

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2022-0004696

(43) 공개일자 2022년01월11일

(51) 국제특허분류(Int. Cl.)

C12N 15/86 (2006.01) A61K 48/00 (2006.01)

A61P 21/00 (2006.01) C12N 9/24 (2006.01)

(52) CPC특허분류

C12N 15/86 (2013.01)

A61K 48/00 (2013.01)

(21) 출원번호 10-2021-7038231

(22) 출원일자(국제) 2020년04월29일

심사청구일자 없음

(85) 번역문제출일자 2021년11월23일

(86) 국제출원번호 PCT/US2020/030493

(87) 국제공개번호 WO 2020/223362

국제공개일자 2020년11월05일

(30) 우선권주장

62/913,401 2019년10월10일 미국(US)

62/840,911 2019년04월30일 미국(US)

(71) 출원인

더 트러스티스 오브 더 유니버시티 오브 펜실베이니아

미국 펜실베이니아 필라델피아 (우편번호 19104) 월
넷 스트리트 3451 7플로어

(72) 발명자

월슨 제임스 엠.

미국 펜실베이니아 19103 필라델피아 델란시 스트리
트 1831

호르도 줄리에트

미국 펜실베이니아 19147 필라델피아 헤드 하우스
코트 203

(74) 대리인

특허법인와이에스장

전체 청구항 수 : 총 52 항

(54) 발명의 명칭 **폼페병의 치료에 유용한 조성물****(57) 요약**

II형 글리코겐 축적 질환(폼페병)을 치료하는 데 유용한 재조합 아데노-연관 바이러스(rAAV)가 제공된다. rAAV는 근육, 심장, 신장, 및 중추신경계 중 적어도 하나의 세포를 표적화하고 발현을 지시하는 조절 서열의 제어 하에 인간 산-a-글루코시다제 hGAA780I 단백질 또는 융합 단백질을 인코딩하는 핵산 서열을 포함하는 벡터 계통을 내부에 패키징한 AAV 캡시드를 포함한다. 또한 이러한 rAAV를 제조 및 사용하는 방법이 제공된다.

(52) CPC특허분류

A61P 21/00 (2018.01)

C12N 9/2402 (2013.01)

C12Y 302/0102 (2013.01)

C12N 2750/14143 (2013.01)

명세서

청구범위

청구항 1

발현 카세트로서, 발현을 지시하는 조절 서열의 제어 하에 적어도 hGAA780I의 활성 부위를 포함하는 인간 산- α -글루코시다제(hGAA)를 인코딩하는 핵산 서열을 포함하되, 780번 위치는 서열번호 3의 아미노산 위치의 넘버링을 기준으로 하는, 발현 카세트.

청구항 2

제1항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3(hGAA780I)의 적어도 204번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, 발현 카세트.

청구항 3

제1항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 204번 아미노산 내지 952번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, 발현 카세트.

청구항 4

제1항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 123번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, 발현 카세트.

청구항 5

제1항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 952번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, 발현 카세트.

청구항 6

제1항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, 발현 카세트.

청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 상기 hGAA780I는 외인성 신호 서열을 더 포함하는 키메라 융합 단백질인, 발현 카세트.

청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 상기 hGAA780I는 서열번호 3의 전장으로 발현되는, 발현 카세트.

청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 hGAA780I는 서열번호 4, 또는 이와 적어도 95% 동일한 서열에 의해 인코딩되는, 발현 카세트.

청구항 10

제1항 내지 제8항 중 어느 한 항에 있어서, 상기 hGAA780I는 서열번호 5, 또는 이와 적어도 95% 동일한 서열에 의해 인코딩되는, 발현 카세트.

청구항 11

제1항 내지 제10항 중 어느 한 항에 있어서, 상기 조절 요소는 다중 표적 조직에서 발현을 지시하는, 발현 카세트.

청구항 12

제11항에 있어서, 상기 다중 표적 조직은 간, 골격근, 심장 및 중추신경계 세포 중 둘 이상을 포함하는, 발현 카세트.

청구항 13

제1항 내지 제12항 중 어느 한 항에 있어서, miR 표적 서열의 적어도 2개의 탠덤 반복부를 더 포함하되, 적어도 2개의 탠덤 반복부는 동일하거나 상이할 수 있는 적어도 제1 miRNA 표적 서열 및 적어도 제2 miRNA 표적 서열을 포함하고 융합 단백질을 인코딩하는 서열에 대해 3'으로 작동 가능하게 연결되는, 발현 카세트.

청구항 14

제13항에 있어서, 상기 miR 표적 서열은 서열번호 26 및 서열번호 27로부터 독립적으로 선택되는, 발현 카세트.

청구항 15

제13항 또는 제14항에 있어서, 상기 miRNA 표적 서열 중 둘 이상은 스페이서에 의해 분리되고 하나 이상의 스페이서는 (i) GGAT; (ii) CACGTG; 및 (iii) GCATGC로부터 독립적으로 선택되는, 발현 카세트.

청구항 16

제1항 내지 제15항 중 어느 한 항에 있어서, 상기 발현 카세트는 재조합 파보바이러스, 재조합 렌티바이러스, 재조합 레트로바이러스, 재조합 아데노바이러스로부터 선택되는 바이러스 벡터에 의해 운반되는, 발현 카세트.

청구항 17

제16항에 있어서, 상기 재조합 파보바이러스는 아데노-연관 바이러스인, 발현 카세트.

청구항 18

제17항에 있어서, 상기 아데노-연관 바이러스는 AAVhu.68 또는 이의 변이체, AAV9 또는 이의 변이체, 또는 AAVrh.10 또는 이의 변이체인, 발현 카세트.

청구항 19

제1항 내지 제15항 중 어느 한 항에 있어서, 상기 발현 카세트는 네이키드(naked) DNA, 네이키드 RNA, 무기 입자, 지질 입자, 중합체-기반 벡터, 또는 키토산-기반 제형으로부터 선택되는 비-바이러스 벡터에 운반되는, 발현 카세트.

청구항 20

II형 글리코겐 축적 질환(포페병)을 치료하는 데 유용한 재조합 아데노-연관 바이러스(rAAV)로서,

(a) 근육, 심장 및 중추신경계 중 적어도 하나의 세포를 표적화하는 AAV 캡시드; 및

(b) AAV 캡시드에 패키징되는 벡터 게놈으로서, 상기 벡터 게놈은 발현을 지시하는 조절 서열의 제어 하에 적어도 hGAA780I의 활성 부위를 포함하는 인간 산- α -글루코시다제(hGAA)를 인코딩하는 핵산 서열을 포함하고, 780번 위치는 서열번호 3의 아미노산 위치의 넘버링을 기준으로 하는, 상기 벡터 게놈

을 포함하는, rAAV.

청구항 21

제20항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3(hGAA780I)의 적어도 204번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, rAAV.

청구항 22

제20항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 204번 아미노산 내지 952번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, rAAV.

청구항 23

제20항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 123번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, rAAV.

청구항 24

제20항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 952번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, rAAV.

청구항 25

제20항에 있어서, 상기 hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함하는, rAAV.

청구항 26

제20항 내지 제25항 중 어느 한 항에 있어서, 상기 hGAA780I는 외인성 신호 서열을 더 포함하는 키메라 융합 단백질인, rAAV.

청구항 27

제20항 내지 제26항 중 어느 한 항에 있어서, 상기 AAV 캡시드는 간, 근육, 심장 및 중추신경계 중 둘 이상의 세포를 표적화하는, rAAV.

청구항 28

제27항에 있어서, 상기 조절 요소는 표적 세포에서 발현을 지시하는, rAAV.

청구항 29

제20항 내지 제28항 중 어느 한 항에 있어서, 상기 캡시드는 AAVhu.68 또는 이의 변이체, AAV9 또는 이의 변이체, 또는 AAVrh.10 또는 이의 변이체인, rAAV.

청구항 30

제20항 내지 제29항 중 어느 한 항에 있어서, 상기 hGAA780I는 서열번호 3의 전장으로 발현되는, rAAV.

청구항 31

제20항 내지 제30항 중 어느 한 항에 있어서, 상기 hGAA780I는 서열번호 4, 또는 이와 적어도 95% 동일한 서열에 의해 인코딩되는, rAAV.

청구항 32

제20항 내지 제30항 중 어느 한 항에 있어서, 상기 hGAA780I는 서열번호 5, 또는 이와 적어도 95% 동일한 서열에 의해 인코딩되는, rAAV.

청구항 33

제20항 내지 제32항 중 어느 한 항에 있어서, 상기 벡터 계놈은 miR 표적 서열의 적어도 2개의 탠덤 반복부를 더 포함하되, 적어도 2개의 탠덤 반복부는 동일하거나 상이할 수 있는 적어도 제1 miRNA 표적 서열 및 적어도 제2 miRNA 표적 서열을 포함하는, rAAV.

청구항 34

제33항에 있어서, 상기 miR 표적 서열은 서열번호 26 및 서열번호 27로부터 독립적으로 선택되는, rAAV.

청구항 35

제33항 또는 제34항에 있어서, 상기 miRNA 표적 서열 중 둘 이상은 스페이서에 의해 분리되고 하나 이상의 스페이서는 (i) GGAT; (ii) CACGTG; 및 (iii) GCATGC로부터 독립적으로 선택되는, rAAV.

청구항 36

제1항 내지 제19항 중 어느 한 항에 따른 발현 카세트를 인코딩하는 서열을 포함하는 핵산 분자.

청구항 37

제36항에 있어서, 상기 분자는 플라스미드인, 핵산 분자.

청구항 38

제36항 또는 제37에 따른 핵산 분자를 포함하는 숙주 세포.

청구항 39

제1항 내지 제19항 중 어느 한 항에 따른 발현 카세트 및 약제학적으로 허용 가능한 담체, 부형제 및/또는 현탁화제 중 적어도 하나를 포함하는 조성물.

청구항 40

제20항 내지 제35항 중 어느 한 항에 따른 rAAV 및 약제학적으로 허용 가능한 담체, 부형제 및/또는 현탁화제 중 적어도 하나를 포함하는 조성물.

청구항 41

제39항 또는 제40항에 있어서, 정맥내 전달용으로 제형화되는 현탁액인, 조성물.

청구항 42

제39항 또는 제40항에 있어서, 척추강내, 수조내 또는 뇌실내 투여용으로 제형화되는 현탁액인, 조성물.

청구항 43

폼페병을 가진 환자를 치료하는 방법으로서, 제1항 내지 제19항 중 어느 한 항에 따른 발현 카세트 또는 제20항 내지 제35항 중 어느 한 항에 따른 rAAV를 상기 환자에게 전달하는 단계를 포함하는, 방법.

청구항 44

제43항에 있어서, 상기 발현 카세트 및/또는 rAAV는 별도의 경로를 통해 공동-전달되는, 방법.

청구항 45

제43항 또는 제44항에 있어서, 상기 환자는 정맥내 및/또는 척추강내 전달을 통해 발현 카세트 및/또는 rAAV를 투여받는, 방법.

청구항 46

알파-글루코시다제(GAA)의 결핍이 있는 환자에서 심장, 호흡기 및/또는 골격근 기능을 개선시키는 방법으로서, 제1항 내지 제19항 중 어느 한 항에 따른 발현 카세트 또는 제20항 내지 제35항 중 어느 한 항에 따른 rAAV를 상기 환자에게 전달하는 단계를 포함하는, 방법.

청구항 47

폼페병을 가진 환자에 대한 치료 요법으로서, 제1항 내지 제19항 중 어느 한 항에 따른 발현 카세트 또는 제20항 내지 제35항 중 어느 한 항에 따른 rAAV를 면역조절제와 조합하여 사용하는 것을 포함하는, 치료 요법.

청구항 48

제47항에 있어서, 상기 환자는 후기 발병형 폼페병을 가진, 치료 요법.

청구항 49

제47항에 있어서, 상기 환자는 영아 발병형 폼페병을 가진, 치료 요법.

청구항 50

제47항 내지 제49항 중 어느 한 항에 있어서, 상기 환자는 기관지확장제, 아세틸콜린에스터라제 저해제, 호흡근 강화 훈련(respiratory muscle strength training: RMST), 효소 대체 요법, 및/또는 횡경막 페이스팅 요법과의 공동-요법을 받는, 치료 요법.

청구항 51

폼페병을 가진 환자를 치료함에 있어서의, 제1항 내지 제19항 중 어느 한 항에 따른 발현 카세트 또는 제20항 내지 제35항 중 어느 한 항에 따른 rAAV의 용도.

청구항 52

약제의 제조에 있어서의 제1항 내지 제19항 중 어느 한 항에 따른 발현 카세트 또는 제18항 내지 제35항 중 어느 한 항에 따른 rAAV의 용도.

발명의 설명

배 경 기 술

[0001] II형 글리코겐증으로도 알려진 폼페병은 심장(심근병증), 근육, 및 운동 뉴런(신경근육계 질환)에 글리코겐 축적을 야기하는 산- α -글루코시다제(GAA) 유전자의 돌연변이에 의해 유발된 리소좀 축적 질환이다. 고전적인 영아 폼페병에서, 심각한 GAA 활성 손실은, 특히 심장과 근육에서, 다계통 및 조기-발병 글리코겐 저장을 유발하고 심폐 부전으로 첫 몇 년 동안 사망을 유발한다. 영아 폼페병은 또한 뉴런(특히, 운동 뉴런) 및 신경아교세포 내의 현저한 글리코겐 저장을 특징으로 한다. 현재의 치료 기준인 효소 대체 요법(ERT)은 근육을 교정하는 데 최적이지 아닌 효율을 가지고 혈액-뇌 장벽을 통과할 수 없어 고전적인 영아 폼페병의 장기 생존자에서 점진적인 신경학적 악화로 이어진다. 심장 교정으로 인해 더 오래 사는 ERT를 받은 환자는 진행성 신경학적 표현형과 함께 새로운 자연사를 보인다. 추가적으로, 재조합 인간 GAA는 면역원성이 높으며 골격근에 의한 흡수 불량으로 인해 매우 많은 양으로 투약되어야 한다.

[0002] 질환의 CNS 구성요소의 교정에 대한 필요성, 개선된 근육 교정에 대한 필요성, 및 더 효과적이고 덜 면역원성이 고/이거나 더 편리한 현재 ERT에 대한 대안에 대한 필요성을 포함하여, 폼페병의 치료에 대한 몇 가지 충족되지 않은 요구가 있다.

발명의 내용

[0003] 특정 실시형태에서, 발현을 지시하는 조절 서열의 제어 하에 적어도 hGAA780I의 활성 부위를 포함하는 인간 산- α -글루코시다제(hGAA)를 인코딩하는 핵산 서열을 포함하는 발현 카세트가 제공되며, 여기서 780번 위치는 서열 번호 3의 아미노산 위치의 넘버링을 기준으로 한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열 번호 3(hGAA780I)의 적어도 204번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 204번 아미노산 내지 952번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 123번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 952번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, 발현 카세트는 miR 표적 서열의 적어도 2개의 탠덤 반복부를 더 포함하며, 여기서 적어도 2개의 탠덤 반복부는 동일하거나 상이할 수 있는 적어도 제1 miRNA 표적 서열 및 적어도 제2 miRNA 표적 서열을 포함하고 hGAA780I 단백질을 인코딩하는 서열에 대해 3'으로 작동 가능하게 연결된다.

[0004] 특정 실시형태에서, 본 명세서에서 제공되는 발현 카세트는 재조합 파코바이러스, 재조합 렌티바이러스, 재조합 레트로바이러스, 및 재조합 아데노바이러스로부터 선택되는 바이러스 벡터에 의해 운반된다. 특정 실시형태에서, 재조합 파코바이러스는 클레드(clade) F 아데노-연관 바이러스, 선택적으로 AAVhu68이다. 특정 실시형태에서, 본 명세서에서 제공되는 발현 카세트는 네이키드(naked) DNA, 네이키드 RNA, 무기 입자, 지질 입자, 중합체-기반 벡터, 또는 키토산-기반 제형으로부터 선택되는 비-바이러스 벡터에 의해 운반된다.

- [0005] 특정 실시형태에서, II형 글리코겐 축적 질환(포페병)을 치료하는 데 유용한 재조합 아데노-연관 바이러스(rAAV)가 본 명세서에서 제공되며, rAAV는 (a) 근육, 심장 및 중추신경계 중 적어도 하나의 세포를 표적화하는 AAV 캡시드; 및 (b) AAV 캡시드에 패키징된 벡터 게놈이되, 발현을 지시하는 조절 서열의 제어 하에 적어도 hGAA780I의 활성 부위를 포함하는 인간 산- α -글루코시다제(hGAA)를 인코딩하는 핵산 서열을 포함하는, 벡터 게놈을 포함하고, 여기서 780번 위치는 서열번호 3의 아미노산 위치의 넘버링을 기준으로 한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 204번 아미노산 내지 890번 아미노산(hGAA780I), 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 204번 아미노산 내지 952번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 123번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 952번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, rAAV 벡터 게놈은 miR 표적 서열의 적어도 2개의 탠덤 반복부를 더 포함하며, 여기서 적어도 2개의 탠덤 반복부는 동일하거나 상이할 수 있는 적어도 제1 miRNA 표적 서열 및 적어도 제2 miRNA 표적 서열을 포함하고 hGAA780I 단백질을 인코딩하는 서열에 대해 3'으로 작동 가능하게 연결된다.
- [0006] 특정 실시형태에서, 본 명세서에 기재된 바와 같은 hGAA780I 단백질을 인코딩하는 발현 카세트 및 각각의 약제학적으로 허용 가능한 담체, 부형제 및/또는 현탁화제 중 적어도 하나를 포함하는 조성물이 제공된다.
- [0007] 특정 실시형태에서, 본 명세서에 기재된 바와 같은 hGAA780I 단백질을 인코딩하는 발현 카세트를 포함하는 rAAV 및 각각의 약제학적으로 허용 가능한 담체, 부형제 및/또는 현탁화제 중 적어도 하나를 포함하는 조성물이 제공된다.
- [0008] 특정 실시형태에서, 포페병을 가진 환자를 치료하고/하거나 알파-글루코시다제(GAA)의 결핍이 있는 환자에서 심장, 호흡기 및/또는 골격근 기능을 개선시키는 방법이 제공된다. 이 방법은 본 명세서에 기재된 바와 같은 발현 카세트, rAAV, 또는 조성물을 환자에게 전달하는 단계를 포함한다. 발현 카세트, rAAV, 또는 조성물은 정맥내로 및/또는 척추강내, 수조내 또는 뇌실내 투여를 통하여 전달될 수 있다. 추가적으로 또는 대안적으로, 이와 같은 유전자 요법은 심장으로의 직접 전달(심장), 폐로의 전달(비강내, 흡입, 기관내), 및/또는 근육내 주사를 포함할 수 있다. 이들 중 하나는 발현 카세트, 벡터, 또는 조성물의 단독 투여 경로일 수 있거나, 다른 전달 경로와 공동 투여될 수 있다.
- [0009] 포페병을 가진 환자를 치료하기 위한 치료 요법은 본 명세서에 기재된 바와 같은 발현 카세트, rAAV, 또는 조성물을 단독으로, 또는 공동 요법과 조합하여, 예를 들어, 면역조절제, 기관지확장제, 아세틸콜린에스터라제 저해제, 호흡근 강화 훈련(respiratory muscle strength training: RMST), 효소 대체 요법, 및/또는 횡경막 페이스팅 요법 중 하나 이상과 조합하여 환자에게 전달하는 것을 포함할 수 있다.
- [0010] 특정 실시형태에서, 본 명세서에 기재된 발현 카세트 및/또는 rAAV의 생성을 위한 핵산 분자 및 숙주 세포가 제공된다.
- [0011] 특정 실시형태에서, 약제의 제조에 있어서 발현 카세트, rAAV, 및/또는 조성물의 용도가 제공된다.
- [0012] 특정 실시형태에서, 포페병을 가진 환자를 치료하고/하거나 알파-글루코시다제(GAA)의 결핍이 있는 환자에서 심장, 호흡기 및/또는 골격근 기능을 개선시키는 데 적합한 발현 카세트, rAAV, 및/또는 조성물이 제공된다.
- [0013] 본 발명의 다른 양태 및 이점은 본 발명의 하기 상세한 설명으로부터 용이하게 명백할 것이다.

