



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2021-0132109
(43) 공개일자 2021년11월03일

- (51) 국제특허분류(Int. Cl.)
C12N 15/86 (2006.01) A61K 48/00 (2006.01)
A61P 25/08 (2006.01) C12N 15/113 (2010.01)
C12N 9/22 (2006.01)
- (52) CPC특허분류
C12N 15/86 (2013.01)
A61K 48/00 (2013.01)
- (21) 출원번호 10-2021-7030158
- (22) 출원일자(국제) 2020년02월24일
심사청구일자 없음
- (85) 번역문제출일자 2021년09월17일
- (86) 국제출원번호 PCT/US2020/019546
- (87) 국제공개번호 WO 2020/176426
국제공개일자 2020년09월03일
- (30) 우선권주장
62/810,005 2019년02월25일 미국(US)
- (71) 출원인
유니버시티 오브 매사추세츠
미국 매사추세츠주 02108 보스턴 31 플로어 윈 비컨 스트리트
- (72) 발명자
에스테베스, 미구엘, 세나
미국 01886 매사추세츠주 웨스트포드 쉬폴리 서클 7
울프, 스콧, 에이.
미국 01890 매사추세츠주 윈체스터 할리우드 로드 28
- (74) 대리인
양영준, 김영

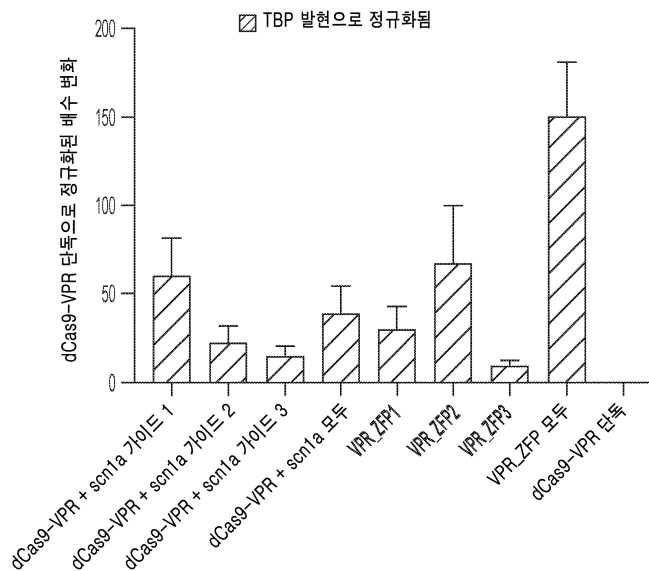
전체 청구항 수 : 총 88 항

(54) 발명의 명칭 DNA-결합 도메인 전사활성화제 및 이의 용도

(57) 요약

일부 양태에서, 본 개시내용은 DNA-결합 도메인 및 전사 조절자 도메인을 포함하는 융합 단백질을 암호화하는 핵산을 포함하는 재조합 아데노 관련 바이러스 (rAAV) 및 이를 사용하는 방법에 관한 것이다. 일부 구현예에서, 융합 단백질의 발현은 세포에서 표적 유전자의 변형된 발현을 초래한다.

대표도 - 도8



(52) CPC특허분류

A61P 25/08 (2018.01)

C12N 15/113 (2013.01)

C12N 9/22 (2013.01)

C07K 2319/80 (2013.01)

C12N 2750/14143 (2013.01)

명세서

청구범위

청구항 1

적어도 하나의 전사 조절자 도메인에 융합된 적어도 하나의 DNA 결합 도메인을 발현하도록 구성된 이식유전자를 포함하는 단리된 핵산으로서, 상기 DNA 결합 도메인은 표적 유전자 또는 표적 유전자의 조절 영역에 결합하고, 상기 표적 유전자는 전압 의존성 나트륨 채널을 암호화하는 단리된 핵산.

청구항 2

제1항에 있어서, 상기 이식유전자는 아데노 관련 바이러스 (AAV)로부터 유래된 역위 말단 반복 (ITR) 측면에 있는 단리된 핵산.

청구항 3

제1항 또는 제2항에 있어서, 상기 전사 조절자 도메인은 표적 유전자의 발현을 상향 조절하는 단리된 핵산.

청구항 4

제1항 내지 제3항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 상기 표적 유전자의 비번역 영역에 결합하는 단리된 핵산.

청구항 5

제4항에 있어서, 상기 비번역 영역은 인핸서, 프로모터, 인트론, 및/또는 억제자인 단리된 핵산.

청구항 6

제4항 또는 제5항에 있어서, 상기 DNA 결합 도메인은 표적 유전자의 조절 영역의 2-2000 bp 업스트림 또는 2-2000 bp 다운스트림에 결합하는 단리된 핵산.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 징크 핑거 단백질 (ZFP), 전사-활성화제 유사 효과기 (TALE), dCas 단백질 (예를 들어, dCas9 또는 dCas12a), 및/또는 호메오도메인을 암호화하는 단리된 핵산.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 5-7 중 어느 하나에 제시된 핵산 서열에 결합하는 단리된 핵산.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 11-16, 23-28, 또는 35-40 중 어느 하나에 제시된 서열을 갖는 핵산에 의해 암호화되는 인식 나선을 포함하는 징크 핑거 단백질인 단리된 핵산.

청구항 10

제9항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 11을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 12를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 13을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 14를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 15를 포함하는 핵산에 의해 암호화된 인식 나선, 및/또는 서열 번호 16을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질인 단리된 핵산.

청구항 11

제9항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 23을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 24를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 25를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 26을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 27을 포함하는 핵산에 의해 암호화된 인식 나선, 및/또는 서열 번호 28을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질인 단리된 핵산.

청구항 12

제9항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 35를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 36을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 37을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 38을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 39를 포함하는 핵산에 의해 암호화된 인식 나선, 및/또는 서열 번호 40을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질인 단리된 핵산.

청구항 13

제1항 내지 12항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 17-22, 29-34, 또는 41-46 중 어느 하나에 제시된 아미노산 서열을 포함하는 징크 핑거 단백질인 단리된 핵산.

청구항 14

제13항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 17을 포함하는 인식 나선, 서열 번호 18을 포함하는 인식 나선, 서열 번호 19를 포함하는 인식 나선, 서열 번호 20을 포함하는 인식 나선, 서열 번호 21을 포함하는 인식 나선, 및/또는 서열 번호 22를 포함하는 인식 나선을 포함하는 징크 핑거 단백질인 단리된 핵산.

청구항 15

제13항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 29를 포함하는 인식 나선, 서열 번호 30을 포함하는 인식 나선, 서열 번호 31을 포함하는 인식 나선, 서열 번호 32를 포함하는 인식 나선, 서열 번호 33을 포함하는 인식 나선, 및/또는 서열 번호 34를 포함하는 인식 나선을 포함하는 징크 핑거 단백질인 단리된 핵산.

청구항 16

제13항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 41을 포함하는 인식 나선, 서열 번호 42를 포함하는 인식 나선, 서열 번호 43을 포함하는 인식 나선, 서열 번호 44를 포함하는 인식 나선, 서열 번호 45를 포함하는 인식 나선, 및/또는 서열 번호 46을 포함하는 인식 나선을 포함하는 징크 핑거 단백질인 단리된 핵산.

청구항 17

제1항 내지 제7항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 dCas 단백질, 선택적으로 dCas9 단백질이고, 상기 단리된 핵산은 적어도 하나의 가이드 핵산을 추가로 포함하는 단리된 핵산.

청구항 18

제17항에 있어서, 상기 가이드 핵산은 *SCN1A*를 표적으로 하는 스페이서 서열을 포함하는 단리된 핵산.

청구항 19

제17항 또는 제18항에 있어서, 상기 가이드 핵산은 서열 번호 85, 86, 89, 90, 93, 또는 94 중 어느 하나의 뉴클레오티드 서열을 갖는 스페이서 서열을 포함하는 단리된 핵산.

청구항 20

제17항 내지 제19항 중 어느 한 항에 있어서, 상기 가이드 핵산은 서열 번호 83-94 중 어느 하나의 뉴클레오티드 서열을 포함하는 단리된 핵산.

청구항 21

제1항 내지 제16항 중 어느 한 항에 있어서, 상기 적어도 하나의 전사 조절자 도메인은 VPR, Rta, p65, 또는 Hsf1 전사활성화제, 또는 이의 임의의 조합을 포함하는 단리된 핵산.

청구항 22

제1항 내지 제21항 중 어느 한 항에 있어서, 상기 적어도 하나의 전사활성화 도메인은 서열 번호 47에 제시된 핵산 서열에 의해 암호화되는 단리된 핵산.

청구항 23

제1항 내지 제22항 중 어느 한 항에 있어서, 상기 적어도 하나의 전사활성화 도메인은 서열 번호 48에 제시된 아미노산 서열에 의해 암호화되는 단리된 핵산.

청구항 24

제1항 내지 제23항 중 어느 한 항에 있어서, 상기 핵산은 AAV2 ITR을 포함하는 단리된 핵산.

청구항 25

제24항에 있어서, 상기 ITR은 Δ TR 및/또는 mTR인 단리된 핵산.

청구항 26

제1항 내지 제25항 중 어느 한 항에 있어서, 상기 이식유전자는 프로모터에 작동 가능하게 연결된 단리된 핵산.

청구항 27

제26항에 있어서, 상기 프로모터는 조직 특이적 프로모터이고, 선택적으로 상기 프로모터는 SST, NPY, 포스페이트 활성화 글루타미나제 (PAG), 소포 글루타메이트 수송체-1 (VGLUT1), 글루탐산 탈탄산효소 65 및 57 (GAD65, GAD67), 시냅신 I, α -CamKII, Dock10, Prox1, 파브알부민 (PV), 소마토스타틴 (SST), 콜레시스토키닌 (CCK), 칼레티닌 (CR), 또는 뉴로펩티드 Y (NPY)와 같은 신경 프로모터인 단리된 핵산.

청구항 28

제1항 내지 제27항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 링커 도메인에 의해 적어도 하나의 전사 조절자 도메인에 융합된 단리된 핵산.

청구항 29

제28항에 있어서, 상기 링커 도메인은 선택적으로:

- (i) 선택적으로 글리신으로 구성된, 가요성 링커, 또는
- (ii) 절단 가능한 링커인 단리된 핵산.

청구항 30

제1항 내지 제29항 중 어느 한 항에 있어서, 상기 이식유전자는 1개의 DNA 결합 도메인, 2개의 DNA 결합 도메인, 3개의 DNA 결합 도메인, 4개의 DNA 결합 도메인, 5개의 DNA 결합 도메인, 6개의 DNA 결합 도메인, 7개의 DNA 결합 도메인, 8개의 DNA 결합 도메인, 9개의 DNA 결합 도메인, 또는 10개의 DNA 결합 도메인을 암호화하는 단리된 핵산.

청구항 31

제1항 내지 제30항 중 어느 한 항에 있어서, 상기 이식유전자는 1개의 전사 조절자 도메인, 2개의 전사 조절자 도메인, 3개의 전사 조절자 도메인, 4개의 전사 조절자 도메인, 5개의 전사 조절자 도메인, 6개의 전사 조절자 도메인, 7개의 전사 조절자 도메인, 8개의 전사 조절자 도메인, 9개의 전사 조절자 도메인, 또는 10개의 전사 조절자 도메인을 암호화하는 단리된 핵산.

청구항 32

하기를 포함하는 재조합 AAV (rAAV):

(i) 적어도 하나의 전사 조절자 도메인에 융합된 적어도 하나의 DNA 결합 도메인을 암호화하는 이식유전자를 포함하는 핵산으로서, 상기 DNA 결합 도메인은 상기 표적 유전자가 전압 의존성 나트륨 채널을 암호화하는 표적 유전자 또는 유전자의 조절 영역에 결합하는 핵산,

(ii) 적어도 하나의 캡시드 단백질.

청구항 33

제32항에 있어서, 상기 이식유전자는 아데노 관련 바이러스 (AAV)로부터 유래된 역위 말단 반복 (ITR) 측면에 있는 rAAV.

청구항 34

제32항 또는 제33항에 있어서, 상기 전사 조절자는 표적 유전자의 발현을 상향 조절하는 rAAV.

청구항 35

제32항 내지 제34항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 표적 유전자의 비번역 영역에 결합하는 rAAV.

청구항 36

제35항에 있어서, 상기 비번역 영역은 인핸서, 프로모터, 인트론, 및/또는 억제자인 rAAV.

청구항 37

제35항 또는 제36항에 있어서, 상기 DNA 결합 도메인은 표적 유전자의 조절 영역의 2-2000 bp 업스트림 또는 2-2000 bp 다운스트림에 결합하는 rAAV.

청구항 38

제32항 내지 제37항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 징크 핑거 단백질 (ZFP), 전사-활성화제 유사 효과기 (TALE), dCas 단백질 (예를 들어, dCas9 또는 dCas12a), 및/또는 호메오도메인을 암호화하는 rAAV.

청구항 39

제32항 내지 제38항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 5-7 중 어느 하나에 제시된 핵산 서열에 결합하는 rAAV.

청구항 40

제32항 내지 제39항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 11-16, 23-28, 또는 35-40 중 어느 하나에 제시된 서열을 갖는 핵산에 의해 암호화되는 인식 나선을 포함하는 징크 핑거 단백질인 rAAV.

청구항 41

제40항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 11을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 12를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 13을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 14를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 15를 포함하는 핵산에 의해 암호화된 인식 나선, 및/또는 서열 번호 16을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질인 rAAV.

청구항 42

제40항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 23을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 24를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 25를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 26을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 27을 포함하는 핵산에 의

해 암호화된 인식 나선, 및/또는 서열 번호 28을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질인 rAAV.

청구항 43

제40항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 35를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 36을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 37을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 38을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 39를 포함하는 핵산에 의해 암호화된 인식 나선, 및/또는 서열 번호 40을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질인 rAAV.

청구항 44

제32항 내지 제43항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 17-22, 29-34, 또는 41-46 중 어느 하나에 제시된 서열을 포함하는 징크 핑거 단백질인 rAAV.

청구항 45

제44항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 17을 포함하는 인식 나선, 서열 번호 18을 포함하는 인식 나선, 서열 번호 19를 포함하는 인식 나선, 서열 번호 20을 포함하는 인식 나선, 서열 번호 21을 포함하는 인식 나선, 및/또는 서열 번호 22를 포함하는 인식 나선을 포함하는 징크 핑거 단백질인 rAAV.

청구항 46

제44항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 29를 포함하는 인식 나선, 서열 번호 30을 포함하는 인식 나선, 서열 번호 31을 포함하는 인식 나선, 서열 번호 32를 포함하는 인식 나선, 서열 번호 33을 포함하는 인식 나선, 및/또는 서열 번호 34를 포함하는 인식 나선을 포함하는 징크 핑거 단백질인 rAAV.

청구항 47

제44항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 41을 포함하는 인식 나선, 서열 번호 42를 포함하는 인식 나선, 서열 번호 43을 포함하는 인식 나선, 서열 번호 44를 포함하는 인식 나선, 서열 번호 45를 포함하는 인식 나선, 및/또는 서열 번호 46을 포함하는 인식 나선을 포함하는 징크 핑거 단백질인 rAAV.

청구항 48

제32항 내지 제37항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 dCas 단백질, 선택적으로 dCas9 단백질이고, 상기 rAAV는 적어도 하나의 가이드 핵산을 추가로 포함하는 rAAV.

청구항 49

제48항에 있어서, 상기 가이드 핵산은 *SCN1A*를 표적으로 하는 스페이서 서열을 포함하는 rAAV.

청구항 50

제48항 또는 제49항에 있어서, 상기 가이드 핵산은 서열 번호 85, 86, 89, 90, 93, 또는 94 중 어느 하나의 뉴클레오티드 서열을 갖는 스페이서 서열을 포함하는 rAAV.

청구항 51

제48항 내지 제50항 중 어느 한 항에 있어서, 상기 가이드 핵산은 서열 번호 83-94 중 어느 하나의 뉴클레오티드 서열을 포함하는 rAAV.

청구항 52

제32항 내지 제47항 중 어느 한 항에 있어서, 상기 적어도 하나의 전사 조절자 도메인은 VRP, Rta, p65, Hsf1 또는 이의 임의의 조합으로부터 유래된 전사활성화제인 rAAV.

청구항 53

제32항 내지 제52항 중 어느 한 항에 있어서, 상기 적어도 하나의 전사 조절자 이식유전자를 암호화하는 이식유전자는 서열 번호 47에 제시된 서열을 포함하는 rAAV.

청구항 54

제32항 내지 제53항 중 어느 한 항에 있어서, 상기 적어도 하나의 전사 조절자 도메인은 서열 번호 48에 제시된 아미노산 서열에 의해 암호화되는 핵산.

청구항 55

제32항 내지 제54항 중 어느 한 항에 있어서, 상기 적어도 하나의 DNA 결합 도메인은 링커 도메인에 의해 적어도 하나의 전사 조절자 도메인에 융합되는 rAAV.

청구항 56

제55항에 있어서, 상기 링커 도메인은 선택적으로:

- (i) 선택적으로 글리신으로 구성된, 가요성 링커, 또는
- (ii) 절단 가능한 링커인 rAAV.

청구항 57

제32항 내지 제56항 중 어느 한 항에 있어서, 상기 이식유전자는 1개의 DNA 결합 도메인, 2개의 DNA 결합 도메인, 3개의 DNA 결합 도메인, 4개의 DNA 결합 도메인, 5개의 DNA 결합 도메인, 6개의 DNA 결합 도메인, 7개의 DNA 결합 도메인, 8개의 DNA 결합 도메인, 9개의 DNA 결합 도메인, 또는 10개의 DNA 결합 도메인을 암호화하는 rAAV.

청구항 58

제32항 내지 제57항 중 어느 한 항에 있어서, 상기 이식유전자는 1개의 전사 조절자 도메인, 2개의 전사 조절자 도메인, 3개의 전사 조절자 도메인, 4개의 전사 조절자 도메인, 5개의 전사 조절자 도메인, 6개의 전사 조절자 도메인, 7개의 전사 조절자 도메인, 8개의 전사 조절자 도메인, 9개의 전사 조절자 도메인, 또는 10개의 전사 조절자 도메인을 암호화하는 rAAV.

청구항 59

제32항 내지 제58항 중 어느 한 항에 있어서, 상기 AAV 캡시드 혈청형은 AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV9, AAV10, AAVrh10, 또는 AAV.PHPB로 이루어진 그룹으로부터 선택되는 rAAV.

청구항 60

제32항 내지 제59항 중 어느 한 항에 있어서, 상기 AAV 캡시드 혈청형은 AAV9인 rAAV.

청구항 61

제32항 내지 제59항 중 어느 한 항에 있어서, 상기 AAV 캡시드 혈청형은 AAV.PHPB인 rAAV.

청구항 62

제32항 내지 제61항 중 어느 한 항에 있어서, 상기 핵산은 AAV2 ITR을 포함하는 rAAV.

청구항 63

제62항에 있어서, 상기 ITR은 Δ TR 및/또는 mTR인 rAAV.

청구항 64

제32항 내지 제63항 중 어느 한 항에 있어서, 상기 이식유전자는 프로모터에 작동 가능하게 연결된 rAAV.

청구항 65

제64항에 있어서, 상기 프로모터는 조직 특이적 프로모터인 rAAV.

청구항 66

제64항 또는 제65항에 있어서, 상기 조직 특이적 프로모터는 신경 프로모터, 예를 들어 SST, NPY, 포스페이트 활성화 글루타미나제 (PAG), 소포 글루타메이트 수송체-1 (VGLUT1), 글루탐산 탈탄산효소 65 및 57 (GAD65, GAD67), 시냅신 I, a-CamKII, Dock10, Prox1, 파브알부민 (PV), 소마토스타틴 (SST), 콜레시스토키닌 (CCK), 칼레티닌 (CR), 또는 뉴로펩티드 Y (NPY)인 rAAV.

청구항 67

표적 유전자의 발현을 증가시키는 방법으로서, 상기 방법은 표적 유전자를 포함하는 세포 또는 대상체에 제1항 내지 제31항 중 어느 한 항의 단리된 핵산 또는 제32항 내지 제66항 중 어느 한 항의 rAAV를 투여하는 것을 포함하는 방법.

청구항 68

제57항에 있어서, 상기 대상체는 표적 유전자에 대해 반가불충분한 것인 방법.

청구항 69

제67항 또는 제68항에 있어서, 상기 표적 유전자는 SCN1A인 방법.

청구항 70

제67항 내지 제69항 중 어느 한 항에 있어서, 상기 세포는 뉴런, 선택적으로 GABA성 뉴런인 방법.

청구항 71

제67항 내지 제70항 중 어느 한 항에 있어서, 상기 제1항 내지 제31항 중 어느 한 항의 단리된 핵산 또는 제32항 내지 제66항 중 어느 한 항의 rAAV의 투여는 투여 전 대상체에서 이식유전자의 발현에 비해 적어도 2배, 적어도 10배, 적어도 20배, 적어도 30배, 적어도 40배, 적어도 50배, 적어도 60배, 적어도 70배, 적어도 80배, 적어도 90배, 또는 적어도 100배 증가된 표적 유전자 발현을 초래하는 방법.

청구항 72

대상체에서 드라베 증후군을 치료하는 방법으로서, 표적 유전자를 발현하는 대상체에 제1항 내지 제31항 중 어느 한 항의 단리된 핵산 또는 제32항 내지 제66항 중 어느 한 항의 rAAV를 투여하는 것은 표적 유전자를 발현하는 대상체에 투여되는 것을 포함하는 방법.

청구항 73

제72항에 있어서, 상기 대상체에서 표적 유전자의 발현은 정상 대상체에 비해 감소되는 것인 방법.

청구항 74

제72항 또는 제73항에 있어서, 상기 대상체는 정상 대상체에 비해 표적 유전자 발현에서 반가불충분하거나 반가불충분으로 의심되는 것인 방법.

청구항 75

제72항 내지 제74항 중 어느 한 항에 있어서, 상기 대상체는 표적 유전자의 반가불충분 발현에 의해 야기된 병태를 갖거나 갖는 것으로 의심되는 방법.

청구항 76

제72항 내지 제75항 중 어느 한 항에 있어서, 상기 표적 유전자는 SCN1A인 방법.

청구항 77

제72항 내지 제76항 중 어느 한 항에 있어서, 상기 단리된 핵산 또는 rAAV는 정맥 내 주사, 근육 내 주사,

흡입, 피하 주사, 및/또는 두개 내 주사에 의해 투여되는 방법.

청구항 78

제72항 내지 제77항 중 어느 한 항에 있어서, 상기 rAAV의 단리된 핵산의 투여는 투여 전 대상체에서 이식유전자의 발현에 비해 적어도 2배, 적어도 10배, 적어도 20배, 적어도 30배, 적어도 40배, 적어도 50배, 적어도 60배, 적어도 70배, 적어도 80배, 적어도 90배, 또는 적어도 100배 증가된 표적 유전자 발현을 초래하는 방법.

청구항 79

제1항 내지 제31항 중 어느 한 항의 단리된 핵산 또는 제32항 내지 제66항 중 어느 한 항의 rAAV를 포함하는 조성물.

청구항 80

제79항에 있어서, 약학적으로 허용 가능한 담체를 추가로 포함하는 조성물.

청구항 81

하기를 포함하는 키트:

제1항 내지 제31항 중 어느 한 항의 단리된 핵산 또는 제32항 내지 제66항 중 어느 한 항의 rAAV를 수용하는 용기.

청구항 82

제81항에 있어서, 상기 키트는 약학적으로 허용 가능한 담체를 수용하는 용기를 추가로 포함하는 키트.

청구항 83

제81항 또는 제82항에 있어서, 상기 rAAV의 단리된 핵산 및 약학적으로 허용 가능한 담체가 동일한 용기에 수용되는 키트.

청구항 84

제81항 내지 제83항 중 어느 한 항에 있어서, 상기 용기는 주사기인 키트.

청구항 85

제1항 내지 제31항 중 어느 한 항의 단리된 핵산 또는 제32항 내지 제66항 중 어느 한 항의 rAAV를 포함하는 숙주 세포.

청구항 86

제85항에 있어서, 상기 숙주 세포가 진핵 세포인 숙주 세포.

청구항 87

제85항에 있어서, 상기 숙주 세포가 포유 동물 세포인 숙주 세포.

청구항 88

제87항에 있어서, 상기 숙주 세포는 인간 세포, 선택적으로 뉴런, 선택적으로 GABA성 뉴런인 숙주 세포.

발명의 설명

기술 분야

관련 출원

[0001]

본 출원은 "ZINC FINGER PROTEIN TRANSACTIVATORS AND USES THEREOF"라는 제목으로 2019년 2월 25일에 출원된, 미국 가출원 일련 번호 62/810,005의 출원일의35 U.S.C. § 119(e)에 따른 이익을 주장하고, 이의 전체 내용은

[0002]

본원에 참고로 포함된다.

배경 기술

- [0003] 표적 유전자 발현의 조절은 의생명 연구의 주요 분야로 떠올랐다. 유전자 발현의 상향 조절은 감소된 유전자 발현으로 인한 반가불충분 조건을 수정할 수 있다. 반가불충분성은 일반적으로 하나 이상의 기능 상실 돌연변이가 유전자의 적어도 하나의 카피에 존재할 때 발생한다. 반가불충분성과 관련된 질환의 치료를 위한 유전자 증강의 AAV-기반 접근법은 전통적인 rAAV 벡터의 패키징 능력으로 인해 방해받는다.

발명의 내용

- [0004] 본 개시내용의 양태는 유전자 전달을 위한 단리된 핵산 및 재조합 AAV 벡터에 관한 것이다. 본 개시내용은 부분적으로, 조성물 (예를 들어, rAAV 벡터 및 rAAV) 및 표적 유전자의 발현을 조절하기 위한 방법을 기반으로 하고, 상기 표적 유전자는 반가불충분, 예를 들어 *SCN1A*이다. 일부 구현예에서, 본 개시내용은 DNA 결합 도메인, 예를 들어 Cys2-His2 징크 핑거 단백질 (ZFP), 및 전사 조절자 도메인을 포함하는 융합 단백질을 제공한다. 일부 구현예에서, 본 개시내용에 의해 기술된 조성물은 전사 조절자 도메인에 융합된 DNA 결합 도메인 (예를 들어, ZFP, 전사 활성화자 유사 효과기 (TALE) 도메인 등)을 포함하는 융합 단백질을 포함한다. 일부 구현예에서, 본 개시내용에 의해 기술된 융합 단백질은 표적 유전자 (예를 들어, *SCN1A*)의 발현을 증가시키고, 따라서 정상 세포 또는 대상체와 비교하여 세포 또는 대상체에서 표적 유전자의 결핍된 발현을 특징으로 질환 (예를 들어, 표적 유전자의 반가불충분성과 관련된 질환)을 치료하는데 유용하다.
- [0005] 따라서, 일부 양태에서, 본 개시내용은 적어도 하나의 전사 조절자 도메인에 융합된 적어도 하나의 DNA 결합 도메인을 발현하도록 구성된 이식유전자를 포함하는 단리된 핵산을 제공하고, 상기 DNA 결합 도메인은 표적 유전자 또는 표적 유전자의 조절 영역 (예를 들어, 인핸서 서열, 프로모터 서열, 억제자 서열, 등)에 결합하고 (예를 들어 대상체 또는 세포에서), 상기 표적 유전자는 전압 의존성 나트륨 채널 (예를 들어, *Nav1.1*)을 암호화한다. 일부 구현예에서, 표적 유전자는 *SCN1A* 유전자이다. 일부 구현예에서, 이식유전자는 아데노 관련 바이러스 (AAV) 역위 말단 반복 (ITR)이 측면에 있다. 일부 구현예에서, 적어도 하나의 DNA 결합 도메인은 표적 유전자에 결합하고 (예를 들어, 대상체 또는 세포에서) 상기 전사 조절자 도메인을 표적 유전자의 발현을 변경, 예를 들어, 상향 조절한다.
- [0006] 일부 양태에서, 본 개시내용은 하기를 포함하는 재조합 AAV (rAAV)를 제공하고: 적어도 하나의 전사 조절자 도메인에 융합된 적어도 하나의 DNA 결합 도메인을 암호화하는 이식유전자를 포함하는 핵산, 상기 DNA 결합 도메인은 (예를 들어 대상체 또는 세포에서) 표적 유전자 또는 표적 유전자의 조절 영역에 결합하고, 상기 표적 유전자는 전압 의존성 나트륨 채널 (예를 들어, *Nav1.1*) 및 적어도 하나의 캡시드 단백질을 암호화한다. 일부 구현예에서, 표적 유전자는 *SCN1A* 유전자이다. 일부 구현예에서, 이식유전자는 AAV 역위 말단 반복 (ITR)이 측면에 있다.
- [0007] 일부 구현예에서, 적어도 하나의 DNA 결합 도메인은 (예를 들어, 대상체 또는 세포에서) 표적 유전자에 결합하고 전사 조절자 도메인은 대상체에서 표적 유전자의 발현을 변경, 예를 들어, 상향 조절한다.
- [0008] 일부 구현예에서, 적어도 하나의 DNA 결합 도메인은 표적 유전자의 비번역 영역에 결합한다. 일부 구현예에서, DNA 결합 도메인은 표적 유전자의 조절 영역, 선택적으로 인핸서 서열, 프로모터 서열, 및/또는 억제자 서열에 결합한다.
- [0009] 일부 구현예에서, DNA 결합 도메인은 표적 유전자의 조절 영역 (예를 들어, 인핸서 서열, 프로모터 서열, 억제자 서열 등)의 2 내지 2000 bp 업스트림 또는 2 내지 2000 bp 업스트림 또는 다운스트림에 결합한다.
- [0010] 일부 구현예에서, 적어도 하나의 DNA 결합 도메인은 징크 핑거 단백질 (ZFP), 전사 활성화자 유사 효과기 (TALE), dCas 단백질 (예를 들어, dCas9 또는 dCas12a), 및/또는 호메오도메인을 암호화한다. 일부 구현예에서, 적어도 하나의 DNA 결합 도메인은 서열 번호 5-7 중 어느 하나에 제시된 핵산 서열에 결합한다. 일부 구현예에서, 적어도 하나의 DNA 결합 도메인은 서열 번호 11-16, 23-28, 또는 35-40 중 어느 하나에 제시된 서열을 갖는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질이다. 일부 구현예에서, 적어도 하나의 DNA 결합 도메인은 서열 번호 17-22, 29-34, 또는 41-46 중 어느 하나에 제시된 아미노산 서열을 포함하는 징크 핑거 단백질이다.
- [0011] 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 11을 포함하는 핵산에 의해 암호화된 인식

나선, 서열 번호 12를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 13을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 14를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 15를 포함하는 핵산에 의해 암호화된 인식 나선, 및/또는 서열 번호 16을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질이다. 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 57의 아미노산 서열을 포함하는 징크 핑거 단백질이다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 57의 아미노산 서열에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 또는 적어도 99% 서열 동일성을 포함한다.

[0012] 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 23을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 24를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 25를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 26을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 27을 포함하는 핵산에 의해 암호화된 인식 나선, 및/또는 서열 번호 28을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질이다. 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 59의 아미노산 서열을 포함하는 징크 핑거 단백질이다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 59의 아미노산 서열에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 또는 적어도 99% 서열 동일성을 포함한다.

[0013] 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 35를 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 36을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 37을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 38을 포함하는 핵산에 의해 암호화된 인식 나선, 서열 번호 39를 포함하는 핵산에 의해 암호화된 인식 나선, 및/또는 서열 번호 40을 포함하는 핵산에 의해 암호화된 인식 나선을 포함하는 징크 핑거 단백질이다. 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 61의 아미노산 서열을 포함하는 징크 핑거 단백질이다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 61의 아미노산 서열에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 또는 적어도 99% 서열 동일성을 포함한다.

[0014] 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 17의 아미노산 서열을 포함하는 인식 나선, 서열 번호 18의 아미노산 서열을 포함하는 인식 나선, 서열 번호 19의 아미노산 서열을 포함하는 인식 나선, 서열 번호 20의 아미노산 서열을 포함하는 인식 나선, 서열 번호 21의 아미노산 서열을 포함하는 인식 나선, 및/또는 서열 번호 22의 아미노산 서열을 포함하는 인식 나선을 포함하는 징크 핑거 단백질이다.

[0015] 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 29를 포함하는 인식 나선, 서열 번호 30을 포함하는 인식 나선, 서열 번호 31을 포함하는 인식 나선, 서열 번호 32를 포함하는 인식 나선, 서열 번호 33을 포함하는 인식 나선, 및/또는 서열 번호 34를 포함하는 인식 나선을 포함하는 징크 핑거 단백질이다.

[0016] 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 서열 번호 41을 포함하는 인식 나선, 서열 번호 42를 포함하는 인식 나선, 서열 번호 43을 포함하는 인식 나선, 서열 번호 44를 포함하는 인식 나선, 서열 번호 45를 포함하는 인식 나선, 및/또는 서열 번호 46을 포함하는 인식 나선을 포함하는 징크 핑거 단백질이다.

[0017] 일부 구현예에서, 상기 적어도 하나의 DNA 결합 도메인은 촉매적 불활성 CRISPR 관련 단백질 (Cas 단백질)이다. 일부 구현예에서, 촉매적 불활성 Cas 단백질 (또는 "죽은 Cas 단백질")은 dCas9 또는 dCas12 단백질이다. 일부 구현예에서, 핵산 또는 rAAV는 적어도 하나의 가이드 핵산 (예를 들어, 가이드 RNA, 또는 gRNA)을 추가로 포함한다. 일부 구현예에서, 상기 가이드 핵산은 *SCN1A*를 표적으로 하는 스페이서 서열을 포함한다. 일부 구현예에서, 상기 가이드 핵산은 서열 번호 85, 86, 89, 90, 93, 또는 94 중 어느 하나의 뉴클레오티드 서열을 갖는 스페이서 서열을 포함한다. 일부 구현예에서, 상기 가이드 핵산은 서열 번호 83-94 중 어느 하나의 뉴클레오티드 서열을 포함한다. 일부 구현예에서, 상기 가이드 핵산은 서열 번호 83-94 중 어느 하나에 제시된 핵산 서열에 의해 암호화된다.

[0018] 일부 구현예에서, 적어도 하나의 전사 조절자 도메인은 VP16 도메인, VP64 도메인, Rta 도메인, p65 도메인, Hsf1 도메인, 또는 이의 임의의 조합, 예를 들어 VPR 도메인 (VP64+p65+Rta1 도메인)을 포함하는 전사활성화제 도메인이다. 일부 구현예에서, 적어도 하나의 전사 조절자 도메인은 서열 번호 47에 제시된 바와 같은 핵산 서열에 의해 암호화된다. 일부 구현예에서, 적어도 하나의 전사활성화 도메인은 서열 번호 48에 제시된 아미노산 서열을 포함한다.

[0019] 일부 구현예에서, 이식유전자 측면에 있는 ITR은 하기로 이루어진 그룹으로부터 선택된 ITR을 포함한다: AAV1

ITR, AAV2 ITR, AAV3 ITR, AAV4 ITR, AAV5 ITR, AAV6 ITR, AAV8 ITR, AAVrh8 ITR, AAV9 ITR, AAV10 ITR, 또는 AAVrh10 ITR. 일부 구현예에서, 상기 ITR은 Δ TR 또는 mTR이다.

