호. 서열번호 166은 하기와 같이 C-말단 Abe8e 서열을 코딩한다: nt 1-522의 인간 CMV 인핸서 및 프로모터; nt 523의 추정적 전사 개시 부위; nt 525-636의 3' 삼중 키싱 루프 이합체화 도메인; nt 637-747의 3' 합성 인트론 서열; nt 748-3399의 3' Abe8e 코딩 서열; nt 3400-3427의 링커; nt 3428-3560의 폴리 아데닐화 신호.

서열번호 **167**은 예시적 키싱 루프 도메인 (GATTTTTGACCTGCTCGATTGTCCACTGCGAGCAGGTCTTTTGGAGTCGGGCGAGGCGGAAGCCCGACTCCTTTTGGCATGCACGCTAGCCGCGTCGTGCATGCCTTTTATC)이다.

서열번호 **168**은 예시적 ISE, M2 (GGGTTATGGGACC)이다.

서열번호 **169**는 예시적 ISE, cTNT (GGCTGAGGGAAGGACTGTCCTGGG)이다.

서열번호 **170**은 예시적 DISE, Rat FGFR2 (CTCTTTCTTTCCATGGGTTGGCCT)이다.

서열번호 **171** 및 **172**는 전장 YFP 코딩 서열을 발현하기 위해 사용될 수 있는 예시적 구조체이다. 서열번호 171은 하기와 같이 N-말단 YFP 서열을 코딩한다: nt 1-522의 인간 CMV 인핸서 및 프로모터; nt 523의 추정적 전사 개시 부위; nt 523-540의 Kozak 서열을 포함하는 5' 비번역 영역; nt 544-3654의 5' 스터퍼(Stuffer) ORF; nt 3655-3729의 자가 절단 2A 서열; nt 3730-4224의 5' YFP 세그먼트; nt 4225-4294의 5' 합성 인트론 서열 (가변); nt 4295-4406의 5' 삼중 키싱 루프 이합체화 도메인 (대문자); nt 4407-4414의 링커; nt 4415-4463의 폴리 아데닐화 신호. 서열번호 172는 하기와 같이 C-말단 YFP 서열을 코딩한다: 명칭: 3' 인트론 스크리닝 분할 YFP; 인간 CMV 인핸서 및 프로모터 nt 1-522; 추정적 전사 개시 부위 nt 523; 3' 삼중 키싱 루프 이합체화 도메인 nt 525-636; 3' 합성 인트론 서열 (가변) nt 637-706; 3' yfp 코딩 서열 nt 707-940; 자가 절단 2A 서열 nt 941-1006; 3' 스터퍼 ORF nt 1007-4228; 링커 nt 4229-4265; 폴리 아데닐화 신호 nt 4257-4388.

서열번호 **173-180**은 예시적 ISE 서열이다.

서열번호 **181**은 스크램블된 서열이다.

서열번호 **182-196**은 예시적 ISE 서열이다.

서열번호 **197-198**은 스크램블된 서열이다.

서열번호 **199-203**은 예시적 ISE 서열이다.

서열번호 **204**는 스크램블된 서열이다.

서열번호 **205**는 예시적 분지점 서열 (TACTAACA)이다.

서열번호 **206**은 예시적 폴리아데닐화 신호 AATAAAATATCTTTATTTTCATTACATCTGTGTGTTGGTTTTTTGTGTG 이다.

발명을 실시하기 위한 구체적인 내용

[0396] **실시예 1**

[0397] **합성 RNA 이합체화 및 재조합 도메인**

[0398] 도 1a는 벡터 설계 (좌측) 및 RNA 상호작용 및 스플라이싱 (우측)의 개략도이다. 좌측: 5' 트랜스-스플라이스 (trsp) DNA 벡터: 개방 화살표(open arrow)는 2개의 대향하는 프로모터이다. RFP 코딩 도메인 및 폴리 아데닐화 요소를 갖는 3' UTR이 SD (splice donor sequence), DISE (downstream intronic splicing enhancer), 및 2개의 ISE (intronic splicing enhancer) (2xISE), 결합 도메인 (BD, 이합체화 도메인으로도 지칭됨), 및 안정한

스템 루프(stem loop) BoxB 요소 (boxB), 자가-절단(self-cleaving) 해머헤드(hammerhead) 리보자임 (HHrz)으로 이어지고, 폴리아데닐화 요소를 포함하는 3' UTR로 종료되는, YFP의 N-말단 부분 (n-yfp)으로부터 반대편에 발현된다. n-yfp 세그먼트는 삽입된 작은 인트론 (n-yfp 내의 백색 세그먼트)을 갖는다. 3' trsp DNA 벡터: 개방 화살표는 2개의 대향하는 프로모터이다. BFP 코딩 도메인 및 폴리 아데닐화 요소를 갖는 3'UTR이, 상보적인 결합 도메인 (안티-BD, 이합체화 도메인으로도 지칭됨)으로부터 반대편에 발현되고, 상기 도메인은 3개의 ISE (intronic splicing enhancer sequence)(3xISE), 분지점 (BP), 폴리피리미딘 트랙트 (PPT), 스플라이스 억셉터 서열 (SA), YFP 코딩 서열의 c-말단 부분으로 이어지고, 폴리아데닐화 요소를 포함하는 3' UTR로 종료된다. 우측: YFP 단백질을 코딩하는 mRNA를 생성하기 위한 프리-mRNA (pre-mRNA) 상호작용 (5' trsp-RNA + 3' trsp-RNA) 및 트랜스-스플라이싱이 도시된다.

[0399] 도 1b는 N-말단 발현 플라스미드 단독의 형질감염은 YFP 형광을 초래하지 않는다는 것을 도시한다. 20k RFP+ 세포를 보여주는 유동 세포 분석.

[0400] 도 1c는 C-말단 발현 플라스미드 단독의 형질감염은 YFP 형광을 초래하지 않는다는 것을 도시한다. 20k BFP+ 세포를 보여주는 유동 세포 분석.

[0401] 도 1d는 결합 도메인이 없는, N-말단 및 C-말단 단편의 발현이 낮은 수준의 YFP 유도를 보인다는 것을 도시한다. 20k BFP+ 세포에 대한 레드 및 그린 형광값을 보여주는 유동 세포 분석.

[0402] 도 1e는 루프 형태(looped configuration)의 합리적으로 설계된 이합체화/결합 도메인을 도시한다. 피리미딘만 또는 퓨린만을 포함하는 저다양성 서열의 세그먼트가 안정한 스템 구조 사이에 존재한다. RNA 폴딩 예측은 결합 도메인과 그의 상보적 서열 간 염기쌍 형성을 위해 이용가능한 개방 서열의 6개의 구간 (1 내지 6으로 번호를 부여함)을 보여준다.

[0403] 도 1f는 개방 서열의 6개의 구간 (1 내지 6으로 번호를 부여함)을 보여주는, 개방 서열의 "루프형(looped)" 이합체화 도메인 형태의 3D 표현을 도시한다.

[0404] 도 1g는 C-말단 절반에 결합 도메인을 갖지 않는 음성 대조군을 도시한다. 20k BFP+ 세포에 대한 레드 및 그린 형광값을 보여주는 유동 세포 분석.

[0405] 도 1h는 N-말단 절반에 결합 도메인을 갖지 않는 음성 대조군을 도시한다. 20k BFP+ 세포에 대한 레드 및 그린 형광값을 보여주는 유동 세포 분석.

[0406] 도 1i는 N-말단 절반 및 C-말단 절반 모두에서 매칭되는 결합 도메인이 세포의 90%에서 강한 YFP 유도를 보인다는 것을 도시한다. 20k BFP+ 세포에 대한 레드 및 그린 형광값을 보여주는 유동 세포 분석.

[0407] 도 1j-1n은 완전히 열린 형태(fully open configuration)를 초래한, 150개 저다양성 피리미딘만으로 또는 퓨린만 포함하는 서열의 구간을 갖는 결합 도메인의 형태에 대한 도 1e-1i에 도시된 데이터에 균등한 데이터를 도시한다.

[0408] 도 1o는 도 1g에서 표시된 세포에 대한 대표적인 형광 이미지를 보여준다.

[0409] 도 1p는 도 1l에서 표시된 세포에 대한 대표적인 형광 이미지를 보여준다.

[0410] 도 1q는 도 1d, 도 1g 내지 1i, 및 도 1l 내지 1n에 표시된 조건의 비교를 도시한다. YFP 유도 계수(induction coefficient)가 계산된다: (#R+Y+ ÷ #R+Y-) x 100 x med.Y-fluor(R+Y+). 비교를 위해, N-말단 상의 원형 인트론(마우스 파르브알부민 유전자의 인트론 I)과 C-말단 단편 상의 그 인트론에 대한 최적화된 결합 도메인의 재조합 효율이 표시된다 (백색 막대). 이는 최적화된 합성 DNA 이합체화 및 재조합 도메인의 유익을 예시한다.

[0411] **실시예 2**

[0412] **3개의 합성 단편으로부터의 단백질의 재구성**

[0413] 도 2a는 벡터 설계의 개략도를 보여준다. YFP의 단백질 코딩 서열이 N-말단, 중간 단편 (m-yfp) 및 C-말단 단편으로 분할된다. n 및 m 단편의 연결부가 루프형 설계 결합 도메인 (BD1)에 의해 연결되고, m 단편과 c 단편 사이의 연결부가 루프형 결합 도메인 (BD2)에 의해 연결된다. 피리미딘 (Y) 및 퓨린 (R) 서열이 m-단편의 자가-고리화(self-circularization)를 방지하고 N-단편과 C-단편의 직접적인 재조합을 방지하도록 배열된다. N-말단 단편은 형질감염 대조군으로서 RFP (red fluorescent protein)와 공-발현되고, C-말단 단편은 형질감염 대조군으로서 BFP (blue fluorescent protein)와 공-발현된다.

[0414] 도 2b는 모든 3개의 단편의 매칭 결합 도메인이 세포의 80%에서 강한 YFP 유도를 보인다는 것을 도시한다. 20k BFP+ 세포에 대한 레드 및 그린 형광값을 보여주는 유동 세포 분석.

[0415] 도 2c는 n 및 m 단편만의 발현의 대표적인 형광 이미지가 yfp 형광을 보이지 않는다 (음성 대조군)는 것을 도시한다.

[0416] 도 2d는 m 및 c 단편만의 발현의 대표적인 형광 이미지가 yfp 형광을 보이지 않는다 (음성 대조군)는 것을 도시한다.

[0417] 도 2e는 3개의 단편 모두의 공-형질감염에 의해 강한 YFP 형광이 유도된다는 것을 보여주는 대표적인 형광 이미지를 도시한다.

[0418] **실시예 3**

[0419] **2개의 부분으로 분할된, 재구성된 전장 YFP의 인 비보 전달**

[0420] 2개의 단편으로부터의 YFP 코딩 서열의 재구성을, 하나는 YFP의 n-말단 코딩 절반 단편을 포함하고, 하나는 c-말단 코딩 절반 단편을 포함하는 것인 2개의 합성 RNA 서열을 이용하여 달성한다 (도 3a) (서열번호 1 및 2). 신생 (P3) 마우스 새끼에서 전신 (iv) 투여 후, 각 단편을 AAV2/8로부터 발현시켰다. 마우스당 2개의 단편 각각에 대해 총 1.88E11 바이러스 게놈을 투여했다. 형광 현미경을 이용하여, 3주 뒤에 간, 심장 근육, 및 골격 근육에서 YFP의 발현을 검출했다.

[0421] 도 3b에 도시된 바와 같이, 미성년 마우스(juvenile mouse)의 간에서 전장 YFP의 발현이 검출되었으나, 주사되지 않았던 마우스의 간(uninjected liver)은 YFP 발현을 보이지 않았다.

[0422] 도 3c에 도시된 바와 같이, 미성년 마우스의 심장 근육에서 전장 YFP의 발현이 검출되었으나, 주사되지 않았던 마우스의 심장 근육은 YFP 발현을 보이지 않았다.

[0423] 도 3d에 도시된 바와 같이, 다리의 골격근에서 전장 YFP의 발현이 검출되었으나, 주사되지 않았던 마우스의 간은 YFP 발현을 보이지 않았다.

[0424] 따라서, 본 명세서에서 개시되는 시스템은 2개 이상의 개별적인 합성 RNA 분자로부터, 인 비보로 전장 단백질을 발현하기 위해 이용될 수 있다.

[0425] **실시예 4**

[0426] **3개의 부분으로 분할된, 재구성된 전장 YFP의 인 비보 전달**

[0427] 3개의 단편으로부터의 YFP 코딩 서열의 재구성을, 하나는 YFP의 n-말단 단편을 포함하고, 하나는 YFP의 중간 단편을 포함하며, 하나는 c-말단 단편을 포함하는 것인 3개의 합성 RNA 서열을 이용하여 달성한다 (도 4a) (각각 서열번호 145, 146, 및 2).

[0428] 각 단편이 신생 (P3) 마우스 새끼의 전경근으로의 근육내 주사 후 AAV2/8로부터 발현되었다. 각각의 단편에 대해 총 1E11 바이러스 게놈을 근육내로 투여했다. 형광 현미경을 이용하여, 3주 뒤에 골격 근육에서 YFP의 발현을 검출했다.

[0429] 도 4b에 도시된 바와 같이, 전장 YFP 형광의 발현이 전경근에서 관찰되었다.

[0430] 따라서, 상기 개시되는 시스템은 3개 이상의 개별적인 합성 RNA 분자로부터, 인 비보로 전장 단백질을 발현하기 위해 이용될 수 있다.

[0431] **실시예 5**

[0432] **재구성된 전장 단백질의 인 비보 전달**

[0433] 인 비보에서 3-파트 sRdR 시스템의 가능성을 입증하기 위해, YFP의 단편을 포함하는 2개 또는 3개의 AAV-전달 플라스미드 (AAV의 DNA 전구체 플라스미드)의 조합을 성체 마우스의 전경 (TA) 후지(hindlimb)근에 경피로 전기천공시켰다. 근육내 전기천공 후 5일차에 2-파트 분할-YFP 시스템 및 3-파트 분할 YFP 시스템 모두의 효율적인 재구성을 관찰했다 (도 5a-5f).

[0434] 도 5a-5f는 성체 마우스 전경근에서 2개의 단편 및 3개의 단편으로부터의 YFP의 효율적인 재구성을 도시한다. 도 5a는 YFP 코딩 서열의 N-말단 절반 및 C-말단 절반이 합성 RNA-이합체화 및 재조합 도메인을 갖는다는 것을

도시한다. 도 5b는 이러한 2개의 단편을 발현하는 2종의 AAV 전달 플라스미드를 성체 마우스 TA(tibialis anterior) 근육 내로 경피로 전기천공시키고, 전기천공 후 5일차에 강한 형광이 검출되었다는 것을 보여준다. 도 5c는 반대쪽 비-주사(contralateral non-injected) TA에서 형광이 검출되지 않았다는 것을 보여준다. 도 5d는 N-말단, 중간, 및 C-말단 YFP 코딩 서열이 각 단편을 그의 인접한 단편에 연결시키는 합성 RNA-이합체화 및 재조합 도메인을 갖는다는 것을 보여준다. 도 5e는 이러한 3개의 단편을 발현하는 3종의 AAV 전달 플라스미드의 경피 전기천공을 도시한다. 강한 YFP 형광이 검출되어, 3개의 단편으로부터 YFP의 효율적인 재구성을 나타낸다. 5f는 반대쪽 비-주사 TA에서의 형광을 도시한다. 형광 채널(fluorescent channel)이 상황(context)을 위해 그레이 스케일 사진 위에 오버레이된다.

[0435] 2개 또는 3개의 벡터를 이용하여 간, 심장 근육 및 골격 근육 (2개의 AAV 벡터), 및 골격 근육 (3개의 AAV 벡터)에서 YFP를 성공적으로 발현시켰다.

[0436] 따라서, 본 명세서에서 제공되는 합성 RNA-이합체화 및 재조합 시스템이 근육에서 사용될 수 있다. 이러한 결과에 근거하여, 원하는 개체 및/또는 조직으로의 AAV로부터의 치료적 전장 디스트로핀 (또는 기타 유전자) 발현을 달성하기 위해 YFP 코딩 서열을 디스트로핀 (또는 기타 유전자) 코딩 서열로 치환할 수 있다.

[0437] **실시예 6**

[0438] **DMD를 치료하기 위한 재구성된 전장 디스트로핀의 전달**

[0439] 듀센 근이영양증 (DMD)을 앓는 환자를 위한 전장 디스트로핀을 이용한 효과적인 유전자 치료법은, 이 큰 단백질의 코딩 서열이 대부분의 바이러스 벡터의 수용력을 초과하므로, 과제로 남아있다. AAV(Adeno-associated viruses)는 유전자 대체 요법에서 일반적이고 선호되는 유전자 전달 방법이다. AAV는 무독성이고 잘 관용되고(well tolerated), 게놈 내로의 랜덤 통합 없이 대체 유전자의 장기적 발현을 가져온다. 그러나, 디스트로핀 유전자는 너무 커서 단일 바이러스에 의해 전달될 수 없다. 단편으로 세분되면, 전장 디스트로핀은 최소 3개의 바이러스를 이용하여서만 전달될 수 있다.

[0440] "마이크로-디스트로핀(micro-Dystrophin)" 또는 "미니-디스트로핀(mini-Dystrophin)"으로 불리는, 디스트로핀의 더 작은 버전이 현재 디스트로핀 유전자 대체 요법을 위해 테스트되고 있으나, 이러한 디스트로핀의 절단된(truncated) 버전은 디스트로핀 단백질의 로드 및 힌지 섹션에서 주요한 도메인이 결여되므로, 이들이 완전한 기능성을 가질 것으로 기대되지 않는다. 현재까지, 이러한 한계를 극복하기 위한 과거의 시도들이 DMD의 치료를 위해 요구되는 효율을 가져오지 못했다.

[0441] 복수의 연속된 단편들로부터 디스트로핀을 포함한, 큰 유전자의 코딩 서열을 효율적으로 재구성하기 위해 이용될 수 있는 신규한 기술이 본 명세서에 제공된다. 전달 벡터로서 AAV와 함께 이 기술을 이용하여, 전장 디스트로핀이 DMD를 위한 마우스 모델 (및 돼지 및 개 모델)에서 발현될 것이다. 일 구체예에서, 개체는 DMD를 갖는 인간 개체, 청소년, 또는 유아이다. 예를 들면, 본 명세서에서에서 개시되는 방법 및 시스템이 2개 또는 3개의 AAV를 통해 전장 디스트로핀을 코딩하는 합성 RNA-이합체화 및 재조합 도메인을 전달하기 위해 이용될 수 있다 (예를 들면, 각각의 AAV가 전장 코딩 서열의 절반 또는 1/3을 전달함). 일 구체예에서, AAV는 근육친화성(myotropic) AAV이다 (예를 들면, 우선적으로 근육을 감염시키는 AAV)이다. 이 접근방식은 DMD를 위한 마우스 또는 개 모델, 및 인간 개체에서 이영양증 증상을 개선하거나 또는 발생을 예방하기 위해 이용될 수 있다.

[0442] 파트 1: 효율적으로 재구성된 삼원 분할(three-way split) 발현 카세트를 작제한다. 인 비트로에서 전장 디스트로핀 코딩 서열을 효율적으로 재구성하고, 각각의 개별적인 카세트는 통상적인 AAV 벡터의 패키징 한계 내에 속하는 것인 3개의 발현 카세트를 작제한다. 디스트로핀의 치료 유효 수준을 달성하기 위해, 대략적으로 디스트로핀의 생리적 수준, 또는 적당하게 초생리적(supraphysiological) 수준을 달성하도록 발현 시스템을 최적화시킬 수 있다. 디스트로핀의 최대 50배 과발현은 유해한 효과없이 관용된다. 디스트로핀 코딩 서열은 그의 길이를 따라 다수의 상이한 지점에서 분할될 수 있다. 그러나, 재구성의 효율은 국소 RNA 미세환경에 의해 영향받고, 여러 가능한 분할 지점(split point)의 효율을 비교하는 것에 의해 재구성 효율의 최대화가 경험적으로 이루어진다. 천연 디스트로핀 코딩 서열을 최적 발현을 위해 코돈 최적화시키고, 최대 재구성 효율을 수용하도록 변형시킬 수 있다. 본 명세서에서 개시된 합성 RNA-이합체화 및 재조합 접근방법을 이용하여 삼원 분할 전구체(three-way split precursor)로부터 전장 디스트로핀 코딩 서열이 재구성될 수 있을 것으로 예상된다. 상이한 구성(구조)들의 스크리닝에서, 디스트로핀의 가장 효율적인 재구성(예를 들면, 대략적으로 생리적 수준 또는 적절하게 초생리적 수준)을 가져오는 3개의 발현 카세트의 세트를 선택한다. 실험은 HEK293T 또는 인간 골격 근육 세포(HSkMC, 일차 또는 트랜스-분화(trans-differentiated))에서 수행할 수 있다. 내생 vs. 외래 특이적 정량적

RT-PCR 프로브를 이용하여, 및 외래 디스트로핀 단백질 중 에피토프 태그 검출 및 웨스턴 블롯 분석에 의해, 분할/재구성 디스트로핀의 상이한 구성에 대해 재구성 효율을 결정할 것이다.

[0443] 파트 2: 비-재구성 단편 대비 전장 디스트로핀 발현을 최대화한다. 합성 RNA-이합체화 및 재조합 도메인의 변형에 의해 비-재구성 디스트로핀(non-reconstituted dystrophin)의 단편화 백그라운드 발현의 억제가 달성될 수 있다. RNA 재조합에서 비효율에 의해 유발된 비-재구성 단편 발현이 디스트로핀 단편의 백그라운드 발현을 초래할 수 있다. 또한, 이러한 단편화 백그라운드 발현의 억제가 합성 RNA-이합체화 및 재조합 도메인의 변형에 의해 달성될 수 있다. 본 명세서에서 개시되는 방식을 이용하여, 디스트로핀의 각각의 단편이 개별적으로 전사된다. 재구성은 RNA 수준에서 일어난다. 따라서, 각각의 개별적인 단편이 잠재적으로 재구성되지 않으면서, 번역될 수 있다. 웨스턴 블롯에서, 전장 디스트로핀은 약 430kDa에 이르나, 이러한 단편들은 전장 디스트로핀의 약 2/3 (~290kDa) 및 1/3 (~140kDa)의 크기일 것이다. 비-재구성 단편 발현을 방지하고, 디스트로핀의 전장 발현에 유리하도록 상기 합성 RNA-이합체화 및 재조합 도메인을 최적화시킬 수 있다. 이는 예를 들면, 전략적으로 데그론 서열을 배치하고, 비-재조합 단편의 RNA 핵 수송을 중단시키고, 데코이 번역 개시 지점(decoy translation initiation points)을 도입하는 것에 의해 달성될 수 있다. 실험은 HEK293T 및 HSkMC에서 수행된다. 웨스턴 블롯 분석을 이용하여 디스트로핀의 완전히 재구성되지 않은 단편들의 식별 및 정량을 가능하게 하는 에피토프 태그로 디스트로핀 코딩 서열을 표지할 수 있다(bookend). 인간 골격 근육 세포에서 이러한 디스트로핀 단편의 세포내 분포를 면역조직화학을 이용하여 평가할 것이다. 추가적으로, RNA 수준에서의 재구성의 효율을 결정하기 위해 이용될 재조합 연결부를 포괄하는 정량적 RT PCR을 포함한, 통상적인 분자 생물학 기법을 이용하여, 억제의 정량적 평가가 수행될 것이다. 단편화 디스트로핀 발현의 낮은 수준이 관찰될 것으로 예상된다. 합성 RNA-이합체화 및 재조합 도메인을 변형시키는 것에 의해, 이러한 단편들이 억제될 수 있다.

[0444] 파트 3. 인 비트로 및 인 비보 발현을 위한 전장 디스트로핀 모듈의 고-역가 AAV 스톡을 생성한다. 디스트로핀 발현 AAV를 고 순도 및 3E13 GC/ml 보다 높은 바이러스 게놈 카운트로 생산할 것이다. 3종의 근육친화성 AAV 혈청형을 생산할 것이다: AAV2/8, AAV2/9, 및 AAV2/rh10. 3원(tripartite) 분할 형광 단백질, 에피토프 태그로 표지된 전장 디스트로핀의 3원 분할 (전술된 파트 2 참조), 및 비-표지(non-tagged) 전장 디스트로핀의 3원 분할을 생산하여, 27개의 고-역가 AAV 제제를 준비할 것이다. 치료용 AAV 입자의 전신 전달은 고농도 대형 바이러스 제제(high concentration large virus preparations)를 요구한다. 3개의 별개의 바이러스로부터 디스트로핀의 재구성 발현을 달성하기 위해, 바이러스의 반복 투여가 수행될 수 있다. HEK293T 세포에서의 AAV 생산. 이오딕사놀(Idixanol) 또는 CsCl 정제. 모든 배치를 인 비트로에서 HEK293T 및 인간 골격 근육 세포에서 테스트할 것이다. 파트 1 및 파트 2에서 약술된 바와 같이, 재구성 효율 및 원치않는 단편 발현을 평가할 것이다.

[0445] 파트 4. 인 비보 FLD-AAV 모듈의 발현/재구성 수준 및 전장 디스트로핀 발현 AAV 모듈의 인 비보 조직 분포를 측정한다. 대리 지표(surrogate indicator)로서 3원 분할 형광 단백질에 대해서도 동일한 것을 평가한다. 인 비보 전달에 대해, 신생 및 미성년 마우스에서 직접적인 근육내 (심장 근육 및 골격 근육) 및 전신 정맥내 전달이 비교될 것이다. FLD-AAV의 직접적인 근육 주사는 전술된 실시예에서 나타난 바와 같이 전장 디스트로핀의 효율적인 발현을 가져올 수 있다. FLD-AAV의 전신 전달은 면역조직화학 및 웨스턴 블롯 분석을 이용하여 조사될 것이다. 신생 및 미성년 마우스에서, 직접적인 근육내 전달 및 전신 정맥내 전달을 포함한 다른 투여 경로가 비교될 것이다. 분석은 하기에 중점을 둘 것이다: (1) 골격 근육 (skeletal muscles (대(major) 전지, 후지, 어깨, 복부, 및 안면 근육) 및 빠른 vs. 느린 경련 근육(fast vs. slow twitch muscles)의 차등적 감염성(differential infectivity)을 전경골근과 비장근을 비교하는 것에 의해 평가함, (2) 심장 근육 발현, 및 (3) 간 발현. 고-역가 AAV 주사의 가능한 부작용에 대해 이 코호트의 동물을 모니터링할 것이다.

[0446] AAV의 직접적인 근육 주사가 FLD-AAV 모듈을 전달하는 접근 방식을 나타내나 (도 5a-5f에서의 결과를 고려할 때 성공적일 것임), 그럼에도 불구하고, 상기 바이러스의 전신 i.v. 전달을 이용하여 전장 디스트로핀 발현을 달성하는 것이 임상적 관점에서 바람직하다. 인 비트로 FLD-AAV 테스트를 이용하여 AAV 카피 수와 재구성된 디스트로핀 수준이 어떻게 상관되는지를 결정할 것이다. 조직 분포 및 재구성의 효율을 인 비보에서 평가하고, 최적 조직 분포를 달성하기 위해 상이한 전달 패러다임 (예를 들면, 혈청형, 바이러스 역가, 적용 경로, 반복 적용의 횟수)을 조사할 것이다. 조직 커버리지(tissue coverage) 및 발현 수준을 평가할 것이다. 근육 섬유의 일부만 디스트로핀을 발현하더라도 유익한 결과가 달성될 수 있다 (예를 들면, 비-스트레스 조건 하에서 약 50%의 심근 세포만 디스트로핀 결함인 경우 정상적인 심장 기능). 디스트로핀의 생리적 및 초생리적 수준이 모두 치료적으로 유용하다. 파트 1 & 2에 약술된 바와 같이 정량적 평가를 수행할 것이다. 무균 조건 하에 신생 또는 미성년 마우스에서 인 비보 근육내 및 전신 바이러스 적용을 수행할 것이다.

[0447] 파트 5. DMD 마우스 모델 (mdx)을 FLD-AAV로 처리하고 질병 발병/진행을 평가한다. 신생 mdx 마우스에서 FLD-

AAV 전달은 근육병증 및 심근병증의 발병 및 진행을 예방할 수 있다. 재구성된 전장 디스트로핀의 바이러스 전달의 최적화 후 (파트 1-4), FLD-AAV 치료를 DMD의 마우스 모델에 투여할 것이다. 그들이 사육된 유전적 배경에 따라, 이러한 마우스들은 인간 DMD에서보다 현저하게 덜 발현되는 근육병증을 갖는다. 보다 심각한 표현형을 제공하는 유전적 배경을 갖는 마우스 (D2.B10-Dmdmdx)는 증가된 후지 약화(hind-limb weakness), 더 낮은 근육 중량, 더 적은 근섬유, 및 증가된 지방 및 섬유증을 보인다. 이러한 파라미터를 야생형 대조군, 처리된 mdx, 및 미처리 mdx 마우스 간에 비교할 수 있다. 원하는 결과는 질병 발병/진행의 개선 또는 예방이다.

[0448] 디스트로핀 유전자 중 돌연변이를 갖는 2개의 마우스 계통, C57BL/10ScSn-Dmdmdx/J, 및 D2.B10-Dmdmdx/J를 이용한다. 파트 4에서 기술된 바와 같이 확립된 파라미터에 따라 FLD-AAV를 전달한다. mdx 마우스에서 근괴사의 발생 전 시간 범위(time window)에, 생후 1주차에 동물에 주사한다. 골격 및 심장 근육병증의 행동 및 해부학적 징후에 대해 야생형, 처리된-mdx 및 비히클/모조(sham)-처리-mdx 마우스를 평가한다. 운동학 및 근전도 검사 테스트 장비를 이용하여, 다양한 운동 작업, 예를 들면, 평균대, 악력, 수평 사다리(horizontal ladder), 트레드밀 속도 검사(treadmill speed challenge), 지상 보행 운동학적 평가(over ground locomotor kinematic assessment), 수영 운동학적 평가 (주변 온도 및 냉수 검사)에서 이들 마우스의 성능을 평가한다. 화합물 접종(chemical challenge) 후 mdx 마우스에서 FLD-AAV 요법이 심근병증의 발현을 예방할 수 있는지 여부가 결정될 것이다.

[0449] 이러한 실험의 원하는 결과는 질병 발병/진행의 개선 또는 예방일 것이다.

[0450] **실시예 7**

[0451] **어셔 증후군을 치료하기 위한 재구성된 전장 MYO7A의 전달**

[0452] MYO7A 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. MYO7A의 제2 절반을 상보적인 합성 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, MYO7A의 절반 2개는 재조합되어 전장 MYO7A 전사물을 형성하고, 단백질로 번역된다.

[0453] **실시예 8**

[0454] **전사/발현 로직 게이트(Transcriptional/expressional logic gate)**

[0455] 표적 유전자를 2개의 비기능성(nonfunctional) 절반으로 세분하고, 2개의 상이한 프로모터로부터 또는 2개의 상이한 전달 비히클을 이용하여 발현되게 하면 교차점 발현 패턴(intersectional expression pattern)을 초래할 수 있다.

[0456] 예를 들면, 본 명세서에서 제공되는 제1 합성 핵산 분자의 프로모터 1이 예를 들면, 세포 타입 A, B, 및 C에서 코딩 서열의 N-말단 절반의 번역을 구동시키고, 본 명세서에서 제공되는 제2 합성 핵산 분자의 프로모터 2는 세포 A, D, E, 및 F의 서브세트에서 C-말단 절반의 발현을 구동시킬 수 있다. 그러한 구체예에서, 표적 단백질을 코딩하는 이펙터 유전자는 중첩 영역 (이 실시예에서, 세포 집단 A)에서만 발현된다.

[0457] 유사한 교차성(intersectionality)이 2개의 절반을 조건적으로, 예를 들면, 재조합 효소의 존재의 조건 하에서, 발현되게 하는 것에 의해, 이용될 수 있다. 교차성이 달성될 수 있는 또 다른 수준은 2개의 절반을 상이한 친화성(tropism)을 갖는 2개의 바이러스로 전달하는 것에 의한다.

[0458] **실시예 9**

[0459] **상보성(complementation)**

[0460] 본 명세서에서 개시되는 방법 및 시스템은 2개의 비-기능성 절반을 2개의 플라스미드가 모두 존재하는 경우에만 활성이 될 수 있는 별개의 플라스미드에 코딩시키는 것에 의해 임의의 유전자 (및 상응하는 표적 단백질)를 상보성 파트로 만들기 위해 이용될 수 있다(LacZ의 알파 상보성의 원칙과 유사함).

[0461] **실시예 10**

[0462] **트리거 RNA**

[0463] 본 명세서에서 개시되는 시스템 및 방법은 표적 단백질의 코딩 서열의 2개 이상의 부분의 재구성이 특정한 "트리거(trigger)" RNA 분자의 존재에 의존적이도록 구성될 수 있다. 도 7b에 도시된 바와 같이, 이 구체예에서, 각각의 합성 핵산 분자의 이합체화 도메인은 상호 간에 역 상보체가 아니나, 대신에 제3 RNA 분자, "트리거

RNA"의 인접 영역에 특이적으로 혼성화되고, 트리거 RNA는 2개의 합성 핵산 분자들을 모으는 브릿지로 작용한다. 이 구체예에서, 상기 시스템은 리포터/이펙터 단백질의 "세포 타입 특이적 유발(triggering)"을 가능하게 하는 특정한 RNA 분자의 존재를 "보고"할 수 있다.

[0464] **실시예 11**

[0465] **3'-UTR 중 안정화 요소의 포함**

[0466] 본 실시예는 RNA를 안정화시키는 3'-UTR 중 서열의 존재에서 분할 코딩 서열(split coding sequence)의 재조합을 평가하기 위해 이용되는 방법을 기술한다. WPRE3 (Woodchuck hepatitis posttranscriptional regulatory element 3)을 예시적인 안정화 서열로 이용했다. 당업자는 다른 RNA 서열 안정화제가 WPRE3 대신에 이용될 수 있다는 것을 이해할 것이다.

[0467] YFP 형광 중앙값을 본 명세서에서 개시된 합성 RNA 이합체화 및 재조합 방식을 이용하여 재구성되는 2원 분할 YFP에 대한 유동 세포측정법에 의해 측정했다. C-말단 YFP 코딩 단편이 폴리 아데닐화 신호 단독 (w/o WPRE3) 또는 WPRE의 절단 버전, WPRE3 및 뒤이은 폴리 아데닐화 신호 (w/WPRE3로 표시됨)로 이어진다. N-말단 YFP 코딩 단편은 형질감염 대조군으로서 양방향성 프로모터로부터 RFP와 공발현된다. C-말단 단편은 형질감염 대조군으로서 양방향성 프로모터로부터 BFP와 공발현된다. 조건들 간에 동일한 레드 및 블루 형광 대조군 값을 갖는 세포들을 비교한다.

[0468] 도 8에 도시된 바와 같이, 3'-UTR에 안정화 요소를 포함시키는 것은 재조합된 전장 YFP의 발현 효율을 약 50-60% 증가시켰다. 이러한 증진은 WPRE 서열이 그들이 포함되어 있는 RNA 분자의 핵 수송을 촉진하는 경우에도 관찰되고, 이는 스플라이세오솜 매개 RNA 연결이 일어날 수 있기 전에 핵의 외부에서 도 6a의 분자 (150)를 셔틀링시켜서 RNA 연결 반응 (및 따라서, 유전자 발현)에 부정적으로 영향을 미쳐서 비-기능성이 되게 했을 수도 있다.

[0469] 따라서, 본 명세서에서 개시되는 합성 분자 (예를 들면, 서열번호 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 145, 146, 147, 148, 155, 156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 및 166)는 RNA 서열 안정화제를 더 포함하도록 변형될 수 있다.

[0470] **실시예 12**

[0471] **재구성 효율에 대한 결합 도메인 길이의 효과**

[0472] 결합 도메인 길이를 하기와 같이 평가했다. YFP를 2개의 비-형광 절반으로 분할했다(서열번호 1 및 2, 그러나 각각 상이한 길이의 결합 도메인을 가짐). 상이한 길이의 결합 도메인(50 내지 500 nt 범위)에 대한 재구성 효율을 배양된 HEK 293t 세포에서 평가했다. RFP (Red Fluorescent Protein)를 형질감염 대조군으로 하여, 양방향성 CMV 프로모터로부터 N-말단 YFP를 발현시킨다. BFP (Blue Fluorescent Protein)를 형질감염 대조군으로 하여, 양방향성 CMV 프로모터로부터 C-말단 YFP를 발현시킨다. 상이한 결합 도메인 길이에 대해, YFP 형광 강도 중앙값을 비교했다. 매칭되는 RFP 및 BFP 형질감염을 갖는 세포를 조건들 간에 비교한다.

[0473] 도 11에 도시된 바와 같이, 모든 분자들이 다양한 정도의 재구성 효율로 일정(some) 수준의 전장 YFP의 발현을 달성했다. 최대 성능은 150 bp 이하(예를 들면, 50-150 bp)의 결합 도메인 길이에서 관찰되었으나, 최대 500 bp의 결합 도메인은 여전히 전장 YFP를 재조합하고 발현시킬 수 있었다.

[0474] **실시예 13**

[0475] **스플라이싱 인핸서 서열의 효과**

[0476] 본 실시예는 본 명세서에서 개시되는 합성 인트론에 하나 이상의 ISE 서열 (예를 들면, 도 6a의 118, 120, 156)을 포함시키는 것의 효과를 평가하기 위해 이용되는 방법을 기술한다.

[0477] YFP를 2개의 비-형광 절반으로 분할했다 (도 12a). 상이한 인트론 구성에 대한 재구성 효율을 배양된 HEK 293t 세포에서 평가했다. RFP (Red Fluorescent Protein)를 형질감염 대조군으로 하여, 양방향성 CMV 프로모터로부터 N-말단 YFP를 발현시켰다. BFP (Blue Fluorescent Protein)를 형질감염 대조군으로 하여, 양방향성 CMV 프로모터로부터 C-말단 YFP를 발현시켰다. 상이한 인트론 구성에 대해, YFP 형광 강도 중앙값을 비교한다. 매칭되는 RFP 및 BFP 형질감염을 갖는 세포를 조건들 간에 비교한다.

[0478] 도 12a에 도시된 바와 같이, 5' 분자 (서열번호 1)는 YFP의 N-말단 부분의 코딩 영역 (n-yfp), 뒤이은 SD

(splice donor sequence), DISE (downstream intronic splicing enhancer), 및 2개의 ISE(intronic splicing enhancers) (2xISE), BD(binding domain), 자가-절단 HHrz(hammerhead ribozyme)을 포함하고, pA(poly adenylation signal)로 종료된다. 3' 분자 (서열번호 2)는 상보적 결합 도메인 (안티-BD), 뒤이은 3개의 ISE (3xISE), BP (branch point), PPT (polypyrimidine tract), SA (splice acceptor sequence), YFP 코딩 서열의 c-말단 부분을 포함하고, pA로 종료된다.

[0479] 도 12b에 도시된 바와 같이, 5' 및 3' 분자 모두에 스플라이스 인핸서의 포함은 전장 YFP의 재구성 효율을 증가시킨다. 스플라이스 인핸서의 제거는 2개의 코딩 서열의 재구성 효율을 약 50-90% 감소시킨다. 제1 열에서, YFP는 기준 구성 (서열번호 1 및 2)을 이용하여 재구성되고, 제2 열은 5' 단편에서 ISE 요소의 결실에 따른 재구성 효율을 보여주고, 제3 열은 5' 단편에서 ISE 및 DISE의 결실 후 재구성 효율을 보여준다. 제4 열은 5' 단편에서 HHrz의 결실 후 재구성 효율을 보여준다. 제5 열은 기준 구성을 이용한 재구성 효율을 보여준다. 제6 열은 3' 단편에서 ISE 요소의 결실 후 재구성 효율을 보여준다. 제7 열은 5' 및 3' 단편 모두에서 ISE 요소의 결실 및 5' 단편에서 DISE의 결실 후 재구성 효율을 보여준다.

[0480] **실시예 14**

[0481] **이중 투영 트레이싱(Dual Projection Tracing)**

[0482] 본 실시예는 2개의 단편 (서열번호 147 및 148)으로부터 전장 flp 재조합효소 (Flpo)의 재구성에 의해 이중 투영 트레이싱을 수행하기 위해 이용되는 방법을 기술한다. 도 13a에 도시된 바와 같이, Flp 재조합효소 유전자를 2개의 비-기능성 절반으로 분할했다. Flpo 유전자의 N-말단 절반을 3' 말단에서 이합체화 도메인 서열 (RNA 말단 연결 모듈, REJ)로 이어지는 합성 인트론 서열과 연결시켰다. Flpo 유전자의 C-말단 절반을 그의 5' 말단에서 합성 인트론 및 이합체화 도메인 (REJ-모듈)과 연결시켰다. 두 구조체에 의한 세포의 감염, 및 각 구조체로부터의 프리-mRNA의 발현 시, 프리-mRNA는 이합체화 도메인(도 13a에서 흑색 평행 막대로 표시됨)에 결합되고, 결과적으로 수득된 복합체가 스플라이싱되어 전장 Flpo 재조합효소 mRNA 전사물을 생성했다. 따라서, 상기 2개의 단편으로부터 기능성 재조합효소 단백질이 생성되었다. 도 13b는 flpo 의존성 레드 형광 단백질 (RFP) (tdTomato)을 운반하는 flp 활성 리포터 마우스의 개략도를 보여준다 (Rosa-CAG-frt-STOP-frt-tdTomato). 2개의 합성 핵산 (DNA) 구조체를 별개의 AAV 내로 패키징시켰다 (역행으로 수송되는 혈청형 AAV2/retro). 제1 구조체를 운반하는 바이러스, AAV2/retro-n-flpo를 마우스의 좌측 일차 운동 피질에 주사하고, 제2 구조체를 운반하는 바이러스, AAV2/retro-c-flpo를 마우스의 우측 일차 운동 피질에 주사했다.

[0483] 도 13c 및 13d에 도시된 바와 같이, 정중선을 교차하는 축삭돌기(axon)를 갖는 일차 운동 피질 세포가 RFP로 표지된다 (도 13c 및 13d에서 백색으로 나타남). Hoechst 염색 (핵)이 상황을 위해 표시된다.

[0484] **실시예 15**

[0485] **긴 단백질의 인 비보 발현**

[0486] 본 실시예는 세포 배양 및 인 비보에서 마우스 일차 운동 피질에서 거대크기(oversized) 카고의 효율적인 발현을 달성하기 위해 이용되는 방법을 기술한다.

[0487] 2개의 바이러스의 AAV 카고 수용력(cargo capacity)을 채우는 큰 질병-유발 유전자(즉, 단일 AAV 패키징 수용력을 초과함)를 모사하기 위해, 분할된 YFP 코딩 서열을 큰 비중단(uninterrupted) 개방 해독 프레임 내에 임베딩시켰다. N-말단에(즉, 5' 쪽에), YFP 코딩 서열의 제1 부분이 긴 스터퍼(stuffer) 서열(즉, 비중단 개방 해독 프레임)로 플랭킹되고, 2A 자가-절단 펩티드를 코딩하는 서열로 이어진다. C-말단 (즉, 3' 쪽)에서, YFP 코딩 서열의 제2 부분이 2A 자가-절단 펩티드 코딩 서열로 이어지고, 그 후, 긴 스터퍼 서열 (즉, 비중단 개방 해독 프레임)로 이어진다 (도 14a). 프리-mRNA 분자를 코딩하는 제1 및 제2 합성 DNA 분자가 프로모터 서열을 제외한, 서열번호 22 및 23으로 표시된다. 결과적으로 발현되는 RNA 분자는 서열번호 22의 1번 위치에 있는 전사 개시 부위 및 서열번호 23의 1번 위치에 있는 전사 개시 부위와 폴리A 테일 사이에 각각 약 4000 nt이다. 결과적으로 전사되는 프리-mRNA 분자 (5' 단편; 서열번호 22로부터 전사됨)는 자가-절단 2A 펩티드 코딩 서열, YFP의 N-말단 부분을 코딩하는 서열, 합성 인트론, 및 (키싱 루프 아키텍처를 갖는) 이합체화 도메인, 및 폴리A 테일로 이어지는, 스터퍼 개방 해독 프레임을 포함한다. C-말단 프리-mRNA 분자 (3' 단편; 서열번호 23으로부터 전사됨)는 상보적인 키싱 루프 이합체화 도메인, 합성 인트론 서열, 뒤이은 C-말단 YFP 코딩 서열, 뒤이은 자가-절단 2A 펩티드 코딩 서열, 뒤이은 스터퍼 개방 해독 프레임, 뒤이은 폴리A 테일로 구성된다.

[0488] 프리-mRNA 분자의 생성 후에, 이합체화 도메인이 결합하고, 스플라이싱이 프리-mRNA를 연결시켜서 전장 mRNA를 생성한다. 번역 동안, YFP를 플랭킹하는 2A 절단 서열이 N- 및 C-말단 스터퍼 서열의 절단 및 기능성 YFP 단백

질의 생성을 가져온다.

[0489] RNA 수준에서 재구성 효율을 결정하기 위해, 2개의 프로브 기반 (5'-가수분해) 정량적 실시간 PCR 분석을 이용한다. 제1 분석은 3' 엑손 YFP 서열에 완전히 포함된 서열(3' 프로브로 표시됨)을 포괄한다. 제2 분석은 5'엑손 YFP 서열과 3' 엑손 YFP 서열간 연결부 (연결부 프로브로 표시됨)를 포괄한다. 재구성 효율은 (연결부 프로브 카운트)/(3' 프로브 카운트)의 비로 계산된다.

[0490] HEK 293t 세포에서 거대 YFP 구조체의 재구성 효율의 정량적 실시간 PCR 분석을 수행했다. 전장 거대 YFP를 기준으로 이용한다. 전장 거대 YFP 비율을 1로 설정한다 (도 14b). 재구성된 것(reconstituted)의 비율은 전장의 분율(fraction)로 표현된다 (분할 -REJ (분할 RNA 말단 연결)로 표시됨). 재구성 효율은 하기와 같이 계산된다: 연결부/3'프라임(3'prime). 도 14b에 도시된 바와 같이, 약 60%의 RNA가 분할-REJ 시스템에서 연결되었다.

[0491] 전장 거대 YFP 발현 및 분할-REJ 발현으로부터의 재구성된 YFP 단백질 발현을 일시적으로 형질감염된 HEK 293t 세포의 유동 세포측정법에 의해 평가한다. 도 14c에 도시된 바와 같이, 분할 REJ 시스템은 큰 카고에 대해서도, 약 45%의 연결 효율을 달성했다.

[0492] 큰 YFP 단백질의 재구성의 인 비보 분석을 하기와 같이 수행했다. 3E9 vg/주사/단편을 포함하는 60nl의 AAV 2/8을 마우스의 일차 운동 피질에 주사했다. 주사 후 10일차에 조직을 수집했다. 도 14d에 도시된 바와 같이, YFP 형광이 벌크 조직에서 용이하게 검출가능하다 (상단 좌측, 상단 중간 패널, 마우스 뇌의 육안 평면도 (macroscopic top view), YFP 형광 + 자가-형광이 상황을 위해 표시됨). 운동 피질의 층 (5) 중 바이러스 주사 부위 및 주변에서 강한 YFP 신호가 검출된다 (우측 패널, 피층에 1 내지 6으로 번호를 부여하고, 대략적인 주사 깊이를 회색 막대로 표시함, 스케일 바(scale bar) = 100 마이크로미터). 따라서, 본 명세서에서 개시되는 시스템은 인 비보에서 큰 단백질을 발현시키기 위해 이용될 수 있다.

[0493] **실시예 16**

[0494] **인자 VIII의 발현**

[0495] 본 실시예는 전장 인간 응고 인자 VIII (FVIII)의 효율적인 재구성을 달성하기 위해 이용되는 방법을 기술한다.

[0496] 실험을 위해 이용된 5' 및 3' 핵산 분자의 개략도가 도 15a에 도시된다 (프리-RNA 분자를 코딩하는 DNA가 각각 서열번호 24 및 25로 기재됨). 각각의 절반은 약 3.8 kb의 FVIII 코딩 서열을 포함한다. FVIII 코딩 서열의 N-말단 절반을 포함하는, 결과적으로 수득되는 RNA 5'-서열 (도 6a의 110에 도식적으로 도시됨)이 효율적인 합성 인트론 및 이합체화 도메인 (키싱 루프 아키텍처), 및 폴리A 테일로 이어진다. FVIII 코딩 서열의 C-말단 절반 (예를 들면, 도 6a의 150)을 포함하는 3'-서열은 상보적 키싱 루프 이합체화 도메인 및 효율적인 합성 인트론 서열에 의해 선행된다.