도면의 간단한 설명

- [0014] 도 1A 및 도 1B는 CB6(세 번째 칼럼), CAG(네 번째 칼럼) 또는 UbC 프로모터(마지막 칼럼)의 지시 하에 hGAAV780I(서열번호 4)에 대한 조작된 코딩 서열을 갖는 다양한 AAVhu68.hGAA의 정맥내 투여 4주 후 포페(-/-) 마우스 간에서의 hGAA 활성을 나타내는 도면. (도 1A) 저용량(1×10^{11} 개 GC). (도 1B) 고용량(1×10^{12} 개).
- 도 2A 및 도 2B는 CB6(세 번째 칼럼), CAG(네 번째 칼럼) 또는 UbC 프로모터(마지막 칼럼)의 지시 하에 hGAAV780I에 대한 조작된 코딩 서열을 갖는 다양한 AAVhu68.hGAA의 정맥내 투여 4주 후 포페(-/-) 마우스 심장에서의 hGAA 활성을 나타내는 도면. (도 2A) 저용량(1×10^{11} 개 GC). (도 2B) 고용량(1×10^{12} 개).

도 3A 및 도 3B는 CB6(세 번째 칼럼), CAG(네 번째 칼럼) 또는 UbC 프로모터(마지막 칼럼)의 지시 하에 hGAAV780I에 대한 조작된 코딩 서열을 갖는 다양한 AAVhu68.hGAA의 정맥내 투여 4주 후 품페(-/-) 마우스 골격근(대퇴사두근)에서의 hGAA 활성을 나타내는 도면. (도 3A) 저용량(1×10^{11} 개 GC). (도 3B) 고용량(1×10^{12} 개).

도 4A 및 도 4B는 CB6(세 번째 칼럼), CAG(네 번째 칼럼) 또는 UbC 프로모터(마지막 칼럼)의 지시 하에 hGAAV780I에 대한 조작된 코딩 서열을 갖는 다양한 AAVhu68.hGAA의 정맥내 투여 4주 후 품페(-/-) 마우스 뇌에서의 hGAA 활성을 나타내는 도면. (도 4A) 저용량(1×10^{11} 개 GC). (도 4B) 고용량(1×10^{12} 개). CB7 활성 하에 발현하는 벡터는 두 가지 용량 모두에서 더 낮은 활성을 갖는 반면, CAG 또는 UbC 프로모터 하에서 발현하는 벡터는 더 높은 용량에서 비슷한 활성을 가진다.

도 5A 내지 도 5H는 AAVhu68.hGAA의 전달 4주 후 품페 마우스에서 심장의 조직학을 나타내는 도면(PAS 염색은 글리코겐 저장을 나타냄). 5가지 상이한 hGAA 발현 카세트를 포함하는 rAAVhu68 벡터를 생성하고 평가하였다. 비히클 대조군 품페(-/-)(도 5D) 및 야생형(+/-)(도 5A) 마우스는 PBS 주사를 받았다. "hGAA"는 천연 신호 펩타이드를 갖는 야생형 서열에 의해 인코딩되는 참조 자연 효소(hGAAV780)를 지칭한다(도 5B). "BiP-vIGF2.hGAAco"는 처음 35개 AA의 결실을 포함하고, 인슐린 수용체에 대한 낮은 친화성으로 IGF2 변이체와의 융합체인 BiP 신호 펩타이드를 추가로 갖는 참조 hGAAV780 단백질에 대한 조작된 코딩 서열을 지칭한다(도 5C). "hGAAcoV780I"는 조작된 서열에 의해 인코딩되고 천연 신호 펩타이드를 포함하는 hGAAV780I 변이체를 지칭한다(도 5E). "BiP-vIGF2.hGAAcoV780I"는 처음 35개 AA의 결실을 포함하고, 인슐린 수용체에 대한 낮은 친화성으로 IGF2 변이체와 융합된 BiP 신호 펩타이드 및 조작된 서열에 의해 인코딩되는 hGAAV780I를 추가로 갖는 hGAAcoV780I를 지칭한다(도 5F).

"Sp7. Δ8.hGAAcoV780I"는 이전 작제물과 동일한 조작된 서열에 의해 인코딩되는 처음 35개 AA의 결실이 있지만 천연 신호 펩타이드 대신에 B2 키모트립시노겐 신호 펩타이드를 인코딩하는 서열을 포함하는 hGAAV780I 변이체를 지칭한다(도 5G). (도 5H) 맹검 조직병리학 반-정량적 중증도 스코어링. 공인 수의 병리학자는 블라인드 방식으로 슬라이드를 검토하고 글리코겐 저장 및 자가포식 축적을 기반으로 중증도 점수를 설정하였다.

도 6A 내지 도 6H는 다양한 hGAA를 인코딩하는 AAVhu68(2.5×10^{13} 개 GC/kg)의 투여 4주 후 품페 마우스에서 대퇴사두근 조직학의 결과(PAS 염색)를 나타내는 도면. 대조군 품페(-/-)(도 6D) 및 야생형(+/-)(도 6A) 마우스는 PBS 주사를 받았다. "hGAA"는 천연 신호 펩타이드를 갖는 야생형 서열에 의해 인코딩되는 참조 자연 효소(hGAAV780)를 지칭한다(도 6B). "hGAAcoV780I"는 조작된 서열에 의해 인코딩되고 천연 신호 펩타이드를 포함하는 hGAAV780I 변이체를 지칭한다(도 6E). "Sp7. Δ8.hGAAcoV780I"는 이전 작제물과 동일한 조작된 서열에 의해 인코딩되는 처음 35개 AA의 결실이 있지만 천연 신호 펩타이드 대신에 B2 키모트립시노겐 신호 펩타이드를 인코딩하는 서열을 포함하는 hGAAV780I 변이체를 지칭한다(도 6F). "BiP-vIGF2.hGAAco"는 처음 35개 AA의 결실을 포함하고, 인슐린 수용체에 대한 낮은 친화성으로 IGF2 변이체와의 융합체인 BiP 신호 펩타이드를 추가로 가지고 조작된 서열에 의해 인코딩되는 참조 hGAAV780을 지칭한다(도 6C). "BiP-vIGF2.hGAAcoV780I"는 처음 35개 AA의 결실을 포함하고, 인슐린 수용체에 대한 낮은 친화성으로 IGF2 변이체와 융합된 BiP 신호 펩타이드 및 조작된 서열에 의해 인코딩되는 hGAAV780I를 추가로 갖는 hGAAV780I를 지칭한다(도 6G). (도 6H) 맹검 조직병리학 반-정량적 중증도 스코어링. 공인 수의 병리학자는 블라인드 방식으로 슬라이드를 검토하고 글리코겐 저장 및 자가포식 축적을 기반으로 중증도 점수를 설정하였다. 0점은 병변이 없음을 의미하고; 1점은 평균적으로 저장에 의해 영향을 받는 근육 섬유의 9% 미만을 의미하며; 2점은 10 내지 49%를 의미하고; 3점은 50 내지 75%를 의미하며 4점은 76 내지 100%를 의미한다.

도 7A 내지 도 7H는 2.5×10^{12} 개 GC/kg(즉, 도 6A 내지 도 6H에서보다 10배 더 적은 용량)으로 다양한 hGAA를 인코딩하는 AAVhu68의 투여 4주 후 품페 마우스 유래의 대퇴사두근 조직학의 결과(파오요드산-설프(PAS) 염색)를 나타내는 도면. 대조군 품페(-/-)(도 7D) 및 야생형(+/-)(도 7A) 마우스는 PBS 주사를 받았다. "hGAA"는 천연 신호 펩타이드를 갖는 야생형 서열에 의해 인코딩되는 참조 자연 효소(hGAAV780)를 지칭한다(도 7B). "hGAAcoV780"는 조작된 서열에 의해 인코딩되고 천연 신호 펩타이드를 포함하는 hGAAV780I 변이체를 지칭한다(도 7E). "Sp7. Δ8.hGAAcoV780I"는 이전 작제물과 동일한 조작된 서열에 의해 인코딩되는 처음 35개 AA의 결실이 있지만 천연 신호 펩타이드 대신에 B2 키모트립시노겐 신호 펩타이드를 인코딩하는 서열을 포함하는 hGAAV780I 변이체를 지칭한다(도 7F). "BiP-vIGF2.hGAAco"는 처음 35개 AA의 결실을 포함하고, 인슐린 수용체에 대한 낮은 친화성으로 IGF2 변이체와의 융합체인 BiP 신호 펩타이드를 추가로 가지고 조작된 서열에 의해 인코딩되는 참조 hGAAV780을 지칭한다(도 7C). "BiP-vIGF2.hGAAcoV780I"는 처음 35개 AA의 결실을 포함하고, 인

슬린 수용체에 대한 낮은 친화성으로 IGF2 변이체와 융합된 BiP 신호 펩타이드 및 조작된 서열에 의해 인코딩되는 hGAAV780I를 추가로 갖는 hGAAV780I를 지칭한다(도 7G). (도 7H) 맹검 조직병리학 반-정량적 중증도 스코어링. 공인 수의 병리학자는 블라인드 방식으로 슬라이드를 검토하고 글리코겐 저장 및 자가포식 축적을 기반으로 중증도 점수를 설정하였다. 0점은 병변이 없음을 의미하고; 1점은 평균적으로 저장에 의해 영향을 받는 근육 섬유의 9% 미만을 의미하며; 2점은 10 내지 49%를 의미하고; 3점은 50 내지 75%를 의미하며 4점은 76 내지 100%를 의미한다.

도 8은 천연 hGAA 또는 처음 35개 AA의 결실을 포함하고, 인슬린 수용체에 대한 낮은 친화성으로 IGF2 변이체와 융합된 BiP 신호 펩타이드 및 조작된 서열에 의해 인코딩되는 hGAAV780I를 추가로 갖는 hGAAV780I("BiP-vIGF2.hGAAcoV780I")를 인코딩하는 서열을 갖는 AAVhu68의 투여(2.5×10^{12} 개 GC/kg) 4주 후 품페 마우스 유래의 척수 조직학의 결과(PAS 및 룩솔 패스트 블루(luxol fast blue) 염색)를 나타내는 도면. 맹검 조직병리학 반-정량적 중증도 스코어링을 척수 절편에 대해 수행하였다.

도 9는 혈장에서 hGAA 활성을 나타내는 도면. 품페 마우스에 야생형 hGAA 또는 BiP-vIGF2.hGAA를 인코딩하는 벡터를 저용량(2.5×10^{12} 개 GC)으로 투여하였다. 정맥내 투여 4주 후 높은 수준의 야생형 및 조작된 hGAA 활성이 혈장에서 검출되었다.

발명을 실시하기 위한 구체적인 내용

- [0015] hGAA780I 효소를 품페병을 가진 환자에게 전달하기 위한 조성물이 제공된다. 이들 조성물로 환자를 치료하기 위한 요법을 포함하여, 상기 조성물을 제조 및 사용하는 방법이 본 명세서에 기재되어 있다.
- [0016] 본 명세서에서 사용되는 바와 같이, 말타제 결핍증, 글리코겐 축적 질환 II형(GSDII), 또는 글리코겐증 II형으로도 지칭되는 "품페병"은 산 α -글루코시다제를 코딩하는 산 α -글루코시다제(GAA) 유전자의 돌연변이에 의해 유발되는 리소좀 효소인 GAA의 전체 부재 또는 부분적 결핍을 특징으로 하는 유전적 리소좀 축적 장애를 지칭하는 것으로 의도된다. 상기 용어는 영아, 청소년 및 성인-발병형 품페병을 포함하지만, 이들로 제한되지 않는 질환의 초기 및 후기 발병 형태를 포함하지만, 이들로 제한되지 않는다.
- [0017] 그리스 문자 "알파" 및 기호 " α "는 본 명세서 전반에 걸쳐 호환 가능하게 사용됨이 이해될 것이다. 유사하게, 그리스 문자 "델타" 및 " Δ "는 본 명세서 전반에 걸쳐 호환 가능하게 사용된다.
- [0018] 본 명세서에서 사용되는 바와 같이, 용어 "산- α -글루코시다제" 또는 "GAA"는 글리코겐, 말토스, 및 아이소말토스의 D-글루코스 단위 사이의 α -1,4 연결을 가수분해하는 리소좀 효소를 지칭한다. 대체 명칭은 리소좀 α -글루코시다제(EC:3.2.1.20); 글루코아밀라제; 1,4- α -D-글루칸 글루코하이드롤라제; 아밀로글루코시다제; 감마-아밀라제 및 엑소-1,4- α -글루코시다제를 포함하지만, 이들로 제한되지 않는다. 인간 산 α -글루코시다제는 GAA 유전자(미국 국립생물정보센터(National Centre for Biotechnology Information: NCBI) 유전자 ID 2548)에 의해 인코딩되며, 상기 유전자는 17번 염색체의 긴 팔에 대해 매핑되었다(위치 17q25.2-q25.3). 아미노산 잔기 516 내지 521번에서 보존된 헥사펩타이드 WIDMNE는 산 α -글루코시다제 단백질의 활성화에 필요하다. 용어 "hGAA"는 인간 GAA에 대한 코딩 서열을 지칭한다.
- [0019] 본 명세서에서 사용되는 바와 같이, "rAAV.hGAA"는 최소한 GAA 효소에 대한 코딩 서열을 포함하는 벡터 계통(예를 들어, 780I 변이체)을 내부에 패키징한 AAV 캡시드를 갖는 rAAV를 지칭한다. rAAVhu68.hGAA 또는 rAAVhu68.hGAA는 AAV 캡시드가 본 명세서에 정의된 AAVhu68 캡시드인 rAAV를 지칭한다.
- [0020] 전장 hGAA의 넘버링을 참조하여, 아미노산 위치 1 내지 27번에 신호 펩타이드가 있다. 추가적으로, 효소는 다수의 성숙한 단백질, 즉, 아미노산 위치 70 내지 952번의 성숙한 단백질, 아미노산 위치 123 내지 952번에 위치한 76 kD 성숙한 단백질, 및 204번 아미노산 내지 952번 아미노산의 70 kD 성숙한 단백질과 연관되어 있다. "활성 촉매 부위"는 헥사펩타이드 WIDMNE(서열번호 3의 아미노산 잔기 516 내지 521번)를 포함한다. 특정 실시형태에서, 더 긴 단편, 예를 들어, 위치 516 내지 616번이 선택될 수 있다. 다른 활성 부위는 위치 376, 404, 405, 441, 481, 516, 518, 519, 600, 613, 616, 649, 674번 중 하나 이상에 위치할 수 있는 리간드 결합 부위를 포함한다.
- [0021] 달리 명시되지 않는 한, 용어 "hGAA780I" 또는 "hGAAV780I"는 서열번호 3으로 재현되는 아미노산 서열을 갖는 전장 프리-프로-단백질을 지칭한다. 일부 경우에, 용어 hGAAco780I 또는 hGAAcoV780I는 hGAA780I를 인코딩하는 조작된 서열을 지칭하는 데 사용된다. 앞의 단락에 기재된 hGAA 참조 단백질과 비교하여, hGAA780I는 참조 hGAA가 발린(Val 또는 V)을 포함하는 780번 위치에 아이소류신(Ile 또는 I)을 가진다. 이러한 hGAA780I는 "참조 서

열"로 문헌에 널리 기재되어 있는 780번 위치에 발린을 갖는 hGAA 서열(hGAAV780)보다 더 우수한 효과 및 개선된 안전성 프로파일을 갖는 것으로 예기치 않게 발견되었다. 예를 들어, 도 5A 내지 5H에서 볼 수 있는 바와 같이, hGAAV780 참조 서열은 hGAA780I 효소에 의해 동일한 용량으로 보이지 않는 독성(섬유화 심근염)을 유도한다. 따라서, hGAA780I의 사용으로 hGAA를 이용한 요법을 받는 환자의 섬유화 심근염을 감소시키거나 없앨 수 있다. hGAA 신호 펩타이드, 성숙한 단백질, 활성 촉매 부위, 및 결합 부위, 즉, 아미노산 위치 1 내지 27번의 신호 펩타이드; 아미노산 위치 70 내지 952번의 성숙한 단백질; 아미노산 위치 123 내지 952번에 위치한 76 kDa 성숙한 단백질, 및 아미노산 204번 아미노산 내지 952번의 70 kD 성숙한 단백질; 아미노산 잔기 516 내지 521번의 헥사펩타이드 WIDMNE를 포함하는 "활성 촉매 부위"(서열번호 28)의 위치는 서열번호 3으로 재현되는 hGAA780I에서 유사한 위치를 기반으로 결정될 수 있으며; 다른 활성 부위는 위치 376, 404, 405, 441, 481, 516, 518, 519, 600, 613, 616, 649, 674번 중 하나 이상에 위치할 수 있는 리간드 결합 부위를 포함한다.

[0022] 특정 실시형태에서, 서열번호 3의 hGAA780I와 적어도 95% 동일, hGAA780I와 적어도 97% 동일, 또는 hGAA780I와 적어도 99% 동일한 서열을 갖는 hGAA780I가 선택될 수 있다. 특정 실시형태에서, 서열번호 3의 성숙한 hGAA780I 단백질과 적어도 95%, 적어도 97%, 또는 적어도 99% 동일한 서열이 제공된다. 특정 실시형태에서, hGAA780I와 적어도 95% 내지 적어도 99% 동일성을 갖는 서열은 임의의 변화 없이 유지된 활성 촉매 부위에 대한 서열을 갖는다. 특정 실시형태에서, 서열번호 3에 대한 hGAA780I에 대해 적어도 95% 내지 적어도 99% 동일성을 갖는 서열은 적절한 동물 모델에서 테스트할 때 참조 hGAAV780보다 개선된 생물학적 효과 및 더 우수한 안전성 프로파일을 갖는 것을 특징으로 한다. 특정 실시형태에서, GAA 활성 분석은 이전에 기재된 바와 같이(예를 들어, 문헌 [J. Hordeaux, et. al., Acta Neuropathological Communications, (2107) 5: 66] 참조) 또는 다른 적합한 방법을 사용하여 수행될 수 있다. 특정 실시형태에서, hGAA780I 효소는 hGAA 아미노산 서열의 다른 위치에 변형을 포함한다. 돌연변이체의 예는, 예를 들어, 미국 특허 제9,920,307호에 기재된 것을 포함할 수 있다. 특정 실시형태에서, 이와 같은 돌연변이체 hGAA780I는 최소한 활성 촉매 부위의 WIDMNE(서열번호 28) 및 하기 기재된 바와 같은 780I 영역의 아미노산을 유지할 수 있다.

[0023] 특정 실시형태에서, 천연 hGAA 신호 펩타이드 이외의 리더 펩타이드를 포함하는 신규 hGAA780I 융합 단백질이 제공된다. 특정 실시형태에서, 이와 같은 외인성 리더 펩타이드는 바람직하게는 인간 기원이고, 예를 들어, IL-2 리더 펩타이드를 포함할 수 있다. 특정 실시형태에서 작동 가능한 특정 외인성 신호 펩타이드는 키모트립시노겐 B2 유래의 아미노산 1 내지 20개, 인간 알파-1-항트립신의 신호 펩타이드, 이두로네이트-2-설파타제 유래의 아미노산 1 내지 25개, 및 프로테아제 CI 저해제 유래의 아미노산 1 내지 23개를 포함한다. 예를 들어, WO 제 2018046774호를 참조한다. 다른 신호/리더 펩타이드는 특히 면역글로불린(예를 들어, IgG), 사이토카인(예를 들어, IL-2, IL12, IL18 등), 인슐린, 알부민, β -글루쿠로니다제, 알칼리성 프로테아제 또는 피브로넥틴 분비 신호 펩타이드에서 천연적으로 발견될 수 있다. 또한, 예를 들어, signalpeptide.de/index.php?m=listspdb_mammalia를 참조한다.

[0024] 이와 같은 키메라 hGAA780I는 전체 27개 aa 천연 신호 펩타이드 대신에 외인성 리더를 가질 수 있다. 선택적으로, hGAA780I 효소의 N-말단 절단으로 신호 펩타이드(예를 들어, 약 2 내지 약 25개 아미노산, 또는 그 사이 값의 결실), 전체 신호 펩타이드, 또는 신호 펩타이드보다 더 긴 단편(서열번호 3의 넘버링을 기준으로 하여 최대 70개 아미노산)의 단지 일부만 결합될 수 있다. 선택적으로, 이와 같은 효소는 길이가 약 5, 10, 15 또는 20개 아미노산인 C-말단 절단을 포함할 수 있다.

[0025] 특정 실시형태에서, 융합 파트너에 결합되는 성숙한 hGAA780I 단백질(aa 70 내지 952번), 성숙한 70 kD 단백질(aa 123 내지 aa 952번), 또는 성숙한 76 kD 단백질(aa 204 내지 952번)을 포함하는 신규 융합 단백질이 제공된다. 선택적으로, 융합 단백질은 hGAA에 고유하지 않은 신호 펩타이드를 더 포함한다. 추가로 선택적으로, 이들 실시형태 중 하나는 길이가 약 5, 10, 15 또는 20개 아미노산인 C-말단 절단을 더 포함할 수 있다.