- [0020] 일부 구현예에서, 단리된 핵산의 이식유전자는 프로모터에 작동 가능하게 연결된다. 일부 구현예에서, 프로모터는 조직-특이적 프로모터이다. 일부 구현예에서, 조직 특이적 프로모터는 신경 프로모터, 예를 들어 SST, NYP 포스페이트 활성화 글루타미나제 (PAG), 소포 글루타메이트 수송체-1 (VGLUT1), 글루탐산 탈탄산효소 65 및 57 (GAD65, GAD67), 시냅신 I, a-CamKII, Dock10, Prox1, 파브알부민 (PV), 소마토스타틴 (SST), 콜레키스토키닌 (CCK), 칼레티닌 (CR), 또는 뉴로펩티드 Y (NPY)이다.
- [0021] 일부 구현예에서, 이식유전자의 DNA 결합 도메인은 링커 도메인에 의해 전사 조절자 도메인에 융합된다. 일부 구현예에서, 링커 도메인은 가요성 링커, 예를 들어 글리신-풍부 링커 또는 글리신-세린 링커, 또는 절단 가능한 링커, 예를 들어 광절단 가능 링커 또는 효소 (예를 들어 프로테아제) 절단가능한 링커이다.
- [0022] 일부 구현예에서, 단리된 핵산은 다중 DNA 결합 도메인, 예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 DNA 결합 도메인을 암호화하는 이식유전자를 포함한다. 일부 구현예에서, 단리된 핵산은 다중 전사 조절자 도메인, 예를 들어 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 전사 조절자 도메인을 암호화하는 이식유전자를 포함한다.
- [0023] 일부 구현예에서, 단리된 핵산 또는 rAAV는 정상 세포 또는 대상체에 대하여 표적 유전자의 비정상적 발현 또는 반가불충분성 (예를 들어, 증가된 발현, 또는 감소된 발현)을 특징으로 하는 세포 또는 대상체에서 발현된다. 일부 구현예에서, 단리된 핵산 또는 rAAV는 정상 세포 또는 대상체에 대하여 표적 유전자의 결핍된 (예를 들어, 감소된) 발현을 특징으로 하는 세포 또는 대상체에서 발현된다. 일부 구현예에서, 단리된 핵산 또는 rAAV의 표적 유전자는 *SCN1A*이다.
- [0024] 일부 구현예에서, AAV 캡시드 혈청형은 AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV9, AAV10, AAVrh10, 또는 AAV.PHPB로 이루어진 그룹으로부터 선택된다.
- [0025] 일부 양태에서, 본 개시내용은 표적 유전자의 발현을 조절하는 방법을 제공한다. 일부 구현예에서, 본 개시내용의 방법은 표적 유전자를 발현하는 세포 또는 대상체로 본원에 기술된 바와 같이 단리된 핵 또는 rAAV를 투여하는 것을 포함하고, 상기 대상체는 표적 유전자에 대해 반가불충분 (예를 들어, *SCN1A*에 대해 반가불충분)이다. 예를 들어, 일부 구현예에서, 세포 또는 대상체에서, 표적 유전자, 예를 들어 *SCN1A*의 발현은 정상 세포 또는 대상체에서의 표적 유전자 발현에 대하여 결핍 (예를 들어, 감소)된다. 일부 구현예에서, 단리된 핵산 또는 rAAV가 투여되는 세포는 뉴런이다. 일부 구현예에서, 뉴런은 GABA성 뉴런이다.
- [0026] 일부 구현예에서, 단리된 핵산 또는 rAAV의 투여는 단리된 핵산 또는 rAAV가 투여되지 않은 대상체에 비해 적어도 2배, 적어도 10배, 적어도 20배, 적어도 30배, 적어도 40배, 적어도 50배, 적어도 60배, 적어도 70배, 적어도 80배, 적어도 90배, 또는 적어도 100배 증가된 표적 유전자 발현 (예를 들어, *SCN1A* 발현)을 초래한다. 일부 구현예에서, 단리된 핵산 또는 rAAV의 투여는 단리된 핵산 또는 rAAV가 투여되기 전 대상체에서의 표적 유전자 (예를 들어, *SCN1A*) 발현에 비해 적어도 2배, 적어도 10배, 적어도 20배, 적어도 30배, 적어도 40배, 적어도 50배, 적어도 60배, 적어도 70배, 적어도 80배, 적어도 90배, 또는 적어도 100배 증가된 표적 유전자 발현 (예를 들어, *SCN1A* 발현)을 초래한다.
- [0027] 일부 양태에서, 본 개시내용은 대상체에서 유전자 발현 (예를 들어, *SCN1A*의 발현)을 조절하는 방법을 제공하고, 여기서 본원에 기술된 바와 같은 단리된 핵산 또는 rAAV는 표적 유전자를 발현하는 대상체에 투여된다. 일부 구현예에서, 대상체에서 표적 유전자의 발현은 건강한 대상체에 대하여 비정상 (예를 들어, 증가 또는 감소)이다. 일부 구현예에서, 대상체는 건강한 대상체에 비해 표적 유전자의 발현에 관하여 반가불충분이거나 의심된다.
- [0028] 일부 구현예에서, 대상체는 표적 유전자의 반가불충분 발현에 의해 유발된 질환 또는 병태를 갖거나 의심된다. 예를 들어 *SCN1A* 발현에 대해 반가불충분한 대상체는, 일부 구현예에서, 드라베 증후군을 앓는다. 일부 구현예에서, 단리된 핵산 또는 rAAV는 정맥 내 주사, 근육 내 주사, 흡입, 피하 주사, 및/또는 두개 내 주사에 의해 대상체에게 투여된다.
- [0029] 일부 양태에서, 본 개시내용은 본 개시내용에 의해 기술된 바와 같은 단리된 핵산 또는 rAAV를 포함하는 조성물을 제공한다. 일부 구현예에서, 조성물은 약학적으로 허용 가능한 담체를 포함한다.
- [0030] 일부 양태에서, 본 개시내용은 본 개시내용에 의해 기술된 바와 같은 단리된 핵 또는 rAAV를 수용하는 용기를

포함하는 키트를 제공한다. 일부 구현예에서, 키트는 약학적으로 허용 가능한 담체를 수용하는 용기를 포함한다. 일부 구현예에서, 단리된 핵산 또는 rAAV 및 약학적으로 허용 가능한 담체는 동일한 용기에 수용된다. 일부 구현예에서, 용기는 주사기이다.

[0031] 일부 양태에서, 본 개시내용은 본 개시내용에 의해 기술된 바와 같은 단리된 핵산 또는 rAAV를 포함하는 숙주 세포를 제공한다. 일부 구현예에서, 숙주 세포는 진핵 세포이다. 일부 구현예에서, 숙주 세포는 포유동물 세포이다. 일부 구현예에서, 숙주 세포는 인간 세포, 선택적으로 뉴런, 예를 들어 GABA성 뉴런이다.

도면의 간단한 설명

[0032] **도 1**은 인간 (HEK)과 마우스 (HEPG2) SCN1A 유전자 사이 서열 보존을 표시하는 크로마토그래피 서열 분석 데이터를 나타낸다 (공통 서열 - 서열 번호 98; 표적 서열 - 서열 번호 99; Hep-SCN1A_R4 서열 (상단) - 서열 번호 100; Hep-SCN1A_R4 서열 (하단) - 서열 번호 101)

도 2는 인간 (서열 번호 1) 및 마우스 (서열 번호 2) SCN1A 유전자의 근위 프로모터 영역의 서열 정렬을 나타내고, 여기서 보존된 서열은 강조 표시되어 있다. 이 보존된 서열 내에 징크 핑거 단백질 (ZFP) 결합 영역에 대한 관심 표적 영역이 있고, 이는 굵게 표시되어 있다 (서열 번호 4).

도 3은 SCN1A 유전자의 근위 프로모터 영역에서 3개의 중첩 표적 ZFP (ZFP-1, ZFP-2, ZFP-3) (서열 번호 5-7) 결합 부위의 위치 (서열 번호 3)를 나타내는 개략도이다.

도 4a-4d는 SCN1A 유전자의 근위 프로모터 영역 (서열 번호 2) 내의 개별 3개의 염기 영역 ("·"로 구분된 빨간색으로 표시된 DNA 삼중체)을 인식할 ZFP-1에서 개별 징크 핑거 (핑거 1 내지 핑거 6; F1-F6)에 대한 6개의 인식 나선 서열의 정렬을 나타낸다. **도 4a**는 ZFP-1의 징크 핑거 1 내지 6 (F1-F6)이 결합할 뉴클레오티드 서열을 강조한다 (서열 번호 3). **도 4b**는 ZFP-1에 대한 핑거 1 내지 6의 각각의 인식 나선 (7개의 아미노산)에 의해 인식되는 3개의 뉴클레오티드 서열을 나타낸다 (서열 번호 17-22). **도 4c**는 6개의 핑거를 각 라인에 하나, 함유하는, ZFP-1의 아미노산 서열을 나타내고 여기서 핑거 사이의 링커는 표준 (TGEKP) 및 비표준 (TGSQKP) 링커 서열을 지정하기 위해 강조된다 (서열 번호 65-70). **도 4d**는 ZFP-1 (F1-F6)의 뉴클레오티드 서열을 나타낸다 (서열 번호 102-107).

도 5a-5d는 ZFP-2에서 개별 징크 핑거 (핑거 1 내지 핑거 6; F1-F6)에 대한 6개의 인식 나선 서열의 정렬을 나타내고 이는 SCN1A 유전자의 근위 프로모터 영역 내에 개별 3개의 염기 영역 ("*"로 구분된 빨간색으로 표시된 DNA 삼중체)을 인식할 것이다 (서열 번호 3). **도 5a**는 ZFP-2의 징크 핑거 1 내지 6 (F1-F6)에 결합할 뉴클레오티드 서열 (서열 번호 3)을 강조한다. **도 5b**는 ZFP-2에 대한 핑거 1 내지 6의 각각의 인식 나선 (7개의 아미노산)에 의해 인식되는 처음 3개의 뉴클레오티드를 나타낸다 (서열 번호 29-34). **도 5c**는 6개의 핑거를 각 라인에 하나, 함유하는, ZFP-2의 아미노산 서열을 나타내고 (서열 번호 69-74), 여기서 핑거 사이의 링커는 표준 (TGEKP) 및 비표준 (TGSQKP) 링커 서열을 지정하기 위해 강조된다. **도 5d**는 ZFP-2 (F1-F6)의 뉴클레오티드 서열을 나타낸다 (서열 번호 108-113).

도 6a-6d는 ZFP-3에서 개별 징크 핑거 (핑거 1 내지 핑거 6; F1-F6)에 대한 6개의 인식 나선 서열의 정렬을 나타내고 이는 SCN1A 유전자의 근위 프로모터 영역 내의 개별 3개의 염기 영역 ("*"로 구분된 빨간색으로 표시된 DNA 삼중체)을 인식할 것이다 (서열 번호 4). **도 6a**는 ZFP-3의 징크 핑거 1 내지 6 (F1-F6)에 결합할 뉴클레오티드 서열 (서열 번호 3)을 강조한다. **도 6b**는 ZFP-3에 대한 핑거 1 내지 6의 각각의 인식 나선 (7개의 아미노산)에 의해 인식되는 처음 3개의 뉴클레오티드를 나타낸다 (서열 번호 41-46). **도 6c**는 6개의 핑거를 각 라인에 하나, 함유하는, ZFP-3의 아미노산 서열을 나타내고 (서열 번호 75-80), 여기서 핑거 사이의 링커는 표준 (TGEKP) 및 비표준 (TGSQKP) 링커 서열을 지정하기 위해 강조된다. **도 6d**는 ZFP-3 (F1-F6)의 뉴클레오티드 서열을 나타낸다 (서열 번호 114-119).

도 7은 정량적 실시간 중합효소 연쇄 반응 (qRT-PCR)에 의해 측정되는, **도 4-6**에 기술된 SCN1A-결합 ZFP가 HEK293T 세포에서 SCN1A 유전자 발현을 증가시키는 것을 표시하는 데이터를 나타낸다. 이러한 발현 작제물은 하기 전사 조절자를 암호화하는 발현 플라스미드의 일시적 형질감염을 통해 세포에 전달되었다: 스트렙토코쿠스 피오케네스 Cas9 + SCN1A 가이드 RNA (SpCas9 + Scn1a); 엔도뉴클레아제 활성이 없는 Cas9 (dCas9); VPR 활성화 도메인 + SCN1A 가이드 RNA (dCas9_VPR + Scn1a); VPR 활성화 도메인 + ZFP1 (VPR_ZFP1); VPR 활성화 도메인 + ZFP2 (VPR_ZFP2); VPR 활성화 도메인 + ZFP3 (VPR_ZFP3); SpCas9 + ASCL1 가이드 RNA (SpCas9 + Asc11); 3개의 VPR_ZFP (VPR_ZFP1 + VPR_ZFP2 + VPR_ZFP3). 발현 수준은 각 샘플에서 qRT-PCR에 의해 결정된 TBP 발현

수준으로 정규화되었다.

도 8은 정량적 실시간 증합효소 연쇄 반응 (qRT-PCR)에 의해 측정되는, 도 4-6에 기술된 SCN1A-결합 ZFP 및 Cas9+SCN1A 가이드 RNA가 HEK293T 세포에서 SCN1A 유전자 발현을 증가시키는 것을 표시하는 데이터를 나타낸다.

발명을 실시하기 위한 구체적인 내용

[0033] 본 개시내용의 양태는 세포 또는 대상체에서 표적 유전자의 발현을 조절하시 (예를 들어, 증가시키기) 위한 방법 및 조성물에 관한 것이고, 상기 표적 유전자는 반가불충분이다 (즉, 표적 유전자는 하나의 기능적 카피를 포함한다). 일부 구현예에서, 표적 유전자는 *SCN1A*이다.

[0034] 일부 구현예에서, 본 개시내용은 DNA 결합 도메인, 예를 들어 ZFP, 및 전사 조절자 도메인을 포함하는 융합 단백질을 제공한다. 일부 구현예에서 본 개시내용은 DNA 결합 도메인, 예를 들어 ZFP, 및 전사활성화제 도메인 (예를 들어, VPR 도메인)을 포함하는 융합 단백질을 제공한다. 일부 구현예에서, DNA 결합 단백질은 표적 유전자의 서열 또는 표적 유전자의 조절 영역에 결합한다. 일부 구현예에서, 조절 영역은 인핸서 서열, 프로모터 서열, 또는 억제자 서열이다. 일부 구현예에서, 프로모터 서열은 내부 프로모터 (예를 들어, 표적 유전자의 인트론에 위치함) 또는 외부 프로모터 (예를 들어, 표적 유전자의 전사 시작 부위의 업스트림에 위치함)일 수 있다. 일부 구현예에서, 본원에 기술된 융합 단백질의 DNA 결합 도메인은 표적 유전자 (예를 들어, *SCN1A*)의 프로모터 영역의 보존 서열에 결합하고, 그래서 전사활성화제 도메인이 유전자 발현을 증가시킨다.

[0035] 일부 양태에서, 본 개시내용은 세포 또는 대상체에서 표적 유전자 (예를 들어, *SCN1A*)의 발현을 증가시키는 방법에 관한 것이다. 일부 구현예에서, 표적 유전자는 세포 또는 대상체가 표적 유전자에 대해 반가불충분하게 하는 돌연변이를 함유한다. 따라서, 본 개시내용의 방법 및 조성물은 일부 구현예에서, 전압 의존성 나트륨 채널 알파 서브유닛 Nav1.1의 반가불충분성을 초래하는 *SCN1A* 유전자의 하나의 카피에서의 돌연변이로부터 일반적으로 유래되는, 표적 유전자 생성물의 반가불충분성과 관련된 질환 및 장애, 예를 들어 드라베 증후군을 치료하기 위해 사용될 수 있다.

[0036] **전사활성화제 융합 단백질**

[0037] 본 개시내용의 일부 양태는 DNA 결합 도메인 (DBD) 및 전사활성화제 도메인을 포함하는 융합 단백질에 관한 것이다. 본원에 사용된 바와 같이, 융합 단백질은 2개 이상의 개별 아미노산 서열에 의해 암호화되는 2개 이상의 연결된 폴리펩티드를 포함한다. 본원에 사용된 바와 같이, 키메라 단백질은 2개 이상의 연결된 유전자가 상이한 종으로부터 유래된 융합 단백질이다. 융합 단백질은 일반적으로 재조합으로 생산되고, 여기서 융합 단백질을 암호화하는 유전자는 2개 이상의 연결된 유전자의 발현 및 생성된 mRNA의 재조합 단백질로의 번역을 지원하는 시스템에 있다. 일부 구현예에서, 융합 단백질은 원핵 또는 진핵 세포에서 재조합적으로 생산된다. 융합 단백질은 다중 배열로 구성될 수 있다. 예를 들어, 하나의 단백질 (단백질 A)은 두 번째 단백질 (단백질 B)의 업스트림에 위치한다. 다른 융합 단백질 구성에서, 단백질 B는 단백질 A의 업스트림에 위치한다. 일부 구현예에서, DNA 결합 도메인을 암호화하는 핵산 서열은 전사활성화제 도메인을 암호화하는 핵산 서열의 업스트림에 위치하고, 전사활성화제에 연결된 DBD를 포함하는 융합 단백질을 생성한다. 일부 구현예에서, 전사활성화제 도메인을 암호화하는 핵산 서열은 DNA 결합 도메인을 암호화하는 핵산 서열의 업스트림에 위치하고, DNA 결합 도메인에 연결된 전사활성화제 도메인을 포함하는 융합 단백질을 생성한다. 일부 구현예에서, 융합 단백질은 DNA 결합 도메인의 업스트림에 위치한 전사활성화제 도메인을 포함한다. 일부 구현예에서, 융합 단백질은 전사활성화제 도메인의 업스트림에 위치한 DNA 결합 도메인을 포함한다.

[0038] 일부 구현예에서, 본 개시내용에 의해 기술된 융합 단백질은 DNA 결합 도메인을 포함한다. 본원에 사용된 바와 같이, "DNA 결합 도메인 (DBD)"은 이중 또는 단일 가닥 DNA (dsDNA 또는 ssDNA)를 인식하는 적어도 하나의 구조적 모티프를 포함하는 독립적으로 접힌 단백질을 지칭한다. 특정 DBD는 특정 서열 (인식 서열 또는 모티프)을 인식하는 반면, 다른 유형의 DBD는 DNA에 대한 일반적인 친화성을 갖는다. 일부 구현예에서, 본 개시내용에 의해 기술된 융합 단백질은 서열 특이적 DBD를 포함한다. 일부 구현예에서, DBD는 *SCN1A* 단백질 (예를 들어, Nav1.1)을 암호화하는 유전자 내 또는 이에 인접한 핵산 서열을 인식 (예를 들어, 특이적으로 결합)한다. DBD를 함유하는 단백질은 일반적으로 전사, 복제, 수리, 및 DNA 저장과 같은 세포 과정에 관여한다. 전사 인자의 DBD는 프로모터 영역 또는 인핸서 요소에서 특정 DNA 서열을 인식하여 유전자 발현을 촉진한다. 전사 인자 DBD는 표적 유전자의 발현을 조절하기 위해 유전 공학에서 융합 단백질로서 활용되고 DNA 결합 특이성 또는 DNA 결합 친화도를 변경하기 위해 돌연변이되어 발현 원하는 표적 유전자의 발현을 조절할 수 있다. DBD의 예시는 나선-회전-나선 모티프, 징크 핑거 모티프 (Cys2-His2 징크 핑거를 포함), 전사 활성화제-유사 효과기 (TALE), 날개

달린 나선 모티프, HMG-박스, dCas 단백질 (예를 들어, dCas9 또는 dCas12a), 호메오도메인 및 OB-폴드 도메인을 포함하지만 이에 제한되지 않는다.

[0039] 일부 구현예에서, 본 개시내용은 징크 핑거 DBD 융합 단백질에 관한 것이다. 본원에 사용된 바와 같이, "징크 핑거 단백질 (ZFP)"은 단백질 접힘을 안정화시키는 하나 이상의 징크 이온의 배위를 특징으로 하는 적어도 하나의 구조적 모티프를 함유하는 단백질을 지칭한다. 징크 핑거는 단백질에서 발견되는 가장 다양한 구조적 모티프 중 하나이고, 인간 유전자의 최대 3%가 징크 핑거를 암호화한다. 대부분의 ZFP는 DNA, RNA, 및 소단백질 유비퀴틴을 포함하는, 표적 분자와 탠덤 접촉을 만드는 다중 징크 핑거를 함유한다. "고전적" 징크 핑거 모티프는 2개의 시스테인 아미노산 및 2개의 히스티딘 아미노산 (C₂H₂)으로 구성되고 서열 특이적 방식으로 DNA에 결합한다. 전사 인자 IIIIA (TFIIIA)를 포함하는, 이러한 ZFP는 일반적으로 유전자 발현에 관여한다. DNA 결합 단백질의 다중 징크 핑거 모티프는 DNA 이중 나선에 결합하고 외부로 돌출한다. 비교적 작은 크기 (예를 들어, 각각의 핑거는 약 25-40개, 보통 27-35개 아미노산)로 인해, 징크 핑거 도메인 융합 단백질은 신규한 DNA 결합 특이성을 갖는 DBD를 생성하는데 활용된다. 이들 DBD는 표적 유전자의 전사 조절을 변경하기 위해 다른 융합된 도메인 (예를 들어, 전사 활성화 또는 억제 도메인 또는 후생적 변형 도메인)을 전달할 수 있다. 일부 구현예에서, 징크 핑거 단백질은 2 내지 8개의 핑거를 포함하고, 여기서 각각의 핑거는 27 내지 40개 아미노산 (예를 들어, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 또는 40개 아미노산)을 함유한다.

[0040] 일부 구현예에서, ZFP는 1, 2, 3, 4, 5, 6, 7, 또는 8개의 징크 핑거를 포함한다. 각각의 징크 핑거는 25-40, 25-30, 30-35, 35-40, 또는 40-45개의 아미노산을 포함할 수 있다. 일부 구현예에서, 징크 핑거는 27-35개의 아미노산을 포함한다. 일부 구현예에서, 징크 핑거는 27, 28, 29, 30, 31, 32, 33, 34, 또는 35개의 아미노산을 포함한다. 징크 핑거는 표적 서열, 대상체에서 반가불충분인, 예를 들어, 표적 유전자 또는 표적 유전자의 조절 영역을 특이적으로 인식하거나 이에 결합할 수 있다. 일부 구현예에서, 징크 핑거는 *SCN1A* 유전자, 예를 들어, 서열 번호 49에 제시된 바와 같은, 인간 *SCN1A*의 표적 서열에 결합한다. 일부 구현예에서, *SCN1A* 유전자의 표적 서열에 결합하는 징크 핑거는 서열 번호 63-80의 하나 이상의 아미노산 서열, 또는 이의 조합을 포함한다. 일부 구현예에서, 징크 핑거는 트리뉴클레오티드 서열을 포함하는 표적 서열을 특이적으로 인식하거나 이에 결합한다.

[0041] 일부 구현예에서, 징크 핑거는 표적 서열, 예를 들어, 트리뉴클레오티드 서열을 포함하는 표적 서열을 인식하거나 이에 결합하는 인식 나선을 포함한다. 일부 구현예에서, 인식 나선은 트리뉴클레오티드에 결합한다. 일부 구현예에서, 인식 나선은 4-10개의 아미노산을 포함한다. 일부 구현예에서, 인식 나선은 4, 6, 7, 8, 9, 또는 10개의 아미노산을 포함한다. 일부 구현예에서, 인식 나선은 *SCN1A* 유전자의 트리뉴클레오티드 서열에 결합한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 인식 서열은 서열 번호 17-22, 29-34, 또는 41-46 중 어느 하나의 아미노산 서열을 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 인식 서열은 서열 번호 11-16, 23-28, 또는 35-40 중 어느 하나에 의해 암호화된다. 일부 구현예에서, 징크 핑거는 서열 번호 17-22, 29-34, 또는 41-46 중 어느 하나의 아미노산 서열을 포함하는 인식 나선과 동일한 뉴클레오티드 서열에 결합한다.

[0042] 일부 구현예에서, 징크 핑거는 이의 C-말단에서 상기 징크 핑거를 추가 징크 핑거에 연결하거나 접촉시키는 역할을 할 수 있는 링커 서열을 포함한다. 일부 구현예에서, 링커 서열은 예를 들어, TGEKP의 아미노산 서열을 포함하는, 표준 링커일 수 있다 (서열 번호 120). 일부 구현예에서, 링커 서열은 예를 들어, TGSQKP의 아미노산 서열을 포함하는, 비표준 링커일 수 있다 (서열 번호 121). 일부 구현예에서, 링커 서열은 2-10개의 아미노산, 예를 들어, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개의 아미노산일 수 있다.

[0043] 일부 구현예에서, 표적 유전자, 예를 들어, *SCN1A* 유전자에 결합하는 ZFP는 6개의 징크 핑거를 포함하고, 이들 각각은 표적 유전자, 예를 들어, *SCN1A* 유전자의 상이한 트리뉴클레오티드 서열을 인식하거나 이에 결합한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 57의 아미노산 서열을 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 63, 64, 65, 66, 67, 및/또는 68의 아미노산 서열을 포함하는 징크 핑거를 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 17, 18, 19, 20, 21, 및/또는 22의 아미노산 서열을 포함하는 인식 나선을 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 59의 아미노산 서열을 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 69, 70, 71, 72, 73, 및/또는 74의 아미노산 서열을 포함하는 징크 핑거를 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 29, 30, 31, 32, 33, 및/또는 34의 아미노산 서열을 포함하는 인식 나선을 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 61의 아미노산 서열을 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 75, 76, 77, 78, 79, 및/또는 80의 아미노산 서열을 포함하는 징크

핑거를 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 서열 번호 41, 42, 43, 44, 45, 및/또는 46의 아미노산 서열을 포함하는 인식 나선을 포함한다. 일부 구현예에서, *SCN1A* 유전자에 결합하는 ZFP는 하기 나 타낸, 서열 번호 57, 59, 또는 61에 대해 적어도 60%, 적어도 65%, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 또는 적어도 99% 서열 동일성을 포함한다.

[0044] 서열 번호 57 (ZFP1 단백질의 아미노산 서열)

[0045] RPFQCRICMRNFSQRGNLVRHIRTHTGKPFACDICGKKFALSFNLTRHTKIHTGSQKPFQCRICMRNFSRSDNLTRHIRTHTGKPFACDICGKKFADRSHLARHTKIHTGSQKPFQCRICMRNFSQKAHLTAHIRTHTGKPFACDICGRKFARSDNLTRHTKIHLRQKD

[0046] 서열 번호 59 (ZFP2 단백질의 아미노산 서열)

[0047] RPFQCRICMRNFSRSDNLTRHIRTHTGKPFACDICGKKFADKRTLIRHTKIHTGSQKPFQCRICMRNFSQRGNLVRHIRTHTGKPFACDICGKKFALSFNLTRHTKIHTGSQKPFQCRICMRNFSRSDNLTRHIRTHTGKPFACDICGRKFADRSHLARHTKIHLRQKD

[0048] 서열 번호 61 (ZFP3 단백질의 아미노산 서열)

[0049] RPFQCRICMRNFSRSDNLTRHIRTHTGKPFACDICGKKFARSDNLTRHTKIHTGSQKPFQCRICMRNFSQSGDLTRHIRTHTGKPFACDICGKKFAVRQTLKQHTKIHTGSQKPFQCRICMRNFSRSDNLTRHIRTHTGKPFACDICGRKFARSDNLTRHTKIHLRQKD

[0050] 일부 구현예에서, DBD는 전사 활성화제-유사 효과기 단백질 (TALE)이다. TALE은 표적 서열, 예를 들어, 표적 유전자 또는 표적 유전자의 조절 영역을 특이적으로 인식하거나 이에 결합할 수 있다. 일부 구현예에서, 대상체는 표적 유전자에 대해 반가불충분하다. 일부 구현예에서, TALE은 *SCN1A* 유전자, 예를 들어, 서열 번호 49에서 제공된 인간 *SCN1A*의 표적 서열에 결합한다. TALE 단백질은 세균에 의해 분비되고 세균 감염에서 도움이 되는 식물 유전자의 발현을 활성화하기 위해 숙주 식물에서 프로모터 서열에 결합한다. 일반적으로, TALE 단백질은 ~30-35 아미노산 반복의 가변 수로 이루어진 중앙 반복 도메인을 통해 표적 서열을 인식하는 새로운 DNA 서열에 결합하도록 조작되고, 여기서 각각의 반복은 표적 서열 내에서 단일 염기 쌍을 인식한다. 이러한 반복의 배열은 일반적으로 DNA 서열을 인식하는데 필요하다.

[0051] 일부 구현예에서, DBD는 호메오도메인이다. 호메오도메인은 표적 서열, 예를 들어, 표적 유전자 또는 표적 유전자의 조절 영역을 특이적으로 인식하거나 이에 결합할 수 있다. 일부 구현예에서, 대상체는 표적 유전자에 대해 반가불충분하다. 일부 구현예에서, 호메오도메인은 *SCN1A* 유전자, 예를 들어, 서열 번호 49에 제공된 인간 *SCN1A*의 표적 서열에 결합한다. 호메오도메인은 표적 서열을 인식하는데 필요한 3개의 알파 나선 및 N-말단 암을 함유하는 단백질이다. 호메오도메인은 일반적으로 작은 DNA 서열 (~4 내지 8개 염기 쌍)을 인식하지만, 이들 도메인은 더 긴 연장된 서열 (12 내지 24개 염기 쌍)을 인식하기 위해 다른 DNA-결합 도메인 (또는 다른 호메오도메인 또는 징크 핑거 단백질)과 함께 융합될 수 있다. 따라서, 호메오도메인은 인간 게놈 내에 고유한 서열을 인식하는 DBD의 구성성분일 수 있다.

[0052] 일부 구현예에서, 적어도 하나의 DNA 결합 도메인은 촉매적 불활성 CRISPR 관련 단백질 (Cas 단백질)이다. 촉매적 불활성 Cas 단백질 (dCas 또는 "죽은 Cas 단백질"로도 알려짐)은 뉴클레아제 활성 (예를 들어, 엔도뉴클레아제 활성)이 감소되거나 모든 뉴클레아제 활성 (예를 들어, 엔도뉴클레아제 활성)이 결여되도록 변경되거나 돌연변이된 Cas 단백질이다. 일부 구현예에서, 촉매적 불활성 Cas 단백질은 dCas9 또는 dCas12 단백질이다. 일부 구현예에서, DBD는 dCas9 또는 dCas12a와 같은 dCas 단백질 ('죽은 Cas'로도 알려짐)이다. dCas 단백질은 촉매적으로 비활성화, 즉, 뉴클레오티드 절단이 불가능하도록 돌연변이된 CRISPR 관련 단백질 (Cas, 예를 들어, Cas9 또는 Cas12a)의 돌연변이 변이체이다. dCas는 표적 서열, 예를 들어, 표적 유전자 또는 표적 유전자의 조절 영역을 특이적으로 인식하거나 이에 결합할 수 있다. dCas 단백질 및 가이드 핵산 (예를 들어, gRNA)을 포함하는 복합체는 가이드 핵산에 상보적인 특정 뉴클레오티드 서열 또는 유전자를 표적으로 하고/거나 이에 결합할 수 있다. 일부 구현예에서, 대상체는 표적 유전자에 대해 반가불충분하다. 일부 구현예에서, dCas는 *SCN1A* 유전자, 예를 들어, 서열 번호 49에 제공된 인간 *SCN1A*의 표적 서열에 결합한다. 그러나, dCas 단백질은 상기 표적 DNA 서열에 상보적이거나 부분적으로 상보적인 가이드 핵산 (예를 들어, 가이드 RNA, gRNA, 또는 sgRNA)에 결합할 때 표적 DNA 서열을 인식하고 이에 결합하는 이들의 능력을 보유한다. 일부 구현예에서, dCas (예를 들어, dCas9) 단백질을 *SCN1A*에 대해 표적화하기 위한 가이드 핵산은 서열 번호 85, 86, 89, 90, 93, 또는 94 중 어느 하나를 갖는 스페이서 서열을 포함한다. 일부 구현예에서, dCas (예를 들어, dCas9) 단백질을 *SCN1A*에 표적화하기 위한 가이드 핵산은 서열 번호 85, 86, 89, 90, 93, 또는 94 중 어느 하나의 적어도 15개 (예를 들어, 적어도 16, 17, 18, 19, 또는 20개) 연속 뉴클레오티드를 갖는 스페이서 서열을 포함한다. 일부 구현예에서, dCas (예를 들어, dCas9) 단백질을 *SCN1A*에 표적화하기 위한 가이드 핵산은 서열 번호 83, 84, 87, 88, 91, 또

는 92 중 어느 하나를 포함한다. 일부 구현예에서, dCas (예를 들어, dCas9) 단백질을 SCN1A에 표적화하기 위한 가이드 핵산은 서열 번호 83-94 중 어느 하나를 포함하거나 이로 이루어진다. 따라서, dCas 엔도뉴클레아제는 인간 게놈 내의 고유한 서열을 인식하는 DBD의 구성성분일 수 있다. 일부 구현예에서, 융합 단백질은 dCas9 단백질 및 전사활성화 도메인 (예를 들어, VPR 도메인)을 포함한다.

[0053] 본 개시내용은 일부 양태에서, 전압 의존성 나트륨 채널 (예를 들어, Na_v1.1)을 암호화하는 유전자에 결합하는 DNA 결합 도메인에 관한 것이다. 일부 구현예에서, 전압 의존성 나트륨 채널을 암호화하는 유전자는 *SCN1A* 유전자이고, 서열 번호 49에 제시된 서열을 포함한다. 일부 구현예에서, DNA 결합 도메인은 표적 유전자의 비번역 영역, 예를 들어 3'-비번역 영역 (3'UTR) 또는 5'-비번역 영역 (5'UTR)에 결합한다. 일부 구현예에서, 비번역 영역은 조절 서열, 예를 들어 인핸서, 프로모터, 인트론, 또는 억제자 서열을 포함한다. 일부 구현예에서, DNA 결합 도메인은 서열 번호 57-62에 제시된 서열을 포함하는 징크 핑거 단백질이다. 일부 구현예에서, DNA 결합 도메인은 서열 번호 5-7 중 어느 하나에 제시된 핵산 서열에 결합한다.

[0054] 이식유전자에 의해 암호화된 DNA 결합 도메인의 수는 다양할 수 있다. 일부 구현예에서, 이식유전자는 하나의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 2개의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 3개의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 4개의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 5개의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 6개의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 7개의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 8개의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 9개의 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 10개 DNA 결합 도메인을 암호화한다. 일부 구현예에서, 이식유전자는 10개 초과 (예를 들어, 20, 30, 50, 100개 등) DNA 결합 도메인을 암호화한다. DNA 결합 도메인은 동일한 DNA 결합 도메인 (예를 들어, 동일한 DBD의 다중 카피), 상이한 DNA 결합 도메인 (예를 들어, 각각의 DBD는 고유한 서열에 결합함), 또는 이의 조합일 수 있다.

[0055] 본 개시내용은 일부 양태에서, 전사활성화제 도메인을 포함하는 융합 단백질에 관한 것이다. 본원에 사용된 바와 같이, "전사활성화 도메인"은 유전자 발현을 조절하는 다른 단백질, 예를 들어 전사 공동 조절자에 대한 결합 부위를 함유하는 전사 인자에서 스캐폴드 도메인을 지칭한다. 일부 구현예에서, 전사활성화 도메인 (전사 활성화 도메인으로도 알려짐)은 전사 인자와의 접촉을 통해 직접적으로 또는 보조활성화제 단백질을 통해 간접적으로 프로모터 또는 인핸서로부터 전사를 활성화하기 위해 DBD와 함께 작용한다. 전사활성화 도메인 (TAD)은 일반적으로 이들의 아미노산 조성물에 따라 명명되고, 상기 아미노산은 활성화에 필수적이거나 TAD에서 가장 풍부하다. TAD는 표적 유전자의 발현을 조절하기 위해 유전 공학에서 융합 단백질로서 활용되고 전사 활성화의 수준을 변경하여 표적 유전자의 발현을 변경하도록 돌연변이될 수 있다. 전사활성화 도메인의 예는 GAL4, HAP1, VP16, P65, RTA, 및 GCN4를 포함하지만 이에 제한되지 않는다.

[0056] 일부 구현예에서, 전사활성화제 도메인은 VP64 도메인을 포함한다. VP64는 VP16 단백질의 4개의 탠덤 카피로 구성된 산성 TAD이고, 이는 단순 헤르페스 바이러스에 의해 자연적으로 발현된다. 유전자의 프로모터에 또는 근처에 결합하는 DBD에 융합될 때, VP64는 강한 전사 활성화제로 작용하므로 표적 유전자 (예를 들어, *SCN1A*)의 발현을 조절하기 위해 활용될 수 있다. VP64 도메인은 일반적으로 단순 헤르페스 단백질 VP16의 최소 활성화 도메인의 사랑채 반복으로 이루어진다. 일부 구현예에서, VP64 도메인은 VP16에서 아미노산 잔기 437-448개의 반복부를 포함한다. 일부 구현예에서, VP16 단백질은 인간 헤르페스 바이러스 2 UL48 유전자에 의해 암호화되고, 이는 NCBI 참조 서열 수탁 번호: NC_001798.2에 제시된 서열을 포함한다. 일부 구현예에서, VP16 유전자는 NCBI 참조 서열 수탁 번호: YP_009137200.1에 제시된 핵산 서열로 암호화된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 뉴클레오티드 서열을 포함한다. 일부 구현예에서, VP16 단백질은 NCBI 참조 서열 수탁 번호 Q69113-1에 제시된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, VP16 유전자는 서열 번호 51에 제시된 핵산 서열로 암호화된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 뉴클레오티드 서열을 포함한다. 일부 구현예에서, VP16 단백질은 서열 번호 52에 제시된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다.

[0057] 일부 구현예에서, 전사활성화제 도메인은 P65 활성화 도메인을 포함한다. P65는 이의 C-말단에서 2개의 인접한 산성 TAD를 함유하는 NF- κ β 전사 인자의 서브유닛이다. 유전자의 프로모터에 또는 근처에 결합하는 DBD에 융합될 때, p65 단백질은 강한 전사 활성화제로 작용하므로 예를 들어 Urlinger, et al. "The p65 domain from

NF-kappaB is an efficient human activator in the tetracycline-regulatable gene expression system," *Gene*, 2000에 기술된 바와 같이, 표적 유전자의 발현을 조절하기 위해 활용될 수 있다. 일부 구현예에서, p65 단백질은 인간 *RELA* 유전자에 의해 암호화되고, 이는 NCBI 참조 서열 수탁 번호: NM_001145138.1, NM_001243984.1, NM_001243985.1, 또는 NM_021975.3에 제시된 서열을 포함한다. 일부 구현예에서, *RELA* 유전자는 NCBI 참조 서열 번호 NM_001145138.1, NM_001243984.1, NM_001243985.1, 또는 NM_021975.3 중 하나에 제시된 핵산 서열로 암호화된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 뉴클레오티드 서열을 포함한다. 일부 구현예에서, p65 단백질은 NP_001138610.1, NP_001230913.1, NP_001230914.1, 및 NP_068110.3에 제시된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, *RELA* 유전자는 서열 번호 53에 제시된 핵산 서열로 암호화된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 뉴클레오티드 서열을 포함한다. 일부 구현예에서, p65 단백질은 서열 번호 54에 제시된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다.