[0497] RNA 수준에서 재구성 효율을 결정하기 위해, 2개의 프로브 기반 (5'-가수분해) 정량적 실시간 PCR 분석을 이용한다. 제1 분석은 3' 엑손 FVIII 서열에 완전히 포함된 서열(3' 프로브로 표시됨)을 포괄한다. 제2 분석은 5'엑손 FVIII 서열과 3' 엑손 FVIII 서열간 연결부 (연결부 프로브로 표시됨)를 포괄한다. 재구성 효율은 (연결부 프로브 카운트)/(3' 프로브 카운트)의 비로 계산된다.

[0498] HEK 293t 세포에서 2일의 발현 후 재구성 효율의 PCR 정량을 수행했다. 전장 FVIII을 기준으로 이용한다. 전장 FVIII 비율을 1로 설정한다. 재구성된 FVIII 분석 비율은 전장의 분율로 표현된다 (분할 -REJ로 표시됨). 도 15b에 도시된 바와 같이, 약 40-60%의 재구성 효율이 달성되었다 (즉, 2개의 RNA의 약 40-60%가 분할-REJ 시스템에서 연결됨).

[0499] 인 비트로 FVIII의 발현을 입증하기 위해, 웨스턴 블롯팅을 이용하였다. FVIII을 N-말단에서 HA-태그로 태깅했다. 구조체를 HEK 293t 세포에서 2일 동안 발현시킨다. 도 15c에 도시된 바와 같이, 본 명세서에서 개시되는 분할-REJ 시스템은 인 비트로에서 전장 FVIII을 성공적으로 발현시켰다.

[0500] 이러한 관찰에 근거하여, 인 비보에서 전장 FVIII 단백질의 발현이, 예를 들면, A형 혈우병을 치료하기 위해, 달성될 수 있다. 예를 들면, FVIII 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. FVIII의 제2 절반을 상보적인 합성 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, FVIII의 절반 2개는 재조합되어 전장 FVIII 전사물을 형성하고, 단백질로 번역된다. 예를 들면, N-말단 FVIII 코딩 서열을 포함하는 서열번호 24, 및 C-말단 FVIII 코딩 서열을 포함하는 서열번호 25에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 서열이 인 비보 발현을 위해 이용될

수 있다.

[0501] **실시예 17**

[0502] **Abca4의 발현**

[0503] 본 실시예는 전장 인간 Abca4 (ATP binding cassette subfamily A member 4)의 효율적인 재구성을 달성하기 위해 이용되는 방법을 기술한다.

[0504] 이용된 5' 및 3' 분자의 개략도가 도 16a에 도시된다 (프리-RNA 분자를 코딩하는 DNA가 각각 서열번호 20 및 21로 기재됨). 5' 절반은 약 3.6kb의 Abca4 코딩 서열을 포함하고, 3' 절반은 약 3.2kb의 Abca4 코딩 영역 + C-말단 3xFLAG 태그를 포함한다. 5'-서열은 상기 코딩 서열의 N-말단 절반 및 뒤이은 효율적인 합성 인트론 서열 및 제1 이합체화 도메인 (키싱 루프)을 포함한다. 상기 코딩 서열의 C-말단 절반을 포함하는 3'-서열은 상보적 (키싱 루프) 이합체화 도메인 및 효율적인 합성 인트론 서열에 의해 선행된다. 연결부를 포함한 생거 시퀀싱 트레이스가 도시된다.

[0505] 도 16b에 도시된 바와 같이, 연결부의 PCR 증폭은 2개의 코딩 서열의 충실한 연결(faithful joining)을 입증한다. RNA 수준에서 재구성 효율을 결정하기 위해, 2개의 프로브 기반 (5'-가수분해) 정량적 실시간 PCR 분석을 이용한다 (도 16c). 제1 분석은 3' 엑손 Abca4 서열에 완전히 포함된 서열(3' 프로브로 표시됨)을 포괄한다. 제2 분석은 5'엑손 Abca4 서열과 3' 엑손 Abca4 서열간 연결부 (연결부 프로브로 표시됨)를 포괄한다. 재구성 효율은 (연결부 프로브 카운트)/(3' 프로브 카운트)의 비로 계산된다. HEK 293t 세포에서 2일의 발현 후 재구성 효율의 PCR 정량이 도 16d에 도시된다. 전장 Abca4를 기준으로 이용한다. 평균 전장 Abca4 비율을 1로 설정한다. 재구성된 Abca4 분석 비율은 전장의 분율로 표현된다 (분할 -REJ로 표시됨). 도 16d에 도시된 바와 같이, 약 35%의 재구성 효율이 달성되었다 (즉, 2개의 RNA의 약 30-40%가 분할-REJ 시스템에서 연결됨).

[0506] 인 비트로 Abca4의 발현을 입증하기 위해, 웨스턴 블롯팅을 이용하였다. Abca4를 C-말단에서 3xFLAG-태그로 태깅했다. 구조체를 HEK 293t 세포에서 2일 동안 발현시킨다. 도 16e에 도시된 바와 같이, 본 명세서에서 개시되는 분할-REJ 시스템은 인 비트로에서 전장 Abca4를 성공적으로 발현시켰다.

[0507] 웨스턴 블롯의 정량이 도 16f에 표시된다. 조건들간 차등적인 형질감염 효율을 정규화시키기 위해, 전장 플라스미드와 C-말단 플라스미드는 형질감염 대조군으로 BFP (Blue Fluorescent Protein)를 공-발현한다. 각 시료 중 BFP 농도를 도트 블롯(dot blot)에 의해 결정하고 조건들 간 정규화를 위해 이용했다. 도 16f에 도시된 바와 같이, 재구성된 Abca4는 직접적인 전장 발현과 비교시 수준의 약 40%로 발현된다. 따라서, 웨스턴 블롯에 의해 결정된 단백질 수준은 qPCR에 의해 결정된 RNA 재구성 효율을 잘 추적한다.

[0508] 이러한 관찰에 근거하여, 인 비보에서 전장 ABCA4 단백질의 발현이, 예를 들면, 스타르가르트병을 치료하기 위해, 달성될 수 있다. 예를 들면, ABCA4 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. ABCA4의 제2 절반을 상보적인 합성 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, ABCA4의 절반 2개는 재조합되어 전장 ABCA4 전사물을 형성하고, 단백질로 번역된다. 예를 들면, N-말단 Abca4 코딩 서열을 포함하는 서열번호 20 (도 10r-10u), 및 C-말단 Abca4 코딩 서열을 포함하는 서열번호 21 (도 10v-10z)에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 서열이 인 비보 발현을 위해 이용될 수 있다.

[0509] **실시예 18**

[0510] **Otof의 발현**

[0511] 본 실시예는 전장 마우스 Otof(Otoferlin)의 효율적인 재구성을 달성하기 위해 이용되는 방법을 기술한다.

[0512] 이용된 5' 및 3' DNA 분자의 서열은 각각 서열번호 155 및 156으로 표시된다. 5' 절반은 약 3.5kb의 Otof 코딩 서열을 포함하고, 3' 절반은 약 2.5kb의 Otof 코딩 영역 + C-말단 3xFLAG 태그를 포함한다. 상기 C-말단 절반을 포함하는 3'-서열(예를 들면, 도 6a의 150)은 상보적 결합 도메인 및 효율적인 합성 인트론 서열에 의해 선행된다. 당업자는 인간 OTOF 코딩 서열 (예를 들면, GenBank Accession No. NM_001287489.2 또는 NM_194248.3)이 서열번호 155 및 156의 마우스 코딩 서열을 대체할 수 있다는 것을 이해할 것이다.

[0513] 인 비트로 Otof의 발현을 입증하기 위해, 웨스턴 블롯팅을 이용하였다. 웨스턴 블롯 검출을 위해 Otof를 C-말단에서 3xFLAG-태그로 태깅했다. 구조체를 HEK 293t 세포에서 2일 동안 발현시킨다. 도 18a에 도시된 바와 같이,

본 명세서에서 개시되는 분할-REJ 시스템은 인 비트로에서 전장 Otof를 성공적으로 발현시켰다.

[0514] 웨스턴 블롯의 정량이 도 18b-18c에 표시된다. 원시 정량(raw quantification)이 전장 대조군의 분획으로서 좌측 막대 플롯에 표시된다 (도 18b). 조건들간 차등적인 형질감염 효율을 정규화시키기 위해, 전장 플라스미드와 C-말단 플라스미드는 형질감염 대조군으로 BFP (Blue Fluorescent Protein)를 공-발현한다. 세포를 회수하기 전에 각 시료 중 BFP 농도를 공초점 형광 현미경 관찰에 의해 결정하고 조건들간 정규화를 위해 이용했다. 정규화된 정량이 전장 대조군의 정규화된 분획으로서 우측 막대 플롯에 표시된다 (도 18c). 도 18c에 도시된 바와 같이, 재구성된 Otof가 직접적인 전장 발현과 비교시 수준의 약 30%로 발현된다.

[0515] 이러한 관찰에 근거하여, 인 비보에서 전장 OTOF 단백질의 발현이, 예를 들면, 상염색체 열성 난청(autosomal recessive deafness) 9를 치료하기 위해, 달성될 수 있다. 예를 들면, OTOF 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인(즉 인트론 및 결합 도메인)을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. OTOF의 제2 절반을 상보적인 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, 2개의 RNA 분자가 표적 세포에서 발현되고, OTOF 코딩 전사물의 2개의 절반이 재조합되어 전장 OTOF 전사물을 형성하고, 단백질로 번역된다. 예를 들면, N-말단 Otof 코딩 서열을 포함하는 서열번호 155, 및 C-말단 Otof 코딩 서열을 포함하는 서열번호 156에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 서열이 인 비보 발현을 위해, 예를 들면, 난청을 치료하기 위해 이용될 수 있다.

[0516] **실시예 19**

[0517] **Myo7a의 발현**

[0518] 본 실시예는 전장 인간 Myo7a (MYOSIN VIIA)의 효율적인 재구성을 달성하기 위해 이용되는 방법을 기술한다.

[0519] 이용된 5' 및 3' DNA 분자의 서열은 각각 서열번호 157 및 158로 표시된다. 5' 절반은 약 3.6kb의 Myo7a 코딩 서열을 포함하고, 3' 절반은 약 3.1kb의 Myo7a 코딩 영역 + C-말단 3xFLAG 태그를 포함한다. C-말단 절반을 포함하는 3'-서열(예를 들면, 도 6a의 150)은 상보적 결합 도메인 및 효율적인 합성 인트론 서열에 의해 선행된다.

[0520] 인 비트로 Myo7a의 발현을 입증하기 위해, 웨스턴 블롯팅을 이용하였다. 웨스턴 블롯 검출을 위해 Myo7a를 C-말단에서 3xFLAG-태그로 태깅했다. 구조체를 HEK 293t 세포에서 2일 동안 발현시킨다. 도 19a에 도시된 바와 같이, 본 명세서에서 개시되는 분할-REJ 시스템은 인 비트로에서 전장 Myo7a를 성공적으로 발현시켰다.

[0521] 웨스턴 블롯의 정량이 도 19b-19c에 표시된다. 원시 정량이 전장 대조군의 분획으로서 좌측 막대 플롯에 표시된다 (도 19b). 조건들간 차등적인 형질감염 효율을 정규화시키기 위해, 전장 플라스미드와 C-말단 플라스미드는 형질감염 대조군으로 BFP (Blue Fluorescent Protein)를 공-발현한다. 세포를 회수하기 전에 각 시료 중 BFP 농도를 공초점 형광 현미경 관찰에 의해 결정하고 조건들간 정규화를 위해 이용했다. 정규화된 정량이 전장 대조군의 정규화된 분획으로서 우측 막대 플롯에 표시된다 (도 19c). 도 19c에 도시된 바와 같이, 재구성된 Myo7a가 직접적인 전장 발현과 비교시 수준의 약 60%로 발현된다.

[0522] 이러한 관찰에 근거하여, 인 비보에서 전장 Myo7a 단백질의 발현이, 예를 들면, 1B 타입의 어셔 증후군를 치료하기 위해, 달성될 수 있다. 예를 들면, MYO7A 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인(즉 인트론 및 결합 도메인)을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. MYO7A의 제2 절반을 상보적인 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, 2개의 RNA 분자가 표적 세포에서 발현되고, MYO7A 코딩 전사물의 2개의 절반이 재조합되어 전장 MYO7A 전사물을 형성하고, 단백질로 번역된다. 예를 들면, N-말단 Myo7a 코딩 서열을 포함하는 서열번호 157, 및 C-말단 Myo7a 코딩 서열을 포함하는 서열번호 158에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 서열이 인 비보 발현을 위해, 예를 들면, 난청을 치료하기 위해 이용될 수 있다.

[0523] **실시예 20**

[0524] **dCas9-VPR의 발현**

[0525] 본 실시예는 전장 dCas9-VPR (enzymatically dead Cas9 fused to a VPR-transcriptional activator domain)의 효율적인 재구성을 달성하기 위해 이용되는 방법을 기술한다.

[0526] 이용된 5' 및 3' DNA 분자의 서열은 각각 서열번호 159 및 160으로 표시된다. 5' 절반은 약 3.3kb의 DCas9-VPR 코딩 서열을 포함하고, 3' 절반은 약 2.5kb의 DCas9-VPR 코딩 영역을 포함한다. C-말단 절반을 포함하는 3'-서열(예를 들면, 도 6a의 150)은 상보적 결합 도메인 및 효율적인 합성 인트론 서열에 의해 선행된다.

[0527] 인 비트로 DCas9-VPR의 발현을 입증하기 위해, 웨스턴 블롯팅을 이용하였다. 구조체를 HEK 293t 세포에서 2일 동안 발현시킨다. 도 20a에 도시된 바와 같이, 본 명세서에서 개시되는 분할-REJ 시스템은 인 비트로에서 전장 DCas9-VPR를 성공적으로 발현시켰다.

[0528] 웨스턴 블롯의 정량이 도 20b-20c에 표시된다. 원시 정량이 전장 대조군의 분획으로서 좌측 막대 플롯에 표시된다 (도 20b). 조건들간 차등적인 형질감염 효율을 정규화시키기 위해, 전장 플라스미드와 C-말단 플라스미드는 형질감염 대조군으로 BFP (Blue Fluorescent Protein)를 공-발현한다. 세포를 회수하기 전에 각 시료 중 BFP 농도를 공초점 형광 현미경 관찰에 의해 결정하고 조건들간 정규화를 위해 이용했다. 정규화된 정량이 전장 대조군의 정규화된 분획으로서 우측 막대 플롯에 표시된다 (도 20c). 도 20c에 도시된 바와 같이, 재구성된 DCas9-VPR이 직접적인 전장 발현과 비교시 수준의 약 35%로 발현된다. HEK 293t 세포에서 UAS 표적화 가이드 RNA와 함께 발현시 (도 20d), 전장 및 이원 분할 재구성 dCas9-VPR(two-way split reconstituted dCas9-VPR)이 모두 UAS-YFP 플라스미드로부터 YFP 발현을 유도하여, 재구성 dCas9-VPR의 기능성을 입증한다.

[0529] 이러한 관찰에 근거하여, 인 비보에서 전장 DCAS9-VPR 단백질의 발현이, 예를 들면, 유전자를 활성화시키거나 또는 과발현시키기 위해, 달성될 수 있다. 예를 들면, DCAS9-VPR 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인(즉 인트론 및 결합 도메인)을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. DCAS9-VPR의 제2 절반을 상보적인 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, 2개의 RNA 분자가 표적 세포에서 발현되고, DCAS9-VPR 코딩 전사물의 2개의 절반이 재조합되어 전장 DCAS9-VPR 전사물을 형성하고, 단백질로 번역된다. 예를 들면, N-말단 DCAS9-VPR 코딩 서열을 포함하는 서열번호 159, 및 C-말단 DCAS9-VPR 코딩 서열을 포함하는 서열번호 160에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 서열이 인 비보 발현을 위해 이용될 수 있다.

[0530] **실시예 21**

[0531] **Prime Editor의 발현**

[0532] 본 실시예는 전장 인간화 Cas9 프라임 에디터 (Prime Editor)의 효율적인 재구성을 달성하기 위해 이용되는 방법을 기술한다.

[0533] 이용된 5' 및 3' DNA 분자의 서열은 각각 서열번호 161 및 162로 표시된다. 5' 절반은 약 3.3kb의 Prime Editor 코딩 서열을 포함하고, 3' 절반은 약 3.0kb의 Prime Editor 코딩 영역을 포함한다. C-말단 절반을 포함하는 3'-서열(예를 들면, 도 6a의 150)은 상보적 결합 도메인 및 효율적인 합성 인트론 서열에 의해 선행된다.

[0534] 인 비트로 Prime Editor의 발현을 입증하기 위해, 웨스턴 블롯팅을 이용하였다. 구조체를 HEK 293t 세포에서 2일 동안 발현시킨다. 도 21a에 도시된 바와 같이, 본 명세서에서 개시되는 분할-REJ 시스템은 인 비트로에서 전장 Prime Editor를 성공적으로 발현시켰다.

[0535] 웨스턴 블롯의 정량이 도 21b-21c에 표시된다. 원시 정량이 전장 대조군의 분획으로서 좌측 막대 플롯에 표시된다 (도 21b). 조건들간 차등적인 형질감염 효율을 정규화시키기 위해, 전장 플라스미드와 C-말단 플라스미드는 형질감염 대조군으로 BFP (Blue Fluorescent Protein)를 공-발현한다. 세포를 회수하기 전에 각 시료 중 BFP 농도를 공초점 형광 현미경 관찰에 의해 결정하고 조건들간 정규화를 위해 이용했다. 정규화된 정량이 전장 대조군의 정규화된 분획으로서 우측 막대 플롯에 표시된다 (도 21c). 도 21c에 도시된 바와 같이, 재구성된 Prime Editor가 직접적인 전장 발현과 비교시 수준의 약 60%로 발현된다. 도 21d는 표적화된 G의 T로의 전환 돌연변이가 전장 및 이원 분할 prime editor를 이용하여 도입될 수 있다는 것을 보여주어, 이원 분할 primer editor 구조체의 기능성을 입증한다.

[0536] 이러한 관찰에 근거하여, 인 비보에서 전장 PRIME EDITOR 단백질의 발현이, 예를 들면, 게놈 점 돌연변이를 치료하기 위해, 달성될 수 있다. 예를 들면, PRIME EDITOR 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인(즉 인트론 및 결합 도메인)을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. PRIME EDITOR의 제2 절반을 상보적인 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, 2개의 RNA 분자가 표적 세포에서 발현되고, PRIME EDITOR 코딩 전사물의 2개의 절반

이 재조합되어 전장 PRIME EDITOR 전사물을 형성하고, 단백질로 번역된다. 예를 들면, N-말단 PRIME EDITOR 코딩 서열을 포함하는 서열번호 161, 및 C-말단 PRIME EDITOR 코딩 서열을 포함하는 서열번호 162에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 서열이 인 비보 발현을 위해 이용될 수 있다.

[0537] **실시예 22**

[0538] **AncBE4의 발현**

[0539] 본 실시예는 전장 인간화 AncBE4 (Cas9 Cytosine Base Editor)의 효율적인 재구성을 달성하기 위해 이용되는 방법을 기술한다.

[0540] 이용된 5' 및 3' DNA 분자의 서열은 각각 서열번호 163 및 164로 표시된다. 5' 절반은 약 2.4kb의 AncBE4 코딩 서열을 포함하고, 3' 절반은 약 3.2kb의 AncBE4 코딩 영역을 포함한다. C-말단 절반을 포함하는 3'-서열(예를 들면, 도 6a의 150)은 상보적 결합 도메인 및 효율적인 합성 인트론 서열에 의해 선행된다.

[0541] 인 비트로 AncBE4의 발현을 입증하기 위해, 웨스턴 블롯팅을 이용하였다. 구조체를 HEK 293t 세포에서 2일 동안 발현시킨다. 도 22a에 도시된 바와 같이, 본 명세서에서 개시되는 분할-REJ 시스템은 인 비트로에서 전장 AncBE4를 성공적으로 발현시켰다.

[0542] 웨스턴 블롯의 정량이 도 22b에 표시된다. 원시 정량이 전장 대조군의 분획으로서 좌측 막대 플롯에 표시된다(도 22b). 도 22b에 도시된 바와 같이, 재구성된 AncBE4가 직접적인 전장 발현과 비교시 수준의 약 40-50%로 발현된다. 도 22c는 표적화된 C의 T로의 전이 돌연변이가 전장 및 이원 분할 AncBE4를 이용하여 도입될 수 있다는 것을 보여주어, 이원 분할 AncBE4 구조체의 기능성을 입증한다.

[0543] 이러한 관찰에 근거하여, 인 비보에서 전장 ANCBE4 단백질의 발현이, 예를 들면, 게놈 점 돌연변이를 치료하기 위해, 달성될 수 있다. 예를 들면, ANCBE4 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인(즉 인트론 및 결합 도메인)을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. ANCBE4의 제2 절반을 상보적인 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, 2개의 RNA 분자가 표적 세포에서 발현되고, ANCBE4 코딩 전사물의 2개의 절반이 재조합되어 전장 ANCBE4 전사물을 형성하고, 단백질로 번역된다. 예를 들면, N-말단 AncBE4 코딩 서열을 포함하는 서열번호 163, 및 C-말단 AncBE4 코딩 서열을 포함하는 서열번호 164에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 서열이 인 비보 발현을 위해 이용될 수 있다.

[0544] **실시예 23**

[0545] **Abe8e의 발현**

[0546] 본 실시예는 전장 인간화 Abe8e (Cas9 Adenosine Base Editor)의 효율적인 재구성을 달성하기 위해 이용되는 방법을 기술한다.

[0547] 이용된 5' 및 3' DNA 분자의 서열은 각각 서열번호 165 및 166으로 표시된다. 5' 절반은 약 2.4kb의 Abe8e 코딩 서열을 포함하고, 3' 절반은 약 3.2kb의 Abe8e 코딩 영역을 포함한다. C-말단 절반을 포함하는 3'-서열(예를 들면, 도 6a의 150)은 상보적 결합 도메인 및 효율적인 합성 인트론 서열에 의해 선행된다.

[0548] 인 비트로 Abe8e의 발현을 입증하기 위해, 웨스턴 블롯팅을 이용하였다. 구조체를 HEK 293t 세포에서 2일 동안 발현시킨다. 도 23a에 도시된 바와 같이, 본 명세서에서 개시되는 분할-REJ 시스템은 인 비트로에서 전장 Abe8e를 성공적으로 발현시켰다.

[0549] 웨스턴 블롯의 정량이 도 23b에 표시된다. 원시 정량이 전장 대조군의 분획으로서 좌측 막대 플롯에 표시된다(도 23b). 도 23b에 도시된 바와 같이, 재구성된 Abe8e가 직접적인 전장 발현과 비교시 수준의 약 70%로 발현된다. 도 23c는 표적화된 C의 T로의 전이 돌연변이가 전장 및 이원 분할 Abe8e를 이용하여 도입될 수 있다는 것을 보여주어, 이원 분할 Abe8e 구조체의 기능성을 입증한다.

[0550] 이러한 관찰에 근거하여, 인 비보에서 전장 ABE8E 단백질의 발현이, 예를 들면, 게놈 점 돌연변이를 치료하기 위해, 달성될 수 있다. 예를 들면, ABE8E 코딩 서열의 제1 절반에 합성 RNA 이합체화 및 재조합 도메인(즉 인트론 및 결합 도메인)을 부가하고 제1 벡터/플라스미드로부터 발현시킨다. ABE8E의 제2 절반을 상보적인 RNA 이합체화 및 재조합 도메인에 부가하고 제2 벡터/플라스미드로부터 발현시킨다. 동일한 세포에서 함께 발현되는 경우, 2개의 RNA 분자가 표적 세포에서 발현되고, ABE8E 코딩 전사물의 2개의 절반이 재조합되어 전장 ABE8E 전사

물을 형성하고, 단백질로 번역된다. 예를 들면, N-말단 Abe8e 코딩 서열을 포함하는 서열번호 165, 및 C-말단 Abe8e 코딩 서열을 포함하는 서열번호 166에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 서열이 인 비보 발현을 위해 이용될 수 있다.

[0551] **실시예 24**

[0552] **RNA 단편 길이의 증가가 이원 분할 유전자 재구성의 감소를 초래한다**

[0553] 5' 단편을 코딩하는 RNA 분자 및 3' 단편을 코딩하는 RNA 분자의 길이의 영향을 평가했다.

[0554] yfp (yellow fluorescent protein) 코딩 서열을 2개의 단편으로 분할했다. RNA 코딩 서열을 연장하기 위해, 스터퍼 개방 해독 프레임(ORF)을 각각 5' 단편의 5' 말단 및 3' 단편의 3' 말단에 설치했다. yfp의 5' yfp 코딩 서열을 자가-절단 2A 서열을 통해 연장된 스터퍼 ORF에 융합시켰다. yfp의 3' yfp 코딩 서열을 자가-절단 2A 서열을 통해 연장된 스터퍼 ORF에 연결시켰다. yfp의 5' 단편과 yfp의 3' 단편의 분할점에, RNA 말단 연결 모듈(합성 인트론 + 결합 도메인)을 설치했다. 자가-절단 2A 서열은 번역 후에 YFP 단백질이 개별적인 스터퍼 ORF로부터 분리될 수 있게 한다. 상이한 길이의 스터퍼 개방 해동 프레임을 포함시키는 것에 의해, 4종의 5' 단편 코딩 구조체와 4종의 3' 단편 코딩 구조체를 조립했다. 이러한 구조체로부터 전사된 RNA (단백질 코딩 서열 + 합성 인트론 및 결합 도메인)의 길이는 5' 단편에 대해 1000nt, 2000nt, 3000nt, 및 4000nt이고, 3' 단편에 대해 1000nt, 2000nt, 3000nt, 및 4000nt이었다.

[0555] YFP 재구성의 효율을 모든 16종의 5'-3' 단편 쌍에 대해 비교했다. 이 비교에서, YFP는 가장 짧은 구조체 (즉, 5'-1000nt와 3'-1000nt)가 쌍을 이룬 경우 가장 효율적으로 재구성되었다. 더 긴 스터퍼 서열을 갖는 단편이 쌍을 이룬 경우, 재구성 효율의 감소가 관찰되었다. 가장 짧은 쌍 형성 (5'-1000nt과 3'-1000nt)의 비율로서, 하기 YFP 재조합 효율이 관찰되었다:

[0556] 5'-1000nt와 3'-1000nt: 100%

[0557] 5'-1000nt와 3'-2000nt: ~40%

[0558] 5'-1000nt와 3'-3000nt: ~20%

[0559] 5'-1000nt와 3'-4000nt: ~16%

[0560] 5'-2000nt와 3'-1000nt: ~55%

[0561] 5'-2000nt와 3'-2000nt: ~30%

[0562] 5'-2000nt와 3'-3000nt: ~20%

[0563] 5'-2000nt와 3'-4000nt: ~15%

[0564] 5'-3000nt와 3'-1000nt: ~60%

[0565] 5'-3000nt와 3'-2000nt: ~40%

[0566] 5'-3000nt와 3'-3000nt: ~25%

[0567] 5'-3000nt와 3'-4000nt: ~20%

[0568] 5'-4000nt와 3'-1000nt: ~40%

[0569] 5'-4000nt와 3'-2000nt: ~35%

[0570] 5'-4000nt와 3'-3000nt: ~20%

[0571] 5'-4000nt와 3'-4000nt: ~15%.

[0572] 이러한 데이터는 분할 유전자의 5' 및 3' 코딩 서열을 코딩하는 단편의 길이를 증가시키는 것이 점진적으로 분할 유전자 재구성의 효율을 저하시킨다는 것을 보여준다.

[0573] **실시예 25**

[0574] **DISE 및 ISE 서열에 의한 RNA 말단 연결 반응의 증진**

[0575] 본 실시예는 특이적 스플라이싱 인핸서 서열을 포함시키는 것에 의해 2개의 RNA 분자의 효율적인 연결을 달성하

기 위해 이용되는 방법을 기술한다.

[0576] 분할 YFP가 삼중 키싱 루프 RNA 이합체화 도메인 및 인트론 세그먼트의 가변 라이브러리(variable library)로 구성된 RNA 말단 연결 모듈을 이용하여 재구성되는 것인 스크리닝 플랫폼을 이용하여, 선택된 인트론 스플라이싱 인핸서 (ISE) 서열의 구체적인 유효성을 조사했다. 이용된 5' 및 3' DNA 분자의 서열이 각각 서열번호 171 및 172로 표시된다 (서열 중 일련의 N은 인트론 라이브러리 배치(intronic library placement), 예를 들면, 하기 표 2의 서열 중 적어도 하나, 예를 들면, 이러한 서열의 1, 2, 3, 4 또는 5개의 부위를 나타냄).

[0577] 인 비트로에서 재구성된 yfp의 발현을 입증하기 위해, 유동 세포측정법을 이용하여 5' 및 3' DNA 분자로 형질감염된 HEK293t 세포에서 yfp 형광 강도를 결정하였다. 도 24a에 도시된 바와 같이, RNA 연결 반응을 촉진하는 효율적인 인트론 스플라이싱 인핸서 서열을 찾기 위해 본 명세서에서 개시되는 분할-REJ 시스템의 인트론 부분을 개별적인 세그먼트로 세분하였다. 구조체의 5' 인트론 부분의 3개의 위치 및 3' 인트론 부분의 3개의 위치에서 이용된 서열이 서열번호 173 내지 204로 제시되고 (표 2), 도 24c에 열거된다.

[0578] 표 2: 예시적인 인트론 스플라이싱 인핸서 서열

표 2

[0579]

표지 (label)	명칭	서열 (서열번호)
ds1	FGFR-2 pre-mRNA, IAS1	GTAAGTATTgctttcatttttgtctttttttaa (173)
ds2	Fas URI6	GTAagttcttgctttgttcaaactgtctat (174)
ds3	CFTR E9 PY1/2	GTAAGTATTCTTTTGTTCTTCACtcat (175)
ds4	TIA1-preferred	GTAAGTATTTTTTTTACTCCtcaTTTTTACTCC (176)
ds5	FAS intron5	GTAAGTATTTTTTTTACGGTTATATTCTCCTTTCCCC (177)
ds6	CD46-D1/9	GTAAGTATTTTCTGTTGTTTATTttcag (178)
ds7	B19V ISE	GTAAGTATTGGGGTTGATTATGTGTGGGACGGTGTAAGG (179)
ds8	ratFGFR2DISE	GTAAGTATTtcctctttctttccatgggttggcct (180)
ds9	just donor scramble	GTAAGTATTaccagagattcgtagacctgcttgac (181)
m1	6xWGGG	TGGGGCTGGGCAGAGGGTTGAGGGGAGAGGGTCCTGGGG (182)
m2	C9-E6-ISE	tcaTGGGTGGGTtcatTGGGTGGGTtca (183)
m3	AdMLBPadj	Tagggcgcagtagtccagggttt (184)
m4	bcl2-I2-BPadj	Ttctctgtggggtggcattctctgctctct (185)
m5	M2	GGGttatGGGACCtcaGGGataaGGGACC (186)
m6	GH1ivs	CGGGGATGGGGGtca (187)
m7	WangGrich	TGGGGGGAGGtcaTGGGGGGAGG (188)
m8	WangISE2	GTTGGTGGTTtcatGTTGGTGGTT (189)
m9	WangA	GGGTTTCGGGTTTtcaGGTGGTCGTTGGT (190)
m10	WangB	GGTGGTCGTTGGTtcaTTTGGGCTATTGG (191)
m11	WangC	TTTGGGCTATTGGtcaAGGGGGCGAGGGG (192)
m12	WangD	AGGGGGCGAGGGGtcaGGTATTCGGTATT (193)
m13	WangE	GGTATTCGGTATTtcaaggtaaCaggtaa (194)
m14	WangFmod	aggtaaCaggtaatcaGGGTTTCGGGTTT (195)
m15	SMN-URC2/3	TCTTACTTTTGTaaacTTTATGGTTTGTg (196)
m16	just scramble	Cacgtattctcggtacggacgttacaga (197)
dd1	scramble	Taagctggtatcc (198)
ap2	4.1R-E16-uISE	CACTAACTCTTTTTCCCCCCttttttttttACAG (199)
ap3	P6-cons-to30	TACTAACtctttctttttttCCTTTCCTTCTTCACAG (200)
ap4	AdMLSA	CACTAACTCTgtcatacttatcctgtccctttttttttccaCAG (201)
ap5	bcl2-I2-SA	CACTAACTCTctttctttttcttccctcctctcccccaactgCAG (202)
ap6	perfectT	CACTAACTCTttttttttttttttttttttttACAGCAG (203)
ad1	scramble	Taagctggtatcc (204)

[0580] 유동 세포 측정법의 정량이 도 24b에 도시된다. 5' 스플라이스 부위 선택 촉진 스플라이싱 인자 TIA-1 (T-Cell-Restricted Intracellular Antigen-1)의 동원을 촉진하는 인트론 서열의 포함이 RNA 말단 연결을 증가시킬 수 있다. 일부 구체예에서, WGGG 모티프를 포함하는 서열이 RNA 말단 연결을 증진시킨다.

[0581] 이러한 관찰에 근거하여, 전장 분할 단백질의 인 비보 발현이 RNA 말단 연결 모듈의 인트론 부분에 특이적 ISE 서열을 포함시키는 것에 의해 증진될 수 있다. 예를 들면, 서열번호 173 내지 180, 182-196, 또는 199 내지 203 중 어느 하나에 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 서열 동일성을 갖는 하나 이상의 서열 (예를 들면, 1개, 2개, 또는 3개의 서열)이 RNA 말단 연결 반응 산물의 인 비보 발현을 위해 이용될 수 있다 (예를 들면, 본 명세서에서 제공되는 구현예를 위한 ISE로서 이용될 수 있음).

[0582] 본 개시의 원리가 적용될 수 있는 다수의 가능한 구현예를 고려하여, 예시된 구현예는 본 발명의 예시에 불과하고 본 발명의 범위를 한정하는 것으로 간주되어서는 안된다는 것이 인식되어야 한다. 오히려, 본 발명의 범위는 하기 청구항에 의해 정의된다. 따라서, 본 발명자들은 이러한 청구항의 범위 및 사상 내에 속하는 모든 것을 본 발명으로 청구한다.

부호의 설명

[0583] 110, 150, 200, 220, 500, 600, 700: 핵산 분자

112, 152, 202, 210: 프로모터

114, 154, 502: 표적 단백질의 N-말단 부분에 대한 코딩 서열

116, 156, 220, 506: 스플라이스 도너

118, 222, 508: DISE

120, 160, 206, 224, 510, 604: ISE

158, 208, 606: 분지점

122, 154, 204, 226: 이합체화 도메인

124, 166, 228, 616: 폴리아데닐화 서열

130, 170, 230, 240: 인트론

160, 608: 폴리피리미딘 트랙트

164, 614: 표적 단백질의 C-말단 부분에 대한 코딩 서열

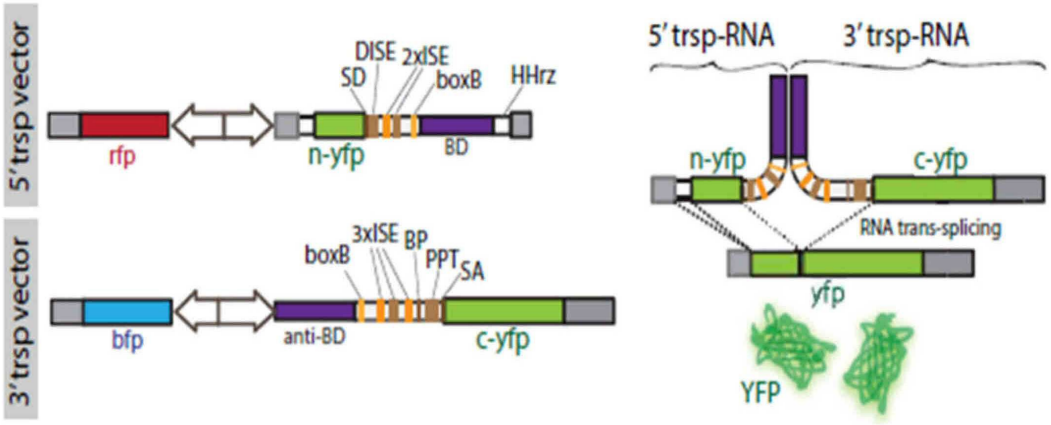
216: 표적 단백질의 중간 부분에 대한 코딩 서열

212, 610: 스플라이스 억셉터

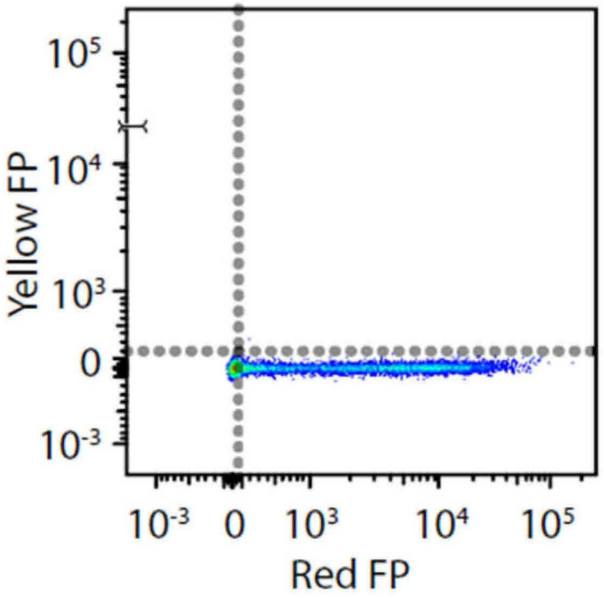
512, 602: 압타머

도면

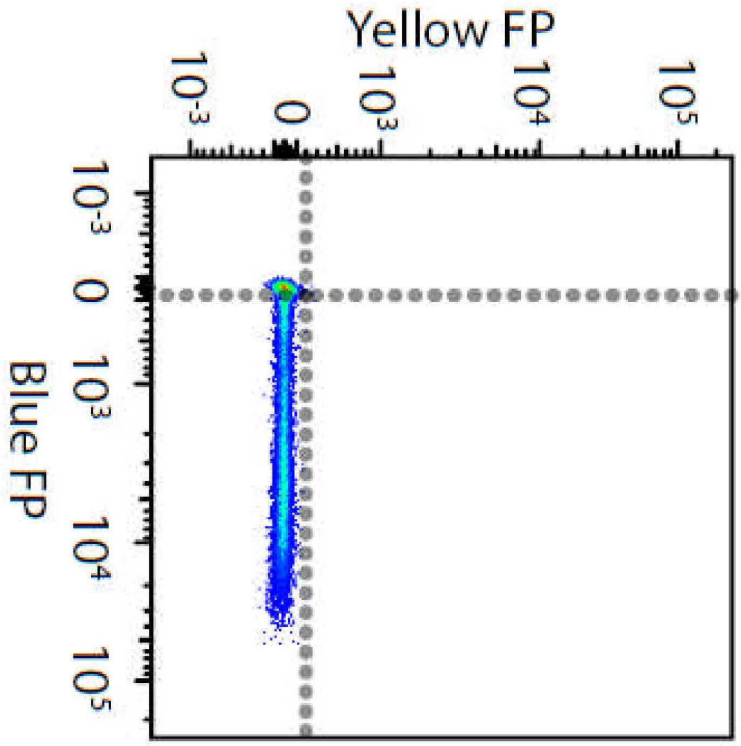
도면1a


5'trsp vector
rfp
DISE
SD
2xISE
boxB
HHrz
n-yfp
BD
3'trsp vector
bfp
boxB
3xISE
BP
PPT
SA
anti-BD
c-yfp
5'trsp-RNA
3'trsp-RNA
n-yfp
c-yfp
RNA trans-splicing
yfp
YFP