[0026] 특정 실시형태에서, hGAA780I 단백질을 포함하는 융합 단백질은 780번 위치에 Ile를 갖는 서열번호 3(hGAA780I)의 적어도 204번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA780I 단백질은 780번 위치에 Ile를 갖는 서열번호 3의 적어도 204번 아미노산 내지 952번 아미노산 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA780I 단백질은 780번 위치에 Ile를 갖는 서열번호 3의 적어도 123번 아미노산 내지 890번 아미노산 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA780I 효소는 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 952번 아미노산 또는 이와 적어도 95% 동일한 서열을 포함한다. 특정 실시형태에서, hGAA780I 단백질은 780번 위치에 Ile를 갖는 서열번호 3의 적어도 70번 아미노산 내지 890번 아미노산, 또는 이와 적어도 95% 동일한 서열을 포

함한다.

- [0027] 특정 실시형태에서, 융합 단백질은 신호 및 리더 서열 및 서열번호 3과 적어도 95% 동일성, 적어도 97% 동일성, 또는 적어도 99% 동일성을 갖는 hGAA780I 서열을 포함하고, 활성 부위에 변화가 없고/없거나 hGAA780I에 대한 N-말단 및/또는 C-말단의 3개 아미노산 내지 12개 아미노산에 변화가 없다. 바람직한 실시형태에서, 조작된 hGAA 발현 카세트는 적어도, T - Val(V) - P - Ile(780I) - Glu(E) - Ala(A) - Leu(L)(서열번호 29)의 인간 hGAA780I 단편을 인코딩한다. 특정 실시형태에서, 조작된 hGAA 발현 카세트는, Gln(Q) - T - V - P - 780I - E - A - L - Gly(G)(서열번호 30)의 더 긴 인간 hGAA780I 단편을 인코딩한다. 특정 실시형태에서, 조작된 hGAA 발현 카세트는 적어도 PLGT - Trp(W) - Tyr(Y) - Asp(D) - LQTVP - 780I - EALG - (Ser 또는 S) - L - PPPAA 서열(서열번호 31)에 해당하는 단편을 인코딩한다. 유사하게, 바람직한 실시형태에서, 활성 결합 부위(서열번호 3의 aa 518 내지 521번)에 아미노산 변화가 없다. 특정 실시형태에서, 600, 616, 및/또는 674번 위치에서의 결합 부위는 변화되지 않은 상태로 유지된다. 특정 실시형태에서, 명명된 융합 단백질은 신호 펩타이드, 선택적인 vIGF+2GS 연장, 선택적인 ER 단백질분해 펩타이드, 및 hGAA의 처음 35개 아미노산의 결실이 있는(즉, 천연 신호 펩타이드 및 28 내지 35번 아미노산이 결여됨) hGAA780I 변이체를 포함한다.
- [0028] 특정 실시형태에서, 융합 단백질은 다른 절단된 hGAA780I 단백질을 사용하여(예를 들어, 본 명세서에 참조에 의해 인용되는 WO 제2018/0467741호(Genethon)에 기재된 절단을 사용하여) 삭제될 수 있다. 이와 같은 일 실시형태는, 키모트립신 B2 신호 펩타이드, ER 단백질분해 펩타이드, 및 hGAA780I의 57 내지 952번 아미노산(hGAA780I의 1 내지 56번 아미노산의 결실이 있음)을 포함하는 sp7-Δ8.hGAA780I로 명명된 분비형 hGAA이다. 특정 실시형태에서, hGAA780I는 서열번호 3의 단편을 포함하는 융합 단백질이다. 특정 실시형태에서, 서열번호 3의 단편은 서열번호 4의 서열, 또는 이와 적어도 95% 동일한 서열에 의해 인코딩된다. 특정 실시형태에서, 서열번호 3의 단편은 서열번호 5의 서열, 또는 이와 적어도 95% 동일한 서열에 의해 인코딩된다.
- [0029] 본 명세서 전반에 걸쳐, 다양한 발현 카세트, 벡터 게놈, 벡터, 및 조성물이 hGAA780I 코딩 서열 또는 hGAA780I 단백질 또는 융합 단백질을 포함하는 것으로 기재되어 있다. 달리 명시되지 않는 한, 본 명세서에 기재된 것과 같은 N-말단 절단, C-말단 절단, 및 융합 단백질을 포함하여 조작된 hGAA780I 단백질, 또는 이에 대한 코딩 서열 중 임의의 것이 발현 카세트, 벡터 게놈, 벡터, 및 조성물로 유사하게 조작될 수 있음을 이해할 것이다.
- [0030] 적합하게는, 본 명세서에 기재된 핵산 서열을 포함하는 발현 카세트가 제공된다.
- [0031] 발현 카세트
- [0032] 본 명세서에서 사용되는 바와 같이, "발현 카세트"는 표적 세포(예를 들어, 프로모터)에서 발현을 지시하는 조절 서열에 작동 가능하게 연결된 기능성 유전자 산물을 인코딩하는 핵산 서열(예를 들어, hGAA780I 코딩 서열)을 포함하는 핵산 분자를 지칭하고, 이에 대한 다른 조절 서열을 포함할 수 있다. 필요한 조절 서열은 표적 세포에서 hGAA780I 코딩 서열의 전사, 번역 및/또는 발현을 허용하는 방식으로 당해 코딩 서열에 작동 가능하게 연결되어 있다.
- [0033] 특정 실시형태에서, 발현 카세트는 비번역 영역(들)에 하나 이상의 miRNA 표적 서열을 포함할 수 있다. miRNA 표적 서열은 이식유전자 발현이 바람직하지 않고/않거나 감소된 수준의 이식유전자 발현이 요구되는 세포에 존재하는 miRNA에 의해 특이적으로 인식되도록 설계된다. 특정 실시형태에서, 발현 카세트는 배근신경절에서 hGAA780I 융합 단백질의 발현을 특이적으로 감소시키는 miRNA 표적 서열을 포함한다. 특정 실시형태에서, miRNA 표적 서열은 3' UTR, 5' UTR, 및/또는 3' 및 5' UTR 둘 다에 위치한다. 특정 실시형태에서, 발현 카세트는 배근신경절(DRG)-특이적 miRNA 서열의 적어도 2개의 탠덤 반복부를 포함하며, 여기서 적어도 2개의 탠덤 반복부는 동일하거나 상이할 수 있는 적어도 제1 miRNA 표적 서열 및 적어도 제2 miRNA 표적 서열을 포함한다. 특정 실시형태에서, 적어도 2개의 drg-특이적 miRNA 탠덤 반복부 중 첫 번째의 시작은 hGAA780I 융합 단백질-코딩 서열의 3' 말단으로부터 20개의 뉴클레오타이드 이내이다. 특정 실시형태에서, 적어도 2개의 DRG-특이적 miRNA 탠덤 반복부 중 첫 번째의 시작은 hGAA780I 융합 단백질 코딩 서열의 3' 말단으로부터 적어도 100개의 뉴클레오타이드이다. 특정 실시형태에서, miRNA 탠덤 반복부는 200 내지 1200개의 뉴클레오타이드 길이를 포함한다. 특정 실시형태에서, miR 표적의 포함은 miR 표적 서열이 결여된 발현 카세트 또는 벡터 게놈에 비해, 하나 이상의 표적 조직에서 치료용 이식유전자의 발현 또는 효능을 변형시키지 않는다.
- [0034] 특정 실시형태에서, 벡터 게놈 또는 발현 카세트는 miR-183 표적 서열인 적어도 하나의 miRNA 표적 서열을 포함한다. 특정 실시형태에서, 벡터 게놈 또는 발현 카세트는 AGTGAATTCTACCAGTGCCATA(서열번호 26)를 포함하는 miR-183 표적 서열을 포함하며, 여기서 miR-183 시드 서열에 상보적인 서열은 밑줄이 그어져 있다. 특정 실시

태에서, 벡터 게놈 또는 발현 카세트는 miR-183 시드 서열에 100% 상보적인 서열의 하나 초과와 복제물(예를 들어, 2개 또는 3개 복제물)을 포함한다. 특정 실시형태에서, miR-183 표적 서열은 길이가 약 7개의 뉴클레오타이드 내지 약 28개의 뉴클레오타이드이고 miR-183 시드 서열에 적어도 100% 상보적인 적어도 하나의 영역을 포함한다. 특정 실시형태에서, miR-183 표적 서열은 서열번호 26에 부분적인 상보성을 갖는 서열을 포함하고, 따라서 서열번호 2에 대해 정렬될 때, 하나 이상의 미스매치가 있다. 특정 실시형태에서, miR-183 표적 서열은 서열번호 26에 대해 정렬될 때 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 미스매치를 갖는 서열을 포함하며, 여기서 미스매치는 비-연속적일 수 있다. 특정 실시형태에서, miR-183 표적 서열은 miR-183 표적 서열 길이의 적어도 30%를 또한 포함하는 100% 상보성의 영역을 포함한다. 특정 실시형태에서, 100% 상보성의 영역은 miR-183 시드 서열에 100% 상보성을 갖는 서열을 포함한다. 특정 실시형태에서, miR-183 표적 서열의 나머지는 miR-183에 적어도 약 80% 내지 약 99% 상보성을 가진다. 특정 실시형태에서, 발현 카세트 또는 벡터 게놈은 절단된 서열번호 26, 즉, 서열번호 26의 5' 또는 3' 말단 중 어느 하나 또는 둘 다에서 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 뉴클레오타이드가 결여된 서열을 포함하는 miR-183 표적 서열을 포함한다. 특정 실시형태에서, 발현 카세트 또는 벡터 게놈은 이식유전자 및 하나의 miR-183 표적 서열을 포함한다. 또 다른 실시형태에서, 발현 카세트 또는 벡터 게놈은 적어도 2, 3 또는 4개의 miR-183 표적 서열을 포함한다.

[0035] 특정 실시형태에서, 벡터 게놈 또는 발현 카세트는 miR-182 표적 서열인 적어도 하나의 miRNA 표적 서열을 포함한다. 특정 실시형태에서, 벡터 게놈 또는 발현 카세트는 AGTGTGAGTCTACCATGCGAAA(서열번호 27)를 포함하는 miR-182 표적 서열을 포함한다. 특정 실시형태에서, 벡터 게놈 또는 발현 카세트는 miR-182 시드 서열에 100% 상보적인 서열의 하나 초과와 복제물(예를 들어, 2개 또는 3개 복제물)을 포함한다. 특정 실시형태에서, miR-182 표적 서열은 길이가 약 7개의 뉴클레오타이드 내지 약 28개의 뉴클레오타이드이고 miR-182 시드 서열에 적어도 100% 상보적인 적어도 하나의 영역을 포함한다. 특정 실시형태에서, miR-182 표적 서열은 서열번호 27에 부분적인 상보성을 갖는 서열을 포함하고, 따라서 서열번호 27에 대해 정렬될 때, 하나 이상의 미스매치가 있다. 특정 실시형태에서, miR-183 표적 서열은 서열번호 27에 대해 정렬될 때 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개 미스매치를 갖는 서열을 포함하며, 여기서 미스매치는 비-연속적일 수 있다. 특정 실시형태에서, miR-182 표적 서열은 miR-182 표적 서열 길이의 적어도 30%를 또한 포함하는 100% 상보성의 영역을 포함한다. 특정 실시형태에서, 100% 상보성의 영역은 miR-182 시드 서열에 100% 상보성을 갖는 서열을 포함한다. 특정 실시형태에서, miR-182 표적 서열의 나머지는 miR-182에 적어도 약 80% 내지 약 99% 상보성을 가진다. 특정 실시형태에서, 발현 카세트 또는 벡터 게놈은 절단된 서열번호 27, 즉, 서열번호 27의 5' 또는 3' 말단 중 어느 하나 또는 둘 다에서 적어도 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 뉴클레오타이드가 결여된 서열을 포함하는 miR-182 표적 서열을 포함한다. 특정 실시형태에서, 발현 카세트 또는 벡터 게놈은 이식유전자 및 하나의 miR-182 표적 서열을 포함한다. 또 다른 실시형태에서, 발현 카세트 또는 벡터 게놈은 적어도 2, 3 또는 4개의 miR-182 표적 서열을 포함한다.

[0036] 용어 "탠덤 반복부"는 본 명세서에서 2개 이상의 연속적인 miRNA 표적 서열의 존재를 지칭하는 데 사용된다. 이러한 miRNA 표적 서열은 연속적일 수 있으며, 즉, 하나의 3' 말단이 개재 서열 없이 다음의 5' 말단의 바로 상류에 있거나, 그 반대이도록 서로 바로 뒤에 위치할 수 있다. 또 다른 실시형태에서, 2개 이상의 miRNA 표적 서열은 짧은 스페이스 서열에 의해 분리된다.

[0037] 본 명세서에서 사용되는 바와 같이, "스페이스"는 2개 이상의 연속적인 miRNA 표적 서열 사이에 위치한, 예를 들어, 길이가 1, 2, 3, 4, 5, 6, 7, 8, 9 또는 10개의 뉴클레오타이드인 임의의 선택된 핵산 서열이다. 특정 실시형태에서, 스페이스는 길이가 1 내지 8개의 뉴클레오타이드, 길이가 2 내지 7개의 뉴클레오타이드, 길이가 3 내지 6개의 뉴클레오타이드, 길이가 4개의 뉴클레오타이드, 4 내지 9개의 뉴클레오타이드, 3 내지 7개의 뉴클레오타이드, 또는 더 긴 값이다. 적합하게는, 스페이스는 비-코딩 서열이다. 특정 실시형태에서, 스페이스는 네(4)개의 뉴클레오타이드를 가질 수 있다. 특정 실시형태에서, 스페이스는 GGAT이다. 특정 실시형태에서, 스페이스는 여섯(6)개의 뉴클레오타이드이다. 특정 실시형태에서, 스페이스는 CACGTG 또는 GCATGC이다.

[0038] 특정 실시형태에서, 탠덤 반복부는 2, 3, 4개 또는 그 이상의 동일한 miRNA 표적 서열을 포함한다. 특정 실시형태에서, 탠덤 반복부는 적어도 2개의 상이한 miRNA 표적 서열, 적어도 3개의 상이한 miRNA 표적 서열, 또는 적어도 4개의 상이한 miRNA 표적 서열 등을 포함한다. 특정 실시형태에서, 탠덤 반복부는 2개 또는 3개의 동일한 miRNA 표적 서열 및 상이한 제4 miRNA 표적 서열을 포함할 수 있다.

[0039] 특정 실시형태에서, 발현 카세트에 탠덤 반복부의 적어도 2개의 상이한 세트가 있을 수 있다. 예를 들어, 3' UTR은 이식유전자의 바로 하류에 탠덤 반복부, UTR 서열, 및 UTR의 3' 말단에 더 가까운 2개 이상의 탠덤 반복부를 포함할 수 있다. 또 다른 예에서, 5' UTR은 1, 2개 또는 그 이상의 miRNA 표적 서열을 포함할 수 있다. 또

다른 예에서, 3'은 탠덤 반복부를 포함할 수 있고 5' UTR은 적어도 1개의 miRNA 표적 서열을 포함할 수 있다.

- [0040] 특정 실시형태에서, 발현 카세트는 이식유전자에 대한 정지 코돈의 약 0 내지 20개의 뉴클레오타이드 내에서 시작하는 2, 3, 4개 또는 그 이상의 탠덤 반복부를 포함한다. 다른 실시형태에서, 발현 카세트는 이식유전자에 대한 정지 코돈으로부터 적어도 100 내지 약 4000개의 뉴클레오타이드의 miRNA 탠덤 반복부를 포함한다.
- [0041] 미국 가특허 출원 제62/783,956호(출원일: 2018년 12월 21일, 본 명세서에 참조에 의해 인용됨)에 대해 우선권을 주장하고, 본 명세서에 참조에 의해 인용되는 PCT/US19/67872호(출원일: 2019년 12월 20일)를 참조한다.
- [0042] 본 명세서에서 사용되는 바와 같이, "작동 가능하게 연결된" 서열은 hGAA780I 코딩 서열과 인접한 발현 제어 서열 및 hGAA780I 코딩 서열을 제어하기 위해 트랜스로 또는 떨어져서 작용하는 발현 제어 서열을 둘 다 포함한다. 이와 같은 조절 서열은 전형적으로, 예를 들어, 프로모터, 인핸서, 인트론, 코작(Kozak) 서열, 폴리 아데닐화 서열, 및 TATA 신호 중 하나 이상을 포함한다.
- [0043] 특정 실시형태에서, 조절 요소는 다수의 표적 세포를 치료하는 데 적합한 단일 발현 카세트의 작제 및 전달을 허용하기 위해, 폼페병에 의해 이환된 다수의 세포 및 조직에서 발현을 지시한다. 예를 들어, 간, 골격근, 심장 및 중추신경계 세포 중 둘 이상에서 발현하는 조절 요소(예를 들어, 프로모터)가 선택될 수 있다. 예를 들어, 중추신경계(예를 들어, 뇌) 세포, 및 골격근에서 발현하는 조절 요소(예를 들어, 프로모터)가 선택될 수 있다. 다른 실시형태에서, 조절 요소는 CNS, 골격근 및 심장에서 발현한다. 다른 실시형태에서, 발현 카세트는 간, 골격근, 심장 및 중추신경계 세포 모두에서 인코딩된 hGAA780I의 발현을 허용한다. 다른 실시형태에서, 조절 요소는 (예를 들어, 본 명세서에 기재된 drg-탈표적화 시스템의 사용 및/또는 조직-특이적 프로모터의 선택에 의해) 특정 조직을 표적화하고 특정 세포 또는 조직에서 발현을 회피하도록 선택될 수 있다. 특정 실시형태에서, 상이한 조직을 우선적으로 표적화하는 본 명세서에서 제공되는 상이한 발현 카세트가 환자에게 투여된다.
- [0044] 조절 서열은 프로모터를 포함한다. 표적화된 세포에서 hGAA780I 단백질을 발현할 프로모터를 포함하지만, 이로 제한되지 않는 적합한 프로모터가 선택될 수 있다.
- [0045] 특정 실시형태에서, 구성 프로모터 또는 유도성/조절 프로모터가 선택된다. 구성 프로모터의 예는 닭 베타-액틴 프로모터이다. 다양한 닭 베타-액틴 프로모터가 단독으로, 또는 다양한 인핸서 요소와 조합하여 기재되었다(예를 들어, CB7은 거대세포바이러스 인핸서 요소; 프로모터, 닭 베타 액틴의 제1 엑손 및 제1 인트론, 및 토끼 베타-글로빈 유전자의 스플라이스 수용체를 포함하는 CAG 프로모터; CBh 프로모터가 있는 닭 베타-액틴 프로모터 임, 문헌[SJ Gray et al, Hu Gene Ther, 2011 Sep; 22(9): 1143-1153]). 특정 실시형태에서, 조절 가능한 프로모터가 선택될 수 있다. 예를 들어, 본 명세서에 참조에 의해 인용되는 WO 제2011/126808B2호를 참조한다.
- [0046] 특정 실시형태에서, 조직-특이적 프로모터가 선택될 수 있다. 조직-특이적인 프로모터의 예는 간(알부민, 문헌[Miyatake et al., (1997) J. Virol., 71:5124-32]; B형 간염 바이러스 코어 프로모터, 문헌[Sandig et al., (1996) Gene Ther., 3:1002-9]; 알파-태아단백질(AFP), 문헌[Arbutnot et al., (1996) Hum. Gene Ther., 7:1503-14]), 중추신경계, 예를 들어, 뉴런(예컨대, 뉴런-특이적 에놀라제(NSE) 프로모터, 문헌[Andersen et al., (1993) Cell. Mol. Neurobiol., 13:503-15]; 신경미세섬유 경쇄 유전자, 문헌[Piccioli et al., (1991) Proc. Natl. Acad. Sci. USA, 88:5611-5]; 및 뉴런-특이적 vgf 유전자, 문헌[Piccioli et al., (1995) Neuron, 15:373-84]), 심장 근육, 골격근, 폐, 및 다른 조직에 대하여 잘 알려져 있다. 또 다른 실시형태에서, 적합한 프로모터는 연장 인자 1 알파(EF1 알파) 프로모터(예를 들어, 문헌[Kim DW et al, Use of the human elongation factor 1 alpha promoter as a versatile and efficient expression system. Gene. 1990 Jul 16;91(2):217-23] 참조), 시냅신 1 프로모터(예를 들어, 문헌[Kügler S et al, Human synapsin 1 gene promoter confers highly neuron-specific long-term transgene expression from an adenoviral vector in the adult rat brain depending on the transduced area. Gene Ther. 2003 Feb;10(4):337-47] 참조), 뉴런-특이적 에놀라제(NSE) 프로모터(예를 들어, 문헌[Kim J et al, Involvement of cholesterol-rich lipid rafts in interleukin-6-induced neuroendocrine differentiation of LNCaP prostate cancer cells. Endocrinology. 2004 Feb;145(2):613-9. Epub 2003 Oct 16] 참조), 또는 CB6 프로모터(예를 들어, 문헌[Large-Scale Production of Adeno-Associated Viral Vector Serotype-9 Carrying the Human Survival Motor Neuron Gene, Mol Biotechnol. 2016 Jan;58(1):30-6. doi: 10.1007/s12033-015-9899-5] 참조)를 포함할 수 있지만, 이들로 제한되지 않는다. 조직-특이적 프로모터를 이용하는 특정 실시형태에서, 상이한 세포 유형을 표적화하는 조직-특이적 프로모터를 갖는 상이한 발현 카세트를 포함하는 공동-요법이 선택될 수 있다.
- [0047] 일 실시형태에서, 조절 서열은 인핸서를 더 포함한다. 일 실시형태에서, 조절 서열은 하나의 인핸서를

포함한다. 또 다른 실시형태에서, 조절 서열은 2개 이상의 발현 인핸서를 포함한다. 이들 인핸서는 동일할 수 있거나 상이할 수 있다. 예를 들어, 인핸서는 알파 mic/bik 인핸서 또는 CMV 인핸서를 포함할 수 있다. 이러한 인핸서는 서로 인접하게 위치한 2개의 복제물로 존재할 수 있다. 대안적으로, 인핸서의 이중 복제물은 하나 이상의 서열에 의해 분리될 수 있다.