[0058] 일부 구현예에서, 전사활성화제 도메인은 RTA 도메인을 포함한다. RTA는 여러 바이러스 유전자의 발현을 촉진하기 위해 인헨서 영역에 결합하는 강력한 전사활성화 도메인인 엡스테인 바 바이러스에서 파생된 소수성 TAD이다. 유전자의 프로모터에 또는 근처에 결합하는 DBD에 융합될 때, RTA 단백질은 강한 전사 활성화제로서 작용하므로 예를 들어 Miyazawa, et al., "IL-10 promoter transactivation by the viral K-RTA protein involves the host-cell transcription factors, specificity proteins 1 and 3," *Journal of Biological Chemistry*, 2018에 기술된 바와 같이, 표적 유전자의 발현을 조절하기 위해 활용될 수 있다. 일부 구현예에서, RTA 단백질은 엡스테인 바 바이러스 BRLF1 유전자에 의해 암호화되고, 이는 NCBI 참조 서열 수탁 번호: YP_041674.1에 제시된 서열을 포함한다. 일부 구현예에서, BRLF1 유전자는 NCBI 참조 서열 번호 YP_041674.1 중 하나에 제시된 핵산 서열로 암호화된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 뉴클레오티드 서열을 포함한다. 일부 구현예에서, RTA 단백질은 YP_041674.1에 제시된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, BRLF1 유전자는 서열 번호 55에 제시된 핵산 서열로 암호화된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 뉴클레오티드 서열을 포함한다. 일부 구현예에서, RTA 단백질은 서열 번호 56에 제시된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다.

[0059] 본 개시내용은 부분적으로, 하이브리드 전사활성화제 도메인을 포함하는 융합 단백질을 기반으로 한다. 본원에 사용된 바와 같이, "하이브리드 전사활성화제 도메인"은 하나 초과 of 전사 활성화 단백질 또는 이의 부분 (예를 들어, 2, 3, 4, 5, 또는 이상의 전사 활성화 단백질, 또는 이의 부분) 포함하는 융합 단백질을 지칭한다. 하이브리드 전사활성화 도메인은 표적 유전자의 발현을 증가시키기 위해 유전 공학에서 활용된다. 본 개시의 일부 구현예에서, Chavez, et al. "Highly efficient Cas9-mediated transcriptional programming", *Nat Methods*, 2015에 기술된, VP64-P65-RTA (VPR)에 대한 뉴클레오티드 서열 (서열 번호 47)을 포함하는 3부 하이브리드 전사 활성화 도메인은 표적 유전자 (예를 들어 *SCN1A*) 발현을 증가시키기 위해 활용된다.

[0060] 일부 구현예에서, 본원에 기술된 융합 단백질은 DBD (예를 들어, ZFP) 및 전사 억제자 단백질을 포함할 수 있다. 일부 양태에서, 본 개시내용은 전사 억제자 도메인을 포함하는 융합 단백질에 관한 것이다. 본원에 사용된 바와 같이, "전사 억제자" 단백질은 일반적으로 표적 유전자의 발현을 하향 조절하는 폴리펩티드를 지칭한다. 전사 억제자의 예시는 KRAB, SMRT/TRAC-2, 및 NCoR/RIP-13을 포함하지만 이에 제한되지 않는다. 일부 구현예에서, 이러한 전사 억제자 융합 단백질은 표적 유전자 (예를 들어, 기능 획득 질환에서 과발현된 유전자)의 발현 수준을 감소시키는데 유용하다.

[0061] 단리된 핵산

[0062] 단리된 핵산 서열은 DNA 또는 RNA 서열을 지칭한다. 일부 구현예에서, 본 개시내용의 단백질 및 핵산은 단리된다. 본원에 사용된 바와 같이, 용어 "단리된"은 인공적으로 생성된 것을 의미한다. 핵산과 관련하여 본원에 사용된, 용어 "단리된"은 하기 단계를 의미한다: (i) 예를 들어, 증합효소 연쇄 반응 (PCR)에 의해 시험관 내 증폭되는 단계; (ii) 클로닝에 의해 재조합적으로 생산되는 단계; (iii) 절단 및 겔 분리에 의해 정제되는 단계; 또는 (iv) 예를 들어, 화학적 합성에 의해 합성되는 단계. 단리된 핵산은 당 업계에 잘 알려진 재조합 DNA 기술에 의해 쉽게 조작할 수 있는 것이다. 따라서, 5' 및 3' 제한 부위가 알려져 있거나 증합효소 연쇄 반응 (PCR)

프라이머 서열이 개시된 벡터에 함유된 뉴클레오티드 서열은 단리된 것으로 간주되지만 자연 상태로 자연 숙주에 존재하는 핵산 서열은 그렇지 않다. 단리된 핵산은 실질적으로 정제될 수 있지만, 반드시 그럴 필요는 없다. 예를 들어, 클로닝 또는 발현 벡터 내에서 단리된 핵산은 그것이 존재하는 세포 내 물질의 아주 작은 비율만을 포함할 수 있다는 점에서 순수하지 않다. 그러나, 당업자에게 공지된 표준 기술에 의해 용이하게 조작될 수 있기 때문에 용어가 본원에서 사용된 바와 같이 이러한 핵산은 단리된다. 단백질 또는 펩티드와 관련하여 본원에서 사용된, 용어 "단리된"은 그의 자연 환경에서 단리되거나 (예를 들어, 화학적 합성, 재조합 DNA 기술 등에 의해) 인공적으로 생성된 단백질을 지칭한다.

[0063] 일부 양태에서, 본 개시내용은 하나 이상의 ZFP-전사활성화 도메인 융합 단백질을 발현하도록 구성된 단리된 핵산 (예를 들어, 발현 작제물, 예를 들어 rAAV 벡터)에 관한 것이다. 일부 구현예에서, 융합 단백질은 1 내지 10개 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 DBD 및/또는 1 내지 10개 (예를 들어, 1, 2, 3, 4, 5, 6, 7, 8, 9, 또는 10개)의 전사활성화제 도메인을 포함한다. 일부 구현예에서, 융합 단백질은 10개 초과 DBS 및/또는 10개 초과 전사활성화제 도메인을 포함한다.

[0064] 본 개시의 일부 양태에서, DNA 결합 도메인은 링커를 통해 간접적으로 전사 조절자 도메인에 융합된다. 본원에서 사용된 "링커"는 단일 이식유전자 내에서 2개의 별개의 폴리펩티드를 구조적으로 연결하는 폴리펩티드의 스트레치이다. 일부 구현예에서, 링커는 별개의 폴리펩티드의 이동을 허용하도록 유연하다. 일부 구현예에서, 가요성 링커는 글리신 잔기를 포함한다. 일부 구현예에서, 가요성 링커는 글리신 및 세린 잔기의 혼합물을 포함한다. 일부 구현예에서, 링커는 절단 가능하고, 폴리펩티드가 분리되도록 허용한다. 일부 구현예에서, 절단 가능한 링커는 프로테아제에 의해 절단된다. 일부 구현예에서, 프로테아제는 트립신 또는 인자 X이다.

[0065] 일부 구현예에서 링커는 5 내지 30개 아미노산 (예를 들어, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개 아미노산)을 포함한다. 일부 구현예에서, 링커는 3 내지 30개 아미노산 (예를 들어, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 또는 30개 아미노산)을 포함한다. 일부 구현예에서, 링커는 3 내지 20개 아미노산 (예를 들어, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 또는 20개 아미노산)을 포함한다.

[0066] 본 개시내용은 부분적으로, 전압 의존성 나트륨 이온 채널 서브유닛 단백질 (SCN 단백질로서도 지칭됨)을 암호화하는 유전자, 예를 들어 *SCN1A*의 발현을 증가시키기 위해 조작된 융합 단백질을 기반으로 한다. 본원에서 사용된 바와 같이, "SCN 단백질"은 흥분성 막의 전압 의존성 나트륨 이온 투과성을 매개하여, 나트륨 이온이 막을 통과할 수 있도록 하는 나트륨 이온 채널 단백질을 지칭한다. 인간의 SCN 단백질의 예는 SCN1A, SCN3A, SCN5A, SCN10A, 및 SCN11A를 포함하지만 이에 제한되지 않는다. 일부 구현예에서, SCN 단백질은 SCN1A (Nav1.1로도 지칭됨)이고, 이는 유형 1 α_1 이온 채널 서브유닛을 암호화한다. 일부 구현예에서, SCN 단백질은 SCN1B 단백질이고, 이는 유형 1 β_1 이온 채널 서브유닛 또는 SCN1C 단백질을 암호화한다. 일부 구현예에서, SCN 단백질은 SCN1A, SCN1B, 및/또는 SCN1C 단백질의 조합이다. 본원에 개시된 바와 같이, SCN 단백질은 SCN 단백질의 부분 또는 단편일 수 있다. 일부 구현예에서, 본원에 개시된 SCN 단백질은 SCN 단백질의 변이체, 예를 들어 점 돌연변이 또는 절단 돌연변이이다.

[0067] 인간에서, SCN1A는 SCN1A 유전자 (Gene ID: 6323, 인간)에 의해 암호화되고, 침팬지, 붉은털원숭이, 개, 소, 마우스, 랫트, 및 닭에서 보존되어 있다. 인간의 SCN1A 유전자는 뇌, 폐 및 고환에서 주로 발현된다. 일부 구현예에서, SCN1A 단백질은 5개의 구조적 반복부 (I, II, III, IV, Q)를 포함한다.

[0068] 일부 구현예에서, SCN1A 단백질은 인간 SCN1A 유전자에 의해 암호화되고, 이는 NCBI 참조 서열 번호 NM_001165963.2, NM_00165964.2, NM_001202435.2, NM_001353948.1, NM_001353949.1, NM_001353950.1, NM_001353951.1, NM_001353952.1, NM_001353954.1, NM_00353955.1, NM_001353957.1, NM_001353958.1, NM_001353960.1, NM_001353961.1, 또는 NM_006920.5에 제시된 서열을 포함한다. 일부 구현예에서, SCN1A 단백질은 마우스 SCN1A 유전자에 의해 암호화되고, 이는 NCBI 참조 서열 번호 NM_001313997.1 또는 NM_018733.2에 제시된 서열을 포함한다. 일부 구현예에서, SCN1A 단백질은 NCBI 참조 서열 번호 NG_011906.1, NM_001313997.1 또는 NM_018733.2에 제시된 핵산 서열로 암호화된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, SCN1A 유전자는 서열 번호 50에 제시된 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, 인간 SCN1A 단백질은 NCBI 참조 서열 번호 NP_001159435.1, NP_0011159436.1, NP_001189364.1, NP_001340877.1, NP_001340878.1, NP_001340879.1,

NP_001340880.1, NP_001340881.1, NP_001340883.1, NP_001340884.1, NP_001340886.1, NP_001340887.1, NP_001340889.1, NP_001340890.1, NP_00851.3에 제시된 서열을 포함한다. 일부 구현예에서, SCN1A 단백질은 NCBI 참조 서열 번호 NG_011906.1, NM_001313997.1 또는 NM_018733.2에 제시된 핵산 서열로 암호화된 아미노산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다. 일부 구현예에서, 마우스 SCN1A 단백질은 NCBI 참조 서열 번호 NP_001300926.1 또는 NP_061203.2에 제시된 서열을 포함한다. 일부 구현예에서, 인간 SCN1A 단백질은 서열 번호 49에 제시된 핵산 서열에 대해 99% 동일한, 95% 동일한, 90% 동일한, 80% 동일한, 70% 동일한, 60% 동일한, 또는 50% 동일한 아미노산 서열을 포함한다.

[0069] 본 개시내용의 단리된 핵산은 재조합 아데노 관련 바이러스 (AAV) 벡터 (rAAV 벡터)일 수 있다. 일부 구현예에서, 본 개시내용에 의해 기술된 단리된 핵산은 제1 아데노 관련 바이러스 (AAV) 역위 말단 반복 (ITR), 또는 이의 변이체를 포함하는 영역 (예를 들어, 제1 영역)을 포함한다. 단리된 핵산 (예를 들어, 재조합 AAV 벡터)은 캡시드 단백질로 패키징되고 대상체에 투여 및/또는 선택된 표적 세포로 전달될 수 있다. "재조합 AAV (rAAV) 벡터"는 일반적으로 최소한, 이식유전자 및 이의 조절 서열, 및 5' 및 3' AAV 역위 말단 반복 (ITR)으로 구성된다. 이식유전자는 본 개시내용의 다른 곳에 기재된 바와 같이, 예를 들어, 단백질 및/또는 발현 제어 서열 (예를 들어, 폴리-A 꼬리)을 암호화하는 영역을 포함할 수 있다.

[0070] 일반적으로, ITR 서열은 약 145 bp 길이이다. 바람직하게는, 실질적으로 ITR을 암호화하는 전체 서열이 분자에서 사용되지만, 어느 정도 이들 서열의 작은 변형은 허용된다. 이들 ITR 서열을 변형시키는 능력은 당 업계의 기술 범위 내에 있다. (예를 들어, Sambrook et al., "Molecular Cloning. A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory, New York (1989); 및 K. Fisher et al., J Virol., 70:520-532 (1996)과 같은 교과서 참조). 본 개시내용에 사용된 이러한 분자의 예는 이식유전자를 함유하는 "시스-작용" 플라스미드이고, 여기서 선택된 이식유전자 서열 및 관련된 조절 요소는 5' 및 3' AAV ITR 서열 측면에 있다. AAV ITR 서열은 현재 확인된 포유동물 AAV 유형을 포함하는, 임의의 공지된 AAV로부터 얻을 수 있다. 일부 구현예에서, 단리된 핵산은 제2 AAV ITR을 포함하는 영역 (예를 들어, 제2 영역, 제3 영역, 제4 영역 등)을 추가로 포함한다.

[0071] 재조합 AAV 벡터에 대해 위에서 확인된 주요 요소에 더하여, 벡터는 또한 벡터로 감염되거나 본 개시내용에 의해 생성된 바이러스에 감염된 세포에서 그의 전사, 번역 및/또는 발현을 허용하는 방식으로 이식유전자의 요소와 작동 가능하게 연결된 통상적인 제어 요소를 포함한다. 본원에 사용된 바와 같이, "작동 가능하게 연결된" 서열은 관심 유전자와 인접한 발현 제어 서열 및 관심 유전자를 제어하기 위해 트랜스로 또는 멀리서 작용하는 발현 제어 서열을 모두 포함한다. 발현 제어 서열은 적절한 전사 개시, 종결, 프로모터 및 인핸서 서열; 스플라이싱 및 폴리아데닐화 (폴리아) 신호와 같은 효율적인 RNA 처리 신호; 세포질 mRNA를 안정화시키는 서열; 번역 효율을 향상시키는 서열 (예를 들어, 코작 공통 서열); 단백질 안정성을 향상시키는 서열; 및 원하는 경우, 암호화된 생성물의 분비를 향상시키는 서열을 포함한다. 천연, 구성적, 유도성 및/또는 조직-특이적인 프로모터를 포함하는, 다수의 발현 제어 서열이 당 업계에 공지되어 있고 활용될 수 있다.

[0072] 본원에 사용된 바와 같이, 핵산 서열 (예를 들어, 코딩 서열) 및 조절 서열은 핵산 서열의 발현 또는 전사가 조절 서열의 영향 또는 제어하에 놓이도록 하는 방식으로 공유 연결될 때 작동 가능하게 연결된 것으로 언급된다. 핵산 서열이 기능적 단백질로 번역되는 것이 바람직한 경우, 5' 조절 서열에서 프로모터의 유도가 코딩 서열의 전사를 초래하고 2개의 DNA 서열 사이 연결의 특성이 (1) 프레임 이동 돌연변이의 도입을 일으키지 않거나, (2) 코딩 서열의 전사를 지시하는 프로모터 영역의 능력을 방해하거나, (3) 단백질로 번역되는 상응하는 RNA 전사체의 능력을 방해하지 않으면, 2개의 DNA 서열은 작동 가능하게 연결된 것으로 언급된다. 따라서, 프로모터 영역이 생성된 전사체가 원하는 단백질 또는 폴리펩타이드로 번역될 수 있도록 DNA 서열의 전사 효과를 낼 수 있다면 프로모터 영역은 핵산 서열에 작동 가능하게 연결될 것이다. 유사하게 2개 이상의 코딩 영역은 공통 프로모터로부터 이들의 전사가 프레임 내에서 번역된 2개 이상의 단백질의 발현을 초래하는 방식으로 연결될 때 작동 가능하게 연결된다. 일부 구현예에서, 작동 가능하게 연결된 코딩 서열은 융합 단백질을 생성한다.

[0073] 이식유전자를 포함 (예를 들어, 융합 단백질 등을 포함)하는 영역은 융합 단백질의 발현을 가능하게 할 단리된 핵산의 임의의 적합한 위치에 위치할 수 있다.

[0074] 이식유전자가 하나 초과인 폴리펩티드를 암호화하는 경우, 각각의 폴리펩티드는 이식유전자 내의 임의의 적합한 위치에 위치할 수 있음을 이해해야 한다. 예를 들어, 제1 폴리펩티드를 암호화하는 핵산은 이식유전자의 인트론에 위치할 수 있고 제2 폴리펩티드를 암호화하는 핵산 서열은 다른 비번역 영역 (예를 들어, 단백질 코딩

서열의 마지막 코돈과 이식유전자의 폴리-A 신호의 첫 번째 염기 사이)에 위치할 수 있다.

- [0075] "프로모터"는 유전자의 특정 전사를 개시하는 데 필요한 세포의 합성 기계, 또는 도입된 합성 기계에 의해 인식되는 DNA 서열을 지칭한다. 구절 "작동적으로 연결된," "작동적으로 위치된," "제어하에" 또는 "전사 제어하에"는 프로모터가 RNA 중합효소 개시 및 유전자 발현을 제어하기 위해 핵산과 관련하여 올바른 위치 및 방향에 있음을 의미한다.
- [0076] 핵산 암호화 단백질의 경우, 폴리아네닐화 서열은 일반적으로 이식유전자 서열 다음에 3' AAV ITR 서열 앞에 삽입된다. 본 개시내용에서 유용한 rAAV 작제물은 또한 바람직하게는 프로모터/인핸서 서열과 이식유전자 사이에 위치하는 인트론을 함유할 수 있다. 하나의 가능한 인트론 서열은 SV-40으로부터 유래되고, SV-40 T 인트론 서열로 지칭된다. 사용될 수 있는 또 다른 벡터 요소는 내부 리보솜 진입 부위 (IRES)이다. IRES 서열은 단일 유전자 전사체로부터 하나 초과 폴리펩티드를 생성하는데 사용된다. IRES 서열은 하나 초과 폴리펩티드 사슬을 함유하는 단백질을 생성하는데 사용될 수 있다. 이들 및 다른 공통 벡터 요소의 선택은 통상적이고 많은 이러한 서열이 이용 가능하다 [예를 들어, Sambrook et al., 및 예를 들어, 페이지 3.18 3.26 및 16.17 16.27에 인용된 참고 문헌 및 Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989 참조]. 일부 구현예에서, 구역적 바이러스 2A 서열이 폴리단백질에 포함되고; 이것은 폴리단백질의 절단을 매개하는 것으로 나타난 작은 펩티드 (대략 18개 아미노산 길이)이다 (Ryan, M D et al., EMBO, 1994; 4: 928-933; Mattion, N M et al., J Virology, November 1996; p. 8124-8127; Furler, S et al., Gene Therapy, 2001; 8: 864-873; 및 Halpin, C et al., The Plant Journal, 1999; 4: 453-459). 2A 서열의 절단 활성은 이전에 플라스미드 및 유전자 치료 벡터 (AAV 및 레트로바이러스)를 포함하는 인공 시스템에서 입증되었다 (Ryan, M D et al., EMBO, 1994; 4: 928-933; Mattion, N M et al., J Virology, November 1996; p. 8124-8127; Furler, S et al., Gene Therapy, 2001; 8: 864-873; 및 Halpin, C et al., The Plant Journal, 1999; 4: 453-459; de Felipe, P et al., Gene Therapy, 1999; 6: 198-208; de Felipe, P et al., Human Gene Therapy, 2000; 11: 1921-1931.; 및 Klump, H et al., Gene Therapy, 2001; 8: 811-817).
- [0077] 구성적 프로모터의 예시는 레트로바이러스 라우스 육종 바이러스 (RSV) LTR 프로모터 (선택적으로 RSV 인핸서를 가짐), 시토메갈로바이러스 (CMV) 프로모터 (선택적으로 CMV 인핸서를 가짐) [예를 들어, Boshart et al., Cell, 41:521-530 (1985) 참조], SV40 프로모터, 디히드로폴레이트 환원효소 프로모터, β-액틴 프로모터, 포스 포글리세롤 키나제 (PGK) 프로모터, 및 EF1α 프로모터 [Invitrogen]를 제한없이, 포함한다. 일부 구현예에서, 프로모터는 P2 프로모터이다. 일부 구현예에서, 프로모터는 닭 β-액틴 (CBA) 프로모터이다. 일부 구현예에서, 프로모터는 2개의 CBA 프로모터이다. 일부 구현예에서, 프로모터는 CMV 인핸서에 의해 분리되는 2개의 CBA 프로모터이다. 일부 구현예에서, 프로모터는 CAG 프로모터이다.
- [0078] 유도성 프로모터는 유전자 발현의 조절을 허용하고 외인성 공급 화합물, 온도와 같은 환경 인자, 또는 특정 생리학적 상태, 예를 들어, 급성기, 세포의 특정 분화 상태의 존재, 또는 복제 세포 단독에 의해 조절될 수 있다. 유도성 프로모터 및 유도성 시스템은 Invitrogen, Clontech 및 Ariad를 포함하지만, 이에 제한되지 않는 다양한 상업적 공급원으로부터 입수할 수 있다. 많은 다른 시스템이 기술되었으며 당업자에 의해 쉽게 선택될 수 있다. 외인성 공급 프로모터에 의해 조절되는 유도성 프로모터의 예시는 징크-유도성 양 메탈로티오닌 (MT) 프로모터, 텍사메타손 (Dex)-유도성 마우스 유방 종양 바이러스 (MMTV) 프로모터, T7 중합효소 프로모터 시스템 (WO 98/10088); 엑디손 곤충 프로모터 (No et al., Proc. Natl. Acad. Sci. USA, 93:3346-3351 (1996)), 테트라사이클린-억제 시스템 (Gossen et al., Proc. Natl. Acad. Sci. USA, 89:5547-5551 (1992)), 테트라사이클린-유도성 시스템 (Gossen et al., Science, 268:1766-1769 (1995), 또한 Harvey et al., Curr. Opin. Chem. Biol., 2:512-518 (1998) 참조), RU486-유도성 시스템 (Wang et al., Nat. Biotech., 15:239-243 (1997) 및 Wang et al., Gene Ther., 4:432-441 (1997)) 및 라파마이신-유도성 시스템 (Magari et al., J. Clin. Invest., 100:2865-2872 (1997))을 포함한다. 이와 관련하여 유용할 수 있는 또 다른 유형의 유도성 프로모터는 특정 생리학적 상태, 예를 들어, 온도, 급성기, 세포의 특정 분화 상태, 또는 복제 세포 단독에 의해 조절된다.
- [0079] 다른 구현예에서, 이식유전자에 대한 천연 프로모터가 사용될 것이다. 이식유전자의 발현이 천연 발현을 모방해야 하는 것이 바람직한 경우 천연 프로모터가 바람직할 수 있다. 천연 프로모터는 이식유전자의 발현이 일시적으로 또는 발달적으로, 또는 조직-특이적 방식으로, 또는 특정 전사 자극에 대한 반응으로 조절되어야 할 때 사용될 수 있다. 추가 구현예에서, 다른 천연 발현 제어 요소, 예를 들어 인핸서 요소, 폴리아네닐화 부위 또는 코작 공통 서열이 또한 천연 발현을 모방하기 위해 사용될 수 있다.
- [0080] 일부 구현예에서, 조절 서열은 조직 특이적 유전자 발현 능력을 부여한다. 일부 경우에, 조직 특이적 조절 서열

은 조직 특이적 방식으로 전사를 유도하는 조직 특이적 전사 인자에 결합한다. 이러한 조직 특이적 조절 서열 (예를 들어, 프로모터, 인핸서 등)은 당 업계에 잘 알려져 있다. 예시적 조직 특이적 조절 서열은 하기 조직 특이적 프로모터를 포함하지만, 이에 제한되지 않는다: 간 특이적 티록신 결합 글로블린 (TBG) 프로모터, 인슐린 프로모터, 글루카곤 프로모터, 소마토스타틴 프로모터, 췌장 폴리펩티드 (PPY) 프로모터, 시냅신-1 (Syn) 프로모터, 크레아틴 키나제 (MCK) 프로모터, 포유동물 데스민 (DES) 프로모터, α -미오신 중쇄 (α -MHC) 프로모터, 또는 심장 트로포닌 T (cTnT) 프로모터. 다른 예시적인 프로모터는 숙련된 기술자에게 명백할 다른 것들 중에서, 베타-액틴 프로모터, B형 간염 바이러스 코어 프로모터, Sandig et al., Gene Ther., 3:1002-9 (1996); 알파-태아단백질 (AFP) 프로모터, Arbutnot et al., Hum. Gene Ther., 7:1503-14 (1996)), 골 오스테오칼신 프로모터 (Stein et al., Mol. Biol. Rep., 24:185-96 (1997)); 골 시알로프로테인 프로모터 (Chen et al., J. Bone Miner. Res., 11:654-64 (1996)), CD2 프로모터 (Hansal et al., J. Immunol., 161:1063-8 (1998); 면역글로블린 중쇄 프로모터; T 세포 수용체 α -사슬 프로모터, 뉴런 특이적 에놀라제 (NSE) 프로모터와 같은 신경원 (Andersen et al., Cell. Mol. Neurobiol., 13:503-15 (1993)), 뉴로필라멘트 경쇄 유전자 프로모터 (Piccioli et al., Proc. Natl. Acad. Sci. USA, 88:5611-5 (1991)), 및 뉴런 특이적 vgf 유전자 프로모터 (Piccioli et al., Neuron, 15:373-84 (1995))를 포함한다.

[0081] 일부 구현예에서, DBD 및 전사활성화제를 포함하는 융합 단백질을 암호화하는 이식유전자는 프로모터에 작동 가능하게 연결된다. 일부 구현예에서, 프로모터는 구성적 프로모터이다. 일부 구현예에서, 프로모터는 유도성 프로모터이다. 일부 구현예에서, 프로모터는 조직 특이적 프로모터이다. 일부 구현예에서, 프로모터는 신경 조직에 특이적이다. 일부 구현예에서, 프로모터는 SST 또는 NPY 프로모터이다.

[0082] 본 개시내용의 양태는 하나 초과와 프로모터 (예를 들어, 2, 3, 4, 5, 또는 그 이상의 프로모터)를 포함하는 단리된 핵산에 관한 것이다. 예를 들어, 단백질을 암호화하는 제1 영역 및 단백질을 암호화하는 제2 영역을 포함하는 이식유전자를 갖는 작제물의 맥락에서 제1 프로모터 서열 (예를 들어, 단백질 코딩 영역에 작동 가능하게 연결된 제1 프로모터 서열)을 사용하는 제1 단백질 코딩 영역의 발현을 유도하고, 제2 프로모터 서열 (예를 들어, 제2 단백질 코딩 영역에 작동 가능하게 연결된 제2 프로모터 서열)을 사용하여 제2 단백질 코딩 영역의 발현을 유도하는 것이 바람직할 수 있다. 일반적으로, 제1 프로모터 서열 및 제2 프로모터 서열은 동일한 프로모터 서열 또는 상이한 프로모터 서열일 수 있다. 일부 구현예에서, 제1 프로모터 서열 (예를 들어, 단백질 코딩 영역의 발현을 유도하는 프로모터)은 RNA 중합효소 III (pol III) 프로모터 서열이다. pol III 프로모터 서열의 비제한적 예시는 U6 및 H1 프로모터 서열을 포함한다. 일부 구현예에서, 제2 프로모터 서열 (예를 들어, 제2 단백질의 발현을 유도하는 프로모터 서열)은 RNA 중합효소 II (pol II) 프로모터 서열이다. pol II 프로모터 서열의 비제한적 예시는 T7, T3, SP6, RSV, 및 시토크갈로바이러스 프로모터 서열을 포함한다. 일부 구현예에서, pol III 프로모터 서열은 제1 단백질 코딩 영역의 발현을 유도한다. 일부 구현예에서, pol II 프로모터 서열은 제2 단백질 코딩 영역의 발현을 유도한다.

[0083] *제조합 아데노 관련 바이러스 (rAAV)*

[0084] 일부 양태에서, 본 개시내용은 단리된 아데노 관련 바이러스 (AAV)를 제공한다. AAV와 관련하여 본원에 사용된 바와 같이, 용어 "단리된"은 인공적으로 생성되거나 수득된 AAV를 지칭한다. 단리된 AAV는 제조합 방법을 사용하여 생성될 수 있다. 이러한 AAV는 본원에서 "제조합 AAV"로 지칭된다. 제조합 AAV (rAAV)는 바람직하게는 rAAV의 뉴클레아제 및/또는 이식유전자가 하나 이상의 미리 결정된 조절(들)에 특이적으로 전달될 수 있도록, 조직 특이적 표적화 능력을 갖는다. AAV 캡시드는 이러한 조직 특이적 표적화 능력을 결정하는 중요한 요소이다. 따라서, 표적화되는 조직에 적절한 캡시드를 갖는 rAAV가 선택될 수 있다.

[0085] 원하는 캡시드 단백질을 갖는 제조합 AAV를 수득하는 방법은 당 업계에 잘 알려져 있다. (예를 들어, US 2003/0138772 참조), 이의 내용은 그 전문이 본원에 참고로 포함됨). 일반적으로 방법은 AAV 캡시드 단백질을 암호화하는 핵산 서열; 기능적 rep 유전자; AAV 역위 말단 반복 (ITR) 및 이식유전자로 구성된 제조합 AAV 벡터; 및 제조합 AAV 벡터의 AAV 캡시드 단백질로의 패키징을 허용하기에 충분한 헬퍼 기능을 함유하는 숙주 세포를 배양하는 것을 포함한다. 일부 구현예에서, 캡시드 단백질은 AAV의 cap 유전자에 의해 암호화되는 구조적 단백질이다. AAV는 3개의 캡시드 단백질, 비리온 단백질 1 내지 3 (VP1, VP2 및 VP3로 명명됨)을 포함하고, 이들 모두는 대체 스플라이싱을 통해 단일 cap 유전자로부터 전사된다. 일부 구현예에서, VP1, VP2 및 VP3의 분자량은 각각 약 87 kDa, 약 72 kDa 및 62 kDa이다. 일부 구현예에서, 번역시, 캡시드 단백질은 바이러스 게놈 주위에 구형 60-mer 단백질 쉘을 형성한다. 일부 구현예에서, 캡시드 단백질의 기능은 바이러스 게놈을 보호하고, 게놈을 전달하고 숙주와 상호작용하는 것이다. 일부 양태에서, 캡시드 단백질은 조직 특이적 방식으로 바이

러스 계놈을 숙주에 전달한다.

- [0086] 일부 구현예에서, AAV 캡시드 단백질은 AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV9, AAV10, AAVrh10, 및 AAV.PHP.B로 이루어진 그룹으로부터 선택된 AAV 혈청형의 것이다. 일부 구현예에서, AAV 캡시드 단백질은 비인간 영장류로부터 유래된 혈청형, 예를 들어 AAVrh8 혈청형의 것이다. 일부 구현예에서, AAV 캡시드 단백질은 광범위하고 효율적인 CNS 형질도입을 위해 유래된 혈청형, 예를 들어 AAV.PHP.B의 것이다. 일부 구현예에서, 캡시드 단백질은 AAV 혈청형 9의 것이다.
- [0087] rAAV 벡터를 AAV 캡시드에 패키징하기 위해 숙주 세포에서 배양될 구성성분은 트랜스로 숙주 세포에 제공될 수 있다. 대안적으로, 필수 구성성분 (예를 들어, 재조합 AAV 벡터, *rep* 서열, *cap* 서열, 및/또는 헬퍼 기능) 중 임의의 하나 이상은 당업자에게 공지된 방법을 사용하여 필수 구성성분 중 임의의 하나 이상을 함유하도록 조작된 안정한 숙주 세포에 의해 제공될 수 있다. 가장 적합하게는, 이러한 안정한 숙주 세포는 유도성 프로모터의 제어하에 필수 구성성분(들)을 함유할 것이다. 그러나, 필수 구성성분(들)은 구성적 프로모터의 제어하에 있을 수 있다. 적합한 유도성 및 구성적 프로모터의 예시는 이식유전자와 함께 사용하기에 적합한 조절 요소의 논의에서 본원에 제공된다. 또 다른 대안에서, 선택된 안정한 숙주 세포는 구성적 프로모터의 제어하에 선택된 구성성분(들) 및 하나 이상의 유도성 프로모터의 제어하에 다른 선택된 구성성분(들)을 함유할 것이다. 예를 들어, 안정한 숙주 세포는 293 세포 (이는 구성적 프로모터의 제어하에 E1 헬퍼 기능을 함유함)로부터 유래되어 생성될 수 있지만, 이는 유도성 프로모터의 제어하에 *rep* 및/또는 *cap* 단백질을 함유한다. 또 다른 안정한 숙주 세포는 당업자에 의해 생성될 수 있다.
- [0088] 일부 구현예에서, 본 개시내용은 이식유전자 (예를 들어, 전사 조절자 도메인에 융합된 DNA 결합 도메인)를 암호화하는 코딩 서열을 포함하는 핵산을 함유하는 숙주 세포에 관한 것이다. 일부 구현예에서, 숙주 세포는 포유 동물 세포, 효모 세포, 박테리아 세포, 곤충 세포, 식물 세포, 또는 진균 세포이다.
- [0089] 본 개시내용의 rAAV를 생성하는데 필요한 재조합 AAV 벡터, *rep* 서열, *cap* 서열, 및 헬퍼 기능은 임의의 적절한 유전 요소 (벡터)를 사용하여 패키징 숙주 세포에 전달될 수 있다. 선택된 유전 요소는 본원에 기술된 방법을 비롯한 임의의 적합한 방법에 의해 전달될 수 있다. 본 개시내용의 임의의 구현예를 구성하기 위해 사용된 방법은 핵산 조작의 당업자에게 알려져 있고 유전 공학, 재조합 공학, 및 합성 기술을 포함한다. 예를 들어, Sambrook et al., *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, N.Y 참조. 유사하게, rAAV 비리온을 생성하는 방법은 잘 알려져 있으며 적절한 방법의 선택은 본 개시내용에서 제한이 없다. 예를 들어, K. Fisher et al., *J. Virol.*, 70:520-532 (1993) 및 미국 특허 번호 5,478,745 참조.
- [0090] 일부 구현예에서, 재조합 AAV는 삼중 형질감염 방법을 사용하여 생성될 수 있다 (미국 특허 번호 6,001,650에 상세히 기술됨). 일반적으로, 재조합 AAV는 숙주 세포를 AAV 입자, AAV 헬퍼 기능 벡터, 및 부속 기능 벡터로 패키징될 AAV 벡터 (ITR 요소 측면에 있는 이식유전자를 포함)를 사용하여 형질감염시킴으로써 생성된다. AAV 헬퍼 기능 벡터는 "AAV 헬퍼 기능" 서열 (예를 들어, *rep* 및 *cap*)을 암호화하고, 이는 생산적인 AAV 복제 및 캡슐화를 위해 트랜스로 기능한다. 바람직하게는, AAV 헬퍼 기능 벡터는 임의의 검출 가능한 야생형 AAV 비리온 (예를 들어, 기능적 *rep* 및 *cap* 유전자를 함유하는 AAV 비리온)을 생성하지 않고 효율적인 AAV 벡터 생산을 지원한다. 본 개시내용과 함께 사용하기에 적합한 벡터의 비제한적 예시는 미국 특허 번호 6,001,650에 기술된, pHLP19 및 미국 특허 번호 6,156,303에 기술된, pRep6cap6 벡터를 포함하고, 둘 모두 그 전문이 본원에 참고로 포함된다. 부속 기능 벡터는 AAV가 복제에 의존하는 비-AAV 유래 바이러스 및/또는 세포 기능 (예를 들어, "부속 기능")에 대한 뉴클레오티드 서열을 암호화한다. 부속 기능은 AAV 유전자 전사의 활성화, 단계 특이적 AAV mRNA 스플라이싱, AAV DNA 복제, *cap* 발현 산물의 합성, 및 AAV 캡시드 조립에 관련된 모이어티를 포함하지만, 이에 제한되지 않는 AAV 복제에 필요한 기능을 포함한다. 바이러스 기반 부속 기능은 아데노바이러스, 헤르페스 바이러스 (단순 헤르페스 바이러스 유형-1 제외), 및 백시니아 바이러스와 같은 공지된 헬퍼 바이러스 중 하나로부터 유래될 수 있다.
- [0091] 일부 양태에서, 본 개시내용은 형질감염된 숙주 세포를 제공한다. 용어 "형질감염"은 세포에 의한 외래 DNA 흡수를 지칭하기 위해 사용되고, 외인성 DNA가 세포막 내부에 도입되었을 때 세포가 "형질감염된다". 다수의 형질감염 기술이 일반적으로 당업계에 공지되어 있다. 예를 들어, Graham et al. (1973) *Virology*, 52:456, Sambrook et al. (1989) *Molecular Cloning, a laboratory manual*, Cold Spring Harbor Laboratories, New York, Davis et al. (1986) *Basic Methods in Molecular Biology*, Elsevier, 및 Chu et al. (1981) *Gene* 13:197 참조. 이러한 기술은 하나 이상의 외인성 핵산, 예를 들어 뉴클레오티드 통합 벡터 및 다른 핵산 분자를

적합한 숙주 세포로 도입하는데 사용될 수 있다.