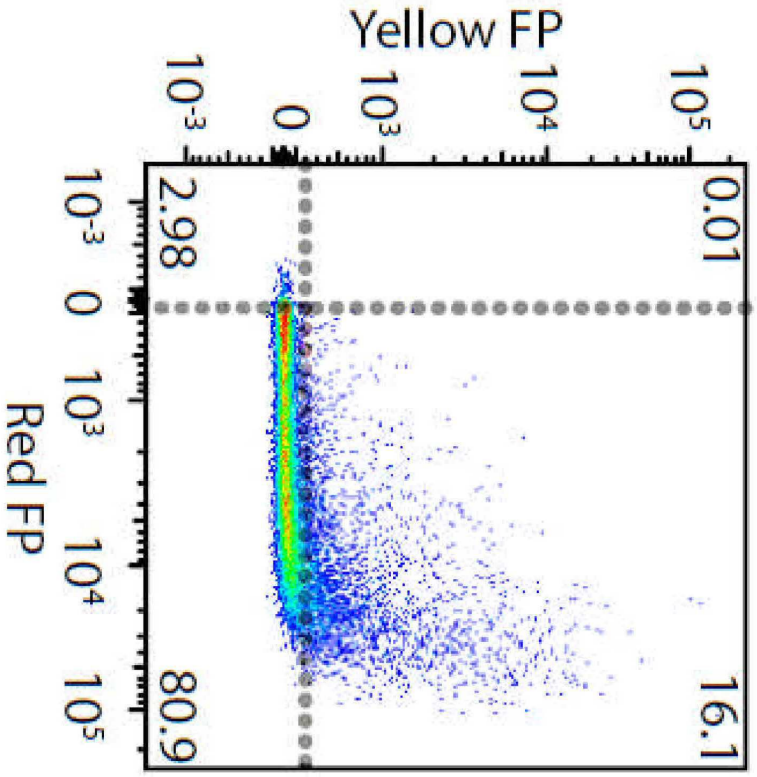
도면1b


Yellow FP
Red FP
10^5
10^4
10^3
0
10^-3
10^-3
0
10^3
10^4
10^5

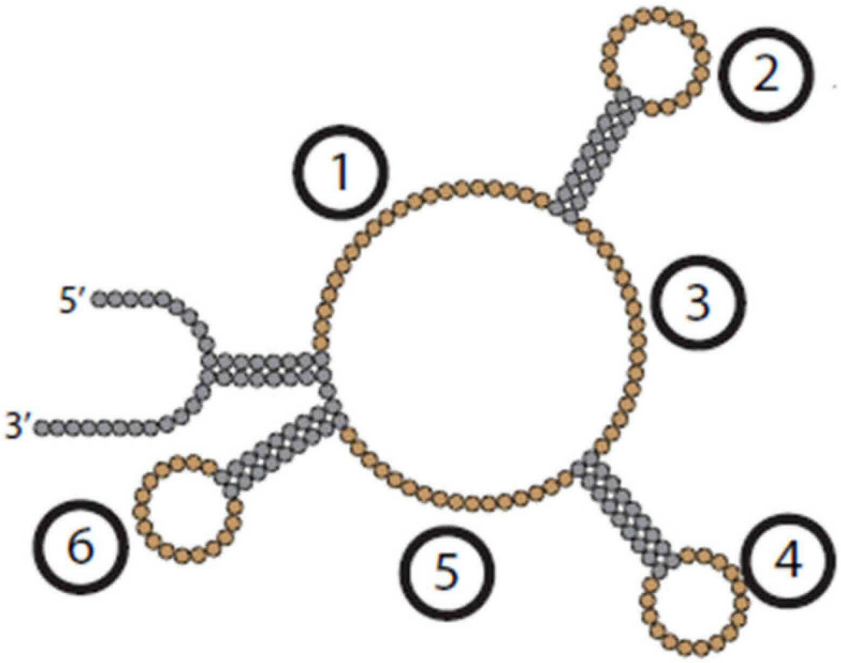
도면1c


Yellow FP
10-3
0
103
104
105
10-3
0
103
104
105
Blue FP

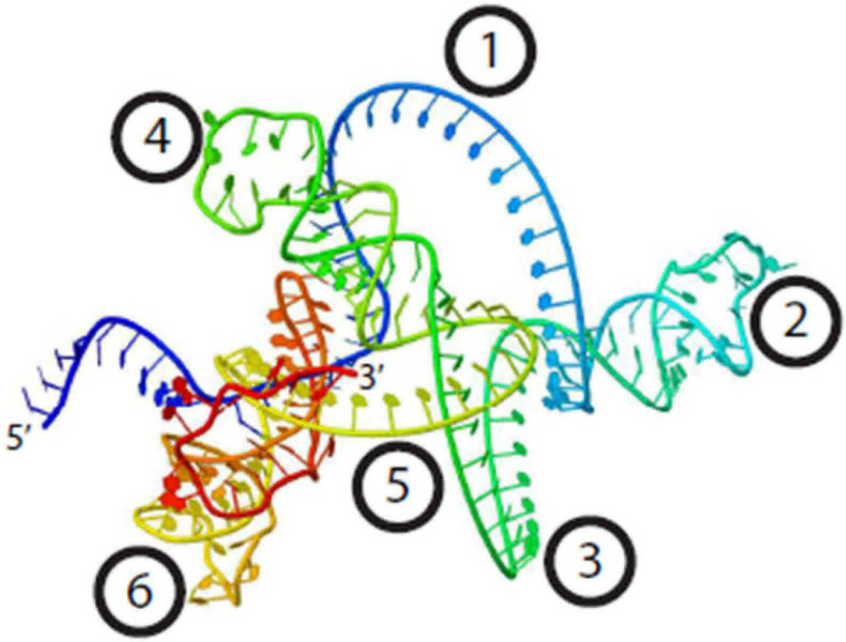
도면1d


Yellow FP
10-3
0
103
104
105
2.98
0.01
80.9
16.1
10-3
0
103
104
105
Red FP

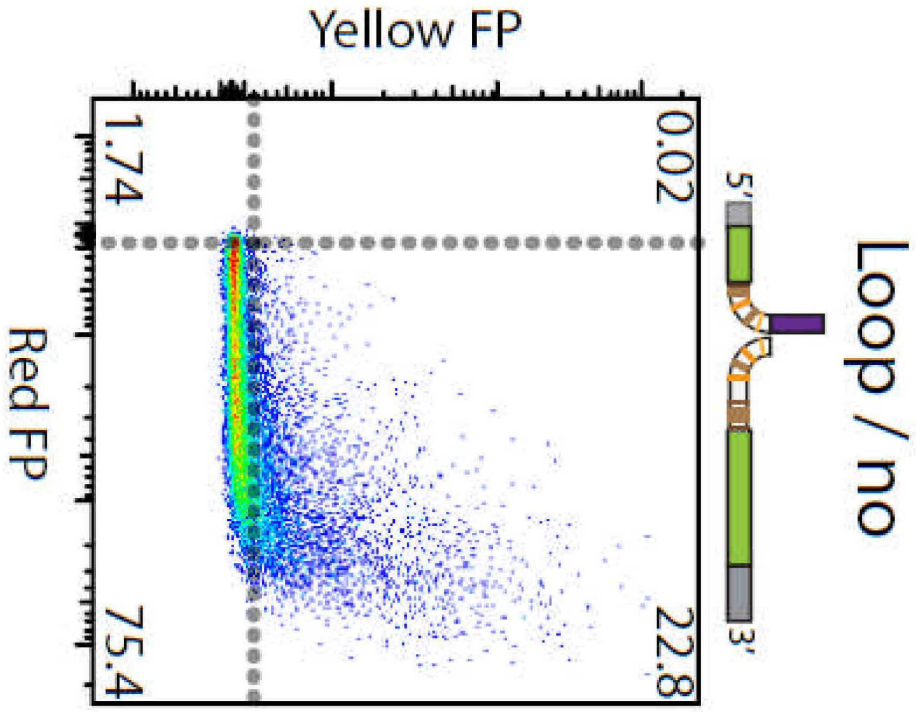
도면1e



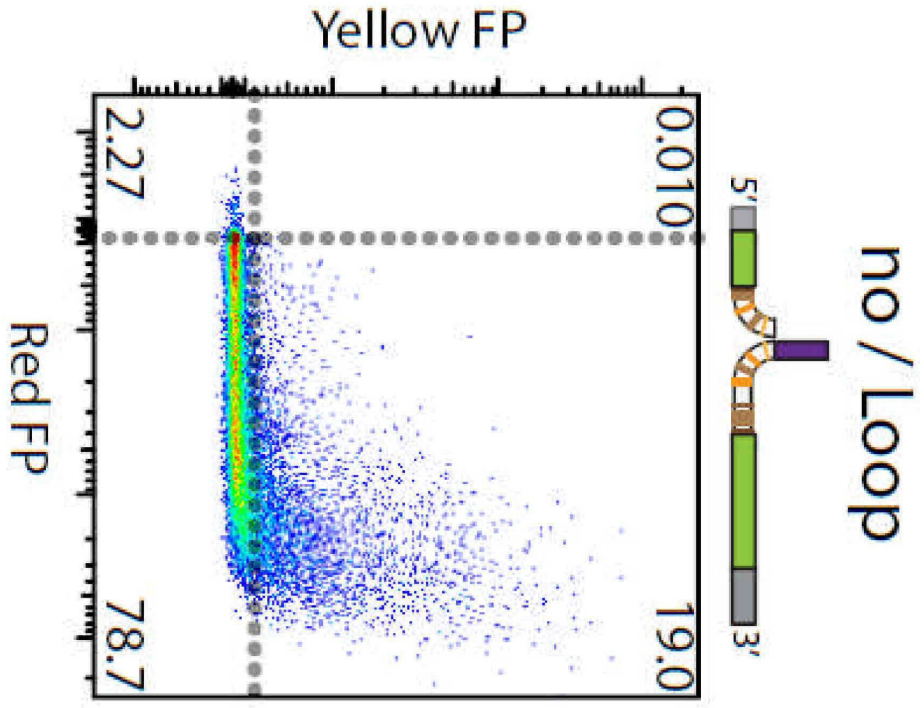
도면1f



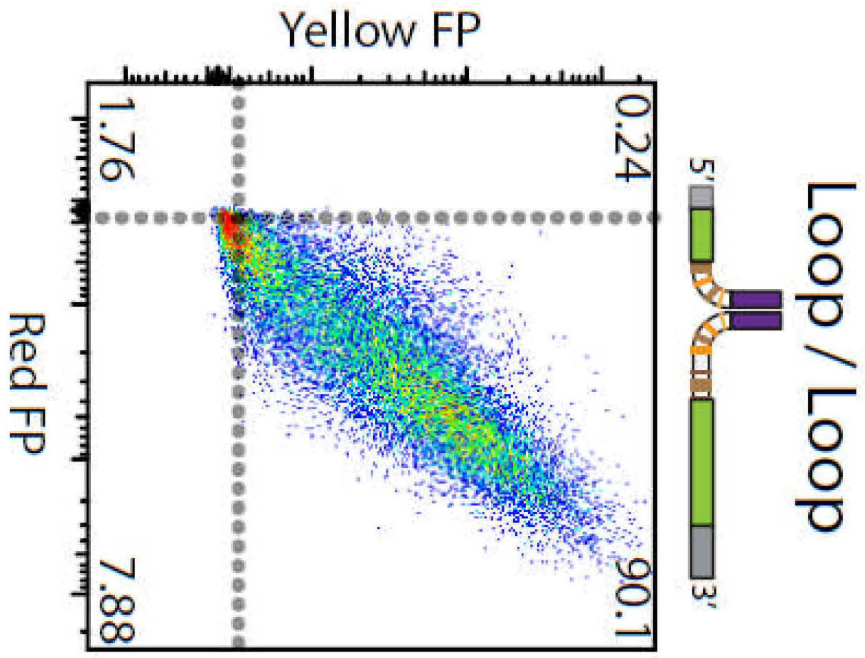
도면1g


Yellow FP
Red FP
1.74
75.4
0.02
22.8
5'
3'
Loop / no

도면1h


Yellow FP
Red FP
2.27
78.7
0.010
19.0
5'
3'
no / Loop

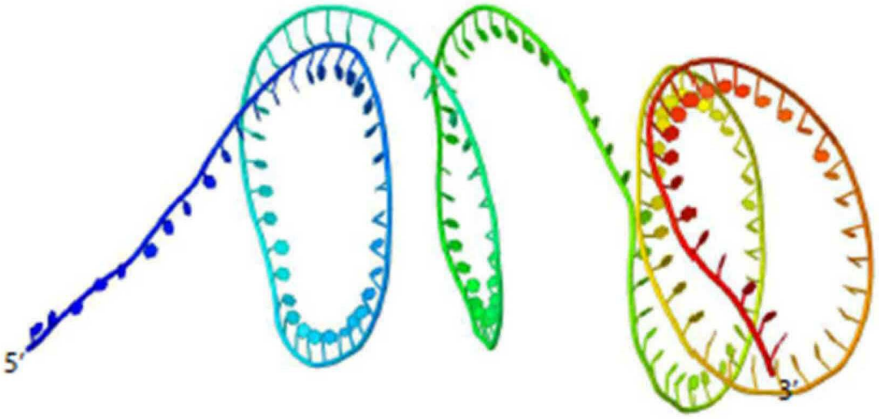
도면1i



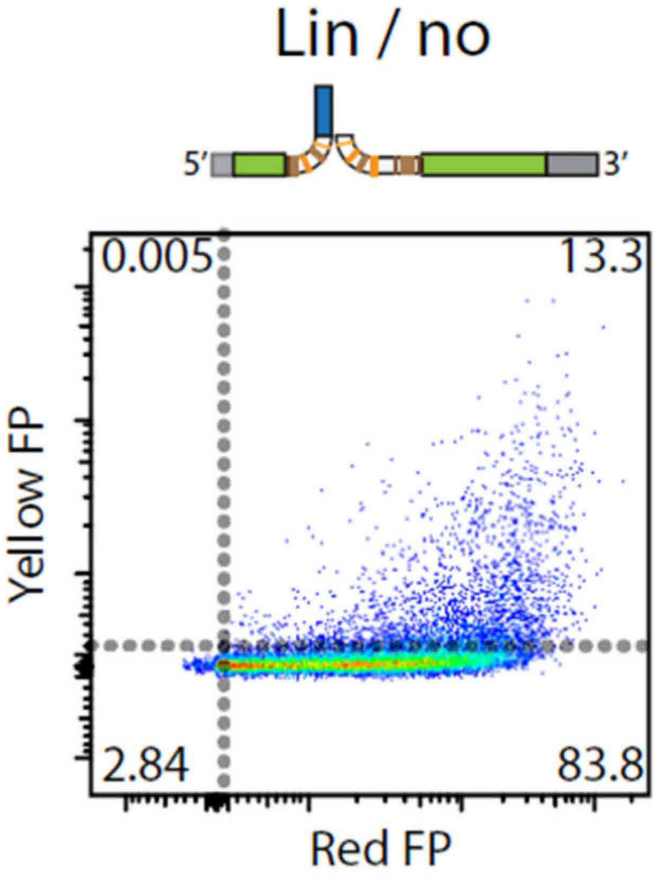
도면1j



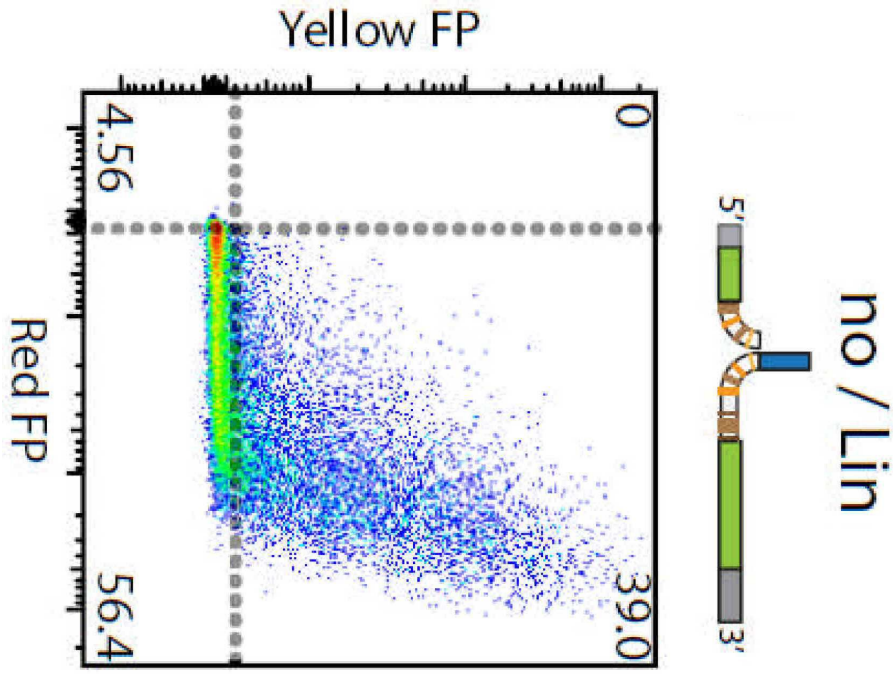
도면1k



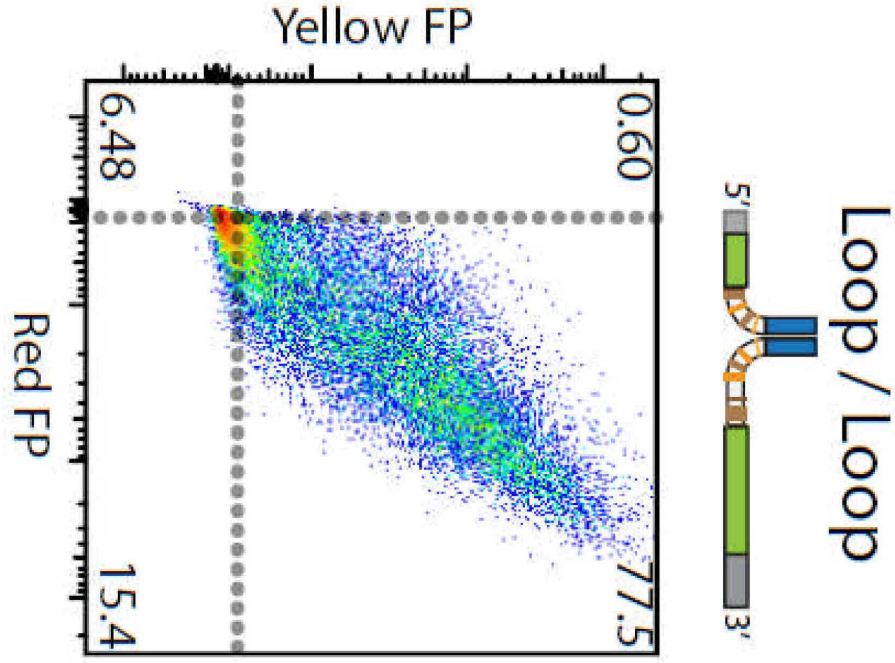
도면11



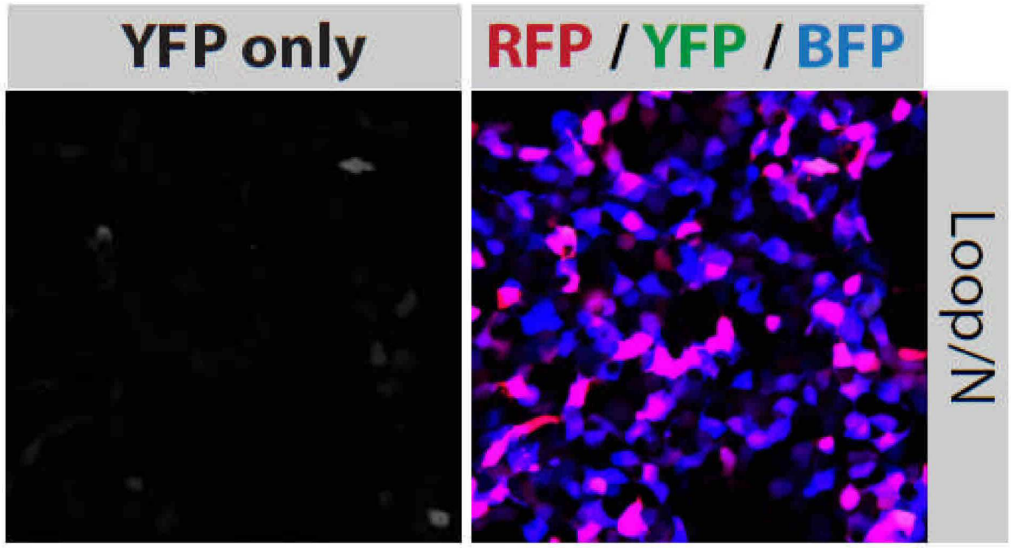
도면1m


Yellow FP
Red FP
4.56
56.4
0
39.0
5'
3'
no / Lin

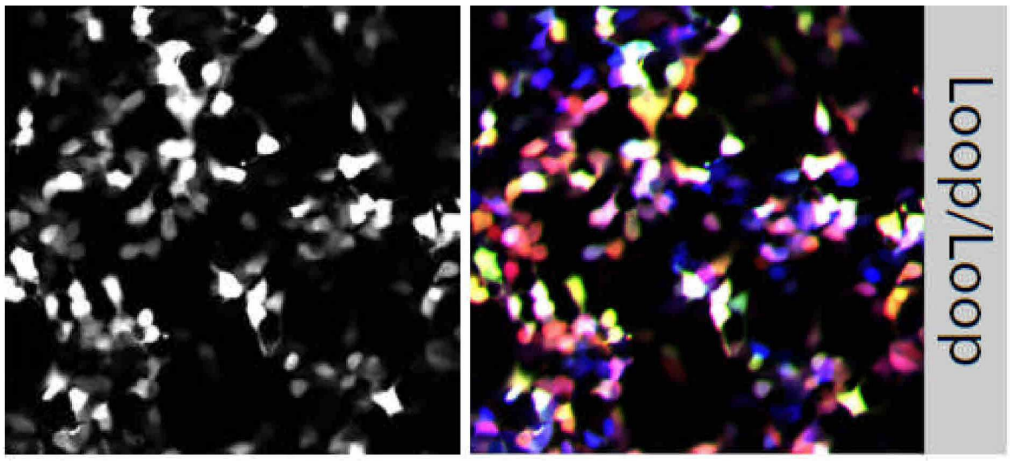
도면1n


Yellow FP
Red FP
6.48
15.4
0.60
77.5
5'
3'
Loop / Loop

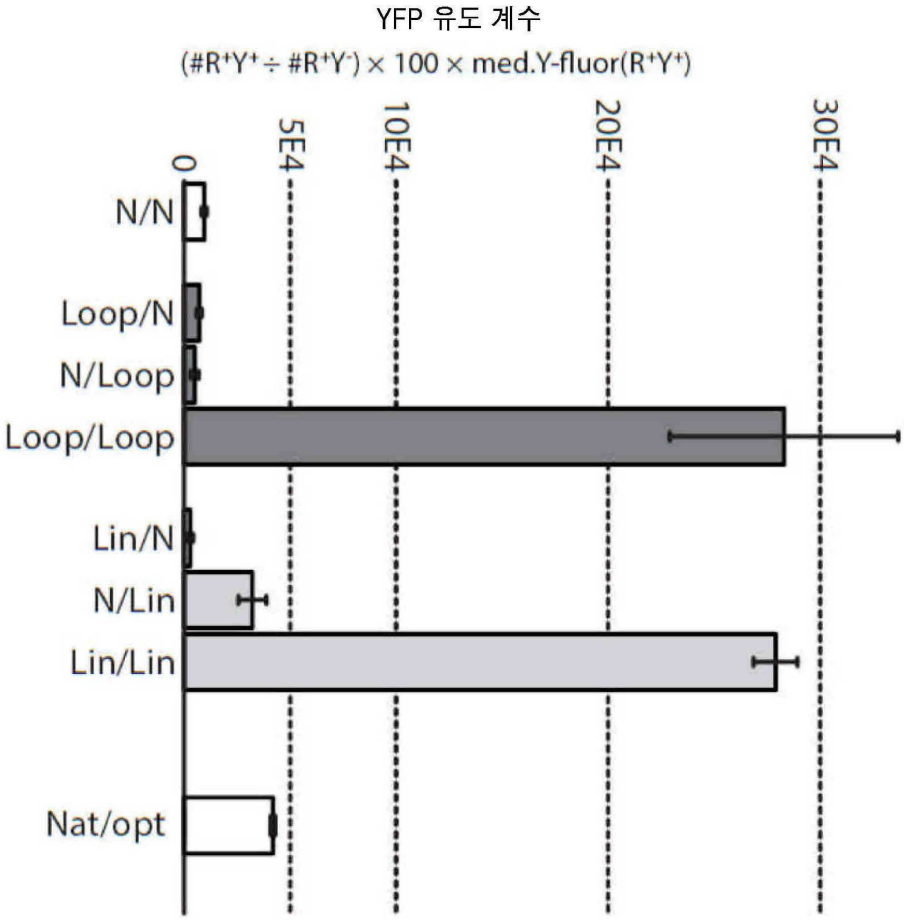
도면1o



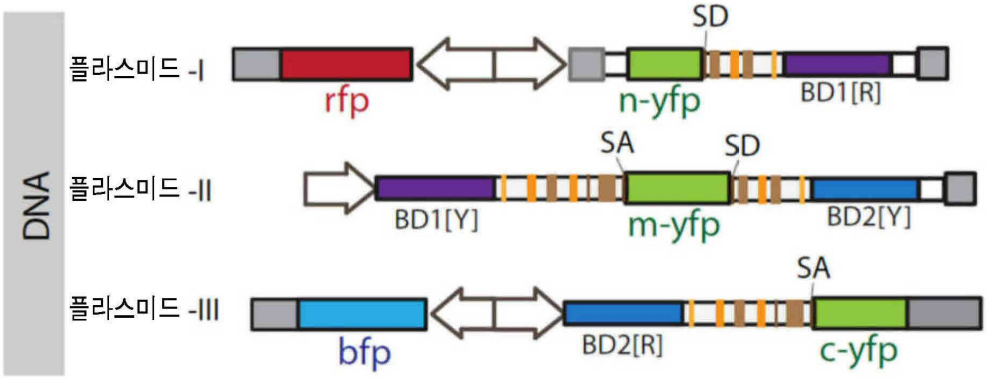
도면1p



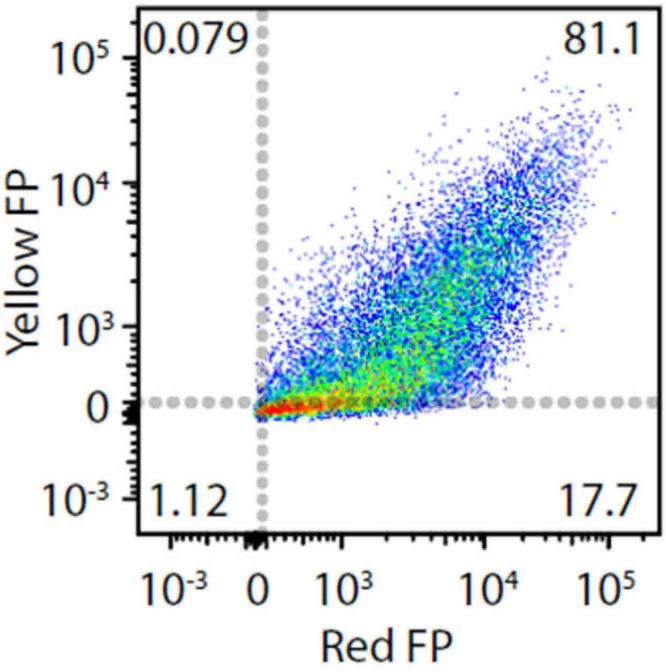
도면1q


YFP 유도 계수
(#R+Y+ ÷ #R+Y-) × 100 × med.Y-fluor(R+Y+)
0
5E4
10E4
20E4
30E4
N/N
Loop/N
N/Loop
Loop/Loop
Lin/N
N/Lin
Lin/Lin
Nat/opt

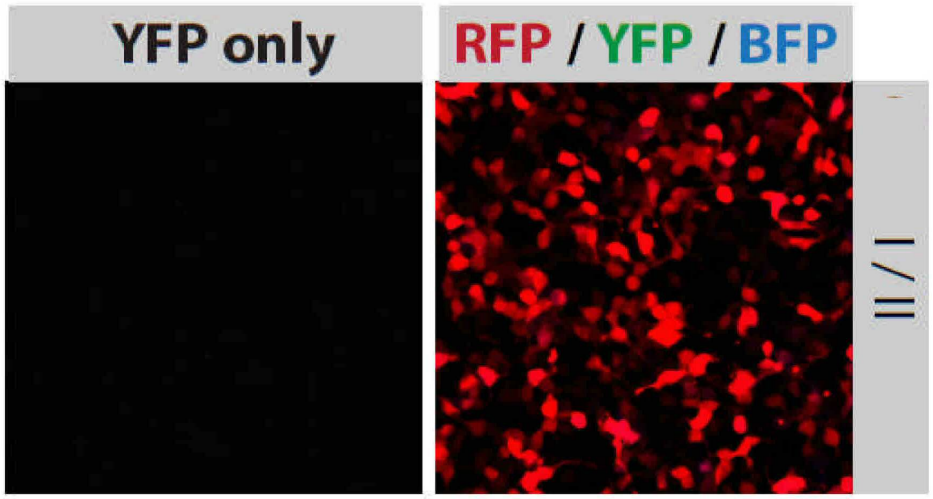
도면2a


DNA
플라스미드 -I
rfp
SD
n-yfp
BD1[R]
플라스미드 -II
BD1[Y]
SA
SD
m-yfp
BD2[Y]
플라스미드 -III
bfp
BD2[R]
SA
c-yfp

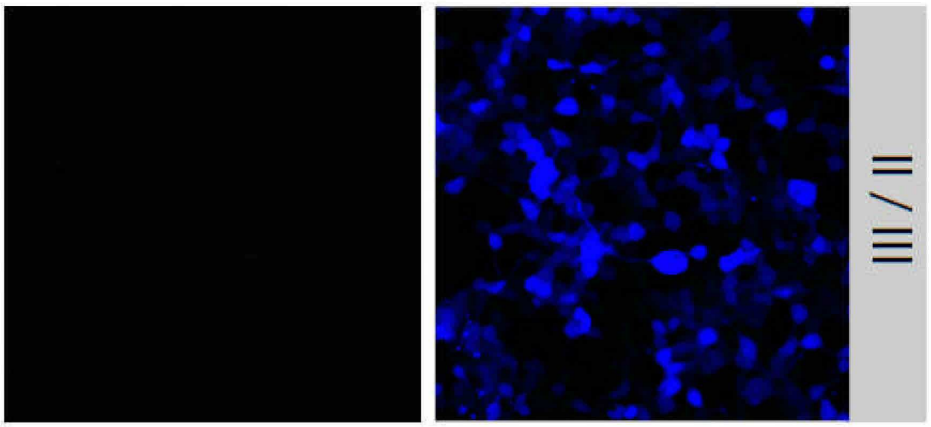
도면2b


0.079
81.1
1.12
17.7
Yellow FP
Red FP
10⁵
10⁴
10³
0
10⁻³
10⁻³
0
10³
10⁴
10⁵

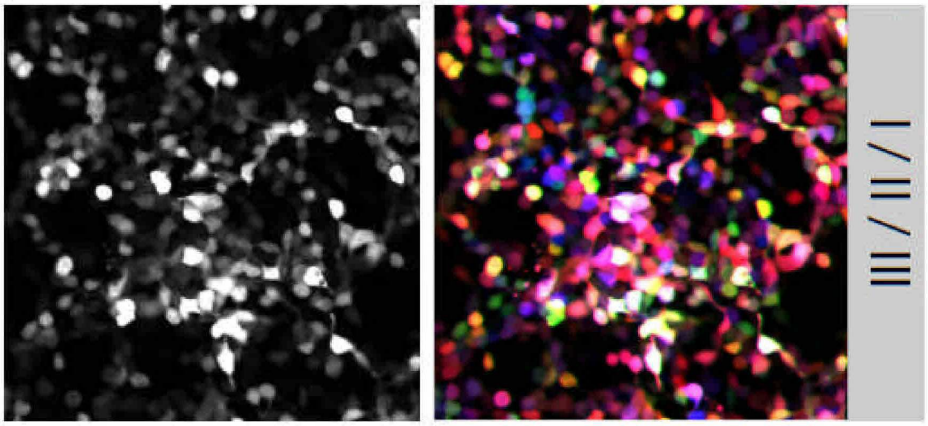
도면2c


YFP only
RFP / YFP / BFP

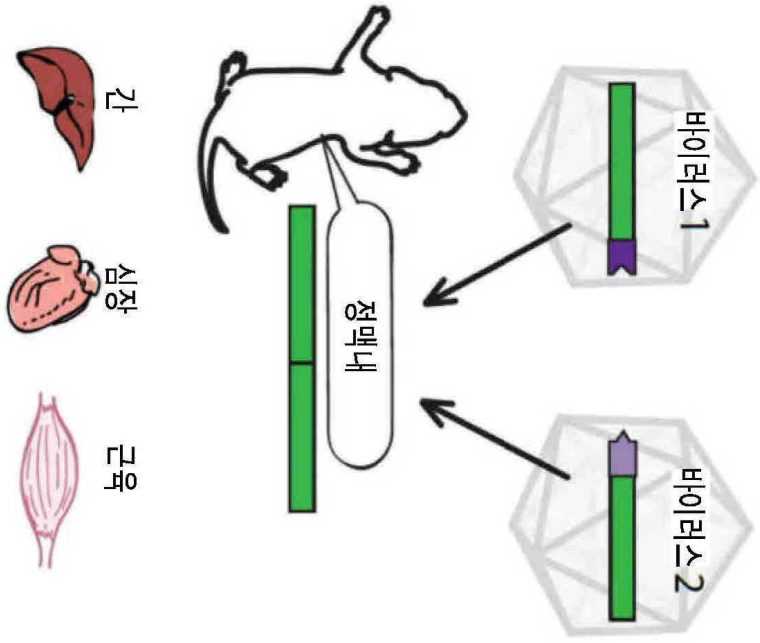
도면2d



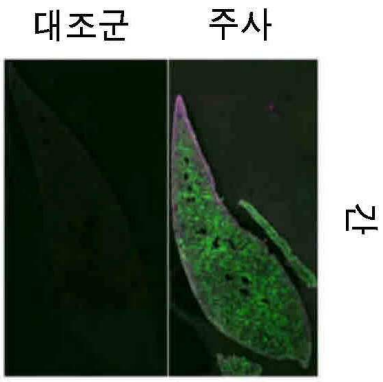
도면2e



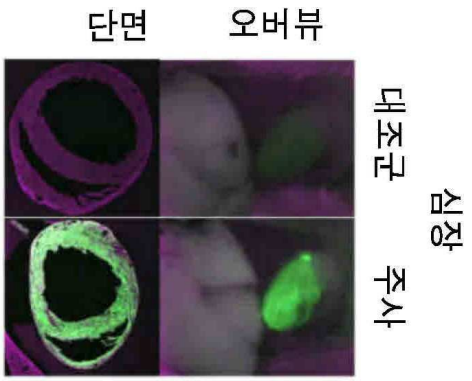
도면3a



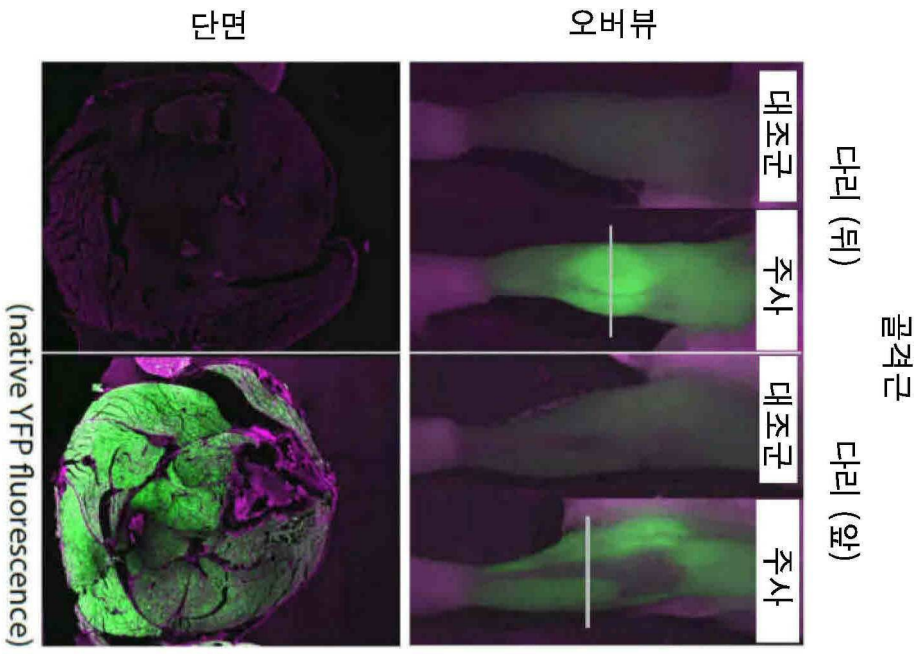
도면3b



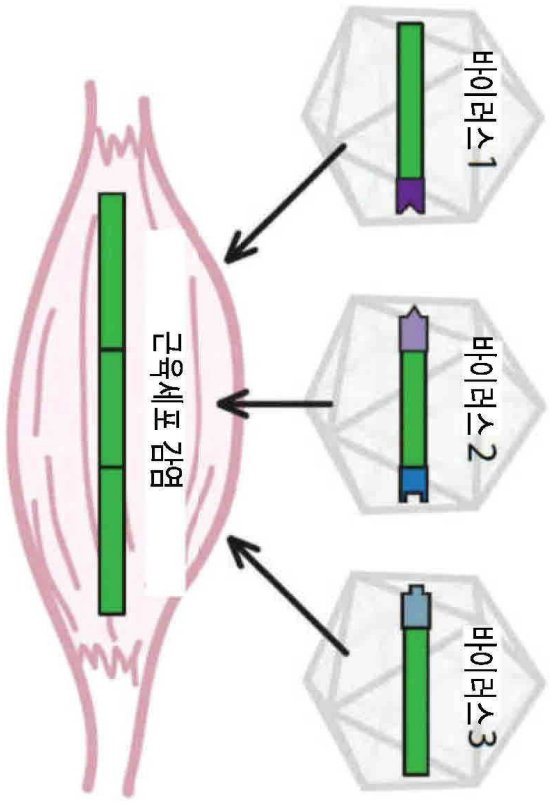
도면3c



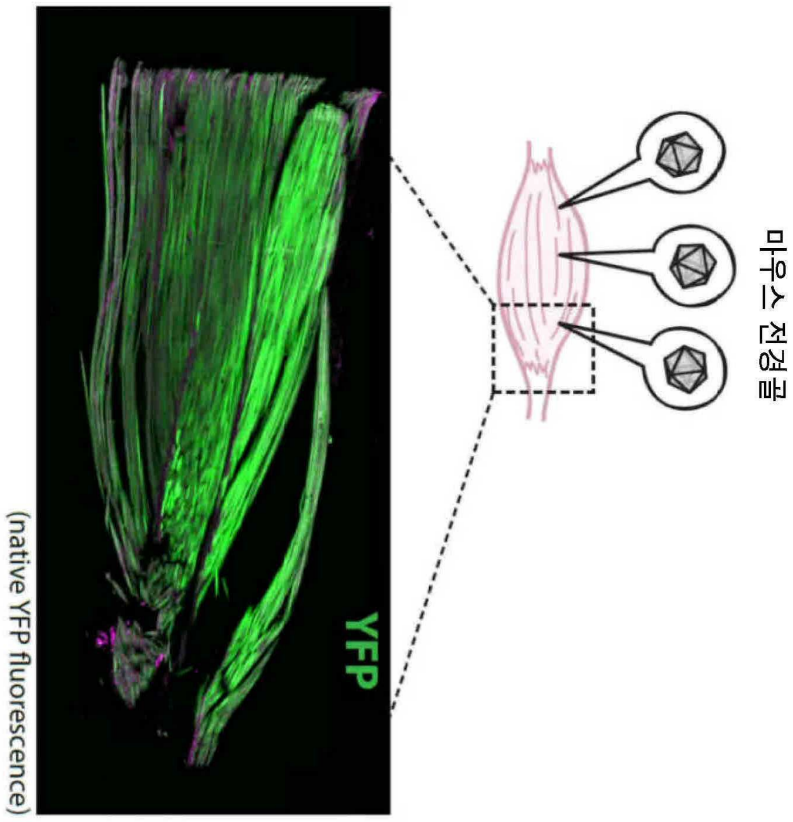
도면3d



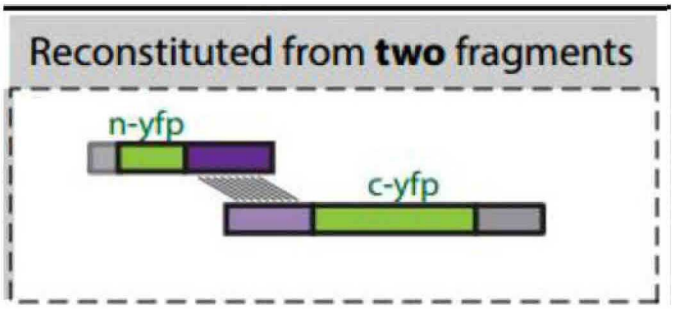
도면4a


바이러스1
바이러스2
바이러스3
근육세포 감염

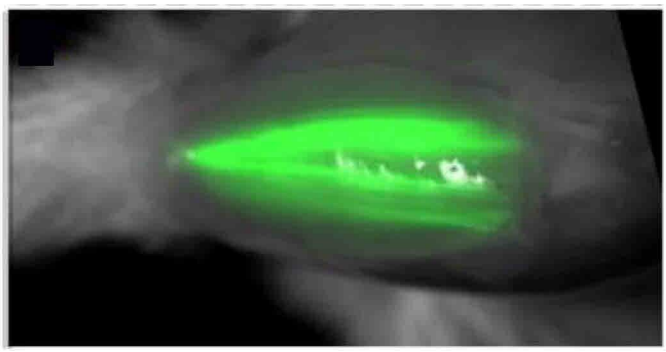
도면4b


마우스 전경골
YFP
(native YFP fluorescence)

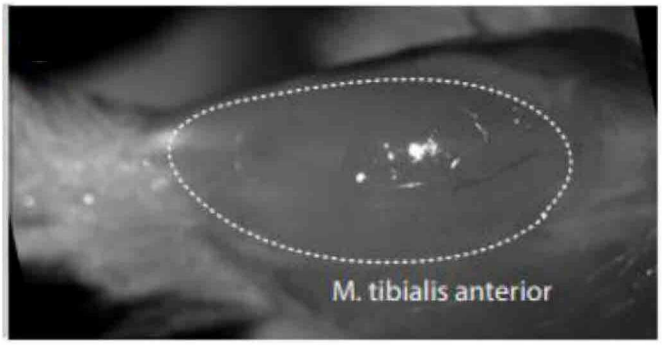
도면5a



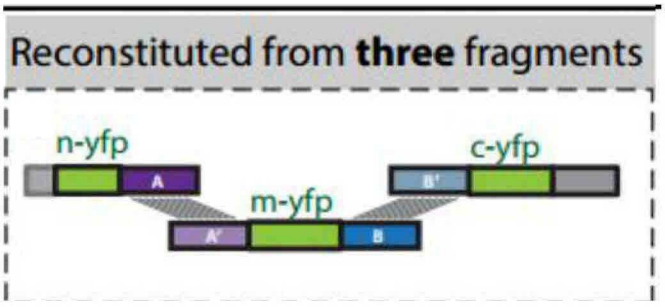
도면5b



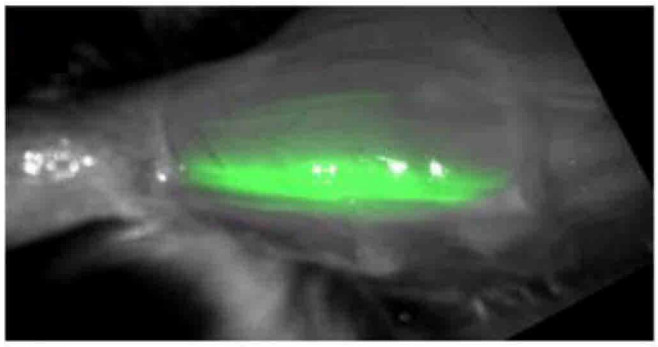
도면5c



도면5d



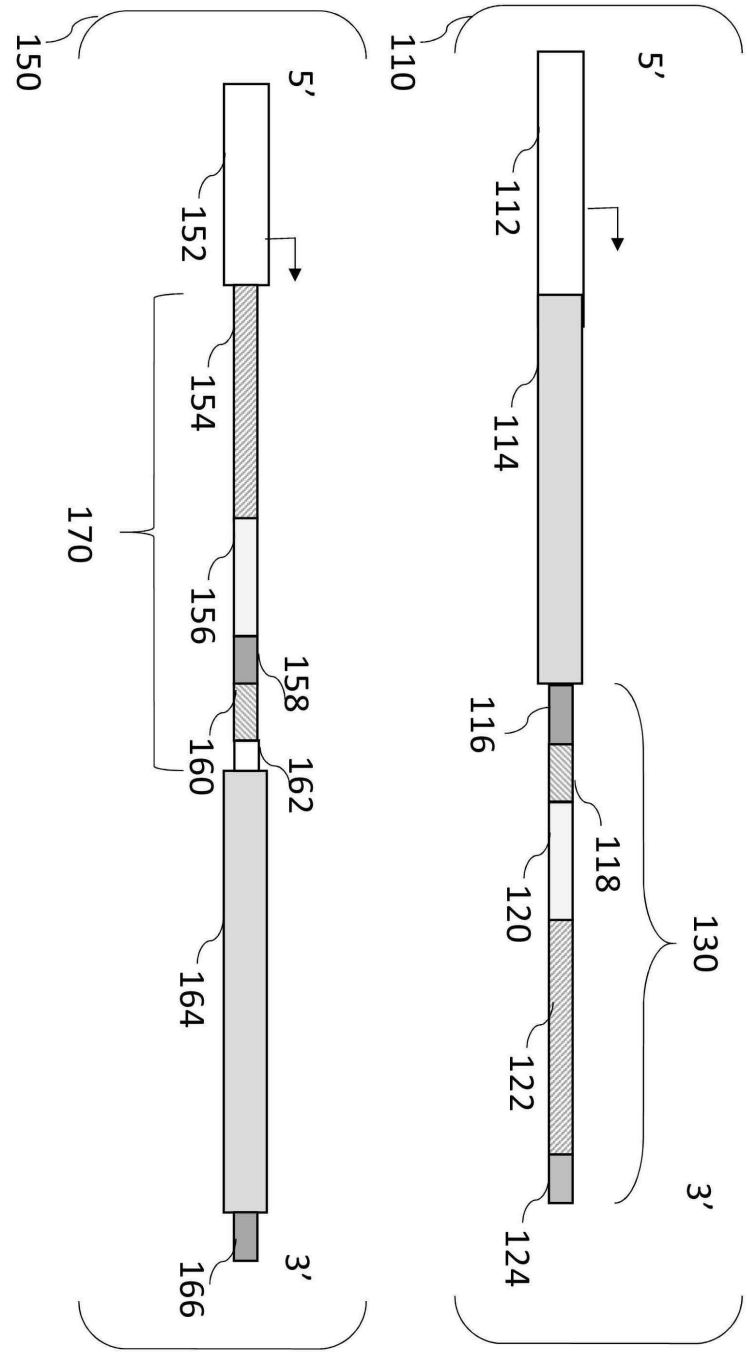
도면5e



도면5f



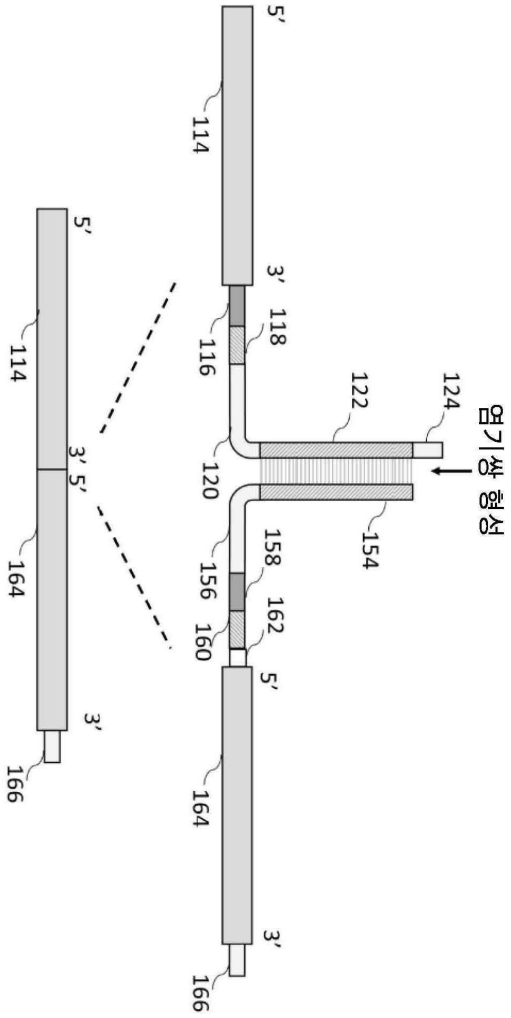
도면6a


150
5′
152
154
170
156
158
160
162
164
3′
166
110
5′
112
114
116
118
120
130
122
3′
124

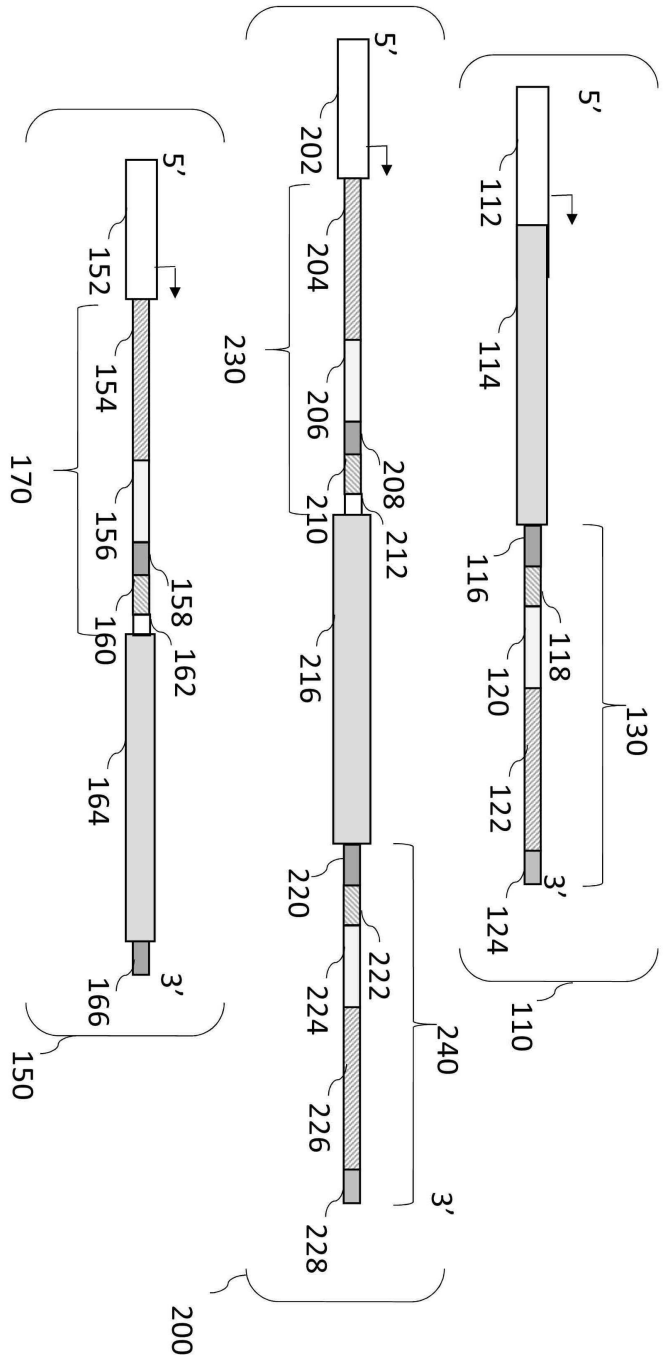
도면6b



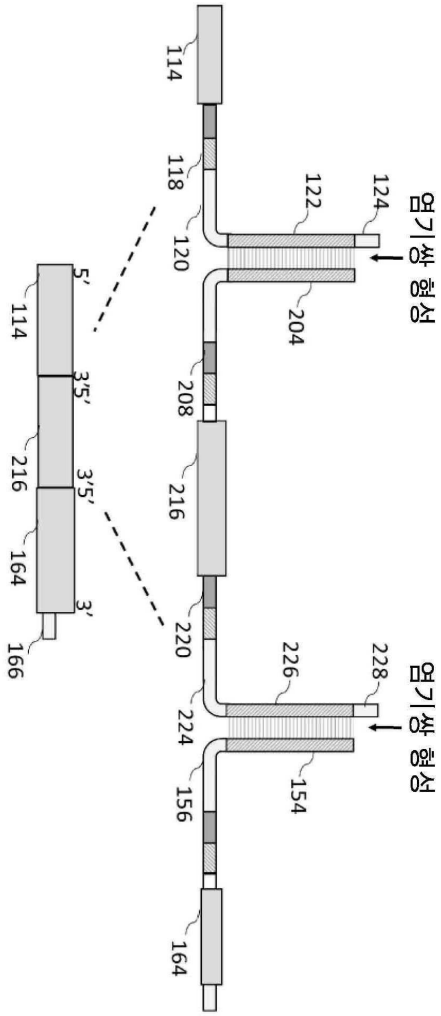
도면6c


5′
114
3′
118
116
122
124
염기쌍 형성
120
154
158
156
162
160
5′
164
3′
166
5′
114
3′ 5′
164
3′
166

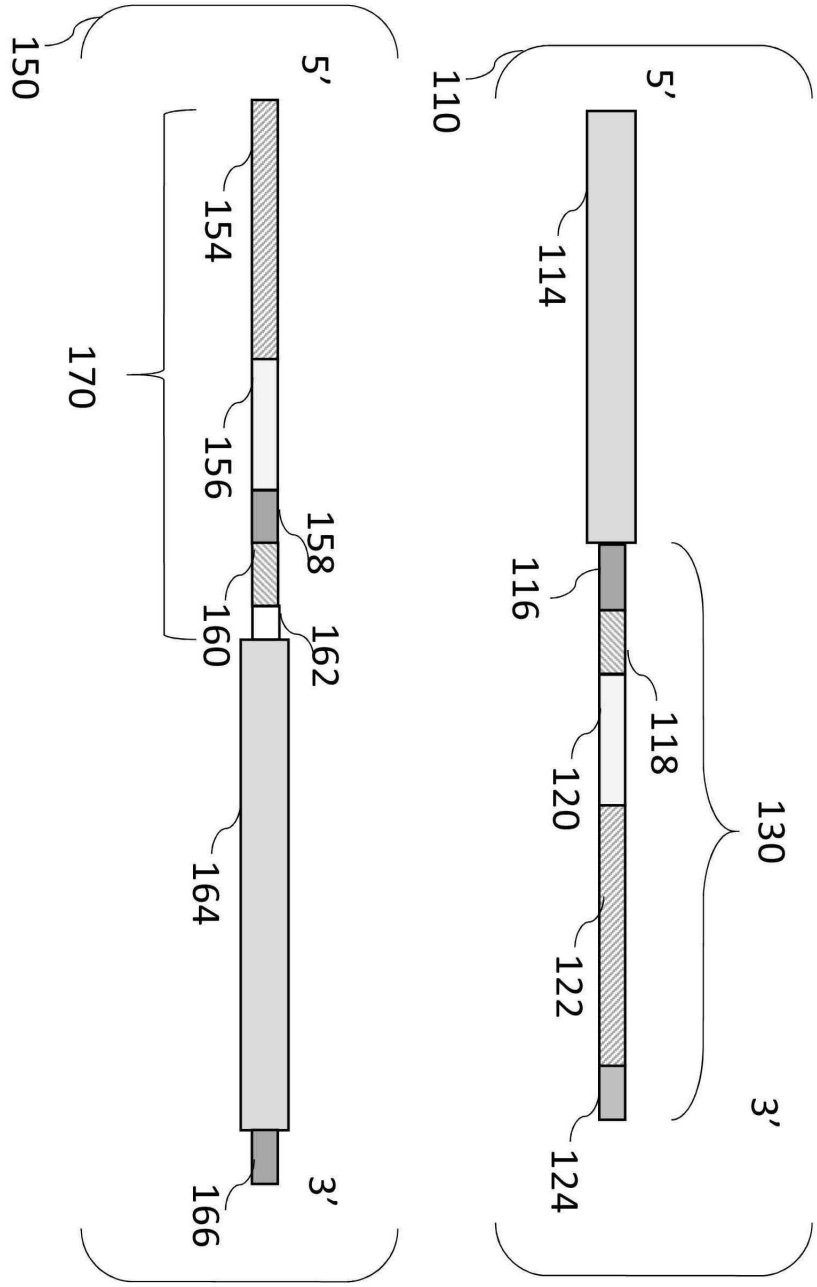
도면6d


5′
112
114
116
118
120
122
124
130
3′
110
5′
202
204
206
208
210
212
216
220
222
224
226
228
230
240
3′
200
5′
152
154
156
158
160
162
164
166
3′
170
150