[0048] 일 실시형태에서, 조절 서열은 인트론을 더 포함한다. 추가 실시형태에서, 인트론은 닭 베타-액틴 인트론이다. 다른 적합한 인트론은 인간 β -글로불린 인트론, 및/또는 상업적으로 입수가 가능한 Promega® 인트론, 및 WO 제 2011/126808호에 기재된 것에 의해 당업계에 알려진 것을 포함한다.

[0049] 일 실시형태에서, 조절 서열은 폴리아데닐화 신호(폴리A)를 더 포함한다. 추가 실시형태에서, 폴리A는 토끼 글로빈 폴리 A이다. 예를 들어, WO 제2014/151341호를 참조한다. 대안적으로, 또 다른 폴리A, 예를 들어, 인간 성장 호르몬(hGH) 폴리아데닐화 서열, SV40 폴리A, 또는 합성 폴리A가 발현 카세트에 포함될 수 있다.

[0050] 발현 카세트는 임의의 적합한 전달 시스템을 통해 전달될 수 있다. 적합한 비-바이러스 전달 시스템은 당업계에 알려져 있으며(예를 들어, 문헌[Ramamoorth and Narvekar. J Clin Diagn Res. 2015 Jan; 9(1):GE01-GE06] 참조, 이는 본 명세서에 참조에 의해 인용됨) 당업자에 의해 용이하게 선택될 수 있고, 예를 들어, 네이키드 DNA, 네이키드 RNA, 덴드리머, PLGA, 폴리메트아크릴레이트, 무기 입자, 지질 입자(예를 들어, 지질 나노입자 또는 LNP), 또는 키토산-기반 제형을 포함할 수 있다.

[0051] 일 실시형태에서, 벡터는 이의 기재된 발현 카세트, 예를 들어, "네이키드 DNA", "네이키드 플라스미드 DNA", RNA 및 mRNA; 예를 들어, 미셀, 리소솜, 양이온성 지질-핵산 조성물, 폴리-글리칸 조성물 및 다른 중합체, 지질 및/또는 콜레스테롤-기반 핵산 컨쥬게이트를 포함하는 나노 입자 및 다양한 조성물과 커플링된 발현 카세트, 및 본 명세서에 기재된 바와 같은 다른 작제물을 포함하는 비-바이러스 플라스미드이다. 예를 들어, 문헌[X. Su et al, Mol. Pharmaceutics, 2011, 8 (3), pp 774-787; 웹 공개: 2011년 3월 21일]; WO 제2013/182683호, WO 제 2010/053572호 및 WO 제2012/170930호를 참조하며, 이들 모두는 본 명세서에 참조에 의해 인용된다.

[0052] 특정 실시형태에서, 본 명세서에 기재된 바와 같은 hGAA780I 변이체, 융합 단백질, 또는 절단된 단백질을 인코딩하는 서열을 갖는 핵산 분자가 본 명세서에서 제공된다. 바람직한 일 실시형태에서, hGAA780I는 서열번호 4의 조작된 서열 또는 hGAA780I 변이체를 인코딩하는 이와 적어도 95% 동일한 서열에 의해 인코딩된다. 특정 실시형태에서, 서열번호 4는 780I 위치에서 Ile를 인코딩하는 코돈이 ATT 또는 ATC가 되도록 변형된다. 특정 실시형태에서, 서열번호 4의 조작된 서열을 포함하는 핵산, 또는 이의 단편은 융합 단백질 또는 절단된 hGAA780I를 발현하는 데 사용된다. 덜 바람직하지만, 특정 실시형태에서, hGAA780I는 서열번호 5에 의해 인코딩된다. 특정 실시형태에서, 핵산 분자는 플라스미드이다.

[0053] 본 명세서에 기재된 발현 카세트의 구성은 본 명세서 전반에 걸쳐 기재된 다른 구성, 요법, 양태, 실시형태 및 방법에 적용되는 것으로 의도됨이 이해되어야 한다.

[0054] 벡터

[0055] 본 명세서에서 사용되는 바와 같이 "벡터"는 핵산 서열의 복제 또는 발현을 위해 적절한 표적 세포로 도입될 수 있는 핵산 서열을 포함하는 생물학적 또는 화학적 모이어티이다. 벡터의 예는 재조합 바이러스, 플라스미드, 리포플렉스, 폴리머솜, 폴리플렉스, 덴드리머, 세포 투과 펩타이드(CPP) 컨쥬게이트, 자성 입자, 또는 나노입자를 포함하지만, 이들로 제한되지 않는다. 일 실시형태에서, 벡터는 이후에 적절한 표적 세포 내로 도입될 수 있는, 기능성 유전자 산물을 인코딩하는 외인성 또는 이중성 조작된 핵산을 갖는 핵산 분자이다. 이와 같은 벡터는 바람직하게는 하나 이상의 복제 기원, 및 재조합 DNA가 삽입될 수 있는 하나 이상의 부위를 가진다. 벡터는 종종 벡터를 갖는 세포가 벡터가 없는 세포로부터 선택될 수 있는 수단을 가지며, 예를 들어, 벡터는 약물 내성 유전자를 인코딩한다. 일반적인 벡터는 플라스미드, 바이러스 계통, 및 "인공 염색체"를 포함한다. 벡터의 생성, 생산, 특성화, 또는 정량화의 통상적인 방법은 당업자에게 이용 가능하다.

[0056] 특정 실시형태에서, 본 명세서에 기재된 벡터는 기능성 hGAA780I 유전자 산물 또는 hGAA780I 융합 단백질을 인코딩하는 핵산 서열을 포함하는 발현 카세트가 바이러스 캡시드 또는 외피에 패키징되는 합성 또는 인공 바이러스 입자를 지칭하는 "복제-결함 바이러스" 또는 "바이러스 벡터"이며, 여기서 또한 바이러스 캡시드 또는 외피 내에 패키징되는 임의의 바이러스 계통 서열은 복제-결함이고; 즉, 이들은 자손 비리온을 생성할 수 없지만 표적 세포를 감염시키는 능력은 보유한다. 일 실시형태에서, 바이러스 벡터의 계통은 복제에 필요한 효소를 인코딩하는 유전자를 포함하지 않지만(계통은 "무기력"(인공 계통의 증폭 및 패키징에 필요한 신호가 축적한 인코딩하는 핵산 서열만을 포함함)하도록 조작될 수 있음), 이들 유전자는 생성 동안 공급될 수 있다. 따라서, 복제에

필요한 바이러스 효소가 존재하는 경우를 제외하고 자손 비리온에 의한 복제 및 감염이 일어나지 못하므로 유전자 요법에 사용하기에 안전하다고 여겨진다.

[0057] 본 명세서에서 사용되는 바와 같이, 재조합 바이러스 벡터는 원하는 세포(들)를 표적화하는 임의의 적합한 바이러스 벡터이다. 따라서, 재조합 바이러스 벡터는 바람직하게는, 중추신경계(예를 들어, 뇌), 골격근, 심장, 및/또는 간을 포함하는 품폐병에 의해 이환된 세포 및 조직 영향 중 하나 이상을 표적화한다. 특정 실시형태에서, 바이러스 벡터는 적어도 중추신경계(예를 들어, 뇌), 폐, 심장 세포, 또는 골격근을 표적화한다. 다른 실시형태에서, 바이러스 벡터는 CNS(예를 들어, 뇌), 골격근 및/또는 심장을 표적화한다. 다른 실시형태에서, 바이러스 벡터는 간, 골격근, 심장 및 중추신경계 세포를 모두 표적화한다. 예는 예시적인 재조합 아데노-연관 바이러스(rAAV)를 제공한다. 그러나, 다른 적합한 바이러스 벡터는, 예를 들어, 재조합 아데노바이러스, 재조합 파코바이러스, 에코바이러스, 재조합 보카바이러스, 하이브리드 AAV/보카바이러스, 재조합 단순 헤르페스 바이러스, 재조합 레트로바이러스, 또는 재조합 렌티바이러스를 포함할 수 있다. 바람직한 실시형태에서, 이들 재조합 바이러스는 복제-불능이다.

[0058] 본 명세서에서 사용되는 바와 같이, 용어 "숙주 세포"는 벡터(예를 들어, 재조합 AAV)가 생성되는 패키징 세포주를 지칭할 수 있다. 숙주 세포는 임의의 수단, 예를 들어, 전기천공, 인산칼슘 침전, 미세주입, 형질전환, 바이러스 감염, 형질감염, 리포솜 전달, 막 융합 기법, 고속 DNA-코팅 펄스, 바이러스 감염 및 원형질체 융합에 의해 세포 내로 도입된 외인성 또는 이중성 DNA를 포함하는 원핵 또는 진핵 세포(예를 들어, 인간, 곤충, 또는 효모)일 수 있다. 숙주 세포의 예는 단리된 세포, 세포 배양물, 대장균(*Escherichia coli*) 세포, 효모 세포, 인간 세포, 비-인간 세포, 포유동물 세포, 비-포유동물 세포, 곤충 세포, HEK-293 세포, 간 세포, 신장 세포, 중추신경계의 세포, 뉴런, 신경아교세포, 또는 줄기 세포를 포함할 수 있지만, 이들로 제한되지 않는다.

[0059] 특정 실시형태에서, 숙주 세포는 단백질이 단리 또는 정제를 위해 시험관내에서 충분한 양으로 생성되도록 hGAA780I의 생성을 위한 발현 카세트를 포함한다. 특정 실시형태에서, 숙주 세포는 hGAA780I, 또는 이의 단편을 인코딩하는 발현 카세트를 포함한다. 본 명세서에서 제공되는 바와 같이, hGAA780I는 치료제(즉, 효소 대체 요법)로서 대상체에게 투여되는 약제학적 조성물에 포함될 수 있다.

[0060] 본 명세서에서 사용되는 바와 같이, 용어 "표적 세포"는 기능성 유전자 산물의 발현이 요구되는 임의의 표적 세포를 지칭한다.

[0061] 본 명세서에서 사용되는 바와 같이, "벡터 계놈"은 바이러스 벡터 내부에 패키징된 핵산 서열을 지칭한다. 일 예에서, "벡터 계놈"은 최소한 5'에서 3'으로 벡터-특이적 서열, 표적 세포에서 발현을 지시하는 조절 제어 서열에 작동 가능하게 연결된 기능성 유전자 산물(예를 들어, hGAA780I, 융합 단백질 hGAA780I, 또는 또 다른 단백질)을 인코딩하는 핵산 서열, 벡터-특이적 서열, 및 선택적으로 비번역 영역(들) 및 벡터-특이적 서열의 miRNA 표적 서열을 포함한다. 벡터-특이적 서열은 벡터 계놈을 바이러스 벡터 캡시드 또는 외피 단백질 내로 특이적으로 패키징하는 말단 반복 서열일 수 있다. 예를 들어, AAV 말단 말단 반복부는 AAV 및 다른 특정 파코바이러스 캡시드 내로 패키징하는 데 이용된다. 렌티바이러스 긴 말단 반복부는 렌티바이러스 벡터 내로의 패키징이 요구되는 경우에 이용될 수 있다. 유사하게, 다른 말단 반복부(예를 들어, 레트로바이러스 긴 말단 반복부) 등이 선택될 수 있다.

[0062] 본 명세서에 기재된 벡터의 구성은 본 명세서 전반에 걸쳐 기재된 다른 구성, 요법, 양태, 실시형태, 및 방법에 적용되는 것으로 의도됨이 이해되어야 한다.

[0063] 아데노-연관 바이러스(AAV)

[0064] 일 양태에서, 융합 단백질을 포함하여 본 명세서에 기재된 바와 같은 hGAAV780I 단백질(효소)을 인코딩하는 AAV 캡시드 및 내부에 패키징된 벡터 계놈을 포함하는 재조합 AAV(rAAV)가 본 명세서에서 제공된다. 특정 실시형태에서, 선택된 AAV 캡시드는 간, 근육, 신장, 심장 및/또는 중추신경계 세포 유형 중 둘 이상의 세포를 표적화한다. 특정 실시형태에서, 간, 골격근, 심장, 신장 및/또는 적어도 하나의 중추신경계 세포 유형 중 적어도 둘 이상에서 hGAA780I 단백질을 발현시키는 것이 바람직하다. 따라서, 일 실시형태에서 선택된 AAV 캡시드는 심장 조직을 표적화한다. 특정 실시형태에서, 심장 조직을 표적화하도록 선택된 AAV 캡시드는 AAV 1, 6, 8, 및 9로부터 선택된다(예를 들어, 문헌[Katz et al. Hum Gene Ther Clin Dev. 2017 Sep 1; 28(3): 157-164] 참조). 또 다른 실시형태에서, 선택된 AAV 캡시드는 신장의 세포를 표적화한다. 일 실시형태에서, 신장 세포를 표적화하기 위한 캡시드는 AAV1, 2, 6, 8, 9, 및 Anc80으로부터 선택된다(예를 들어, 문헌[Ikeda Y et al. J Am Soc Nephrol.

2018 Sep; 29(9):2287-2297 및 Ascio et al. Biochem Biophys Res Commun. 2018 Feb 26; 497(1): 19-24] 참조).

[0065] 일 실시형태에서, 벡터 계놈은 AAV 5' 반전 말단 반복부(ITR), 본 명세서에 기재된 바와 같은 발현 카세트, 및 AAV 3' ITR을 포함한다. 일 실시형태에서, 벡터 계놈은 rAAV 벡터를 형성하는 rAAV 캡시드 내부에 패키징된 핵산 서열을 지칭한다. 이와 같은 핵산 서열은 발현 카세트에 측접하는 AAV 반전 말단 반복부 서열(ITR)을 포함한다. 일 예에서, AAV 또는 보카바이러스 캡시드로 패키징하기 위한 "벡터 계놈"은 최소한 5'에서 3'으로 AAV 5' ITR, 표적 세포에서 발현을 지시하는 조절 제어 서열에 작동 가능하게 연결된 본 명세서에 기재된 바와 같은 기능성 hGAA780I(예를 들어, hGAA780I 융합 단백질)를 인코딩하는 핵산 서열 및 AAV 3' ITR을 포함한다. 특정 실시형태에서, ITR은 AAV2 유래이고 캡시드는 상이한 AAV 유래이다. 대안적으로, 다른 ITR이 사용될 수 있다. 특정 실시형태에서, 벡터 계놈은 이식유전자 발현이 바람직하지 않고/않거나 감소된 수준의 이식유전자 발현이 요구되는 세포에서 miRNA 서열에 의해 특이적으로 인식되도록 설계된 비번역 영역(들)에 miRNA 표적 서열을 더 포함한다.

[0066] ITR은 벡터 생성 동안 계놈의 복제 및 패키징을 담당하는 유전 요소이며 rAAV를 생성하는 데 필요한 유일한 바이러스 시스 요소이다. 일 실시형태에서, ITR은 캡시드를 공급하는 것과 상이한 AAV 유래이다. 바람직한 실시형태에서, AAV2 유래의 ITR 서열, 또는 이의 결실된 형태(Δ ITR)는 편의를 위해 그리고 규제 승인을 가속화하기 위해 사용될 수 있다. 그러나, 다른 AAV 출처 유래의 ITR이 선택될 수 있다. ITR의 출처가 AAV2이고 AAV 캡시드가 또 다른 AAV 출처 유래인 경우, 생성된 벡터는 위형(pseudotyped)으로 불릴 수 있다. 전형적으로, AAV 벡터 계놈은 AAV 5' ITR, hGAA780I 코딩 서열 및 임의의 조절 서열, 및 AAV 3' ITR을 포함한다. 그러나, 이들 요소의 다른 구성이 적합할 수 있다. D-서열 및 말단 분해 부위(terminal resolution site; trs)가 결실된 Δ ITR이라고 하는 5' ITR의 단축 형태가 설명된 바 있다. 다른 실시형태에서, 전장 AAV 5' 및 3' ITR이 사용된다.

[0067] 본 명세서에서 사용되는 바와 같이 용어 "AAV"는 자연 발생 아데노-연관 바이러스, 당업자에게 이용 가능한 그리고/또는 본 명세서에 기재된 구성(들) 및 방법(들)을 고려한 아데노-연관 바이러스뿐만 아니라, 인공 AAV를 지칭한다. 아데노-연관 바이러스(AAV) 바이러스 벡터는 표적 세포로 전달하기 위해 AAV 반전 말단 반복부 서열(ITR)이 측접한 발현 카세트에 패키징된 AAV 단백질 캡시드를 갖는 AAV 뉴클레아제(예를 들어, DNase)-내성 입자이다. 뉴클레아제-내성 재조합 AAV(rAAV)는 AAV 캡시드가 완전히 조립되었고 생성 공정에서 존재할 수 있는 오염 핵산을 제거하도록 설계된 뉴클레아제 인큐베이션 단계 동안 이들 패키징된 벡터 계놈 서열을 분해(소화)로부터 보호함을 나타낸다. 많은 경우에, 본 명세서에 기재된 rAAV는 DNase 내성이다.

[0068] AAV 캡시드의 출처는 임의의 수십 가지의 자연 발생 및 이용 가능한 아데노-연관 바이러스뿐만 아니라 조작된 AAV 중 하나일 수 있다. 아데노-연관 바이러스(AAV) 바이러스 벡터는 표적 세포로 전달하기 위해 내부에 핵산 서열이 패키징된 AAV 단백질 캡시드를 갖는 AAV DNase-내성 입자이다. AAV 캡시드는 선택된 AAV에 따라 대략 1:1:10 내지 1:1:20의 비율로 정20면체 대칭으로 배열된 60개의 캡시드(cap) 단백질 서브유닛, 즉, VP1, VP2, 및 VP3으로 구성된다. 다양한 AAV가 상기 확인된 바와 같이 AAV 바이러스 벡터의 캡시드에 대한 출처로서 선택될 수 있다. 예를 들어, 미국 특허 출원 공개 제2007/0036760 A1호; 미국 특허 출원 공개 제2009/0197338 A1호; EP 제1310571호를 참조한다. 또한, WO 제2003/042397호(AAV7 및 다른 유인원 AAV), 미국 특허 제7790449호 및 미국 특허 제7282199호(AAV8), WO 제2005/033321호 및 미국 제7,906,111호(AAV9), 및 WO 제2006/110689호, WO 제2003/042397호(rh.10) 및 WO 제2018/160582호(AAVhu68)를 참조한다. 이들 문헌은 또한 AAV를 생성하기 위해 선택될 수 있는 다른 AAV를 기재하며, 이들은 참조에 의해 원용된다. 달리 명시되지 않는 한, 본 명세서에 기재된 AAV 캡시드, ITR, 및 다른 선택된 AAV 구성요소는 AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV8bp, AAV7M8, AAVAnc80, AAVrh10, 및 AAVPHP.B로 일반적으로 식별되는 AAV 및 알려진 또는 언급된 AAV 중 임의의 것의 변이체 또는 아직 발견되지 않은 AAV 또는 이의 변이체 또는 혼합물을 포함하지만, 이들로 제한되지 않는 임의의 AAV로부터 용이하게 선택될 수 있다. 예를 들어, 본 명세서에 참조에 의해 원용된 WO 제2005/033321호를 참조한다. 일 실시형태에서, AAV 캡시드는 AAV1 캡시드 또는 이의 변이체, AAV8 캡시드 또는 이의 변이체, AAV9 캡시드 또는 이의 변이체, AAVrh.10 캡시드 또는 이의 변이체, AAVrh64R1 캡시드 또는 이의 변이체, AAVhu.37 캡시드 또는 이의 변이체, 또는 AAV3B 또는 이의 변이체이다. 일 양태에서, 캡시드는 AAVhu.37 캡시드이다. 특정 실시형태에서, 캡시드 단백질은 rAAV 벡터의 명칭에서 용어 "AAV" 다음에 오는 숫자 또는 숫자 및 문자의 조합에 의해 지정된다. 또한, 각각 명칭이 "Novel Adeno-Associated Virus (AAV) Vectors, AAV Vectors Having Reduced Capsid Deamidation And Uses Therefor"이고 2019년 2월 27일에 출원된 PCT/US19/19804호 및 PCT/US19/19861호를 참조하며, 이들은 본 명세서에 전문이 참조에 의해 원용된다.