- [0092] "숙주 세포"는 관심 물질을 보유하거나, 보유할 수 있는 임의의 세포를 지칭한다. 종종 숙주 세포는 포유동물 세포이다. 일부 구현예에서, 숙주 세포는 뉴런, 선택적으로 GABA성 뉴런이다. 본원에 사용된 바와 같이, "GABA성 뉴런"은 감마 아미노부티르산 (GABA)을 생성하는 신경 세포이다. 포유동물에서, GABA는 이것이 결합하는 뉴런을 결합하고 억제하는 신경계에 널리 분포되어 있는 신경 전달물질이다. 이와 같이, GABA는 간질, 자폐 및 불안을 포함하는, 신경계에 영향을 미치는 수많은 장애와 관련이 있다. *SCN1A* 반접합체 및 녹아웃 마우스 연구는 뇌의 GABA성 뉴런에서 깊은 나트륨 전류 결손을 관찰하였다. 숙주 세포는 AAV 헬퍼 작제물, AAV 미니유전자 플라스미드, 부속 기능 벡터, 또는 재조합 AAV의 생산과 연관된 다른 전달 DNA의 수용자로서 사용될 수 있다. 용어는 형질감염된 원래 세포의 자손을 포함한다. 따라서, 본원에 사용된 "숙주 세포"는 외인성 DNA 서열로 형질감염된 세포를 지칭할 수 있다. 단일 모 세포의 자손은 자연적, 우발적 또는 고의적 돌연변이로 인해 형태 또는 게놈 또는 전체 DNA 보체가 원래의 모체와 반드시 완전히 동일하지 않을 수 있음이 이해되어야 한다.
- [0093] 본원에 사용된 바와 같이, 용어 "세포주"는 시험관 내에서 연속적 또는 연장된 성장 및 분열이 가능한 세포 집단을 지칭한다. 종종, 세포주는 단일 전구 세포로부터 유래된 클론 집단이다. 자발적 또는 유도된 변화가 이러한 클론 집단의 저장 또는 전달 동안 핵형에서 발생할 수 있다는 것이 당 업계에 추가로 공지되어 있다. 따라서, 언급된 세포주로부터 유래된 세포는 조상 세포 또는 배양물과 정확히 동일하지 않을 수 있고, 언급된 세포주는 이러한 변이체를 포함한다.
- [0094] 본원에 사용된 바와 같이, 용어 "재조합 세포"는 외인성 DNA 세그먼트, 예를 들어 생물학적 활성 폴리펩티드의 전사 또는 RNA와 같은 생물학적 활성 핵산의 생산을 유도하는 DNA 세그먼트가 도입된 세포를 지칭한다.
- [0095] 본원에 사용된 바와 같이, 용어 "벡터"는 임의의 유전적 요소, 예를 들어 플라스미드, 파지, 트랜스포존, 코스미드, 염색체, 인공 염색체, 바이러스, 비리온 등을 포함하고, 이는 적절한 제어 요소와 연관된 때 복제할 수 있고 세포 간에 유전자 서열을 전달할 수 있다. 일부 구현예에서, 벡터는 바이러스 벡터, 예를 들어 rAAV 벡터, 렌티바이러스 벡터, 아데노바이러스 벡터, 레트로바이러스 벡터, 등이다. 따라서, 용어는 클로닝 및 발현 비히클, 뿐만 아니라 바이러스 벡터를 포함한다. 일부 구현예에서, 유용한 벡터는 전사될 핵산 세그먼트가 프로모터의 전사 제어 하에 위치하는 벡터인 것으로 고려된다.
- [0096] "프로모터"는 유전자의 특정 전사를 개시하는 데 필요한 세포의 합성 기계, 또는 도입된 합성 기계에 의해 인식되는 DNA 서열을 지칭한다. 구절 "작동적으로 연결된," "작동적으로 위치된," "제어하에" 또는 "전사 제어하에"는 프로모터가 RNA 중합효소 개시 및 유전자 발현을 제어하기 위해 핵산과 관련하여 올바른 위치 및 방향에 있음을 의미한다. 용어 "발현 벡터 또는 작제물"은 서열을 암호화하는 핵산의 부분 또는 전부가 전사될 수 있는 핵산을 함유하는 임의의 유형의 유전적 작제물을 의미한다. 일부 구현예에서, 발현은 예를 들어, 전사된 유전자로부터 생물학적 활성 폴리펩티드 생성물을 생성하기 위한 핵산의 전사를 포함한다. 본 개시내용의 rAAV를 생산하기 위해 원하는 AAV 캡시드에서 재조합 벡터를 패키징하는 전술한 방법은 제한을 의미하지 않으며 다른 적합한 방법이 당업자에게 명백할 것이다.
- [0097] 표적 유전자 발현을 조절하기 위한 방법
- [0098] 세포 또는 대상체에서 유전자 발현을 조절하는 방법이 본 개시내용에 의해 제공된다. 방법은 일반적으로 세포 또는 대상체에 단리된 핵산 또는 DNA 결합 도메인 (예를 들어, ZFP 도메인) 및 전사활성화 도메인을 포함하는 융합 단백질을 암호화하는 이식유전자를 포함하는 rAAV를 투여하는 것을 포함한다. 일부 구현예에서, 융합 단백질은 ZFP 및 VP64 전사활성화제를 포함한다. 일부 구현예에서, 융합 단백질은 ZFP 및 p65 전사활성화제를 포함한다. 일부 구현예에서, 융합 단백질은 ZFP 및 RTA 전사활성화제를 포함한다. 일부 구현예에서, 융합 단백질은 ZFP 및 VPR 전사활성화제를 포함한다. 일부 구현예에서, 방법은 dCas9 단백질 및 *SCN1A*를 표적으로 하는 적어도 하나의 가이드 핵산 (예를 들어, 서열 번호 83-94 중 어느 하나를 포함하거나 서열 번호 83-94 중 어느 하나에 의해 암호화되는 가이드 핵산)을 세포 또는 대상체에 투여하는 것으로 포함한다.
- [0099] 단리된 핵산 또는 융합 단백질 (예를 들어, 전사활성화제를 포함하는 융합 단백질)을 암호화하는 rAAV를 세포 또는 대상체에 투여하는 것은 일부 구현예에서, 표적 유전자 (예를 들어, *SCN1A*)의 증가된 발현을 초래한다. 따라서, 일부 구현예에서, 본 개시내용에 의해 기술된 조성물 및 방법은 표적 유전자의 반가불충분성으로 유래된 병태, 예를 들어 *SCN1A* 유전자의 반가불충분성으로부터 초래된 드라베 증후군을 치료하는데 유용하다.
- [0100] 본원에 사용된 바와 같이, "반가불충분성"은 한 카피의 유전자 (예를 들어, *SCN1A*)가 예를 들어, 유전적 돌연변이에 의해 불활성화되거나, 결실되고, 유전자의 남은 기능적 카피는 유전자의 정상적 기능을 보존하기에 충분한

양의 유전자 생성물을 생산하기에 충분하지 않는 유전적 병태를 지칭한다.

- [0101] 영아기 중증 근간대성 간질로도 알려진, 드라베 증후군은 일반적으로 생후 3년 이내에 나타나는, 드문, 평생 형태의 간질이다. 드라베 증후군은 장기간의 빈번한 발작, 행동 및 발달 지연, 움직임 및 균형 문제, 지연된 언어 및 대화 문제, 및 자율신경계의 교란을 특징으로 한다. 일부 구현예에서, 대상체는 예를 들어 한 카피의 *SCN1A* 유전자가 돌연변이되고, 세포 또는 대상체에서 감소된 *SCN1A* 단백질을 초래하는, 드라베 증후군과 관련된 반가불충분성을 갖는다. 대다수의 드라베 증후군 환자는 절단된 단백질로 번역되는 *SCN1A* 돌연변이; 스플라이스 부위 및 미스센스 돌연변이, 뿐만 아니라 *SCN1A* 유전자에 걸쳐 임의로 분포하는 돌연변이를 포함하는 드라베 증후군과 관련된 다른 *SCN1A* 돌연변이를 보유한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *SCN1A* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *SCN1A*를 표적화하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *SCN1A* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다.
- [0102] 일부 구현예에서, 대상체는 *MED13L* 반가불충분성 증후군과 관련된 반가불충분성을 갖고, 여기서 대상체는 *MED13L* 유전자의 단일 기능적 카피만을 갖는다. *MED13L* 반가불충분성 증후군을 앓는 대상체는 일반적으로 *MED13L* 유전자의 제2, 비기능적 카피의 돌연변이를 갖는다. *MED13L* 반가불충분성 증후군은 지적 장애, 언어 문제, 독특한 얼굴 특징, 및 발달 지연을 특징으로 한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *MED13L* 유전자 및 전사활성화 도메인 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *MED13L*을 표적화하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *MED13L* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다.
- [0103] 일부 구현예에서, 대상체는 골수이형성 증후군과 관련된 반가불충분성을 갖는다. 골수이형성 증후군을 앓는 대상체는 일반적으로 이소시트레이트 탈수소효소 1 (*IDH1*), 이소시트레이트 탈수소효소 2 (*IDH2*), 및/또는 *GATA2* 유전자의 한 카피에 돌연변이를 갖는다. 골수이형성 증후군은 골수의 미성숙 혈액 세포가 건강한 혈액 세포로 성숙하지 않는 한 그룹의 암이다. 때때로, 이 증후군은 급성 골수성 백혈병을 초래할 수 있다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *IDH1* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *IDH1*을 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *IDH1* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *IDH2* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *IDH2*을 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *IDH2* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *GATA2* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *GATA2*를 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *GATA2* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다.
- [0104] 일부 구현예에서, 대상체는 디조지 증후군과 관련된 반가불충분성을 갖는다. 디조지 증후군을 앓는 대상체는 일반적으로 22q11.2로 알려진 위치에서 22번 염색체의 중간에 30 내지 40개 유전자 결손을 갖는다. 특히, 질환은 *TBX* 유전자의 반가불충분성을 특징으로 할 수 있다. 디조지 증후군은 선천성 심장 문제, 특이적 얼굴 특징, 빈번한 감염, 발달 지연, 학습 문제 및 구개열을 특징으로 한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *TBX* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *TBX*를 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *TBX* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다.
- [0105] 일부 구현예에서, 대상체는 *CHARGE* 증후군과 관련된 반가불충분성을 갖는다. 대부분의 경우, *CHARGE* 증후군을 앓는 대상체는 *CHD7* 유전자에 대해 반가불충분하다. *CHARGE* 증후군은 눈의 안조직 결손, 심장 결함, 비강 폐쇄 증, 성장 및/또는 발달의 지체, 생식기 및/또는 비뇨기 이상, 및 귀 이상 및 난청을 특징으로 한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *CHD7* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *CHD7*을 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *CHD7* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는)

가이드 핵산 (예를 들어, gRNA)을 포함한다.

- [0106] 일부 구현예에서, 대상체는 엘러스-단로스 증후군과 관련된 반가불충분성을 갖는다. 엘러스-단로스 증후군을 앓는 대상체는 *COL1A1*, *COL1A2*, *COL3A1*, *COL5A1*, *COL5A2*, *TNXB*, *ADAMTS2*, *PLOD1*, *BAGALT7*, *DSE*, 및/또는 *DAST1/CHST14* 유전자에 대해 반가불충분일 수 있다. 엘러스-단로스 증후군은 피부의 과탄성을 특징으로 하고 대동맥 박리, 척추 측만증, 조기 발병 골관절염을 유발할 수 있다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *COL1A1*, *COL1A2*, *COL3A1*, *COL5A1*, *COL5A2*, *TNXB*, *ADAMTS2*, *PLOD1*, *BAGALT7*, *DSE*, 또는 *DAST1/CHST14* 유전자 중 어느 하나 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *COL1A1*, *COL1A2*, *COL3A1*, *COL5A1*, *COL5A2*, *TNXB*, *ADAMTS2*, *PLOD1*, *BAGALT7*, *DSE*, 또는 *DAST1/CHST14* 중 어느 하나를 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *COL1A1*, *COL1A2*, *COL3A1*, *COL5A1*, *COL5A2*, *TNXB*, *ADAMTS2*, *PLOD1*, *BAGALT7*, *DSE*, 또는 *DAST1/CHST14* 유전자 중 어느 하나를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다.
- [0107] 일부 구현예에서, 대상체는 전두측두엽 치매 (FTD)와 관련된 반가불충분성을 갖는다. FTD를 앓는 대상체는 *MAPT* 유전자에 대해 반가불충분하고, 이는 타우 단백질, 및/또는 *GRN* 유전자를 암호화한다. FTD는 기억 상실, 사회적 인식 부족, 충동 조절 부족, 및 언어 장애를 특징으로 한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *MAPT* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *MAPT*를 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *MAPT* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *GRN* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *GRN*을 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *GRN* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다.
- [0108] 일부 구현예에서, 대상체는 홀트-오람 증후군과 관련된 반가불충분성을 갖는다. 홀트-오람 증후군을 앓는 대상체는 *TBX5* 유전자에 대해 반가불충분하다. 홀트-오람 증후군은 선천성 심장 결함 및 심장 전도 질환을 포함하는 심장 합병증을 특징으로 한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *TBX5* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *TBX5*를 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *TBX5* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다.
- [0109] 일부 구현예에서, 대상체는 마르판 증후군과 관련된 반가불충분성을 갖는다. 마르판 증후군을 앓는 대상체는 일반적으로 피브릴린-1 단백질을 암호화하는, *FBNI* 유전자에 대해 반가불충분하다. 마르판 증후군은 불균형적인 사지 길이, 조기 발병 관절염, 심장 합병증 및/또는 자율 신경계 기능 장애를 특징으로 한다. 일부 구현예에서, 본 개시내용의 융합 단백질은 *FBNI* 유전자 및 전사활성화 도메인을 특이적으로 표적화하는 (예를 들어, 결합하는) ZFP 도메인을 포함한다. 일부 구현예에서, *FBNI*을 표적으로 하는 조성물은 (i) dCas 단백질 및 전사활성화 도메인을 포함하는 융합 단백질, 및 (ii) *FBNI* 유전자를 특이적으로 표적화하는 (예를 들어, 결합하는) 가이드 핵산 (예를 들어, gRNA)을 포함한다.
- [0110] 본 개시내용은 부분적으로, 본원에 기술된 바와 같이 대상체에 융합 단백질을 투여하는 방법을 기반으로 한다. 일부 구현예에서, 융합 단백질은 DBD 및 전사 활성화제를 포함한다. 일부 구현예에서, DBD는 ZNF, TALE, dCas 단백질 (예를 들어, dCas9 또는 dCas12a), 또는 *SCN1A* 유전자에 결합하는 호메오도메인이다. 일부 구현예에서, 전사 활성화제는 VP64, p65, RTA, 또는 VP64-p65-RTA (VPR)를 포함하는 3부 전사 활성화제이다. 일부 구현예에서, 융합 단백질은 AAV 역위 말단 반복 (ITR) 서열 측면에 있다. 일부 구현예에서, 융합 단백질은 프로모터에 작동 가능하게 연결된다. 일부 구현예에서, 대상체는 *SCN1A* 단백질 반가불충분성을 유발하는 *SCN1A*에서의 돌연변이를 갖거나 갖는 것으로 의심된다. 일부 구현예에서, 대상체는 드라베 증후군을 갖거나 갖는 것으로 의심된다.
- [0111] 일부 양태에서, 본 개시내용은 세포에서 표적 유전자의 발현을 조절 (예를 들어, 증가, 감소 등)하는 방법을 제공한다. 일부 구현예에서, 본 개시내용은 세포에서 표적 유전자 (예를 들어, *SCN1A*)의 발현을 증가시키는 방법을 제공한다. 일부 구현예에서, 세포는 포유동물 세포이다. 일부 구현예에서, 세포는 대상체 (예를 들어, 생체 내)에 있다. 일부 구현예에서, 대상체는 포유동물 대상체, 예를 들어 인간이다. 일부 구현예에서, 세포는 신경계 세포 (중추 신경계 세포 또는 말초 신경계 세포), 예를 들어 뉴런 (예를 들어, GABA성 뉴런, 단극성 뉴런,

양극성 뉴런, 바스켓 세포, 베헤 세포, 루가로 세포, 가시 뉴런, 푸르키네 세포, 피라미드 세포, 렌쇼 세포, 파립 세포, 운동 뉴런, 방추 세포 등) 또는 신경교 세포 (예를 들어, 성상교세포, 희돌기교세포, 뇌실막 세포, 방사 신경교, 슈반 세포, 위성 세포, 등)이다.

[0112] "정상" 세포 또는 대상체에서, 표적 유전자 (예를 들어, SCN1A)의 발현은 세포 또는 대상체가 표적 유전자 (예를 들어, SCN1A)에 관해서 반가불충분하지 않으면 충분하다. 일부 구현예에서, 이식유전자의 "개선된" 또는 "증가된" 발현 또는 활성은 본원에 기술된 하나 이상의 단리된 핵산, rAAV, 또는 조성물이 투여되지 않은 세포 또는 대상체에서 이식유전자의 발현 또는 활성에 비해 측정된다. 일부 구현예에서, 이식유전자의 "개선된" 또는 "증가된" 발현 또는 활성은 본원에 기술된 하나 이상의 단리된 핵산, rAAV, 또는 조성물이 대상체에 투여된 후 대상체에서 이식유전자의 발현 또는 활성에 비해 측정된다 (예를 들어, 유전자 발현은 이의 투여 전 및 투여 후 측정됨). 예를 들어, 일부 구현예에서, 세포 또는 대상체에서 SCN1A의 "개선된" 또는 "증가된" 발현은 융합 ZFP-전사활성화제를 암호화하는 이식유전자가 투여되지 않는 세포 또는 대상체에 비해 측정된다. 일부 구현예에서, 본 개시내용에 의해 기술된 방법은 본 개시내용에 의해 기술된 하나 이상의 조성물을 투여받지 않은 대상체의 SCN1A 발현 및/또는 활성에 비해 2배 내지 100배 (예를 들어, 2배, 5배, 10배, 50배, 100배 등) 증가한 대상체에서 SCN1A 발현 및/또는 활성을 유발한다.

[0113] 본원에 사용된 바와 같이, 용어 "치료", "치료하는", 및 "치료법"은 치료적 처치 및 예방적 또는 방지적 조작을 지칭한다. 용어는 추가적으로 기존 증상 개선, 추가 증상 방지, 증상의 기저 원인을 개선 또는 방지, 증상, 예를 들어, 반가불충분 유전자, 예를 들어, 반가불충분 SCN1A 유전자와 관련된 증상의 원인을 방지 또는 반전을 포함한다. 따라서, 유의한 결과를 나타내는 용어는 장애 (예를 들어, 반가불충분 유전자, 예를 들어, 드라베 증후군과 관련된 질환 또는 병태), 또는 이러한 장애가 발달할 잠재력이 있는 대상체에 부여된다. 더욱이, 용어 "치료"는 또한 질환, 질환의 증상 또는 질환에 대한 소인을 치료, 치유, 개선, 완화, 변경, 치료, 개선, 향상 또는 영향을 주고자 질환, 질환의 증상 또는 질환에 대한 소인을 가질 수 있는 대상체, 또는 대상체로부터 단리된 조직 또는 세포주에 대해 체제 (예를 들어, 치료제 또는 치료 조성물, 예를 들어, 표적 유전자 또는 표적 유전자의 조절 영역을 표적으로 하거나 이에 결합하는 단리된 핵산 또는 rAAV)의 적용 또는 투여를 포함한다.

[0114] 치료제 또는 치료적 조성물은 특정 질환 (예를 들어, 반가불충분 유전자, 예를 들어, 드라베 증후군과 관련된 질환 또는 병태)의 증상을 방지 및/또는 감소시키는 약학적으로 허용 가능한 형태인 화합물을 포함할 수 있다. 예를 들어 치료적 조성물은 반가불충분 유전자, 예를 들어, 드라베 증후군과 관련된 질환 또는 병태의 증상을 방지 및/또는 감소시키는 약학 조성물일 수 있다. 본 발명의 치료적 조성물은 임의의 적합한 형태로 제공될 것으로 고려된다. 치료적 조성물의 형태는 본원에 기술된 바와 같은 투여 방식을 포함하는, 다수의 인자에 따라 다를 수 있다. 치료적 조성물은 본원에 기술된 다른 성분 중에서, 희석제, 보조제 및 부형제를 함유할 수 있다.

[0115] 투여 방식

[0116] 본 개시내용의 단리된 핵산, rAAV 및 조성물은 당 업계에 공지된 임의의 적절한 방법에 따른 조성물로 대상체에 전달될 수 있다. 예를 들어, 바람직하게는 생리학적으로 호환 가능한 담체 내 (예를 들어, 조성물 내)에 현탁된, rAAV는 대상체, 즉 숙주 동물, 예를 들어 인간, 마우스, 랫트, 고양이, 개, 양, 토끼, 말, 소, 염소, 돼지, 기니피그, 햄스터, 닭, 칠면조, 또는 비인간 영장류 (예를 들어, 마카크 원숭이)에 투여될 수 있다. 일부 구현예에서 숙주 동물은 인간을 포함하지 않는다.

[0117] 포유동물 대상체에 대한 rAAV의 전달은 예를 들어, 근육 내 주사에 의해 또는 포유동물 대상체의 혈류로의 투여에 의해 의할 수 있다. 혈류로의 투여는 정맥, 동맥, 또는 다른 혈관 도관에 주사에 의해 의할 수 있다. 일부 구현예에서, rAAV는 외과 분야에 잘 알려진 기술인, 단리된 사지 관류에 의해 혈류 내로 투여되고, 방법은 본질적으로 rAAV 비리온의 투여 전에 기술자가 전신 순환으로부터 사지를 단리할 수 있게 한다. 미국 특허 번호 6,177,403에 기술된, 단리된 사지 관류 기술의 변형이 또한 근육 세포 또는 조직으로의 형질도입을 잠재적으로 향상시키기 위해 단리된 사지의 혈관계에 비리온을 투여하기 위해 숙련된 기술자에 의해 사용될 수 있다. 더욱이, 어떤 경우에는, 대상체의 CNS로 비리온을 전달하는 것이 바람직할 수 있다. "CNS"는 척추동물의 뇌 및 척수의 모든 세포 및 조직을 의미한다. 따라서, 용어는 신경 세포, 신경교 세포, 성상교세포, 뇌척수액 (CSF), 간질 공간, 뼈, 연골 등을 포함하지만, 이에 제한되지 않는다. 재조합 AAV는 바늘, 카테터 또는 관련 장치를 사용하여, 당 업계에 공지된 신경외과 기법, 예를 들어 정위 주사를 사용하여, 예를 들어, 심실 영역, 뿐만 아니라 선조체 (예를 들어, 선조체의 미상 핵 또는 피막), 시상, 척수 및 신경근 접합부, 또는 소뇌 소엽으로의 주사에 의해 CNS 또는 뇌로 직접 전달될 수 있다 (예를 들어, Stein et al., J Virol 73:3424-3429, 1999; Davidson et al., PNAS 97:3428-3432, 2000; Davidson et al., Nat. Genet. 3:219-223, 1993; 및 Alisky and Davidson, Hum. Gene

Ther. 11:2315-2329, 2000 참조). 일부 구현예에서, 본 개시내용에 기술된 바와 같은 rAAV는 정맥 내 주사에 의해 투여된다. 일부 구현예에서, rAAV는 뇌 내 주사에 의해 투여된다. 일부 구현예에서, rAAV는 경막 내 주사에 의해 투여된다. 일부 구현예에서, rAAV는 선조체 내 주사에 의해 투여된다. 일부 구현예에서, rAAV는 두개 내 주사에 의해 전달된다. 일부 구현예에서, rAAV는 대조 주사에 의해 전달된다. 일부 구현예에서, rAAV는 대뇌 측뇌실 주사에 의해 전달된다.

- [0118] 본 개시내용의 양태는 캡시드 단백질 및 이식유전자를 암호화하는 핵산을 포함하는 재조합 AAV를 포함하는 조성물에 관한 것이고, 상기 이식유전자는 하나 이상의 단백질을 암호화하는 핵산 서열을 포함한다. 일부 구현예에서, 핵산은 AAV ITR을 추가로 포함한다. 일부 구현예에서, 조성물은 약학적으로 허용 가능한 담체를 추가로 포함한다.
- [0119] 본 개시내용의 조성물은 rAAV를 단독으로, 또는 하나 이상의 다른 바이러스 (예를 들어, 하나 이상의 상이한 이식유전자를 암호화하는 제2 rAAV)와 조합하여 포함할 수 있다. 일부 구현예에서, 조성물은 각각 하나 이상의 상이한 이식유전자를 갖는 1, 2, 3, 4, 5, 6, 7, 8, 9, 10개, 또는 그 이상의 상이한 rAAV를 포함한다.
- [0120] 적합한 담체는 rAAV가 향하는 적응증을 고려하여 당업자에 의해 용이하게 선택될 수 있다. 예를 들어, 하나의 적합한 담체는 다양한 완충 용액과 제형화될 수 있는 식염수 (예를 들어, 인산염 완충 식염수)를 포함한다. 다른 예시적인 담체는 멸균 식염수, 락토스, 수크로스, 인산칼슘, 젤라틴, 텍스트란, 한천, 펙틴, 땅콩유, 참기름, 및 물을 포함한다. 담체의 선택은 본 개시내용의 제한 사항이 아니다.
- [0121] 선택적으로, 본 개시내용의 조성물은 rAAV 및 담체(들)에 더하여, 다른 통상적인 제약 성분, 예를 들어 보존제, 또는 화학적 안정화제를 함유할 수 있다. 적합한 예시적인 보존제는 클로로부탄올, 칼륨 소르베이트, 소르브산, 이산화황, 프로필 갈레이트, 파라벤, 에틸 바닐린, 글리세린, 페놀, 파라클로로페놀, 및 Pluronic® F-68과 같은 폴록사머 (비이온성 계면 활성제)를 포함한다. 적합한 화학적 안정제는 젤라틴 및 알부민을 포함한다.
- [0122] rAAV는 원하는 조직의 세포를 형질감염시키고 과도한 부작용 없이 충분한 수준의 유전자 전달 및 발현을 제공하기에 충분한 양으로 투여된다. 통상적이고 약학적으로 허용 가능한 투여 경로는 선택된 기관으로의 직접 전달 (예를 들어, 간으로의 문맥 내 전달), 경구, 흡입 (비강 내 및 기관 내 전달을 포함), 안 내, 정맥 내, 근육 내, 피하, 피내, 중앙 내, 및 다른 비경구 투여 경로를 포함하지만, 이에 제한되지 않는다. 원하는 경우, 투여 경로를 조합할 수 있다.
- [0123] 특정 "치료적 효과"를 달성하기 위해 필요한 rAAV 비리온의 용량, 예를 들어, 게놈 카피/체중 킬로그램당 (GC/kg)의 용량 단위는 하기를 포함하지만, 이에 제한되지 않는 여러 인자에 따라 다를 수 있다: rAAV 비리온 투여의 경로, 치료적 효과를 달성하기 위해 필요한 유전자 또는 RNA 발현의 수준, 치료되는 특정 질환 또는 장애, 및 유전자 또는 RNA 산물의 안정성. 당업자는 전술한 인자, 뿐만 아니라 당업계에 널리 공지된 다른 인자에 기초하여 특정 질환 또는 장애를 갖는 환자를 치료하기 위한 rAAV 비리온 용량 범위를 용이하게 결정할 수 있다.
- [0124] rAAV의 유효량은 동물을 표적 감염시키고, 원하는 조직을 표적화하기에 충분한 양이다. 일부 구현예에서, rAAV의 유효량은 리소좀 축적병의 전증상 단계 동안 대상체에게 투여된다. 일부 구현예에서, 리소좀 축적병의 전증상 단계는 출생 (예를 들어, 주산기) 내지 4주령 사이에 발생한다.
- [0125] 일부 구현예에서, rAAV 조성물은 특히 높은 rAAV 농도가 존재하는 경우 (예를 들어, $\sim 10^{13}$ GC/mL 또는 초과), 조성물에서 AAV 입자의 응집을 감소시키도록 제형화된다. rAAV의 응집을 감소시키는 방법은 당 업계에 잘 알려져 있고, 예를 들어, 계면활성제의 첨가, pH 조정, 염 농도 조정 등을 포함한다 (예를 들어, Wright FR, et al., Molecular Therapy (2005) 12, 171-178 참조, 그 내용은 본원에 참고로 포함됨.)
- [0126] 약학적으로 허용 가능한 부형제 및 담체 용액의 제형화는 다양한 치료 요법에서 본원에 기술된 특정 조성물을 사용하기에 적합한 투여 및 치료 요법의 개발과 마찬가지로, 당업자에게 잘 알려져 있다.
- [0127] 일반적으로, 이들 제형은 활성 화합물의 적어도 약 0.1% 또는 그 이상을 함유할 수 있지만, 활성 성분(들)의 백분율은 물론 다를 수 있고 편리하게는 전체 제형의 중량 또는 부피의 약 1 또는 2% 내지 약 70% 또는 80% 또는 그 이상일 수 있다. 당연히, 제조될 수 있는 각각 치료학적으로 유용한 조성물에서 활성 화합물의 양은 적합한 투여량이 화합물의 임의의 주어진 단위 용량에서 얻어질 수 있는 방식이다. 용해도, 생체이용률, 생물학적 반감기, 투여 경로, 제품 저장 수명, 뿐만 아니라 다른 약리학적 고려사항과 같은 인자는 이러한 약제학적 제형을

제조하는 당업자에 의해 고려될 것이고, 이와 같이, 다양한 투여량 및 치료 요법이 바람직할 수 있다.

- [0128] 특정 상황에서 본원에 개시된 적합하게 제형화된 약학 조성물의 rAAV-기반 치료 작제물을 피하, 척장 내, 비강 내, 비경구, 정맥 내, 근육 내, 척수강 내, 또는 경구, 복강 내, 또는 흡입에 의해 전달하는 것이 바람직할 것이다. 일부 구현예에서, 미국 특허 번호 5,543,158; 5,641,515 및 5,399,363 (각각 그 전문이 참고로 구체적으로 포함됨)에 기술된 바와 같은 투여 양상이 rAAV를 전달하기 위해 사용될 수 있다. 일부 구현예에서, 바람직한 투여 방식은 문맥 주사에 의한 것이다.
- [0129] 주사용으로 적합한 약제학적 형태는 멸균 수용액 또는 분산액 및 멸균 주사액 또는 분산액의 즉석 제조를 위한 멸균 분말을 포함한다. 분산액은 또한 글리세롤, 액체 폴리에틸렌 글리콜, 및 이들의 혼합물 및 오일에서 제조될 수 있다. 일반적인 보관 및 사용 조건에서, 이들 제제는 미생물의 성장을 방지하기 위한 방부제를 함유한다. 많은 경우에 형태는 쉽게 주사할 수 있을 정도로 무균 상태이며 유동적이다. 이는 제조 및 보관 조건에서 안정해야 하고 박테리아 및 곰팡이와 같은 미생물의 오염 작용에 대해 보존되어야 한다. 담체는 예를 들어, 물, 에탄올, 폴리올 (예를 들어, 글리세롤, 프로필렌 글리콜, 및 액체 폴리에틸렌 글리콜 등), 이의 적합한 혼합물, 및/또는 식물성 오일을 함유하는 용매 또는 분산 매질일 수 있다. 적절한 유동성은 예를 들어, 레시틴과 같은 코팅의 사용, 분산액의 경우 필요한 입자 크기의 유지 및 계면 활성제의 사용에 의해 유지될 수 있다. 미생물 작용의 방지는 다양한 항균제 및 항진균제, 예를 들어, 파라벤, 클로로부탄올, 페놀, 소르브산, 티메로살 등에 의해 야기될 수 있다. 많은 경우에, 등장화제, 예를 들어, 설탕 또는 염화나트륨을 포함하는 것이 바람직할 것이다. 주사 가능한 조성물의 연장된 흡수는 흡수 지연 제제, 예를 들어, 알루미늄 모노스테아레이트 및 젤라틴의 조성물에서의 사용에 의해 야기될 수 있다.
- [0130] 주사 가능한 수용액의 투여를 위해, 예를 들어 용액은 필요한 경우, 적절하게 완충될 수 있고, 액체 희석제는 먼저 충분한 식염수 또는 포도당으로 등장성이 된다. 이들 특정 수용액은 정맥 내, 근육 내, 피하 및 복강 내 투여에 특히 적합하다. 이와 관련하여, 사용될 수 있는 멸균 수성 매질은 당업자에게 공지되어 있을 것이다. 예를 들어, 1회 용량을 1 mL의 등장성 NaCl 용액에 용해시키고 1000 mL의 피하주사액을 첨가하거나 제안된 주입 부위에 주사할 수 있다 (예를 들어, "Remington's Pharmaceutical Sciences" 15th Edition, 페이지 1035-1038 및 1570-1580 참조). 투여량의 약간의 변화는 숙주의 상태에 따라 필연적으로 발생할 것이다. 투여 책임자는 어쨌든, 개별 숙주에 대한 적절한 용량을 결정할 것이다.
- [0131] 멸균 주사 용액은 적절한 용매 중 필요한 양의 활성 rAAV를 필요에 따라, 본원에 열거된 다양한 다른 성분과 함께 혼합한 후 여과 멸균하여 제조한다. 일반적으로, 분산액은 다양한 멸균된 활성 성분을 기본 분산 매질 및 상기 열거된 것들로부터 필요한 다른 성분을 함유하는 멸균 비히클에 혼합함으로써 제조된다. 멸균 주사 용액의 제조를 위한 멸균 분말의 경우, 바람직한 제조 방법은 이의 사전 멸균 여과된 용액으로부터 활성 성분과 임의의 추가의 원하는 성분의 분말을 생성하는 진공 건조 및 동결 건조 기술이다.
- [0132] 본원에 개시된 rAAV 조성물은 또한 중성 또는 염 형태로 제형화될 수 있다. 약학적으로 허용 가능한 염은 산 부가염 (단백질의 유리 아미노기와 함께 형성됨)을 포함하고 예를 들어, 염산 또는 인산과 같은 무기산, 또는 아세트산, 옥살산, 타르타르산, 만델산 등과 같은 유기산으로 형성된다. 유리 카르복실기로 형성된 염은 또한 예를 들어, 나트륨, 칼륨, 암모늄, 칼슘, 또는 제2철 수산화물과 같은 무기 염기, 및 이소프로필아민, 트리메틸아민, 히스티딘, 프로카인 등과 같은 유기 염기로부터 유래될 수 있다. 제제화 시, 용액은 투여 제형과 양립 가능한 방식으로 치료적으로 효과적인 양으로 투여될 것이다. 제형은 주사용 용액, 약물 방출 캡슐 등과 같은 다양한 제형으로 용이하게 투여된다.
- [0133] 본원에 사용된 바와 같이, "담체"는 임의의 및 모든 용매, 분산 매질, 비히클, 코팅, 희석제, 항균제 및 항진균제, 등장화제 및 흡수 지연제, 완충제, 담체 용액, 현탁액, 콜로이드 등을 포함한다. 약제학적 활성 물질을 위한 이러한 매질 및 제제의 사용은 당업계에 잘 알려져 있다. 보충 활성 성분이 또한 조성물에 포함될 수 있다. 구절 "약학적으로 허용 가능한"은 숙주에 투여될 때 알레르기 반응 또는 유사한 부작용을 일으키지 않는 분자 개체 및 조성물을 지칭한다.
- [0134] 리포솜, 나노캡슐, 마이크로입자, 마이크로스피어, 지질 입자, 소포 등과 같은 전달 비히클은 본 개시내용의 조성물의 숙주 세포로의 도입을 위해 사용될 수 있다. 특히, rAAV 벡터 전달 이식유전자는 지질 입자, 리포솜, 소포, 나노스피어, 또는 나노 입자 등에 캡슐화된 전달을 위해 제형화될 수 있다.
- [0135] 이러한 제형은 본원에 개시된 핵산 또는 rAAV 작제물의 약학적으로 허용 가능한 제형의 도입을 위해 바람직할 수 있다. 리포솜의 형성 및 사용은 일반적으로 당업자에게 알려져 있다. 최근, 혈청 안정성 및 순환 반감기가

개선된 리포솜이 개발되었다 (미국 특허 번호 5,741,516). 추가로, 잠재적 약물 담체로서 리포솜 및 리포솜 유사 체계의 다양한 방법이 기술되어 있다 (미국 특허 번호 5,567,434; 5,552,157; 5,565,213; 5,738,868 및 5,795,587).