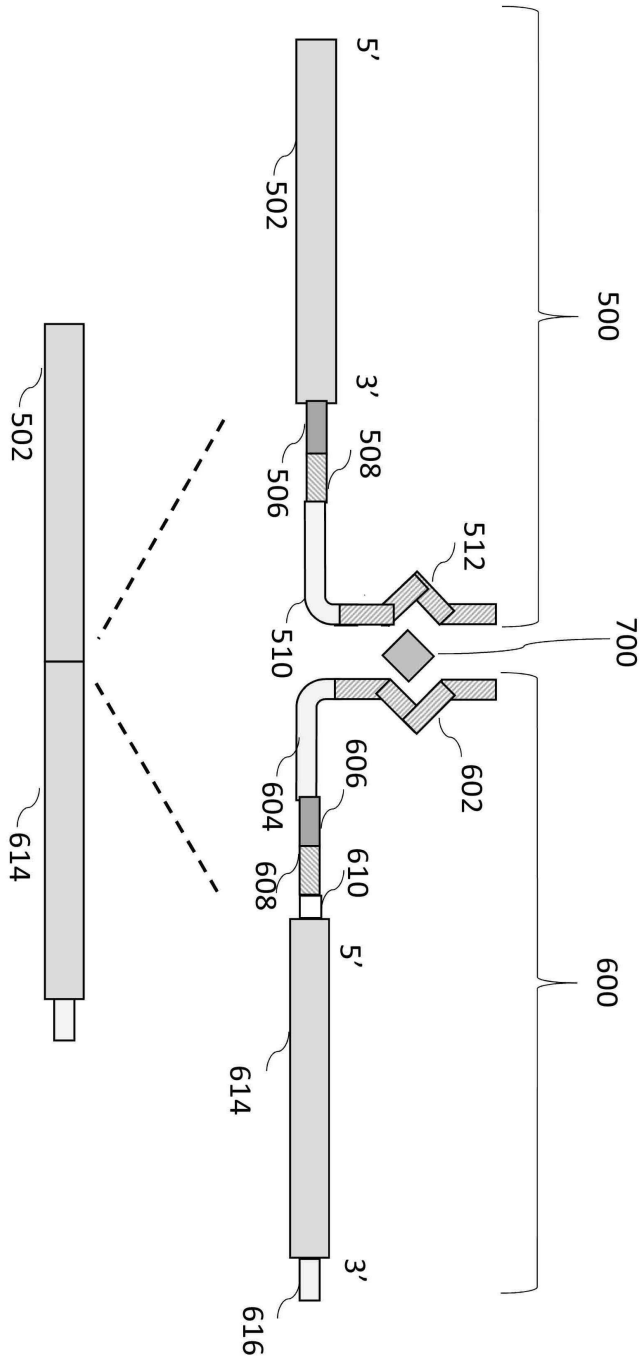
도면6e


114
118
120
122
124
염기쌍 형성
204
208
216
220
226
228
염기쌍 형성
224
156
154
164
5′
114
3′5′
216
3′5′
164
3′
166

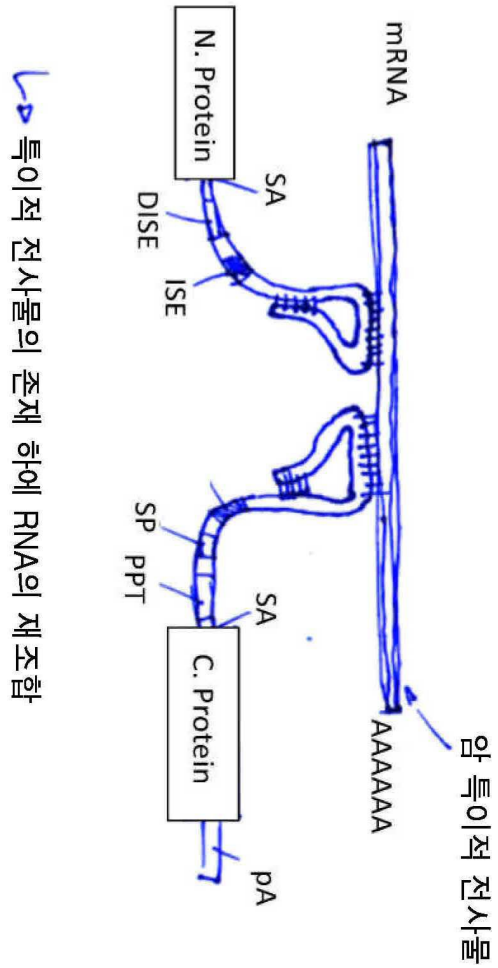
도면6f


150
5'
154
170
156
158
160
162
164
3'
166
110
5'
114
116
118
120
130
122
3'
124

도면7a

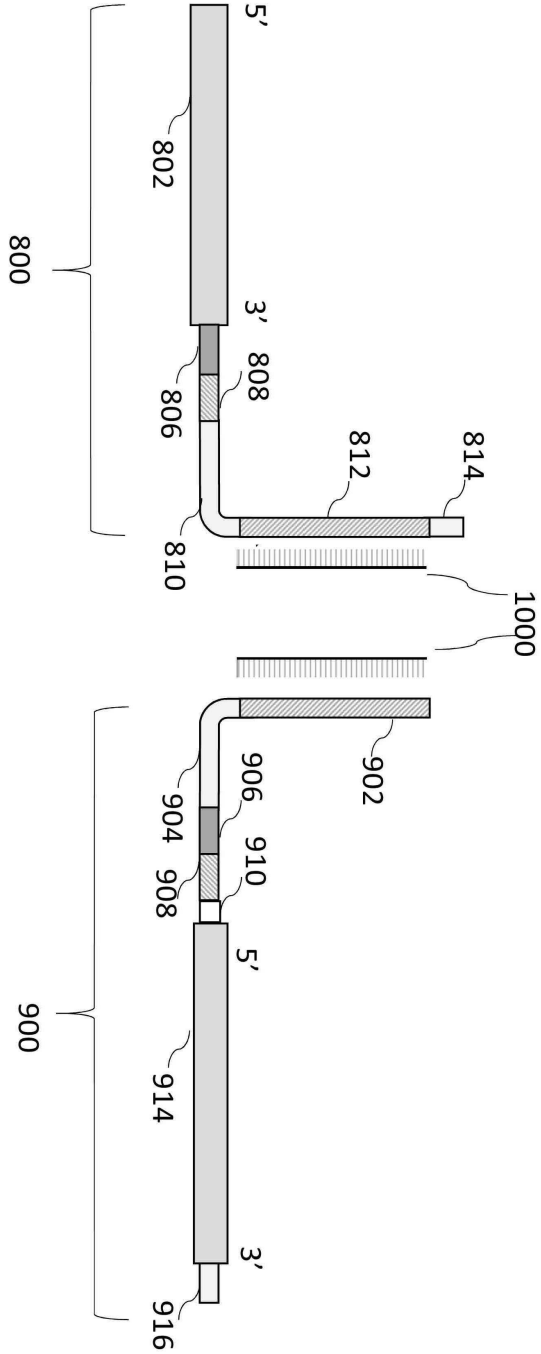

5′
502
500
3′
508
506
512
510
700
502
602
606
604
614
608
610
5′
600
614
3′
616

도면7b

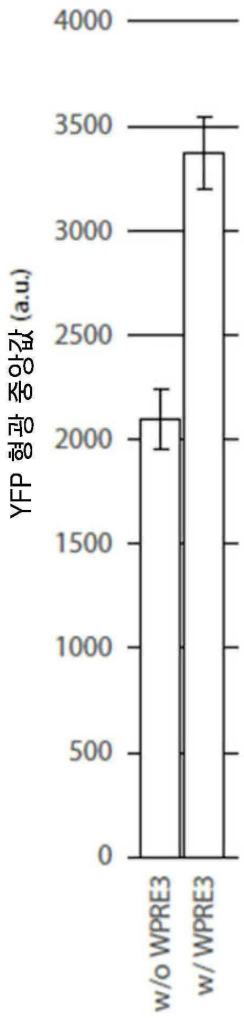


특이적 전사물의 존재 하에 RNA의 재조합

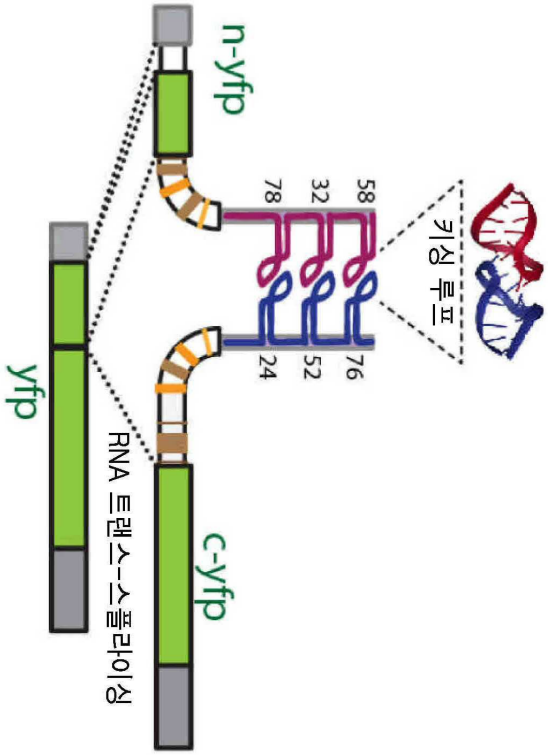
도면7c


800
802
5'
3'
806
808
810
812
814
1000
900
902
904
906
908
910
914
916

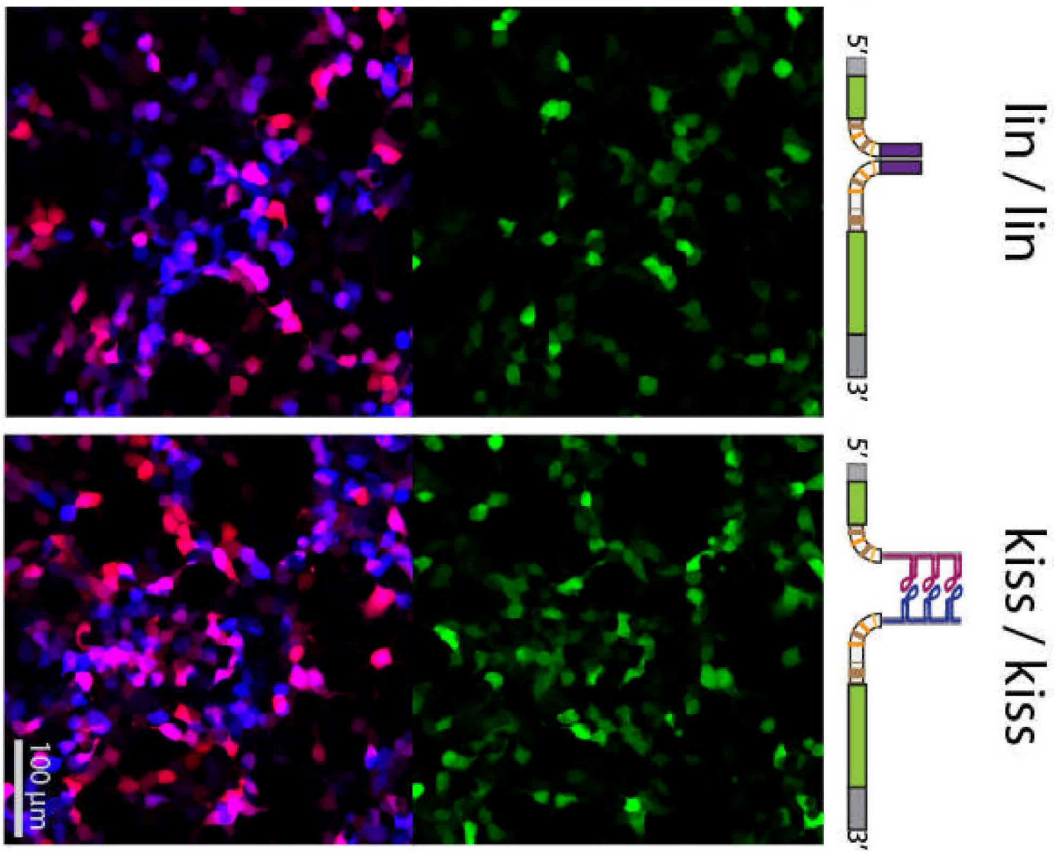
도면8


4000
3500
3000
2500
2000
1500
1000
500
0
YFP 형광 중앙값 (a.u.)
w/o WPRE3
w/ WPRE3

도면9a


n-yfp
58
32
78
키싱 루프
76
52
24
yfp
RNA 트랜스-스플라이싱
c-yfp

도면9b


lin / lin
kiss / kiss
5′
3′
5′
3′
100 μm

도면10a

CMV-N.YFP-5'.Synthetic.Intron-DimerizationDomain(Loop2)-pA

```
101 GCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCCTGGAGGGGTGGAGTCGTGACCTAGGcgttacataact 200
    CMV enhancer
    ITR
201 tacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagggactttccat 300
    CMV enhancer
301 tgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggta 400
    CMV enhancer
401 aatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcg 500
    CMV enhancer
    CMV promoter
501 gttttggcagtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaa 600
    CMV promoter
601 aatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctTGGA 700
    CMV promoter
    SV40int spacer
    TATA
701 TGTTGCCTTTACTTCTAGGCGCGCCgccaccatggtgagcaagggcgaggagctgttcaccggggtggtgcccatcctggtcgagctggacggcgacgta 800
    citrine
    SV40int spacer
    putative transcription start site
801 aacggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgc 900
    citrine
901 cctggcccaccctcgtgaccaccttcggctacggcctgatgtgcttcgcccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcc 1000
    citrine
1001 cgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgaggtgaagttcgagggcgacaccctggtgaaccgc 1100
    citrine
1101 atcgagctgaagggcatcgacttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggccgaca 1200
    citrine
1201 agcagaagaacggcatcaagGTAAGTATTagctctttctttccatgggttggcctcgccgcgtgGGCTGAGGGAAGGACTGTCCTGGGACTGGACAGGCG 1300
    citrine    ratFgfr2 DISE    cTNT (chicken cardiac troponin T ISE)
    SD consesus    ISE (Human molecular genetics 7.9 (1998): 1491-1496)
    GGGttatGGGACC
```

도면10b

CMV-N.YFP-5'.Synthetic.Intron-DimerizationDomain(Loop2)-pA
SD
boxBr
ISE (Human molecular genetics 7.9 (1998): 1491-1496)
LoopDim2R
StemI
B1
L2
StemII
StemIII
3'HHrz (leaving 6xT-GCGGTC)
Cleavage site (-1)
3'HHrz (leaving 6xA-GCGGTC)
SaCas9 sgRNA scaffold
ITR
KpnI
polyA (pGL4.26)

도면10c

CMV-DimerizationDomain(Loop2)-3'.Synthetic.Intron-C.YFP-pA

도면10d

CMV-DimerizationDomain(Loop2)-3'.Synthetic.Intron-C.YFP-pA
Branchpoint
PPT
ISE (Human molecular genetics 7.9 (1998): 1491-1496)
GGGttatGGGACC
cTNT (chicken cardiac troponin T ISE)
SA
citrine
SA
PPT
citrine
STOP
SV40 polyA
AATAAA
SV40 polyA

도면10e

도면10f

LCB290 - simpleSD-gfphybrid
SD (consensus)
spacer
SacII
EcoRV
eGFP homology
Lox71
XbaI

도면10g

LCB291 - simpleSA-gfphybrid

```
      *   10    *   20    *   30    *   40    *   50    *   60    *   70    *   80    *   90    *   100
      *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
1   ATGCATataacttcgtataGCATACATtatagcaatttatAGCTAGCgtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgagggg 100
                                            <<<<<<<
                                              NheI
                                                V  E  L  D  G  D  V  N  G  H  K  F  S  V  S  G  E  G
                                                1
                                            >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                                            eGFP homology
                                  <<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<<
                                     LoxJTZ17
                                  <<<<<<
                                  NsiI
      *   110   *   120   *   130   *   140   *   150   *   160   *   170   *   180   *   190   *   200
      *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
101 cgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccaccctcgtgaccaccctgacctac 200
19    E  G  D  A  T  Y  G  K  L  T  L  K  F  I  C  T  T  G  K  L  P  V  P  W  P  T  L  V  T  T  L  T  Y
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      eGFP homology
      *   210   *   220   *   230   *   240   *   250   *   260   *   270   *   280   *   290   *   300
      *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
201 ggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttCCG 300
                                                                                                    >>>
                                                                                                    SacII
    G  V  Q  C  F  S  R  Y  P  D  H  M  K  Q  H  D  F  F  K  S  A  M  P  E  G  Y  V  Q  E  R  T  I
52  >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
    eGFP homology
                                                                                                    >>>
                                                                                                    spacer
      *   310   *   320   *   330   *   340   *   350   *   360   *
      *    *    *    *    *    *    *    *    *    *    *    *    *
301 CGGAACATTATTATAACGTTGCTCGAATACTAACTGGTACCTCTTCTTTTTTTTTTGATATCTGCAG 367
    >>>                                                         >>>
    SacII                                                       3'SpliceSite
                                                             >>
                                                             PPT
                                                        <<<<<<
                                                        EcoRV
                                                  <<<<<<<
                                                  intronic
                                            >>>>>>>>>>
                                            >>>>>>>>>>
                                      Branchpoint (BP)
                                  >>>>>>>>>>>>>>>>>>>>
                                  PPT
                  >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
                  spacer
```

도면10h

LCB300 - boxBloops with ISE
putative transcription start site
ISE (Human molecular genetics 7.9 (1998): 1491-1496)
Start(out-of-frame-decoy)

도면10i

LCB414 - 3′intron 3xISE without hybridization domain

도면10j

LCB470
SD consesus
rat rGrz DISE
cTNNT (chicken cardiac troponin T ISE)
ISE (Human molecular genetics 7.9 (1998): 1491-1496)
spacer
boxBr
EcoRV
LoopDim2R
XPD DR36
StemI
Cleavage site (-1)
L1
3'HHrz (leaving 6xA-GCGGTC)
3'HHrz (leaving 6xT-GCGGTC)
StemII
StemIII
B1
L2
gBlock-SA-sg-HDV
SaCas9 sgRNA scaffold
KpnI
SapI.gBlock.splitter.LCB

도면10k

LCB469
putative transcription start site
LoopDim2R
ISE (Human molecular genetics 7.9 (1998): 1491-1496)
cTNT (chicken cardiac troponin T ISE)
spacer
boxBr
EcoRV
ClaI
FPT
SA

도면101

도면10m

LCB571
ratFgfr2 DISE
SD consesus
spacer
boxBr
EcoRV
LoopDim2R
StemI
XPD DR36
Cleavage site (-1)
3'HHrz (leaving 6xA-GCGGTC)
StemIII
StemII
B1
StemII
gBlock-SA-sg-HDV
L2
KpnI
SaCas9 sgRNA scaffold
3'HHrz (leaving 6xT-GCGGTC)
SapI.gBlock.splitter.LCB

도면10n

LCB572
SD consesus
ratFgfr2 DiSE
cTNT (chicken cardiac troponin T ISE)
ISE (Human molecular genetics 7.9 (1998): 1491-1496)
Spacer
GGGttatGGGACC
boxBr
LoopDim2R
EcoRV
KpnI
SapI.gBlock.splitter.LCB
EcoRV
LoopDim2R

도면10o

LCB573
AscI
spacer
SV40int
StuI AGG^CCT
spacer
SacII
EcoRV
LoopDim2R
5452_1stCreF
5480_SV40f
5731_347P1F
boxBr
EcoRV
spacer
ClaI
PPT
Branchpoint
SA
ISE (Human molecular genetics 7.9 (1998): 1491-1496)

도면10p

LCB605
deconserverd split point
EcoRV
LoopDim2R
overhang
U12atac-consensus
hNOP2i6
gBlock-SA-sg-HDV
SaCas9 sgRNA scaffold
3'HHrz (leaving 6xT-GCGGTC)
KpnI
StemIII
StemIII
StemIII
StemII
StemII
L2
StemI
L1
StemI
StemII
SapI.gBlock.splitter.LCB
3'HHrz (leaving 6xA-GCGGTC)
Cleavage_site (-1)
XPD DR36
B1
XPD DR36
EcoRV
StemI

도면10q

LCB606
LoopDim2.1
AscI
SV40int spacer
StuI AGG^CCT
putative transcription start site
SacII
5452_1stCreF
5480_SV40f
5731_347PIF
EcoRV
hNOP2i6
U12-branchpoint (consensus)
Branchsite
deconserverd split point
overhang
U12atac-acceptor(consensus)

도면10r

LCB610 N.Abca4
putative transcription start site
New Feature
SV40int spacer
Abca4 (co,sd)
AscI

도면10s

도면10t

N.Abca4-continued

도면10u

N.Abca4-continued

도면10v

C.Abca4

도면10w

C. Abca4

도면10x

C.Abca4 continued

```
      *   1810    *   1820    *   1830    *   1840    *   1850    *   1860    *   1870    *   1880    *   1890    *   1900
1801  GTGCTGTACCTGATTCAAGAGAGAGTGAACAAGAGCAAGCACCTCCAGTTCATCTCCGGGGTGTCCCCAACCACCTACTGGGTCACCAATTTTCTGTGGG 1900
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Abca4 (co,sd)
      *   1910    *   1920    *   1930    *   1940    *   1950    *   1960    *   1970    *   1980    *   1990    *   2000
1901  ACATCATGAACTACAGCGTGTCAGCCGGCCTGGTCGTGGGCATCTTTATCGGCTTTcaaAAGAAGCCTACACGAGCCCCGAGAACCTGCCTGCTTTGGT 2000
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Abca4 (co,sd)
      *   2010    *   2020    *   2030    *   2040    *   2050    *   2060    *   2070    *   2080    *   2090    *   2100
2001  TGCTCTGCTGCTCCTGTATGGCTGGGCCGTGATTCCCATGATGTACCCCGCCAGCTTTCTGTTTGACGTGCCCAGCACAGCCTACGTGGCCCTGTCTTGC 2100
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Abca4 (co,sd)
      *   2110    *   2120    *   2130    *   2140    *   2150    *   2160    *   2170    *   2180    *   2190    *   2200
2101  GCCAATCTGTTCATCGGCATCAACAGCAGCGCCATCACATTCATCCTGGAACTGTTCGAGAACAGGAGCCCTGCTGCGGTTCAACGCCGTGCTGCGGA 2200
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Abca4 (co,sd)
      *   2210    *   2220    *   2230    *   2240    *   2250    *   2260    *   2270    *   2280    *   2290    *   2300
2201  AACTGCTGATCGTGTTCCCTCACTTCTGTCTCGGCcggGGCCTGATCGACCTGGCTCTGTCTCaaGCCGTGACCGATGTGTACGCCAGATTTGGCGAGGA 2300
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Abca4 (co,sd)
      *   2310    *   2320    *   2330    *   2340    *   2350    *   2360    *   2370    *   2380    *   2390    *   2400
2301  ACACTCCGCCAATCCATTCCACTGGGACCTGATCGGCAAGAACCTGTTCGCCATGGTGGTGGAAGGCGTCGTGTACTTCCTGCTCACTCTGCTGGTGCAG 2400
      >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
      Abca4 (co,sd)
                  >>>>>>>>>>>>>>>>>>>
                  overhang
```

도면10y

C.Abca4 continued

```
              2410      2420      2430      2440      2450      2460      2470      2480      2490      2500
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
2401 AGACACTTTTTCTGTCCCAATGGATCGCCGAGCCTACCAAAGAACCCATTGTGGACGAGGACGACGATGTGGCCGAGGAAAGACAGAGAATCATCACCG 2500
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

              2510      2520      2530      2540      2550      2560      2570      2580      2590      2600
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
2501 GCGGCAACAAGACCGATATCCTGAGACTGCACGAGCTGACAAAGATCTACCCCGGCACAAGCTCCCCAGCCGTGGATAGGCTTTGTGTGGGAGTTAGACC 2600
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

              2610      2620      2630      2640      2650      2660      2670      2680      2690      2700
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
2601 CGGCGAGTGCTTTGGCCTGCTGGGAGTTAATGGCGCCGGAAAGACCACCACCTTCAAGATGCTGACCGGCGACACCACAGTGACAAGCGGAGATGCTACA 2700
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

              2710      2720      2730      2740      2750      2760      2770      2780      2790      2800
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
2701 GTGGCCGGCAAGAGCATCCTGACCAACATCAGCGAAGTGCATCAGAACATGGGCTACTGCCCTCAGTTCGACGCCATCGACGAACTGCTGACAGGCCGCG 2800
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

              2810      2820      2830      2840      2850      2860      2870      2880      2890      2900
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
2801 AACACCTGTATCTGTATGCCAGACTGAGAGGCGTGCCCGCTGAAGAGATCGAGAAGGTGGCCAACTGGTCCATCAAGTCTCTGGGCCTGACAGTGTACGC 2900
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

              2910      2920      2930      2940      2950      2960      2970      2980      2990      3000
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
2901 CGACTGTCTGGCCGGAACATACAGCGGAGGAAACAAGCGGAAGCTGAGCACCGCCATTGCTCTGATCGGATGCCCACCTCTGGTCCTGCTGGATGAACCC 3000
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

              3010      3020      3030      3040      3050      3060      3070      3080      3090      3100
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
3001 ACCACCGGAATGGATCCCCAGGCTAGAAGAATGCTCTGGAACGTGATCGTGTCTATCATCCGCGAGGGCAGAGCTGTGGTGCTGACCTCTCACtccATGG 3100
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

              3110      3120      3130      3140      3150      3160      3170      3180      3190      3200
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
3101 AAGAGTGCGAGGCTCTGTGTACCCGGCTGGCCATTATGGTCAAGGGCGCCTTCAGATGCATGGGCACCATTCAGCATCTGAAAAGCAAGTTCGGCGACGG 3200
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

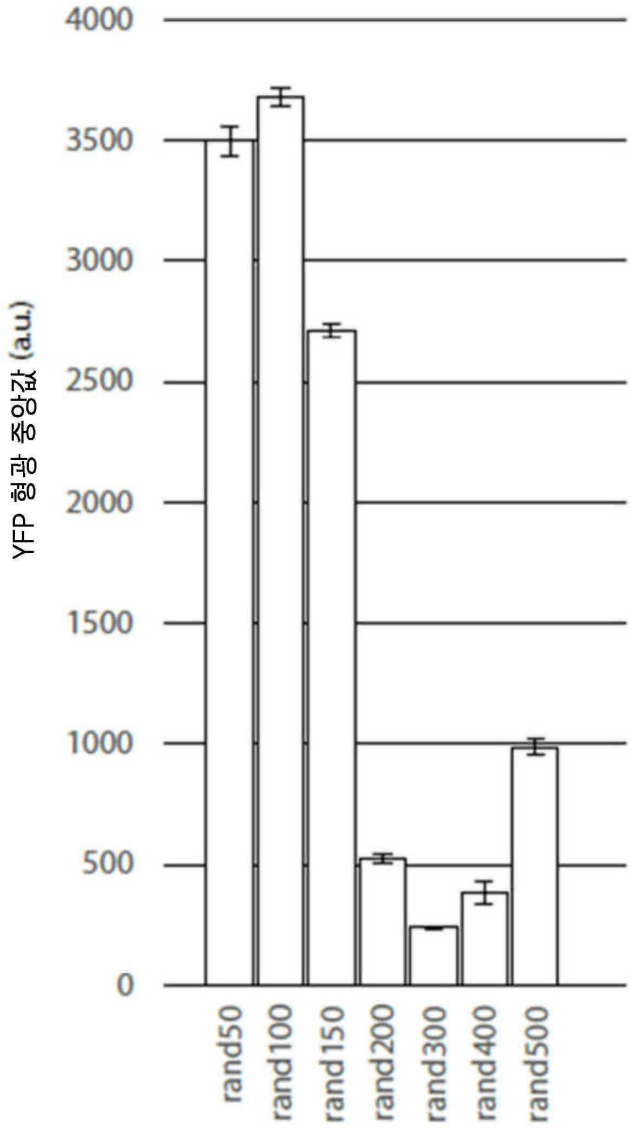
              3210      3220      3230      3240      3250      3260      3270      3280      3290      3300
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
3201 CTACATCGTGACAATGAAGATCAAGAGCCCCAAGGACGACCTCCTGCCTGATCTGAACCCCGTGGAACAGTTTTTCAGGGCAACTTCCCCGGCTCCGTG 3300
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)

              3310      3320      3330      3340      3350      3360      3370      3380      3390      3400
         *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *    *
3301 CAGCGGGAAAGACACTATAACATGCTGCAGTTTCAGGTGTCCTCCTCCAGCCTGGCTCGGATCTTTCAACTGCTGCTCTCTCACAAGGACAGCCTGCTGA 3400
     >>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>>
     Abca4 (co,sd)
```

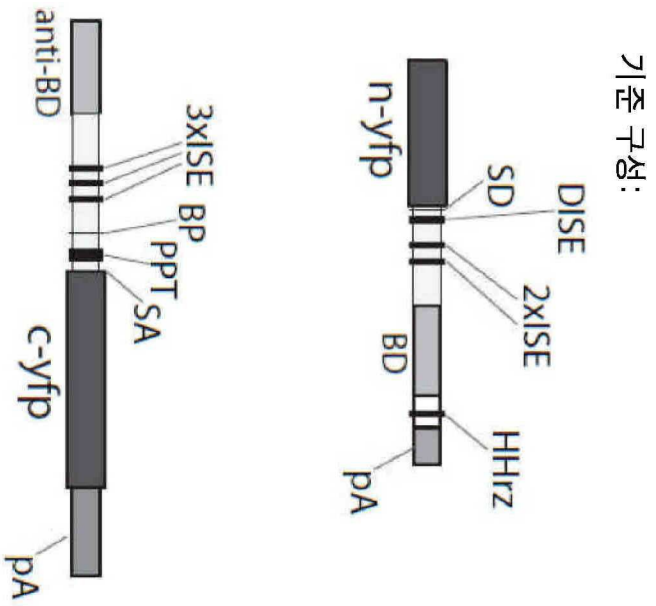
도면10z

C. Abca4 continued

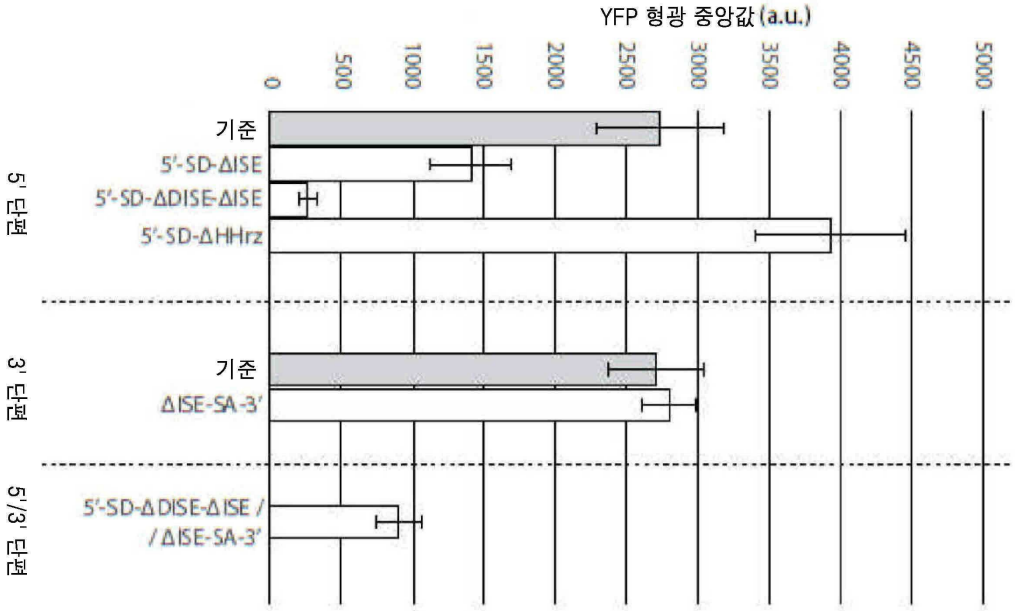
도면11


4000
3500
3000
2500
2000
1500
1000
500
0
YFP 형광 중앙값 (a.u.)
rand50
rand100
rand150
rand200
rand300
rand400
rand500

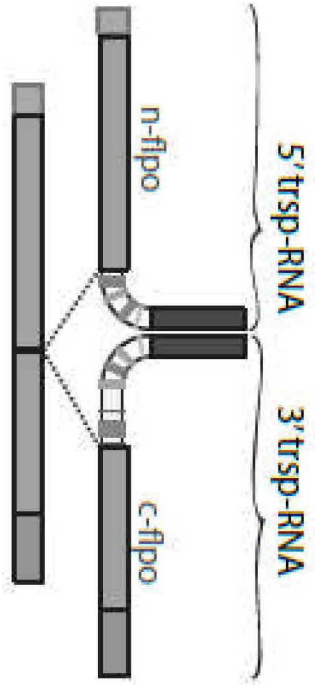
도면12a


anti-BD
3xISE
BP
PPT
SA
c-yfp
pA
기준 구성:
n-yfp
SD
DISE
2xISE
BD
HHrz
pA

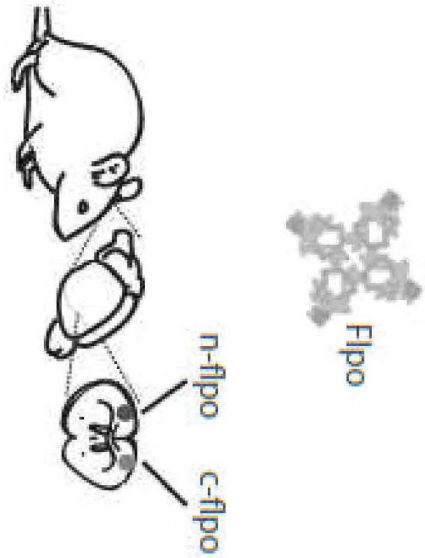
도면12b


YFP 형광 중앙값 (a.u.)
0
500
1000
1500
2000
2500
3000
3500
4000
4500
5000
기준
5'-SD-ΔISE
5'-SD-ΔDISE-ΔISE
5'-SD-ΔHHrz
5' 단편
기준
ΔISE-SA-3'
3' 단편
5'-SD-ΔDISE-ΔISE /
/ΔISE-SA-3'
5'/3' 단편

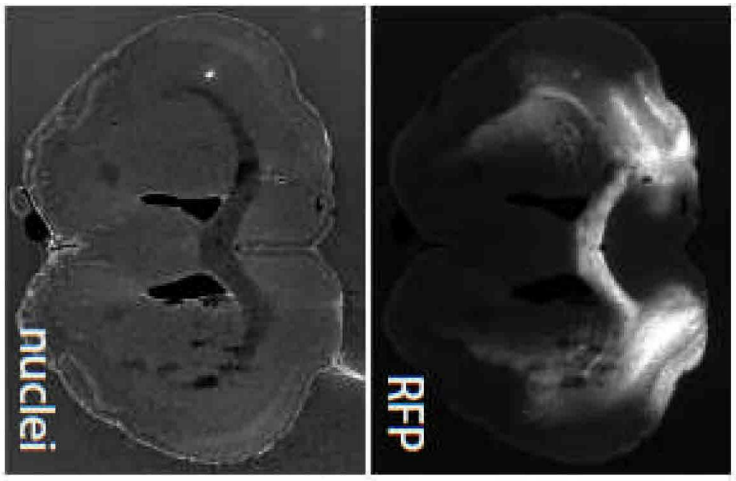
도면13a



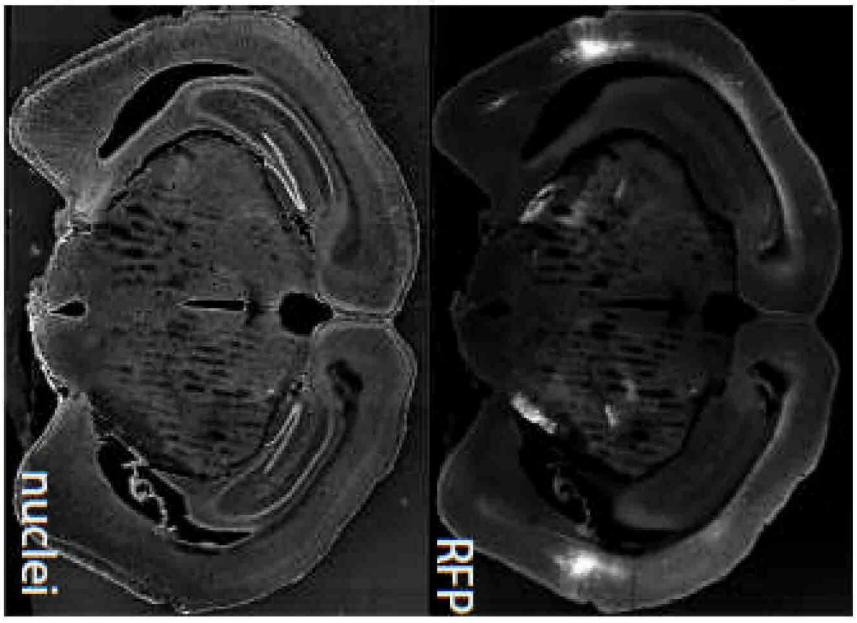
도면13b



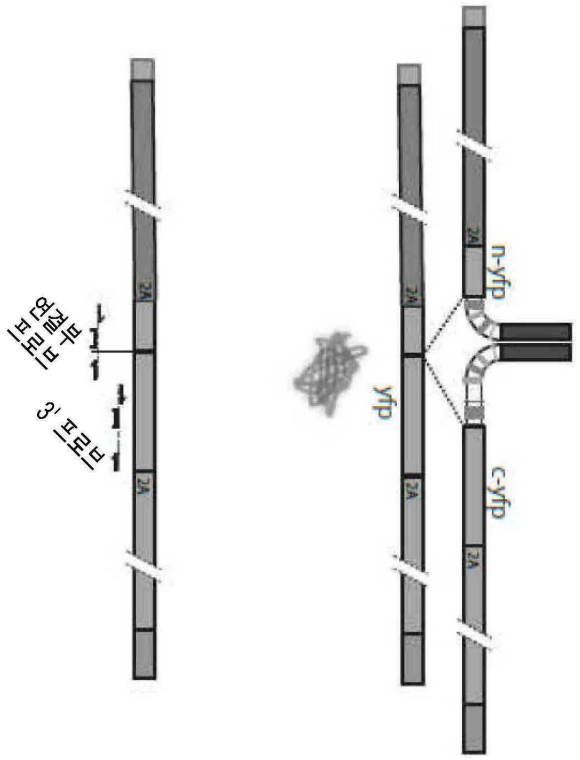
도면13c



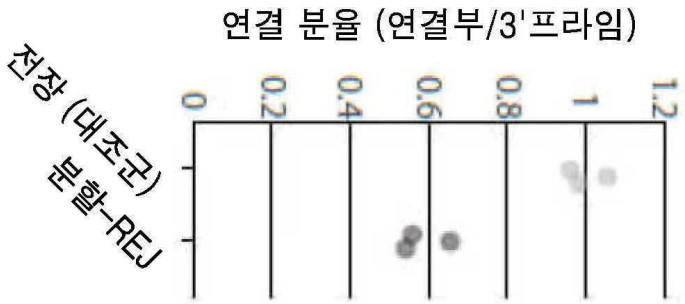
도면13d



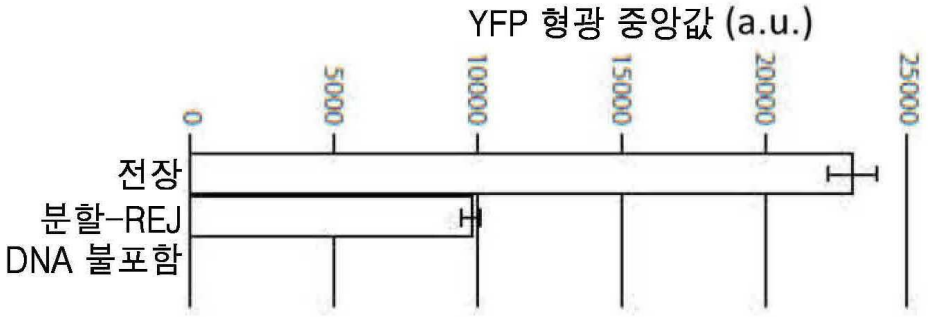
도면14a



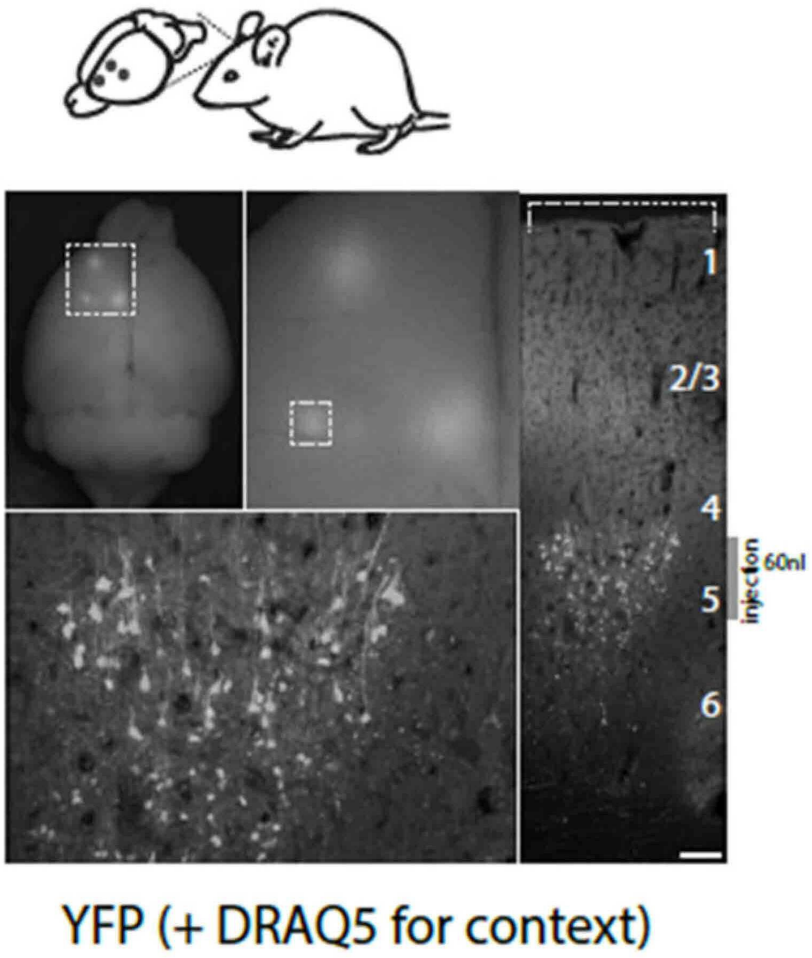
도면14b



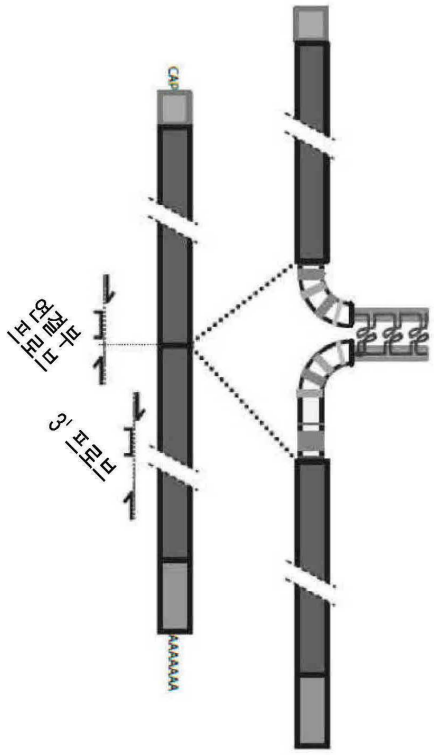
도면14c



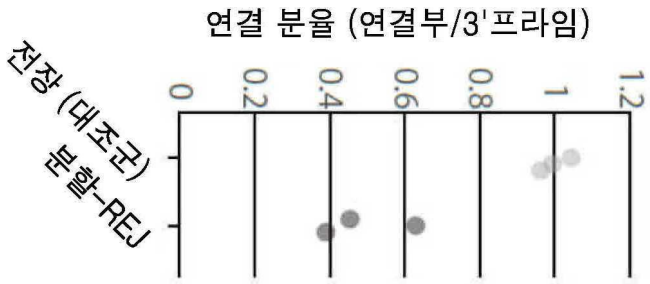
도면14d



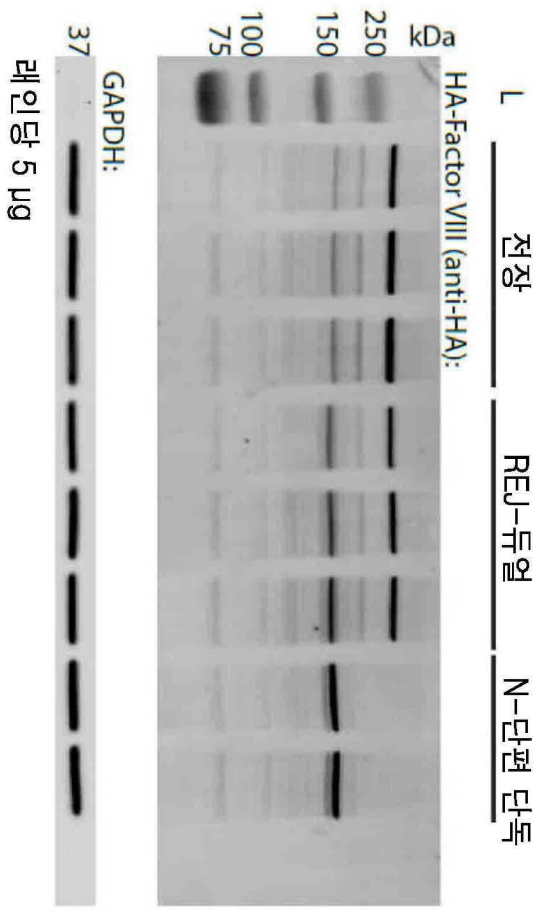
도면15a


CAP
연결부 프로브
3' 프로브
AAAAAAA

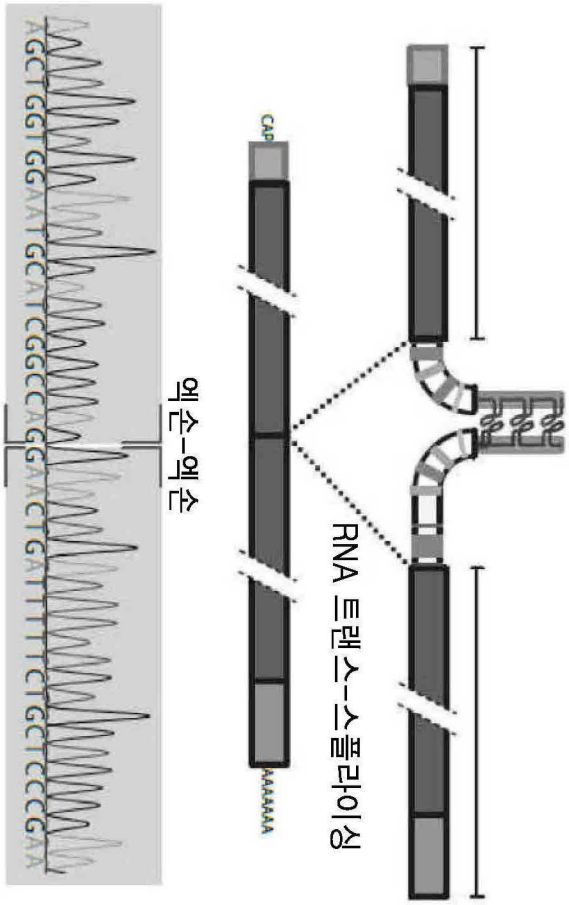
도면15b


연결 분율 (연결부/3'프라임)
전장 (대조군)
분할-REJ
0
0.2
0.4
0.6
0.8
1
1.2

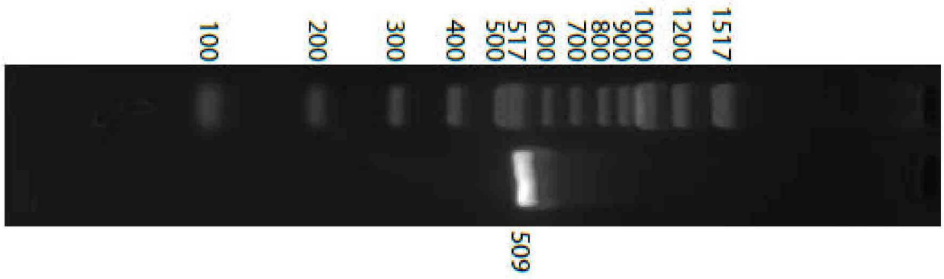
도면15c


L
전장
REJ-듀얼
N-단편 단독
kDa
250
150
100
75
HA-Factor VIII (anti-HA):
37
GAPDH:
레인당 5 μg