[0069] ITR 또는 다른 AAV 구성요소는 당업자에게 이용 가능한 기법을 사용하여 AAV로부터 용이하게 단리되거나 조작될

수 있다. 이와 같은 AAV는 학문적, 상업적, 또는 공공 출처(예를 들어, 아메리칸 타입 컬처 컬렉션(American Type Culture Collection), 미국 버지니아주 매너서스 소재)로부터 단리되거나, 조작되거나, 입수할 수 있다. 대안적으로, AAV 서열은, 예를 들어, GenBank, PubMed 등과 같은 데이터베이스 또는 문헌에서 이용 가능한 것과 같은 공개된 서열을 참조하여 합성 또는 다른 적합한 수단을 통해 조작될 수 있다. AAV 바이러스는 통상적인 분자 생물학 기법에 의해 조작되어, 핵산 서열의 세포 특이적 전달, 면역원성의 최소화, 안정성 및 입자 수명 조정, 효율적인 분해, 핵으로의 정확한 전달 등을 위해 이러한 입자를 최적화하는 것을 가능하게 한다.

[0070] 본 명세서에서 사용되는 바와 같이, 호환 가능하게 사용되는 용어 "rAAV" 및 "인공 AAV"는 캡시드 단백질 및 그 안에 패키징된 벡터 게놈을 포함하는 AAV를 의미하지만, 이로 제한되지 않으며, 여기서 벡터 게놈은 AAV에 이종성인 핵산을 포함한다. 일 실시형태에서, 캡시드 단백질은 비-자연 발생 캡시드이다. 이와 같은 인공 캡시드는 선택된 AAV 서열(예를 들어, vp1 캡시드 단백질의 단편)을 상이한 선택된 AAV, 동일한 AAV의 비-연속적인 부분, 비-AAV 바이러스 출처, 또는 비-바이러스 출처로부터 얻을 수 있는 이종성 서열과 조합하여 사용함으로써 임의의 적합한 기법에 의해 생성될 수 있다. 인공 AAV는 위형 AAV, 키메라 AAV 캡시드, 재조합 AAV 캡시드, 또는 "인간화" AAV 캡시드일 수 있지만, 이들로 제한되지 않는다. 하나의 AAV의 캡시드가 이종 캡시드 단백질로 대체된 위형 벡터가 본 발명에 유용하다. 일 실시형태에서, AAV2/5 및 AAV2/8은 예시적인 위형 벡터이다. 선택된 유전 요소는 형질감염, 전기천공, 리포솜 전달, 막 융합 기법, 고속 DNA-코팅 펄스, 바이러스 감염 및 원형질체 융합을 포함하는 임의의 적합한 방법에 의해 전달될 수 있다. 이와 같은 작제물을 제조하는 데 사용되는 방법은 핵산 조작의 숙련자에게 알려져 있으며, 유전 공학, 재조합 공학, 및 합성 기법을 포함한다. 예를 들어, 문헌 [Green and Sambrook, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY (2012)]을 참조한다.

[0071] 특정 실시형태에서, AAV 캡시드는 자연 및 조작된 클레드 F 아데노-연관 바이러스로부터 선택된다. 하기 예에서, 클레드 F 아데노-연관 바이러스는 AAVhu68이다. 본 명세서에 전문이 참조에 의해 인용되는 WO 제 2018/160582호를 참조한다. 그러나, 다른 실시형태에서, AAV 캡시드는 상이한 클레드, 예를 들어, 클레드 A, B, C, D, 또는 E로부터, 또는 이들 클레드 중 임의의 것의 외부 AAV 출처로부터 선택된다.

[0072] 본 명세서에서 사용되는 바와 같이, AAV의 그룹과 관련되는 용어 "클레드"는 (적어도 1000개 복제물의) 적어도 75%의 부트스트랩 값에 의해 이웃 결합 알고리즘(Neighbor-Joining algorithm) 및 AAV vp1 아미노산 서열의 정렬을 기반으로 0.05 이하의 푸아송 보정 거리 측정을 사용하여 결정된 서로 계통 발생적으로 관련된 AAV의 그룹을 지칭한다. 이웃 결합 알고리즘은 문헌에 설명된 바 있다. 예를 들어, 문헌[M. Nei and S. Kumar, *Molecular Evolution and Phylogenetics* (Oxford University Press, New York (2000))]을 참조한다. 이 알고리즘을 구현하는 데 사용될 수 있는 컴퓨터 프로그램이 이용 가능하다. 예를 들어, MEGA v2.1 프로그램은 변형된 Nei-Gojobori 방법을 구현한다. 이러한 기법 및 컴퓨터 프로그램, 및 AAV vp1 캡시드 단백질의 서열을 사용하여, 당업자는 선택된 AAV가 본 명세서에서 식별된 클레드 중 하나에 포함되는지, 다른 클레드에 포함되는지, 또는 이들 클레드의 외부에 있는지 여부를 용이하게 결정할 수 있다. 예를 들어, 클레드 A, B, C, D, E 및 F를 식별하고, GenBank 수탁 번호 AY530553 내지 AY530629인 신규 AAV의 핵산 서열을 제공하는, 문헌[G Gao, *et al*, *J Virol*, 2004 Jun; 78(10): 6381-6388]을 참조한다. 또한, WO 제2005/033321호를 참조한다.

[0073] 본 명세서에서 사용되는 바와 같이, "AAV9 캡시드"는 (a) 본 명세서에 참조에 의해 인용되는 GenBank 수탁 번호 AAS99264의 아미노산 서열 및 AAV vp1 캡시드 단백질 및/또는 (b) GenBank 수탁 번호 AY530579.1(nt 1..2211)의 뉴클레오타이드 서열에 의해 인코딩되는 아미노산 서열을 갖는 AAV9를 지칭한다. 이러한 인코딩된 서열로부터의 일부 변이는 본 발명에 포함되며, 이는 GenBank 수탁 번호 AAS99264 및 US 제7906111호(또한 WO 제2005/033321호)의 참조된 아미노산 서열과 약 99% 동일성(즉, 참조된 서열로부터 약 1% 미만의 변이)을 갖는 서열을 포함할 수 있다. 이와 같은 AAV는, 예를 들어, 자연 분리주(예를 들어, hu31 또는 hu32), 또는 예를 들어, AAV9 캡시드와 정렬된 임의의 다른 AAV 캡시드에서 상응하는 위치로부터 "모집된" 대체 잔기로부터 선택되는 아미노산 치환을 포함하지만, 이로 제한되지 않는 아미노산 치환, 결실 또는 첨가가 있는 AAV9의 변이체; 예를 들어, US 제9,102,949호, US 제8,927,514호, US 제2015/349911호, WO 제2016/049230A1호, US 제9,623,120호, 및 제9,585,971호에 기재된 것을 포함할 수 있다. 그러나, 다른 실시형태에서, AAV9의 다른 변이체, 또는 상기 참조된 서열과 적어도 약 95% 동일성을 갖는 AAV9 캡시드가 선택될 수 있다. 예를 들어, US 제2015/0079038호를 참조한다. 캡시드를 생성하는 방법, 이에 대한 코딩 서열, 및 rAAV 바이러스 벡터의 생성 방법은 설명된 바 있다. 예를 들어, 문헌[Gao, *et al*, *Proc. Natl. Acad. Sci. U.S.A.* 100 (10), 6081-6086 (2003)] 및 US 제2013/0045186A1호를 참조한다.

[0074] 특정 실시형태에서, AAVhu68 캡시드는 명칭이 "Novel Adeno-associated virus (AAV) Clade F Vector and Uses

Therefor"인 WO 제2018/160582호에 기재된 바와 같으며, 이는 본 명세서에 참조에 의해 인용된다. 특정 실시형태에서, AAVhu68 캡시드 단백질은 서열번호 2의 1 내지 736번의 예상되는 아미노산 서열을 인코딩하는 핵산 서열로부터 발현에 의해 생성되는 AAVhu68 vp1 단백질, 서열번호 2로부터 생성되는 vp1 단백질 또는 서열번호 2의 1 내지 736번의 예상되는 아미노산 서열을 인코딩하는 서열번호 1과 적어도 70% 동일한 핵산 서열로부터 생성되는 vp1 단백질; 서열번호 2의 적어도 약 138 내지 736번 아미노산의 예상되는 아미노산 서열을 인코딩하는 핵산 서열로부터 발현에 의해 생성되는 AAVhu68 vp2 단백질, 서열번호 1의 적어도 412 내지 2211번 뉴클레오타이드를 포함하는 서열로부터 생성되는 vp2 단백질, 또는 서열번호 2의 적어도 약 138 내지 736번 아미노산의 예상되는 아미노산 서열을 인코딩하는 서열번호 1의 적어도 412 내지 2211번 뉴클레오타이드와 적어도 70% 동일한 핵산 서열로부터 생성되는 vp2 단백질, 및/또는 서열번호 2의 적어도 약 203 내지 736번 아미노산의 예상되는 아미노산 서열을 인코딩하는 핵산 서열로부터 발현에 의해 생성되는 AAVhu68 vp3 단백질, 서열번호 1의 적어도 607 내지 2211번 뉴클레오타이드를 포함하는 서열로부터 생성되는 vp3 단백질, 또는 서열번호 2의 적어도 약 203 내지 736번 아미노산의 예상되는 아미노산 서열을 인코딩하는 서열번호 1의 적어도 607 내지 2211번 뉴클레오타이드와 적어도 70% 동일한 핵산 서열로부터 생성되는 vp3 단백질을 포함한다.

[0075] AAVhu68 vp1, vp2 및 vp3 단백질은 전형적으로 서열번호 2의 전장 vp1 아미노산 서열(1 내지 736번 아미노산)을 인코딩하는 동일한 핵산 서열에 의해 인코딩되는 대안적인 스플라이스 변이체로서 발현된다. 선택적으로 vp1-인코딩 서열은 단독으로 vp1, vp2, 및 vp3 단백질을 발현하는 데 사용된다. 대안적으로, 이 서열은 vp1-고유 영역(약 aa 1번 내지 약 aa 137번) 및/또는 vp2-고유 영역(약 aa 1번 내지 약 aa 202번), 또는 이에 상보적인 가닥, 상응하는 mRNA(서열번호 1의 약 nt 607번 내지 약 nt 2211번), 또는 서열번호 2의 aa 203 내지 736번을 인코딩하는 서열번호 1과 적어도 70% 내지 적어도 99%(예를 들어, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 적어도 98% 또는 적어도 99%) 동일한 서열 없이 서열번호 2의 AAVhu68 vp3 아미노산 서열(약 aa 203 내지 736번)을 인코딩하는 핵산 서열 중 하나 이상과 공동 발현될 수 있다. 추가적으로, 또는 대안적으로, vp1-인코딩 및/또는 vp2-인코딩 서열은 vp1-고유 영역(약 aa 1번 내지 약 137번), 또는 이에 상보적인 가닥, 상응하는 mRNA(서열번호 1의 nt 412번 내지 nt 2211번), 또는 서열번호 2의 aa 138 내지 736번을 인코딩하는 서열번호 1의 nt 412 내지 2211번과 적어도 70% 내지 적어도 99%(예를 들어, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 적어도 98% 또는 적어도 99%) 동일한 서열 없이 서열번호 2의 AAVhu68 vp2 아미노산 서열(약 aa 138 내지 736번)을 인코딩하는 핵산 서열과 공동-발현될 수 있다.

[0076] 본 명세서에 기재된 바와 같이, rAAVhu68은 서열번호 2의 vp1 아미노산 서열을 인코딩하는 AAVhu68 핵산으로부터 캡시드를 발현하는 생성 시스템에서 생성되는 rAAVhu68 캡시드, 및 선택적으로, 예를 들어, vp1 및/또는 vp2-고유 영역이 없는 vp3 단백질을 인코딩하는, 추가적인 핵산 서열을 갖는다. 단일 핵산 서열 vp1을 사용하는 생성으로부터 생성된 rAAVhu68은 vp1 단백질, vp2 단백질 및 vp3 단백질의 이중 집단을 생성한다. 보다 구체적으로, AAVhu68 캡시드는 서열번호 2의 예상되는 아미노산 잔기로부터 변형을 갖는 vp1 단백질 내, vp2 단백질 내 및 vp3 단백질 내의 하위집단을 포함한다. 이들 하위집단은 최소한 탈아미드화된 아스파라긴(N 또는 Asn) 잔기를 포함한다. 예를 들어, 아스파라긴-글리신 쌍의 아스파라긴은 고도로 탈아미드화되어 있다.

[0077] 일 실시형태에서, AAVhu68 vp1 핵산 서열은 서열번호 1의 서열, 또는 이에 상보적인 가닥, 예를 들어, 상응하는 mRNA를 가진다. 특정 실시형태에서, vp2 및/또는 vp3 단백질은, 예를 들어, 선택된 발현 시스템에서 vp 단백질의 비율을 변경하기 위해, vp1과 상이한 핵산 서열로부터 추가적으로 또는 대안적으로 발현될 수 있다. 특정 실시형태에서, 또한 vp1-고유 영역(약 aa 1번 내지 약 aa 137번) 및/또는 vp2-고유 영역(약 aa 1번 내지 약 aa 202번)이 없는 서열번호 2의 AAVhu68 vp3 아미노산 서열(약 aa 203 내지 736번)을 인코딩하는 핵산 서열, 또는 이에 상보적인 가닥, 상응하는 mRNA(서열번호 2의 약 nt 607번 내지 약 nt 2211번)가 제공된다. 특정 실시형태에서, 또한 vp1-고유 영역(약 aa 1번 내지 약 137번)이 없는 서열번호 2의 AAVhu68 vp2 아미노산 서열(약 aa 138 내지 736번)을 인코딩하는 핵산 서열, 또는 이에 상보적인 가닥, 상응하는 mRNA(서열번호 1의 nt 412번 내지 2211번)가 제공된다.

[0078] 그러나, 서열번호 2의 아미노산 서열을 인코딩하는 다른 핵산 서열이 rAAVhu68 캡시드를 생성하는 데 사용하기 위해 선택될 수 있다. 특정 실시형태에서, 핵산 서열은 서열번호 1의 핵산 서열 또는 서열번호 2를 인코딩하는 서열번호 1과 적어도 70% 내지 99% 동일, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 적어도 99% 동일한 서열을 갖는다. 특정 실시형태에서, 핵산 서열은 서열번호 1의 핵산 서열 또는 서열번호 2의 vp2 캡시드 단백질(약 aa 138 내지 736번)을 인코딩하는 서열번호 1의 약 nt 412번 내지 약 nt 2211번과 적어도 70% 내지 99%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 또는 적어도 99% 동일한 서열을 갖는다. 특정 실시형태에서, 핵산 서열은 서열번호 1의 약 nt 607번 내지 약 nt 2211번

의 핵산 서열 또는 서열번호 1의 vp3 캡시드 단백질(약 aa 203번 내지 736번)을 인코딩하는 서열번호 1의 nt 412번 내지 약 nt 2211번과 적어도 70% 내지 99%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 또는 적어도 99% 동일한 서열을 갖는다.

[0079] DNA(게놈 또는 cDNA), 또는 RNA(예를 들어, mRNA)를 포함하여, 이러한 AAVhu68 캡시드를 인코딩하는 핵산 서열을 설계하는 것은 당업계의 기술 범위 내에 있다. 특정 실시형태에서, AAVhu68 vp1 캡시드 단백질을 인코딩하는 핵산 서열은 서열번호 2에서 제공된다. 특정 실시형태에서, AAVhu68 캡시드는 본 명세서에 기재된 바와 같은 변형(예를 들어, 탈아미드화된 아미노산)이 있는 서열번호 1의 핵산 서열 또는 서열번호 2의 vp1 아미노산 서열을 인코딩하는 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 또는 적어도 99% 서열을 사용하여 생성된다. 특정 실시형태에서, vp1 아미노산 서열은 서열번호 2로 재현된다.

[0080] 특정 실시형태에서, 캡시드 탈아미드화가 감소된 AAV 캡시드가 선택될 수 있다. 예를 들어, 둘 다 2019년 2월 27일에 출원되고, 전문이 참조에 의해 인용되는 PCT/US19/19804호 및 PCT/US18/19861호를 참조한다.

[0081] 본 명세서에서 사용되는 바와 같이 vp 캡시드 단백질을 지칭하는 데 사용되는 경우, 용어 "이중" 또는 이의 임의의 문법적 변형은, 예를 들어, 상이한 변형된 아미노산 서열이 있는 vp1, vp2 또는 vp3 단량체(단백질)를 갖는, 동일하지 않은 요소로 이루어진 집단을 지칭한다. 서열번호 2는 AAVhu68 vp1 단백질의 인코딩된 아미노산 서열을 제공한다. vp1, vp2 및 vp3 단백질(대안적으로 아이소형(isoform)으로 불림)과 관련하여 사용될 때 용어 "이중"은 캡시드 내 vp1, vp2 및 vp3 단백질의 아미노산 서열의 차이를 지칭한다. AAV 캡시드는 vp1 단백질 내, vp2 단백질 내 및 vp3 단백질 내에 예상되는 아미노산 잔기로부터 변형이 있는 하위집단을 포함한다. 이러한 하위집단은, 최소한 특정 탈아미드화된 아스파라긴(N 또는 Asn) 잔기를 포함한다. 예를 들어, 특정 하위집단은 아스파라긴-글리신 쌍에서 적어도 1, 2, 3 또는 4개의 고도로 탈아미드화된 아스파라긴(N) 위치를 포함하고 선택적으로 다른 탈아미드화된 아미노산을 더 포함하며, 여기서 탈아미드화는 아미노산 변화 및 다른 선택적인 변형을 초래한다.

[0082] 본 명세서에서 사용되는 바와 같이, vp 단백질의 "하위집단"은 달리 명시되지 않는 한, 공통적으로 적어도 하나의 정의된 특징을 가지고, 적어도 하나의 그룹 구성원 내지 참조 그룹의 모든 구성원보다 더 적은 수의 그룹 구성원으로 이루어진 vp 단백질의 그룹을 지칭한다. 예를 들어, vp1 단백질의 "하위집단"은 달리 명시되지 않는 한, 조립된 AAV 캡시드 내에 적어도 하나(1)의 vp1 단백질이 있고 모든 vp1 단백질보다 더 적은 수의 vp1 단백질이 있다. vp3 단백질의 "하위집단"은 달리 명시되지 않는 한, 조립된 AAV 캡시드 내에 하나(1)의 vp3 단백질 내지 모든 vp3 단백질보다 더 적은 수의 vp3 단백질이 있을 수 있다. 예를 들어, vp1 단백질은 vp 단백질의 하위집단일 수 있고; vp2 단백질은 vp 단백질의 별개의 하위집단일 수 있으며, vp3은 또한 조립된 AAV 캡시드 내의 vp 단백질의 추가의 하위집단이다. 또 다른 예에서, vp1, vp2 및 vp3 단백질은, 예를 들어, 아스파라긴-글리신 쌍에서 상이한 변형, 예를 들어, 적어도 1, 2, 3 또는 4개의 고도로 탈아미드화된 아스파라긴을 갖는 하위집단을 포함할 수 있다.