[0136] 리포솜은 일반적으로 다른 절차에 의한 형질감염에 내성이 있는 다수의 세포 유형에 성공적으로 사용되었다. 또한, 리포솜은 바이러스 기반 전달 시스템의 전형적인 DNA 길이 제한이 없다. 리포솜은 유전자, 약물, 방사선 치료제, 바이러스, 전사 인자 및 알로스테릭 효과기를 다양한 배양된 세포주 및 동물로 도입하는 데 효과적으로 사용되었다. 또한, 리포솜 매개 약물 전달의 효과를 조사하는 여러 성공적인 임상 시험이 완료되었다.

[0137] 리포솜은 수성 매질에 분산되어 다중층 동심 이중층 소포 (다중층 소포 (MLV)라고도 함)를 자발적으로 형성하는 인지질로부터 형성된다. MLV는 일반적으로 25 nm 내지 4 μm의 직경을 갖는다. MLV의 초음파 처리는 코어에 수용액을 포함하는, 200 내지 500 Å 범위의 직경을 갖는 작은 단층 소포 (SUV)의 형성을 초래한다.

[0138] 대안적으로, rAAV의 나노캡슐 제형이 사용될 수 있다. 나노캡슐은 일반적으로 안정적이고 재현 가능한 방식으로 물질을 포획할 수 있다. 세포 내 중합체 과부하로 인한 부작용을 피하기 위해, 이러한 초미세 입자 (약 0.1 μm 크기)는 생체 내에서 분해될 수 있는 중합체를 사용하여 설계해야 한다. 이러한 요구 사항을 충족하는 생분해성 폴리알킬-시아노아크릴레이트 나노입자가 사용에 고려된다.

[0139] 상기 기술된 전달 방법 이외에, rAAV 조성물을 숙주에 전달하는 대안적 방법으로서 하기 기술이 또한 고려된다. 초음파 영동법 (즉, 초음파)이 사용되었고 순환계로 및 순환계를 통한 약물 투과의 속도 및 효능을 향상시키기 위한 장치로서 미국 특허 번호 5,656,016에 기술되었다. 고려되는 다른 약물 전달 대안은 골 내 주사 (미국 특허 번호 5,779,708), 마이크로칩 장치 (미국 특허 번호 5,797,898), 안과 제형 (Bourlais et al., 1998), 경피 매트릭스 (미국 특허 번호 5,770,219 및 5,783,208) 및 피드백 제어 전달 (미국 특허 번호 5,697,899)이다.

[0140] **실시예**

[0141] *실시예 1. SCN1A 유전자 발현을 상향 조절하기 위한 징크 핑거 단백질의 설계*

[0142] 인간 (HEK293T 세포)과 마우스 (HEPG2 세포) SCN1A 프로모터 서열 사이 상동 영역을 각 종에 대한 RIKEN CAGE-seq 데이터 세트에서 확인된 2개의 두드러진 전사 시작 부위를 둘러싸는 서열의 정렬에 의해 확인하였다 (도 1). 인간 (HEK)과 마우스 (HEPG2) 사이에 고도로 보존된 서열은 SCN1A의 근위 프로모터 영역에 존재한다 (도 2). 6개의 핑거로 이루어진 3개의 ZFP는 미리 정의된 DNA-결합 특이성을 갖는 1개 및 2개의 핑거 모듈의 조합을 통해 SCN1A의 근위 프로모터 영역에서 상동성의 중첩 15-22개 뉴클레오티드 영역에 결합하도록 설계되었다 (도 3). 각각 6개의 핑거로 이루어진 3개의 ZFP (ZFP1-ZFP3)는 도 3에서 확인된 중첩되는 고도로 보존된 서열에 결합하도록 설계되었다. 각각의 핑거는 SCN1A의 근위 프로모터의 고도로 보존된 영역에서 3개의 염기 영역 (삼중체)에 결합하도록 설계되었다.

[0143] 도 4a에 나타낸 바와 같이, ZFP-1은 SCN1A 유전자의 근위 프로모터 영역 내의 각각의 3개의 염기 영역 ("·"로 구분된 적색으로 표시된 DNA 삼중체)을 인식한다 (서열 번호 2). 도 4b에 나타낸 바와 같이, ZFP-1에 대한 핑거 1 내지 6의 각각의 인식 나선 (7개의 아미노산)은 3개의 뉴클레오티드 서열에 결합한다. ZFP-1의 6개의 핑거의 아미노산 서열 (서열 번호 17-22)은 도 4c에 나타내고; 핑거 사이의 링커는 표준 (TGEKP) 및 비표준 (TGSQKP) 링커 서열을 지정하기 위해 강조 표시된다. ZFP-1의 6개의 핑거의 뉴클레오티드 서열 (서열 번호 11-16)은 도 4d에 나타낸다.

[0144] **표 1. SCN1A를 표적으로 하는 ZFP-1의 인식 나선**

	아미노산 서열	뉴클레오티드 서열
ZFP-1 인식 나선 1	QRGNLVR (서열 번호 17)	CAGCGGGGAAACCTGGTGAGG (서열 번호 11)
ZFP-1 인식 나선 2	LSFNLTR (서열 번호 18)	CTGAGCTTCAATCTAACCAGA (서열 번호 12)
ZFP-1 인식 나선 3	RSDNLTR (서열 번호 19)	CGGAGTGACAACCTAACGCGG (서열 번호 13)
ZFP-1 인식 나선 4	DRSHLAR (서열 번호 20)	GACCGGTCTCACCTTGCCCGA (서열 번호 14)
ZFP-1 인식 나선 5	QKAHLTA (서열 번호 21)	CAGAAGGCCCATTTGACTGCC (서열 번호 15)
ZFP-1 인식 나선 6	RSDNLTR (서열 번호 22)	CGGTCGGACAACCTCACACGC (서열 번호 16)

[0145]

[0146] 도 5a에 나타낸 바와 같이, ZFP-2는 SCN1A 유전자의 근위 프로모터 영역 내에 각각의 3개의 염기 영역 ("·"로 구분된 적색으로 표시된 DNA 삼중체)을 인식한다 (서열 번호 3). 도 5b에 나타낸 바와 같이, ZFP-2에 대한 핑거 1 내지 6의 각각의 인식 나선 (7개의 아미노산)은 3개의 뉴클레오티드 서열에 결합한다. ZFP-2의 6개의 핑거의 아미노산 서열 (서열 번호 29-34)은 도 5c에 나타낸다; 핑거 사이의 링커는 표준 (TGEKP) 및 비표준 (TGSQKP) 링커 서열을 지정하기 위해 강조 표시된다. ZFP-1의 6개의 핑거의 뉴클레오티드 서열 (서열 번호 23-28)은 도 5d에 나타낸다.

[0147] 표 2. SCN1A를 표적으로 하는 ZFP-2의 인식 나선

	아미노산 서열	뉴클레오티드 서열
ZFP-2 인식 나선 1	RSSNLTR (서열 번호 29)	CGAAGTTCCAACCTGACACGG (서열 번호 23)
ZFP-2 인식 나선 2	DKRTLIR (서열 번호 30)	GACAAGCGGACCTTAATCCGC (서열 번호 24)
ZFP-2 인식 나선 3	QRGNLVR (서열 번호 31)	CAGCGGGGAAATCTAGTGCGA (서열 번호 25)
ZFP-2 인식 나선 4	LSFNLTR (서열 번호 32)	CTGAGCTTCAACTTGACTCGT (서열 번호 26)
ZFP-2 인식 나선 5	RSDNLTR (서열 번호 33)	CGGAGTGACAATCTTACGAGA (서열 번호 27)
ZFP-2 인식 나선 6	DRSHLAR (서열 번호 34)	GACCGGAGCCACTTAGCCAGG (서열 번호 28)

[0148]

[0149] 도 6a에 나타낸 바와 같이, ZFP-3은 SCN1A 유전자의 근위 프로모터 영역 내에 각각의 3개의 염기 영역 ("·"로 구분된 적색으로 표시된 DNA 삼중체)을 인식한다 (서열 번호 4). 도 6b에 나타낸 바와 같이, ZFP-3에 대한 핑거 1 내지 6의 각각의 인식 나선 (7개의 아미노산)은 3개의 뉴클레오티드 서열에 결합한다. ZFP-3의 6개의 핑거의 아미노산 서열 (서열 번호 41-46)은 도 6c에 나타내고; 핑거 사이의 링커는 표준 (TGEKP) 및 비표준 (TGSQKP) 링커 서열을 지정하기 위해 강조 표시된다. ZFP-1의 6개의 핑거의 뉴클레오티드 서열 (서열 번호 35-40)은 도 6d에 나타낸다.

[0150] 표 3. SCN1A를 표적으로 하는 ZFP-3의 인식 나선

	아미노산 서열	뉴클레오티드 서열
ZFP-3 인식 나선 1	DRSALAR (서열 번호 41)	GACCGGAGCGCGCTGGCACGG (서열 번호 35)
ZFP-3 인식 나선 2	RSDNLTR (서열 번호 42)	CGAAGTGACAACTTAACGCGC (서열 번호 36)
ZFP-3 인식 나선 3	QSGDLTR (서열 번호 43)	CAGTCAGGGGACCTCACTCGT (서열 번호 37)
ZFP-3 인식 나선 4	VRQTLKQ (서열 번호 44)	GTACGACAGACGCTTAAACAA (서열 번호 38)
ZFP-3 인식 나선 5	AAGNLTR (서열 번호 45)	GCCGCTGGTAACTTGACACGA (서열 번호 39)
ZFP-3 인식 나선 6	RSDNLTR (서열 번호 46)	AGATCTGATAATCTAACGCGT (서열 번호 40)

[0151]

[0152] SCN1A 유전자의 근위 프로모터 영역에 보존된 서열을 표적화하기 위해 설계된 추가 ZFP는 각각 5 또는 6개의 핑거 도메인을 포함할 것이고 인간과 마우스 SCN1A 사이 고도로 보존된 15-22개 뉴클레오티드의 영역에 결합할 것이다.

[0153] 표 4. *SCN1A*를 표적으로 하는 징크 핑거 단백질

	아미노산 서열	뉴클레오티드 서열
ZFP-1	RPFQCRICMRNFSQRGNLVR HIRTHTGKPFACDICGKKF ALSFNLTRHTKIHTGSQKPF QCRICMRNFSRSDNLTRHIRT HTGEKPFACDICGKKFADRS HLARHTKIHTGSQKPFQCRIC MRNFSQKAHLTAHIRTHTGE KPFACDICGRKFARSDNLTR HTKIHLRQKD (서열 번호 57)	CGACCATTCCAGTGTGCAATCTGCATGCGCAAC TTCAGCCAGCGGGAAAACCTGGTGAGGCATATC CGCACCCACACGGGAGAGAAGCCTTTTGCCTGC GATATTTGTGGAAAGAAGTTTGTCTCTGAGCCTT CAATCTAACCCAGACACACCAAGATTTCATACTGG GTCCAGAAAACCGTTCCAGTGTAGGATATGCAT GAGGAATTTCTCTCGGAGTGACAACCTTAACGC GGCATATAAGGACGCACACAGGTGAAAAACCAT TTGCATGCGACATCTGTGGCAAAAAGTTTGGC GACCGGTCTCACCTTGCCCGACACAAAAATC CATACCGGCAGTCAAAAAGCCCTTCAATGTCCG ATTTGCATGCGAAAACCTTCTCACAGAAGGCCAT TTGACTGCCCATATTCTACTCATACTGGCGAG AAACCTTTTCGCTTGCATATATGTGGTCGTAA GTTTGCACGGTCGGACAACCTCACACGCCACAC TAAGATACACCTGCGGCAGAAGGAC (서열 번호 58)
ZFP-2	RPFQCRICMRNFSRSSNLTRH IRHTHTGKPFACDICGKKFA DKRTLIRHTKIHTGSQKPFQC RICMRNFSQRGNLVRHIRTH TGEKPFACDICGKKFALSFN LTRHTKIHTGSQKPFQCRIC MRNFSRSDNLTRHIRTHTGE KPFACDICGRKFADRSHLAR HTKIHLRQKD (서열 번호 59)	CGACCATTCCAGTGTGCAATCTGCATGCGCAAC TTCAGCCGAAAGTTCCAACCTGACACGGCATATC CGCACCCACACGGGAGAGAAGCCTTTTGCCTGC GATATTTGTGGAAAGAAGTTTGTGACAAGCG GACCTTAATCCGCGCACACCAAGATTTCATACTGG GTCCAGAAAACCGTTCCAGTGTAGGATATGCAT GAGGAATTTCTCTCAGCGGGAAATCTAGTGC GACATATAAGGACGCACACAGGTGAAAAACCAT TTGCATGCGACATCTGTGGCAAAAAGTTTGGC CTGAGCTTCAACTTGACTCGTCACAAAAATC CATACCGGCAGTCAAAAAGCCCTTCAATGTCCG ATTTGCATGCGAAAACCTTCTCACGGAGTGACAA TCTTACGAGACATATTCTACTCATACTGGCGA GAAACCTTTTCGCTTGCATATATGTGGTCGTA AGTTTGCAGACCGGAGCCACTTAGCCAGGCACA CTAAGATACACCTGCGGCAGAAGGAC (서열 번호 60)
ZFP-3	RPFQCRICMRNFSDRSALAR HIRTHTGKPFACDICGKKF ARSDNLTRHTKIHTGSQKPF QCRICMRNFSQSGDLTRHIRT HTGEKPFACDICGKKFAVRQ TLKQHTKIHTGSQKPFQCRI CMRNFSAAGNLTRHIRTHTG EKPFACDICGRKFARSDNLT RHTKIHLRQKD (서열 번호 61)	CGACCATTCCAGTGTGCAATCTGCATGCGCAAC TTCAGCGACCGGAGCGCTGGCACGGCATATC CGCACCCACACGGGAGAGAAGCCTTTTGCCTGC GATATTTGTGGAAAGAAGTTTGTGCAAGTGA CAACTTAACCGCGCCACACCAAGATTTCATACTGG GTCCAGAAAACCGTTCCAGTGTAGGATATGCAT GAGGAATTTCTCTCAGTCAAGGACCTCACTCG TCATATAAGGACGCACACAGGTGAAAAACCAT TGCAATGCGACATCTGTGGCAAAAAGTTTGGC TACGACAGACGCTTAAACAACACAAAAATCC ATACCGGCAGTCAAAAAGCCCTTCAATGTCCG TTTGCATGCGAAAACCTTCTCAGCCGCTGGTAACT TGACACGACATATTCGTACTCATACTGGCGAGA AACCTTTTCGCTTGCATATATGTGGTCGTAAG TTTGCAGATCTGATAATCTAACCGGTCACACT AAGATACACCTGCGGCAGAAGGAC (서열 번호 62)

[0154] 실시예 2. ZFP는 인간 세포에서 *SCN1A* 유전자 발현을 증가시킨다

[0156] ZFP1-ZFP3의 *SCN1A*의 전사를 상향 조절하는 능력을 조사하기 위해, ZFP1-ZFP3 DNA 결합 도메인은 키메라 전사활성화제를 형성하기 위해 하이브리드 VP64, p53, 및 RTA (VPR) 3부 강한 전사 활성화제 도메인에 융합된다. VPR 융합 활성화제 도메인은 전사 조절 복합체를 모집하고 염색질 접근성을 증가시키고 높은 수준의 유전자 발현을 달성하기 위해 돕는다. 따라서, ZFP 도메인은 *SCN1A* 유전자 발현을 증가시키기 위해 VPR 활성화제를 근위 프로모터 영역에서 고투로 보존된 서열에 대해 표적화할 것이다.

[0157] VPR-ZFP1, VPR-ZFP2, 및/또는 VPR-ZFP3 융합 단백질을 암호화하는 발현 플라스미드는 일시적 형질감염을 통해 HEK293 세포로 형질감염되고 *SCN1A* 유전자 발현을 (정규화에 대해 기준으로 TBP 발현을 사용하여) qRT-PCR로 측정하였다. VPR-ZFP 융합은 VPR에 융합된 ZFP1, ZFP2, 및/또는 ZFP3을 포함한다. 각각 VPR에 융합된 ZFP1, ZFP2, 및 ZFP3 DNA 결합 도메인을 함유하는, 다중 조절을 위한 3개의 작제물의 형질감염은 형질감염되지 않은 세포에 비해 45배 증가된 *SCN1A* 유전자 발현을 초래하였고, VPR-ZFP 키메라 전사활성화제는 유전자의 프로모터 근위 영역에서 결합함으로써 *SCN1A* 유전자 발현을 증가시킬 수 있음을 나타낸다 (도 7).

[0158] VPR-[ZFP1-ZFP3] 융합 단백질, 뿐만 아니라 ZFP DNA 결합 도메인이 현재 설계된 VPR-ZFP 융합 단백질은 둘 다 낮은 수준의 *SCN1A* 발현을 갖는, HeLa 및 HEPG2 세포에서 형질감염되고 있다. VPR-ZFP 융합 단백질은 VPR 전사 활성화제에 융합된 단일뿐만 아니라 다중 ZFP DNA 결합 도메인의 조합을 함유한다. *SCN1A* 유전자 발현은 이들 VPR-ZFP 융합이 유전자 발현을 증가시킬 수 있는지 결정하기 위해 qRT-PCR로 측정된다. 가장 유망한 VPR-ZFP 융합 후보를 *SCN1A* 발현을 증가시키는 능력을 위한 융합 단백질의 아레노 관련 바이러스 (AAV) 전달 후 1차 마

우스 대뇌 피질 뉴런에서 검사하였다.

- [0159] ZFP 도메인의 특이성은 DNA 결합에 중요한 잔기가 다양한 무작위 라이브러리의 이상적인 ZFP를 확인하기 위해 박테리아 윈-하이브리드 선택 시스템 (예를 들어, Meng, et al., "Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases," *Nat Biotechnol*, 2008 참조)을 사용하여 추가로 최적화되고 있다. 새롭게 선택된 ZFP는 개별적으로 뿐만 아니라 다중 ZFP의 조합 모두 VPR 전사활성화제 도메인에 융합될 것이고, qRT-PCR 분석 후 가장 *SCN1A* 유전자 발현을 증가시키는 후보 ZFP 도메인을 확인하기 위해 HEK293, HeLa, 및 HEPG2 세포, 뿐만 아니라 1차 마우스 대뇌 피질 뉴런에서 형질감염될 것이다.
- [0160] 실시예 3. 다양한 효능을 갖는 ZFP^{SCN1A} 전사활성화제 시리즈를 생성한다
- [0161] 실시예 2로부터 *SCN1A* 유전자 발현을 상향 조절하는데 가장 효과적인 ZFP를 AAV 감염다중도 (MOI)의 범위에 걸쳐 *SCN1A* 유전자 발현의 2배 상향 조절을 달성하기 위한 조립체를 확인하기 위해 예상 효능의 구배를 갖는 인간 전사활성화 도메인 (예를 들어, Rta, p65, Hsf1 등) 시리즈에 융합한다. 정상 및 *SCN1A*^{+/-} 마우스의 마우스 1차 대뇌 피질 뉴런을 ZFP^{SCN1A} 융합 전사활성화제를 발현하는 AAV 벡터로 감염시킨다. Nav1.1 단백질의 발현 수준을 웨스턴 블롯 및 qPCR을 사용하여 평가한다. TGF- α 로 8시간 처리된 1차 뉴런을 이 처리가 Nav1.1 단백질 발현을 ~6 내지 8배 증가시키기 때문에 양성 대조군으로 사용한다 (Chen et al., 2015, *Neuroinflammation* 12: 126). ZFP^{SCN1A} 전사활성화의 특정화를 입증하기 위해 다른 Nav 알파 서브유닛 유전자의 발현 수준의 변화를 또한 평가하였다. ZFP^{SCN1A} (HA 태그)에 대한 항체 및 GABA성 뉴런에 특이적인 마커 (예를 들어, 파브알부민⁺ 또는 소마토스타틴⁺) 또는 범용 신경 마커 (예를 들어, NeuN, TUBIII, 및/또는 Map2)를 사용하는 이중 면역형광 염색을 통해 Nav1.1 발현이 GABA성 중간뉴런에 대해 제한되는지 여부를 결정하기 위해 면역형광법을 사용한다. *SCN1A* 유전자의 전사활성화에 대한 ZFP^{SCN1A}의 특이성을 게놈 결합 부위 및 유전자 전달시 생성된 전사체학 프로필을 매핑하기 위해 또한 ChIP-Seq 및 RNA-Seq에 의해 평가한다.
- [0162] 실시예 4. 프로모터 활성 의존성 *SCN1A*-ZFP 전사활성화제의 설계를 안내하는 GABA성 억제제에서 *SCN1A* 프로모터의 히스톤 조직 및 후성유전체 풍경
- [0163] 게놈 표적에 결합하는 ZFP의 능력은 표적 서열 (예를 들어, 뉴클레오솜이 없는 영역의 존재)의 접근성에 달려 있다. DNA 접근성에 대한 이러한 요구 사항은 DNA 표적 서열 접근성의 존재를 기반으로 하는 세포 유형의 서브셋에서만 기능적인 ZFP 전사활성화제를 설계하기 위해 이용된다. 세포 유형 활성화에서의 추가 제한은 ZFP 전사활성화제 발현에 대한 조직 특이적 프로모터의 사용을 통해 달성된다. 복어 (*타키후구 루브림스*) 소마토스타틴 및 뉴로펩티드 Y 유전자의 작은 프로모터는 대뇌 피질 및 해마 억제성 중간뉴런 AAV 벡터 및 렌티바이러스의 맥락에서 고도로 특이적인 이식유전자 발현을 유도하는 것으로 나타났다. 일부 구현예에서, DNA 접근성에 민감한 *SCN1A* 특이적 ZFP의 AAV-기반 전사 제한의 조합은 뇌 전체의 억제성 중간뉴런에서 Nav1.1 단백질 발현의 고도로 특이적인 상향 조절을 초래한다. 이러한 이중 조절 접근법은 정상적으로 발현되지 않는 세포에서 Nav1.1 단백질의 이소성 발현으로 인해 발생할 수 있는 부작용을 최소화할 것이다.
- [0164] *SCN1A* 프로모터의 뉴클레오솜 구조 및 후성유전적 풍경을 마우스 및 인간 GABA성 억제 및 글루타민성 흥분 뉴런 모두에서 분석하였다. 이 정보는 이러한 세포 유형의 *SCN1A* 유전자와 주위에서만 접근 가능한 서열의 표적화를 통해 GABA성 억제 뉴런 제한된 ZFP 전사활성화제를 설계하기 위해 사용된다.
- [0165] Emx1-IRES-Cre를 ROSA26/stop/EGFP 마우스와 교차시켜 생성된 GAD67 프로모터 및 GFP-양성 글루타민성 흥분 뉴런 하에서 TdTomato를 발현하는 형질전환 마우스의 GABA성 억제 뉴런은 형광 활성화 세포 분류 (FACS)를 사용하여 단리된다. 인간 GABA성 및 흥분 뉴런은 유도 만능 줄기 세포 (iPS) 세포로부터 생성되고 이들 세포 유형, 뿐만 아니라 전기생리학적 활성화에 대한 특이적인 마커에 대한 면역염색 및 RT-PCR 모두를 사용하여 확인되었다. 마우스 및 인간 뉴런 집단에서 *SCN1A* 프로모터 주변의 접근 가능한 게놈 영역은 전위효소 접근성 염색질 (ATAC-Seq)에 대한 분석을 사용하여 특성화된다.
- [0166] GABA성 뉴런에서만 접근 가능한 서열을 인식하는 ZFP^{SCN1A} 전사활성화제는 억제 및 흥분 뉴런에서 *SCN1A* 프로모터 주위 게놈 영역의 차등 염색질 접근성에 기초하여 설계되고 있다. 일련의 후보 ZFP-VPR 전사활성화제 융합은 전사활성화제의 결합이 억제 영역에서 Nav1.1 발현을 잠재적으로 상향 조절할 뿐만 아니라 흥분 뉴런에서 Nav1.1

발현의 임의의 바람직하지 않은 유도된 발현을 드러낼 것으로 기대되는 상이한 SCN1A 접근 가능 영역을 표적화하기 위해 생성되고 있다.

- [0167] 범-뉴런 인간 시냅신-1 또는 억제 중간뉴런 특이적 프로모터 하에서 AAV 벡터로부터 발현될 때 어떤 모델 드라베 증후군이 억제 뉴런에서만 접근 가능한 DNA 서열을 인식하도록 설계된 ZFP^{SCN1A} 전사활성화제가 필요한 특이성을 제공하는지 결정하기 위해 배양된 인간 iPS-유래 뉴런 및 마우스 SCN1A^{+/-} 1차 뉴런에서 발현 연구를 수행하였다. Nav1.1 발현 수준은 억제 GABA성 (예를 들어, GABA⁺, GAD65/67⁺, 소마토스타틴, 및/또는 파브알부민) 및 흥분 글루타민성 (예를 들어, Cux1+, FoxG1+, GABA_A 수용체, GABA⁻) 뉴런에 대한 뉴런-유형 특이적 마커를 사용하여 qRT-PCR, 웨스턴 블롯, 및 이중 면역 형광에 의해 측정한다. 동염색체 영역 내 염색질 구조 및 DNA 서열로서 마우스 및 인간 SCN1A 프로모터에서 상이한 서열을 표적화하도록 설계된 ZFP^{SCN1A} 전사활성화제의 세포 유형 특이성은 종에 따라 다르다. 이러한 실험의 대조군은 전사활성화 도메인이 없는 GFP, ZFP를 암호화하는 유사한 AAV 벡터, 또는 ZFP DNA-결합 도메인이 없는 전사활성화제로 감염된 뉴런 배양물을 포함한다.
- [0168] 마이크로RNA (miRNA) 결합 부위는 바람직하지 않은 발현이 발생하는 세포 유형에 제한된 ZFP^{SCN1A} 전사활성화제의 3' 비번역 영역 (3' UTR)에 포함되어 있다 (예를 들어, 글루타민성 흥분 뉴런). 이 접근법은 이전에 AAV- 전달된 이식유전자의 발현을 제한하는 데 사용되었다 (Xie, et al., "MicroRNA-regulated, systematically delivered rAAV9: a step closer to CNS-restricted transgene expression," *Mol. Ther.* 2011). GABA성 억제 뉴런 및 다른 세포 유형의 miRNA 발현 프로파일의 차이는 소 RNA 서열 분석으로 결정된다.
- [0169] 실시예 5. 환자 유래 iPS-생성된 GABA성 중간뉴런의 나트륨 전류 결함을 수정하기 위한 AAV-ZFP^{SCN1A} 유전자 요법의 가능성을 평가한다
- [0170] 드라베 증후군에 대한 ZFP^{SCN1A} 전사활성화제(들)의 개발을 위한 중요한 단계는 이들 인공 전사활성화제가 인간 뉴런에서 원하는 기능을 갖는지 입증하는 것이다. 이를 위해, 드라베 환자 (n= 4-6) 및 비-드라베 환자 (n= 4)로부터 iPS 세포를 수득한다. 비-드라베 유전적 배경이 이들 세포 내에서 표현되어, 유전자 발현을 인위적으로 조작할 필요가 없으므로, iPS 세포는 생물 의학 연구를 위한 최첨단 세포주로서 부상하였다. CRISPR-Cas9 게놈 편집 기술은 SCN1A의 유전적 돌연변이를 야생형 서열로 수리하거나, 대조군 세포주 내 정상 대립유전자로 드라베-관련 돌연변이를 도입함으로써 동종 세포주를 생성하는데 활용되고 있다. 따라서 동종 세포주는 상이한 인간 대상체의 세포주를 비교하여 발생하는 자연적 가변성을 제거하고 질환 특이적 표현형을 확인하고 증대시키는 데 가치가 있다. 확립된 억제 뉴런 분화 프로토콜 및 검증 파이프라인은 iPS 세포주를 전뇌 GABA성 억제제 중간뉴런으로 분화시키는데 사용되고 있다.
- [0171] 드라베 환자로부터 유래된 억제 뉴런은 전체 세포 패치 클램프 전기생리학 측정에 의해 결정된 바와 같이 감소된 나트륨 전류 손상된 활동 전위 발화를 나타낸다. 본원에 기술된 드라베-유래 뉴런이 이러한 질환 관련 표현형을 개괄한다는 것을 확인하기 위해 유사한 측정을 수행한다. 나트륨 전류 결함이 억제제에서 발생하지만, 드라베 환자의 흥분 뉴런에서는 발생하지 않으므로 (Sun et al) 억제 뉴런만이 본 개시내용에서 활용되고 있다. 드라베 환자-유래된 억제 뉴런에서의 돌연변이 유도된 나트륨 채널 결함은 야생형 SCN1A의 이소성 발현에 의해 구제될 수 있다 (참고 20). 따라서, 본 개시내용에서 기술된 방법은 드라베 증후군의 맥락에서 야생형 나트륨 채널 기능 및 생리학을 복원하는데 ZFP^{SCN1A} 전사활성화제의 효능을 시험하기에 적합하다.
- [0172] GABA성 억제 뉴런 배양물은 범용 뉴런 또는 억제 뉴런 특이적 프로모터 하에서 ZFP^{SCN1A} 전사활성화제를 암호화하는 AAV 벡터로 감염된다. Nav1.1 발현 수준의 변화는 웨스턴 블롯으로 평가된다. 억제 뉴런에서 기능적 나트륨 전류의 복원은 형질감염된 세포에 비해 형질감염되지 않은 세포의 전체 세포 패치 클램핑에 의해 평가된다. 모든 환자-유래된 억제 뉴런의 게놈에 걸친 ZFP^{SCN1A} 전사활성화제의 결합은 ChIP-seq에 의해 분석되고 RNA-seq에 의해 검출되는 임의의 식별된 전사체와 상관관계가 있다. 이 실험에서 대조군은 VPR 전사활성화제 도메인 없이 GFP, ZFP를 암호화하는 유사한 AAV 벡터, 및 ZFP DNA 결합 도메인이 없는 VPR 전사활성화제 도메인으로 감염된 뉴런 배양물이다.
- [0173] 실시예 6. SCN1A 마우스에서 상이한 연령 및 전달 경로에서 AAV-ZFP^{SCN1A} 중재의 치료 가능성 평가

[0174] AAV의 광범위한 지향성은 광범위하게 발현되는 유전자에 대한 유전자 요법 적용을 위한 중요한 특성이지만, 관심 이식유전자가 세포-유형 특이적 방식으로 발현되는 경우 상당한 도전이 될 수 있다. 이는 각각, 티록신 결합 단백질 (TBP), 크레아틴 키나아제 및 트로포닌 T과 같은 조직 특이적 프로모터의 사용을 통해 간, 근육, 및 심장과 같은 신체의 주요 조직에 대해 크게 해결되었다. 이들 조직에서 매우 풍부한 마이크로RNA, 예를 들어 간에서 miR-122 및 골격근에서 miR-1에 대한 결합 부위의 다중 카피의 통합을 통해 특정 조직에서 더 높은 수준의 표적화를 달성하기 위해 조직 특이적 프로모터에 추가 수준의 제어를 중첩할 수 있다. 최근에 기술된 AAV-PHP.B 혈청형은 전신 전달 후 CNS 유전자 전달에 매우 효율적이고, 그것은 광범위한 세포 유형을 형질도입한다. 더욱이 이의 말초 조직에 대한 지향성은 대부분 AAV9 만큼 광범위하다. 드라베 증후군에 대한 유전자 요법 접근법의 목표는 GABA성 억제 중간뉴런에서만 Nav1.1 발현을 복원하면서 다른 뉴런 및 다른 곳에서 이소성 발현으로 인한 해로운 영향을 방지하는 것이다. 복어 (*타키후구 루브립스*) 소마토스타틴 (fSST), 및 뉴로펩티드 Y (fNYP) 유전자로부터 유래된 작은 프로모터 (<2.8 kb) 하에서 GFP를 암호화하는 AAV 및 렌티바이러스 벡터는 두개 내 주입시 마우스 뇌에서 억제 뉴런 특이적 발현을 유도하는 것으로 나타났다. GFP 발현을 유도하는 이러한 프로모터를 보유하는 AAV-PHP.B 벡터를 이식유전자 발현이 보편적인 강한 CAG 프로모터 및 최소의 상대적으로 약한 마우스 MeCP2 프로모터에 의해 유도되는 대조군 벡터와 비교한다. GABA성 억제 중간뉴런에 대한 fSST 및 fNYP 프로모터를 갖는 AAV-PHP.B-GFP 벡터의 특이성은 6주령 (꼬리 정맥) 및 출생 후 1일 (안구 뒤) 마우스, 신생아의 CSF 전달, 그리고 마지막으로 치상회 (DG)를 표적화하는 단측 주입에서의 전신 투여에 의한 CNS로 전달시 연구되고 있다 (표 5). 상이한 연령의 *Scn1a*^{+/-} 마우스가 처리되기 때문에 CNS 유전자 전달의 효율성은 전달 경로에 따라 상당히 다르고, 각각의 전달 경로에 대한 CNS 전체에 걸쳐 GABA성 억제 중간뉴런에 대한 뉴런 형질도입 효율 및 프로모터 특이성의 기준선을 확립하기 위해 광범위한 분석이 수행되고 있다. 짧은 fSST 및 fNYP 프로모터로부터 GFP 발현을 유도하는 AAV 벡터는 이전에 직접 주사 후 해마의 억제 중간뉴런에 대해 매우 특이적인 것으로 나타났다. 본 개시내용의 AAV-PHP.B 벡터는 후속 연구에서와 동일한 방식으로 검증되고, 여기서 치상회 및 과립 세포층의 내벽에 구체적으로 위치하는 *Scn1a*^{+/-} 마우스의 해마 형성에서 억제 뉴런에서 Nav1.1 발현 복원의 치료적 영향이 평가되고 있다 (이론은 아래에 명시됨). 실험은 129SvJ를 Jackson Laboratories (Bar Harbor, ME)로부터 얻은 C57BL/6 마우스와 교배하여 UMMS에서 생산된 129SvJ/C57BL/6 마우스에서 수행되고 있다. 주사 후 1개월에 마우스를 안락사시키고 후 세포 특이적 마커 및 GFP에 대한 항체를 사용하는 이중 면역형광법을 사용하여 형질도입 효율 및 특이성의 조직학적 분석을 위해 뇌 및 척수를 수집한다. GABA성 억제 중간뉴런에 대한 유전자 전달 효율 및 특이성은 글루탐산 탈탄산효소 (GAD; GABA성 뉴런에 대한 마커) 및 GFP에 대한 항체를 사용하는 이중 면역형광 염색에 의해 뇌와 척수 전반에 걸쳐 평가되고 있다. 또한 이들 단백질 및 GFP에 대해 특이적인 항체를 사용하여 소마토스타틴 (SST), 파브알부민 (PV), 칼레티닌 (CR), 혈관활성 장 펩티드 (VIP) 또는 뉴로펩티드 Y (NPY)를 발현하는 억제 중간뉴런의 서브셋에 대한 프로모터 및/또는 AAV-PHP.B의 우선적 특이성을 평가한다. 조직학적으로 GFP 발현을 평가하기 위해 전신 및 ICV 투여로 처리된 마우스로부터 간, 심장 및 골격근을 수집하고 말초 조직에서 이소성 발현 가능성을 결정하기 위해 웨스턴 블롯을 활용한다.

[0175] 표 5. 실험군

전달 경로	코호트당 마우스의 수			
	전신		ICV	IC
연령	6 주*	PND1 #	PND1 #	8 주*
용량 (vg)	2x10 ¹²	4x10 ¹¹	4x10 ¹⁰	1x10 ¹⁰
AAV-fSST-GFP	6	6-8	6-8	4
AAV-fNYP-GFP	6	6-8	6-8	4
AAV-CAG-GFP	6	6-8	6-8	4
AAV-MeCP2-GFP	6	6-8	6-8	-
비히클 (PBS)	2	-	-	-

* 그룹은 두 성별의 동일한 수의 마우스로 구성된다.