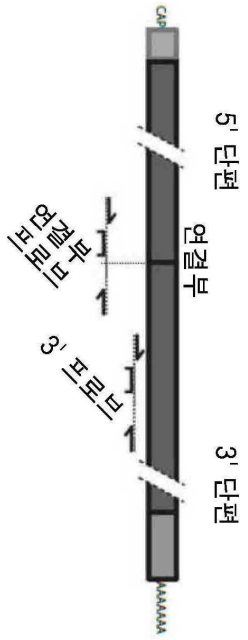
도면16a


엑손-엑손
RNA 트랜스-스플라이싱
CAP
AAAAAA

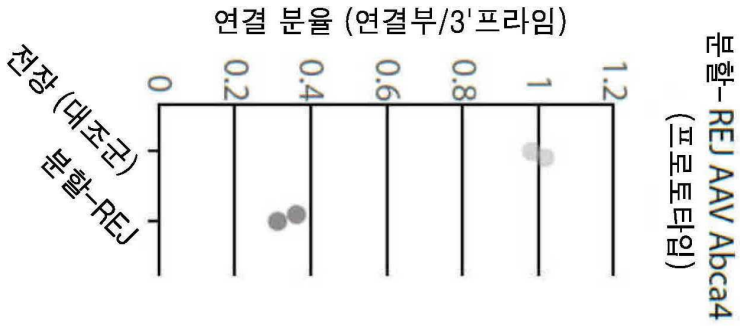
도면16b


100
200
300
400
500
517
600
700
800
900
1000
1200
1517
509

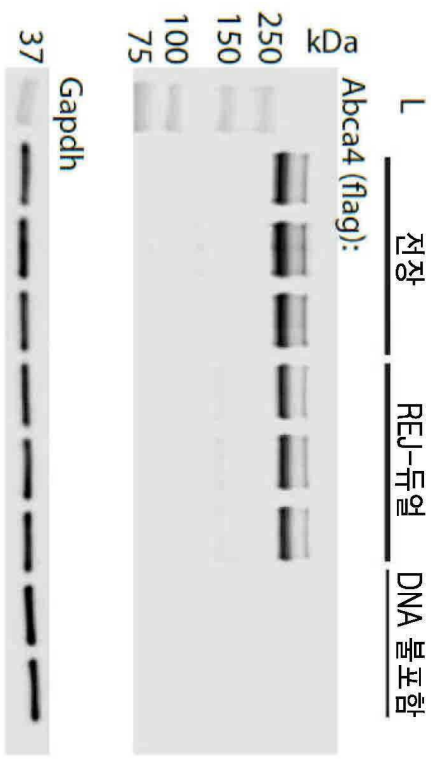
도면16c


5' 단편
연결부
연결부 프로브
3' 프로브
3' 단편

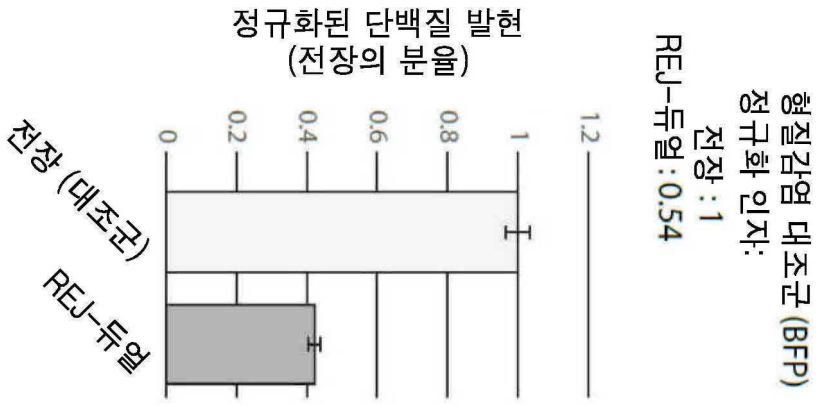
도면16d


연결 분율 (연결부/3'프라임)
0
0.2
0.4
0.6
0.8
1
1.2
전장 (대조군)
부활-REJ
부활-REJ AAV Abca4
(프로토타입)

도면16e


kDa
250
150
100
75
37
Abca4 (flag):
Gapdh
L
전장
REJ-듀얼
DNA 불포함

도면16f


형질감염 대조군 (BFP)
정규화 인자:
전장 : 1
REJ-듀얼 : 0.54
정규화된 단백질 발현
(전장의 분율)
0
0.2
0.4
0.6
0.8
1
1.2
전장 (대조군)
REJ-듀얼

도면17a

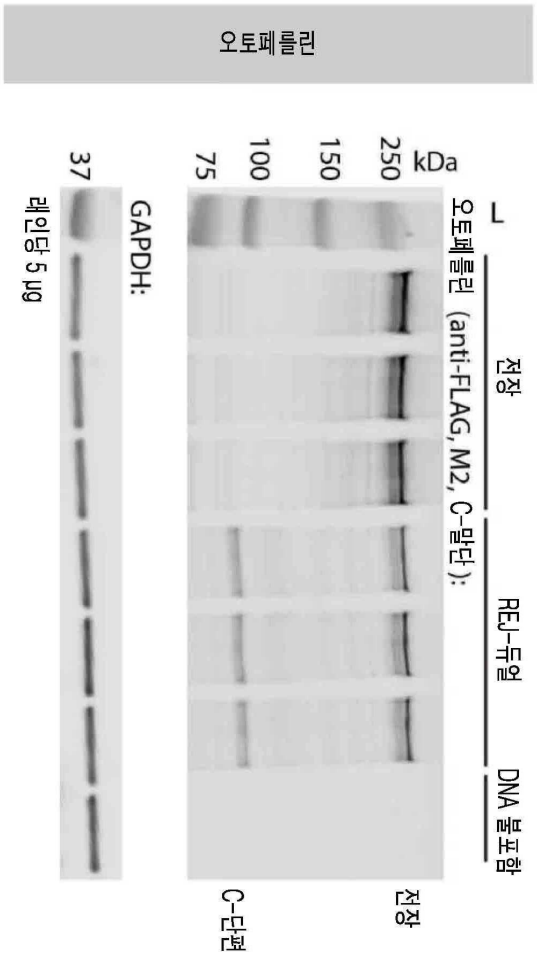
HIV-1 기반 키싱 루프 이합체화 도메인

도면17b

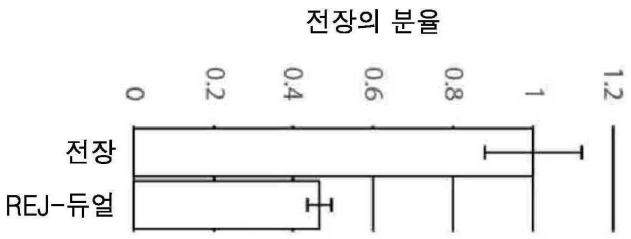
HIV-2 기반 키싱 루프 이합체화 도메인

```
5'-[...]GUCGAGGCCGAgcgggcAAAGGCCUCGAC[...]-3' (N-단편)
5'-[...]((((((((((.............)))))))))[...]-3'
          /////////   |||||   /////////
3'-[...]((((((((((.............)))))))))[...]-5'
3'-[...]CAGCUCCGGAAAcgcccgAGCCGGAGCUG[...]-5' (C-단편)
```

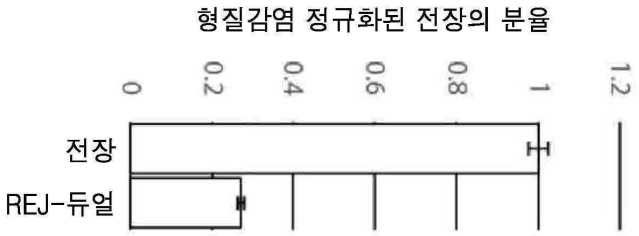
도면18a



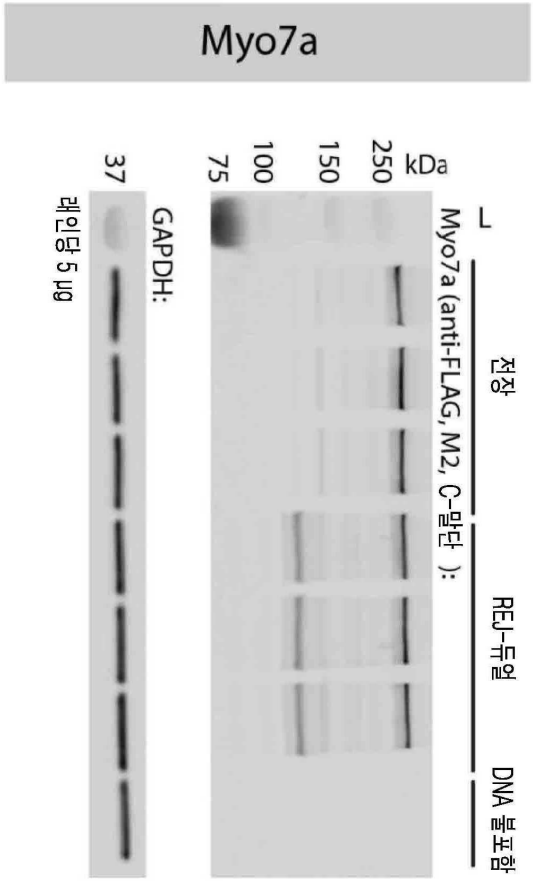
도면18b



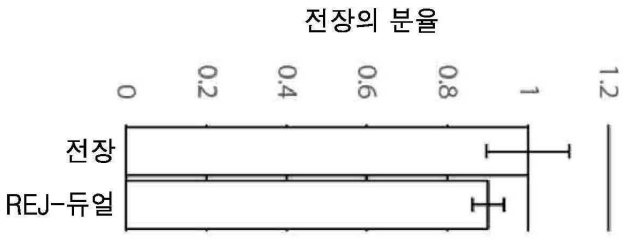
도면18c



도면19a



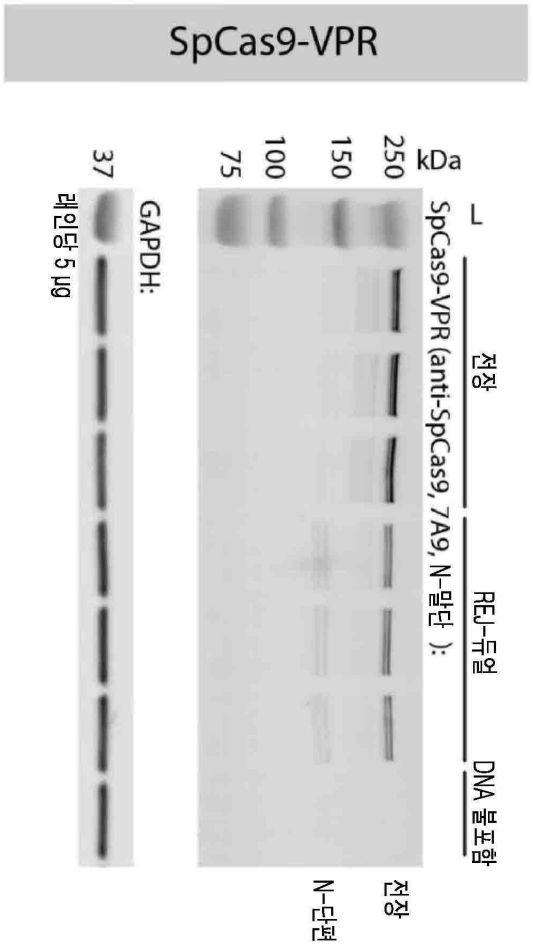
도면19b



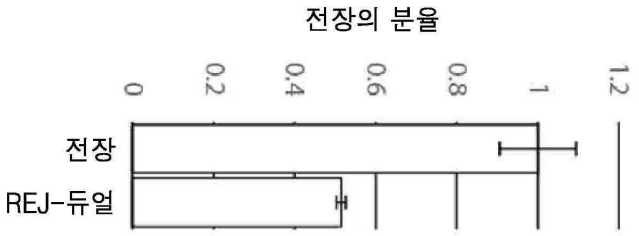
도면19c



도면20a



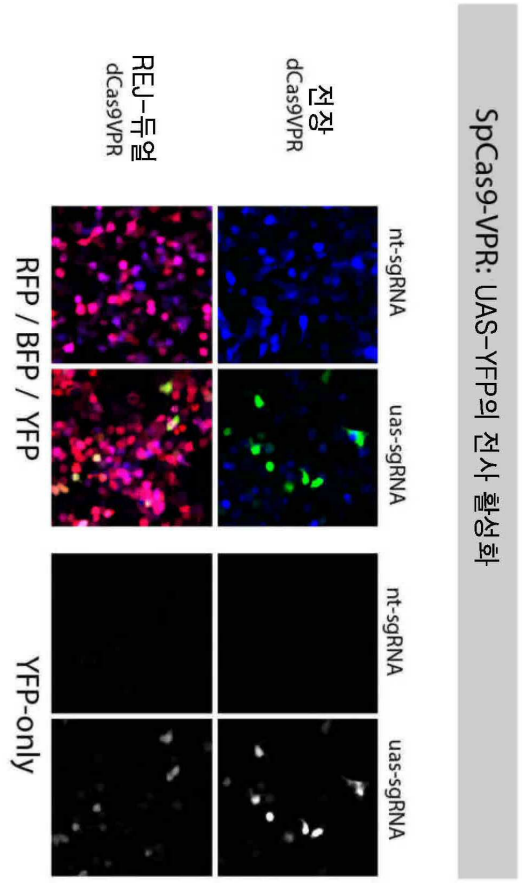
도면20b



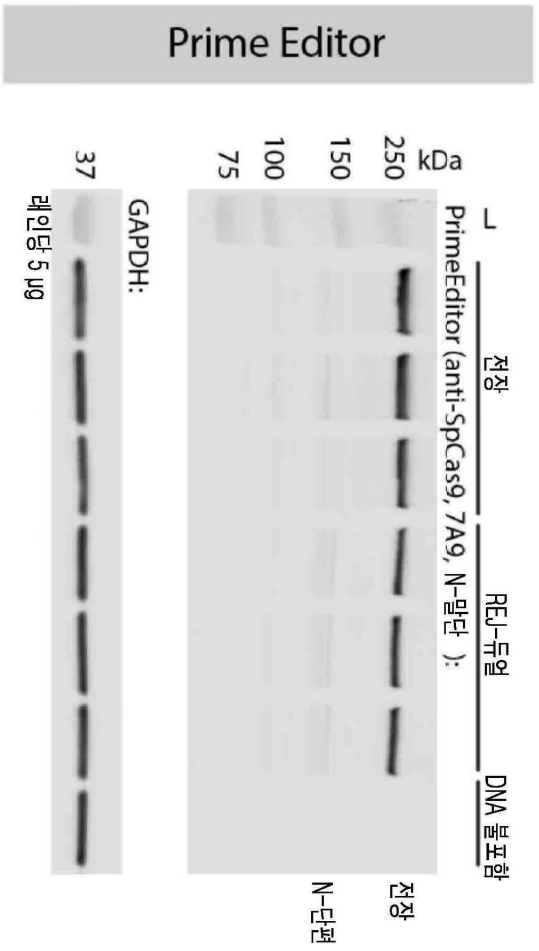
도면20c



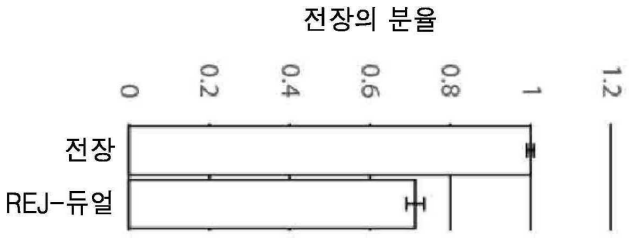
도면20d


SpCas9-VPR: UAS-YFP의 전사 활성화
RFP / BFP / YFP
YFP-only
전장
dCas9VPR
REJ-듀얼
dCas9VPR
nt-sgRNA
uas-sgRNA
nt-sgRNA
uas-sgRNA

도면21a



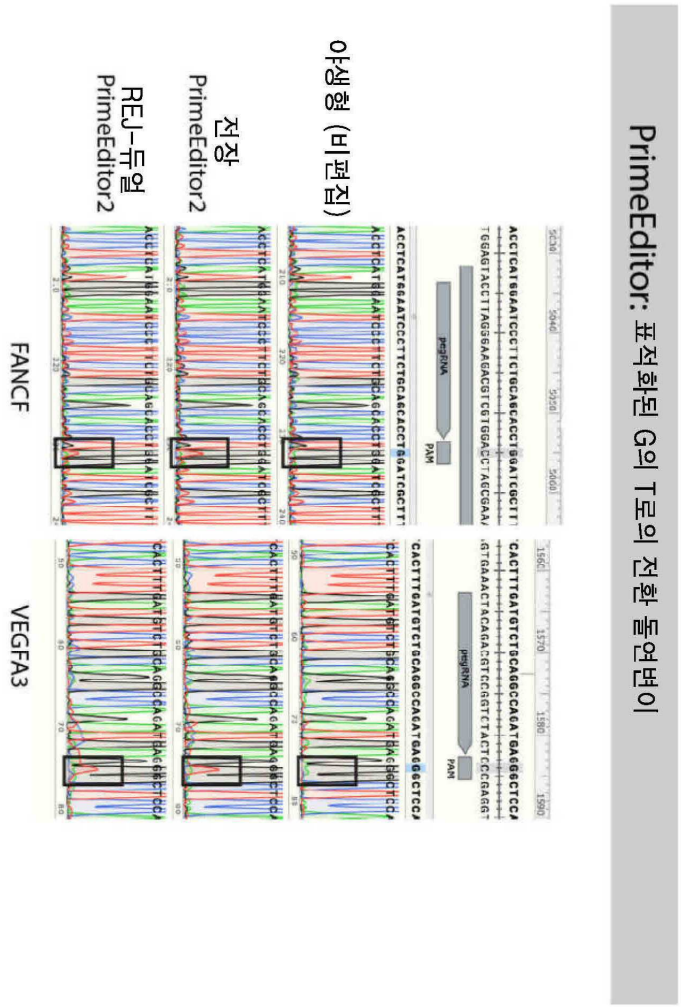
도면21b



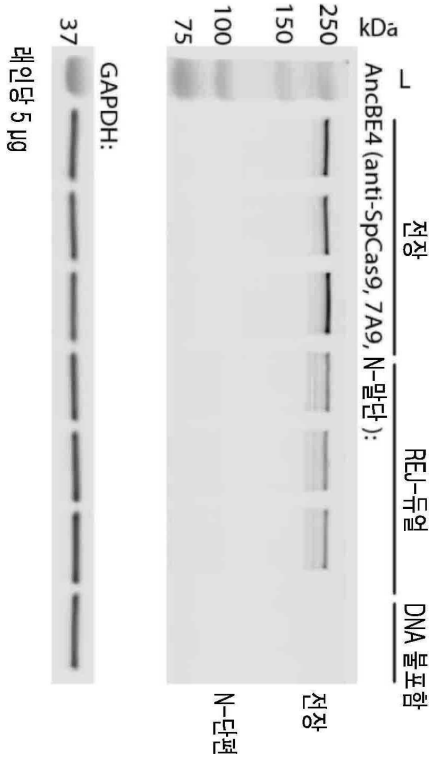
도면21c



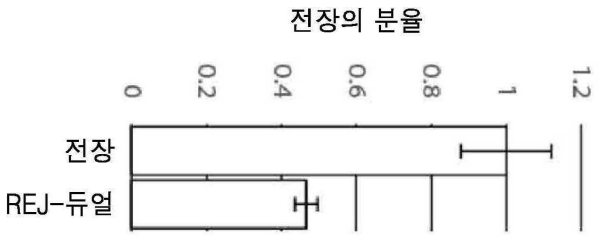
도면21d


PrimeEditor: 표적화된 G의 T로의 전환 돌연변이
야생형 (비편집)
전장
PrimeEditor2
REJ-듀얼
PrimeEditor2
FANCF
VEGFA3

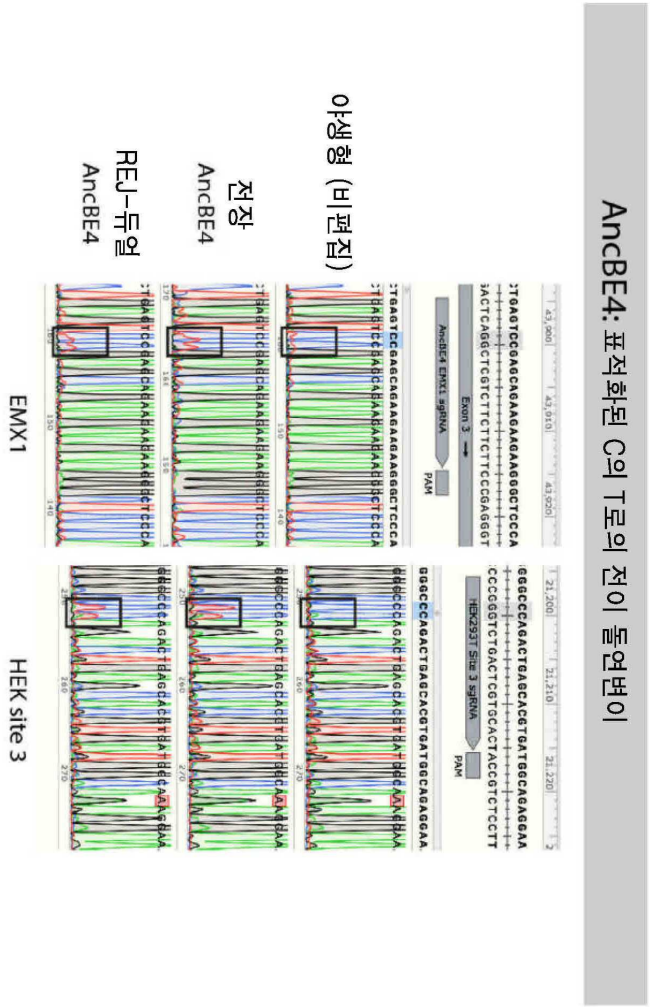
도면22a


kDa
250
150
100
75
37
AncBE4 (anti-SpCas9, 7A9, N-말단):
전장
REJ-듀얼
DNA 불포함
N-단편
전장
GAPDH:
레인당 5 μg

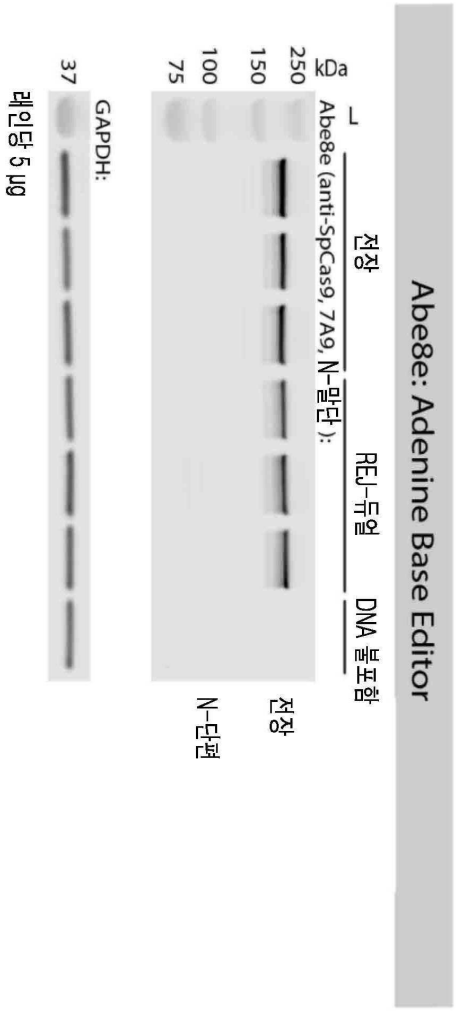
도면22b


전장의 분율
0
0.2
0.4
0.6
0.8
1
1.2
전장
REJ-듀얼

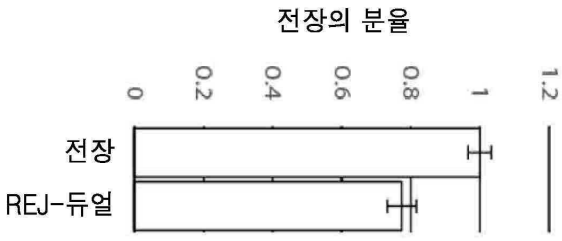
도면22c


AncBE4: 표적화된 C의 T로의 전이 돌연변이
야생형 (비편집)
전장
AncBE4
REJ-듀얼
AncBE4
EMX1
HEK site 3

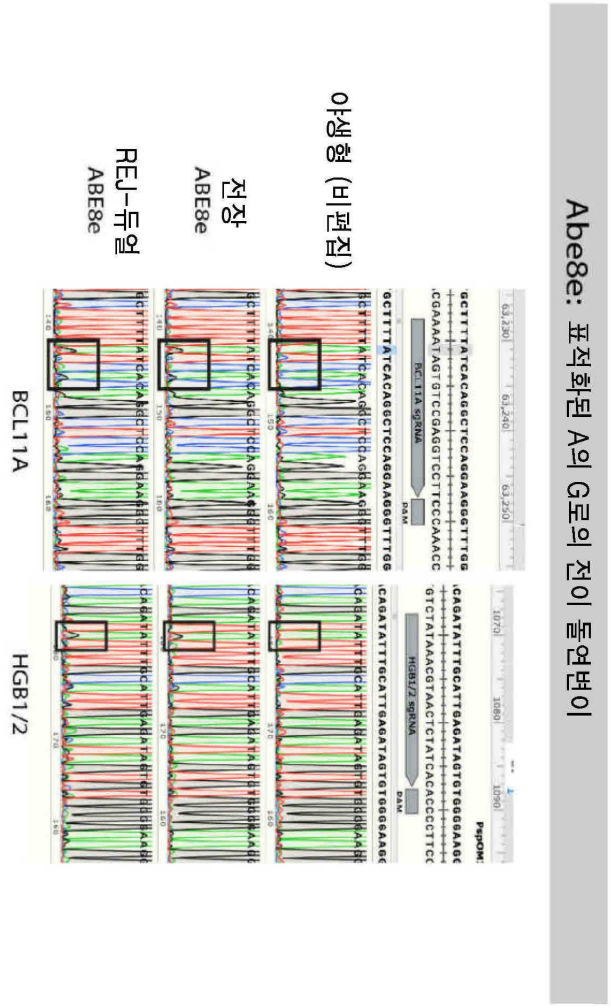
도면23a


Abe8e: Adenine Base Editor
kDa
250
150
100
75
37
Abe8e (anti-SpCas9, 7A9, N-말단):
전장
REJ-듀얼
DNA 불포함
전장
N-단편
GAPDH:
레인당 5 μg

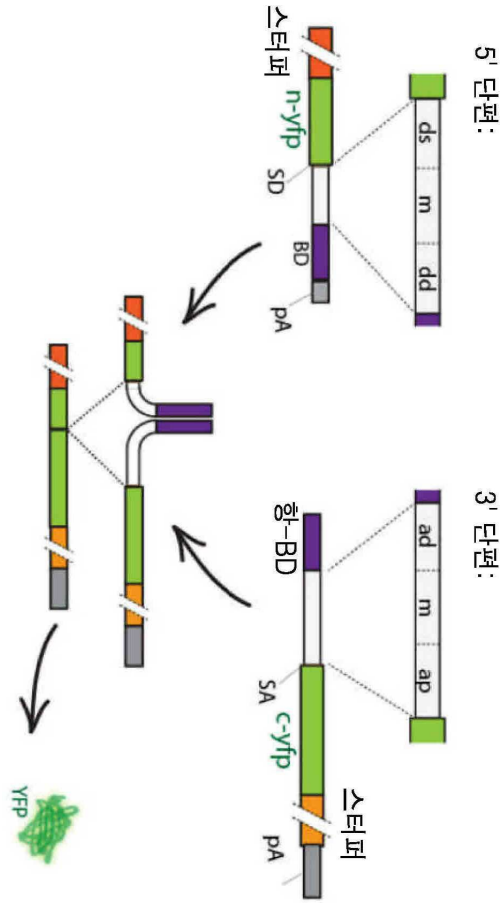
도면23b


전장의 분율
0
0.2
0.4
0.6
0.8
1
1.2
전장
REJ-듀얼

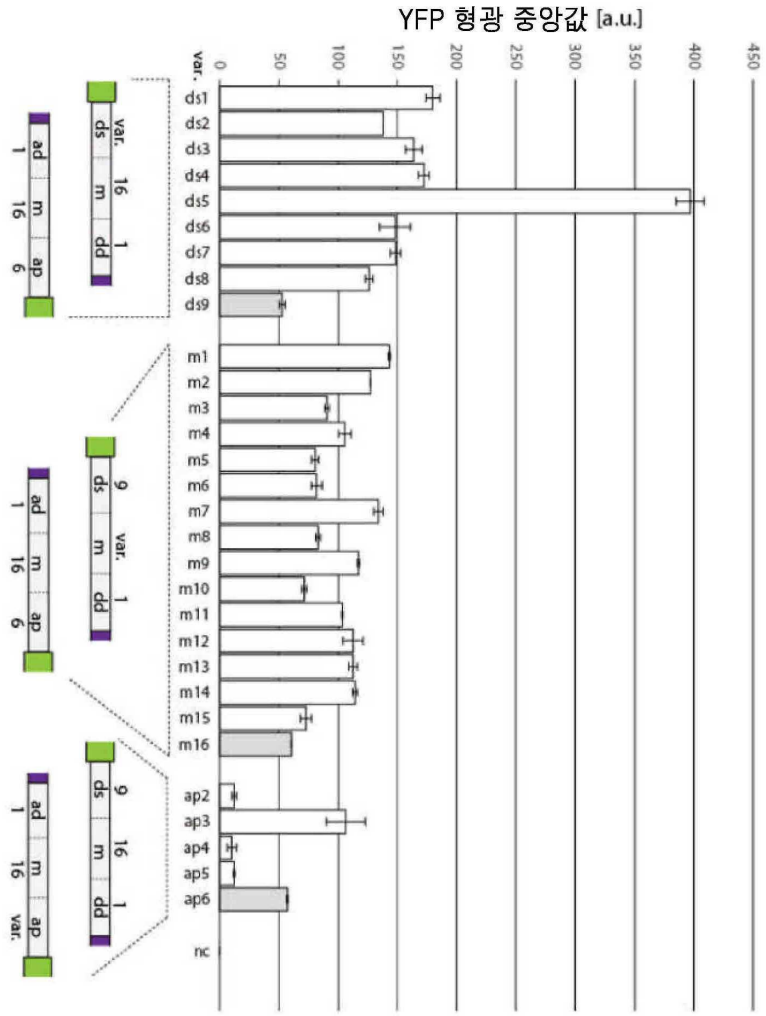
도면23c


Abe8e: 표적화된 A의 G로의 전이 돌연변이
야생형 (비편집)
전장
ABE8e
REJ-듀얼
ABE8e
BCL11A
HGB1/2

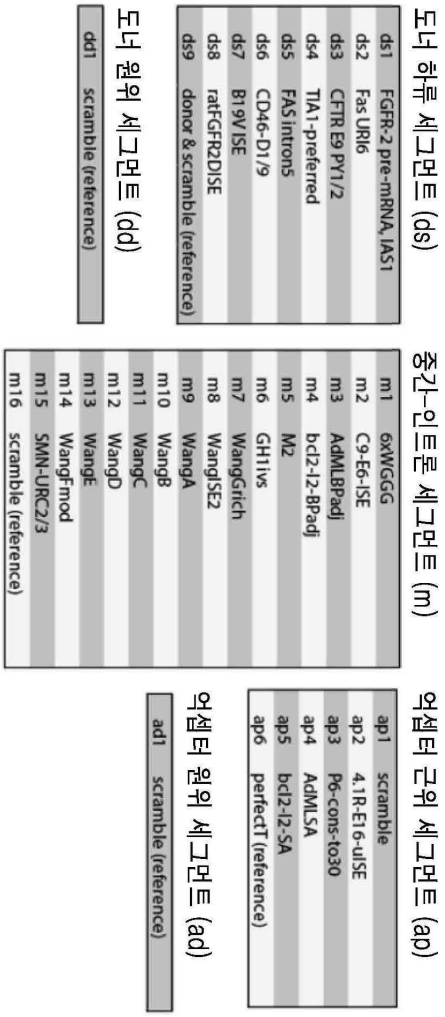
도면24a



도면24b


YFP 형광 중앙값 [a.u.]
var.
0
50
100
150
200
250
300
350
400
450
ds1
ds2
ds3
ds4
ds5
ds6
ds7
ds8
ds9
m1
m2
m3
m4
m5
m6
m7
m8
m9
m10
m11
m12
m13
m14
m15
m16
ap2
ap3
ap4
ap5
ap6
nc
ad
m
ap
ds
dd
var.
1
16
6
9

도면24c



서 열 목 록

SEQUENCE LISTING

<110> Salk Institute for Biological Studies

<120> COMPOSITIONS AND METHODS FOR HIGH-EFFICIENCY RECOMBINATION OF RNA MOLECULES

<130> 7158-102574-07

<150> 62/933,714

<151> 2019-11-11

<150> PCT/US2020/025430

<151> 2020-03-27

<160> 206

<170> PatentIn version 3.5

<210> 1

<211> 1491

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 1

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcc 540

accatggtga gcaagggcga ggagctgttc accggggtgg tgcccatcct ggtcgagctg 600

gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc 660

tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gccctggccc 720

accctcgtga ccaccttcgg ctacggcctg atgtgcttcg cccgctaccc cgaccacatg 780

aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc 840

ttcttcaagg acgacggcaa ctacaagacc cgcgccgagg tgaagttcga gggcgacacc 900

ctggtgaacc gcatcgagct gaagggcatc gacttcaagg aggacggcaa catcctgggg 960

cacaagctgg agtacaacta caacagccac aacgtctata tcatggccga caagcagaag 1020

aacggcatca aggtaagtat tagctctttc tttccatggg ttggcctcgc cgcgtgggct 1080

gagggaagga ctgtcctggg actggacagg cgggttatgg gacctgaaaa gcggccctga 1140

aaaagggccg cgatgaaaac gaagcgagct aaagcctcct ctctcttctt cagaactcct 1200

ctcttttctc tcctccagga gttcttcctc tctcccttct tctcaaatgc tttctccctc 1260

tctcctgcat ttgagctcct tctttcctct ctcgacaatc ccctttctc cctcttgatt 1320

gtcgactagc tcgcaatcat cgcggtatca aaaagcggtc aggcagctaa accaaaaggt 1380

ttagcaattg cctctgatga gtcgctgaaa tgcgacgaaa accgcttttt ggtaccaata 1440

aaatatcttt attttcatta catctgtgtg ttggtttttt gtgtgactag t 1491

<210> 2

<211> 1302

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 2

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcg 540

gaaaaccgcg ggataccgcg atgattgcga gctagtcgac aatcaagagg gagaaaaggg 600

gattgtcgag agaggaaaga aggagctcaa atgcaggaga gagggagaaa gcatttgaga 660

agaagggaga gaggaacaac tcgtggagga gagaaaagag acgagttgtg aagaagagag 720

aggaggcttt agctcgcttc gttttcatca ttattgcggc cctgaaaaag ggccgcttat 780

aacgttgctc gaattcgggt tatgggacca gtgaaggctg agggaaggac tgtcctggga 840

ctggacaggc gggttatggg acctgaaaat actaacaatc gatttttttt ccctttttt 900

ccaggtgaac ttcaagatcc gccacaacat cgaggacggc agcgtgcagc tcgccgacca 960

ctaccagcag aacaccccca tcggcgacgg ccccgtgctg ctgcccgaca accactacct 1020

gagctaccag tccgccctga gcaaagaccc caacgagaag cgcgatcaca tggtcctgct 1080

ggagttcgtg accgccgccg ggatcactct cggcatggac gagctgtaca aggacctttg 1140

agaattcctc acctgcgatc tcgatgcttt atttgtgaaa tttgtgatgc tattgcttta 1200

tttgtaacca ttataagctg caataaacaa gttaacaaca acaattgcat tcattttatg 1260

tttcaggttc agggggaggt gtgggaggtt ttttaaacta gt 1302

<210> 3

<211> 404

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 3

gtaagtatta gctctttctt tccatgggtt ggcctcgccg cgtgggctga gggaaggact 60

gtcctgggac tggacaggcg ggttatggga cctgaaaagc ggccctgaaa aagggccgcg 120

atgaaaacga agcgagctaa agcctcctct ctcttcttca gaactcctct cttttctctc 180

ctccaggagt tcttcctctc tcccttcttc tcaaatgctt tctccctctc tcctgcattt 240

gagctccttc tttcctctct cgacaatccc cttttctccc tcttgattgt cgactagctc 300

gcaatcatcg cggtatcaaa aagcggtcag gcagctaaac caaaaggttt agcaattgcc 360

tctgatgagt cgctgaaatg cgacgaaaac cgctttttgg tacc 404

<210> 4

<211> 382

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 4

acttctaggc gcgccgcgga aaaccgcggg ataccgcgat gattgcgagc tagtcgacaa 60

tcaagaggga gaaaagggga ttgtcgagag aggaaagaag gagctcaaat gcaggagaga 120

gggagaaagc atttgagaag aagggagaga ggaacaactc gtggaggaga gaaaagagac 180

gagttgtgaa gaagagagag gaggctttag ctcgcttcgt tttcatcatt attgcggccc 240

tgaaaaaggg ccgcttataa cgttgctcga attcgggtta tgggaccagt gaaggctgag 300

ggaaggactg tcctgggact ggacaggcgg gttatgggac ctgaaaatac taacaatcga 360

ttttttttcc ctttttttcc ag 382

<210> 5

<211> 489

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 5

atggtgagca agggcgagga gctgttcacc ggggtggtgc ccatcctggt cgagctggac 60

ggcgacgtaa acggccacaa gttcagcgtg tccggcgagg gcgagggcga tgccacctac 120

ggcaagctga ccctgaagtt catctgcacc accggcaagc tgcccgtgcc ctggcccacc 180

ctcgtgacca ccttcggcta cggcctgatg tgcttcgccc gctaccccga ccacatgaag 240

cagcacgact tcttcaagtc cgccatgccc gaaggctacg tccaggagcg caccatcttc 300

ttcaaggacg acggcaacta caagacccgc gccgaggtga agttcgaggg cgacaccctg 360

gtgaaccgca tcgagctgaa gggcatcgac ttcaaggagg acggcaacat cctggggcac 420

aagctggagt acaactacaa cagccacaac gtctatatca tggccgacaa gcagaagaac 480

ggcatcaag 489

<210> 6

<211> 237

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 6

gtgaacttca agatccgcca caacatcgag gacggcagcg tgcagctcgc cgaccactac 60

cagcagaaca cccccatcgg cgacggcccc gtgctgctgc ccgacaacca ctacctgagc 120

taccagtccg ccctgagcaa agaccccaac gagaagcgcg atcacatggt cctgctggag 180

ttcgtgaccg ccgccgggat cactctcggc atggacgagc tgtacaagga cctttga 237

<210> 7

<211> 382

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 7

acttctaggc gcgccgcgga aaaccgcggg ataccgcgat gattgcgagc tagggagaga 60

gaggggaaag aaaagagaaa gaggaggagg aaagagggga gagaggggag ggaaaggaga 120

gaagggagga agggaagaaa gaaagaagag gaaaagaggg gaggaggagg agaaaggaga 180

aaaaaagaag ggaagggaga aaggctttag ctcgcttcgt tttcatcatt attgcggccc 240

tgaaaaaggg ccgcttataa cgttgctcga attcgggtta tgggaccagt gaaggctgag 300

ggaaggactg tcctgggact ggacaggcgg gttatgggac ctgaaaatac taacaatcga 360

ttttttttcc cttttttcc ag 382

<210> 8

<211> 301

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 8

gtaagtgtcc cgcggaacat tattataacg ttgctcgaag atatcagatg gtgcgctcct 60

ggacgtagcc ttcgggcatg gcggacttga agaagtcgtg ctgcttcatg tggtcggggt 120

agcggctgaa gcactgcacg ccgtaggtca gggtggtcac gagggtgggc cagggcacgg 180

gcagcttgcc ggtggtgcag atgaacttca gggtcagctt gccgtaggtg gcatcgccct 240

cgccctcgcc ggacacgctg aacttgtggc cgtttacgtc gccgtccagc tcgactctag 300

a 301

<210> 9

<211> 326

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 9

gctagcgtcg agctggacgg cgacgtaaac ggccacaagt tcagcgtgtc cggcgagggc 60

gagggcgatg ccacctacgg caagctgacc ctgaagttca tctgcaccac cggcaagctg 120

cccgtgccct ggcccaccct cgtgaccacc ctgacctacg gcgtgcagtg cttcagccgc 180

taccccgacc acatgaagca gcacgacttc ttcaagtccg ccatgcccga aggctacgtc 240

caggagcgca ccatctccgc ggaacattat tataacgttg ctcgaatact aactggtacc 300

tcttcttttt tttttgatat ctgcag 326

<210> 10

<211> 278

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 10

gttgccttta cttctggcgc gccaaaaggc gtgccagaag taccgggcta ataatgtttc 60

gcggtcctct taaatctgcc taaatacgta taaatttgat cgccctgaaa aagggcgatc 120

aaagccctga aaaagggcat acgtagccct gaaaaagggc aggcagagcc ctgaaaaagg 180

gcaagaggac cgcggaacat tattagccgc caccatggac aggcgggtta tgggacctga 240

aaatactaac aatcgatttt ttttcccttt ttttccag 278

<210> 11

<211> 190

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 11

acttctaggc gcgccgcgga aaaccgcggg atatcattat tgcggccctg aaaaagggcc 60

gcttataacg ttgctcgaat tcgggttatg ggaccagtga aggctgaggg aaggactgtc 120

ctgggactgg acaggcgggt tatgggacct gaaaatacta acaatcgatt ttttttccct 180

ttttttccag 190

<210> 12

<211> 459

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 12

gtaagtatta gctctttctt tccatgggtt ggcctcgccg cgtgggctga gggaaggact 60

gtcctgggac tggacaggcg ggttatggga cctgaaaagc ggccctgaaa aagggccgcg 120

atgaaaacga agcgagctaa agcctcctct ctcttcttca gaactcctct cttttctctc 180

ctccaggagt tcttcctctc tcccttcttc tcaaatgctt tctccctctc tcctgcattt 240

gagctccttc tttcctctct cgacaatccc cttttctccc tcttgattgt cgactagctc 300

gcaatcatcg cggtatcaaa aagcggtcag gcagctaaac caaaaggttt agcaattgcc 360

tctgatgagt cgctgaaatg cgacgaaaac cgctttttgg taccaataaa atatctttat 420

tttcattaca tctgtgtgtt ggtttttgt gtgactagt 459

<210> 13

<211> 382

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 13

acttctaggc gcgccgcgga aaaccgcggg ataccgcgat gattgcgagc tagtcgacaa 60

tcaagaggga gaaaagggga ttgtcgagag aggaaagaag gagctcaaat gcaggagaga 120

gggagaaagc atttgagaag aagggagaga ggaagaactc ctggaggaga gaaaagagag 180

gagttctgaa gaagagagag gaggctttag ctcgcttcgt tttcatcatt attgcggccc 240

tgaaaaaggg ccgcttataa cgttgctcga attcgggtta tgggaccagt gaaggctgag 300

ggaaggactg tcctgggact ggacaggcgg gttatgggac ctgaaaatac taacaatcga 360

tttttttttcc ctttttttcc ag 382

<210> 14

<211> 372

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 14

gtaagtatta agcggccctg aaaaagggcc gcgatgaaaa cgaagcgagc taaagcctcc 60

tctctcttct tcagaactcc tctcttttct ctcctccagg agttcttcct ctctcccttc 120

ttctcaaatg ctttctccct ctctcctgca tttgagctcc ttctttcctc tctcgacaat 180

ccccttttct ccctcttgat tgtcgactag ctcgcaatca tcgcggtatc aaaaagcggt 240

caggcagcta aaccaaaagg tttagcaatt gcctctgatg agtcgctgaa atgcgacgaa 300

aaccgctttt tggtaccaat aaaatatctt tattttcatt acatctgtgt gttggttttt 360

tgtgtgacta gt 372

<210> 15

<211> 407

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 15

gtaagtatta gctctttctt tccatgggtt ggcctcgccg cgtgaagcgg ccctgaaaaa 60

gggccgcgat gaaaacgaag cgagctaaag cctcctctct cttcttcaga actcctctct 120

tttctctcct ccaggagttc ttcctctctc ccttcttctc aaatgctttc tccctctctc 180

ctgcatttga gctccttctt tcctctctcg acaatcccct tttctccctc ttgattgtcg 240

actagctcgc aatcatcgcg gtatcaaaaa gcggtcaggc agctaaacca aaaggtttag 300

caattgcctc tgatgagtcg ctgaaatgcg acgaaaaccg ctttttggta ccaataaaat 360

atctttattt tcattacatc tgtgtgttgg ttttttgtgt gactagt 407

<210> 16

<211> 378

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 16

gtaagtatta gctctttctt tccatgggtt ggcctcgccg cgtgggctga gggaaggact 60

gtcctgggac tggacaggcg ggttatggga cctgaaaagc ggccctgaaa aagggccgcg 120

atgaaaacga agcgagctaa agcctcctct ctcttcttca gaactcctct cttttctctc 180

ctccaggagt tcttcctctc tcccttcttc tcaaatgctt tctccctctc tcctgcattt 240

gagctccttc tttcctctct cgacaatccc cttttctccc tcttgattgt cgactagctc 300

gcaatcatcg cggtatcggt accaataaaa tatctttatt ttcattacat ctgtgtgttg 360

gtttttgtg tgactagt 378

<210> 17

<211> 309

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 17

acttctaggc gcgccgcgga aaaccgcggg ataccgcgat gattgcgagc tagtcgacaa 60

tcaagaggga gaaaagggga ttgtcgagag aggaaagaag gagctcaaat gcaggagaga 120

gggagaaagc atttgagaag aagggagaga ggaagaactc ctggaggaga gaaaagagag 180

gagttctgaa gaagagagag gaggctttag ctcgcttcgt tttcatcatt attgcggccc 240

tgaaaaaggg ccgcttataa cgttgctcga attctactaa caatcgattt tttttccctt 300

tttttccag 309

<210> 18

<211> 419

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 18

atatcctttt agggcagagt gaagagttag gaggaaggtg gttgggagag ggatttccag 60

gccttaggac atcatgacag atgaaaacga agcgagctaa agcctcctct ctcttcttca 120

gaactcctct cttttctctc ctccaggagt tcttcctctc tcccttcttc tcaaatgctt 180

tctccctctc tcctgcattt gagctccttc tttcctctct cgacaatccc cttttctccc 240

tcttgattgt cgactagctc gcaatcatcg cggtatcaaa aagcggtcag gcagctaaac 300

caaaaggttt agcaattgcc tctgatgagt cgctgaaatg cgacgaaaac cgctttttgg 360

taccaataaa atatctttat tttcattaca tctgtgtgtt ggtttttttgt gtgactagt 419

<210> 19

<211> 275

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 19

acttctaggc gcgccgcgga aaaccgcggg ataccgcgat gattgcgagc tagtcgacaa 60

tcaagaggga gaaaagggga ttgtcgagag aggaaagaag gagctcaaat gcaggagaga 120

gggagaaagc atttgagaag aagggagaga ggaacaactc gtggaggaga gaaaagagac 180

gagttgtgaa gaagagagag gaggctttag ctcgcttcgt tttcatcatt tccaggcctt 240

aggacatcat gacatttttc cttaactttg ctcac 275

<210> 20

<211> 3975

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 20

acttctaggc gcgccgccac catgggattc gtgcggcaga ttcagctgct gctgtggaag 60

aactggaccc tgcggaagcg gcagaaaatc agattcgtgg tggaactcgt gtggcccctg 120

agcctgtttc tggtgctgat ctggctgcgg aacgccaatc ctctgtacag ccaccacgag 180

tgtcacttcc ccaacaaggc catgccttct gccggaatgc tgccttggct gcagggcatc 240

ttctgcaacg tgaacaaccc ctgctttcaa agccccacac ctggcgaaag ccctggcatc 300

gtgtccaact acaacaacag catcctggcc agagtgtacc gggacttcca agagctgctg 360

atgaacgccc ctgagtctca gcacctgggc agaatctgga ccgagctgca catcctgagc 420

cagttcatgg acaccctgag aacacacccc gagagaatcg ccggcagggg catcagaatc 480

cgggacatcc tgaaggacga ggaaaccctg acactgttcc tcatcaagaa catcggcctg 540

agcgacagcg tggtgtacct gctgatcaac agccaagtgc ggcccgagca gtttgctcat 600

ggcgtgccag atctcgccct gaaggatatc gcctgttctg aggccctgct ggaacggttc 660

atcatcttca gccagcggag aggcgccaag accgtcagat atgccctgtg cagtctgagc 720

cagggaaccc tgcagtggat cgaggatacc ctgtacgcca acgtggactt cttcaagctg 780

ttccgggtgc tgcccacact gctggattct cggtcccaag gcatcaacct gagaagctgg 840

ggcggcatcc tgtccgacat gagcccaaga atccaagagt tcatccaccg gcctagcatg 900

caggacctgc tgtgggttac cagacctctg atgcagaacg gcggacccga gacattcacc 960

aagctgatgg gcattctgag cgatctgctg tgcggctacc ctgaaggcgg aggatctaga 1020

gtgctgagct tcaattggta cgaggacaac aactacaagg ccttcctggg catcgactcc 1080

accagaaagg accccatcta cagctacgac cggcggacaa ccagcttctg caatgccctg 1140

atccagagcc tggaaagcaa ccctctgacc aagatcgctt ggagggccgc caaacctctg 1200

ctgatgggaa agatcctgta cacccctgac agccctgccg ccagaagaat cctgaagaac 1260

gccaacagca ccttcgagga actggaacac gtgcgcaagc tggtcaaggc ctgggaagaa 1320

gtgggacctc agatctggta cttcttcgac aatagcaccc agatgaacat gatcagagac 1380

accctgggca accctaccgt gaaggacttc ctgaacagac agctgggcga agagggcatt 1440

accgccgagg ccatcctgaa ctttctgtac aagggcccca gagagtccca ggccgacgac 1500

atggccaact tcgattggcg ggacatcttc aacatcaccg acagaaccct gcggctggtc 1560

aaccagtacc tggaatgcct ggtgctggac aagttcgaga gctacaacga cgagacacag 1620

ctgacccaga gagccctgtc tctgctggaa gagaatatgt tctgggctgg cgtggtgttc 1680

cccgacatgt acccttggac aagcagcctg cctcctcacg tgaagtacaa gatccggatg 1740

gacatcgacg tggtcgaaaa gaccaacaag atcaaggacc ggtactggga cagcggccct 1800

agagctgatc ccgtggaaga ttttcgctac atctggggcg gattcgcata cctgcaggac 1860

atggtggaac agggaatcac acggtcccag gtgcaggctg aagctcctgt gggaatctac 1920

ctgcagcaga tgccttatcc ttgcttcgtg gacgacagct tcatgatcat cctgaatcgg 1980

tgcttcccca tcttcatggt gctggcctgg atctactccg tgtctatgac cgtgaagtcc 2040

atcgtgctgg aaaaagagct gcggctgaaa gagacactga agaaccaggg cgtgtccaat 2100

gccgtgatct ggtgcacctg gtttctggac agcttctcca ttatgagcat gagcatcttt 2160

ctgctgacga tcttcatcat gcacggccgg atcctgcact acagcgaccc ctttatcctc 2220

ttcctgttcc tgctggcctt ctccaccgct acaatcatgc tgtgttttct gctgtccacc 2280

ttcttctcca aagcctctct ggccgctgct tgtagcggcg tgatctactt caccctgtac 2340

ctgcctcaca tcctgtgctt cgcatggcag gacagaatga ccgccgagct gaagaaagct 2400

gtgtccctgc tgagccctgt ggcctttggc tttggcaccg agtacctcgt cagatttgag 2460

gaacaaggac tgggactgca gtggtccaac atcggcaata gccctacaga gggcgacgag 2520
ttcagcttcc tgctgtctat gcaaatgatg ctgctggacg ccgccgtgta tggactgctg 2580
gcttggtatc tggaccaggt gttccctgcc gattacggca ctcctctgcc ttggtatttc 2640
ctgctgcaag agagctactg gctcggcggc gagggatgta gcaccagaga agaaagagcc 2700
ctggaaaaga ccgagcctct gaccgaggaa acagaggacc ctgaacaccc agagggcatc 2760
cacgatagct ttttcgagag agaacacccc ggctgggtgc caggcgtgtg tgtgaagaat 2820
ctggtcaaga tcttcgagcc ctgcggcaga cctgccgtgg acagactgaa catcaccttc 2880