[0083] 달리 명시되지 않는 한, 고도로 탈아미드화된 것은 참조된 아미노산 위치에서 예상되는 아미노산 서열과 비교하여, 참조된 아미노산 위치에서 적어도 45% 탈아미드화된, 적어도 50% 탈아미드화된, 적어도 60% 탈아미드화된, 적어도 65% 탈아미드화된, 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 97%, 적어도 99%, 또는 최대 약 100%의 탈아미드화된 것을 지칭한다(예를 들어, 서열번호 2[AAVhu68]의 넘버링을 기준으로 57번 아미노산에서 아스파라긴의 적어도 80%가 총 vp1 단백질을 기준으로 탈아미드화될 수 있고 총 vp1, vp2 및 vp3 단백질을 기준으로 탈아미드화될 수 있음). 이와 같은 백분율은 2D-겔, 질량분석법 기법, 또는 다른 적합한 기법을 사용하여 결정될 수 있다.

[0084] 따라서, rAAV는, 최소한 적어도 하나의 고도로 탈아미드화된 아스파라긴을 포함하는 적어도 하나의 하위집단을 포함하는, 탈아미드화된 아미노산이 있는 vp1, vp2, 및/또는 vp3 단백질의 rAAV 캡시드 내의 하위집단을 포함한다. 추가적으로, 다른 변형은 특히 선택된 아스파르트산(D 또는 Asp) 잔기 위치에서 이성질체화를 포함할 수 있다. 또 다른 실시형태에서, 변형은 Asp 위치에서의 아미드화를 포함할 수 있다.

[0085] 특정 실시형태에서, AAV 캡시드는 적어도 4 내지 적어도 약 25개의 탈아미드화된 아미노산 잔기 위치를 갖는 vp1, vp2 및 vp3의 하위집단을 포함하며, 이의 적어도 1 내지 10%는 vp 단백질의 인코딩된 아미노산 서열과 비교하여 탈아미드화된다. 이들 대부분은 N 잔기일 수 있다. 그러나, Q 잔기가 또한 탈아미드화될 수 있다.

[0086] 특정 실시형태에서, rAAV는 실시예 1에서 제공되고 본 명세서에서 참조에 의해 인용되는 표에 제시된 위치에서 2, 3, 4개 또는 그 이상의 탈아미드화된 잔기의 조합을 포함하는 하위집단을 갖는 vp1, vp2 및 vp3 단백질을

갖는 AAV 캡시드를 가진다. rAAV에서의 탈아미드화는 2D 겔 전기영동법, 및/또는 질량 분석법, 및/또는 단백질 모델링 기법을 사용하여 결정될 수 있다. 온라인 크로마토그래피는 Acclaim PepMap 칼럼 및 NanoFlex 소스가 있는 Q Exactive HF(Thermo Fisher Scientific)에 커플링된 Thermo UltiMate 3000 RSLC 시스템(Thermo Fisher Scientific)을 이용하여 수행될 수 있다. MS 데이터는 조사 스캔(200 내지 2000 m/z)으로부터의 가장 풍부한 서열결정전(not-yet-sequenced) 전구체 이온을 동력학적으로 선택하는 Q Exactive HF에 대한 데이터 의존적인 상위 20 방법을 사용하여 획득된다. 서열결정은 예측된 자동 이득 제어로 결정된 $1e5$ 이온의 목표 값을 이용한 고에너지 충돌 분열 단편화를 통해서 수행되고, 전구체의 단리는 4 m/z의 창으로 수행되었다. 조사 스캔은 m/z 200에서 120,000의 분해능으로 획득되었다. HCD 스펙트럼에 대한 분해능은 50ms의 최대 이온 주입 시간 및 30의 정규화된 충돌 에너지를 이용하여 m/z 200에서 30,000으로 설정될 수 있다. S-렌즈 RF 수준은 50에서 설정되어, 소화로부터 펩타이드에 의해서 점유된 m/z 영역의 최적의 전달을 제공할 수 있다. 전구체 이온은 단편화 선택으로부터 단일, 미배정, 또는 6 및 더 높은 전하 상태로 제외될 수 있다. BioPharma Finder 1.0 소프트웨어(Thermo Fischer Scientific)가 획득된 데이터의 분석에 사용될 수 있다. 펩타이드 맵핑을 위해서, 고정된 변형으로서 설정된 카바미도메틸화; 및 가변 변형으로서 설정된 산화, 탈아미드화, 및 포스포릴화, 10-ppm 질량 정확도, 높은 프로테아제 특이성, 및 MS/MS 스펙트럼에 대한 0.8의 신뢰 수준을 갖는 단기식 단백질 FASTA 데이터베이스를 사용하여 검색을 수행한다. 적합한 프로테아제의 예는, 예를 들어, 트립신 또는 키모트립신을 포함할 수 있다. 탈아미드화된 펩타이드의 질량 분석법적 식별은 비교적 간단한데, 그 이유는 탈아미드화가 무손상 분자 +0.984 Da의 질량(-OH와 -NH₂ 기 간의 질량 차이)에 추가되기 때문이다. 특정 펩타이드의 탈아미드화 백분율은 탈아미드화된 펩타이드의 질량 면적을, 탈아미드화된 펩타이드와 천연 펩타이드의 면적의 합으로 나눔으로써 결정된다. 가능한 탈아미드화 부위의 수를 고려하여, 상이한 부위에서 탈아미드화된 동중 원소종(isobaric species)이 단일 피크에서 동시에 이동할 수 있다. 결론적으로, 다수의 잠재적인 탈아미드화 부위가 있는 펩타이드로부터 기원한 단편 이온을 사용하여 탈아미드화의 다수의 부위를 위치시키거나 구별할 수 있다. 이러한 경우에, 관찰된 동위원소 패턴 내의 상대 강도를 사용하여 상이한 탈아미드화된 펩타이드 이성질체의 상대 풍부도를 구체적으로 결정할 수 있다. 이러한 방법은, 모든 이성질체 중에 대한 단편화 효율성이 동일하고, 탈아미드화 부위에 독립적이라고 가정한다. 이러한 예시적인 방법에 대한 다수의 변경이 사용될 수 있음을 당업자는 이해할 것이다. 예를 들어, 적합한 질량 분석기는, 예를 들어, 사중극자 비행 시간 질량 분석기(quadrupole time of flight mass spectrometer; QTOF), 예컨대, Waters Xevo 또는 Agilent 6530 또는 오비트랩 기기(orbitrap instrument), 예컨대, Orbitrap Fusion 또는 Orbitrap Velos(Thermo Fisher)를 포함할 수 있다. 적합하게 액체 크로마토그래피 시스템은, 예를 들어, Waters 또는 Agilent 시스템으로부터의 Acquity UPLC 시스템(1100 또는 1200 시리즈)을 포함한다. 적합한 데이터 분석 소프트웨어는, 예를 들어, MassLynx(Waters), Pinpoint 및 Pepfinder(Thermo Fischer Scientific), Mascot(Matrix Science), Peaks DB(Bioinformatics Solutions)를 포함할 수 있다. 또 다른 기법은, 예를 들어, 문헌[X. Jin et al, Hu Gene Therapy Methods, Vol. 28, No. 5, pp. 255-267](2017년 6월 16일자로 온라인 공개됨)에 기재된 것일 수 있다.

[0087] 탈아미드화에 추가적으로, 다른 변형이 일어날 수 있는데, 이는 하나의 아미노산의 상이한 아미노산 잔기로의 전환을 초래하지 않는다. 이와 같은 변형은 아세틸화된 잔기, 이성질체화, 포스포릴화, 또는 산화를 포함할 수 있다.

[0088] 탈아미드화의 조절: 특정 실시형태에서, AAV는 아스파라긴-글리신 쌍에서 글리신을 변화시키도록 변형되어, 탈아미드화를 감소시킨다. 다른 실시형태에서, 아스파라긴은 상이한 아미노산, 예를 들어, 더 느린 속도로 탈아미드화시키는 글루타민으로; 또는 아미드기가 결여된 아미노산(예를 들어, 글루타민 및 아스파라긴은 아미드기를 포함함)으로; 그리고/또는 아민기가 결여된 아미노산(예를 들어, 라이신, 아르기닌 및 히스티딘은 아민기를 포함함)으로 변경된다. 본 명세서에서 사용되는 바와 같이, 아미드 또는 아민 측쇄가 결여된 아미노산은, 예를 들어, 글리신, 알라닌, 발린, 류신, 아이소류신, 세린, 트레오닌, 시스틴, 페닐알라닌, 타이로신, 또는 트립토판, 및/또는 프롤린을 지칭한다. 예컨대, 기재된 변형은 인코딩된 AAV 아미노산 서열에서 발견되는 아스파라긴-글리신 쌍 중 1, 2 또는 3개에 존재할 수 있다. 특정 실시형태에서, 이와 같은 변형은 아스파라긴-글리신 쌍 중 4개 모두에서 일어나지는 않는다. 따라서, AAV 및/또는 조작된 AAV 변이체의 탈아미드화를 감소시키는 방법은 더 낮은 탈아미드화 속도를 가진다. 추가적으로, 또는 대안적으로 하나 이상의 다른 아미드 아미노산은 비-아미드 아미노산으로 변화되어 AAV의 탈아미드화를 감소시킬 수 있다. 특정 실시형태에서, 본 명세서에 기재된 바와 같은 돌연변이체 AAV 캡시드는 아르기닌-글리신 쌍에서 돌연변이를 포함하여, 글리신은 알라닌 또는 세린으로 변화된다. 돌연변이체 AAV 캡시드는 1, 2 또는 3개의 돌연변이체를 포함할 수 있고, 여기서 참조 AAV는 본래 4개의 NG 쌍을 포함한다. 특정 실시형태에서, AAV 캡시드는 1, 2, 3 또는 4개의 이와 같은 돌연변이체를 포함할 수 있고, 여기서 참조 AAV는 본래 5개의 NG 쌍을 포함한다. 특정 실시형태에서, 돌연

변이체 AAV 캡시드는 NG 쌍 내에 단일 돌연변이만을 포함한다. 특정 실시형태에서, 돌연변이체 AAV 캡시드는 2개의 상이한 NG 쌍 내에 돌연변이를 포함한다. 특정 실시형태에서, 돌연변이체 AAV 캡시드는 AAV 캡시드에서 구조적으로 분리된 위치에 위치한 2개의 상이한 NG 쌍이다. 특정 실시형태에서, 돌연변이는 VP1-고유 영역에 존재하지 않는다. 특정 실시형태에서, 돌연변이 중 하나는 VP1-고유 영역에 존재한다. 선택적으로, 돌연변이체 AAV 캡시드는 NG 쌍 내에 어떠한 변형도 포함하지 않지만, NG 쌍의 외부에 위치한 하나 이상의 아스파라긴 또는 글루타민에서 탈아미드화를 최소화 또는 제거하기 위해서 돌연변이를 포함한다. AAVhu68 캡시드 단백질에서, 4개의 잔기(N57, N329, N452, N512)는 일상적으로 70% 초과 탈아미드화 수준을 나타내고 대부분의 경우 다양한 로트에 걸쳐 90% 초과 수준을 나타낸다. 추가적인 아스파라긴 잔기(N94, N253, N270, N304, N409, N477 및 Q599)는 또한 다양한 로트에 걸쳐 최대 약 20%의 탈아미드화 수준을 나타낸다. 탈아미드화 수준은 초기에 트립신 소화를 사용하여 확인하였고, 키모트립신 소화를 이용하여 검증하였다.

[0089] AAVhu68 캡시드는 서열번호 2의 예상되는 아미노산 잔기로부터 변형을 갖는 vp1 단백질 내, vp 단백질 내 및 vp3 단백질 내의 하위집단을 포함한다. 이들 하위집단은 최소한 특정 탈아미드화된 아스파라긴(N 또는 Asn) 잔기를 포함한다. 예를 들어, 특정 하위집단은 서열번호 2의 아스파라긴-글리신 쌍에서 적어도 1, 2, 3 또는 4개의 고도로 탈아미드화된 아스파라긴(N) 위치를 포함하고 선택적으로 다른 탈아미드화된 아미노산을 더 포함하며, 여기서 탈아미드화는 아미노산 변화 및 다른 선택적인 변형을 초래한다. 이들 및 다른 변형의 다양한 조합이 본 명세서에 기재되어 있다.

[0090] 특정 실시형태에서, 본 명세서에 기재된 바와 같은 rAAV는 자가-상보성 AAV이다. "자가-상보성 AAV"는 재조합 AAV 핵산 서열에 의해 운반되는 코딩 영역이 분자내 이중-가닥 DNA 주형을 형성하도록 설계된 작제물을 지칭한다. 감염시, 두번째 가닥의 세포 매개 합성을 기다리기 보다는, scAAV의 2개의 상보적인 절반이 회합하여 즉각적인 복제 및 전사 준비가 된 하나의 이중 가닥 DNA(dsDNA) 단위를 형성할 것이다. 예를 들어, 문헌[D M McCarty et al, "Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis", Gene Therapy, (August 2001), Vol 8, Number 16, Pages 1248-1254]을 참조한다. 자가-상보성 AAV는, 예를 들어, 미국 특허 제6,596,535호; 제7,125,717호; 및 제7,456,683호에 기재되어 있으며, 이들 각각은 본 명세서에 전문이 참조에 의해 인용된다.

[0091] 본 명세서에 기재된 rAAV의 구성은 본 명세서 전반에 걸쳐 기재된 다른 구성, 요법, 양태, 실시형태, 및 방법에 적용되는 것으로 의도됨이 이해되어야 한다.

[0092] 약제학적 조성물

[0093] hGAA780I 단백질 또는 융합 단백질 또는 hGAA780I 단백질 또는 융합 단백질 이식유전자를 포함하는 발현 카세트를 포함하는 약제학적 조성물은 액체 현탁액, 동결건조 또는 동결된 조성물, 또는 또 다른 적합한 제형일 수 있다. 특정 실시형태에서, 조성물은 hGAA780I 단백질 또는 융합 단백질 또는 발현 카세트 및 현탁액을 형성하는 생리학적으로 양립 가능한 액체(예를 들어, 용액, 희석제, 담체)를 포함한다. 이와 같은 액체는 바람직하게는 수성 기반이고, 완충제(들), 계면활성제(들), pH 조정제(들), 보존제(들), 또는 다른 적합한 부형제 중 하나 이상을 포함할 수 있다. 적합한 구성요소는 하기에 보다 상세하게 논의되어 있다. 약제학적 조성물은 수성 현탁액 및 임의의 선택된 부형제, 및 hGAA780I 단백질 또는 융합 단백질 또는 발현 카세트를 포함한다.

[0094] 특정 실시형태에서, 약제학적 조성물은 이식유전자 및 비-바이러스 전달 시스템을 포함하는 발현카세트를 포함한다. 이는, 예를 들어, 네이키드 DNA, 네이키드 RNA, 무기 입자, 지질 또는 지질-유사 입자, 키토산-기반 제형 및 당업계에 알려지고, 예를 들어, Ramamoorth 및 Narvekar(상기 인용된 바와 같음)에 의해 설명된 기타 다른 것을 포함할 수 있다. 다른 실시형태에서, 약제학적 조성물은 바이러스 벡터 시스템에서 조작된 이식유전자를 포함하는 발현 카세트를 포함하는 현탁액이다. 특정 실시형태에서, 약제학적 조성물은 비-복제 바이러스 벡터를 포함한다. 적합한 바이러스 벡터는 예를 들어, 재조합 아데노바이러스, 재조합 렌티바이러스, 재조합 보카바이러스, 재조합 아데노-연관 바이러스(AAV), 또는 또 다른 재조합 파코바이러스와 같은 임의의 적합한 전달 벡터를 포함할 수 있다. 특정 실시형태에서, 바이러스 벡터는 이를 필요로 하는 환자에게 유전자 산물을 전달하기 위한 재조합 AAV이다.

[0095] 일 실시형태에서, 약제학적 조성물은 hGAA780I 단백질 또는 융합 단백질 또는 hGAA780I 단백질 또는 융합 단백질에 대한 코딩 서열을 포함하는 발현 카세트 및 뇌실내(ICV), 척추강내(IT), 수조내, 또는 정맥내(IV) 주사를 통한 전달에 적합한 제형 완충제를 포함한다. 일 실시형태에서, 발현 카세트는 벡터 게놈이 패키징된 재조합 바이러스 벡터(즉, rAAV.hGAA780I)의 일부이다.

- [0096] 일 실시형태에서, 약제학적 조성물은 효소 대체 요법(ERT)로서 대상체에게 전달하기 위한 hGAA780I 단백질, 또는 이의 기능성 단편을 포함한다. 이와 같은 약제학적 조성물은 보통 정맥내로 투여되지만, 일부 상황에서는 피내, 근육내 또는 경구 투여가 또한 가능하다. 조성물은 폼페병을 앓고 있거나 이에 걸릴 위험이 있는 개체의 예방적 치료를 위해 투여될 수 있다. 치료적 적용을 위해, 약제학적 조성물은 확립된 질환을 앓고 있는 환자에게 축적된 대사산물의 농도를 감소시키고/시키거나 대사산물의 추가 축적을 방지하거나 저지하기에 충분한 양으로 투여된다. 리소좀 효소 결핍 질환의 위험이 있는 개체의 경우, 약제학적 조성물은 대사산물의 축적을 방지하거나 저지하기에 충분한 양으로 예방적으로 투여된다. 본 명세서에 기재된 변형된 GAA 조성물은 치료학적 유효량으로 투여된다. 일반적으로, 치료학적 유효량은 대상체의 의학적 상태의 중증도뿐만 아니라, 대상체의 연령, 전반적인 상태, 및 성별에 따라 달라질 수 있다. 투약량은 의사에 의해 결정될 수 있으며, 관찰된 치료의 효과에 맞게 필요에 따라 조정될 수 있다. 일 양태에서, hGAA780I 단백질 또는 융합 단백질, 또는 이의 기능성 단편의 단위 투약량을 포함하도록 제형화된 ERT용 약제학적 조성물이 본 명세서에서 제공된다.
- [0097] 일 실시형태에서, 조성물은 대상체에게 전달하기에 적합한 최종 제형을 포함하고, 예를 들어, 생리화적으로 양립 가능한 pH 및 염 농도로 완충된 수성 액체 현탁액이다. 선택적으로, 하나 이상의 계면활성제가 제형에 존재한다. 또 다른 실시형태에서, 조성물은 대상체에 투여하기 위해 희석되는 농축물로서 수송될 수 있다. 다른 실시형태에서, 조성물은 동결건조되고 투여 시에 재구성될 수 있다.
- [0098] 일 실시형태에서, 본 명세서에서 제공된 바와 같은 조성물은 수성 현탁액 액체에 용해된 계면활성제, 보존제, 부형제, 및/또는 완충제를 포함한다. 일 실시형태에서, 완충제는 PBS이다. 또 다른 실시형태에서, 완충제는 인공 뇌척수액(aCSF), 예를 들어, 엘리엇 제형 완충제(Elliott's formulation buffer); 또는 하바드 장치 관류액(Harvard apparatus perfusion fluid)(다음과 같은 최종 이온 농도(단위: mM)를 갖는 인공 CSF: Na 150; K 3.0; Ca 1.4; Mg 0.8; P 1.0; Cl 155). 완충 식염수, 계면활성제, 및 약 100mM 염화나트륨(NaCl) 내지 약 250mM 염화나트륨과 동등한 이온 강도로 조정된 생리화적으로 양립 가능한 염 또는 염의 혼합물, 또는 동등한 이온 농도로 조정된 생리화적으로 양립 가능한 염을 포함하는 다양한 적합한 용액이 알려져 있다.
- [0099] 적합하게는, 제형은 생리화적으로 허용 가능한 pH, 예를 들어, pH 6 내지 8, 또는 pH 6.5 내지 7.5, pH 7.0 내지 7.7, 또는 pH 7.2 내지 7.8의 범위로 조정된다. 뇌척수액의 pH가 약 7.28 내지 약 7.32이므로, 척추강내 전달을 위해 이 범위 내의 pH가 요구될 수 있으며; 반면 정맥내 전달의 경우, 6.8 내지 약 7.2의 pH가 요구될 수 있다. 그러나, 가장 넓은 범위 내의 다른 pH 및 이러한 하위 범위가 다른 전달 경로에 대해 선택될 수 있다.
- [0100] 적합한 계면활성제, 또는 계면활성제의 조합이 무독성인 비-이온성 계면활성제 중에서 선택될 수 있다. 일 실시형태에서, 예를 들어, 중성 pH를 가지고, 평균 분자량이 8400인 폴록사머(Poloxamer) 188로도 알려진 Pluronic® F68[BASF]과 같은 1차 하이드록실기로 종결되는 이작용성 블록 공중합체 계면활성제가 선택된다. 다른 계면활성제 및 다른 폴록사머, 즉, 폴리옥시에틸렌(폴리(에틸렌 옥사이드))의 2개 친수성 사슬이 측접한 폴리옥시프로필렌(폴리(프로필렌 옥사이드))의 중심 소수성 사슬로 구성된 비이온성 삼중블록 공중합체, Solutol HS 15(Macrogon-15 하이드록시스테아레이트), Labrasol(폴리옥시 카프릴산 글리세라이드), 폴리옥시 10 올레일 에터, Tween(폴리옥시에틸렌 솔비탄 지방산 에스터), 에탄올 및 폴리에틸렌 글라이콜이 선택될 수 있다. 일 실시형태에서, 제형은 폴록사머를 포함한다. 이러한 공중합체는 일반적으로 문자 "P"(폴록사머의 경우) 뒤에 3자리 숫자가 있는 명칭으로 명명되며: 처음 2자리 숫자×100은 폴리옥시프로필렌 코어의 대략적인 분자량을 제공하고, 마지막 숫자×10은 폴리옥시에틸렌 함량의 백분율을 제공한다. 일 실시형태에서, 폴록사머 188이 선택된다. 계면활성제는 현탁액의 최대 약 0.0005% 내지 약 0.001%의 양으로 존재할 수 있다.
- [0101] 일 예에서, 제형은, 예를 들어, 염화나트륨, 중탄산나트륨, 텍스트로스, 황산마그네슘(예를 들어, 황산마그네슘·7H₂O), 염화칼륨, 염화칼슘(예를 들어, 염화칼슘·2H₂O), 이염기성 인산나트륨, 및 이들의 혼합물 중 하나 이상을 물에 포함하는 완충된 식염수 용액을 포함할 수 있다. 적합하게는, 척추강내 전달의 경우, 삼투압농도는 뇌척수액과 양립 가능한 범위(예를 들어, 약 275 내지 약 290) 내에 있으며; 예를 들어, emedicine.medscape.com/article/2093316-overview를 참조한다. 선택적으로, 척추강내 전달의 경우, 상업적으로 입수 가능한 희석제가 현탁화제로서, 또는 또 다른 현탁화제 및 다른 선택적인 부형제와 함께 사용될 수 있다. 예를 들어, Elliotts B® 용액[Lukare Medical]을 참조한다.
- [0102] 다른 실시형태에서, 제형은 하나 이상의 투과 증진제를 포함할 수 있다. 적합한 투과 증진제의 예는, 예를 들어, 만니톨, 소듐 글리콜레이트, 소듐 타우로콜레이트, 소듐 데옥시콜레이트, 소듐 살리실레이트, 소듐 카프릴레이트, 소듐 카프레이트, 소듐 라우릴 설페이트, 폴리옥시에틸렌-9-라우릴 에터, 또는 EDTA를 포함할 수 있다.