벡터당 1 리터 주입

약어: ICV - 뇌실 내 주사; IC - 두개 내 주사; PND1 - 출생 후 1 일

[0176]

[0177] 6주령 *Scn1a*^{+/-} 마우스는 상이한 ZFP^{Scn1a} 전사활성화제 단백질을 암호화하는 AAV-PHP.B 벡터의 치상회로, 활성화제 단독, 또는 동일한 부피의 인산염 완충 식염수 (PBS)의 영향에 대한 제어를 위한 ZFP^{Scn1a} 활성화 도메인을 갖지만 DNA-결합 도메인이 없는 작제물을 양측 주사 투여한다 (n=3 수컷 + 3 암컷/군). 이 실험에 사용된 단일 가닥 AAV 벡터는 또한 형질도입 세포의 식별을 용이하게 하는 ZFP^{Scn1a} cDNA의 IRES-GFP 카세트 다운스트림을 보유

한다. 적어도 2개의 ZFP^{Scn1a} 전사활성화제를 시험하고, 이는 다양한 뉴런, 뿐만 아니라 상기 기술된 2개의 가장 유망한 GABA성 억제 뉴런 제한된 ZFP^{Scn1a} 전사활성화제에서 더 넓은 활성화가 있을 수 있다. 주사 후 1개월에, 뇌를 수확하고 로딩 대조군으로서 베타-액틴 또는 튜불린을 사용하는 웨스턴 블롯에 의해 ZFP^{Scn1a}, Nav1.1, Nav1.3, GAD65, GAD67 단백질의 발현 수준을 평가하기 위해 뇌의 반구에서 해마를 해부한다. 다른 뇌 반구를 GAD 및 GFP, 또는 GAD에 대한 항체 및 모든 ZFP^{Scn1a} 단백질에 포함된 에피토프 태그 (HA 또는 myc 태그)를 사용하는 이중 면역 형광 염색에 의해 치상회 및 과립 세포층의 내부 소엽에서 % 형질도입된 억제 중간뉴런을 분석하기 위해 연속 뇌 절편 (10 μm)을 사용하는 조직학적 연구에 의해 조사한다. 또한, 나트륨 채널 발현의 정상 패턴의 회복을 입증하기 위해 Nav1.1 및 Nav1.3을 발현하는 GAD-양성 뉴런의 백분율을 측정한다. Nav1.1 및 Nav1.3 단백질 발현의 면역 형광 검출 이외에, GABA성 중간뉴런의 mRNA 수준의 변화를 Nav1.1, Nav1.3, ZFP^{Scn1a} 및 GAD에 대한 RNAScope 프로브를 사용하여 평가한다. RNAScope는 뇌의 뉴런에서 mRNA 수준을 분석하는 매우 민감한 제자리 혼성화 기술이다. ZFP^{Scn1a} 발현이 유발하는 Nav1.1 수준의 변화를 평가하기 위한 이들 두 가지 접근법의 조합은 중간뉴런에서 본 개시내용의 유전자 요법 접근법에 의해 변화가 달성되는 방법의 포괄적인 이해를 제공한다.

[0178] AAV-PHP.B-ZFP^{Scn1a} 유전자 요법의 치료적 효능은 출생 후 1일, 또는 꼬리 정맥을 통해 6주령에 개시된 두 성별의 *Scn1a*^{+/-} 마우스에서 분석된다. 대조군은 ZFP DNA-결합 도메인 없이 ZFP 유사 단백질을 암호화하는 AAV 벡터로 처리된 마우스, 뿐만 아니라 연령 일치된 처리되지 않은 *Scn1a*^{+/-} 마우스 및 야생형 한배 새끼 (그룹당 n=15 수컷 및 15 암컷)을 포함한다. 각 그룹의 마우스 서브셋 (n=3 수컷 및 3 암컷)은 12주령에 안락사시키고 웨스턴 블롯뿐만 아니라 GAD (및 다른 뉴런 유형 특이적 마커, 예를 들어, GAD65, GAD67) 및 ZFP에 대한 항체를 사용하는 면역 형광, 및 뇌 및 척수에 걸친 이들 세포에서 Nav1.1 발현의 복원을 사용하여 GABA성 중간뉴런에 대한 유전자 전달 효율을 평가한다. 또한, 말초 조직에서 Nav1.1 발현과 함께, ZFP의 이소성 발현이 평가된다. 각 그룹의 동물의 다른 서브셋 (n=24)은 생존 (최대 1세), 운동 수행 및 행동에 대한 영향을 연구하기 위해 사용되고, 이는 생후 2-12개월에서 2개월마다 시험된다. *Scn1a*^{+/-} 마우스가 PND21에 의해 앞다리와 뒷다리의 손상된 협응을 나타내기 때문에, 운동 기능 및 조정을 가속 로타로드 및 빔 교차 검사를 사용하여 평가한다. 또한, *Scn1a*^{+/-} 마우스가 손상된 수행을 나타낼 때, *Scn1a*^{+/-} 마우스에서 심각하게 손상된 것으로 보이는 공간 학습 및 기억을 시험하는 개방 필드, 고위 플러스 미로, 등지 건설, 구슬 매립, 및 반스 미로를 포함하는 행동 검사가 활용된다. 드라베 증후군 환자의 자발적인 발작 특성이 또한 *Scn1a*^{+/-} 마우스에서 명백하고 연령 및 체중에 따라 빈도가 증가한다. 더욱이 *Scn1a*^{+/-} 마우스의 조기 돌연사가 강직-간대성 발작 직후에 발생한다. 따라서 발작 빈도 및 지속기간을 평가하기 2, 6 및 12개월령에 24시간 동안 지속적인 비디오 모니터링을 활용한다. 상기 기술된 검사에서 측정된 1차 결과에서 유의한 변화가 검출되면 새로운 물체, 냄새 및 마우스에 대한 반응으로 챔버 선호도 판독값을 사용하는 사회적 상호 작용 연구가 고려된다. 위에 설명된 분자 및 조직학적 분석을 수행하기 위해 인도적 종말점 실험에서 뇌, 척수, 및 말초 기관을 수집하고 평가한다.

[0179] 실시예 7. 인간 세포에서 SCN1A 유전자 발현을 증가시키는 ZFP 및 dCas9 시스템

[0180] SCN1A의 전사를 상향 조절하는 ZFP1-ZFP3의 능력을 조사하기 위해, ZFP1-ZFP3 DNA 결합 도메인은 키메라 전사활성화제를 형성하기 위해 하이브리드 VP64, p53, 및 RTA (VPR) 3부 강한 전사 활성화제 도메인에 융합된다. VPR 융합 활성화제 도메인은 전사 조절 복합체를 모집하고 염색질 접근성을 증가시키고 높은 수준의 유전자 발현을 달성하기 위해 돕는다. 따라서, ZFP 도메인은 SCN1A 유전자 발현을 증가시키기 위해 근위 프로모터 영역에서 VPR 활성화제를 고도로 보존된 서열에 대해 표적화할 것이다.

[0181] 추가로, SCN1A의 전사를 상향 조절하기 위해 SCN1A를 표적으로 하는 dCas9 시스템의 능력을 조사하기 위해, SCN1A를 표적으로 하는 3개의 가이드 RNA를 dCas9 단백질과 복합체화하였다.

[0182] HEK293T 세포를 일시적으로 다음의 실험 조건 중 하나로 감염시켰다 - (1) VPR-ZFP1 작제물; (2) VPR-ZFP2 작제물; (3) VPR-ZFP3 작제물; (4) VPR-ZFP1, VPR-ZFP2, 및 VPR-ZFP3 작제물 세 가지 모두; (5) dCas9-VPR 작제물 및 SCN1A 가이드 RNA 1; (6) dCas9-VPR 작제물 및 SCN1A 가이드 RNA 2; (7) dCas9-VPR 작제물 및 SCN1A 가

이드 RNA 3; (8) dCas9-VPR 작제물 및 SCN1A 가이드 RNA 1, SCN1A 가이드 RNA 2, 및 SCN1A 가이드 RNA 3 세 가지 모두; 및 (9) 임의의 가이드 RNA가 없는 dCas9-VPR 작제물 (대조군). SCN1A 유전자 발현을 qRT-PCR로 평가하였다. SCN1A의 배수 활성화를 대조군 실험으로 정규화하였다 (임의의 가이드 RNA가 없는 dCas9-VPR 작제물).

[0183] 모든 시험된 실험 조건은 대조군 실험에 비해 SCN1A의 유전자 활성화 증가를 발생시켰다 (도 8). 이들 데이터는 본 실시예 및 본 개시내용 전반에 걸쳐 기술된 징크 핑거 단백질이 유전자 발현에 영향을 미치도록 SCN1A를 표적화할 수 있음을 입증한다. 이들 데이터는 추가로 본 실시예의 가이드 RNA 서열 (서열 번호 83-94)이 유전자 발현에 영향을 주기 위해 dCas9를 SCN1A에 대해 표적화할 수 있음을 입증한다.

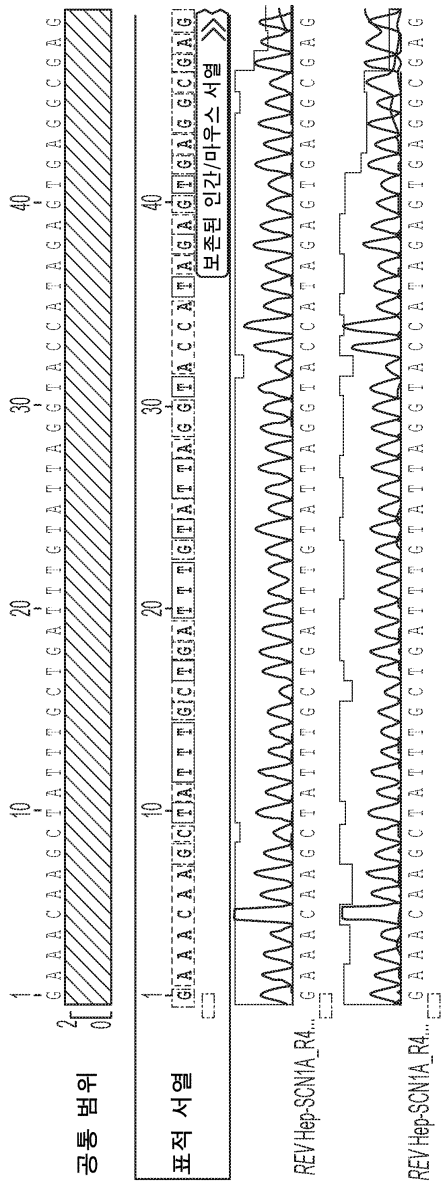
[0184] 표 6. SCN1A를 표적으로 하는 가이드 핵산 (굵은 글씨 스페이서 서열)

	뉴클레오티드 서열 (DNA)	뉴클레오티드 서열 (RNA)
SCN1A 가이드 1	GAGGTACC ATAGAGTGAGGCG GTTTTAGAGCTAGAAATAGCA AGTTAAAATAAGGCTAGTCCG TTATCAACTTGAAAAAGTGGC ACCGAGTCGGTGC (서열 번호 83)	GAGGUACCAU AGAGUGAGGCGGUU UUAGAGCUAGAAAUAGCAAGUAA AAUAAGGCUAGUCCGUUAUCAACU UGAAAAAGUGGCACCGAGUCGGUG C (서열 번호 84)
SCN1A 가이드 2	ACCGAGGC GAGGATGAAGCCG AGG TTTTAGAGCTAGAAATAG CAAGTTAAAATAAGGCTAGTC CGTTATCAACTTGAAAAAGTG GCACCGAGTCGGTGC (서열 번호 87)	ACCGAGGC GAGGAUGAAGCCGAGG UUUUAGAGCUAGAAAUAGCAAGUU AAAAUAAGGCUAGUCCGUUAUCAA CUUGAAAAAGUGGCACCGAGUCGG UGC (서열 번호 88)
SCN1A 가이드 3	ACCGAAGCC GAGAGGATACTG CAG GTTTTAGAGCTAGAAATA GCAAGTTAAAATAAGGCTAGT CCGTTATCAACTTGAAAAAGT GGCACCGAGTCGGTGC (서열 번호 91)	ACCGAAGCC GAGAGGAUACUGCAG GUUUUAGAGCUAGAAAUAGCAAGU UAAAAUAAGGCUAGUCCGUUAUCA ACUUGAAAAAGUGGCACCGAGUCG GUGC (서열 번호 92)

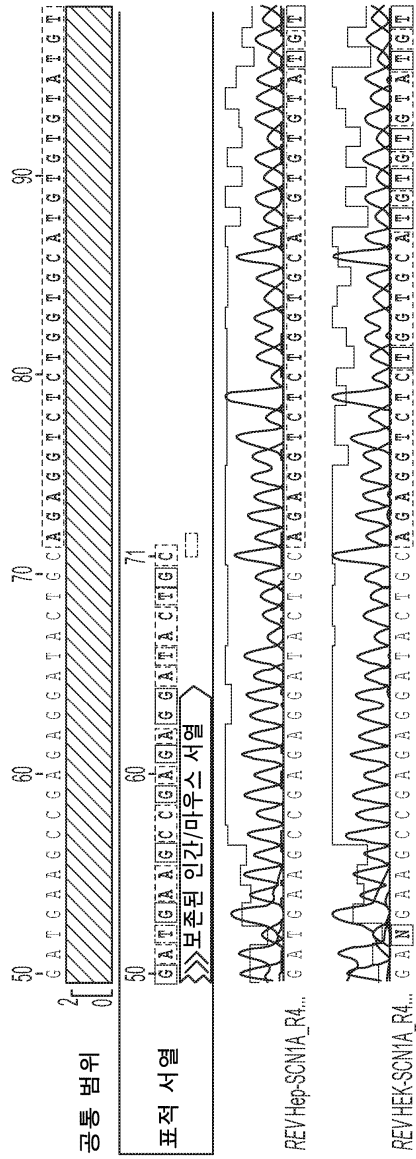
[0185]

도면

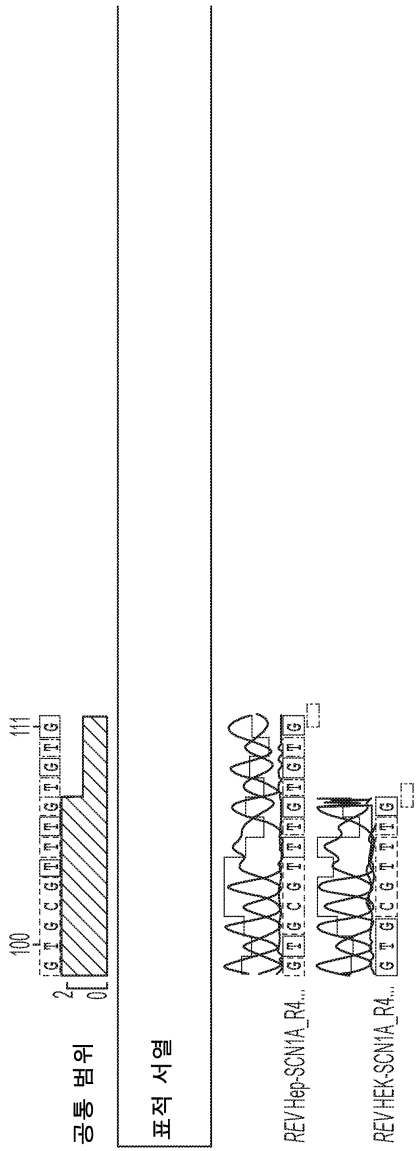
도면1a



도면1b



도면1c



도면6b

F1 GTC DRSLAR
 F2 GAG RSDNLTR
 F3 GCA QSGDLTR
 F4 ACT VRQTLKQ
 F5 GAT AAGNLTR
 F6 GAG RSDNLTR

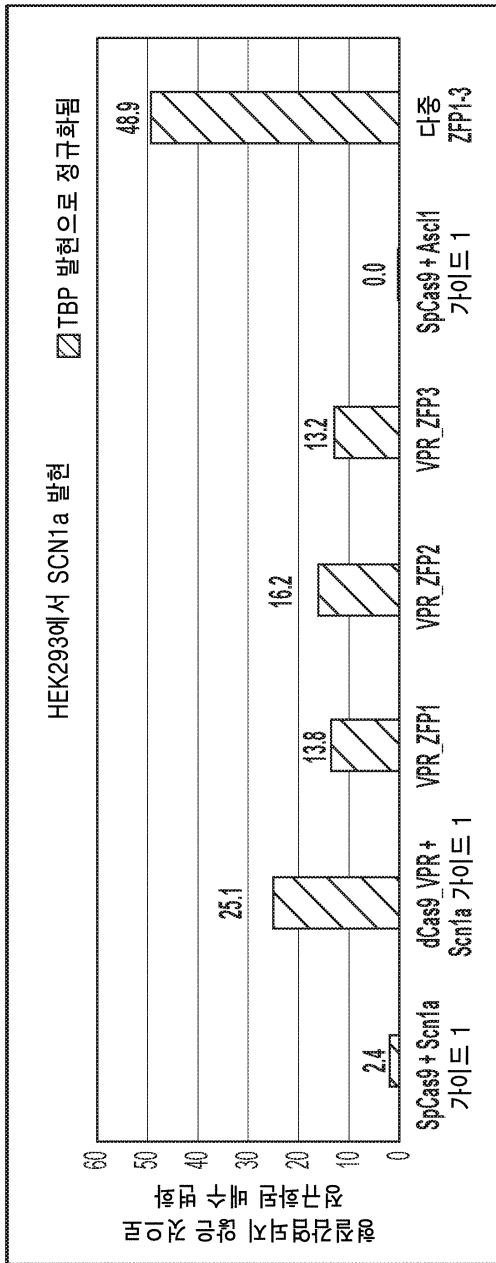
도면6c

RP
 FQCRICMRNFS DRSLAR HIRTHTGEKP
 FACDICGKKFA RSDNLTR HTKIHTGSQKP
 FQCRICMRNFS QSGDLTR HIRTHTGEKP
 FACDICGKKFA VRQTLKQ HTKIHTGSQKP
 FQCRICMRNFS AAGNLTR HIRTHTGEKP
 FACDICGRKFA RSDNLTR HTKIHLRQKD

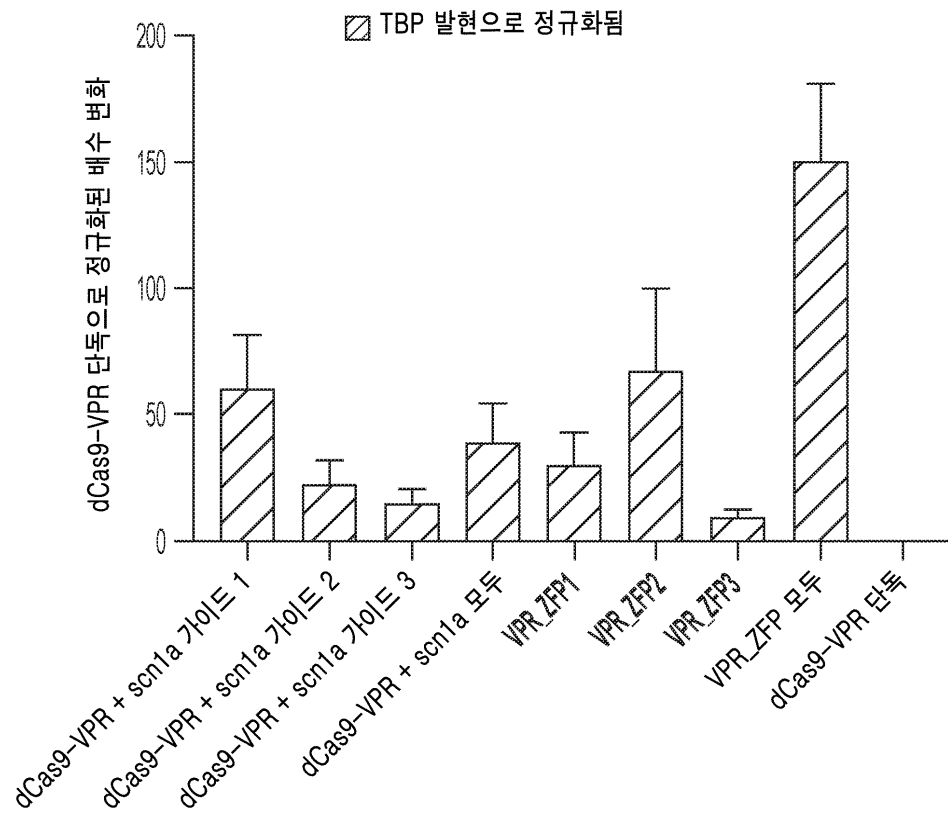
도면6d

CGACCA
 TTCCAGTGTGGAATCTGCATCGGCACTTCAGC GAC CGG AGC GCG CTG GCA CGG CATATCCGCACCCACCCGGGAGAGAGCCCT
 TTTGCCTGGCATATTTCTGGAAGCAGTTTGCTH CGA AGT GAC AAC TTA ACG CGC CACACCAAGATTCATACTGGGTCCCAGAAACCG
 TTCCAGTGTAGGATATGCATGAGCAATTTCTCTH CAG TCA GGG GAC CTC ACT CGT CATATAAGGACGCACACAGCTCAAAAACCA
 TTTGCATGCCACATCTGTGGCAAAAAGTTTGGH GTA CGA CAG ACG CTT AAA CAA CACACAAAAATCCATAACGGCAGTCAAAAGCCC
 TTTCAATGTCCATTTCATGCGAAACTTTCAM GCC GCT GGT AAC TTG ACA CGA CATATTGCTACTCATACTGGCGAGAAACCT
 TTTCCCTTCCATATATGTGGTCCTAAGTTTGCAM AGA TCT GAT AAT CTA ACG CGT CACACTAAGATACACTCCGGCAGAGGAC

도면7



도면8



서열 목록

SEQUENCE LISTING

<110> Univeristy of Massachusetts

<120> ZINC FINGER PROTEIN TRANSACTIVATORS AND USES THEREOF

<130> U0120.70106W000

<140> Not Yet Assigned

<141> Concurrently Herewith

<150> US 62/810,005

<151> 2019-02-25

<160> 121

<170> PatentIn version 3.5

<210> 1

<211> 672

<212> DNA

<213> Homo sapiens

<400> 1

aatttccatg gactcttttt ccaaaggaat aactggaatg aataaactta aaatcaagat 60
gaaacaatta gatggcttac ctgattaaaa ggaaaattat ccatctgcag tgaggaacag 120

catcacccaa agacgagatg ataacaatgt gccttcagtt gcaattgttc agttccttct 180
tgcaaaaggt gtcaaagtat ttacaagggc tgcagtctca ctggggcaga acacacagac 240
acacaaacac acacaaacgc acacatacac acatgcacca gagacctctg cagtatcctc 300
tcggcttcat cctcgcctca ctctatggia cctaatacaa atcagcaaat agcttgtttc 360
aaaaaaaaaaa aaaagtcaag acagcacctt acattacatc gccatctagt ggctaaatat 420
taaacacttt ctcaaatcc agatttatga tttcttctc aacctctttt ctctcagctt 480
ttttcctttc ttctctgtaa tctcccagta ttgcttctcc ttgcttctct ttcattcctt 540

attgctatat aatatcatga acctaatac tcaaagagga aaaggtttga aagtaaatat 600
agctattttc aagtagtact tgaaaaactt agcattattt tagtttgaaa ctgttacttt 660
attcctaata tg 672

<210> 2
<211> 669
<212> DNA
<213> Mus musculus
<400> 2

tatttccgtg ggctcttctc cccaaggatt taccaggtaa gaattcacca ccaaagaaga 60
tcacaatgag ataatcagat ggcttacctg ataaaaagga aaattatcca tctgcagtea 120
ggagcaacat ctccccagca cgagtcgca ccttccgttg caacgattca gattccttct 180

tgcaaaaggt gaccaagtgc ttcacaaggg ctgcagcctc ataggggaga acacacgtac 240
acaaacacac gcacacacac acacacatgc accagagacc tctgcagtat cctctggctt 300
catcctcgcc tactctatg gtacctata caaatcagca aatagcttgt tttaaaaaaa 360
agaaagaaaa aaagcggaga cagcacctaa cgttacagtg ccatctagt gctacatcgt 420
aaataggttc tcacagcctg gatttctgtg ttctttctca accgcttctt tctggttcct 480
ttttcttttt tctctttat tttggtttta ttacttctc agatgccttt ttttcattcc 540
cctttgctct gctacatgg aactattgac ttaaagatta aaacaatcag aactggagag 600

cgttgctttt aagttaaaaa aaaaagggtt gctaattttg tttgtaaag ttactttatt 660
ttctctatt 669

<210> 3
<211> 130

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 3
 tttttttttt ttttttgaa acaagctatt tgctgattg tattaggtac catagagtga 60
 ggcgaggatg aagccgagag gatactgcag aggtctctgg tgcattgtg tatgtgtgcg 120
 tttgtgtgtg 130
 <210> 4
 <211> 41
 <212>
 > DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 4
 gaggtagggc aggatgaagc cgagaggata ctgcagaggt c 41
 <210> 5
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 5
 gaggtagggc aggatgaa 18
 <210> 6
 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 6
 ggcgaggatg aagccgag 18
 <210> 7

 <211> 18
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic

<400> 7

gaggatactg cagaggtc

18

<210> 8

<211> 5

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 8

Glu Gly Glu Asp Glu

1 5

<210> 9

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 9

Gly Glu Asp Glu Ala Glu

1 5

<210> 10

<211> 6

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 10

Glu Asp Thr Ala Glu Val

1 5

<210> 11

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 11

cagcgggaa acctggtgag g

21

<210> 12

<211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 12
 ctgagettca atctaaccag a 21
 <210> 13
 <211> 21
 <212> DNA
 <213> Artificial Sequence

 <220><223> Synthetic
 <400> 13
 cggagtgaca acttaacgcg g 21
 <210> 14
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 14
 gaccggtctc accttgcccg a 21
 <210> 15
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 15
 cagaaggccc attgactgc c 21
 <210> 16
 <211> 21
 <212> DNA
 <213>
 Artificial Sequence
 <220><223> Synthetic
 <400> 16

cggtcggaca acctcacacg c

<210> 17

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 17

Gln Arg Gly Asn Leu Val Arg

1 5

<210> 18

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 18

Leu Ser Phe Asn Leu Thr Arg

1 5

<210> 19

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 19

Arg Ser Asp Asn Leu Thr Arg

1 5

<210> 20

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 20

Asp Arg Ser His Leu Ala Arg

1 5

<210> 21

<220><223> Synthetic

<400> 25
cagcggggaa atctagtgcg a 21

<210> 26
<211> 21
<212> DNA
<213> Artificial Sequence

<220><223> Synthetic

<400> 26
ctgagcttca acttgactcg t 21

<210> 27
<211> 21
<212> DNA
<213> Artificial Sequence

<220><223> Synthetic

<400> 27
cggagtgaca atcttacgag a 21

<210> 28
<211> 21
<212> DNA
<213> Artificial Sequence

<220><223> Synthetic

<400> 28
gaccggagcc acttagccag g 21

<210> 29
<211> 7
<212> PRT
<213> Artificial Sequence

<220><223> Synthetic

<400> 29
Arg Ser Ser Asn Leu Thr Arg
1 5

<210> 30

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 30

Asp Lys Arg Thr Leu Ile Arg

1 5

<210> 31

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 31

Gln Arg Gly Asn Leu Val Arg

1 5

<210> 32

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 32

Leu Ser Phe Asn Leu Thr Arg

1 5

<210> 33

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 33

Arg Ser Asp Asn Leu Thr Arg

1 5

<210> 34

<211> 7

<212> PRT

<213> Artificial Sequence
 <220><223> Synthetic
 <400> 34
 Asp Arg Ser His Leu Ala Arg

1 5

<210> 35

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 35

gaccggagcg cgctggcacg g 21

<210> 36

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 36

cgaagtgaca acttaacgcg c 21

<210> 37

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 37

cagtcagggg acctcactcg t 21

<210> 38

<211> 21

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 38

gtacgacaga cgcttaaaca a 21

<210> 39
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 39
 gccgctggta acttgacacg a 21
 <210> 40
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 40
 agatctgata atctaacgcg t 21

 <210> 41
 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 41
 Asp Arg Ser Ala Leu Ala Arg
 1 5
 <210> 42
 <211> 7
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 42
 Arg Ser Asp Asn Leu Thr Arg
 1 5
 <210> 43
 <211> 7
 <212> PRT
 <213> Artificial Sequence

<220><223> Synthetic

<400> 43

Gln Ser Gly Asp Leu Thr Arg

1 5

<210> 44

<211> 7

<212

> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 44

Val Arg Gln Thr Leu Lys Gln

1 5

<210> 45

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 45

Ala Ala Gly Asn Leu Thr Arg

1 5

<210> 46

<211> 7

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 46

Arg Ser Asp Asn Leu Thr Arg

1 5

<210> 47

<211> 1569

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 47

gaggccagcg gttccggacg ggctgacgca ttggacgatt ttgatctgga tatgctggga 60
 agtgacgccc tcgatgattt tgaccttgac atgcttggtt cggatgccct tgatgacttt 120
 gacctcgaca tgctcggcag tgacgccctt gatgatttcg acctggacat gctgattaac 180
 tctagaagtt ccgatctag ccagtacctg cccgacaccg acgaccggca ccgatcgag 240
 gaaaagcgga agcggaccta cgagacattc aagagcatca tgaagaagtc ccccttcagc 300
 ggccccaccg accctagacc tccacctaga agaatcgccg tgcccagcag atccagcgcc 360
 agcgtgccaa aacctgcccc ccagccttac cctttacca gcagcctgag caccatcaac 420

 tacgacgagt tcctaccat ggtgttcccc agcggccaga tctctcaggc ctctgctctg 480
 gctccagccc ctctcaggt gctgcctcag gctcctgtc ctgcaccagc tccagccatg 540
 gtgtctgcac tggtcaggc accagcacc gtgcctgtgc tggtcctgg acctccacag 600
 gctgtggctc caccagcccc taaacctaca caggccggcg agggcacact gctgaagct 660
 ctgctgcagc tgcagttcga cgacaggat ctgggagccc tgctgggaaa cagcaccgat 720
 cctgccgtgt tcaccgacct ggccagcgtg gacaacagcg agttccagca gctgctgaac 780
 cagggcattc ctgtggcccc tcacaccacc gagcccatgc tgatggaata ccccagggcc 840

 atcaccggcg tcgtgacagg cgctcagagg cctcctgatc cagctcctgc ccctctggga 900
 gcaccaggcc tgctaattgg actgctgtct ggcgacgagg acttcagctc tatcgccgat 960
 atggatttct cagccttctt gggctctggc agcggcagcc gggattccag ggaagggatg 1020
 tttttgccga agcctgaggc cggtccgct attagtacg tgtttgaggg ccgcgaggtg 1080
 tgccagccaa aacgaatccg gccatttcat cctccaggaa gtccatgggc caaccgcca 1140
 ctccccgcca gcctgcacc aacaccaacc ggtccagtac atgagccagt cgggtcactg 1200
 accccggcac cagtccctca gccactggat ccagcggccg cagtgactcc cgaggccagt 1260

 cacctgttgg aggatcccga tgaagagacg agccaggctg tcaaagccct tcgggagatg 1320
 gccgatactg tgattcccca gaaggaagag gctgcaatct gtggccaaat ggacctttcc 1380
 catccgcccc caaggggcca tctggatgag ctgacaacca cacttgagtc catgaccgag 1440
 gatctgaacc tggactcacc cctgaccccc gaattgaacg agattctgga taccttctg 1500
 aacgacgagt gcctcttgca tgccatgcat atcagcacag gactgtccat ctctgacaca 1560
 tctctgttt 1569

<210> 48

<211> 523

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 48

Glu Ala Ser Gly Ser Gly Arg Ala Asp Ala Leu Asp Asp Phe Asp Leu
 1 5 10 15
 Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu
 20 25 30
 Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Ser Asp
 35 40 45
 Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Ile Asn Ser Arg Ser Ser
 50 55 60

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

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

Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala Pro Lys
 195 200 205
 Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu Gln Leu
 210 215 220

Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser Thr Asp
 225 230 235 240

Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu Phe Gln
 245 250 255

Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr Glu Pro
 260 265 270

Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr Gly Ala
 275 280 285

Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly Leu
 290 295 300

Pro Asn Gly Leu Leu Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala Asp
 305 310 315 320