tacgagaacc agattaccgc ctttctgggc cacaacggcg ctggcaagac aaccacactg 2940
agcatcctca ccggcctgct gcctccaaca agcggcacag ttctcgttgg cggcagagac 3000
atcgagacaa gcctggatgc cgtcagacag tccctgggca tgtgccctca gcacaacatc 3060
ctgtttcacc acctgaccgt ggccgagcac atgctgtttt atgcccagct gaagggcaag 3120
agccaagaag aggctcagct ggaaatggaa gccatgctcg aggacaccgg cctgcaccac 3180
aagagaaatg aggaagccca ggatctgagc ggcggcatgc agagaaaact gagcgtggcc 3240
attgccttcg tgggcgacgc caaggttgtg atcctggatg agcctacaag cggcgtggac 3300

ccttacagca gaagatccat ctgggatctg ctgctgaagt acagaagcgg ccggaccatc 3360
atcatgagca cccaccacat ggacgaggcc gatctgctcg gagacagaat cgccatcatt 3420
gctcagggca gactgtactg cagcggcacc ccactgtttc tgaagaactg tttcggcacc 3480
ggactgtatc tgaccctcgt gcggaagatg aagaacatcc agtctcagcg gaagggcagc 3540
gagggcacct gtagctgttc tagcaagggc tttagcacca cctgtccagc tcacgtggac 3600
gatctgaccc ctgaacaggt gctggatggc gacgtgaacg agctgatgga cgtggtgctg 3660
caccatgtgc ctgaggccaa gctggtggaa tgcatcggcc aggtaagtat tagctctttc 3720

tttccatggg ttggcctcgc cgcgtgggct gagggaagga ctgtcctggg actggacagg 3780
cgggttatgg gacctgaagc gataaaaggc atgcacgttt gcggctacgt gcatgccaaa 3840
aggagtcggg cttgcctccg tgcccgactc caaaagacct gctcgaggag gtggacgagc 3900
aggtcaaaaa tccgggtacc aataaaatat ctttattttc attacatctg tgtgttggtt 3960
ttttgtgtga ctagt 3975

<210> 21

<211> 3611

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 21

aggatttttg acctgctcga ttgtccactg cgagcaggtc ttttggagtc gggcgaggcg 60

gaagcccgac tccttttggc atgcacgcta gccgcgtcgt gcatgccttt tatcgaattc 120

gggttatggg accagtgaag gctgagggaa ggactgtcct gggactggac aggcgggtta 180

tgggacctga aaatactaac aatcgatttt ttttcccttt ttttccagga actgatttttt 240

ctgctcccga acaagaactt caagcaccgg gcctacgcca gcctgttcag agagctggaa 300

gaaaccctgg ccgacctggg cctgtctagc tttggcatca gcgacacccc tctcgaagag 360

atcttcctga aagtgacaga ggacagcgat agcggccctc tgtttgctgg cggagcacag 420

caaaagcgcg agaacgtgaa ccctagacac ccctgtctgg gcccaagaga gaaagccgga 480

cagacccctc aggacagcaa tgtgtgctct cctggtgctc ctgccgctca tcctgaggga 540

caacctccac ctgaacctga gtgtcctgga cctcagctga acaccggaac acagctggtt 600

ctgcagcacg tgcaggctct gctcgtgaag agattccagc acaccatcag aagccacaag 660

gactttctgg cccagatcgt gctgcccgcc acctttgttt ttctggctct gatgctgagc 720

atcgtgatcc ctccattcgg cgagtacccc gctctgacac tgcacccttg gatctacggc 780

cagcagtaca ccttttttctc catggacgaa cccggcagcg agcagttcac agtgctggct 840

gatgtcctgc tgaacaagcc cggcttcggc aaccggtgtc tgaaagaagg atggctgcct 900

gagtaccctt gcggcaacag cacaccttgg aaaaccccta gcgtgtcccc taacatcacc 960

cagctgttcc aaaagcagaa atggacccaa gtgaacccct ctccatcctg ccggtgctcc 1020

acaagggaaa agctgaccat gctgcccgag tgtccagaag gcgctggcgg acttcctcca 1080

cctcagagaa cacagagatc caccgagatt ctccaggacc tgaccgaccg gaatatcagc 1140

gacttcctgg ttaagacata ccccgcactg atccggtcca gcctgaagtc caagttctgg 1200

gtcaacgaac agagatacgg cggcatcagc atcggcggaa aactgcctgt ggtgcctatc 1260

acaggcgagg cccttgtggg ctttctgtcc gatctgggga gaatcatgaa cgtgtccggc 1320

ggacctatca ccagggaagc cagcaaagag atccccgatt tcctgaagca cctggaaacc 1380

gaggacaata tcaaagtgtg gttcaacaac aaaggatggc acgccctcgt gtcttttctg 1440

aacgtggccc acaatgccat cctgcgggct agcctgccta aggacagaag ccctgaggaa 1500

tacggcatca ccgtgatctc ccagcctctg aatctgacca aagagcagct gagcgagatc 1560

accgtgctga ccacctctgt ggatgctgtg gtggccatct gcgtgatctt cagcatgagc 1620

ttcgtgcccg cctccttcgt gctgtacctg attcaagaga gagtgaacaa gagcaagcac 1680

ctccagttca tctccggggt gtccccaacc acctactggg tcaccaattt tctgtgggac 1740

atcatgaact acagcgtgtc agccggcctg gtcgtgggca tctttatcgg ctttcaaaag 1800
aaggcctaca cgagccccga gaacctgcct gctttggttg ctctgctgct cctgtatggc 1860
tgggccgtga ttcccatgat gtaccccgcc agctttctgt ttgacgtgcc cagcacagcc 1920
tacgtggccc tgtcttgcgc caatctgttc atcggcatca acagcagcgc catcacattc 1980
atcctggaac tgttcgagaa caacaggacc ctgctgcggt tcaacgccgt gctgcggaaa 2040
ctgctgatcg tgttccctca cttctgtctc ggccggggcc tgatcgacct ggctctgtct 2100
caagccgtga ccgatgtgta cgccagattt ggcgaggaac actccgccaa tccattccac 2160

tgggacctga tcggcaagaa cctgttcgcc atggtggtgg aaggcgtcgt gtacttcctg 2220
ctcactctgc tggtgcagag acacttttt ctgtcccaat ggatcgccga gcctaccaaa 2280
gaacccattg tggacgagga cgacgatgtg gccgaggaaa gacagagaat catcaccggc 2340
ggcaacaaga ccgatatcct gagactgcac gagctgacaa agatctaccc cggcacaagc 2400
tccccagccg tggataggct ttgtgtggga gttagacccg gcgagtgctt tggcctgctg 2460
ggagttaatg gcgccggaaa gaccaccacc ttcaagatgc tgaccggcga caccacagtg 2520
acaagcggag atgctacagt ggccggcaag agcatcctga ccaacatcag cgaagtgcat 2580

cagaacatgg gctactgccc tcagttcgac gccatcgacg aactgctgac aggccgcgaa 2640
cacctgtatc tgtatgccag actgagaggc gtgcccgctg aagagatcga gaaggtggcc 2700
aactggtcca tcaagtctct gggcctgaca gtgtacgccg actgtctggc cggaacatac 2760
agcggaggaa acaagcggaa gctgagcacc gccattgctc tgatcggatg cccacctctg 2820
gtcctgctgg atgaacccac caccggaatg gatccccagg ctagaagaat gctctggaac 2880
gtgatcgtgt ctatcatccg cgagggcaga gctgtggtgc tgacctctca ctccatggaa 2940
gagtgcgagg ctctgtgtac ccggctggcc attatggtca agggcgcctt cagatgcatg 3000

ggcaccattc agcatctgaa aagcaagttc ggcgacggct acatcgtgac aatgaagatc 3060
aagagcccca aggacgacct cctgcctgat ctgaaccccg tggaacagtt ttttcagggc 3120
aacttccccg gctccgtgca gcgggaaaga cactataaca tgctgcagtt tcaggtgtcc 3180
tcctccagcc tggctcggat ctttcaactg ctgctctctc acaaggacag cctgctgatt 3240
gaagagtaca gcgtgacaca gaccacactc gaccaggttt tcgtgaactt cgccaagcag 3300
cagaccgaga gccacgacct gcctctgcat cctcgggccg ctggtgcctc tagacaagct 3360
caggacggcg ctcgggctga ctacaaagac catgacggtg attataaaga tcatgacatc 3420

gactataagg atgacgatga caaatgaggt accaattcct cacctgcgat ctcgagcttt 3480

atttgtgaaa tttgtgatgc tattgcttta tttgtaacca ttataagctg caataaacaa 3540

gttaacaaca acaattgcat tcattttatg tttcaggttc agggggaggt gtgggaggtt 3600

ttttaaacta g 3611

<210> 22

<211> 3975

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 22

acttctaggc gcgccgccac catggcccca aagaagaagc ggaaggtcgg tatccacgga 60

gtcccagcag ccaagcggaa ctacatcctg ggcctggaca tcggcatcac cagcgtgggc 120

tacggcatca tcgactacga gacacgggac gtgatcgatg ccggcgtgcg gctgttcaaa 180

gaggccaacg tggaaaacaa cgagggcagg cggagcaaga gaggcgccag aaggctgaag 240

cggcggaggc ggcatagaat ccagagagtg aagaagctgc tgttcgacta caacctgctg 300

accgaccaca gcgagctgag cggcatcaac ccctacgagg ccagagtgaa gggcctgagc 360

cagaagctga gcgaggaaga gttctctgcc gccctgctgc acctggccaa gagaagaggc 420

gtgcacaacg tgaacgaggt ggaagaggac accggcaacg agctgtccac caaagagcag 480

atcagccgga acagcaaggc cctggaagag aaatacgtgg ccgaactgca gctggaacgg 540

ctgaagaaag acggcgaagt gcggggcagc atcaacagat tcaagaccag cgactacgtg 600

aaagaagcca aacagctgct gaaggtgcag aaggcctacc accagctgga ccagagcttc 660

atcgacacct acatcgacct gctggaaacc cggcggacct actatgaggg acctggcgag 720

ggcagcccct tcggctggaa ggacatcaaa gaatggtacg agatgctgat gggccactgc 780

acctacttcc ccgaggaact gcggagcgtg aagtacgcct acaacgccga cctgtacaac 840

gccctgaacg acctgaacaa tctcgtgatc accagggacg agaacgagaa gctggaatat 900

tacgagaagt tccagatcat cgagaacgtg ttcaagcaga agaagaagcc caccctgaag 960

cagatcgcca aagaaatcct cgtgaacgaa gaggatatta agggctacag agtgaccagc 1020

accggcaagc ccgagttcac caacctgaag gtgtaccacg acatcaagga cattaccgcc 1080

cggaaagaga ttattgagaa cgccgagctg ctggatcaga ttgccaagat cctgaccatc 1140

taccagagca gcgaggacat ccaggaagaa ctgaccaatc tgaactccga gctgacccag 1200

gaagagatcg agcagatctc taatctgaag ggctataccg gcacccacaa cctgagcctg 1260

aaggccatca acctgatcct ggacgagctg tggcacacca acgacaacca gatcgctatc 1320

ttcaaccggc tgaagctggt gcccaagaag gtggacctgt cccagcagaa agagatcccc 1380

accaccctgg tggacgactt catcctgagc cccgtcgtga agagaagctt catccagagc 1440
atcaaagtga tcaacgccat catcaagaag tacggcctgc ccaacgacat cattatcgag 1500
ctggcccgcg agaagaactc caaggacgcc cagaaaatga tcaacgagat gcagaagcgg 1560
aaccggcaga ccaacgagcg gatcgaggaa atcatccgga ccaccggcaa agagaacgcc 1620
aagtacctga tcgagaagat caagctgcac gacatgcagg aaggcaagtg cctgtacagc 1680
ctggaagcca tccctctgga agatctgctg aacaacccct tcaactatga ggtggaccac 1740
atcatcccca gaagcgtgtc cttcgacaac agcttcaaca acaaggtgct cgtgaagcag 1800

gaagaaaaca gcaagaaggg caaccggacc ccattccagt acctgagcag cagcgacagc 1860
aagatcagct acgaaacctt caagaagcac atcctgaatc tggccaaggg caagggcaga 1920
atcagcaaga ccaagaaaga gtatctgctg gaagaacggg acatcaacag gttctccgtg 1980
cagaaagact tcatcaaccg gaacctggtg gataccagat acgccaccag aggcctgatg 2040
aacctgctgc ggagctactt cagagtgaac aacctggacg tgaaagtgaa gtccatcaat 2100
ggcggcttca ccagctttct gcggcggaag tggaagttta agaaagagcg gaacaagggg 2160
tacaagcacc acgccgagga cgccctgatc attgccaacg ccgatttcat cttcaaagag 2220

tggaagaaac tggacaaggc caaaaaagtg atggaaaacc agatgttcga ggaaaagcag 2280
gccgagagca tgcccgagat cgaaaccgag caggagtaca aagagatctt catcaccccc 2340
caccagatca agcacattaa ggacttcaag gactacaagt acagccaccg ggtggacaag 2400
aagcctaata gagagctgat taacgacacc ctgtactcca cccggaagga cgacaagggc 2460
aacaccctga tcgtgaacaa tctgaacggc ctgtacgaca aggacaatga caagctgaaa 2520
aagctgatca acaagagccc cgaaaagctg ctgatgtacc accacgaccc ccagacctac 2580
cagaaactga agctgattat ggaacagtac ggcgacgaga agaatccccc gtacaagtac 2640

tacgaggaaa ccgggaacta cctgaccaag tactccaaaa aggacaacgg ccccgtgatc 2700
aagaagatta agtattacgg caacaaactg aacgcccatc tggacatcac cgacgactac 2760
cccaacagca gaaacaaggt cgtgaagctg tccctgaagc cctacagatt cgacgtgtac 2820
ctggacaatg gcgtgtacaa gttcgtgacc gtgaagaatc tggatgtgat caaaaaagaa 2880
aactactacg aagtgaatag caagtgctat gaggaagcta agaagctgaa gaagatcagc 2940
aaccaggccg agtttatcgc ctccttctac aacaacgatc tgatcaagat caacggcgag 3000
ctgtatagag tgatcggcgt gaacaacgac ctgctgaacc ggatcgaagt gaacatgatc 3060

gacatcacct accgcgagta cctggaaaac atgaacgaca agaggccccc caggatcatt 3120

aagacaatcg ccggaagcgg agctactaac ttcagcctgc tgaagcaggc tggagacgtg 3180

gaggagaacc ctggacctag gcgcgccgcc accatggtga gcaagggcga ggagctgttc 3240

accggggtgg tgcccatcct ggtcgagctg gacggcgacg taaacggcca caagttcagc 3300

gtgtccggcg agggcgaggg cgatgccacc tacggcaagc tgaccctgaa gttcatctgc 3360

accaccggca agctgcccgt gccctggccc accctcgtga ccaccttcgg ctacggcctg 3420

atgtgcttcg cccgctaccc cgaccacatg aagcagcacg acttcttcaa gtccgccatg 3480

cccgaaggct acgtccagga gcgcaccatc ttcttcaagg acgacggcaa ctacaagacc 3540

cgcgccgagg tgaagttcga gggcgacacc ctggtgaacc gcatcgagct gaagggcatc 3600

gacttcaagg aggacggcaa catcctgggg cacaagctgg agtacaacta caacagccac 3660

aacgtctata tcatggccga caagcagaag aacggcatca aggtaagtat tagctctttc 3720

tttccatggg ttggcctcgc cgcgtgggct gagggaagga ctgtcctggg actggacagg 3780

cgggttatgg gacctgaagc gataaaaggc atgcacgttt gcggctacgt gcatgccaaa 3840

aggagtcggg cttgcctccg tgcccgactc caaaagacct gctcgaggag gtggacgagc 3900

aggtcaaaaa tccgggtacc aataaaatat ctttattttc attacatctg tgtgttggtt 3960

ttttgtgtga ctagt 3975

<210> 23

<211> 3912

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 23

aggatttttg acctgctcga ttgtccactg cgagcaggtc ttttggagtc gggcgaggcg 60

gaagcccgac tccttttggc atgcacgcta gccgcgtcgt gcatgccttt tatcttcggg 120

ttatgggacc agtgaaggct gagggaagga ctgtcctggg actggacagg cgggttatgg 180

gacctgaaaa tactaacaat cgatttttttt tccctttttt tccaggtgaa cttcaagatc 240

cgccacaaca tcgaggacgg cagcgtgcag ctcgccgacc actaccagca gaacaccccc 300

atcggcgacg gccccgtgct gctgcccgac aaccactacc tgagctacca gtccgccctg 360

agcaaagacc ccaacgagaa gcgcgatcac atggtcctgc tggagttcgt gaccgccgcc 420

gggatcactc tcggcatgga cgagctgtac aaggaccttg gaagcggagc tactaacttc 480

agcctgctga agcaggctgg agacgtggag gagaaccctg gacctatcac aaagaagcac 540

acagcccact tctccaagaa gggcgaagag gaaaacctgg aaggcctggg caatcagacc 600

aagcagatcg tcgagaagta cgcctgcacc accagaatca gccccaacac aagccagcag 660

aacttcgtga cccagcggag caaaagagcc ctgaagcagt ttcggctgcc cctggaagaa 720
accgagctgg aaaagcggat catcgtggac gacaccagca cacagtggtc caagaacatg 780
aagcacttga cccctagcac actgacccag atcgactaca acgagaaaga gaagggcgct 840
atcacacaga gcccactgag cgactgtctg accagaagcc acagcatccc tcaggccaac 900
agatcccctc tgccaatcgc caaagtgtct agcttcccca gcatcagacc catctacctg 960
accagagtgc tgttccagga caacagcagc catctgccag ccgccagcta ccggaagaaa 1020
gatagcggcg tgcaagagtc cagccacttt ctgcaaggcg ctaagaagaa caatctgagc 1080

ctggctattc tgaccctgga aatgaccggc gatcagagag aagtcggctc tctgggcacc 1140
agcgccacaa atagcgtgac ctacaaaaag gtggaaaaca ccgtgctgcc taagcctgac 1200
ctgccaaaga caagcggcaa ggtggaactg ctgccaaagg tgcacatcta ccagaaggac 1260
ctgtttccta ccgagacaag caacggctct cccggccatc tggatctggt ggaaggatct 1320
ctgctgcagg gaaccgaggg cgccatcaag tggaacgagg ccaatagacc tggcaaggtg 1380
cccttcctga gagtggccac agagtctagc gccaagacac cctccaaact gctggatccc 1440
ctggcctggg ataaccacta cggcactcag atccccaaag aggaatggaa gtcccaagag 1500

aagtcccctg aaaagaccgc cttcaagaag aaggacacca ttctgtccct gaatgcctgc 1560
gagagcaacc acgccattgc cgccatcaat gagggccaga acaagcccga gatcgaagtg 1620
acctgggcca agcagggaag aaccgagaga ctgtgctccc agaatcctcc tgtgctgaag 1680
cggcaccaga gagaaatcac ccggaccaca ctgcagagcg accaagaaga gatcgattac 1740
gacgatacca tcagcgtcga gatgaagaaa gaagatttcg acatctacga cgaggacgag 1800
aatcagagcc ctcggagctt ccagaagaaa accaggcact actttattgc cgccgtcgag 1860
cggctgtggg actacggaat gtctagctct cctcacgtgc tgcggaatag agcccagtct 1920

ggtagcgtgc cccagttcaa aaaggtcgtg ttccaagagt tcaccgacgg cagcttcacc 1980
cagccactgt atagaggcga gctgaacgag catctgggcc tgctgggccc ttatatcaga 2040
gccgaagtgg aagataacat catggtcacc ttccggaatc aggcctctcg gccctacagc 2100
ttctacagct ccctgatctc ctacgaagag gaccagagac agggcgcaga gccccggaag 2160
aatttcgtga agcccaacga gactaagacc tacttttgga aggtgcagca ccatatggcc 2220
cctacaaagg acgagttcga ctgcaaagcc tgggcctact tctccgatgt ggacctcgag 2280
aaggatgtgc acagcggact catcggccca ctgcttgtgt gccacaccaa cacactgaac 2340

cccgctcacg gcagacaagt gacagtgcaa gaattcgccc tgtttttcac catcttcgac 2400

gaaacgaagt cctggtactt caccgaaaac atggaaagaa actgcagggc cccttgcaac 2460

attcagatgg aagatcccac cttcaaagag aactaccggt tccacgccat caacggctac 2520

atcatggaca cactgcccgg cctggttatg gctcaggatc agagaatccg gtggtatctg 2580

ctgtccatgg gctccaacga gaatatccac tccatccact tctccggcca cgtgttcacc 2640

gtgcggaaaa aagaagagta caaaatggcc ctgtacaatc tgtaccctgg ggtgttcgaa 2700

accgttgaga tgctgcctag caaggccgga atttggagag tggaatgtct gattggagag 2760

cacctccacg ccgggatgag caccctgttt ctggtgtact ccaacaagtg tcagacccct 2820

ctcggcatgg cctctggcca cattagagac ttccagatca ccgccagcgg acagtatgga 2880

cagtgggccc ctaaactggc cagactgcac tactccggca gcatcaatgc ctggtccacc 2940

aaagagcctt tcagctggat caaagtggac ctgctggctc ccatgatcat ccacggaatc 3000

aagacccagg gcgccagaca aaagttcagc agcctgtaca tcagccagtt catcatcatg 3060

tacagcctgg acggaaagaa gtggcagacc taccggggca atagcaccgg cacactgatg 3120

gtgttcttcg gcaacgtgga ctccagcggc attaagcaca acatcttcaa ccctccaatc 3180

attgcccgat acatccggct gcaccccaca cactacagca tcaggtctac cctgagaatg 3240

gaactgatgg gctgcgacct gaacagctgc agcatgcccc tcggaatgga aagcaaggcc 3300

atcagcgacg cccagatcac agcctctagc tacttcacca acatgttcgc cacttggagc 3360

ccctctaagg cccggcttca tctgcaaggc agaagcaacg cttggaggcc ccaagtgaac 3420

aaccccaaag aatggctgca ggtcgacttt cagaaaacca tgaaagtgac aggcgtgacc 3480

acacagggcg tcaagtccct gctgacctct atgtacgtga aagagtttct gatcagctcc 3540

agccaggacg gccaccagtg gaccctgttc ttccaaaacg gcaaagtgaa agtgttccag 3600

ggaaatcagg acagcttcac acccgtggtc aactccctgg atcctccact gctgacaaga 3660

tacctgcgga ttcaccctca gtcttgggtg caccagattg ccctgcggat ggaagtgctg 3720

ggctgtgaag ctcaggacct ctactgaggt accaattcct cacctgcgat ctcgatgctt 3780

tatttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca 3840

agttaacaac aacaattgca ttcattttat gtttcaggtt cagggggagg tgtgggaggt 3900

tttttaaact ag 3912

<210> 24

<211> 3828

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 24

acttctaggc gcgccgccac catgtaccca tacgatgttc cagattacgc ttatccttat 60

gacgtgcctg actacgccta tccctacgac gtccccgact atgcagtgta caagaaaacc 120

ctgttcgtgg aattcaccga ccacctgttc aatatcgcca agcctcggcc tccttggatg 180

ggactgctgg gacctacaat tcaggccgag gtgtacgaca ccgtggtcat caccctgaag 240

aacatggcca gccatcctgt gtctctgcac gccgtgggag tgtcttactg gaaggcttct 300

gagggcgccg agtacgacga tcagacaagc cagagagaga aagaggacga caaggttttc 360

cctggcggca gccacaccta tgtctggcaa gtcctgaaag aaaacggccc tatggcctcc 420

gatcctctgt gcctgacata cagctacctg agccacgtgg acctggtcaa ggacctgaat 480

tctggcctga tcggagccct gctcgtgtgt agagaaggca gcctggccaa agagaaaacc 540

cagacactgc acaagttcat cctgctgttc gccgtgttcg acgagggcaa gagctggcac 600

agcgagacaa agaacagcct gatgcaggac agggatgccg cctctgctcg ggcttggcct 660

aagatgcaca ccgtgaacgg ctacgtgaac agaagcctgc ctggactgat cggctgccac 720

agaaagtccg tgtactggca cgtgatcggc atgggcacaa cacctgaggt gcacagcatc 780

tttctggaag gacacacctt cctcgtgcgg aaccatagac aggccagcct ggaaatcagc 840

cctatcacct tcctgaccgc tcagaccctg ctgatggatc tgggccagtt tctgctgttc 900

tgccacatca gctcccacca gcacgatggc atggaagcct acgtgaaggt ggacagctgc 960

cccgaagaac cccagctgcg gatgaagaac aacgaggaag ccgaggacta cgacgacgac 1020

ctgaccgact ctgagatgga cgtcgtcaga ttcgacgacg ataacagccc cagcttcatc 1080

caaatcagaa gcgtggccaa gaagcacccc aagacctggg tgcactatat cgccgccgag 1140

gaagaggact gggattacgc tcctctggtg ctggcccctg acgacagaag ctacaagagc 1200

cagtacctga acaacggccc tcagcggatc ggccggaagt ataagaaagt gcggttcatg 1260

gcctacaccg acgagacatt caagaccaga gaggccatcc agcacgagag cggaattctg 1320

ggccctctgc tgtatggcga agtgggcgat acactgctga tcatcttcaa gaaccaggcc 1380

agcagaccct acaacatcta ccctcacggc atcaccgatg tgcggccccc gtattctaga 1440

aggctgccca agggcgtgaa gcacctgaag gacttcccta tcctgcctgg cgagatcttc 1500

aagtacaagt ggaccgtgac cgtggaagat ggccccacca agagcgaccc tagatgtctg 1560

acacggtact acagcagctt cgtgaacatg gaacgcgacc tggccagcgg cctgattgga 1620

cctctgctga tctgctacaa agaaagcgtg gaccagcggg gcaaccagat catgagcgac 1680

aagcggaacg tgatcctgtt tagcgtgttc gatgagaacc ggtcctggta tctgaccgag 1740

aacatccagc ggtttctgcc caatcctgct ggcgtgcagc tggaagatcc tgagttccag 1800

gcctccaaca tcatgcactc catcaatggc tatgtgttcg acagcctgca gctgagcgtg 1860

tgcctgcacg aagtggccta ctggtacatc ctgagcattg gcgcccagac cgacttcctg 1920

tccgtgttct tttccggcta caccttcaag cacaagatgg tgtacgagga taccctgaca 1980

ctgttcccat tctccggcga gacagtgttc atgagcatgg aaaaccccgg cctgtggatc 2040

ctgggctgtc acaacagcga cttccggaac agaggcatga cagccctgct gaaggtgtcc 2100

agctgcgaca agaacaccgg cgactactac gaggacagct atgaggacat cagcgcctac 2160

ctgctgagca agaacaatgc catcgagccc agaagcttca gccagaatag cagacacccc 2220

tccaccagac agaagcagtt caacgccaca acaatccccg agaacgacat cgagaaaacc 2280

gatccttggt ttgcccaccg gacccctatg cctaagatcc agaacgtgtc ctccagcgat 2340

ctgctgatgc tcctgagaca gagccctaca cctcacggac tgagcctgtc cgatctgcaa 2400

gaggccaaat acgaaacctt cagcgacgac ccttctcctg gcgccatcga cagcaacaat 2460

agcctgagcg agatgaccca cttcagacca cagctgcacc acagcggcga catggtgttt 2520

acacctgaga gcggcctcca gctgagactg aatgagaagc tgggaaccac cgccgccacc 2580

gagctgaaga aactggactt caaggtgtcc tctaccagca acaacctgat cagcacaatc 2640

ccctccgaca acctggctgc cggcaccgac aacacatctt ctctgggccc acctagcatg 2700

cccgtgcact acgatagcca gctggatacc acactgttcg gcaagaagtc tagccctctg 2760

acagagtctg gcggccctct gtctctgagc gaggaaaaca acgacagcaa gctgctggaa 2820

tccggcctga tgaacagcca agagtcctcc tggggcaaga atgtgtccag caccgagtcc 2880

ggcagactgt tcaagggaaa gagagcccac ggacctgctc tgctgaccaa ggataacgcc 2940

ctgttcaaag tgtccatcag cctgctcaag accaacaaga cctccaacaa ctccgccacc 3000

aacagaaaga cccacatcga cggccctagc ctgctgatcg agaatagccc tagcgtctgg 3060

cagaatatcc tggaaagcga caccgagttc aagaaagtga cccctctgat ccacgaccgg 3120

atgctcatgg acaagaacgc caccgctctg cggctgaacc acatgagcaa caagacaacc 3180

agcagcaaga atatggaaat ggtgcagcag aagaaagagg gccccattcc tccagacgct 3240

cagaaccccg atatgagctt cttcaagatg ctctttctgc ccgagagcgc ccggtggatc 3300

cagagaacac acggcaagaa ctccctgaac tccggccagg gaccttctcc aaagcagctg 3360

gtttccctgg gacctgagaa gtccgtggaa ggccagaact tcctgagcga aaagaacaaa 3420

gtggtcgtcg gcaagggcga gttcaccaag gatgtgggcc tgaaagagat ggtctttccc 3480

agcagccgga acctgttcct gaccaacctg gacaacctgc acgagaacaa cacccacaat 3540

caagagaaga agatccaaga ggtaagtatt agctctttct ttccatgggt tggcctcgcc 3600

gcgtgggctg agggaaggac tgtcctggga ctggacaggc gggttatggg acctgaagcg 3660

ataaaaggca tgcacgtttg cggctacgtg catgccaaaa ggagtcgggc ttgcctccgt 3720

gcccgactcc aaaagacctg ctcgaggagg tggacgagca ggtcaaaaat ccgggtacca 3780

ataaaatatc tttatttca ttacatctgt gtgttggttt tttgtgtg 3828

<210> 25

<211> 3802

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 25

aggatttttg acctgctcga ttgtccactg cgagcaggtc ttttggagtc gggcgaggcg 60

gaagcccgac tccttttggc atgcacgcta gccgcgtcgt gcatgccttt tatcttcggg 120

ttatgggacc agtgaaggct gagggaagga ctgtcctggg actggacagg cgggttatgg 180

gacctgaaaa tactaacaat cgatttttt tccctttttt tccaggaaat cgaaaagaaa 240

gagacactca tccaagagaa cgtggtgctg cctcagatcc acacagtgac cggcaccaag 300

aactttatga agaatctgtt cctgctgagt acccggcaga atgtggaagg cagctacgac 360

ggcgcttatg cccctgtgct gcaagacttc agatccctga acgactccac caatcggaca 420

aagaagcaca cagcccactt ctccaagaag ggcgaagagg aaaacctgga aggcctgggc 480

aatcagacca agcagatcgt cgagaagtac gcctgcacca ccagaatcag ccccaacaca 540

agccagcaga acttcgtgac ccagcggagc aaaagagccc tgaagcagtt tcggctgccc 600

ctggaagaaa ccgagctgga aaagcggatc atcgtggacg acaccagcac acagtggtcc 660

aagaacatga agcacttgac ccctagcaca ctgacccaga tcgactacaa cgagaaagag 720

aagggcgcta tcacacagag cccactgagc gactgtctga ccagaagcca cagcatccct 780

caggccaaca gatcccctct gccaatcgcc aaagtgtcta gcttccccag catcagaccc 840

atctacctga ccagagtgct gttccaggac aacagcagcc atctgccagc cgccagctac 900

cggaagaaag atagcggcgt gcaagagtcc agccactttc tgcaaggcgc taagaagaac 960

aatctgagcc tggctattct gaccctggaa atgaccggcg atcagagaga agtcggctct 1020

ctgggcacca gcgccacaaa tagcgtgacc tacaaaaagg tggaaaacac cgtgctgcct 1080

aagcctgacc tgccaaagac aagcggcaag gtggaactgc tgccaaaggt gcacatctac 1140

cagaaggacc tgtttcctac cgagacaagc aacggctctc ccggccatct ggatctggtg 1200

gaaggatctc tgctgcaggg aaccgagggc gccatcaagt ggaacgaggc caatagacct 1260

ggcaaggtgc ccttcctgag agtggccaca gagtctagcg ccaagacacc ctccaaactg 1320

ctggatcccc tggcctggga taaccactac ggcactcaga tccccaaaga ggaatggaag 1380

tcccaagaga agtcccctga aaagaccgcc ttcaagaaga aggacaccat tctgtccctg 1440

aatgcctgcg agagcaacca cgccattgcc gccatcaatg agggccagaa caagcccgag 1500

atcgaagtga cctgggccaa gcagggaaga accgagagac tgtgctccca gaatcctcct 1560

gtgctgaagc ggcaccagag agaaatcacc cggaccacac tgcagagcga ccaagaagag 1620

atcgattacg acgataccat cagcgtcgag atgaagaaag aagatttcga catctacgac 1680

gaggacgaga atcagagccc tcggagcttc cagaagaaaa ccaggcacta ctttattgcc 1740

gccgtcgagc ggctgtggga ctacggaatg tctagctctc ctcacgtgct gcggaataga 1800

gcccagtctg gtagcgtgcc ccagttcaaa aaggtcgtgt tccaagagtt caccgacggc 1860

agcttcaccc agccactgta tagaggcgag ctgaacgagc atctgggcct gctgggccct 1920

tatatcagag ccgaagtgga agataacatc atggtcacct tccggaatca ggcctctcgg 1980

ccctacagct tctacagctc cctgatctcc tacgaagagg accagagaca gggcgcagag 2040

ccccggaaga atttcgtgaa gcccaacgag actaagacct acttttggaa ggtgcagcac 2100

catatggccc ctacaaagga cgagttcgac tgcaaagcct gggcctactt ctccgatgtg 2160

gacctcgaga aggatgtgca cagcggactc atcggcccac tgcttgtgtg ccacaccaac 2220

acactgaacc ccgctcacgg cagacaagtg acagtgcaag aattcgccct gtttttcacc 2280

atcttcgacg aaacgaagtc ctggtacttc accgaaaaca tggaaagaaa ctgcagggcc 2340

ccttgcaaca ttcagatgga agatcccacc ttcaaagaga actaccggtt ccacgccatc 2400

aacggctaca tcatggacac actgcccggc ctggttatgg ctcaggatca gagaatccgg 2460

tggtatctgc tgtccatggg ctccaacgag aatatccact ccatccactt ctccggccac 2520

gtgttcaccg tgcggaaaaa agaagagtac aaaatggccc tgtacaatct gtaccctggg 2580

gtgttcgaaa ccgttgagat gctgcctagc aaggccggaa tttggagagt ggaatgtctg 2640

attggagagc acctccacgc cgggatgagc accctgtttc tggtgtactc caacaagtgt 2700

cagacccctc tcggcatggc ctctggccac attagagact tccagatcac cgccagcgga 2760

cagtatggac agtgggcccc taaactggcc agactgcact actccggcag catcaatgcc 2820

tggtccacca aagagccttt cagctggatc aaagtggacc tgctggctcc catgatcatc 2880

cacggaatca agacccaggg cgccagacaa aagttcagca gcctgtacat cagccagttc 2940

atcatcatgt acagcctgga cggaaagaag tggcagacct accggggcaa tagcaccggc 3000

acactgatgg tgttcttcgg caacgtggac tccagcggca ttaagcacaa catcttcaac 3060

cctccaatca ttgcccgata catccggctg caccccacac actacagcat caggtctacc 3120

ctgagaatgg aactgatggg ctgcgacctg aacagctgca gcatgcccct cggaatggaa 3180

agcaaggcca tcagcgacgc ccagatcaca gcctctagct acttcaccaa catgttcgcc 3240

acttggagcc cctctaaggc ccggcttcat ctgcaaggca gaagcaacgc ttggaggccc 3300

caagtgaaca accccaaaga atggctgcag gtcgactttc agaaaaccat gaaagtgaca 3360

ggcgtgacca cacagggcgt caagtccctg ctgacctcta tgtacgtgaa agagtttctg 3420

atcagctcca gccaggacgg ccaccagtgg accctgttct tccaaaacgg caaagtgaaa 3480

gtgttccagg gaaatcagga cagcttcaca cccgtggtca actccctgga tcctccactg 3540

ctgacaagat acctgcggat tcaccctcag tcttgggtgc accagattgc cctgcggatg 3600

gaagtgctgg gctgtgaagc tcaggacctc tactgaggta ccaattcctc acctgcgatc 3660

tcgatgcttt atttgtgaaa tttgtgatgc tattgcttta tttgtaacca ttataagctg 3720

caataaacaa gttaacaaca acaattgcat tcattttatg tttcaggttc agggggaggt 3780

gtgggaggtt ttttaaacta gt 3802

<210> 26

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 26

tgggggggagg 10

<210

> 27

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 27

gtagtgaggg 10

<210> 28

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 28

gttggtggtt 10

<210> 29

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400

> 29

agttgtggtt 10

<210> 30

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 30

gtattgggtc 10

<210> 31

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 31

agtgtgaggg 10

<210> 32

<211> 10

<212

> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 32

gggtaatggg 10

<210> 33

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 33

tcattggggt 10

<210> 34

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 34

ggtgggggtc 10

<210> 35

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 35

ggttttgttg 10

<210> 36

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 36

tatactcccg 10

<210> 37

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 37

gtattcgatc 10

<210> 38

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 38

gtagttccct 10

<210> 39

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 39

gttaatagta 10

<210> 40

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 40

tgctggttag 10

<210> 41

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 41

ataggtaacg 10

<210> 42

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 42

tctgaattgc 10

<210> 43

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 43

tctgggtttg 10

<210> 44

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 44

cattctcttt 10

<210> 45

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 45

gtattggtgt 10

<210> 46

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 46

tttagatttg 10

<210> 47

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 47

ataagtactg 10

<210> 48

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 48

tagtctatta 10

<210> 49

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 49

aggtattgca 10

<210> 50

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 50

gtagattacg 10

<210> 51

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 51

gggcgggtgc 10

<210> 52

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 52

cgtttacaat 10

<210> 53

<211> 11

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 53

gtacagggat g 11

<210> 54

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 54

aatcagggga 10

<210> 55

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 55

ggaggttttg 10

<210> 56

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 56

gtattccctg 10

<210> 57

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 57

tggtaagatc 10

<210> 58

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 58

gtagttaagt 10

<210> 59

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 59

gttggtttgg 10

<210> 60

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 60

gtatttactt 10

<210> 61

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 61

gtaacggggt 10

<210> 62

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 62

tttttttctg 10

<210> 63

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 63

ggggaaggga 10

<210> 64

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 64

ttaccccggt 10

<210> 65

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 65

gtattctatg 10

<210> 66

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 66

aggtattgtg 10

<210> 67

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 67

tttgggggggg 10

<210> 68

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 68

gttgttagcg 10

<210> 69

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 69

ggtagttggg 10

<210> 70

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 70

ctaagtactg 10

<210> 71

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 71

aaccatcttc 10

<210> 72

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 72

gtacctgggt 10

<210> 73

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 73

gtatctcatt 10

<210> 74

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 74

aaataaaatt 10

<210> 75

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 75

ggtgggttat 10

<210> 76

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 76

taagggaggg 10

<210> 77

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 77

tatgggaggg 10

<210> 78

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 78

gatgggaggg 10

<210> 79

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 79

tggggggggt 10

<210> 80

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 80

ggggaagggg 10

<210> 81

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 81

tggtaagagg 10

<210> 82

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 82

gggttagggt 10

<210> 83

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 83

gtatcggggg 10

<210> 84

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 84

ggttttgctg 10

<210> 85

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 85

tgggggtgga 10

<210> 86

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 86

actttagag 10

<210> 87

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 87

gtaacgggtt 10

<210> 88

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 88

gtttggggga 10

<210> 89

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 89

attttagag 10

<210> 90

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 90

ttaaagtagg 10

<210> 91

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 91

gtattaatat 10

<210> 92

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 92

ggtttgggtg 10

<210> 93

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 93

tatgggaaag 10

<210> 94

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 94

ggttgggagg 10

<210> 95

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 95

gtatttagtg 10

<210> 96

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 96

gagttaaatg 10

<210> 97

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 97

ttgtaagttg 10

<210> 98

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 98

tgggggtagg 10

<210> 99

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 99

gttcttaggg 10

<210> 100

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 100

gtattctaag 10

<210> 101

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 101

ggaggttttg 10

<210> 102

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 102

agaatatgta 10

<210> 103

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 103

atctttcggg 10

<210> 104

<211

> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 104

ttgcattgaa 10

<210> 105

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 105

ggtgggattt 10

<210> 106

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 106

tttatctaat 10

<210> 107

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 107

gcgggtggtg 10

<210> 108

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 108

ggtttagata 10

<210> 109

<211> 10

<212>

DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 109

tttatgcgtt 10

<210> 110

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 110

tgggtaaggc 10

<210> 111

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 111

gggggtggtc 10

<210> 112

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 112

gtagtatatt 10

<210> 113

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 113

ggaggtattt 10

<210> 114

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 114

gtattgtaag 10

<210> 115

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 115

tttacgggag 10

<210> 116

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 116

tagttctggg 10

<210> 117

<211

> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 117

ccacgtctat 10

<210> 118

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 118

agtgggtagg 10

<210> 119

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 119

caatttttac 10

<210> 120

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 120

ggtctggggg 10

<210> 121

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 121

atcaagattg 10

<210> 122

<211> 10

<212>

DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 122

gttagctaaa 10

<210> 123

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 123

agtgtggggt 10

<210> 124

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 124

ggtatgtggg 10

<210> 125

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 125

gtagtgtggg 10

<210> 126

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 126

aggaggtgtt 10

<210> 127

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 127

gttggtaggt 10

<210> 128

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 128

gtaggtggtt 10

<210> 129

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 129

aggtgttggt 10

<210> 130

<211

> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 130

tatggttgtg 10

<210> 131

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 131

ttaggttagt 10

<210> 132

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 132

gattggagtt 10

<210> 133

<211> 10

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 133

gtagagtgga 10

<210> 134

<211> 24

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 134

cucuuucuuu uccauggguu ggcu 24

<210> 135

<211> 24

<212>

RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 135

ggcugaggga aggacugucc uggg 24

<210> 136

<211> 13

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 136

ggguuauggg acc 13

<210> 137

<211> 12

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 137

auauccuuuu ua 12

<210> 138

<211> 12

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 138

guauccuuuu ua 12

<210> 139

<211> 33

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 139

aggcuucgga gcaaggaggc agcuccgaag ccu 33

<210> 140

<211> 33

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 140

aggcuucgga gcaagccucc agcuccgaag ccu 33

<210> 141

<211> 29

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 141

gucgaggccg agcgggcaaa ggccucgac 29

<210> 142

<211> 29

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 142

gucgaggccg agcccgcaaa ggccucgac 29

<210> 143

<211

> 10

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<220><221> misc_feature

<222> (1)..(3)

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

<220><221> misc_feature

<222> (8)..(10)