- [0103] 약제학적으로 허용 가능한 담체 및 본 명세서에 기재된 바와 같은 핵산 서열을 포함하는 벡터를 포함하는 약제학적 조성물이 추가적으로 제공된다. 본 명세서에서 사용되는 바와 같이, "담체"는 임의의 모든 용매, 분산 매질, 비히클, 코팅제, 희석제, 항박테리아 및 항진균제, 등장화제 및 흡수 지연제, 완충제, 담체 용액, 현탁액, 콜로이드 등을 포함한다. 약제학적 활성 물질을 위한 이와 같은 매질 및 작용제의 사용은 당업계에 잘 알려져 있다. 보충 활성 성분이 또한 조성물에 포함될 수 있다. 전달 비히클, 예컨대, 리포솜, 나노캡슐, 마이크로입자, 마이크로스피어, 지질 입자, 소포 등이 본 명세서에 기재된 조성물을 적합한 숙주 세포 내로 도입하는 데 사용될 수 있다. 특히, rAAV 벡터는 지질 입자, 리포솜, 비히클, 나노구체, 또는 나노입자 등에 캡슐화되는 전달을 위해 제형화될 수 있다. 일 실시형태에서, 치료학적 유효량의 벡터가 약제학적 조성물에 포함된다. 담체의 선택은 제한이 없다. 다른 통상적인 약제학적으로 허용 가능한 담체, 예컨대, 보존제, 또는 화학적 안정화제. 적합한 예시적인 보존제는 클로로부탄올, 솔빈산칼륨, 솔빈산, 이산화황, 프로필 갈레이트, 파라벤, 에틸 바닐린, 글리세린, 페놀, 및 파라클로로페놀을 포함한다. 적합한 화학적 안정화제는 젤라틴 및 알부민을 포함한다.
- [0104] 어구 "약제학적으로 허용 가능한"은 숙주에게 투여될 때 알레르기 반응 또는 유사한 부반응을 일으키지 않는 분자 실체 및 조성물을 지칭한다.
- [0105] 본 명세서에서 사용되는 바와 같이, 용어 "투약량" 또는 "양"은 치료 과정에서 대상체에게 전달되는 총 투약량 또는 양, 또는 단일 단위(또는 다중 단위 또는 분할 투약량) 투여로 전달되는 투약량 또는 양을 지칭할 수 있다.
- [0106] 본 명세서에 기재된 수성 현탁액 또는 약제학적 조성물은 임의의 적합한 경로 또는 상이한 경로의 조합에 의해 이를 필요로 하는 대상체에게 전달하도록 설계된다. 일 실시형태에서, 약제학적 조성물은 뇌실내(ICV), 척추강내(IT), 또는 수조내 주사를 통해 전달하도록 제형화된다. 일 실시형태에서, 본 명세서에 기재된 조성물은 정맥내 주사에 의해 이를 필요로 하는 대상체에게 전달하도록 설계된다. 대안적으로, 다른 투여 경로(예를 들어, 경구, 흡입, 비강내, 기관내, 동맥내, 안내, 근육내, 및 다른 비경구 경로)가 선택될 수 있다.
- [0107] 본 명세서에서 사용되는 바와 같이, 용어 "척추강내 전달" 또는 "척추강내 투여"는 약물이 뇌척수액(CSF)에 도달하도록 척추관으로, 보다 구체적으로는 지주막하 공간으로 주사를 통한 약물에 대한 투여 경로를 지칭한다. 척추강내 전달은 요추 천자, 뇌실내, 후두하/수조내, 및/또는 C1-2 천자를 포함할 수 있다. 예를 들어, 물질은 요추 천자에 의해 지주막하 공간 전체에 확산을 위해 도입될 수 있다. 또 다른 예에서, 주사는 대조 내로 이루어질 수 있다. 수조내 전달은 벡터 확산을 증가시키고/시키거나 투여에 의해 유발되는 독성 및 염증을 감소시킬 수 있다. 예를 들어, 문헌[Christian Hinderer et al, Widespread gene transfer in the central nervous system of cynomolgus macaques following delivery of AAV9 into the cisterna magna, Mol Ther Methods Clin Dev. 2014; 1: 14051. Published online 2014 Dec 10. doi: 10.1038/mtm.2014.51]을 참조한다.
- [0108] 본 명세서에서 사용되는 바와 같이, 용어 "수조내 전달" 또는 "수조내 투여"는 뇌실의 뇌척수액으로 직접 또는 소뇌연수조 내에, 보다 구체적으로 후두하 천자를 통한 또는 대조 내로 직접 주사에 의한 또는 영구적으로 위치한 튜브를 통한 약물에 대한 투여 경로를 지칭한다.
- [0109] 일 양태에서, 제형 완충제 중에 본 명세서에 기재된 바와 같은 벡터를 포함하는 약제학적 조성물이 본 명세서에서 제공된다. 특정 실시형태에서, 복제-결합 바이러스 조성물은 인간 환자의 경우 (체중이 70 kg인 평균 대상체를 치료하기 위해) 약 1.0×10^9 개 GC 내지 약 1.0×10^{16} 개 GC의 범위(범위 내의 모든 정수 또는 분수 양을 포함함), 바람직하게는 1.0×10^{12} 개 GC 내지 1.0×10^{14} 개 GC인 복제-결합 바이러스의 양을 포함하도록 투약 단위로 제형화될 수 있다. 일 실시형태에서, 조성물은 용량당 적어도 1×10^9 , 2×10^9 , 3×10^9 , 4×10^9 , 5×10^9 , 6×10^9 , 7×10^9 , 8×10^9 , 또는 9×10^9 개 GC(범위 내의 모든 정수 또는 분수 양을 포함함)를 포함하도록 제형화된다. 또 다른 실시형태에서, 조성물은 용량당 적어도 1×10^{10} , 2×10^{10} , 3×10^{10} , 4×10^{10} , 5×10^{10} , 6×10^{10} , 7×10^{10} , 8×10^{10} , 또는 9×10^{10} 개 GC(범위 내의 모든 정수 또는 분수 양을 포함함)를 포함하도록 제형화된다. 또 다른 실시형태에서, 조성물은 용량당 적어도 1×10^{11} , 2×10^{11} , 3×10^{11} , 4×10^{11} , 5×10^{11} , 6×10^{11} , 7×10^{11} , 8×10^{11} , 또는 9×10^{11} 개 GC(범위 내의 모든 정수 또는 분수 양을 포함함)를 포함하도록 제형화된다. 또 다른 실시형태에서, 조성물은 용량당 적어도 1×10^{12} , 2×10^{12} , 3×10^{12} , 4×10^{12} , 5×10^{12} , 6×10^{12} , 7×10^{12} , 8×10^{12} , 또는 9×10^{12} 개 GC(범위 내의 모든 정수 또는 분수 양을 포함함)를 포함하도록 제형화된다. 또 다른 실시형태에서, 조성물은 용

량당 적어도 1×10^{13} , 2×10^{13} , 3×10^{13} , 4×10^{13} , 5×10^{13} , 6×10^{13} , 7×10^{13} , 8×10^{13} , 또는 9×10^{13} 개 GC(범위 내의 모든 정수 또는 분수 양을 포함함)를 포함하도록 제형화된다. 또 다른 실시형태에서, 조성물은 용량당 적어도 1×10^{14} , 2×10^{14} , 3×10^{14} , 4×10^{14} , 5×10^{14} , 6×10^{14} , 7×10^{14} , 8×10^{14} , 또는 9×10^{14} 개 GC(범위 내의 모든 정수 또는 분수 양을 포함함)를 포함하도록 제형화된다. 또 다른 실시형태에서, 조성물은 용량당 적어도 1×10^{15} , 2×10^{15} , 3×10^{15} , 4×10^{15} , 5×10^{15} , 6×10^{15} , 7×10^{15} , 8×10^{15} , 또는 9×10^{15} 개 GC(범위 내의 모든 정수 또는 분수 양을 포함함)를 포함하도록 제형화된다. 일 실시형태에서, 인간 적용의 경우 용량은 용량당 1×10^{10} 내지 약 1×10^{12} 개 GC(범위 내의 모든 정수 또는 분수 양을 포함함)의 범위일 수 있다.

[0110] 일 실시형태에서, 제형 완충제 중에 본 명세서에 기재된 바와 같은 rAAV를 포함하는 약제학적 조성물이 제공된다. 일 실시형태에서, rAAV는 약 1×10^9 개 게놈 복제물(GC)/ml 내지 약 1×10^{14} 개 GC/ml로 제형화된다. 추가 실시형태에서, rAAV는 약 3×10^9 개 GC/ml 내지 약 3×10^{13} 개 GC/ml로 제형화된다. 또한 추가의 실시형태에서, rAAV는 약 1×10^9 개 GC/ml 내지 약 1×10^{13} 개 GC/ml로 제형화된다. 일 실시형태에서, rAAV는 적어도 약 1×10^{11} 개 GC/ml로 제형화된다.

[0111] 일 실시형태에서, 본 명세서에 기재된 바와 같은 rAAV를 포함하는 약제학적 조성물은 뇌 질량 그램당 약 1×10^9 개 GC 내지 뇌 질량 그램당 약 1×10^{14} 개 GC의 용량으로 투여 가능하다.

[0112] 본 명세서에 기재된 약제학적 조성물의 조성물은 본 명세서 전반에 걸쳐 기재된 다른 조성물, 요법, 양태, 실시형태, 및 방법에 적용되는 것으로 의도됨이 이해되어야 한다.

[0113] 치료 방법

[0114] 본 명세서에 기재된 바와 같은 발현 카세트, rAAV, 및/또는 hGAA780I 단백질 또는 융합 단백질을, 선택적으로 면역조절제와 조합하여 포함하는, 폼페병을 가진 환자를 치료하기 위한 치료 요법이 제공된다. 특정 실시형태에서, 환자는 후기 발병형 폼페병을 가진다. 다른 실시형태에서, 환자는 아동기 발병형 폼페병을 가진다. 특정 실시형태에서, 공동-치료제는 발현 카세트, rAAV, 또는 hGAA780I 단백질 또는 융합 단백질, 예컨대, 면역조절 요법과 함께 전달된다. 추가적으로, 또는 대안적으로, 공동-요법은 기관지확장제, 아세틸콜린에스터라제 저해제, 호흡근 강화 훈련(RMST), 효소 대체 요법, 및/또는 횡격막 페이스팅 요법 중 하나 이상을 포함할 수 있다. 특정 실시형태에서, 환자는 rAAV의 단일 투여를 받는다. 특정 실시형태에서, 환자는 본 명세서에 기재된 바와 같은 발현 카세트 및/또는 rAAV를 포함하는 조성물의 단일 투여를 받는다. 특정 실시형태에서, 유효량의 발현 카세트를 포함하는 조성물의 이러한 단일 투여는 적어도 하나의 공동-치료제를 포함한다. 특정 실시형태에서, 환자는 2가지의 상이한 경로를 통해 실질적으로 동시에 발현 카세트, rAAV, 및/또는 hGAA780I 단백질 또는 융합 단백질 또는 본 명세서에 기재된 바와 같은 것을 투여받는다. 특정 실시형태에서, 2가지 상이한 주사 경로는 정맥내 및 척추강내 투여이다. 일 실시형태에서, 조성물은 뇌실내, 척추강내, 수조내, 또는 정맥내로 대상체에게 전달되는 현탁액이다. 특정 실시형태에서, 알파-글루코시다제의 결핍이 있는 환자는 심장, 호흡기, 및/또는 골격근 기능 중 하나 이상을 개선시키기 위해 본 명세서에서 제공되는 바와 같은 조성물을 투여받는다. 특정 실시형태에서, 치료의 결과로서 심장, CNS(뇌), 및/또는 골격근 중 하나 이상에서 글리코겐 저장 및/또는 자가포식 축적 감소가 있다.

[0115] 특정 실시형태에서, 발현 카세트, rAAV, 바이러스 또는 비-바이러스 벡터는 약제를 제조하는 데 사용된다. 특정 실시형태에서, 폼페병을 치료하기 위한 조성물의 용도가 제공된다.

[0116] 이들 조성물은, 예를 들어, 면역요법, 효소 대체 요법(예를 들어, Genzyme(Sanofi Corporation)에 의해 시판되는 Lumizyme, 미국 외에서는 Myozyme으로 시판됨)을 포함하는 다른 요법과 조합하여 사용될 수 있다. 폼페병의 추가적인 치료는 대중적이고 지지적이다. 예를 들어, 호흡기 지원이 필요할 수 있고; 물리적 요법은 호흡근을 강화시키는 데 도움이 될 수 있으며; 일부 환자는 밤 및/또는 낮 시간 동안 기계적 인공호흡(즉, 바이팍(bipap) 또는 용적 인공호흡기)을 통한 호흡 보조를 필요로 할 수 있다. 추가적으로, 기도 감염 동안 추가적인 지원이 필요할 수 있다. 교정기를 포함하여 정형외과적 장치가 일부 환자에게 권장될 수 있다. 구축 또는 척추 변형과 같은 특정 정형외과적 증상에 대해 수술이 필요할 수 있다. 일부 영아는 코를 통해 식도를 거쳐 위로 이어지는 급식관(비위관)의 삽입을 필요로 할 수 있다. 일부 아동에 있어서, 급식관은 복벽의 작은 수술 구멍을 통해 위로 직접 삽입될 필요가 있을 수 있다. 후기 발병형 폼페병을 가진 일부 개체는 연식을 필요로 할 수 있지만 급

식관을 필요로 하는 개체는 거의 없다.

- [0117] 본 명세서에 기재된 바와 같이, 용어 "증가하다"(예를 들어, 조직, 혈액 등에서 측정된 바와 같이 hGAA780I로 치료한 후 hGAA 수준 증가) 또는 "감소하다", "저하하다", "개선하다", "향상하다", "지연시키다", 또는 이의 임의의 문법적 변형, 또는 변화를 나타내는 임의의 유사한 용어는 달리 명시되지 않는 한, 상응하는 참조(예를 들어, 비치료 대조군 또는 폼페병이 없는 정상 대조군의 대상체)와 비교하여 약 5배, 약 2배, 약 1배, 약 90%, 약 80%, 약 70%, 약 60%, 약 50%, 약 40%, 약 30%, 약 20%, 약 10%, 또는 약 5%의 변동을 의미한다.
- [0118] 본 명세서에서 호환 가능하게 사용되는 바와 같이 "환자" 또는 "대상체"는 인간, 수의 또는 농장 동물, 가축 또는 애완동물, 임상 연구에 일반적으로 사용되는 동물을 포함하여 수컷 또는 암컷 포유류 동물을 의미한다. 일 실시형태에서, 이러한 방법 및 조성물의 대상체는 인간 환자이다. 일 실시형태에서, 이러한 방법 및 조성물의 대상체는 남성 또는 여성 인간이다.
- [0119] 일 실시형태에서, 현탁액은 pH가 7.28 내지 약 7.32이다.
- [0120] 이러한 용량 및 농도의 전달에 적합한 부피는 당업자에 의해 결정될 수 있다. 예를 들어, 약 $1\mu\text{l}$ 내지 150ml의 부피가 선택될 수 있으며, 성인의 경우 더 많은 부피가 선택된다. 전형적으로, 신생아의 경우, 적합한 부피는 약 0.5ml 내지 약 10ml이고, 더 성장한 영아의 경우, 약 0.5ml 내지 약 15ml가 선택될 수 있다. 토끼의 경우, 약 0.5ml 내지 약 20ml의 부피가 선택될 수 있다. 아동의 경우, 최대 약 30ml의 부피가 선택될 수 있다. 심대 초반 및 심대의 경우, 최대 약 50ml의 부피가 선택될 수 있다. 또 다른 실시형태에서, 환자는 선택된 약 5ml 내지 약 15ml, 또는 약 7.5ml 내지 약 10ml의 부피로 척추강내 투여를 받을 수 있다. 다른 적합한 부피 및 투약량이 결정될 수 있다. 투약량은 임의의 부작용에 대한 치료적 이익의 균형을 맞추기 위해 조정될 것이며, 이와 같은 투약량은 재조합 벡터가 이용되는 치료적 적용에 따라 달라질 수 있다.
- [0121] 일 실시형태에서, 본 명세서에 기재된 바와 같은 rAAV를 포함하는 조성물은 뇌 질량 그램당 약 1×10^9 개 GC 내지 뇌 질량 그램당 약 1×10^{14} 개 GC의 용량으로 투여 가능하다. 특정 실시형태에서, rAAV는 체중 kg당 약 1×10^9 개 GC 내지 체중 kg당 약 1×10^{13} 개 GC의 용량으로 전신으로 공동-투여된다.
- [0122] 일 실시형태에서, 대상체는 치료학적 유효량의 본 명세서에 기재된 발현 카세트, rAAV 또는 hGAA780I 효소(즉, 단백질 또는 융합 단백질)를 전달받는다. 본 명세서에서 사용되는 바와 같이, "치료학적 유효량"은 발현 카세트, rAAV, 또는 hGAA780I, 또는 이들의 조합물의 양을 지칭한다. 따라서, 특정 실시형태에서, 방법은 본 명세서에서 제공되는 hGAA780I 효소를 포함하는 조성물을 투여하는 것과 조합하여 hGAA780I-인코딩 핵산 서열의 전달을 위한 rAAV 또는 발현 작제물을 대상체에게 투여하는 단계를 포함한다.
- [0123] 일 실시형태에서, 발현 카세트는 벡터 게놈으로 뇌 질량 그램(g)당 약 1×10^9 개 게놈 복제물(GC) 내지 뇌 질량 그램당 약 1×10^{13} 개 GC(범위 내의 모든 정수 또는 분수 양 및 종점을 포함함)의 양으로 전달된다. 또 다른 실시형태에서, 투약량은 뇌 질량 그램당 1×10^{10} 개 GC 내지 뇌 질량 그램당 약 1×10^{13} 개 GC이다. 구체적인 실시형태에서, 환자에게 투여되는 벡터의 용량은 적어도 약 1.0×10^9 개 GC/g, 약 1.5×10^9 개 GC/g, 약 2.0×10^9 개 GC/g, 약 2.5×10^9 개 GC/g, 약 3.0×10^9 개 GC/g, 약 3.5×10^9 개 GC/g, 약 4.0×10^9 개 GC/g, 약 4.5×10^9 개 GC/g, 약 5.0×10^9 개 GC/g, 약 5.5×10^9 개 GC/g, 약 6.0×10^9 개 GC/g, 약 6.5×10^9 개 GC/g, 약 7.0×10^9 개 GC/g, 약 7.5×10^9 개 GC/g, 약 8.0×10^9 개 GC/g, 약 8.5×10^9 개 GC/g, 약 9.0×10^9 개 GC/g, 약 9.5×10^9 개 GC/g, 약 1.0×10^{10} 개 GC/g, 약 1.5×10^{10} 개 GC/g, 약 2.0×10^{10} 개 GC/g, 약 2.5×10^{10} 개 GC/g, 약 3.0×10^{10} 개 GC/g, 약 3.5×10^{10} 개 GC/g, 약 4.0×10^{10} 개 GC/g, 약 4.5×10^{10} 개 GC/g, 약 5.0×10^{10} 개 GC/g, 약 5.5×10^{10} 개 GC/g, 약 6.0×10^{10} 개 GC/g, 약 6.5×10^{10} 개 GC/g, 약 7.0×10^{10} 개 GC/g, 약 7.5×10^{10} 개 GC/g, 약 8.0×10^{10} 개 GC/g, 약 8.5×10^{10} 개 GC/g, 약 9.0×10^{10} 개 GC/g, 약 9.5×10^{10} 개 GC/g, 약 1.0×10^{11} 개 GC/g, 약 1.5×10^{11} 개 GC/g, 약 2.0×10^{11} 개 GC/g, 약 2.5×10^{11} 개 GC/g, 약 3.0×10^{11} 개 GC/g, 약 3.5×10^{11} 개 GC/g, 약 4.0×10^{11} 개 GC/g, 약 4.5×10^{11} 개 GC/g, 약 5.0×10^{11} 개 GC/g, 약 5.5×10^{11} 개 GC/g, 약 6.0×10^{11} 개 GC/g, 약 6.5×10^{11} 개 GC/g, 약 7.0×10^{11} 개 GC/g, 약 7.5×10^{11} 개 GC/g, 약 8.0×10^{11} 개 GC/g, 약 8.5×10^{11} 개 GC/g, 약 9.0×10^{11} 개 GC/g, 약 9.5×10^{11} 개 GC/g.