Met Asp Phe Ser Ala Leu Leu Gly Ser Gly Ser Gly Ser Arg Asp Ser
 325 330 335

Arg Glu Gly Met Phe Leu Pro Lys Pro Glu Ala Gly Ser Ala Ile Ser
 340 345 350

Asp Val Phe Glu Gly Arg Glu Val Cys Gln Pro Lys Arg Ile Arg Pro
 355 360 365

Phe His Pro Pro Gly Ser Pro Trp Ala Asn Arg Pro Leu Pro Ala Ser
 370 375 380

Leu Ala Pro Thr Pro Thr Gly Pro Val His Glu Pro Val Gly Ser Leu
 385 390 395 400

Thr Pro Ala Pro Val Pro Gln Pro Leu Asp Pro Ala Pro Ala Val Thr
 405 410 415

Pro Glu Ala Ser His Leu Leu Glu Asp Pro Asp Glu Glu Thr Ser Gln
 420 425 430

Ala Val Lys Ala Leu Arg Glu Met Ala Asp Thr Val Ile Pro Gln Lys
 435 440 445

Glu Glu Ala Ala Ile Cys Gly Gln Met Asp Leu Ser His Pro Pro Pro
 450 455 460

Arg Gly His Leu Asp Glu Leu Thr Thr Thr Leu Glu Ser Met Thr Glu

caggatTTTT gggaaaacct gtatcagctg accctgcgcg cggcgggcaa aacatatatg 1200
 atTTTTTTtg tgctggatgat ttttctgggc agcttttatac tgattaacct gattctggcg 1260
 gtggtggcga tggcgtatga agaacagaac caggcgaccc tggaagaagc ggaacagaaa 1320
 gaagcggaat ttcagcagat gattgaacag ctgaaaaaac agcaggaagc ggcgagcagc 1380
 gcggcgaccc cgaccgcgag cgaacatagc cgcgaaccga gcgcggcggg ccgcctgagc 1440
 gatagcagca gcgaagcgag caaactgagc agcaaaagcg cgaaagaagc ccgcaaccgc 1500
 cgcaaaaaac gcaaacagaa agaacagagc ggcggcgaag aaaaagatga agatgaattt 1560

 cagaaaagcg aaagcgaaga tagcattcgc cgcaaaggct ttcgctttag cattgaaggc 1620
 aaccgcctga cciatgaaaa acgctatagc agcccgcatac agagcctgct gagcattcgc 1680
 ggcagcctgt ttagcccgcg ccgcaacagc cgcaccagcc tgtttagctt tcgcggccgc 1740
 gcgaaagatg tgggcagcga aaacgatttt gcggatgatg aacatagcac ctttgaagat 1800
 aacgaaagcc gccgcgatag cctgtttgtg ccgcgccgcc atggcgaacg ccgcaacagc 1860
 aacctgagcc agaccagccg cagcagccgc atgctggcgg tgtttccggc gaacggcaaa 1920
 atgcatagca ccgtggattg caacggcgtg gtgagcctgg tgggcggccc gagcgtgccg 1980

 accagcccgg tgggccagct gctgccgga gtgattattg ataaaccggc gaccgatgat 2040
 aacggcacca ccaccgaac cgaaatgccc aaacgccga gcagcagctt tcatgtgagc 2100
 atggattttc tggaagatcc gagccagcgc cagcgcgca tgagcattgc gagcattctg 2160
 accaacaccg tggaagaact ggaagaaagc cgccagaaat gcccgccgtg ctggtataaa 2220
 tttagcaaca tttttctgat ttgggattgc agcccgtatt ggctgaaagt gaaacatgtg 2280
 gtgaacctgg tggatgatga tccgtttgtg gatctggcga ttaccatttg cattgtgctg 2340
 aacacctgt ttatggcgat ggaacattat ccgatgaccg atcattttaa caacgtgctg 2400

 accgtgggca acctggtgtt taccggcatt tttaccgagg aatgtttct gaaaattatt 2460
 gcgatggatc cglattatta ttttcaggaa ggctggaaca tttttgatgg ctttattgtg 2520
 acctgagcc tggatggaact gggcctggcg aacgtggaag gcctgagcgt gctgcgcagc 2580
 tttgcctgc tgcgcgtgtt taaactggcg aaaagctggc cgaccctgaa catgctgatt 2640
 aaaattattg gcaacagcgt gggcgcgctg ggcaacctga ccctggtgct ggcgattatt 2700
 gtgtttattt ttgcggtggt gggcatgcag ctgtttggca aaagctataa agattgcctg 2760
 tgcaaaattg cgagcgattg ccagctgccg cgctggcata tgaacgattt ttttcatagc 2820

 tttctgattg tgtttcgctg gctgtgcggc gaatggattg aaaccatgtg ggattgcatg 2880
 gaagtggcgg gccaggcgat gtgctgacc gtgtttatga tggatgatgt gattggcaac 2940
 ctggtggtgc tgaacctgtt tctggcgtg ctgctgagca gctttagcgc ggataacctg 3000

gcggcgaccg atgatgataa cgaaatgaac aacctgcaga ttgcggtgga tcgcatgcat 3060
 aaaggcgtgg cgtatgtgaa acgcaaaatt tatgaattta ttcagcagag ctttattcgc 3120
 aaacagaaaa ttctggatga aattaaaccg ctggatgata tgaacaaca aaaagatagc 3180
 tgcattgagca accataaccg ggaaattggc aaagatctgg attatctgaa agatgtgaac 3240

 ggcaccacca gcggcattgg caccggcagc agcgtggaaa aatatattat tgatgaaagc 3300
 gattatata gctttattaa caaccggagc ctgaccgtga ccgtgccgat tgcggtgggc 3360
 gaaagcgatt ttgaaaacct gaacaccgaa gatttttagca gcgaaagcga tctggaagaa 3420
 agcaaagaaa aactgaacga aagcagcagc agcagcgaag gcagcaccgt ggatattggc 3480
 gcgccggtgg aagaacagcc ggtggtggaa ccggaagaaa ccctggaacc ggaagcgtgc 3540
 ttaccgaag gctgcgtgca gcgctttaa tgctgccaga ttaacgtgga agaaggccgc 3600
 ggcaaacagt ggtggaacct gcgccgacc tgctttcgca ttgtggaaca taactggttt 3660

 gaaaccttta ttgtgtttat gattctgctg agcagcggcg cgctggcgtt tgaagatatt 3720
 tatattgac agcgcaaac cattaaaacc atgctggaat atgcggataa agtgtttacc 3780
 tatattttta ttctggaat gctgctgaaa tgggtggcgt atggctatca gacctattt 3840
 accaacgctt ggtgctgctt ggattttctg attgtggatg tgagcctggt gaccctgacc 3900
 gcgaacgcgc tgggctatag cgaactgggc gcgattaaaa gcctgcgcac cctgcgcgcg 3960
 ctgcgccgcg tgcgcgcgct gagecccttt gaaggcatgc gcgtggtggt gaacgcgctg 4020
 ctgggcgcga ttccgagcat tatgaacgtg ctgctggtgt gcctgatttt ttggctgatt 4080

 tttagcatta tggcggtgaa cctgtttgcg ggcaaatTTT atcattgcat taacaccacc 4140
 accggcgatc gctttgatat tgaagatgtg aacaaccata ccgattgcct gaaactgatt 4200
 gaacgcaacg aaaccgcgcg ctggaaaaaac gtgaaagtga actttgataa cgtgggcttt 4260
 ggctatctga gcctgctgca ggtggcgacc tttaaaggct ggatggatat tatgtatgcg 4320
 gcggtggata gccgcaactt ggaactgcag ccgaaatatg aagaaagcct gtatatgtat 4380
 ctgtatTTTg tgatttttat tatttttggc agctttttta ccctgaacct gtttattggc 4440
 gtgattattg ataactttaa ccagcagaaa aaaaaattt gcggccagga tatttttatg 4500

 accgaagaac agaaaaata ttataacgcg atgaaaaaac tgggcagcaa aaaaccgcag 4560
 aaaccgattc cgcgcccggg caacaaattt cagggcatgg tgtttgattt tgtgaccgcg 4620
 caggtgtttg atattagcat tatgattctg atttgcctga acatggtgac catgatggtg 4680
 gaaaccgatg atcagagcga atatgtgacc accattctga gccgattaa cctggtgttt 4740
 attgtgctgt ttaccggcga atgcgtgctg aaactgatta gcctgcgcca ttattatTTT 4800
 accattggct ggaacatttt tgattttgtg gtggtgattc tgagcattgt gggcatgttt 4860

ctggcgaac tgattgaaaa atattttgtg agcccgacc tgtttcgcgt gattcgctg 4920

gcgcgattg gccgattct gcgcctgatt aaaggcgcga aaggcattcg caccctgctg 4980

tttgcgctga tgatgagcct gccggcgtg ttaacattg gcctgctgct gtttctggtg 5040

atgtttatit atgcgatttt tggcatgagc aactttgcgt atgtgaaacg cgaagtgggc 5100

attgatgata tgtttaactt tgaaccttt ggcaacagca tgattgcct gtttcagatt 5160

accaccagcg cgggctggga tggcctgctg gcgccgattc tgaacagcaa accgccgat 5220

tgcatccga acaaagtga cccgggcagc agcgtgaaag gcgattgcgg caaccgagc 5280

gtgggcattt tttttttgt gagctatatt attattagct ttctggtggt ggtgaacatg 5340

tatattgcgg tgattctgga aaactttagc gtggcgaccg aagaaagcgc ggaaccgctg 5400

agcgaagatg attttgaat gttttatgaa gtgtgggaaa aatttgatcc ggatgcgacc 5460

cagtttatgg aatttgaana actgagccag tttgcggcgg cgctggaacc gccgctgaac 5520

ctgccgcagc cgaacaaact gcagctgatt gcgatggatc tgccgatggt gagcggcgat 5580

cgattcatt gcctggatat tctgtttgcg ttaccacaaac gcgtgctggg cgaagcggc 5640

gaaatggatg cgctgcgcat tcagatggaa gaacgcttta tggcgagcaa cccgagcaaa 5700

gtgagctatc agccgattac caccacctg aaacgcaaac aggaagaagt gagcgcggtg 5760

attattcagc gcgcgatcg ccgccatctg ctgaaacgca ccgtgaaaca ggcgagcttt 5820

acctataaca aaaacaaaat taaaggcggc gcgaacctgc tgattaaaga agatatgatt 5880

attgatcgca ttaacgaaaa cagcattacc gaaaaaacg atctgacat gagcaccgcg 5940

gcgtgccgcg cgagctatga tcgcgtgacc aaaccgattg tggaaaaaca tgaacaggaa 6000

ggcaaagatg aaaaagcgaaggcaaa 6027

<210> 50

<211> 2009

<212> PRT

<213> Homo sapiens

<400> 50

Met Glu Gln Thr Val Leu Val Pro Pro Gly Pro Asp Ser Phe Asn Phe

1 5 10 15

Phe Thr Arg Glu Ser Leu Ala Ala Ile Glu Arg Arg Ile Ala Glu Glu

 20 25 30

Lys Ala Lys Asn Pro Lys Pro Asp Lys Lys Asp Asp Asp Glu Asn Gly

35 40 45

Pro Lys Pro Asn Ser Asp Leu Glu Ala Gly Lys Asn Leu Pro Phe Ile
 50 55 60
 Tyr Gly Asp Ile Pro Pro Glu Met Val Ser Glu Pro Leu Glu Asp Leu

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

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

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

 260 265 270
 Leu Arg Asn Lys Cys Ile Gln Trp Pro Pro Thr Asn Ala Ser Leu Glu
 275 280 285
 Glu His Ser Ile Glu Lys Asn Ile Thr Val Asn Tyr Asn Gly Thr Leu

Tyr Glu Lys Arg Tyr Ser Ser Pro His Gln Ser Leu Leu Ser Ile Arg
 545 550 555 560
 Gly Ser Leu Phe Ser Pro Arg Arg Asn Ser Arg Thr Ser Leu Phe Ser
 565 570 575
 Phe Arg Gly Arg Ala Lys Asp Val Gly Ser Glu Asn Asp Phe Ala Asp
 580 585 590
 Asp Glu His Ser Thr Phe Glu Asp Asn Glu Ser Arg Arg Asp Ser Leu
 595 600 605
 Phe Val Pro Arg Arg His Gly Glu Arg Arg Asn Ser Asn Leu Ser Gln
 610 615 620
 Thr Ser Arg Ser Ser Arg Met Leu Ala Val Phe Pro Ala Asn Gly Lys
 625 630 635 640
 Met His Ser Thr Val Asp Cys Asn Gly Val Val Ser Leu Val Gly Gly
 645 650 655
 Pro Ser Val Pro Thr Ser Pro Val Gly Gln Leu Leu Pro Glu Val Ile
 660 665 670
 Ile Asp Lys Pro Ala Thr Asp Asp Asn Gly Thr Thr Thr Glu Thr Glu
 675 680 685
 Met Arg Lys Arg Arg Ser Ser Ser Phe His Val Ser Met Asp Phe Leu
 690 695 700
 Glu Asp Pro Ser Gln Arg Gln Arg Ala Met Ser Ile Ala Ser Ile Leu
 705 710 715 720
 Thr Asn Thr Val Glu Glu Leu Glu Glu Ser Arg Gln Lys Cys Pro Pro
 725 730 735
 Cys Trp Tyr Lys Phe Ser Asn Ile Phe Leu Ile Trp Asp Cys Ser Pro
 740 745 750
 Tyr Trp Leu Lys Val Lys His Val Val Asn Leu Val Val Met Asp Pro
 755 760 765
 Phe Val Asp Leu Ala Ile Thr Ile Cys Ile Val Leu Asn Thr Leu Phe
 770 775 780
 Met Ala Met Glu His Tyr Pro Met Thr Asp His Phe Asn Asn Val Leu

Ile Arg Lys Gln Lys Ile Leu Asp Glu Ile Lys Pro Leu Asp Asp
 1040 1045 1050

Leu Asn Asn Lys Lys Asp Ser Cys Met Ser Asn His Thr Ala Glu
 1055 1060 1065

Ile Gly Lys Asp Leu Asp Tyr Leu Lys Asp Val Asn Gly Thr Thr
 1070 1075 1080

Ser Gly Ile Gly Thr Gly Ser Ser Val Glu Lys Tyr Ile Ile Asp
 1085 1090 1095

Glu Ser Asp Tyr Met Ser Phe Ile Asn Asn Pro Ser Leu Thr Val
 1100 1105 1110

Thr Val Pro Ile Ala Val Gly Glu Ser Asp Phe Glu Asn Leu Asn
 1115 1120 1125

Thr Glu Asp Phe Ser Ser Glu Ser Asp Leu Glu Glu Ser Lys Glu
 1130 1135 1140

Lys Leu Asn Glu Ser Ser Ser Ser Ser Glu Gly Ser Thr Val Asp
 1145 1150 1155

Ile Gly Ala Pro Val Glu Glu Gln Pro Val Val Glu Pro Glu Glu
 1160 1165 1170

Thr Leu Glu Pro Glu Ala Cys Phe Thr Glu Gly Cys Val Gln Arg
 1175 1180 1185

Phe Lys Cys Cys Gln Ile Asn Val Glu Glu Gly Arg Gly Lys Gln
 1190 1195 1200

Trp Trp Asn Leu Arg Arg Thr Cys Phe Arg Ile Val Glu His Asn
 1205 1210 1215

Trp Phe Glu Thr Phe Ile Val Phe Met Ile Leu Leu Ser Ser Gly
 1220 1225 1230

Ala Leu Ala Phe Glu Asp Ile Tyr Ile Asp Gln Arg Lys Thr Ile
 1235 1240 1245

Lys Thr Met Leu Glu Tyr Ala Asp Lys Val Phe Thr Tyr Ile Phe
 1250 1255 1260

Ile Leu Glu Met Leu Leu Lys Trp Val Ala Tyr Gly Tyr Gln Thr

1265	1270	1275
Tyr Phe Thr Asn Ala Trp Cys Trp Leu Asp Phe Leu Ile Val Asp		
1280	1285	1290
Val Ser Leu Val Ser Leu Thr Ala Asn Ala Leu Gly Tyr Ser Glu		
1295	1300	1305
Leu Gly Ala Ile Lys Ser Leu Arg Thr Leu Arg Ala Leu Arg Pro		
1310	1315	1320
Leu Arg Ala Leu Ser Arg Phe Glu Gly Met Arg Val Val Val Asn		
1325	1330	1335
Ala Leu Leu Gly Ala Ile Pro Ser Ile Met Asn Val Leu Leu Val		
1340	1345	1350
Cys Leu Ile Phe Trp Leu Ile Phe Ser Ile Met Gly Val Asn Leu		
1355	1360	1365
Phe Ala Gly Lys Phe Tyr His Cys Ile Asn Thr Thr Thr Gly Asp		
1370	1375	1380
Arg Phe Asp Ile Glu Asp Val Asn Asn His Thr Asp Cys Leu Lys		
1385	1390	1395
Leu Ile Glu Arg Asn Glu Thr Ala Arg Trp Lys Asn Val Lys Val		
1400	1405	1410
Asn Phe Asp Asn Val Gly Phe Gly Tyr Leu Ser Leu Leu Gln Val		
1415	1420	1425
Ala Thr Phe Lys Gly Trp Met Asp Ile Met Tyr Ala Ala Val Asp		
1430	1435	1440
Ser Arg Asn Val Glu Leu Gln Pro Lys Tyr Glu Glu Ser Leu Tyr		
1445	1450	1455
Met Tyr Leu Tyr Phe Val Ile Phe Ile Ile Phe Gly Ser Phe Phe		
1460	1465	1470
Thr Leu Asn Leu Phe Ile Gly Val Ile Ile Asp Asn Phe Asn Gln		
1475	1480	1485
Gln Lys Lys Lys Phe Gly Gly Gln Asp Ile Phe Met Thr Glu Glu		
1490	1495	1500

Gln Lys Lys Tyr Tyr Asn Ala Met Lys Lys Leu Gly Ser Lys Lys
 1505 1510 1515
 Pro Gln Lys Pro Ile Pro Arg Pro Gly Asn Lys Phe Gln Gly Met
 1520 1525 1530
 Val Phe Asp Phe Val Thr Arg Gln Val Phe Asp Ile Ser Ile Met
 1535 1540 1545
 Ile Leu Ile Cys Leu Asn Met Val Thr Met Met Val Glu Thr Asp
 1550 1555 1560
 Asp Gln Ser Glu Tyr Val Thr Thr Ile Leu Ser Arg Ile Asn Leu
 1565 1570 1575
 Val Phe Ile Val Leu Phe Thr Gly Glu Cys Val Leu Lys Leu Ile
 1580 1585 1590
 Ser Leu Arg His Tyr Tyr Phe Thr Ile Gly Trp Asn Ile Phe Asp
 1595 1600 1605
 Phe Val Val Val Ile Leu Ser Ile Val Gly Met Phe Leu Ala Glu
 1610 1615 1620
 Leu Ile Glu Lys Tyr Phe Val Ser Pro Thr Leu Phe Arg Val Ile
 1625 1630 1635
 Arg Leu Ala Arg Ile Gly Arg Ile Leu Arg Leu Ile Lys Gly Ala
 1640 1645 1650
 Lys Gly Ile Arg Thr Leu Leu Phe Ala Leu Met Met Ser Leu Pro
 1655 1660 1665
 Ala Leu Phe Asn Ile Gly Leu Leu Leu Phe Leu Val Met Phe Ile
 1670 1675 1680
 Tyr Ala Ile Phe Gly Met Ser Asn Phe Ala Tyr Val Lys Arg Glu
 1685 1690 1695
 Val Gly Ile Asp Asp Met Phe Asn Phe Glu Thr Phe Gly Asn Ser
 1700 1705 1710
 Met Ile Cys Leu Phe Gln Ile Thr Thr Ser Ala Gly Trp Asp Gly
 1715 1720 1725
 Leu Leu Ala Pro Ile Leu Asn Ser Lys Pro Pro Asp Cys Asp Pro

1730 1735 1740
 Asn Lys Val Asn Pro Gly Ser Ser Val Lys Gly Asp Cys Gly Asn

1745 1750 1755
 Pro Ser Val Gly Ile Phe Phe Phe Val Ser Tyr Ile Ile Ile Ser

1760 1765 1770
 Phe Leu Val Val Val Asn Met Tyr Ile Ala Val Ile Leu Glu Asn

1775 1780 1785
 Phe Ser Val Ala Thr Glu Glu Ser Ala Glu Pro Leu Ser Glu Asp

1790 1795 1800
 Asp Phe Glu Met Phe Tyr Glu Val Trp Glu Lys Phe Asp Pro Asp

1805 1810 1815
 Ala Thr Gln Phe Met Glu Phe Glu Lys Leu Ser Gln Phe Ala Ala

1820 1825 1830
 Ala Leu Glu Pro Pro Leu Asn Leu Pro Gln Pro Asn Lys Leu Gln

1835 1840 1845
 Leu Ile Ala Met Asp Leu Pro Met Val Ser Gly Asp Arg Ile His

1850 1855 1860
 Cys Leu Asp Ile Leu Phe Ala Phe Thr Lys Arg Val Leu Gly Glu

1865 1870 1875
 Ser Gly Glu Met Asp Ala Leu Arg Ile Gln Met Glu Glu Arg Phe

1880 1885 1890
 Met Ala Ser Asn Pro Ser Lys Val Ser Tyr Gln Pro Ile Thr Thr

1895 1900 1905
 Thr Leu Lys Arg Lys Gln Glu Glu Val Ser Ala Val Ile Ile Gln

1910 1915 1920
 Arg Ala Tyr Arg Arg His Leu Leu Lys Arg Thr Val Lys Gln Ala

1925 1930 1935
 Ser Phe Thr Tyr Asn Lys Asn Lys Ile Lys Gly Gly Ala Asn Leu

1940 1945 1950
 Leu Ile Lys Glu Asp Met Ile Ile Asp Arg Ile Asn Glu Asn Ser

1955 1960 1965

Ile Thr Glu Lys Thr Asp Leu Thr Met Ser Thr Ala Ala Cys Pro
 1970 1975 1980

Pro Ser Tyr Asp Arg Val Thr Lys Pro Ile Val Glu Lys His Glu

1985 1990 1995

Gln Glu Gly Lys Asp Glu Lys Ala Lys Gly Lys

2000 2005

<210> 51

<211> 1470

<212> DNA

<213> Homo sapiens

<400> 51

atggatctgc tggatgatga actgtttgcg gatatgaacg cggatggcgc gagccccccg 60
 ccgccgcgcc cggcgggcgg cccgaaaaac accccggcgg cgccgccgct gtatgacgacc 120
 ggccgcctga gccaggcgca gctgatgccg agcccccgca tgccgggtgcc gccggcggcg 180
 ctgtttaacc gcctgctgga tgatctgggc tttagcgcgg gcccggcgct gtgcacatg 240

 ctggatacct ggaacgaaga tctgtttagc gcgctgccga ccaacgcgga tctgtatcgc 300
 gaatgcaaat ttctgagcac cctgccgagc gatgtggtgg aatggggcga tgcgtatgtg 360
 ccggaacgca cccagattga tattcgcgcg catggcgatg tggcgtttcc gaccctgccg 420
 gcgaccgcg atggcctggg cctgtattat gaagcgtga gccgcttttt tcatgcggaa 480
 ctgcgcgcgc gcgaagaaag ctatcgacc gtgctggcga acttttgag cgcgctgtat 540
 cgctatctgc gcgcgagcgt gcgccagctg catcgccagg cgcatatcgc cggcccgat 600
 cgcatctgg gcgaaatgct gcgcgcgacc attgcgatc gctattatcg cgaaaccgcg 660

 cgccctggcg gcgtgctgtt tctgcatctg tatctgtttc tgaccgcga aattctgtgg 720
 gcggcgtatg cggaacagat gatgcgccg gatctgtttg attgcctgtg ctgcgatctg 780
 gaaagctggc gccagctggc gggcctgttt cagccgttta tgtttgtgaa cggcgcgctg 840
 accgtgcgcg gcgtgccgat tgaagcgcgc cgctgcgcg aactgaacca tattcgcaa 900
 catctgaacc tgccgctggt gcgcagcgcg gcgaccgaag aaccgggcgc gccgctgacc 960
 accccgccga cctgcatgg caaccaggcg gcgcgagcg gctattttat ggtgctgatt 1020
 cgcgcgaaac tggatagcta tagcagcttt accaccagcc cgagcgaagc ggtgatgcgc 1080

 gaacatgctg atagccgcgc gcgcacccaaa aacaactatg gcagcaccat tgaaggcctg 1140
 ctggatctgc cggatgatga tgcgccgaa gaagcgggcc tggcggcgc gccgctgagc 1200

tttctgccgg cgggccatac ccgccgcctg agcaccgcgc cgccgaccga tgtgagcctg 1260
 ggcatgaac tgcattctgga tggcgaagat gtggcgatgg cgcatgcgga tgcgctggat 1320
 gatattgatac tggatatgct gggcgatggc gatagcccgg gcccgggcctt tacccecgat 1380
 gatagcgcgc cgtatggcgc gctggatatg gcggtatttg aatttgaaca gatgtttacc 1440
 gatgcgctgg gcattgatga atatggcggc 1470

<210> 52

<211> 490

<212> PRT

<213> Homo sapiens

<400> 52

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

Ser Ala Leu Tyr Arg Tyr Leu Arg Ala Ser Val Arg Gln Leu His Arg
 180 185 190
 Gln Ala His Met Arg Gly Arg Asp Arg Asp Leu Gly Glu Met Leu Arg
 195 200 205
 Ala Thr Ile Ala Asp Arg Tyr Tyr Arg Glu Thr Ala Arg Leu Ala Arg
 210 215 220
 Val Leu Phe Leu His Leu Tyr Leu Phe Leu Thr Arg Glu Ile Leu Trp
 225 230 235 240
 Ala Ala Tyr Ala Glu Gln Met Met Arg Pro Asp Leu Phe Asp Cys Leu
 245 250 255
 Cys Cys Asp Leu Glu Ser Trp Arg Gln Leu Ala Gly Leu Phe Gln Pro
 260 265 270
 Phe Met Phe Val Asn Gly Ala Leu Thr Val Arg Gly Val Pro Ile Glu
 275 280 285
 Ala Arg Arg Leu Arg Glu Leu Asn His Ile Arg Glu His Leu Asn Leu
 290 295 300
 Pro Leu Val Arg Ser Ala Ala Thr Glu Glu Pro Gly Ala Pro Leu Thr
 305 310 315 320
 Thr Pro Pro Thr Leu His Gly Asn Gln Ala Arg Ala Ser Gly Tyr Phe
 325 330 335
 Met Val Leu Ile Arg Ala Lys Leu Asp Ser Tyr Ser Ser Phe Thr Thr
 340 345 350
 Ser Pro Ser Glu Ala Val Met Arg Glu His Ala Tyr Ser Arg Ala Arg
 355 360 365
 Thr Lys Asn Asn Tyr Gly Ser Thr Ile Glu Gly Leu Leu Asp Leu Pro
 370 375 380
 Asp Asp Asp Ala Pro Glu Glu Ala Gly Leu Ala Ala Pro Arg Leu Ser
 385 390 395 400
 Phe Leu Pro Ala Gly His Thr Arg Arg Leu Ser Thr Ala Pro Pro Thr
 405 410 415
 Asp Val Ser Leu Gly Asp Glu Leu His Leu Asp Gly Glu Asp Val Ala

ggattgagga gaaacgtaaa aggacatatg agaccttcaa gagcatcatg aagaagagtc 1080
 ctttcagcgg acccaccgac ccccgccctc cacctcgacg cattgctgtg ctttcccga 1140
 gctcagcttc tgtccccaag ccagcacccc agccctatcc ctttactca tccttgagca 1200
 ccatcaacta tgatgagitt cccaccatgg tgtttccttc tgggcagatc agccaggcct 1260
 cggccttggc cccggccctc cccaagtcc tgccccaggc tccagccct gccctgctc 1320

 cagccatggt atcagctctg gcccaggccc cagcccctgt cccagtccta gcccaggccc 1380
 ctctcaggc tgtggcccca cctgccccca agcccacca ggctggggaa ggaacgctgt 1440
 cagaggccct gctcgagctg cagtttgatg atgaagacct gggggccttg ctggcaaca 1500
 gcacagacce agctgtgttc acagacctgg catccgtcga caactccgag tttcagcagc 1560
 tgctgaacca gggcatacct gtggccccc acacaactga gcccatgctg atggagtacc 1620
 ctgaggctat aactgccta gtgacagggg cccagaggcc ccccgacca gctcctgctc 1680
 cactgggggc cccggggctc cccaatggc tcctttcagg agatgaagac ttctcctcca 1740

 ttgcgacat ggacttctca gccctgctga gtcagatcag ctctaaggg ggtgacgct 1800
 gccctccca gagcactggg ttgcagggga ttgaagcct ccaaaagcac ttacggattc 1860
 tggtaggggtg tgttccaact gcccctaact ttgtggatgt cttccttga ggggggagcc 1920
 atatattatt cttttattgt cagtatctgt atctctctct ctttttgag gtgcttaagc 1980
 agaagcatta acttctctgg aaagggggga gctggggaaa ctcaaacttt tcccctgtec 2040
 tgatggtcag ctcccttctc tgtagggaac tctggggtcc cccatccca tctccagct 2100
 tctggtactc tctagagac agaagcaggc tggaggttaag gcctttgagc ccacaaagcc 2160

 ttatcaagt tcttccatca tggattcatt acagcttaat caaataacg ccccagatac 2220
 cagcccctgt atggcactgg cattgtccct gtgcctaaca ccagcgtttg aggggctggc 2280
 ctctctgccc tacagaggtc tctgcccgtt ctttcttgc tcaacctgg ctgaaggaaa 2340
 ccagtgaac agcactggct ctctccagga tccagaagg gtttggctctg ggacttcctt 2400
 gctctccctc ttctcaagt ccttaatagt aggtaagtt gttaagagt ggggagagca 2460
 ggctggcage tctccagtca ggaggcatag ttttactga acaatcaaag cacttggact 2520
 cttgctcttt ctactctgaa ctaataaatc tgttgccaag ctggctagaa 2570

<210> 54

<211> 548

<212> PRT

<213> Homo sapiens

<400> 54

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

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

500 505 510

Ala Pro Leu Gly Ala Pro Gly Leu Pro Asn Gly Leu Leu Ser Gly Asp

515 520 525

Glu Asp Phe Ser Ser Ile Ala Asp Met Asp Phe Ser Ala Leu Leu Ser

530 535 540

Gln Ile Ser Ser

545

<210> 55

<211> 1815

<212> DNA

<213> Homo sapiens

<400> 55

atgcgcccga aaaaagatgg cctggaagat tttctgcgcc tgaccccgga aattaaaaaa 60

cagctgggca gcctggtgag cgattattgc aacgtgctga acaaagaatt taccgcgggc 120

agcgtggaaa ttaccctgcg cagctataaa atttgcaaag cgtttattaa cgaagcgaag 180

gcgcatggcc gcgaatgggg cggcctgatg gcgaccctga acatttgcaa cttttgggcg 240

attctgcgca acaaccgctg gcgcccgcgc gcgaaaaacg cgggcaacga tgcgtgcagc 300

attgctgccc cgattgtgat gcgctatgtg ctggatcatc tgattgtggt gaccgatcgc 360

ttttttattc aggcgccgag caaccgctg atgattccgg cgaccattgg caccgcgatg 420

tataaactgc tgaacatag ccgctgccc gcgtatacct atagcaaagt gctgggcctg 480

gatcgcgccc cgattatggc gagcggcaaa caggtggtgg aacatctgaa ccgcatggaa 540

aaagaaggcc tgctgagcag caaatttaa gcgttttgca aatgggtggt tacctatccg 600

gtgctggaag aaatgtttca gaccatggtg agcagcaaaa ccggccatct gaccgatgat 660

gtgaaagatg tgcgcccgtg gattaaacc ctcccgcgcg cgagctatag cagccatgcg 720

ggccagcgca gctatgtgag cggcgtgctg ccggcgtgcc tgctgagcac caaaagcaaa 780

gcggtggaaa ccccgattct ggtgagcggc gcggatcgca tggatgaaga actgatgggc 840

aacgatggcg gcgcgagcca taccgaagcg cgctatagcg aaagcggcca gtttcatgcg 900

tttaccgatg aactggaaag cctgccgagc ccgaccatgc cgctgaaacc gggcgcgagc 960

agcgcggatt gcggcgatag cagcagcagc agcagcgata gcggcaacag cgataccgaa 1020

cagagcgaac gcgaagaagc gcgcgcggaa gcgccgcgcc tgcgcccgcc gaaaagccgc 1080

cgaccagcc gccgaaccg cggccagacc ccgtgccga gcaacgcggc ggaaccggaa 1140
 cagccgtgga ttgcggcggg gcatcaggaa agcgatgaac gcccatttt tccgcatccg 1200
 agcaaaccga cctttctgcc gccggtgaaa cgcaaaaaag gcctgcgcga tagccgcgaa 1260
 ggcatgtttc tgccgaaacc ggaagcgggc agcgcgatta gcgatgtgtt tgaaggccgc 1320

gaagtgtgcc agccgaaacg cattgcccg tttcatccgc cgggcagccc gtgggcgaac 1380
 cgcccgtgc cggcgagcct ggcgccgacc ccgaccggcc cggatcatga accggtgggc 1440
 agcctgacc cggcgccggg gccgcagccg ctggatccgg cggcggcggg gaccccgaa 1500
 gcgagccatc tgctggaaga tccgatgaa gaaaccagcc aggcggtgaa agcgcctgcgc 1560
 gaaatggcgg ataccgtgat tccgcagaaa gaagaagcgg cgatttgcgg ccagatggat 1620
 ctgagccatc cggcggcgg cggccatctg gatgaactga ccaccacct gaaagcatg 1680
 accgaagatc tgaacctgga tagcccctg accccggaac tgaacgaaat tctggatacc 1740

tttctgaacg atgaatgctt gctgcatgcg atgcatatta gcaccgcct gagcattttt 1800
 gataccagcc tgttt 1815

<210> 56

<211> 605

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 56

Met Arg Pro Lys Lys Asp Gly Leu Glu Asp Phe Leu Arg Leu Thr Pro

1 5 10 15

Glu Ile Lys Lys Gln Leu Gly Ser Leu Val Ser Asp Tyr Cys Asn Val

20 25 30

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

35 40 45

Tyr Lys Ile Cys Lys Ala Phe Ile Asn Glu Ala Lys Ala His Gly Arg

50 55 60

Glu Trp Gly Gly Leu Met Ala Thr Leu Asn Ile Cys Asn Phe Trp Ala

65 70 75 80

Ile Leu Arg Asn Asn Arg Val Arg Arg Arg Ala Glu Asn Ala Gly Asn

85 90 95

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

 Asp Arg Ala Ala Ile Met Ala Ser Gly Lys Gln Val Val Glu His Leu
 165 170 175
 Asn Arg Met Glu Lys Glu Gly Leu Leu Ser Ser Lys Phe Lys Ala Phe
 180 185 190
 Cys Lys Trp Val Phe Thr Tyr Pro Val Leu Glu Glu Met Phe Gln Thr
 195 200 205
 Met Val Ser Ser Lys Thr Gly His Leu Thr Asp Asp Val Lys Asp Val
 210 215 220

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

 Glu Ala Arg Tyr Ser Glu Ser Gly Gln Phe His Ala Phe Thr Asp Glu
 290 295 300
 Leu Glu Ser Leu Pro Ser Pro Thr Met Pro Leu Lys Pro Gly Ala Gln
 305 310 315 320
 Ser Ala Asp Cys Gly Asp Ser Ser Ser Ser Ser Asp Ser Gly Asn
 325 330 335
 Ser Asp Thr Glu Gln Ser Glu Arg Glu Glu Ala Arg Ala Glu Ala Pro

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

Ile Ser Thr Gly Leu Ser Ile Phe Asp Thr Ser Leu Phe
 595 600 605

<210> 57

<211> 172

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 57

Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Arg Gly
 1 5 10 15

Asn Leu Val Arg His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala
 20 25 30

Cys Asp Ile Cys Gly Lys Lys Phe Ala Leu Ser Phe Asn Leu Thr Arg
 35 40 45

His Thr Lys Ile His Thr Gly Ser Gln Lys Pro Phe Gln Cys Arg Ile

50 55 60

Cys Met Arg Asn Phe Ser Arg Ser Asp Asn Leu Thr Arg His Ile Arg
 65 70 75 80

Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys Gly Lys Lys
 85 90 95

Phe Ala Asp Arg Ser His Leu Ala Arg His Thr Lys Ile His Thr Gly
 100 105 110

Ser Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln

115 120 125

Lys Ala His Leu Thr Ala His Ile Arg Thr His Thr Gly Glu Lys Pro
 130 135 140

Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Asn Leu
 145 150 155 160

Thr Arg His Thr Lys Ile His Leu Arg Gln Lys Asp

165 170

<210> 58

<211> 516

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 58

```
cgaccattcc agtgtcgaat ctgcatgcgc aacttcagcc agcggggaaa cctggtgagg      60
catatccgca cccacacggg agagaagcct ttgcctgcg atatttggg aaagaagttt      120
gctctgagct tcaatctaac cagacacacc aagattcata ctgggtccca gaaccgttc      180
cagtgtagga tatgcatgag gaatttctct cggagtgaca acttaacgcg gcatataagg      240
acgcacacag gtgaaaaacc atttgcatgc gacatctgtg gcaaaaagt tgcggaccgg      300
tctcaccttg cccgacacac aaaaatccat accggcagtc aaaagccctt tcaatgtcgc      360
at ttgcatgc gaaacttctc acagaaggcc catttgactg cccatattcg tactcatact      420
```

```
ggcgagaaac ctttcgcttg cgatatatgt ggtcgttaagt ttgcacggtc ggacaacctc      480
acacgccaca ctaagataca cctgcggcag aaggac                                516
```

<210> 59

<211> 172

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 59

Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Ser

1 5 10 15

Asn Leu Thr Arg His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala

20 25 30

Cys Asp Ile Cys Gly Lys Lys Phe Ala Asp Lys Arg Thr Leu Ile Arg

35 40 45

His Thr Lys Ile His Thr Gly Ser Gln Lys Pro Phe Gln Cys Arg Ile

50 55 60

Cys Met Arg Asn Phe Ser Gln Arg Gly Asn Leu Val Arg His Ile Arg

65 70 75 80

Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys Gly Lys Lys

85 90 95

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

Ala Arg His Thr Lys Ile His Leu Arg Gln Lys Asp
 165 170

<210> 60

<211> 516

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 60

cgaccattcc agtgtcgaat ctgcatgcgc aacttcagcc gaagttccaa cctgacacgg 60
 catatccgca cccacacggg agagaagcct ttgcctgcg atatttgtgg aaagaagttt 120
 gctgacaage ggaccttaat cgcacacacc aagattcata ctgggtccca gaaaccgttc 180
 cagtgtagga tatgcatgag gaatttctct cagcggggaa atctagtgcg acatataagg 240
 acgcacacag gtgaaaaacc atttgcatgc gacatctgtg gcaaaaagt tgcgctgagc 300

ttcaacttga ctgctcacac aaaaatccat accggcagtc aaaagccctt tcaatgtcgc 360
 atttgcatgc gaaacttctc acggagtgc aatcttacga gacatattcg tactcact 420
 ggcgagaaac ctttcgcttg cgatatatgt ggtcgttaagt ttgcagaccg gagccactta 480
 gccaggcaca ctaagataca cctgcggcag aaggac 516

<210> 61

<211> 172

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 61

Arg Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser

1 5 10 15

Ala Leu Ala Arg His Ile Arg Thr His Thr Gly Glu Lys Pro Phe Ala

 20 25 30

Cys Asp Ile Cys Gly Lys Lys Phe Ala Arg Ser Asp Asn Leu Thr Arg

 35 40 45

His Thr Lys Ile His Thr Gly Ser Gln Lys Pro Phe Gln Cys Arg Ile

 50 55 60

Cys Met Arg Asn Phe Ser Gln Ser Gly Asp Leu Thr Arg His Ile Arg

65 70 75 80

Thr His Thr Gly Glu Lys Pro Phe Ala Cys Asp Ile Cys Gly Lys Lys

 85 90 95

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

 100 105 110

Ser Gln Lys Pro Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Ala

 115 120 125

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

 130 135 140

Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Asn Leu

145 150 155 160

Thr Arg His Thr Lys Ile His Leu Arg Gln Lys Asp

 165 170

<210> 62

<211> 516

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 62

cgaccattcc agtgtcgaat ctgcatgcgc aacttcagcg accggagcgc gctggcacgg 60

catatccgca cccacacggg agagaagcct ttgcctgcg atatttgtgg aaagaagttt 120

gctcgaagtg acaacttaac gcgccacacc aagattcata ctgggtccca gaaaccgttc 180

cagtgtagga tatgcatgag gaatttctct cagtcagggg acctcactcg tcatataagg 240

acgcacacag gtgaaaaacc atttgcatgc gacatctgtg gcaaaaagt tgcggtacga 300
 cagacgctta aacaacacac aaaaatccat accggcagtc aaaagccctt tcaatgtcgc 360
 atttgcatgc gaaacttctc agccgctggt aacttgacac gacatattcg tactcatact 420
 ggcgagaaac ctttcgcttg cgatatatgt ggtcgttaagt ttgcaagatc tgataatcta 480
 acgcgtcaca ctaagataca cctgcggcag aaggac 516