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

<400> 143

nnnaggunnn 10

<210> 144

<211> 12

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 144

uuuuccuuaa cu 12

<210> 145

<211> 1305

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 145

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420
ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480
acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcc 540
accatggtga gcaagggcga ggagctgttc accggggtgg tgcccatcct ggtcgagctg 600
gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg cgatgccacc 660
tacggcaagc tgaccctgaa gttcatctgc accaccggca agctgcccgt gccctggccc 720

accctcgtga ccaccttcgg ctacggcctg atgtgcttcg cccgctaccc cgaccacatg 780
aagcagcacg acttcttcaa gtccgccatg cccgaaggct acgtccagga gcgcaccatc 840
ttcttcaagg taagtattag ctctttcttt ccatgggttg gcctcgccgc gtgggctgag 900
ggaaggactg tcctgggact ggacaggcgg gttatgggac ctgaaaagcg gccctgaaaa 960
agggccgcga tctgtagaaa gcgagctagt gccggacagt tagaggaaaa ggggaagaac 1020
tgtccgaaaa aaggggggga agacagtgac tagaaaggga agggagaagt cactgtagag 1080
gggaaggaaa aggctagcta gaggagaagg aaagaggcta gctagcagag gagaaggaaa 1140

ggcgccagca gttcggtgct atcaaaaagc ggtcaggcag ctaaaccaaa aggtttagca 1200
attgcctctg atgagtcgct gaaatgcgac gaaaaccgct ttttggtacc aataaaatat 1260
ctttattttc attacatctg tgtgttggtt ttttgtgtga ctagt 1305

<210> 146
<211> 1543
<212> DNA
<213> Artificial Sequence
<220><223> Synthetic nucleic acid sequence
<400> 146

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60
gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180
aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240
catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300
catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360
atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcg 540

gaaaaccgcg ggatagcacc gaactgctgg cgcctttcct tctcctctgc tagctagcct 600

ctttccttct cctctagcta gccttttcct tcccctctac agtgacttct cccttccctt 660

tctagtcact gtcttccccc ccttttttcg gacagttctt ccccttttcc tctaactgtc 720

cggcactagc tcgctttcta cagatcatta ttgcggccct gaaaaagggc cgcttataac 780

gttgctcgaa ttcgggttat gggaccagtg aaggctgagg gaaggactgt cctgggactg 840

gacaggcggg ttatgggacc tgaaaatact aacaatcgat ttttttttccc ttttttttcca 900

ggacgacggc aactacaaga cccgcgccga ggtgaagttc gagggcgaca ccctggtgaa 960

ccgcatcgag ctgaagggca tcgacttcaa ggaggacggc aacatcctgg ggcacaagct 1020

ggagtacaac tacaacagcc acaacgtcta tatcatggcc gacaagcaga agaacggcat 1080

caaggtaagt attagctctt tctttccatg ggttggcctc gccgcgtggg ctgagggaag 1140

gactgtcctg ggactggaca ggcgggttat gggacctgaa aagcggccct gaaaaagggc 1200

cgcagcgaaa acgaagcgag ctaaagcctc ctctctcttc ttcagaactc ctctcttttc 1260

tctcctccag gagttcttcc tctctccctt cttctcaaat gctttctccc tctctcctgc 1320

atttgagctc cttctttcct ctctcgacaa tccccttttc tccctcttga ttgtcgacta 1380

gctcgcaatc atcgcggtgc taaaaagcgg tcaggcagct aaaccaaaag gtttagcaat 1440

tgcctctgat gagtcgctga aatgcgacga aaaccgcttt ttggtaccaa taaaatatct 1500

ttattttcat tacatctgtg tgttggtttt ttgtgtgact agt 1543

<210> 147

<211> 1571

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 147

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccacc 540

atgagccagt tcgacatcct gtgcaagacc ccccccaagg tgctggtgcg gcagttcgtg 600

gagagattcg agaggcccag cggcgagaag atcgccagct gtgccgccga gctgacctac 660

ctgtgctgga tgatcaccca caacggcacc gccatcaaga gggccacctt catgagctac 720

aacaccatca tcagcaacag cctgagcttc gacatcgtga acaagagcct gcagttcaag 780

tacaagaccc agaaggccac catcctggag gccagcctga agaagctgat ccccgcctgg 840

gagttcacca tcatccctta caacggccag aagcaccaga gcgacatcac cgacatcgtg 900

tccagcctgc agctgcagtt cgagagcagc gaggaggccg acaagggcaa cagccacagc 960

aagaagatgc tgaaggccct gctgtccgag ggcgagagca tctgggagat caccgagaag 1020

atcctgaaca gcttcgagta caccagcagg ttcaccaaga ccaagaccct gtaccagttc 1080

ctgttcctgg ccacattcat caactgcggc aggtaagtat tagctctttc tttccatggg 1140

ttggcctcgc cgcgtgggct gagggaagga ctgtcctggg actggacagg cgggttatgg 1200

gacctgaaaa gcggccctga aaaagggccg cgatgaaaac gaagcgagct aaagcctcct 1260

ctctcttctt cagaactcct ctcttttctc tcctccagga gttcttcctc tctcccttct 1320

tctcaaatgc tttctccctc tctcctgcat ttgagctcct tctttcctct ctcgacaatc 1380

ccctttctc cctcttgatt gtcgactagc tcgcaatcat cgcggtatca aaaagcggtc 1440

aggcagctaa accaaaaggt ttagcaattg cctctgatga gtcgctgaaa tgcgacgaaa 1500

accgcttttt ggtaccaata aaatatcttt attttcatta catctgtgtg ttggtttttt 1560

gtgtgactag t 1571

<210> 148

<211> 1765

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 148

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcg 540

gaaaaccgcg ggataccgcg atgattgcga gctagtcgac aatcaagagg gagaaaaggg 600

gattgtcgag agaggaaaga aggagctcaa atgcaggaga gagggagaaa gcatttgaga 660

agaagggaga gaggaagaac tcctggagga gagaaaagag aggagttctg aagaagagag 720

aggaggcttt agctcgcttc gttttcatca ttattgcggc cctgaaaaag ggccgcttat 780

aacgttgctc gaattcgggt tatgggacca gtgaaggctg agggaaggac tgtcctggga 840

ctggacaggc gggttatggg acctgaaaat actaacaatc gatttttttt ccctttttt 900

ccaggttcag cgacatcaag aacgtggacc ccaagagctt caagctggtg cagaacaagt 960

acctgggcgt gatcattcag tgcctggtga ccgagaccaa gacaagcgtg tccaggcaca 1020

tctacttttt cagcgccaga ggcaggatcg acccccctggt gtacctggac gagttcctga 1080

ggaacagcga gcccgtgctg aagagagtga acaggaccgg caacagcagc agcaacaagc 1140

aggagtacca gctgctgaag gacaacctgg tgcgcagcta caacaaggcc ctgaagaaga 1200

acgcccccta ccccatcttc gctatcaaga acggccctaa gagccacatc ggcaggcacc 1260

tgatgaccag ctttctgagc atgaagggcc tgaccgagct gacaaacgtg gtgggcaact 1320

ggagcgacaa gagggcctcc gccgtggcca ggaccaccta cacccaccag atcaccgcca 1380

tccccgacca ctacttcgcc ctggtgtcca ggtactacgc ctacgacccc atcagcaagg 1440

agatgatcgc cctgaaggac gagaccaacc ccatcgagga gtggcagcac atcgagcagc 1500

tgaagggcag cgccgagggc agcatcagat accccgcctg gaacggcatc atcagccagg 1560

aggtgctgga ctacctgagc agctacatca acaggcggat ctgagaattc ctcacctgcg 1620

atctcgatgc tttatttgtg aaatttgtga tgctattgct ttatttgtaa ccattataag 1680

ctgcaataaa caagttaaca acaacaattg cattcatttt atgtttcagg ttcaggggga 1740

ggtgtgggag gtttttttaaa ctagt 1765

<210> 149

<211> 10

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 149

aaagaaggaa 10

<210> 150

<211> 12

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 150

cuuucuuuuc uu 12

<210> 151

<211> 11

<212>

RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<220><221> misc_feature

<222> (1)..(3)

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

<220><221> misc_feature

<222> (8)..(11)

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

<400> 151

nnnaggunnn n 11

<210> 152

<211> 11

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<220><221> misc_feature

<222> (1)..(3)

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

<220><221

> misc_feature

<222> (8)..(11)

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

<400> 152

nnnuggunnn n 11

<210> 153

<211> 11

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<220><221> misc_feature

<222> (3)..(8)

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

<400> 153

gannnnnnaa a 11

<210> 154

<211> 12

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic nucleic acid sequence

<400> 154

gccgccacca tg 12

<210> 155

<211> 4311

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 155

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcc 540

gccaccatgg ctctgatcgt gcacctgaaa accgtgtccg agctgagagg caagggcgac 600

agaatcgcca aagtgacctt cagaggccag agcttctaca gcagagtgct ggaaaactgc 660

gaaggcgtgg ccgacttcga cgagacattc agatggcctg tggccagcag catcgacaga 720

aacgaggtgc tcgagatcca gatcttcaac tacagcaagg tgttcagcaa caagctgatc 780

gggaccttct gcatggtgct gcagaaagtg gtggaagaga accgcgtgga agtgaccgac 840

acactgatgg acgacagcaa cgccatcatc aagaccagcc tgagcatgga agtgcgctac 900

caggccacag atggcacagt cggaccttgg gacgatggcg atttcctggg agatgagagc 960

ctgcaagagg aaaaggacag ccaagagaca gacggcctgc tgcctggctc tcggcctagc 1020

acaagaatca gcggcgagaa gtccttcaga agcaagggca gagaaaagac caaaggcggc 1080

agagatggcg agcacaaggc tggcagatct gtgttcagcg ccatgaagct gggcaagacc 1140

agaagccaca aagaggaacc ccagagacag gacgagccag ccgttctgga aatggaagat 1200

ctcgaccatc tggccatcca gctcggcgac ggacttgacc ctgattctgt gtctctggcc 1260

agcgtgacag ccctgacaag caacgtgtcc aacaagagaa gcaagcccga catcaagatg 1320

gaacccagcg ccggcagacc catggattac caggtgtcca tcaccgtgat cgaggccaga 1380

cagctcgtgg gcctgaacat ggatcctgtc gtgtgtgtgg aagtgggcga cgacaaaaag 1440

tacaccagca tgaaggaaag caccaactgt ccctactaca acgagtactt cgtgttcgac 1500

ttccacgtgt ccccagacgt gatgttcgac aagatcatta agatcagcgt gatccacagc 1560

aagaacctgc tgagaagcgg cacactcgtg ggcagcttta agatggacgt gggcaccgtg 1620

tacagccagc cagagcacca gtttcaccac aagtgggcca tcctgagcga ccccgatgat 1680

atctctgctg gcctgaaggg ctacgtgaag tgtgatgtgg ctgtcgtcgg caaaggcgac 1740

aacatcaaga caccccacaa ggccaacgag actgacgagg acgatatcga gggcaacctg 1800

ctgctgccag aaggcgtgcc accagaaaga cagtgggcca gattctatgt gaagatctac 1860

agagccgagg gcctgcctag aatgaacaca agcctgatgg ccaacgtgaa gaaggctttc 1920

atcggcgaga acaaggacct ggtggacccc tacgtccagg tgttcttcgc tggacagaaa 1980

ggcaagacct ccgtgcagaa gtccagctac gagcccctgt ggaacgaaca ggtggtgttc 2040

accgatctgt tccctccact gtgcaagaga atgaaggtgc agatccggga cagcgacaaa 2100

gtgaacgatg tggccatcgg cacccacttc atcgacctga gaaagatcag caacgacggc 2160

gacaagggct tcctgcctac acttggacct gcctgggtca acatgtacgg cagcaccaga 2220

aactacaccc tgctggacga gcaccaggac ctgaacgaag gactcggaga gggcgtgtcc 2280

ttccgggcta gactgatgct gggactcgcc gtggaaatcc tggacacaag caaccctgag 2340

ctgaccagca gcacagaggt gcaggttgaa caggccacac ctgtgtctga gagctgcacc 2400
ggcagaatgg aagagttctt cctgttcggc gccttcctgg aagcctccat gatcgataga 2460
aagaacggcg ataagcccat caccttcgaa gtgaccatcg gcaactacgg caacgaggtg 2520
gacggcatgt ctagacccct ccggcctaga ccaagaaaag agcccggcga cgaggaagag 2580
gtggacctga tccagaacag cagcgacgat gagggcgacg aagctggcga tctggcaagc 2640
gttagcagca cccctcctat gaggccccag atcaccgacc ggaactactt tcatctgccc 2700
tacctggaaa gaaagccctg catctacatc aagagctggt ggcctgacca gagaaggcgg 2760

ctgtacaacg ctaacatcat ggaccatatc gccgacaagc tggaagaggg actgaacgac 2820
gtccaagaga tgatcaagac cgagaagtct taccccgaga gaaggctgag gggcgtgctc 2880
gaggaactga gctgtggatg ccacagattt ctgagcctgt ccgacaagga ccagggcaga 2940
agcagcagaa ccagactgga tagagagcgg ctgaagtcct gcatgcgcga gctggaatct 3000
atgggccagc aggccaagag cctgagagcc caagtgaaga gacacaccgt gcgggacaag 3060
ctgagatcct gccagaactt cctgcagaag ctgcggttcc tggccgatga gcctcagcac 3120
tctatccccg acgtgttcat ctggatgatg agcaacaaca agaggatcgc ctacgccaga 3180

gtgcccagca aggatctgct gtttagcatc gtggaagagg aactcggcaa ggactgcgcc 3240
aaagtcaaga ccctgttcct gaagctgcca ggcaagagag gcttcggctc tgctggatgg 3300
acagtgcagg ctaagctgga actgtacctg tggctgggcc tgagcaagca gagaaaggac 3360
ttcctgtgcg gcctgccttg cggcttcgaa gaagtgaagg ctgctcaagg cctgggcctg 3420
cacagcttcc ctccaatctc tctggtgtac acaaagaagc aggccttcca gctgagggcc 3480
cacatgtacc aggctagatc tctgttcgcc gccgactcta gcggcctgtc tgatcctttc 3540
gctcgggtgt tcttcatcaa ccagagccag tgcaccgagg tgctgaacga gacactgtgt 3600

cctacctggg accagatgct ggtctttgac aacctcgagc tgtacggcga ggctcacgaa 3660
ctgagagatg accctcctat catcgtcatc gagatctacg accaggacag catgggcaaa 3720
gccgacttca tgggcagaac cttcgccaag cctctggtca agatggccga cgaggcttac 3780
tgccctcctc ggttcccacc tcagctcgag tactaccaga tctaccgggg ctctgctaca 3840
gccggcgatc tgctggctgc ttttgagctg ctgcaaatcg gccctagcgg caaggctgat 3900
ctgcctccaa tcaacggccc tgtggacatg gacagaggcc ccattatgcc tgtgcctgtg 3960
ggcatcagac ccgtgctgag caagtacaga gtggaagtgc tgttttgggg cctgcgcgac 4020

ctgaagagag tgaacctggc tcaggtaagt attagctctt tctttccatg ggttggcctc 4080

gccgcgtggg ctgagggaag gactgtcctg ggactggaca ggcgggttat gggacctgaa 4140

gcgataaaag gcatgcacgt ttgcggctac gtgcatgcca aaaggagtcg ggcttgcctc 4200

cgtgcccgac tccaaaagac ctgctcgagg aggtggacga gcaggtcaaa aatccgggta 4260

ccaataaaat atctttattt tcattacatc tgtgtgttgg ttttttgtgt g 4311

<210> 156

<211> 3467

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 156

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttaggatttt tgacctgctc 540

gattgtccac tgcgagcagg tcttttggag tcgggcgagg cggaagcccg actccttttg 600

gcatgcacgc tagccgcgtc gtgcatgcct tttatcttcg ggttatggga ccagtgaagg 660

ctgagggaag gactgtcctg ggactggaca ggcgggttat gggacctgaa aatactaaca 720

atcgattttt tttccctttt tttccaggtg gacagaccca gagtggatat cgagtgtgct 780

ggcaaagggg tgcagagcag cctgatccat aactacaaga agaaccccaa cttcaacacc 840

ctggtcaagt ggttcgaagt ggatctgccc gagaacgaac tgctgcaccc acctctgaac 900

atcagagtgg tggactgcag agccttcggc agatacaccc tcgtgggatc tcacgccgtg 960

tctagcctga gaagattcat ctacagacct ccagacagaa gcgcccctaa ctggaacaca 1020

acaggcgagg tggtggtgtc catggaaccc gaggaacccg tgaagaaact ggaaaccatg 1080

gtcaagctgg acgccacctc cgatgctgtc gtgaaagtgg acgtggccga ggacgagaaa 1140

gagcgcaaga agaagaaaaa gaagggcccc agcgaggaac ctgaagagga agaacctgac 1200

gagagcatgc tggactggtg gtccaagtac ttcgcctcca tcgacacaat gaaggaacag 1260

ctgagacagc acgagacaag cggcaccgac ctcgaagaga aagaagagat ggaatccgcc 1320
gaaggactga agggccctat gaagtccaaa gagaagtcta gggccgccaa agaagagaaa 1380
aaaaagaaga accagtctcc tggaccaggc cagggatctg aggctcccga aaagaaaaag 1440
gccaagatcg acgagctgaa ggtgtacccc aaagagctgg aaagcgagtt cgacagcttc 1500
gaggactggc tgcacacctt caatctgctg agaggaaaga caggcgacga cgaggatggc 1560
agcactgaag aagagagaat cgtcggcaga ttcaagggca gcctgtgcgt gtacaaggtg 1620
ccactgcctg aggacgtgtc cagagaggct ggctacgatc ctacctacgg catgttccaa 1680

ggcatcccta gcaacgaccc catcaatgtg ctcgtgcgga tctatgtcgt gcgggccact 1740
gatctgcatc ccgccgatat caacggcaag gcagacccct atatcgctat caagctgggg 1800
aaaaccgaca tcagggacaa agagaactac atcagcaagc agctgaaccc cgtgttcggc 1860
aagagcttcg acatcgaggc tagcttcccc atggaatcca tgctgaccgt ggccgtgtac 1920
gactgggatc tcgtgggaac agacgacctg atcggagaga caaagattga cctggaaaac 1980
cggttctact ccaagcaccg ggccacctgt ggaatcgccc agacctactc tatccacggc 2040
tacaacatct ggcgggaccc catgaagcct agccagatcc tgaccaggct gtgcaaagaa 2100

ggcaaggtcg acggccctca ctttggacct cacggccggg tcagagtggc caacagagtg 2160
ttcacaggcc cctccgagat cgaggatgag aacggccaga gaaagcccac cgatgagcat 2220
gtggctctga gcgctctgag acactgggaa gatatcccta gagtgggctg cagactggtg 2280
cccgagcacg tggaaacaag acccctgctg aacccagaca agcccggaat cgaacagggc 2340
agactcgaac tgtgggtcga catgttccct atggacatgc ccgcacctgg cacaccactg 2400
gacatcagcc ctaggaagcc caagaaatac gagctgcgcg tgatcgtgtg gaacaccgac 2460
gaagtggtgc tggaagatga cgacttcttc accggcgaaa agtccagcga catcttcgtc 2520

agaggatggc tgaagggaca gcaagaggat aagcaggaca ccgacgtgca ctaccacagc 2580
cttacaggcg aaggcaactt taactggcgc tacctgtttc ctttcgacta cctggccgcc 2640
gaagagaaga tcgtgatgtc caagaaagaa tctatgttca gctgggacga gacagagtac 2700
aagatccccg ccagactgac cctgcagatc tgggatgccg atcacttcag cgccgacgac 2760
tttctgggag ccatcgagct ggacctgaat agattcccca gaggcgccaa gaccgccaag 2820
cagtgcacaa tggaaatggc cactggcgag gtcgacgtgc cactggtgtc tatcttcaag 2880
cagaagcgcg tcaaaggctg gtggcccctg ctggctagaa acgagaacga cgagttcgag 2940

ctgaccggaa aggtggaagc cgagctgcat ctgctgacag ctgaagaggc cgagaagaat 3000

cctgtgggcc tcgctaggaa tgagcccgat cctctggaaa agcccaacag acccgatacc 3060

gccttcgtgt ggtttctgaa cccactgaag tccatcaagt acctgatctg tacccggtac 3120

aagtggctga ttatcaagat cgtgctggcc ctgctggggc tgctgatgct tgctctgttc 3180

ctgtactccc tgcctggcta tatggtcaag aagctgctgg gcgccggcgc tcgggctgac 3240

tacaaagacc atgacggtga ttataaagat catgacatcg actataagga tgacgatgac 3300

aaatgaggta ccaattcctc acctgcgatc tcgatgcttt atttgtgaaa tttgtgatgc 3360

tattgcttta tttgtaacca ttataagctg caataaacaa gttaacaaca acaattgcat 3420

tcattttatg tttcaggttc agggggaggt gtgggaggtt ttttaaa 3467

<210> 157

<211> 4392

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 157

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgaccttа tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcc 540

accatggtca ttctgcagca gggcgaccac gtgtggatgg atctgagact gggccaagag 600

ttcgacgtgc caatcggcgc cgtggtcaag ctgtgtgatt ctggccaggt gcaagtcgtg 660

gacgacgagg ataatgagca ctggatcagc cctcagaacg ccacacacat caagcctatg 720

caccccacat ctgtgcacgg cgtggaagat atgatccggc tgggcgatct gaacgaggcc 780

ggcatcctga gaaacctgct gatcagatac cgggaccacc tgatctacac ctacaccggc 840

tctatcctgg tggccgtgaa tccctaccag ctgctgagca tctacagccc cgagcacatc 900

cggcagtaca ccaacaagaa aatcggcgag atgcctcctc acatcttcgc cattgccgac 960

aactgctact tcaacatgaa gcggaacagc cgggaccagt gctgcatcat ctctggcgaa 1020

tctggcgccg gaaagaccga gagcacaaag ctgatcctgc agttcctggc cgccatcagc 1080
ggacagcact cttggattga gcagcaggtc ctggaagcca cacctattct ggaagccttc 1140
ggcaacgcca agaccatccg gaacgacaac agcagcagat tcggcaaata catcgacatc 1200
cacttcaaca agagaggcgc cattgagggc gccaagatcg agcagtacct gctggaaaag 1260
tccagagtgt gcagacaggc cctggacgag agaaactacc acgtgttcta ctgcatgctg 1320
gaaggcatga gcgaggacca gaagaagaag ctcggactcg gccaggccag cgactacaat 1380
tatctggcca tgggcaactg catcacatgc gagggcagag tggacagcca agagtacgcc 1440

aacatccgca gcgccatgaa ggtgctgatg ttcaccgaca ccgagaactg ggagatcagc 1500
aaactgctgg ccgctatcct gcatctgggc aacctgcagt acgaggccag aaccttcgag 1560
aacctggatg cctgcgaggt gctgttctct ccttccctgg ctaccgccgc ctctctgctg 1620
gaagtgaacc ctcctgatct gatgagctgc ctgaccagca gaaccctgat caccagaggc 1680
gagacagtgt ctacccctct gagcagagaa caggctctgg atgtgcggga cgccttcgtg 1740
aagggcatct acggcagact gttcgtgtgg atcgtggaca agatcaacgc cgccatctac 1800
aagcctccaa gccaggacgt gaagaacagc agaagatcca tcggcctgct ggacatcttc 1860

ggcttcgaga atttcgccgt gaacagcttc gagcagctgt gcatcaactt cgccaacgag 1920
cacctccagc agttcttcgt gcggcacgtg ttcaagctgg aacaagagga atacgacctg 1980
gaatccatcg actggctgca catcgagttc accgataacc aggacgccct ggacatgatc 2040
gccaacaagc ccatgaacat catcagcctg atcgacgagg aaagcaagtt ccccaagggc 2100
accgatacca ccatgctgca caagctgaac agccagcaca aactgaatgc caactacatc 2160
ccgcctaaga acaaccacga gacacagttc ggcatcaacc acttcgccgg catcgtgtac 2220
tacgaaaccc agggctttct ggaaaagaac cgggacaccc tgcacggcga catcattcag 2280

ctggtgcaca gcagccggaa caagttcatc aagcagatct tccaggccga cgtcgccatg 2340
ggagccgaga caagaaagag aagccccaca ctgagcagcc agttcaagcg gagtctggaa 2400
ctgctgatga gaaccctggg agcctgccag cctttctttg tgcggtgcat caagcccaac 2460
gagttcaaga aacccatgct gttcgaccgg cacctgtgtg tgcggcagct gagatacagc 2520
ggcatgatgg aaaccatcag gattcggaga gccggctatc ccatccggta cagcttcgtg 2580
gaattcgtcg agcggtacag agtgctgctg cctggcgtga agcctgccta caaacagggc 2640
gatctcagag gcacctgtca gagaatggcc gaagccgtgc tgggcaccca tgacgattgg 2700

cagatcggaa agacaaagat cttcctgaag gaccaccacg acatgctgct cgaggtggaa 2760

agagacaagg ccatcaccga cagagtgatc ctgctccaga aagtgatccg gggcttcaag 2820

gacagaagca atttcctgaa gctgaagaat gccgccactc tgatccagag acactggcgg 2880

ggacacaact gccggaagaa ctacggcctg atgaggctgg gcttcctgag actgcaggcc 2940

ctgcacagaa gcagaaagct gcaccagcag tacagactgg cccggcagcg gatcatccag 3000

tttcaagcca gatgtcgggc ctacctcgtg cgcaaggcct tcagacatag actgtgggcc 3060

gtgctgaccg tgcaggccta tgccagagga atgattgccc gcagactgca ccagagactg 3120

agagccgagt atctgtggcg gctggaagcc gagaaaatgc ggctggccga ggaagagaag 3180

ctgcggaaag agatgagcgc caagaaggcc aaagaagagg ccgagcggaa gcaccaagag 3240

agactggctc aactggccag agaggacgcc gagagagagc tgaaagagaa agaggccgcc 3300

agacggaaga aagaactcct ggaacagatg gaacgggcca gacacgagcc cgtgaaccac 3360

agcgatatgg tggataagat gttcggcttc ctgggcacct ctggcggact gcctggacaa 3420

gaaggacagg cccctagcgg ctttgaggac ctggaacgtg ggagaagaga aatggtggaa 3480

gaggatctgg acgccgctct gcctctgcct gacgaggatg aagaagatct gagcgagtac 3540

aagttcgcca agtttgccgc cacctacttt caaggcacca ccacacacag ctacaccaga 3600

aggcctctga agcagcccct gctgtaccac gatgatgagg gcgatcaact ggcagccctg 3660

gccgtgtgga ttaccatcct cagattcatg ggcgacctgc ctgagcctaa gtaccacacc 3720

gccatgtctg acggctccga gaagatcccc gtgatgacca agatctacga gactctgggc 3780

aagaaaacct acaagcgcga gctgcaggct ctccaaggcg aaggcgaagc tcaactgcct 3840

gagggccaga aaaagtcctc tgtgcgccac aaactggtgc acctgacact gaagaagaaa 3900

agcaagctga cagaggaagt gaccaagcgg ctgcacgatg gcgagtctac agtgcagggc 3960

aacagcatgc tcgaggacag acccaccagc aacctggaaa aactgcactt catcatcggc 4020

aacggaatcc tgcggcctgc tctgagggat gagatctact gccagatctc caagcagctg 4080

acacacaacc ccagcaagag cagctacgcc agaggctgga ttctggtaag tattagctct 4140

ttctttccat gggttggcct cgccgcgtgg gctgagggaa ggactgtcct gggactggac 4200

aggcgggtta tgggacctga agcgataaaa ggcatgcacg tttgcggcta cgtgcatgcc 4260

aaaaggagtc gggcttgcct ccgtgcccga ctccaaaaga cctgctcgag gaggtggacg 4320

agcaggtcaa aaatccgggt accaataaaa tatctttatt ttcattacat ctgtgtgttg 4380

gtttttgtg tg 4392

<210> 158

<211> 4055

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 158

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttaggatttt tgacctgctc 540

gattgtccac tgcgagcagg tcttttggag tcgggcgagg cggaagcccg actccttttg 600

gcatgcacgc tagccgcgtc gtgcatgcct tttatcttcg ggttatggga ccagtgaagg 660

ctgagggaag gactgtcctg ggactggaca ggcgggttat gggacctgaa aatactaaca 720

atcgattttt tttccctttt tttccaggtg tctctgtgcg tgggctgttt cgccccaagc 780

gagaagttcg tgaagtacct gaggaacttc atccacggcg gacctccagg ctacgccccc 840

tactgtgaag agaggctgag aaggaccttt gtgaacggca cccggacaca gcctccatcc 900

tggctggaac tccaggccac caagagcaaa aagcccatca tgctgcccgt gacctttatg 960

gatggcacca caaagaccct gctgaccgat agcgccacca ccgccaaaga gctgtgtaac 1020

gccctggctg acaagattag cctgaaggat agattcggct tcagcctgta cattgccctg 1080

ttcgacaagg tgtccagcct cggctctggc tctgaccatg tgatggatgc catcagccag 1140

tgcgagcagt atgccaaaga acagggcgcc caagagagga acgctccttg gcggctgttc 1200

tttcggaaag aggtgttcac cccttggcac agccccagcg aagataacgt ggccaccaat 1260

ctgatctacc agcaagttgt gcggggcgtg aagttcggcg agtacagatg cgaaaaagag 1320

gacgatctgg ccgagctggc ctctcagcag tactttgtgg actacggcag cgagatgatc 1380

ctggaacggc tgctgaatct ggtgcccacc tacattcccg atcgggagat caccccactg 1440

aaaaccctcg agaagtgggc ccagctggcc attgctgccc acaagaaagg catctatgcc 1500

cagcggagaa cagacgccca gaaagtcaaa gaggatgtcg ttagctacgc ccggttcaag 1560

tggcctctgc tgtttagccg gttctacgag gcctacaagt tcagcggccc cagtctgccc 1620

aagaacgatg tgatcgtggc tgtgaactgg accggcgtgt acttcgtgga tgagcaagaa 1680

caagtgctgc ttgagctgag cttccccgag atcatggccg tgtccagctc cagagaatgc 1740

agagtgtggc tgagcctggg ctgtagcgat ctgggatgtg ccgctcctca ttctggatgg 1800

gctggactga caccagccgg accttgtagc ccttgttggt cttgccgggg ggccaagaca 1860

acagccccta gctttaccct ggccaccatt aagggcgacg agtacacctt caccagcagc 1920

aacgccgagg acatcagaga tctggtcgtg accttcctgg aaggcctgcg gaagcggagc 1980

aaatatgtgg tggccctgca ggacaacccc aatcctgctg gcgaggaatc cggctttctg 2040

agctttgcca aaggcgacct gatcatcctg gaccacgaca ccggcgagca agtgatgaat 2100

agcggctggg ccaacggcat caatgagcgg acaaagcagc ggggcgactt ccctaccgat 2160

agcgtgtacg tgatgcccac cgtgaccatg cctccaaggg aaatcgtggc cctggtcacc 2220

atgacacccg accagagaca ggatgttgtg cggctgctgc agctgaggac agccgaacca 2280

gaagtgcggg ccaagcctta cacactggaa gagttcagct acgactactt ccggcctcct 2340

ccaaagcaca ccctgtctag agtgatggtg tccaaggcca gaggcaagga taggctgtgg 2400

tcccacacaa gagagcccct gaaacaggca ctgctgaaaa agctgctggg cagcgaggaa 2460

ctgagccaag aagcctgtct ggcctttatc gccgtgctga agtacatggg cgattacccc 2520

tccaagcgga ccagatccgt gaacgaactg accgaccaga ttttcgaggg cccactgaag 2580

gccgagcctc tgaaagatga ggcctacgtg cagattctga aacagctgac cgacaaccac 2640

atccgctaca gcgaggaacg cggatgggaa ctgctgtggc tgtgtaccgg actgttccca 2700

cctagcaaca ttctgctgcc ccacgtgcag cggtttctgc agtctagaaa gcactgccct 2760

ctggccatcg attgcctgca gaggctgcaa aaggccctga gaaatggctc ccggaagtac 2820

cctcctcacc tggtggaagt ggaagccatc cagcacaaga ccacacagat ctttcacaag 2880

gtctacttcc ccgacgacac agacgaggcc tttgaggtgg aatcctctac caaggccaag 2940

gacttctgcc agaatatcgc caccaggctg ctgctgaagt ccagcgaagg ctttagcctg 3000

tttgtgaaga tcgccgacaa agtgctgagc gtgcccgaga acgacttctt tttcgatttt 3060

gtgcgccatc tgaccgactg gattaagaag gctagaccca tcaaggatgg catcgtgccc 3120

agcctgacct atcaggtgtt ctttatgaag aagctgtgga cgaccaccgt gcctggcaag 3180

gatcctatgg ccgacagcat cttccactac taccaagagc tgcccaagta cctgcggggc 3240

taccacaagt gtaccagaga agaggtcctg cagctgggag ccctgatcta tagagtgaag 3300

tttgaagagg acaagagcta cttccctagc atccccaagc tgctgcgcga actggttccc 3360

caggatctga tccggcaagt gtcccctgat gactggaagc ggtctatcgt ggcctacttt 3420

aacaagcacg ccggcaagag taaagaggaa gccaagctgg cctttctgaa gctcatcttt 3480

aagtggccta ccttcggctc cgccttcttc gaagtgaagc agaccaccga gcctaacttc 3540

cctgagattc tgctgatcgc catcaacaaa tacggcgtgt ccctgatcga tcccaagaca 3600

aaggacatcc tgacaacaca ccccttcacc aaaatcagca actggtccag cggcaacacc 3660

tacttccaca tcaccatcgg caatctcgtg cggggctcta agctgctgtg tgaaaccagc 3720

ctgggataca agatggacga cctgctgaca agctacatct cccagatgct gaccgccatg 3780

agcaaacaga gaggctctcg gagcggcaag tggggcgctc gggctgacta caaagaccat 3840

gacggtgatt ataaagatca tgacatcgac tataaggatg acgatgacaa atgaggtacc 3900

aattcctcac ctgcgatctc gatgctttat ttgtgaaatt tgtgatgcta ttgctttatt 3960

tgtaaccatt ataagctgca ataaacaagt taacaacaac aattgcattc attttatgtt 4020

tcaggttcag ggggaggtgt gggaggtttt ttaaa 4055

<210> 159

<211> 4161

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 159

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcc 540

accatgaaga gaaccgccga cggcagcgag ttcgagagcc ctaagaaaaa gcggaaggtg 600

gacaagaagt acagcatcgg cctggctatc ggcaccaatt ctgttggctg ggccgtgatc 660

accgacgagt acaaggtgcc cagcaagaaa ttcaaggtgc tgggcaacac cgaccggcac 720

agcatcaaga agaatctgat cggcgccctg ctgttcgact ctggcgaaac agccgaagcc 780

accagactga agaggacagc cagacggcgg tacaccagaa gaaagaaccg gatctgctac 840

ctgcaagaga tcttcagcaa cgagatggcc aaggtggacg acagcttctt ccaccggctg 900

gaagagtcct tcctggtgga agaggataag aagcacgagc ggcaccccat cttcggcaac 960

atcgtggatg aggtggccta ccacgagaag taccccacca tctaccacct gagaaagaaa 1020

ctggtggaca gcaccgacaa ggccgacctg agactgatct atctggccct ggctcacatg 1080

atcaagttcc ggggccactt cctgatcgag ggcgacctga atcctgacaa cagcgacgtg 1140

gacaagctgt tcatccagct ggtgcagacc tacaaccagc tgttcgagga aaaccccatc 1200

aacgccagcg gagtggatgc caaggccatc ctgtctgccc ggctgagcaa gagcagacgg 1260

ctggaaaacc tgatcgctca gctgcccggc gagaagaaga atggcctgtt cggcaacctg 1320

attgccctga gcctgggcct gacacctaac ttcaagagca acttcgacct ggccgaggac 1380

gccaaactgc agctgtccaa ggacacctac gacgacgacc tggacaatct gctggcccag 1440

atcggcgatc agtacgccga cttgtttctg gccgccaaga acctgtccga cgccatcctg 1500

ctgagcgaca tcctgagagt gaacaccgag atcacaaagg cccctctgag cgcctctatg 1560

atcaagagat acgacgagca ccaccaggat ctgaccctgc tgaaggccct cgttagacag 1620

cagctgcctg agaagtacaa agagattttc ttcgaccaga gcaagaacgg ctacgccggc 1680

tacattgatg gcggagccag ccaagaggaa ttctacaagt tcatcaagcc catcctcgag 1740

aagatggacg gcaccgagga actgctggtc aagctgaaca gagaggacct gctgcggaag 1800

cagcggacct tcgacaatgg ctctatccct caccaaatcc acctgggaga gctgcacgcc 1860

attctgcgga gacaagagga cttttaccca ttcctgaagg acaaccggga aaagattgag 1920

aagatcctga ccttcaggat cccctactac gtgggaccac tggccagagg caatagcaga 1980

ttcgcctgga tgaccagaaa gagcgaggaa accatcacac cctggaactt cgaggaagtg 2040

gtggataagg gcgccagcgc tcagtccttc atcgagcgga tgaccaactt cgataagaac 2100

ctgcctaacg agaaggtgct gcccaagcac agcctgctgt acgagtactt caccgtgtac 2160

aacgagctga ccaaagtgaa atacgtgacc gagggaatga gaaagcccgc ctttctgagc 2220

ggcgagcaga aaaaggccat tgtggatctg ctgttcaaga ccaaccggaa agtgaccgtg 2280

aagcagctga aagaggacta cttcaagaaa atcgagtgct tcgacagcgt ggaaatcagc 2340

ggcgtggaag atcggttcaa tgccagcctg ggcacatacc acgacctgct gaaaattatc 2400

aaggacaagg acttcctgga caacgaagag aacgaggaca tcctggaaga tatcgtgctg 2460

accctgacac tgtttgagga cagagagatg atcgaggaac ggctgaaaac atacgcccac 2520

ctgttcgacg acaaagtgat gaagcaactg aagcggcgga gatacaccgg ctggggcaga 2580

ctgtctcgga agctgatcaa cggcatccgg gataagcagt ccggcaagac catcctggac 2640

tttctgaagt ccgacggctt cgccaatcgg aacttcatgc agctgatcca cgacgacagc 2700

ctgaccttta aagaggatat ccagaaagcc caggtgtccg gccagggcga ttctctgcat 2760

gagcacattg ccaacctggc cggctctccc gccattaaga agggcattct gcagacagtg 2820
aaggtggtgg acgagctggt caaagtcatg ggcagacaca agcccgagaa catcgtgatc 2880
gaaatggcca gagagaacca gaccacacag aagggccaga agaacagccg cgagagaatg 2940
aagcggatcg aagagggcat caaagagctg ggcagccaga tcctgaaaga acaccccgtg 3000
gaaaacaccc agctgcagaa cgagaagctg tacctgtact acctccaaaa cggccgggat 3060
atgtatgtgg accaagagct ggacatcaac cggctgtccg actacgatgt ggacgctatc 3120
gtgccccagt cttttctgaa agacgactcc atcgacaaca aggtcctgac cagaagcgac 3180

aagaaccggg gcaagagcga taacgtgccc tccgaagagg tcgtgaagaa gatgaagaac 3240
tactggcgac agctgctgaa cgccaagctg attacccagc ggaagttcga taacctgacc 3300
aaggccgaga gaggcggcct gtctgaactg gataaggccg gcttcatcaa gagacagctg 3360
gtggaaaccc ggcagatcac caaacacgtg gcacagattc tggactcccg gatgaacact 3420
aagtacgacg agaatgacaa gctgatccgg gaagtgaaag tgatcaccct gaagtccaag 3480
ctggtgtccg atttccggaa ggatttccag ttctacaaag tgcgcgagat caacaactac 3540
catcacgccc acgacgccta cctgaatgcc gttgttggaa cagccctgat caagaagtat 3600

cccaagctgg aatccgagtt cgtgtacggc gactacaagg tgtacgacgt gcggaagatg 3660
atcgccaaga gcgagcaaga gattggcaag gctaccgcca agtacttctt ctacagcaac 3720
atcatgaact ttttcaagac cgagattacc ctggccaacg gcgagatcag aaagcggcct 3780
ctgatcgaga caaacggcga aaccggcgag attgtgtggg acaagggcag agattttgcc 3840
accgtgcgga aagtgctgag catgccccaa gtgaatatcg tgaaaaagac cgaggtaagt 3900
attagctctt tctttccatg ggttggcctc gccgcgtggg ctgagggaag gactgtcctg 3960
ggactggaca ggcgggttat gggacctgaa gcgataaaag gcatgcacgt ttgcggctac 4020

gtgcatgcca aaaggagtcg ggcttgcctc cgtgcccgac tccaaaagac ctgctcgagg 4080
aggtggacga gcaggtcaaa aatccgggta ccaataaaat atctttattt tcattacatc 4140
tgtgtgttgg ttttttgtgt g 4161

<210> 160
<211> 3410
<212> DNA
<213> Artificial sequence
<220><223> Synthetic nucleic acid sequence
<400> 160

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgaccttа tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttaggatttt tgacctgctc 540

gattgtccac tgcgagcagg tcttttggag tcgggcgagg cggaagcccg actccttttg 600

gcatgcacgc tagccgcgtc gtgcatgcct tttatcttcg ggttatggga ccagtgaagg 660

ctgagggaag gactgtcctg ggactggaca ggcgggttat gggacctgaa aatactaaca 720

atcgattttt tttccctttt tttccaggtg cagacaggcg gcttcagcaa agagtccatt 780

ctgcccaaga gaaacagcga taagctgatc gcccggaaga aggactggga ccctaagaag 840

tacggcggct tcgatagccc taccgtggcc tattctgtgc tggtggtggc caaagtggaa 900

aagggcaagt ccaagaaact caagagcgtg aaagagctgc tggggatcac catcatggaa 960

agaagcagct tcgagaagaa tcctatcgat ttcctcgagg ccaagggcta caaagaagtg 1020

aaaaaggacc tgatcatcaa gctccccaag tactccctgt tcgagctgga aaatggccgg 1080

aagcggatgc tggcttctgc tggcgaactg cagaagggaa acgaactggc cctgcctagc 1140

aaatatgtga acttcctgta cctggccagc cactatgaga agctgaaggg cagccccgag 1200

gacaatgagc aaaagcagct gtttgtggaa cagcacaagc actacctgga cgagatcatc 1260

gagcagatct ccgagttctc caagagagtg atcctggccg acgctaatct ggacaaagtg 1320

ctgtccgcct acaacaagca ccgggacaag cctatcagag agcaggccga gaatatcatc 1380

cacctgttta ccctgaccaa tctgggagcc cctgccgcct tcaagtactt tgacaccacc 1440

atcgaccgga agcgctacac cagcaccaaa gaggtgctgg acgccacact gatccaccag 1500

tctatcaccg gcctgtacga gacacggatc gacctgtctc agctcggagg cgatagcagg 1560

gctgacccca agaagaagag gaaggtgtcg ccagggatcc gtcgacttga cgcgttgata 1620

tcaacaagtt tgtacaaaaa agcaggctac aaagaggcca gcggttccgg acgggctgac 1680

gcattggacg attttgatct ggatatgctg ggaagtgacg ccctcgatga ttttgacctt 1740

gacatgcttg gttcggatgc ccttgatgac tttgacctcg acatgctcgg cagtgacgcc 1800

cttgatgatt tcgacctgga catgctgatt aactctagaa gttccggatc tccgaaaaag 1860

aaacgcaaag ttggtagcca gtacctgccc gacaccgacg accggcaccg gatcgaggaa 1920

aagcggaagc ggacctacga gacattcaag agcatcatga agaagtcccc cttcagcggc 1980

cccaccgacc ctagacctcc acctagaaga atcgccgtgc ccagcagatc cagcgccagc 2040

gtgccaaaac ctgcccccca gccttacccc ttcaccagca gcctgagcac catcaactac 2100

gacgagttcc ctaccatggt gttccccagc ggccagatct ctcaggcctc tgctctggct 2160

ccagcccctc ctcaggtgct gcctcaggct cctgctcctg caccagctcc agccatggtg 2220

tctgcactgg ctcaggcacc agcacccgtg cctgtgctgg ctcctggacc tccacaggct 2280

gtggctccac cagcccctaa acctacacag gccggcgagg gcacactgtc tgaagctctg 2340

ctgcagctgc agttcgacga cgaggatctg ggagccctgc tgggaaacag caccgatcct 2400

gccgtgttca ccgacctggc cagcgtggac aacagcgagt tccagcagct gctgaaccag 2460

ggcatccctg tggcccctca caccaccgag cccatgctga tggaataccc cgaggccatc 2520

acccggctcg tgacaggcgc tcagaggcct cctgatccag ctcctgcccc tctgggagca 2580

ccaggcctgc ctaatggact gctgtctggc gacgaggact tcagctctat cgccgatatg 2640

gatttctcag ccttgctggg ctctggcagc ggcagccggg attccaggga agggatgttt 2700

ttgccgaagc ctgaggccgg ctccgctatt agtgacgtgt ttgagggccg cgaggtgtgc 2760

cagccaaaac gaatccggcc atttcatcct ccaggaagtc catgggccaa ccgcccactc 2820

cccgccagcc tcgcaccaac accaaccggt ccagtacatg agccagtcgg gtcactgacc 2880

ccggcaccag tccctcagcc actggatcca gcgcccgcag tgactcccga ggccagtcac 2940

ctgttggagg atcccgatga agagacgagc caggctgtca aagcccttcg ggagatggcc 3000

gatactgtga ttccccagaa ggaagaggct gcaatctgtg gccaaatgga cctttcccat 3060

ccgcccccaa ggggccatct ggatgagctg acaaccacac ttgagtccat gaccgaggat 3120

ctgaacctgg actcaccccct gaccccggaa ttgaacgaga ttctggatac cttcctgaac 3180

gacgagtgcc tcttgcatgc catgcatatc agcacaggac tgtccatctt cgacacatct 3240

ctgttttgag gtaccaattc ctcacctgcg atctcgatgc tttatttgtg aaatttgtga 3300

tgctattgct ttatttgtaa ccattataag ctgcaataaa caagttaaca acaacaattg 3360

cattcatttt atgtttcagg ttcaggggga ggtgtgggag gtttttttaaa 3410

<210> 161

<211> 4161

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 161

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcc 540

accatgaaga gaaccgccga cggcagcgag ttcgagagcc ctaagaaaaa gcggaaggtg 600

gacaagaagt acagcatcgg cctggacatc ggcaccaatt ctgttggctg ggccgtgatc 660

accgacgagt acaaggtgcc cagcaagaaa ttcaaggtgc tgggcaacac cgaccggcac 720

agcatcaaga agaatctgat cggcgccctg ctgttcgact ctggcgaaac agccgaagcc 780

accagactga agaggacagc cagacggcgg tacaccagaa gaaagaaccg gatctgctac 840

ctgcaagaga tcttcagcaa cgagatggcc aaggtggacg acagcttctt ccaccggctg 900

gaagagtcct tcctggtgga agaggataag aagcacgagc ggcaccccat cttcggcaac 960

atcgtggatg aggtggccta ccacgagaag taccccacca tctaccacct gagaaagaaa 1020

ctggtggaca gcaccgacaa ggccgacctg agactgatct atctggccct ggctcacatg 1080

atcaagttcc ggggccactt cctgatcgag ggcgacctga atcctgacaa cagcgacgtg 1140

gacaagctgt tcatccagct ggtgcagacc tacaaccagc tgttcgagga aaaccccatc 1200

aacgccagcg gagtggatgc caaggccatc ctgtctgccc ggctgagcaa gagcagacgg 1260

ctggaaaacc tgatcgctca gctgcccggc gagaagaaga atggcctgtt cggcaacctg 1320

attgccctga gcctgggcct gacacctaac ttcaagagca acttcgacct ggccgaggac 1380

gccaaactgc agctgtccaa ggacacctac gacgacgacc tggacaatct gctggcccag 1440

atcggcgatc agtacgccga cttgtttctg gccgccaaga acctgtccga cgccatcctg 1500

ctgagcgaca tcctgagagt gaacaccgag atcacaaagg cccctctgag cgcctctatg 1560

atcaagagat acgacgagca ccaccaggat ctgaccctgc tgaaggccct cgttagacag 1620

cagctgcctg agaagtacaa agagattttc ttcgaccaga gcaagaacgg ctacgccggc 1680
tacattgatg gcggagccag ccaagaggaa ttctacaagt tcatcaagcc catcctcgag 1740
aagatggacg gcaccgagga actgctggtc aagctgaaca gagaggacct gctgcggaag 1800
cagcggacct tcgacaatgg ctctatccct caccaaatcc acctgggaga gctgcacgcc 1860
attctgcgga gacaagagga cttttaccca ttcctgaagg acaaccggga aaagattgag 1920
aagatcctga ccttcaggat cccctactac gtgggaccac tggccagagg caatagcaga 1980
ttcgcctgga tgaccagaaa gagcgaggaa accatcacac cctggaactt cgaggaagtg 2040