<210> 63

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 63

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Arg Gly Asn Leu

1 5 10 15

Val Arg His Ile Arg Thr His Thr Gly Glu Lys Pro

20 25

<210> 64

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 64

Phe Ala Cys Asp Ile Cys Gly Lys Lys Phe Ala Leu Ser Phe Asn Leu

1 5 10 15

Thr Arg His Thr Lys Ile His Thr Gly Ser Gln Lys Pro

20 25

<210> 65

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 65

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp Asn Leu

1 5 10 15

Thr Arg His Ile Arg Thr His Thr Gly Glu Lys Pro

20 25

<210> 66

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 66

Phe Ala Cys Asp Ile Cys Gly Lys Lys Phe Ala Asp Arg Ser His Leu

1 5 10 15

Ala Arg His Thr Lys Ile His Thr Gly Ser Gln Lys Pro

20 25

<210> 67

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 67

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Lys Ala His Leu

1 5 10 15

Thr Ala His Ile Arg Thr His Thr Gly Glu Lys Pro

20 25

<

210> 68

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 68

Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Asn Leu

1 5 10 15

Thr Arg His Thr Lys Ile His Leu Arg Gln Lys Asp

20 25

<210> 69

<211> 28
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 69
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Ser Asn Leu
 1 5 10 15

Thr Arg His Ile Arg Thr His Thr Gly Glu Lys Pro
 20 25

<210> 70
 <211> 29
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 70
 Phe Ala Cys Asp Ile Cys Gly Lys Lys Phe Ala Asp Lys Arg Thr Leu
 1 5 10 15

Ile Arg His Thr Lys Ile His Thr Gly Ser Gln Lys Pro
 20 25

<210> 71
 <211> 28
 <212> PRT
 <213> Artificial Sequence
 <220>
 ><223> Synthetic
 <400> 71
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Arg Gly Asn Leu
 1 5 10 15

Val Arg His Ile Arg Thr His Thr Gly Glu Lys Pro
 20 25

<210> 72
 <211> 29
 <212> PRT
 <213> Artificial Sequence

<220><223> Synthetic
 <400> 72
 Phe Ala Cys Asp Ile Cys Gly Lys Lys Phe Ala Leu Ser Phe Asn Leu
 1 5 10 15
 Thr Arg His Thr Lys Ile His Thr Gly Ser Gln Lys Pro
 20 25

<210> 73
 <211> 28
 <212> PRT
 <213> Artificial Sequence

<220><223> Synthetic
 <400> 73
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Arg Ser Asp Asn Leu
 1 5 10 15
 Thr Arg His Ile Arg Thr His Thr Gly Glu Lys Pro
 20 25

<210> 74
 <211> 28
 <212> PRT
 <213> Artificial Sequence

<220><223> Synthetic
 <400> 74
 Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Asp Arg Ser His Leu
 1 5 10 15
 Ala Arg His Thr Lys Ile His Leu Arg Gln Lys Asp
 20 25

<210> 75
 <211> 28
 <212> PRT
 <213> Artificial Sequence

<220><223> Synthetic
 <400> 75
 Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Asp Arg Ser Ala Leu

1 5 10 15
 Ala Arg His Ile Arg Thr His Thr Gly Glu Lys Pro

20 25

<210> 76

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 76

Phe Ala Cys Asp Ile Cys Gly Lys Lys Phe Ala Arg Ser Asp Asn Leu

1 5 10 15

Thr Arg His Thr Lys Ile His Thr Gly Ser Gln Lys Pro

20 25

<210> 77

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 77

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Gln Ser Gly Asp Leu

1 5 10 15

Thr Arg His Ile Arg Thr His Thr Gly Glu Lys Pro

20 25

<210> 78

<211> 29

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 78

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

1 5 10 15

Lys Gln His Thr Lys Ile His Thr Gly Ser Gln Lys Pro

20 25

<210> 79

<211> 28

<212> PRT

<213> Artificial Sequence

<220

><223> Synthetic

<400> 79

Phe Gln Cys Arg Ile Cys Met Arg Asn Phe Ser Ala Ala Gly Asn Leu

1 5 10 15

Thr Arg His Ile Arg Thr His Thr Gly Glu Lys Pro

 20 25

<210> 80

<211> 28

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 80

Phe Ala Cys Asp Ile Cys Gly Arg Lys Phe Ala Arg Ser Asp Asn Leu

1 5 10 15

Thr Arg His Thr Lys Ile His Leu Arg Gln Lys Asp

 20 25

<210> 81

<211> 4104

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 81

atggacaaga agtactccat tgggctcgct atcggtacca acagcgtcgg ctgggccgtc 60

attacggacg agtacaaggt gccgagcaaa aaattcaaag ttctgggcaa taccgatcgc 120

cacagcataa agaagaacct cattggagcc ctctgttcg actccgggga gacggccgaa 180

gccacgcggc tcaaagaac agcacggcgc agatataccc gcagaaagaa tcggatctgc 240

tacctgcagg agatctttag taatgagatg gctaaagtg atgactcttt cttccatagg 300

ctggaggagt cctttttggt ggaggaggat aaaaagcacg agcggccacc aatctttggc 360

aatatcgtgg acgaggtggc gtaccatgaa aagtacccaa ccatatatca tctgaggaag 420
aagctggtag acagtactga taagctgac ttgcggttga tctatctcgc gctggcgcac 480
atgatcaaat ttcggggaca cttcctcadc gagggggacc tgaaccaga caacagcgat 540
gtcgacaaac tctttatcca actggttcag acttacaatc agcttttcga ggagaacccg 600
atcaacgcat ccggcgttga cgccaaagca atcctgagcg ctaggctgtc caaatcccgg 660
cggctcgaaa acctcatcgc acagctccct ggggagaaga agaacggcct gtttgtaat 720

cttatcgccc tgtcactcgg gctgaccccc aactttaaat ctaacttcga cctggccgaa 780
gatgccaaagc tgcaactgag caaagacacc tacgatgatg atctcgacaa tctgctggcc 840
cagatcggcg accagtacgc agaccttttt ttggcggcaa agaacctgtc agacgccatt 900
ctgctgagtg atattctcgc agtgaacacg gagatcacca aagctccgct gagcgtagt 960
atgatcaagc gctatgatga gcaccaccaa gacttgactt tgctgaaggc cttgtcaga 1020
cagcaactgc ctgagaagta caaggaaatt ttcttcgac agtctaaaaa tggctacgcc 1080
ggatacattg acggcggagc aagccaggag gaattttaca aatttattaa gcccatcttg 1140

gaaaaaatgg acggcaccga ggagctgctg gtaaagetga acagagaaga tetgttgcgc 1200
aaacagcgca ctttcgacaa tggaagcatc cccaccaga ttcacctggg cgaactgcac 1260
gctatcctca ggcggcaaga ggatttctac cctttttga aagataacag ggaaaagatt 1320
gagaaaaatc teacatttcg gatacctac tatgtaggcc cctcgtcgc gggaaatcc 1380
agattcgcgt ggatgactcg caaatcagaa gagaccatca ctccctggaa cttcgaggaa 1440
gtcgtggata agggggcctc tgccagtc ttcatcgaag ggatgactaa ctttgataaa 1500
aatctgccta acgaaaaggt gcttctctaaa cactctctgc tgtacgagta cttcacagtt 1560

tataacgagc tcaccaaggt caaatcgtc acagaaggga tgagaaagcc agcattcctg 1620
tctggagagc agaagaaagc tatcgtggac ctctcttca agacgaaccg gaaagttacc 1680
tgaaacagc tcaaagaaga ctatttcaa aagattgaat gtttcgactc tgttgaatc 1740
agcggagtgg aggatcgtt caacgcaccc ctgggaacgt atcacgatct cctgaaaatc 1800
attaaagaca aggacttctt ggacaatgag gagaacgagg acattcttga ggacattgtc 1860
ctcacctta cgttgtttga agataggag atgattgaag aacgcttgaa aacttacgct 1920
catctcttcg acgacaaagt catgaaacag ctcaagagac gccgatatac aggatggggg 1980

cggctgtcaa gaaaactgat caatggcatc cgagacaagc agagtggaaa gacaatcctg 2040
gattttctta agtccgatgg atttgccaac cggaacttca tgcagttgat ccatgatgac 2100
tctctcacct ttaaggagga catccagaaa gcacaagttt ctggccaggg ggacagtctt 2160
cacgagcaca tcgctaactt tgcagtagc ccagctatca aaaagggaat actgcagacc 2220

gttaaaggtcg tggatgaact cgtcaaagta atgggaaggc ataagcccga gaatatcgtt 2280
 atcgagatgg cccgagagaa ccaaactacc cagaaggac agaagaacag tagggaaagg 2340
 atgaagagga ttgaagaggg tataaaagaa ctggggfccc aaatccttaa ggaacaccca 2400

 gttgaaaaca cccagcttca gaatgagaag ctctacctgt actacctgca gaacggcagg 2460
 gacatgtacg tggatcagga actggacatc aaccggttgt ccgactacga cgtggatgct 2520
 atcgtgcccc aaagctttct caaagatgat tctattgata ataaagtgtt gacaagatcc 2580
 gataaaaata gagggaaagag tgataacgtc ccctcagaag aagtgtgcaa gaaaatgaaa 2640
 aattattggc ggcagctgct gaacccaaa ctgatcacac aacggaagtt cgataatctg 2700
 actaaggctg aacgaggtgg cctgtctgag ttggataaag ccggcttcat caaaaggcag 2760
 cttgttgaga cacgccagat caccaagcac gtggcccaaa ttctcgattc acgcatgaac 2820

 accaagtacg atgaaaatga caaactgatt cgagaggtga aagttattac tctgaagtct 2880
 aagctggtct cagatctcag aaaggacttt cagttttata aggtgagaga gatcaacaat 2940
 taccaccatg cgcatgatgc ctacctgaat gcagtggtag gcactgcact tatcaaaaaa 3000
 tatcccaage tggaatctga atttgtttac ggagactata aagtgtacga tgttaggaaa 3060
 atgatcgcaa agtctgagca ggaaataggc aaggccaccg ctaagtactt cttttacagc 3120
 aatattatga atttttcaa gaccgagatt aactggcca atggagagat tcggaagcga 3180
 ccacttatcg aaacaaacgg agaaacagga gaaatcgtgt gggacaaggg tagggatttc 3240

 gcgacagtcc gcaaggtcct gtccatgccg caggtgaaca tcgttaaaaa gaccgaagta 3300
 cagaccggag gcttctccaa ggaaagtatc ctcccgaaaa ggaacagcga caagctgatc 3360
 gcacgcaaaa aagattggga cccaagaaa tacggcggat tcgattctcc tacagtcgct 3420
 tacagtgtac tggttgtggc caaagtggag aaagggaagt ctaaaaaact caaaagcgtc 3480
 aaggaactgc tgggcatcac aatcatggag cgatccagct tcgagaaaaa ccccatcgac 3540
 tttctgaag cgaaggata taaagaggtc aaaaaagacc tcatcattaa gctgccaag 3600
 tactctctct ttgagcttga aaacggccgg aaacgaatgc tcgctagtgc gggcgagctg 3660

 cagaaaggta acgagctggc actgcctct aaatacgtta atttcttgta tctggccagc 3720
 cactatgaaa agctcaaagg gtctccgaa gataatgagc agaagcagct gttcgtggaa 3780
 caacacaaac actacctga tgagatcadc gagcaataa gcgagttctc caaaagagtg 3840
 atcctcgccg acgtaacct cgataaggig ctttctgctt acaataagca cagggataag 3900
 cccatcaggg agcaggcaga aaacattatc cacttgitta ctctgaccaa cttgggcgcg 3960
 cctgcagcct teaagtactt cgacaccacc atagacagaa agcggtacac ctctacaag 4020

gaggtcctgg acgccacact gattcatcag tcaattacgg ggctctatga aacaagaatc 4080

gacctctctc agctcgggtgg agac 4104

<210> 82

<211> 1368

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 82

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

1 5 10 15

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

20 25 30

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

35 40 45

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

50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

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

85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

100 105 110

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

115 120 125

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

130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

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

165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

Leu Lys Asp Asn Arg Glu Lys Ile Glu Lys Ile Leu Thr Phe Arg Ile
 435 440 445
 Pro Tyr Tyr Val Gly Pro Leu Ala Arg Gly Asn Ser Arg Phe Ala Trp
 450 455 460
 Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480
 Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495
 Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510
 Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525
 Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540
 Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560
 Val Lys Gln Leu Lys Glu Asp Tyr Phe Lys Lys Ile Glu Cys Phe Asp
 565 570 575
 Ser Val Glu Ile Ser Gly Val Glu Asp Arg Phe Asn Ala Ser Leu Gly
 580 585 590
 Thr Tyr His Asp Leu Leu Lys Ile Ile Lys Asp Lys Asp Phe Leu Asp
 595 600 605
 Asn Glu Glu Asn Glu Asp Ile Leu Glu Asp Ile Val Leu Thr Leu Thr
 610 615 620
 Leu Phe Glu Asp Arg Glu Met Ile Glu Glu Arg Leu Lys Thr Tyr Ala
 625 630 635 640
 His Leu Phe Asp Asp Lys Val Met Lys Gln Leu Lys Arg Arg Arg Tyr
 645 650 655
 Thr Gly Trp Gly Arg Leu Ser Arg Lys Leu Ile Asn Gly Ile Arg Asp
 660 665 670
 Lys Gln Ser Gly Lys Thr Ile Leu Asp Phe Leu Lys Ser Asp Gly Phe

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

930 935 940

Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser

945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg

965 970 975

Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val

980 985 990

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

995 1000 1005

Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala

1010 1015 1020

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

1025 1030 1035

Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala

1040 1045 1050

Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu

1055 1060 1065

Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val

1070 1075 1080

Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr

1085 1090 1095

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

1100 1105 1110

Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro

1115 1120 1125

Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val

1130 1135 1140

Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys

1145 1150 1155

Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser

<220><223> Synthetic
 <400> 83
 gaggtacat agagtgaggc ggttttagag ctagaaatag caagttaaaa taaggctagt 60
 ccgttatcaa ctigaaaaag tggcaccgag tcggtgc 97
 <210> 84
 <211> 97
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 84
 gagguaccat agagugaggc gguuuuagag cuagaaauag caaguuaaaa uaaggcuagu 60
 ccguuaucac cuugaaaaag uggcaccgag ucggugc 97
 <210> 85
 <211> 21
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 85
 gaggtacat agagtgaggc g 21
 <210> 86
 <211> 21
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 86
 gagguaccat agagugaggc g 21
 <210> 87
 <211> 99
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 87
 accgagcga ggatgaagcc gaggttttag agctagaaat agcaagtaa aataaggcta 60

gtccgttata aacttgaaaa agtggcaccg agtcggtgc 99

<210> 88

<211> 99

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 88

accgaggcga ggaugaagcc gagguuuuag agcuagaaa agcaaguuaa aauaaggcua 60

guccguuau aacuugaaaa aguggcaccg agucggugc 99

<210> 89

<211> 23

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 89

accgaggcga ggatgaagcc gag 23

<

210> 90

<211> 23

<212> RNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 90

accgaggcga ggaugaagcc gag 23

<210> 91

<211> 100

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 91

accgaagccg agaggatact gcaggtttta gagctagaaa tagcaagtta aataaggct 60

agtccgttat caacttgaaa aagtggcacc gattcggtgc 100

<210> 92

<211> 100

<212> RNA

<213> Artificial Sequence
 <220><223> Synthetic

<400> 92
 accgaagccg agaggauacu gcagguuuua gagcuagaaa uagcaaguua aaauaaggcu 60
 aguccguuuu caacuugaaa aaguggcacc gagucggugc 100

<210> 93
 <211> 24
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic

<400> 93
 accgaagccg agaggatact gcag 24

<210> 94
 <211> 24
 <212> RNA
 <213> Artificial Sequence
 <220><223> Synthetic

<400> 94
 accgaagccg agaggauacu gcag 24

<210> 95
 <211> 1569
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic

<400> 95
 gacgcattgg acgattttga tctggatatg ctgggaagtg acgcctcga tgattttgac 60
 cttgacatgc ttggttcgga tgcccttgat gactttgacc tcgacatgct cggcagtgac 120
 gcccttgatg atttcgacct ggacatgctg attaacicta gaagtccgg atctccgaaa 180
 aagaaacgca aagttggtag ccagtacctg cccgacaccg acgaccggca ccggatcgag 240
 gaaaagcgga agcggacctc cgagacattc aagagcatca tgaagaagtc ccccttcagc 300
 ggccccaccg accctagacc tccacctaga agaatcgccg tgcccagcag atccagcgcc 360

agcgtgcaa aacctgcccc ccagccttac cccttcacca gcagcctgag caccatcaac 420

tacgacgagt tcctacat ggtgttcccc agcggccaga tctctcaggc ctctgctctg 480
gctccagccc ctctcaggt gctgcctcag gctcctgtc ctgcaccagc tccagccatg 540
gtgtctgcac tggctcaggc accagcacc gtgcctgtgc tggctcctgg acctccacag 600
gctgtggctc caccagcccc taaacctaca caggccggcg agggcacact gcttgaagct 660
ctgctgcagc tgcagttcga cgacaggat ctgggagccc tgctgggaaa cagcaccgat 720
cctgcccgtg taccgacct ggccagcgtg gacaacagcg agttccagca gctgctgaac 780

cagggcatcc ctgtggcccc tcacaccacc gagcccatgc tgatggaata ccccgaggcc 840
atcaccgggc tegtacagg cgctcagagg cctcctgatc cagctcctgc ccctctggga 900
gcaccaggcc tgctaattgg actgctgtct ggcgacgagg acttcagctc tatcgcgat 960
atggatttct cagccttgct gggctctggc agcggcagcc gggattccag ggaagggatg 1020
tttttccga agcctgaggc cggtccgct attagtgacg tgtttgaggg ccgcgaggtg 1080
tgccagccaa aacgaatccg gccatttcat cctccaggaa gtccatgggc caaccgcca 1140
ctccccgca gcctgcacc aacaccaacc ggtccagtac atgagccagt cgggtcactg 1200

accccggcac cagtccctca gccactggat ccagcgeccg cagtactcc cgaggccagt 1260
cacctgttgg aggatcccga tgaagagacg agccaggctg tcaaagccct tcgggagatg 1320
gccgatactg tgattcccga gaaggaagag gctgcaatct gtggccaaat ggaccttcc 1380
catccgcccc caaggggcca tctggatgag ctgacaacca cacttgagtc catgaccgag 1440
gatctgaacc tggactcacc cctgaccccc gaattgaacg agattctgga taccttctg 1500
aacgacgagt gcctcttga tgccatgcat atcagcacag gactgtccat ctctgacaca 1560
tctctgttt 1569

<210> 96

<211> 523

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 96

Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Gly Ser Asp Ala Leu
1 5 10 15
Asp Asp Phe Asp Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe
 20 25 30
Asp Leu Asp Met Leu Gly Ser Asp Ala Leu Asp Asp Phe Asp Leu Asp
 35 40 45

Met Leu Ile Asn Ser Arg Ser Ser Gly Ser Pro Lys Lys Lys Arg Lys
 50 55 60
 Val Gly Ser Gln Tyr Leu Pro Asp Thr Asp Asp Arg His Arg Ile Glu
 65 70 75 80
 Glu Lys Arg Lys Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met Lys Lys
 85 90 95
 Ser Pro Phe Ser Gly Pro Thr Asp Pro Arg Pro Pro Pro Arg Arg Ile
 100 105 110
 Ala Val Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala Pro Gln
 115 120 125
 Pro Tyr Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp Glu Phe
 130 135 140
 Pro Thr Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser Ala Leu
 145 150 155 160
 Ala Pro Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro Ala Pro
 165 170 175
 Ala Pro Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala Pro Val Pro
 180 185 190
 Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro Pro Ala Pro Lys
 195 200 205
 Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu Ala Leu Leu Gln Leu
 210 215 220
 Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu Leu Gly Asn Ser Thr Asp
 225 230 235 240
 Pro Ala Val Phe Thr Asp Leu Ala Ser Val Asp Asn Ser Glu Phe Gln
 245 250 255
 Gln Leu Leu Asn Gln Gly Ile Pro Val Ala Pro His Thr Thr Glu Pro
 260 265 270
 Met Leu Met Glu Tyr Pro Glu Ala Ile Thr Arg Leu Val Thr Gly Ala
 275 280 285
 Gln Arg Pro Pro Asp Pro Ala Pro Ala Pro Leu Gly Ala Pro Gly Leu

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic

<400> 97

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

1 5 10 15

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

 20 25 30

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

 35 40 45

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

 50 55 60

Lys Arg Thr Ala Arg Arg Arg Tyr Thr Arg Arg Lys Asn Arg Ile Cys

65 70 75 80

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

 85 90 95

Phe Phe His Arg Leu Glu Glu Ser Phe Leu Val Glu Glu Asp Lys Lys

 100 105 110

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

 115 120 125

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

 130 135 140

Ser Thr Asp Lys Ala Asp Leu Arg Leu Ile Tyr Leu Ala Leu Ala His

145 150 155 160

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

 165 170 175

Asp Asn Ser Asp Val Asp Lys Leu Phe Ile Gln Leu Val Gln Thr Tyr

 180 185 190

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

 195 200 205

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

Met Thr Arg Lys Ser Glu Glu Thr Ile Thr Pro Trp Asn Phe Glu Glu
 465 470 475 480

Val Val Asp Lys Gly Ala Ser Ala Gln Ser Phe Ile Glu Arg Met Thr
 485 490 495

Asn Phe Asp Lys Asn Leu Pro Asn Glu Lys Val Leu Pro Lys His Ser
 500 505 510

Leu Leu Tyr Glu Tyr Phe Thr Val Tyr Asn Glu Leu Thr Lys Val Lys
 515 520 525

Tyr Val Thr Glu Gly Met Arg Lys Pro Ala Phe Leu Ser Gly Glu Gln
 530 535 540

Lys Lys Ala Ile Val Asp Leu Leu Phe Lys Thr Asn Arg Lys Val Thr
 545 550 555 560

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

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

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

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

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

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

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

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

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

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

705 710 715 720
 His Glu His Ile Ala Asn Leu Ala Gly Ser Pro Ala Ile Lys Lys Gly
 725 730 735

 Ile Leu Gln Thr Val Lys Val Val Asp Glu Leu Val Lys Val Met Gly
 740 745 750
 Arg His Lys Pro Glu Asn Ile Val Ile Glu Met Ala Arg Glu Asn Gln
 755 760 765
 Thr Thr Gln Lys Gly Gln Lys Asn Ser Arg Glu Arg Met Lys Arg Ile
 770 775 780
 Glu Glu Gly Ile Lys Glu Leu Gly Ser Gln Ile Leu Lys Glu His Pro
 785 790 795 800

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

 Gly Lys Ser Asp Asn Val Pro Ser Glu Glu Val Val Lys Lys Met Lys
 865 870 875 880
 Asn Tyr Trp Arg Gln Leu Leu Asn Ala Lys Leu Ile Thr Gln Arg Lys
 885 890 895
 Phe Asp Asn Leu Thr Lys Ala Glu Arg Gly Gly Leu Ser Glu Leu Asp
 900 905 910
 Lys Ala Gly Phe Ile Lys Arg Gln Leu Val Glu Thr Arg Gln Ile Thr
 915 920 925

 Lys His Val Ala Gln Ile Leu Asp Ser Arg Met Asn Thr Lys Tyr Asp
 930 935 940
 Glu Asn Asp Lys Leu Ile Arg Glu Val Lys Val Ile Thr Leu Lys Ser
 945 950 955 960

Lys Leu Val Ser Asp Phe Arg Lys Asp Phe Gln Phe Tyr Lys Val Arg
 965 970 975
 Glu Ile Asn Asn Tyr His His Ala His Asp Ala Tyr Leu Asn Ala Val
 980 985 990

 Val Gly Thr Ala Leu Ile Lys Lys Tyr Pro Lys Leu Glu Ser Glu Phe
 995 1000 1005
 Val Tyr Gly Asp Tyr Lys Val Tyr Asp Val Arg Lys Met Ile Ala
 1010 1015 1020
 Lys Ser Glu Gln Glu Ile Gly Lys Ala Thr Ala Lys Tyr Phe Phe
 1025 1030 1035
 Tyr Ser Asn Ile Met Asn Phe Phe Lys Thr Glu Ile Thr Leu Ala
 1040 1045 1050

 Asn Gly Glu Ile Arg Lys Arg Pro Leu Ile Glu Thr Asn Gly Glu
 1055 1060 1065
 Thr Gly Glu Ile Val Trp Asp Lys Gly Arg Asp Phe Ala Thr Val
 1070 1075 1080
 Arg Lys Val Leu Ser Met Pro Gln Val Asn Ile Val Lys Lys Thr
 1085 1090 1095
 Glu Val Gln Thr Gly Gly Phe Ser Lys Glu Ser Ile Leu Pro Lys
 1100 1105 1110

 Arg Asn Ser Asp Lys Leu Ile Ala Arg Lys Lys Asp Trp Asp Pro
 1115 1120 1125
 Lys Lys Tyr Gly Gly Phe Asp Ser Pro Thr Val Ala Tyr Ser Val
 1130 1135 1140
 Leu Val Val Ala Lys Val Glu Lys Gly Lys Ser Lys Lys Leu Lys
 1145 1150 1155
 Ser Val Lys Glu Leu Leu Gly Ile Thr Ile Met Glu Arg Ser Ser
 1160 1165 1170

 Phe Glu Lys Asn Pro Ile Asp Phe Leu Glu Ala Lys Gly Tyr Lys
 1175 1180 1185
 Glu Val Lys Lys Asp Leu Ile Ile Lys Leu Pro Lys Tyr Ser Leu

1190	1195	1200
Phe Glu Leu Glu Asn Gly Arg	Lys Arg Met Leu Ala	Ser Ala Gly
1205	1210	1215
Glu Leu Gln Lys Gly Asn Glu	Leu Ala Leu Pro Ser	Lys Tyr Val
1220	1225	1230
Asn Phe Leu Tyr Leu Ala Ser	His Tyr Glu Lys Leu	Lys Gly Ser
1235	1240	1245
Pro Glu Asp Asn Glu Gln Lys	Gln Leu Phe Val Glu	Gln His Lys
1250	1255	1260
His Tyr Leu Asp Glu Ile Ile	Glu Gln Ile Ser Glu	Phe Ser Lys
1265	1270	1275
Arg Val Ile Leu Ala Asp Ala	Asn Leu Asp Lys Val	Leu Ser Ala
1280	1285	1290
Tyr Asn Lys His Arg Asp Lys	Pro Ile Arg Glu Gln	Ala Glu Asn
1295	1300	1305
Ile Ile His Leu Phe Thr Leu	Thr Asn Leu Gly Ala	Pro Ala Ala
1310	1315	1320
Phe Lys Tyr Phe Asp Thr Thr	Ile Asp Arg Lys Arg	Tyr Thr Ser
1325	1330	1335
Thr Lys Glu Val Leu Asp Ala	Thr Leu Ile His Gln	Ser Ile Thr
1340	1345	1350
Gly Leu Tyr Glu Thr Arg Ile	Asp Leu Ser Gln Leu	Gly Gly Asp
1355	1360	1365
Gly Thr Gly Gly Pro Pro Lys	Lys Lys Arg Lys Val	Ala Ala Ala
1370	1375	1380
Ser Arg Tyr Pro Arg Gly Asp	Ala Leu Asp Asp Phe	Asp Leu Asp
1385	1390	1395
Met Leu Gly Ser Asp Ala Leu	Asp Asp Phe Asp Leu	Asp Met Leu
1400	1405	1410
Gly Ser Asp Ala Leu Asp Asp	Phe Asp Leu Asp Met	Leu Gly Ser
1415	1420	1425

Asp Ala Leu Asp Asp Phe Asp Leu Asp Met Leu Ile Asn Ser Arg
 1430 1435 1440
 Ser Ser Gly Ser Pro Lys Lys Lys Arg Lys Val Gly Ser Gln Tyr
 1445 1450 1455
 Leu Pro Asp Thr Asp Asp Arg His Arg Ile Glu Glu Lys Arg Lys
 1460 1465 1470

 Arg Thr Tyr Glu Thr Phe Lys Ser Ile Met Lys Lys Ser Pro Phe
 1475 1480 1485
 Ser Gly Pro Thr Asp Pro Arg Pro Pro Pro Arg Arg Ile Ala Val
 1490 1495 1500
 Pro Ser Arg Ser Ser Ala Ser Val Pro Lys Pro Ala Pro Gln Pro
 1505 1510 1515
 Tyr Pro Phe Thr Ser Ser Leu Ser Thr Ile Asn Tyr Asp Glu Phe
 1520 1525 1530

 Pro Thr Met Val Phe Pro Ser Gly Gln Ile Ser Gln Ala Ser Ala
 1535 1540 1545
 Leu Ala Pro Ala Pro Pro Gln Val Leu Pro Gln Ala Pro Ala Pro
 1550 1555 1560
 Ala Pro Ala Pro Ala Met Val Ser Ala Leu Ala Gln Ala Pro Ala
 1565 1570 1575
 Pro Val Pro Val Leu Ala Pro Gly Pro Pro Gln Ala Val Ala Pro
 1580 1585 1590

 Pro Ala Pro Lys Pro Thr Gln Ala Gly Glu Gly Thr Leu Ser Glu
 1595 1600 1605
 Ala Leu Leu Gln Leu Gln Phe Asp Asp Glu Asp Leu Gly Ala Leu
 1610 1615 1620
 Leu Gly Asn Ser Thr Asp Pro Ala Val Phe Thr Asp Leu Ala Ser
 1625 1630 1635
 Val Asp Asn Ser Glu Phe Gln Gln Leu Leu Asn Gln Gly Ile Pro
 1640 1645 1650

 Val Ala Pro His Thr Thr Glu Pro Met Leu Met Glu Tyr Pro Glu

1655	1660	1665
Ala Ile Thr Arg Leu Val Thr Gly Ala Gln Arg Pro Pro Asp Pro		
1670	1675	1680
Ala Pro Ala Pro Leu Gly Ala Pro Gly Leu Pro Asn Gly Leu Leu		
1685	1690	1695
Ser Gly Asp Glu Asp Phe Ser Ser Ile Ala Asp Met Asp Phe Ser		
1700	1705	1710
Ala Leu Leu Gly Ser Gly Ser Gly Ser Arg Asp Ser Arg Glu Gly		
1715	1720	1725
Met Phe Leu Pro Lys Pro Glu Ala Gly Ser Ala Ile Ser Asp Val		
1730	1735	1740
Phe Glu Gly Arg Glu Val Cys Gln Pro Lys Arg Ile Arg Pro Phe		
1745	1750	1755
His Pro Pro Gly Ser Pro Trp Ala Asn Arg Pro Leu Pro Ala Ser		
1760	1765	1770
Leu Ala Pro Thr Pro Thr Gly Pro Val His Glu Pro Val Gly Ser		
1775	1780	1785
Leu Thr Pro Ala Pro Val Pro Gln Pro Leu Asp Pro Ala Pro Ala		
1790	1795	1800
Val Thr Pro Glu Ala Ser His Leu Leu Glu Asp Pro Asp Glu Glu		
1805	1810	1815
Thr Ser Gln Ala Val Lys Ala Leu Arg Glu Met Ala Asp Thr Val		
1820	1825	1830
Ile Pro Gln Lys Glu Glu Ala Ala Ile Cys Gly Gln Met Asp Leu		
1835	1840	1845
Ser His Pro Pro Pro Arg Gly His Leu Asp Glu Leu Thr Thr Thr		
1850	1855	1860
Leu Glu Ser Met Thr Glu Asp Leu Asn Leu Asp Ser Pro Leu Thr		
1865	1870	1875
Pro Glu Leu Asn Glu Ile Leu Asp Thr Phe Leu Asn Asp Glu Cys		
1880	1885	1890

Leu Leu His Ala Met His Ile Ser Thr Gly Leu Ser Ile Phe Asp
 1895 1900 1905
 Thr Ser Leu Phe Pro Lys Lys Lys Arg Lys Val Arg Ser Lys Arg
 1910 1915 1920
 Pro Ala Ala Thr Lys Lys Ala Gly Gln Ala Lys Lys Lys Lys Leu
 1925 1930 1935
 Asp

<210> 98

<211> 112

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 98

gaaacaagct atttgcctgat ttgtattagg taccatagag tgaggcgagg atgaagccga 60

gaggatactg cagaggtctc tgggtcaatg tgtgtatgtg tgcgtttgtg tg 112

<210> 99

<211> 71

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 99

gaaacaagct atttgcctgat ttgtattagg taccatagag tgaggcgagg atgaagccga 60

gaggatactg c 71

<210> 100

<211> 112

<212> DNA

<213> Artificial Sequence

<220><223> Synthetic

<400> 100

gaaacaagct atttgcctgat ttgtattagg taccatagag tgaggcgagg atgaagccga 60

gaggatactg cagaggtctc tgggtcaatg tgtgtatgtg tgcgtttgtg tg 112

<210> 101

<211> 108
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <220><221> misc_feature
 <222> (52)..(52)
 <223> n is a, c, g, or t
 <400> 101
 gaaacaagct atttgcgat ttgtattagg taccatagag tgaggcgagg angaagccga 60
 gaggatactg cagaggcttc tgggtcaatg tgtgtatgtg tgcgtttg 108
 <210> 102
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 102

 ttccagtgc gaatctgcat gcgcaacttc agccagcggg gaaacctggt gaggcatac 60
 cgcaccaca cgggagagaa gcct 84
 <210> 103
 <211> 87
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 103
 ttgcctgcg atatttggg aaagaagttt gctctgagct tcaatctaac cagacacacc 60
 aagattcata ctgggtccca gaaaccg 87
 <210> 104
 <211> 85
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 104

 ttccagtgta ggatatgcat gaggaatttc tctcggagtg acaacttaac gcggcatata 60

aggacgcaca caggtgaaaa aacaa 85
 <210> 105
 <211> 87
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 105
 ttgcatgcg acatctgtgg caaaaagttt gggaccggg ctcaccttgc cggacacaca 60
 aaaatccata cggcagtca aaagccc 87
 <210> 106
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 106

 tttcaatgtc gcatttgcac gcgaaacttc tcacagaagg cccatttgac tgcccatatt 60
 cgtactcata ctggcgagaa acct 84
 <210> 107
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 107
 ttcgcttgcg atatatgtgg tcgtaagttt gcacggtcgg acaacctcac acgccacact 60
 aagatacacc tgcggcagaa ggac 84
 <210> 108
 <211> 85
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 108

 ttccagtgtc gaatctgcat gcgcaacttc agcccgaatg tccaacctga cacggcatat 60
 ccgcacccac acgggagaga agcct 85

<210> 109
 <211> 87
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 109
 tttgcctgcg atatttgagg aaagaagttt gctgacaagc ggaccttaat cgccacacc 60
 aagattcata ctgggtccca gaaaccg 87

<210> 110
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 110

 ttccagtgtg ggatatgcat gaggaatttc tctcagcggg gaaatctagt gcgacatata 60
 aggacgcaca caggtgaaaa acca 84

<210> 111
 <211> 87
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 111
 tttgcatgcg acatctgtgg caaaaagttt gcgctgagct tcaacttgac tegteacaca 60
 aaaatccata cggcagtc aaagccc 87

<210> 112
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 112

 ttccaatgtc gcatttgcat gcgaaacttc tcacggagtg acaatcttac gagacatatt 60
 cgtactcata ctggcgagaa acct 84

<210> 113

<211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 113
 ttcgcttgcg atatatgtgg tcgtaagttt gcagaccgga gccacttagc caggcacact 60
 aagatacacc tgcggcagaa ggac 84
 <210> 114
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 114

 ttccagtgc gaatctgcat gcgcaacttc agcgaccgga gcgcgctggc acggcatatc 60
 cgcaccaca cgggagagaa gcct 84
 <210> 115
 <211> 87
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 115
 ttgacctgcg atatttgg aaagaagttt gctcgaagtg acaacttaac gcgccacacc 60
 aagattcata ctgggtccca gaaaccg 87
 <210> 116
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 116

 ttccagtgtg gatatgcat gaggaatttc tctcagtcag gggacctcac tcgtcatata 60
 aggacgcaca caggtgaaaa acca 84
 <210> 117
 <211> 87

<212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 117
 ttgcatgcg acatctgigg caaaaagttt gcggtacgac agacgcttaa acaacacaca 60
 aaaatccata ccggcagtca aaagccc 87
 <210> 118
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 118

 tttcaatgtc gcatttgcac gcgaaacttc tcagccgctg gtaacttgac acgacatatt 60
 cgtaactcata ctggcgagaa acct 84
 <210> 119
 <211> 84
 <212> DNA
 <213> Artificial Sequence
 <220><223> Synthetic
 <400> 119
 ttcgcttgcg atatatgigg tcgtaagttt gcaagatctg ataactaac gcgtcacact 60
 aagatacacc tgcggcagaa ggac 84
 <210> 120
 <211> 5
 <212> PRT
 <213> Artificial Sequence
 <220><223> Synthetic Polypeptide

 <400> 120
 Thr Gly Glu Lys Pro
 1 5
 <210> 121
 <211> 6
 <212> PRT

<213> Artificial Sequence

<220><223> Synthetic Polypeptide

<400> 121

Thr Gly Ser Gln Lys Pro

1

5