gtggataagg gcgccagcgc tcagtccttc atcgagcgga tgaccaactt cgataagaac 2100
ctgcctaacg agaaggtgct gcccaagcac agcctgctgt acgagtactt caccgtgtac 2160
aacgagctga ccaaagtgaa atacgtgacc gagggaatga gaaagcccgc ctttctgagc 2220
ggcgagcaga aaaaggccat tgtggatctg ctgttcaaga ccaaccggaa agtgaccgtg 2280
aagcagctga aagaggacta cttcaagaaa atcgagtgct tcgacagcgt ggaaatcagc 2340
ggcgtggaag atcggttcaa tgccagcctg ggcacatacc acgacctgct gaaaattatc 2400
aaggacaagg acttcctgga caacgaagag aacgaggaca tcctggaaga tatcgtgctg 2460

accctgacac tgtttgagga cagagagatg atcgaggaac ggctgaaaac atacgcccac 2520
ctgttcgacg acaaagtgat gaagcaactg aagcggcgga gatacaccgg ctggggcaga 2580
ctgtctcgga agctgatcaa cggcatccgg gataagcagt ccggcaagac catcctggac 2640
tttctgaagt ccgacggctt cgccaatcgg aacttcatgc agctgatcca cgacgacagc 2700
ctgaccttta aagaggatat ccagaaagcc caggtgtccg gccagggcga ttctctgcat 2760
gagcacattg ccaacctggc cggctctccc gccattaaga agggcattct gcagacagtg 2820
aaggtggtgg acgagctggt caaagtcatg ggcagacaca agcccgagaa catcgtgatc 2880

gaaatggcca gagagaacca gaccacacag aagggccaga agaacagccg cgagagaatg 2940
aagcggatcg aagagggcat caaagagctg ggcagccaga tcctgaaaga acaccccgtg 3000
gaaaacaccc agctgcagaa cgagaagctg tacctgtact acctccaaaa cggccgggat 3060
atgtatgtgg accaagagct ggacatcaac cggctgtccg actacgatgt ggacgctatc 3120
gtgccccagt cttttctgaa agacgactcc atcgacaaca aggtcctgac cagaagcgac 3180
aagaaccggg gcaagagcga taacgtgccc tccgaagagg tcgtgaagaa gatgaagaac 3240
tactggcgac agctgctgaa cgccaagctg attacccagc ggaagttcga taacctgacc 3300

aaggccgaga gaggcggcct gtctgaactg gataaggccg gcttcatcaa gagacagctg 3360

gtggaaaccc ggcagatcac caaacacgtg gcacagattc tggactcccg gatgaacact 3420

aagtacgacg agaatgacaa gctgatccgg gaagtgaaag tgatcaccct gaagtccaag 3480

ctggtgtccg atttccggaa ggatttccag ttctacaaag tgcgcgagat caacaactac 3540

catcacgccc acgacgccta cctgaatgcc gttgttggaa cagccctgat caagaagtat 3600

cccaagctgg aatccgagtt cgtgtacggc gactacaagg tgtacgacgt gcggaagatg 3660

atcgccaaga gcgagcaaga gattggcaag gctaccgcca agtacttctt ctacagcaac 3720

atcatgaact ttttcaagac cgagattacc ctggccaacg gcgagatcag aaagcggcct 3780

ctgatcgaga caaacggcga aaccggcgag attgtgtggg acaagggcag agattttgcc 3840

accgtgcgga aagtgctgag catgccccaa gtgaatatcg tgaaaaagac cgaggtaagt 3900

attagctctt tctttccatg ggttggcctc gccgcgtggg ctgagggaag gactgtcctg 3960

ggactggaca ggcgggttat gggacctgaa gcgataaaag gcatgcacgt ttgcggctac 4020

gtgcatgcca aaaggagtcg ggcttgcctc cgtgcccgac tccaaaagac ctgctcgagg 4080

aggtggacga gcaggtcaaa aatccgggta ccaataaaat atctttattt tcattacatc 4140

tgtgtgttgg ttttttgtgt g 4161

<210> 162

<211> 3911

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 162

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttaggatttt tgacctgctc 540

gattgtccac tgcgagcagg tcttttggag tcgggcgagg cggaagcccg actccttttg 600

gcatgcacgc tagccgcgtc gtgcatgcct tttatcttcg ggttatggga ccagtgaagg 660

ctgagggaag gactgtcctg ggactggaca ggcgggttat gggacctgaa aatactaaca 720

atcgattttt tttccctttt tttccaggtg cagacaggcg gcttcagcaa agagtccatt 780

ctgcccaaga gaaacagcga taagctgatc gcccggaaga aggactggga ccctaagaag 840

tacggcggct tcgatagccc taccgtggcc tattctgtgc tggtggtggc caaagtggaa 900

aagggcaagt ccaagaaact caagagcgtg aaagagctgc tggggatcac catcatggaa 960

agaagcagct tcgagaagaa tcctatcgat ttcctcgagg ccaagggcta caaagaagtg 1020

aaaaaggacc tgatcatcaa gctccccaag tactccctgt tcgagctgga aaatggccgg 1080

aagcggatgc tggcttctgc tggcgaactg cagaagggaa acgaactggc cctgcctagc 1140

aaatatgtga acttcctgta cctggccagc cactatgaga agctgaaggg cagccccgag 1200

gacaatgagc aaaagcagct gtttgtggaa cagcacaagc actacctgga cgagatcatc 1260

gagcagatct ccgagttctc caagagagtg atcctggccg acgctaatct ggacaaagtg 1320

ctgtccgcct acaacaagca ccgggacaag cctatcagag agcaggccga gaatatcatc 1380

cacctgttta ccctgaccaa tctgggagcc cctgccgcct tcaagtactt tgacaccacc 1440

atcgaccgga agcgctacac cagcaccaaa gaggtgctgg acgccacact gatccaccag 1500

tctatcaccg gcctgtacga gacacggatc gacctgtctc agctcggagg cgattctggc 1560

ggatctagcg gtggaagctc tggctctgag acacctggca caagcgagtc tgccacacct 1620

gagtctagcg gcggatcttc aggcggcagc agcaccctga atatcgagga tgagtacaga 1680

ctgcacgaga caagcaaaga acccgacgtg tccctgggct ctacctggct gtctgatttt 1740

cctcaagcct gggccgaaac aggcggaatg ggacttgctg ttagacaggc tcccctgatc 1800

attcccctga aggccacaag cacccctgtg tccatcaagc agtaccccat gtctcaagag 1860

gcccggctgg gaatcaagcc ccacattcag agactgctgg accagggcat cctggtgcct 1920

tgtcaaagcc cttggaatac ccctctgctg cctgtgaaga agcccggcac caacgactac 1980

agacccgtgc aggatctgcg cgaagtgaac aagagagtcg aggacattca ccccaccgtg 2040

cctaatcctt acaacctgct gtctggcctg cctccttccc accaatggta cacagtgctg 2100

gacctgaagg atgccttctt ctgcctgcgg ctgcacccta caagccagcc tctgtttgcc 2160

ttcgagtggc gggatccaga gatgggcatt agcggacagc tgacctggac cagactgccc 2220

cagggcttca agaatagccc cacactgttc aacgaggccc tgcacaggga cctcgccgac 2280

tttagaattc agcaccccga cctgattctg ctgcagtatg tggatgatct gctgctggcc 2340

gctaccagcg agctggattg tcagcaggga acaagagccc tgctgcagac cctgggcaat 2400

ctgggctata gagcctctgc caagaaggcc cagatttgcc agaagcaagt taagtacctg 2460

ggctacctgc tcaaagaagg ccagcgttgg ctgaccgagg ccagaaaaga aaccgtgatg 2520

ggccagccta cacctaagac acccagacag ctgagagagt tcctgggcaa agccggattc 2580

tgcaggctgt ttatccctgg cttcgccgag atggctgccc ctctgtatcc tctgacaaag 2640

cccggaactc tgttcaactg gggcccagac cagcagaaag cctaccaaga gatcaagcag 2700

gctctgctga cagcccctgc tctgggactg cctgatctga ccaagccttt cgagctgttc 2760

gtggacgaga agcagggcta tgccaagggc gtgctgacac agaaactcgg cccttggaga 2820

aggcccgtgg cttacctgag caaaaagctg gatcctgtgg ccgctggctg gcctccttgt 2880

ctgagaatgg tggccgctat cgccgtgctg actaaggatg ccggcaagct gacaatggga 2940

cagcctctgg ttattctggc ccctcatgcc gtggaagccc tcgtgaaaca gcctcctgat 3000

cggtggctga gcaacgccag aatgacccac taccaggcac tgctgctcga caccgacaga 3060

gtgcaatttg gccctgtggt ggccctgaat ccagccacat tgctgcctct gcctgaggag 3120

ggactgcagc acaactgcct cgatatcctg gctgaggccc acggcacaag acccgatctg 3180

acagatcagc cactgcctga cgccgaccac acctggtata cagatggcag ctctctgctg 3240

caagagggcc agagaaaagc tggcgccgct gtgaccacag agacagaagt gatttgggcc 3300

aaagctctgc ctgccggcac atctgcccaa agagccgaac tgatcgcact gacacaggcc 3360

ctgaagatgg ccgagggcaa gaaactgaac gtgtacaccg actccagata cgccttcgcc 3420

accgctcaca tccacggcga aatctacaga cgcagaggat ggctgaccag cgagggaaaa 3480

gagattaaga acaaggacga gattctcgcc ctcctcaagg ccctgttcct gcctaagcgg 3540

ctgagcatca tccactgtcc tggccaccag aagggacact ctgccgaggc tagaggcaac 3600

agaatggccg atcaggctgc cagaaaggcc gccattaccg agacacccga taccagcaca 3660

ctgctgattg agaacagcag cccttccggc ggctccaaaa gaacagctga cggctccgag 3720

tttgagccca aaaagaaacg gaaagtgtga ggtaccaatt cctcacctgc gatctcgatg 3780

ctttatttgt gaaatttgtg atgctattgc tttatttgta accattataa gctgcaataa 3840

acaagttaac aacaacaatt gcattcattt tatgtttcag gttcaggggg aggtgtggga 3900

ggtttttaa a 3911

<210> 163

<211> 3159

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 163

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccacc 540

atgaaacgga cagccgacgg aagcgagttc gagtcaccaa agaagaagcg gaaagtcagc 600

agtgaaaccg gaccagtggc agtggaccca accctgagga gacggattga gccccatgaa 660

tttgaagtgt tctttgaccc aagggagctg aggaaggaga catgcctgct gtacgagatc 720

aagtggggca caagccacaa gatctggcgc cacagctcca agaacaccac aaagcacgtg 780

gaagtgaatt tcatcgagaa gtttacctcc gagcggcact tctgcccctc taccagctgt 840

tccatcacat ggtttctgtc ttggagccct tgcggcgagt gttccaaggc catcaccgag 900

ttcctgtctc agcaccctaa cgtgaccctg gtcatctacg tggcccggct gtatcaccac 960

atggaccagc agaacaggca gggcctgcgc gatctggtga attctggcgt gaccatccag 1020

atcatgacag ccccagagta cgactattgc tggcggaact tcgtgaatta tccacctggc 1080

aaggaggcac actggccaag atacccaccc ctgtggatga agctgtatgc actggagctg 1140

cacgcaggaa tcctgggcct gcctccatgt ctgaatatcc tgcggagaaa gcagccccag 1200

ctgacatttt tcaccattgc tctgcaatct tgtcactatc agcggctgcc tcctcatatt 1260

ctgtgggcta ccggcctgaa gtctggagga tctagcggag gatcctctgg cagcgagaca 1320

ccaggaacaa gcgagtcagc aacaccagag agcagtggcg gcagcagcgg cggcagcgac 1380

aagaagtaca gcatcggcct ggccatcggc accaattctg ttggctgggc cgtgatcacc 1440

gacgagtaca aggtgcccag caagaaattc aaggtgctgg gcaacaccga ccggcacagc 1500

atcaagaaga atctgatcgg cgccctgctg ttcgactctg gcgaaacagc cgaagccacc 1560

agactgaaga ggacagccag acggcggtac accagaagaa agaaccggat ctgctacctg 1620

caagagatct tcagcaacga gatggccaag gtggacgaca gcttcttcca ccggctggaa 1680

gagtccttcc tggtggaaga ggataagaag cacgagcggc accccatctt cggcaacatc 1740

gtggatgagg tggcctacca cgagaagtac cccaccatct accacctgag aaagaaactg 1800

gtggacagca ccgacaaggc cgacctgaga ctgatctatc tggccctggc tcacatgatc 1860

aagttccggg gccacttcct gatcgagggc gacctgaatc ctgacaacag cgacgtggac 1920

aagctgttca tccagctggt gcagacctac aaccagctgt tcgaggaaaa ccccatcaac 1980

gccagcggag tggatgccaa ggccatcctg tctgcccggc tgagcaagag cagacggctg 2040

gaaaacctga tcgctcagct gcccggcgag aagaagaatg gcctgttcgg caacctgatt 2100

gccctgagcc tgggcctgac acctaacttc aagagcaact tcgacctggc cgaggacgcc 2160

aaactgcagc tgtccaagga cacctacgac gacgacctgg acaatctgct ggcccagatc 2220

ggcgatcagt acgccgactt gtttctggcc gccaagaacc tgtccgacgc catcctgctg 2280

agcgacatcc tgagagtgaa caccgagatc acaaaggccc ctctgagcgc ctctatgatc 2340

aagagatacg acgagcacca ccaggatctg accctgctga aggccctcgt tagacagcag 2400

ctgcctgaga agtacaaaga gattttcttc gaccagagca agaacggcta cgccggctac 2460

attgatggcg gagccagcca agaggaattc tacaagttca tcaagcccat cctcgagaag 2520

atggacggca ccgaggaact gctggtcaag ctgaacagag aggacctgct gcggaagcag 2580

cggaccttcg acaatggctc tatccctcac caaatccacc tgggagagct gcacgccatt 2640

ctgcggagac aagaggactt ttacccattc ctgaaggaca accgggaaaa gattgagaag 2700

atcctgacct tcaggatccc ctactacgtg ggaccactgg ccagaggcaa tagcagattc 2760

gcctggatga ccagaaagag cgaggaaacc atcacaccct ggaacttcga ggaagtggtg 2820

gataagggcg ccagcgctca gtccttcatc gagcggatga ccaacttcga taagaacctg 2880

cctaacgaga aggtaagtat tagctctttc tttccatggg ttggcctcgc cgcgtgggct 2940

gagggaagga ctgtcctggg actggacagg cgggttatgg gacctgaagc gataaaaggc 3000

atgcacgttt gcggctacgt gcatgccaaa aggagtcggg cttgcctccg tgcccgactc 3060

caaaagacct gctcgaggag gtggacgagc aggtcaaaaa tccgggtacc aataaaatat 3120

ctttattttc attacatctg tgtgttggtt ttttgtgtg 3159

<210> 164

<211> 4115

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 164

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttaggatttt tgacctgctc 540

gattgtccac tgcgagcagg tcttttggag tcgggcgagg cggaagcccg actccttttg 600

gcatgcacgc tagccgcgtc gtgcatgcct tttatcttcg ggttatggga ccagtgaagg 660

ctgagggaag gactgtcctg ggactggaca ggcgggttat gggacctgaa aatactaaca 720

atcgattttt tttccctttt tttccaggtg ctgcccaagc acagcctgct gtacgagtac 780

ttcaccgtgt acaacgagct gaccaaagtg aaatacgtga ccgagggaat gagaaagccc 840

gcctttctga gcggcgagca gaaaaaggcc attgtggatc tgctgttcaa gaccaaccgg 900

aaagtgaccg tgaagcagct gaaagaggac tacttcaaga aaatcgagtg cttcgacagc 960

gtggaaatca gcggcgtgga agatcggttc aatgccagcc tgggcacata ccacgacctg 1020

ctgaaaatta tcaaggacaa ggacttcctg gacaacgaag agaacgagga catcctggaa 1080

gatatcgtgc tgaccctgac actgtttgag gacagagaga tgatcgagga acggctgaaa 1140

acatacgccc acctgttcga cgacaaagtg atgaagcaac tgaagcggcg gagatacacc 1200

ggctggggca gactgtctcg gaagctgatc aacggcatcc gggataagca gtccggcaag 1260

accatcctgg actttctgaa gtccgacggc ttcgccaatc ggaacttcat gcagctgatc 1320

cacgacgaca gcctgacctt taaagaggat atccagaaag cccaggtgtc cggccagggc 1380

gattctctgc atgagcacat tgccaacctg gccggctctc ccgccattaa gaagggcatt 1440

ctgcagacag tgaaggtggt ggacgagctg gtcaaagtca tgggcagaca caagcccgag 1500

aacatcgtga tcgaaatggc cagagagaac cagaccacac agaagggcca gaagaacagc 1560

cgcgagagaa tgaagcggat cgaagagggc atcaaagagc tgggcagcca gatcctgaaa 1620

gaacaccccg tggaaaacac ccagctgcag aacgagaagc tgtacctgta ctacctccaa 1680

aacggccggg atatgtatgt ggaccaagag ctggacatca accggctgtc cgactacgat 1740

gtggaccata tcgtgcccca gtcttttctg aaagacgact ccatcgacaa caaggtcctg 1800

accagaagcg acaagaaccg gggcaagagc gataacgtgc cctccgaaga ggtcgtgaag 1860

aagatgaaga actactggcg acagctgctg aacgccaagc tgattaccca gcggaagttc 1920

gataacctga ccaaggccga gagaggcggc ctgtctgaac tggataaggc cggcttcatc 1980

aagagacagc tggtggaaac ccggcagatc accaaacacg tggcacagat tctggactcc 2040

cggatgaaca ctaagtacga cgagaatgac aagctgatcc gggaagtgaa agtgatcacc 2100

ctgaagtcca agctggtgtc cgatttccgg aaggatttcc agttctacaa agtgcgcgag 2160

atcaacaact accatcacgc ccacgacgcc tacctgaatg ccgttgttgg aacagccctg 2220

atcaagaagt atcccaagct ggaatccgag ttcgtgtacg gcgactacaa ggtgtacgac 2280

gtgcggaaga tgatcgccaa gagcgagcaa gagattggca aggctaccgc caagtacttc 2340

ttctacagca acatcatgaa ctttttcaag accgagatta ccctggccaa cggcgagatc 2400

agaaagcggc ctctgatcga gacaaacggc gaaaccggcg agattgtgtg ggacaagggc 2460

agagattttg ccaccgtgcg gaaagtgctg agcatgcccc aagtgaatat cgtgaaaaag 2520

accgaggtgc agacaggcgg cttcagcaaa gagtccattc tgcccaagag aaacagcgat 2580

aagctgatcg cccggaagaa ggactgggac cctaagaagt acggcggctt cgatagccct 2640

accgtggcct attctgtgct ggtggtggcc aaagtggaaa agggcaagtc caagaaactc 2700

aagagcgtga aagagctgct ggggatcacc atcatggaaa gaagcagctt cgagaagaat 2760

cctatcgatt tcctcgaggc caagggctac aaagaagtga aaaaggacct gatcatcaag 2820

ctccccaagt actccctgtt cgagctggaa aatggccgga agcggatgct ggcttctgct 2880

ggcgaactgc agaagggaaa cgaactggcc ctgcctagca aatatgtgaa cttcctgtac 2940

ctggccagcc actatgagaa gctgaagggc agccccgagg acaatgagca aaagcagctg 3000

tttgtggaac agcacaagca ctacctggac gagatcatcg agcagatctc cgagttctcc 3060

aagagagtga tcctggccga cgctaatctg gacaaagtgc tgtccgccta caacaagcac 3120

cgggacaagc ctatcagaga gcaggccgag aatatcatcc acctgtttac cctgaccaat 3180

ctgggagccc ctgccgcctt caagtacttt gacaccacca tcgaccggaa gcgctacacc 3240

agcaccaaag aggtgctgga cgccacactg atccaccagt ctatcaccgg cctgtacgag 3300

acacggatcg acctgtctca gctcggaggc gatagcggcg ggagcggcgg gagcgggggg 3360

agcactaatc tgagcgacat cattgagaag gagactggga aacagctggt cattcaggag 3420

tccatcctga tgctgcctga ggaggtggag gaagtgatcg gcaacaagcc agagtctgac 3480

atcctggtgc acaccgccta cgacgagtcc acagatgaga atgtgatgct gctgacctct 3540

gacgcccccg agtataagcc ttgggccctg gtcatccagg attctaacgg cgagaataag 3600

atcaagatgc tgagcggagg atccggagga tctggaggca gcaccaacct gtctgacatc 3660

atcgagaagg agacaggcaa gcagctggtc atccaggaga gcatcctgat gctgcccgaa 3720

gaagtcgaag aagtgatcgg aaacaagcct gagagcgata tcctggtcca taccgcctac 3780

gacgagagta ccgacgaaaa tgtgatgctg ctgacatccg acgccccaga gtataagccc 3840

tgggctctgg tcatccagga ttccaacgga gagaacaaaa tcaaaatgct gtctggcggc 3900

tcaaaaagaa ccgccgacgg cagcgaattc gagcccaaga agaagaggaa agtctaaacc 3960

aattcctcac ctgcgatctc gatgctttat ttgtgaaatt tgtgatgcta ttgctttatt 4020

tgtaaccatt ataagctgca ataaacaagt taacaacaac aattgcattc attttatgtt 4080

tcaggttcag ggggaggtgt gggaggtttt ttaaa 4115

<210> 165

<211> 2973

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 165

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccacc 540

atgaaacgga cagccgacgg aagcgagttc gagtcaccaa agaagaagcg gaaagtctct 600

gaggtggagt tttcccacga gtactggatg agacatgccc tgaccctggc caagagggca 660

cgggatgaga gggaggtgcc tgtgggagcc gtgctggtgc tgaacaatag agtgatcggc 720

gagggctgga acagagccat cggcctgcac gacccaacag cccatgccga aattatggcc 780

ctgagacagg gcggcctggt catgcagaac tacagactga ttgacgccac cctgtacgtg 840

acattcgagc cttgcgtgat gtgcgccggc gccatgatcc actctaggat cggccgcgtg 900

gtgtttggcg tgaggaactc aaaaagaggc gccgcaggct ccctgatgaa cgtgctgaac 960

taccccggca tgaatcaccg cgtcgaaatt accgagggaa tcctggcaga tgaatgtgcc 1020

gccctgctgt gcgatttcta tcggatgcct agacaggtgt tcaatgctca gaagaaggcc 1080

cagagctcca tcaactccgg aggatctagc ggaggctcct ctggctctga gacacctggc 1140

acaagcgaga gcgcaacacc tgaaagcagc gggggcagca gcgggggggtc agacaagaag 1200

tacagcatcg gcctggccat cggcaccaat tctgttggct gggccgtgat caccgacgag 1260

tacaaggtgc ccagcaagaa attcaaggtg ctgggcaaca ccgaccggca cagcatcaag 1320

aagaatctga tcggcgccct gctgttcgac tctggcgaaa cagccgaagc caccagactg 1380

aagaggacag ccagacggcg gtacaccaga agaaagaacc ggatctgcta cctgcaagag 1440

atcttcagca acgagatggc caaggtggac gacagcttct tccaccggct ggaagagtcc 1500

ttcctggtgg aagaggataa gaagcacgag cggcacccca tcttcggcaa catcgtggat 1560

gaggtggcct accacgagaa gtaccccacc atctaccacc tgagaaagaa actggtggac 1620

agcaccgaca aggccgacct gagactgatc tatctggccc tggctcacat gatcaagttc 1680

cggggccact tcctgatcga gggcgacctg aatcctgaca acagcgacgt ggacaagctg 1740

ttcatccagc tggtgcagac ctacaaccag ctgttcgagg aaaaccccat caacgccagc 1800

ggagtggatg ccaaggccat cctgtctgcc cggctgagca agagcagacg gctggaaaac 1860

ctgatcgctc agctgcccgg cgagaagaag aatggcctgt tcggcaacct gattgccctg 1920

agcctgggcc tgacacctaa cttcaagagc aacttcgacc tggccgagga cgccaaactg 1980

cagctgtcca aggacaccta cgacgacgac ctggacaatc tgctggccca gatcggcgat 2040

cagtacgccg acttgtttct ggccgccaag aacctgtccg acgccatcct gctgagcgac 2100

atcctgagag tgaacaccga gatcacaaag gcccctctga gcgcctctat gatcaagaga 2160

tacgacgagc accaccagga tctgaccctg ctgaaggccc tcgttagaca gcagctgcct 2220

gagaagtaca aagagatttt cttcgaccag agcaagaacg gctacgccgg ctacattgat 2280

ggcggagcca gccaagagga attctacaag ttcatcaagc ccatcctcga gaagatggac 2340

ggcaccgagg aactgctggt caagctgaac agagaggacc tgctgcggaa gcagcggacc 2400

ttcgacaatg gctctatccc tcaccaaatc cacctgggag agctgcacgc cattctgcgg 2460

agacaagagg acttttaccc attcctgaag gacaaccggg aaaagattga gaagatcctg 2520

accttcagga tcccctacta cgtgggacca ctggccagag gcaatagcag attcgcctgg 2580

atgaccagaa agagcgagga aaccatcaca ccctggaact tcgaggaagt ggtggataag 2640

ggcgccagcg ctcagtcctt catcgagcgg atgaccaact tcgataagaa cctgcctaac 2700

gagaaggtaa gtattagctc tttctttcca tgggttggcc tcgccgcgtg ggctgaggga 2760

aggactgtcc tgggactgga caggcgggtt atgggacctg aagcgataaa aggcatgcac 2820

gtttgcggct acgtgcatgc caaaaggagt cgggcttgcc tccgtgcccg actccaaaag 2880

acctgctcga ggaggtggac gagcaggtca aaaatccggg taccaataaa atatctttat 2940

tttcattaca tctgtgtgtt ggtttttgt gtg 2973

<210> 166

<211> 3560

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 166

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttaggatttt tgacctgctc 540

gattgtccac tgcgagcagg tcttttggag tcgggcgagg cggaagcccg actccttttg 600

gcatgcacgc tagccgcgtc gtgcatgcct tttatcttcg ggttatggga ccagtgaagg 660

ctgagggaag gactgtcctg ggactggaca ggcgggttat gggacctgaa aatactaaca 720

atcgattttt tttccctttt tttccaggtg ctgcccaagc acagcctgct gtacgagtac 780

ttcaccgtgt acaacgagct gaccaaagtg aaatacgtga ccgagggaat gagaaagccc 840

gcctttctga gcggcgagca gaaaaaggcc attgtggatc tgctgttcaa gaccaaccgg 900

aaagtgaccg tgaagcagct gaaagaggac tacttcaaga aaatcgagtg cttcgacagc 960

gtggaaatca gcggcgtgga agatcggttc aatgccagcc tgggcacata ccacgacctg 1020

ctgaaaatta tcaaggacaa ggacttcctg gacaacgaag agaacgagga catcctggaa 1080

gatatcgtgc tgaccctgac actgtttgag gacagagaga tgatcgagga acggctgaaa 1140

acatacgccc acctgttcga cgacaaagtg atgaagcaac tgaagcggcg gagatacacc 1200

ggctggggca gactgtctcg gaagctgatc aacggcatcc gggataagca gtccggcaag 1260

accatcctgg actttctgaa gtccgacggc ttcgccaatc ggaacttcat gcagctgatc 1320

cacgacgaca gcctgacctt taaagaggat atccagaaag cccaggtgtc cggccagggc 1380

gattctctgc atgagcacat tgccaacctg gccggctctc ccgccattaa gaagggcatt 1440

ctgcagacag tgaaggtggt ggacgagctg gtcaaagtca tgggcagaca caagcccgag 1500

aacatcgtga tcgaaatggc cagagagaac cagaccacac agaagggcca gaagaacagc 1560

cgcgagagaa tgaagcggat cgaagagggc atcaaagagc tgggcagcca gatcctgaaa 1620

gaacaccccg tggaaaacac ccagctgcag aacgagaagc tgtacctgta ctacctccaa 1680

aacggccggg atatgtatgt ggaccaagag ctggacatca accggctgtc cgactacgat 1740

gtggaccata tcgtgcccca gtcttttctg aaagacgact ccatcgacaa caaggtcctg 1800

accagaagcg acaagaaccg gggcaagagc gataacgtgc cctccgaaga ggtcgtgaag 1860

aagatgaaga actactggcg acagctgctg aacgccaagc tgattaccca gcggaagttc 1920

gataacctga ccaaggccga gagaggcggc ctgtctgaac tggataaggc cggcttcatc 1980

aagagacagc tggtggaaac ccggcagatc accaaacacg tggcacagat tctggactcc 2040

cggatgaaca ctaagtacga cgagaatgac aagctgatcc gggaagtgaa agtgatcacc 2100

ctgaagtcca agctggtgtc cgatttccgg aaggatttcc agttctacaa agtgcgcgag 2160

atcaacaact accatcacgc ccacgacgcc tacctgaatg ccgttgttgg aacagccctg 2220

atcaagaagt atcccaagct ggaatccgag ttcgtgtacg gcgactacaa ggtgtacgac 2280

gtgcggaaga tgatcgccaa gagcgagcaa gagattggca aggctaccgc caagtacttc 2340

ttctacagca acatcatgaa ctttttcaag accgagatta ccctggccaa cggcgagatc 2400

agaaagcggc ctctgatcga gacaaacggc gaaaccggcg agattgtgtg ggacaagggc 2460

agagattttg ccaccgtgcg gaaagtgctg agcatgcccc aagtgaatat cgtgaaaaag 2520

accgaggtgc agacaggcgg cttcagcaaa gagtccattc tgcccaagag aaacagcgat 2580

aagctgatcg cccggaagaa ggactgggac cctaagaagt acggcggctt cgatagccct 2640

accgtggcct attctgtgct ggtggtggcc aaagtggaaa agggcaagtc caagaaactc 2700

aagagcgtga aagagctgct ggggatcacc atcatggaaa gaagcagctt cgagaagaat 2760

cctatcgatt tcctcgaggc caagggctac aaagaagtga aaaaggacct gatcatcaag 2820

ctccccaagt actccctgtt cgagctggaa aatggccgga agcggatgct ggcttctgct 2880

ggcgaactgc agaagggaaa cgaactggcc ctgcctagca aatatgtgaa cttcctgtac 2940

ctggccagcc actatgagaa gctgaagggc agccccgagg acaatgagca aaagcagctg 3000

tttgtggaac agcacaagca ctacctggac gagatcatcg agcagatctc cgagttctcc 3060

aagagagtga tcctggccga cgctaatctg gacaaagtgc tgtccgccta caacaagcac 3120

cgggacaagc ctatcagaga gcaggccgag aatatcatcc acctgtttac cctgaccaat 3180

ctgggagccc ctgccgcctt caagtacttt gacaccacca tcgaccggaa gcgctacacc 3240

agcaccaaag aggtgctgga cgccacactg atccaccagt ctatcaccgg cctgtacgag 3300

acacggatcg acctgtctca gctcggaggc gattctggcg gctcaaaaag aaccgccgac 3360

ggcagcgaat tcgagcccaa gaagaagagg aaagtctaag gtaccaattc ctcacctgcg 3420

atctcgatgc tttatttgtg aaatttgtga tgctattgct ttatttgtaa ccattataag 3480

ctgcaataaa caagttaaca acaacaattg cattcatttt atgtttcagg ttcaggggga 3540

ggtgtgggag gtttttttaaa 3560

<210> 167

<211> 112

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 167

gatttttgac ctgctcgatt gtccactgcg agcaggtctt ttggagtcgg gcgaggcgga 60

agcccgactc cttttggcat gcacgctagc cgcgtcgtgc atgcctttta tc 112

<210> 168

<211> 13

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 168

gggttatggg acc 13

<210> 169

<211> 24

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 169

ggctgaggga aggactgtcc tggg 24

<210> 170

<211> 24

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 170

ctctttcttt ccatgggttg gcct 24

<210> 171

<211> 4463

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<220><221> misc_feature

<222> (4225)..(4294)

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

<400> 171

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttacttctag gcgcgccgcc 540

accatggccc caaagaagaa gcggaaggtc ggtatccacg gagtcccagc agccaagcgg 600

aactacatcc tgggcctgga catcggcatc accagcgtgg gctacggcat catcgactac 660

gagacacggg acgtgatcga tgccggcgtg cggctgttca aagaggccaa cgtggaaaac 720

aacgagggca ggcggagcaa gagaggcgcc agaaggctga agcggcggag gcggcataga 780

atccagagag tgaagaagct gctgttcgac tacaacctgc tgaccgacca cagcgagctg 840

agcggcatca acccctacga ggccagagtg aagggcctga gccagaagct gagcgaggaa 900

gagttctctg ccgccctgct gcacctggcc aagagaagag gcgtgcacaa cgtgaacgag 960

gtggaagagg acaccggcaa cgagctgtcc accaaagagc agatcagccg gaacagcaag 1020

gccctggaag agaaatacgt ggccgaactg cagctggaac ggctgaagaa agacggcgaa 1080

gtgcggggca gcatcaacag attcaagacc agcgactacg tgaaagaagc caaacagctg 1140

ctgaaggtgc agaaggccta ccaccagctg gaccagagct tcatcgacac ctacatcgac 1200

ctgctggaaa cccggcggac ctactatgag ggacctggcg agggcagccc cttcggctgg 1260

aaggacatca aagaatggta cgagatgctg atgggccact gcacctactt ccccgaggaa 1320

ctgcggagcg tgaagtacgc ctacaacgcc gacctgtaca acgccctgaa cgacctgaac 1380

aatctcgtga tcaccaggga cgagaacgag aagctggaat attacgagaa gttccagatc 1440

atcgagaacg tgttcaagca gaagaagaag cccaccctga agcagatcgc caaagaaatc 1500

ctcgtgaacg aagaggatat taagggctac agagtgacca gcaccggcaa gcccgagttc 1560

accaacctga aggtgtacca cgacatcaag gacattaccg cccggaaaga gattattgag 1620

aacgccgagc tgctggatca gattgccaag atcctgacca tctaccagag cagcgaggac 1680

atccaggaag aactgaccaa tctgaactcc gagctgaccc aggaagagat cgagcagatc 1740

tctaatctga agggctatac cggcacccac aacctgagcc tgaaggccat caacctgatc 1800

ctggacgagc tgtggcacac caacgacaac cagatcgcta tcttcaaccg gctgaagctg 1860

gtgcccaaga aggtggacct gtcccagcag aaagagatcc ccaccaccct ggtggacgac 1920

ttcatcctga gccccgtcgt gaagagaagc ttcatccaga gcatcaaagt gatcaacgcc 1980

atcatcaaga agtacggcct gcccaacgac atcattatcg agctggcccg cgagaagaac 2040

tccaaggacg cccagaaaat gatcaacgag atgcagaagc ggaaccggca gaccaacgag 2100

cggatcgagg aaatcatccg gaccaccggc aaagagaacg ccaagtacct gatcgagaag 2160

atcaagctgc acgacatgca ggaaggcaag tgcctgtaca gcctggaagc catccctctg 2220

gaagatctgc tgaacaaccc cttcaactat gaggtggacc acatcatccc cagaagcgtg 2280

tccttcgaca acagcttcaa caacaaggtg ctcgtgaagc aggaagaaaa cagcaagaag 2340

ggcaaccgga ccccattcca gtacctgagc agcagcgaca gcaagatcag ctacgaaacc 2400

ttcaagaagc acatcctgaa tctggccaag ggcaagggca gaatcagcaa gaccaagaaa 2460

gagtatctgc tggaagaacg ggacatcaac aggttctccg tgcagaaaga cttcatcaac 2520

cggaacctgg tggataccag atacgccacc agaggcctga tgaacctgct gcggagctac 2580

ttcagagtga acaacctgga cgtgaaagtg aagtccatca atggcggctt caccagcttt 2640

ctgcggcgga agtggaagtt taagaaagag cggaacaagg ggtacaagca ccacgccgag 2700

gacgccctga tcattgccaa cgccgatttc atcttcaaag agtggaagaa actggacaag 2760

gccaaaaaag tgatggaaaa ccagatgttc gaggaaaagc aggccgagag catgcccgag 2820

atcgaaaccg agcaggagta caaagagatc ttcatcaccc cccaccagat caagcacatt 2880

aaggacttca aggactacaa gtacagccac cgggtggaca agaagcctaa tagagagctg 2940
attaacgaca ccctgtactc cacccggaag gacgacaagg gcaacaccct gatcgtgaac 3000
aatctgaacg gcctgtacga caaggacaat gacaagctga aaaagctgat caacaagagc 3060
cccgaaaagc tgctgatgta ccaccacgac ccccagacct accagaaact gaagctgatt 3120
atggaacagt acggcgacga gaagaatccc ctgtacaagt actacgagga aaccgggaac 3180
tacctgacca agtactccaa aaaggacaac ggccccgtga tcaagaagat taagtattac 3240
ggcaacaaac tgaacgccca tctggacatc accgacgact accccaacag cagaaacaag 3300

gtcgtgaagc tgtccctgaa gccctacaga ttcgacgtgt acctggacaa tggcgtgtac 3360
aagttcgtga ccgtgaagaa tctggatgtg atcaaaaaag aaaactacta cgaagtgaat 3420
agcaagtgct atgaggaagc taagaagctg aagaagatca gcaaccaggc cgagtttatc 3480
gcctccttct acaacaacga tctgatcaag atcaacggcg agctgtatag agtgatcggc 3540
gtgaacaacg acctgctgaa ccggatcgaa gtgaacatga tcgacatcac ctaccgcgag 3600
tacctggaaa acatgaacga caagaggccc cccaggatca ttaagacaat cgccggaagc 3660
ggagctacta acttcagcct gctgaagcag gctggagacg tggaggagaa ccctggacct 3720

aggcgcgccg ccaccatggt gagcaagggc gaggagctgt tcaccggggt ggtgcccatc 3780
ctggtcgagc tggacggcga cgtaaacggc cacaagttca gcgtgtccgg cgagggcgag 3840
ggcgatgcca cctacggcaa gctgaccctg aagttcatct gcaccaccgg caagctgccc 3900
gtgccctggc ccaccctcgt gaccaccttc ggctacggcc tgatgtgctt cgcccgctac 3960
cccgaccaca tgaagcagca cgacttcttc aagtccgcca tgcccgaagg ctacgtccag 4020
gagcgcacca tcttcttcaa ggacgacggc aactacaaga cccgcgccga ggtgaagttc 4080
gagggcgaca ccctggtgaa ccgcatcgag ctgaagggca tcgacttcaa ggaggacggc 4140

aacatcctgg ggcacaagct ggagtacaac tacaacagcc acaacgtcta tatcatggcc 4200
gacaagcaga agaacggcat caagnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 4260
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnngataaa aggcatgcac gtttgcggct 4320
acgtgcatgc caaaaggagt cgggcttgcc tccgtgcccg actccaaaag acctgctcga 4380
ggaggtggac gagcaggtca aaaatccggg taccaataaa atatctttat tttcattaca 4440
tctgtgtgtt ggtttttgt gtg 4463
<210> 172
<211> 3467
<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 172

cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc cccgcccatt 60

gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc attgacgtca 120

atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt atcatatgcc 180

aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta 240

catgacctta tgggactttc ctacttggca gtacatctac gtattagtca tcgctattac 300

catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg actcacgggg 360

atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc aaaatcaacg 420

ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggcg gtaggcgtgt 480

acggtgggag gtctatataa gcagagcttg gatgttgcct ttaggatttt tgacctgctc 540

gattgtccac tgcgagcagg tcttttggag tcgggcgagg cggaagcccg actccttttg 600

gcatgcacgc tagccgcgtc gtgcatgcct tttatcttcg ggttatggga ccagtgaagg 660

ctgagggaag gactgtcctg ggactggaca ggcgggttat gggacctgaa aatactaaca 720

atcgattttt tttccctttt tttccaggtg gacagaccca gagtggatat cgagtgtgct 780

ggcaaagggg tgcagagcag cctgatccat aactacaaga agaaccccaa cttcaacacc 840

ctggtcaagt ggttcgaagt ggatctgccc gagaacgaac tgctgcaccc acctctgaac 900

atcagagtgg tggactgcag agccttcggc agatacaccc tcgtgggatc tcacgccgtg 960

tctagcctga gaagattcat ctacagacct ccagacagaa gcgcccctaa ctggaacaca 1020

acaggcgagg tggtggtgtc catggaaccc gaggaacccg tgaagaaact ggaaaccatg 1080

gtcaagctgg acgccacctc cgatgctgtc gtgaaagtgg acgtggccga ggacgagaaa 1140

gagcgcaaga agaagaaaaa gaagggcccc agcgaggaac ctgaagagga agaacctgac 1200

gagagcatgc tggactggtg gtccaagtac ttcgcctcca tcgacacaat gaaggaacag 1260

ctgagacagc acgagacaag cggcaccgac ctcgaagaga aagaagagat ggaatccgcc 1320

gaaggactga agggccctat gaagtccaaa gagaagtcta gggccgccaa agaagagaaa 1380

aaaaagaaga accagtctcc tggaccaggc cagggatctg aggctcccga aaagaaaaag 1440

gccaagatcg acgagctgaa ggtgtacccc aaagagctgg aaagcgagtt cgacagcttc 1500

gaggactggc tgcacacctt caatctgctg agaggaaaga caggcgacga cgaggatggc 1560

agcactgaag aagagagaat cgtcggcaga ttcaagggca gcctgtgcgt gtacaaggtg 1620

ccactgcctg aggacgtgtc cagagaggct ggctacgatc ctacctacgg catgttccaa 1680
ggcatcccta gcaacgaccc catcaatgtg ctcgtgcgga tctatgtcgt gcgggccact 1740
gatctgcatc ccgccgatat caacggcaag gcagacccct atatcgctat caagctgggg 1800
aaaaccgaca tcagggacaa agagaactac atcagcaagc agctgaaccc cgtgttcggc 1860
aagagcttcg acatcgaggc tagcttcccc atggaatcca tgctgaccgt ggccgtgtac 1920
gactgggatc tcgtgggaac agacgacctg atcggagaga caaagattga cctggaaaac 1980
cggttctact ccaagcaccg ggccacctgt ggaatcgccc agacctactc tatccacggc 2040

tacaacatct ggcgggaccc catgaagcct agccagatcc tgaccaggct gtgcaaagaa 2100
ggcaaggtcg acggccctca ctttggacct cacggccggg tcagagtggc caacagagtg 2160
ttcacaggcc cctccgagat cgaggatgag aacggccaga gaaagcccac cgatgagcat 2220
gtggctctga gcgctctgag acactgggaa gatatcccta gagtgggctg cagactggtg 2280
cccgagcacg tggaaacaag acccctgctg aacccagaca agcccggaat cgaacagggc 2340
agactcgaac tgtgggtcga catgttccct atggacatgc ccgcacctgg cacaccactg 2400
gacatcagcc ctaggaagcc caagaaatac gagctgcgcg tgatcgtgtg gaacaccgac 2460

gaagtggtgc tggaagatga cgacttcttc accggcgaaa agtccagcga catcttcgtc 2520
agaggatggc tgaagggaca gcaagaggat aagcaggaca ccgacgtgca ctaccacagc 2580
cttacaggcg aaggcaactt taactggcgc tacctgtttc ctttcgacta cctggccgcc 2640
gaagagaaga tcgtgatgtc caagaaagaa tctatgttca gctgggacga gacagagtac 2700
aagatccccg ccagactgac cctgcagatc tgggatgccg atcacttcag cgccgacgac 2760
tttctgggag ccatcgagct ggacctgaat agattcccca gaggcgccaa gaccgccaag 2820
cagtgcacaa tggaaatggc cactggcgag gtcgacgtgc cactggtgtc tatcttcaag 2880

cagaagcgcg tcaaaggctg gtggcccctg ctggctagaa acgagaacga cgagttcgag 2940
ctgaccggaa aggtggaagc cgagctgcat ctgctgacag ctgaagaggc cgagaagaat 3000
cctgtgggcc tcgctaggaa tgagcccgat cctctggaaa agcccaacag acccgatacc 3060
gccttcgtgt ggtttctgaa cccactgaag tccatcaagt acctgatctg tacccggtac 3120
aagtggctga ttatcaagat cgtgctggcc ctgctggggc tgctgatgct tgctctgttc 3180
ctgtactccc tgcctggcta tatggtcaag aagctgctgg gcgccggcgc tcgggctgac 3240
tacaaagacc atgacggtga ttataaagat catgacatcg actataagga tgacgatgac 3300

aaatgaggta ccaattcctc acctgcgatc tcgatgcttt atttgtgaaa tttgtgatgc 3360

tattgcttta tttgtaacca ttataagctg caataaacaa gttaacaaca acaattgcat 3420

tcattttatg tttcaggttc agggggaggt gtgggaggtt ttttaaa 3467

<210> 173

<211> 33

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 173

gtaagtattg ctttcatttt tgtctttttt taa 33

<210> 174

<211> 30

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 174

gtaagttctt gctttgttca aactgtctat 30

<210> 175

<211> 27

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 175

gtaagtattc ttttgttctt cactcat 27

<210> 176

<211> 32

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 176

gtaagtattt ttttactcct catttttact cc 32

<210> 177

<211

> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 177

gtaagtattt ttttacggtt atattctcct ttcccc 36

<210> 178

<211> 28

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 178

gtaagtattt tctgttgttt attttcag 28

<210> 179

<211> 39

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 179

gtaagtattg gggttgatta tgtgtgggac ggtgtaagg 39

<210> 180

<211> 35

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 180

gtaagtattt cctctttctt tccatgggtt ggcct 35

<210> 181

<211> 35

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 181

gtaagtatta ccagagattc gtagacctgc ttgac 35

<210> 182

<211> 39

<212>

DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 182

tggggctggg cagagggttg aggggagagg gtcctgggg 39

<210> 183

<211> 28

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 183

tcatgggtgg gttcattggg tgggttca 28

<210> 184

<211> 23

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 184

tagggcgcag tagtccaggg ttt 23

<210> 185

<211> 30

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 185

ttctctgtgg ggtggcattc tctgctctct 30

<210> 186

<211> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 186

gggttatggg acctcaggga taagggacc 29

<210> 187

<211> 15

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 187

cggggatggg ggtca 15

<210> 188

<211> 23

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 188

tgggggggagg tcatgggggg agg 23

<210> 189

<211> 24

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 189

gttggtggtt tcatgttggt ggtt 24

<210> 190

<211

> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 190

gggtttcggg ttttcaggtg gtcgttggt 29

<210> 191

<211> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 191

ggtggtcgtt ggttcatttg ggctattgg 29

<210> 192

<211> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 192

tttgggctat tggtcaaggg ggcgagggg 29

<210> 193

<211> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 193

aggggggcgag gggtcaggta ttcggtatt 29

<210> 194

<211> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 194

ggtattcggt atttcaaggt aacaggtaa 29

<210> 195

<211> 29

<212>

DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 195

aggtaacagg taatcagggt ttcgggttt 29

<210> 196

<211> 29

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 196

tcttactttt gtaaacttta tggtttgtg 29

<210> 197

<211> 28

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 197

cacgtattct cggtacggac gttacaga 28

<210> 198

<211> 13

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 198

taagctggta tcc 13

<210> 199

<211> 34

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 199

cactaactct ttttcccccc ttttttttt acag 34

<210> 200

<211> 36

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 200

tactaactct ttctttttc ctttccttct tcacag 36

<210> 201

<211> 43

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 201

cactaactct gtcatactta tcctgtccct ttttttttcca cag 43

<210> 202

<211> 45

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 202

cactaactct ctttcttttt cttccctcct ctcccccaac tgcag 45

<210> 203

<211

> 38

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 203

cactaactct ttttttttttt ttttttttttt tacagcag 38

<210> 204

<211> 13

<212> DNA

<213> Artificial sequence

<220><223> Synthetic nucleic acid sequence

<400> 204

taagctggta tcc 13

<210> 205

<211> 8

<212> DNA

<213> Artificial Sequence

<220><223> branch point sequence

<400> 205

tactaaca 8

<210> 206

<211> 49

<212> DNA

<213> Artificial Sequence

<220><223> polyadenylation signal

<400> 206

aataaaatat ctttattttc attacatctg tgtgttggtt ttttgtgtg 49