개 GC/g, 약 1.0×10^{12} 개 GC/g, 약 1.5×10^{12} 개 GC/g, 약 2.0×10^{12} 개 GC/g, 약 2.5×10^{12} 개 GC/g, 약 3.0×10^{12} 개 GC/g, 약 3.5×10^{12} 개 GC/g, 약 4.0×10^{12} 개 GC/g, 약 4.5×10^{12} 개 GC/g, 약 5.0×10^{12} 개 GC/g, 약 5.5×10^{12} 개 GC/g, 약 6.0×10^{12} 개 GC/g, 약 6.5×10^{12} 개 GC/g, 약 7.0×10^{12} 개 GC/g, 약 7.5×10^{12} 개 GC/g, 약 8.0×10^{12} 개 GC/g, 약 8.5×10^{12} 개 GC/g, 약 9.0×10^{12} 개 GC/g, 약 9.5×10^{12} 개 GC/g, 약 1.0×10^{13} 개 GC/g, 약 1.5×10^{13} 개 GC/g, 약 2.0×10^{13} 개 GC/g, 약 2.5×10^{13} 개 GC/g, 약 3.0×10^{13} 개 GC/g, 약 3.5×10^{13} 개 GC/g, 약 4.0×10^{13} 개 GC/g, 약 4.5×10^{13} 개 GC/g, 약 5.0×10^{13} 개 GC/g, 약 5.5×10^{13} 개 GC/g, 약 6.0×10^{13} 개 GC/g, 약 6.5×10^{13} 개 GC/g, 약 7.0×10^{13} 개 GC/g, 약 7.5×10^{13} 개 GC/g, 약 8.0×10^{13} 개 GC/g, 약 8.5×10^{13} 개 GC/g, 약 9.0×10^{13} 개 GC/g, 약 9.5×10^{13} 개 GC/g, 또는 약 1.0×10^{14} 개 GC/뇌 질량 g이다.

[0124] 일 실시형태에서, 치료 방법은 효소 대체 요법으로서 hGAA780I 단백질 또는 융합 단백질의 전달을 포함한다. 특정 실시형태에서, hGAA780I 단백질 또는 융합 단백질은 유전자 요법(본 명세서에서 제공되는 바와 같은 발현 카세트 또는 rAAV를 포함하지만, 이들로 제한되지 않음)과 조합하여 ERT로서 전달된다. 특정 실시형태에서, 방법은 하나 초과와 ERT(예를 들어, ha GAA780I 단백질 또는 융합 단백질을 또 다른 치료용 단백질, 예컨대, Lumizyme과 조합하여 포함하는 조성물)를 대상체에게 투여하는 단계를 포함한다. 본 명세서에 기재된 hGAA780I 단백질 또는 융합 단백질을 포함하는 조성물은 1, 2, 3, 4, 5, 6, 7, 8, 9, 10일, 또는 그 이상마다 대상체에게 투여될 수 있다. 투여는 매주, 매월, 또는 격월 투여가 처방된 외래 환자에게 정맥내 주입에 의할 수 있다. 화합물의 적절한 치료학적 유효 투약량은 치료하는 임상 의에 의해 선택되며 약 $1 \mu\text{g/kg}$ 내지 약 500 mg/kg , 약 10 mg/kg 내지 약 100 mg/kg , 약 20 mg/kg 내지 약 100 mg/kg 및 대략 20 mg/kg 내지 대략 50 mg/kg 을 포함한다. 일부 실시형태에서, 적합한 치료학적 용량은, 예를 들어, 0.1, 0.25, 0.5, 0.75, 1, 5, 10, 15, 20, 30, 40, 50, 60, 70, 및 100 mg/kg 으로부터 선택된다.

[0125] 특정 실시형태에서, 방법은 hGAA780I 단백질 또는 융합 단백질을 환자에게 매주 10 mg/환자 체중 kg 이상의 투약량으로 대상체에게 투여하는 단계를 포함한다. 종종 투약량은 주당 10 mg/kg 초과이다. 투약 요법은 주당 10 mg/kg 내지 주당 적어도 1000 mg/kg 의 범위일 수 있다. 전형적으로 투약 요법은 주당 10 mg/kg , 주당 15 mg/kg , 주당 20 mg/kg , 주당 25 mg/kg , 주당 30 mg/kg , 주당 35 mg/kg , 주당 40 mg/kg , 주당 45 mg/kg , 주당 60 mg/kg , 주당 80 mg/kg 및 주당 120 mg/kg 이다. 바람직한 요법에서, 10 mg/kg , 15 mg/kg , 20 mg/kg , 30 mg/kg 또는 40 mg/kg 이 매주 1회, 2회, 또는 3회 투여된다. 치료는 전형적으로 적어도 4주, 때때로 24주, 때때로 환자의 평생 동안 계속된다. 선택적으로, 인간 알파-글루코시다제의 수준은 치료 후(예를 들어, 혈장 또는 근육에서) 모니터링되고 검출된 수준이 정상인 값의 실질적으로 아래(예를 들어, 20% 미만)로 떨어질 때 추가 투약량이 투여된다. 일 실시형태에서, hGAA780I 단백질 또는 융합 단백질은 초기에 "고"용량(즉, "부하 용량")으로 투여되고, 그 다음 더 낮은 용량(즉, "유지 용량")으로 투여된다. 부하 용량의 예는 주당 1 내지 3회(예를 들어, 1, 2, 또는 3주 동안) 적어도 약 40 mg/환자 체중 kg 이다. 유지 용량의 예는 주당 적어도 약 5 내지 적어도 약 10 mg/환자 체중 kg , 또는 그 이상, 예컨대, 주당 20 mg/kg , 주당 30 mg/kg , 주당 40 mg/kg 이다. 특정 실시형태에서, 투약량은 투약 기간 동안 증가하는 속도로 투여된다. 이는 유동 정맥내 주입의 속도를 증가시킴으로써 또는 일정한 속도로 투여되는 hGAA780I 단백질 또는 융합 단백질의 농도를 증가시키는 구배를 사용함으로써 달성될 수 있다. 이러한 방식으로 투여하면 면역원성 반응의 위험을 감소시킬 수 있다. 특정 실시형태에서, 정맥내 주입은 수 시간(예를 들어, 1 내지 10시간, 바람직하게는 2 내지 8시간, 더 바람직하게는 3 내지 6시간)의 기간에 걸쳐 일어나고, 주입 속도는 투여 기간 동안 간격을 두고 증가된다.

[0126] 일 실시형태에서, 방법은 대상체가 면역억제 공동-요법을 받는 것을 더 포함한다. 이와 같은 공동-요법을 위한 면역억제제는 글루코코르티코이드, 스테로이드, 항대사물질, T-세포 저해제, 매크롤라이드(예를 들어, 라파마이신 또는 라파로그), 및 알킬화제, 항-대사물질, 세포독성 항생제, 항체, 또는 이무노필린에 대한 활성제를 포함하는 세포정지제를 포함하지만, 이들로 제한되지 않는다. 면역억제제는 질소 머스터드, 나이트로소우레아, 백금 화합물, 메토트렉세이트, 아자싸이오프린, 머캅토프린, 플루오로우라실, 닥티노마이신, 안트라사이클린, 미토마이신 C, 블레오마이신, 미트라마이신, IL-2 수용체- 또는 CD3-유도 항체, 항-IL-2 항체, 사이클로스포린, 타크로리무스, 시롤리무스, IFN- β , IFN- γ , 오피오이드, 또는 TNF- α (중양 피사 인자-알파) 결합제를 포함할 수 있다. 특정 실시형태에서, 면역억제제 요법은 유전자 요법 투여 후 0, 1, 2, 7일, 또는 이 이상 전에 시작될 수 있다. 이들 약물 중 하나 이상은 유전자 요법 투여 후 동일한 용량 또는 조정된 용량으로 계속될 수 있다. 이와 같은 요법은 필요에 따라 약 1주(7일), 약 60일, 또는 그 이상 동안 이루어질 수 있다.

- [0127] 일 실시형태에서, 본 명세서에 기재된 바와 같은 발현 카세트를 포함하는 조성물은 이를 필요로 하는 대상체에 1회 투여된다. 특정 실시형태에서, 발현 카세트는 rAAV를 통해 전달된다. 본 명세서에 기재된 조성물 및 방법이 본 명세서 전반에 걸쳐 기재된 다른 조성물, 요법, 양태, 실시형태 및 방법에 적용되는 것으로 의도됨이 이해되어야 한다.
- [0128] 본 명세서에서 제공되는 조성물 및 방법은 영아 발병형-포페병 또는 후기 발병형 포페병 및/또는 이와 연관된 증상을 치료하는 데 사용될 수 있다. 특정 실시형태에서, 효능은 질환의 하나 이상의 증상의 개선에 의해 또는 질환 진행의 둔화에 의해 결정될 수 있다. 영아 발병형-포페병의 증상은 근긴장 저하, 호흡기/호흡 문제, 간비대, 비후성 심근병증뿐만 아니라 심장, 근육, CNS(특히, 운동 뉴런)에의 글리코겐 저장을 포함하지만, 이들로 제한되지 않는다. 후기 발병형-포페병의 증상은 근위근 약화, 호흡기/호흡 문제뿐만 아니라 근육 및 운동 뉴런에의 글리코겐 저장을 포함하지만, 이들로 제한되지 않는다. 투여 경로는 환자의 상태 및/또는 진단을 기반으로 결정될 수 있다. 특정 실시형태에서, IV 및 ICM 경로의 조합을 통한 hGAA780I 단백질 또는 융합 단백질의 전달을 위해 본 명세서에 기재된 rAAV를 투여하는 단계를 포함하는, 영아-발병형 포페병 또는 후기-발병형 포페병으로 진단된 환자의 치료를 위한 방법이 제공된다. 일부 실시형태에서, 후기-발병형 포페병을 가진 것으로 확인된 환자는 rAAV의 전신 전달만을 포함하는 치료(예를 들어, IV만)를 투여받는다. 본 명세서에 기재된 바와 같이, rAAV를 포함하는 조성물의 전달은 효소 대체 요법(ERT)과 조합될 수 있다. 특정 실시형태에서, ERT와 조합하여 본 명세서에 기재된 rAAV를 ICM 전달하는 것을 포함하는, 포페병으로 진단된 대상체를 치료하기 위한 방법이 제공된다. 특정 실시형태에서, 영아-발병형 포페병을 가진 것으로 확인된 대상체는 ICM 주사를 통해 본 명세서에 기재된 rAAV를 투여받고 또한 말초 질환의 양상의 치료를 위해 ERT를 받는다.
- [0129] 본 명세서에 기재된 바와 같은 "핵산"은 RNA, DNA, 또는 이들의 변형일 수 있고, 단일 또는 이중 가닥일 수 있으며, 예를 들어, 관심이 있는 단백질을 인코딩하는 핵산, 올리고뉴클레오타이드, 핵산 유사체, 예를 들어, 펩타이드-핵산(PNA), 유사상보성(pseudocomplementary) PNA(pc-PNA), 잠금 핵산(LNA) 등을 포함하는 군으로부터 선택될 수 있다. 예를 들어, 이와 같은 핵산 서열은, 예를 들어, 전사 억제 인자로 작용하는 단백질을 인코딩하는 핵산 서열, 안티센스 분자, 리보자임, 작은 억제성 핵산 서열, 예를 들어, RNAi, shRNAi, siRNA, 마이크로RNAi(mRNAi), 안티센스 올리고뉴클레오타이드 등(이들로 제한되지 않음)을 포함하지만, 이들로 제한되지 않는다.
- [0130] 단백질, 펩타이드, 또는 폴리펩타이드를 "역번역"하는 방법은 당업자에게 알려져 있다. 일단 단백질의 서열이 알려지면, 아미노산 서열을 핵산 코딩 서열로 역번역하는 웹-기반 및 상업적으로 이용 가능한 컴퓨터 프로그램뿐만 아니라, 서비스-기반 회사가 있다. 예를 들어, EMBOS에 의한 backtranseq(ebi.ac.uk/Tools/st에서 온라인으로 이용 가능함); Gene Infinity(geneinfinity.org/sms/sms_backtranslation.html에서 온라인으로 이용 가능함); ExPasy(expasy.org/tools/에서 온라인으로 이용 가능함)를 참조한다. 일 실시형태에서, RNA 및/또는 cDNA 코딩 서열은 인간 세포에서 최적 발현을 위해 설계된다.
- [0131] 핵산 서열과 관련하여 용어 "동일성 백분율(%)", "서열 동일성", "서열 동일성 백분율", 또는 "동일한 백분율"은 관련성을 위해서 정렬되는 경우 동일한 2개의 서열에서의 잔기를 지칭한다. 서열 동일성 비교의 길이는 계능의 전장, 유전자 코딩 서열의 전장, 또는 적어도 약 500 내지 5000개의 뉴클레오타이드의 단편에 걸친 것일 수 있으며, 이것이 바람직하다. 그러나, 예를 들어, 적어도 약 9개의 뉴클레오타이드, 보통 적어도 약 20 내지 24개의 뉴클레오타이드, 적어도 약 28 내지 32개의 뉴클레오타이드, 적어도 약 36개 이상의 뉴클레오타이드의 더 작은 단편 간의 동일성이 또한 바람직할 수 있다.
- [0132] 동일성 백분율은 단백질의 전장, 폴리펩타이드, 약 32개의 아미노산, 약 330개의 아미노산, 또는 이들의 펩타이드 단편 또는 상응하는 핵산 서열 코딩 서열에 걸친 아미노산 서열에 대하여 용이하게 결정될 수 있다. 적합한 아미노산 단편은 적어도 약 8개의 아미노산 길이일 수 있고, 최대 약 700개의 아미노산일 수 있다. 일반적으로, 2개의 상이한 서열 사이의 "동일성", "상동성", 또는 "유사성"을 지칭하는 경우, "동일성", "상동성" 또는 "유사성"은 "정렬된" 서열을 참조로 결정된다. "정렬된" 서열 또는 "정렬"은, 종종 참조 서열과 비교하여 누락되거나 추가적인 염기 또는 아미노산에 대한 수정을 포함하는, 다수의 핵산 서열 또는 단백질(아미노산) 서열을 지칭한다.
- [0133] 정렬은 다양한 공공 또는 상업적으로 입수 가능한 다중 서열 정렬 프로그램 중 임의의 것을 사용하여 수행된다. 서열 정렬 프로그램, 예를 들어, "Clustal X", "Clustal Omega", "MAP", "PIMA", "MSA", "BLOCKMAKER", "MEME" 및 "Match-Box" 프로그램이 아미노산 서열에 이용 가능하다. 일반적으로, 이들 프로그램 중 임의의 것은 디폴트 설정으로 사용되지만, 당업자는 필요에 따라 이러한 설정을 변경할 수 있다. 대안적으로, 당업자는 적어도 참조

된 알고리즘 및 프로그램에 의해 제공되는 것과 같은 동일성 또는 정렬의 수준을 제공하는 또 다른 알고리즘 또는 컴퓨터 프로그램을 이용할 수 있다. 예를 들어, 문헌[J. D. Thomson et al, Nucl. Acids. Res., "A comprehensive comparison of multiple sequence alignments", 27(13):2682-2690 (1999)]을 참조한다.

[0134] 다중 서열 정렬 프로그램은 또한 핵산 서열에 대해 이용 가능하다. 이와 같은 프로그램의 예는 "Clustal W", "Clustal Omega", "CAP Sequence Assembly", "BLAST", "MAP" 및 "MEME"를 포함하며, 이들은 인터넷 웹 서버를 통해 접근 가능하다. 이와 같은 프로그램에 대한 다른 공급원은 당업자에게 알려져 있다. 대안적으로, Vector NTI 유틸리티가 또한 사용된다. 또한 상기 기재된 프로그램에 포함된 것을 포함하여, 뉴클레오타이드 서열 동일성을 측정하는 데 사용될 수 있는 당업계에 알려진 다수의 알고리즘이 있다. 또 다른 예로서, 폴리뉴클레오타이드 서열은 GCG 버전 6.1의 프로그램인 Fasta™를 사용하여 비교될 수 있다. Fasta™은 질의 서열과 검색 서열 사이의 최상의 중복 영역의 정렬 및 서열 동일성 백분율을 제공한다. 예를 들어, 핵산 서열 사이의 서열 동일성 백분율은 본 명세서에 참조에 의해 인용되는 GCG 버전 6.1에서 제공되는 바와 같이 디폴트 파라미터(단어 크기 6, 득점 매트릭스에 대한 NOPAM 인자)와 함께 Fasta™를 사용하여 결정될 수 있다.

[0135] 본 명세서에서 사용되는 바와 같이, 용어 "조절 서열", 또는 "발현 제어 서열"은 작동 가능하게 연결된 단백질을 인코딩하는 핵산 서열의 전사를 유도, 억제, 또는 다르게는 제어하는 개시자 서열, 인핸서 서열, 및 프로모터 서열과 같은 핵산 서열을 지칭한다.

[0136] 핵산 서열 또는 단백질을 설명하는 데 사용되는 용어 "외인성"은 핵산 또는 단백질이 염색체 또는 숙주 세포에 존재하는 위치에서 자연적으로 발생하지 않음을 의미한다. 외인성 핵산 서열은 또한 동일한 숙주 세포 또는 대상체로부터 유래되고 이들 내로 삽입되지만, 비-자연 상태, 예를 들어, 상이한 복제물 수, 또는 상이한 조절 요소의 제어 하에 존재하는 서열을 지칭한다.

[0137] 핵산 서열 또는 단백질을 설명하는 데 사용되는 용어 "이종"은 핵산 또는 단백질이 단백질이 발현되는 숙주 세포 또는 대상체와 상이한 유기체 또는 동일한 유기체의 상이한 종으로부터 유래되었음을 의미한다. 플라스미드, 발현 카세트, 또는 벡터에서 단백질 또는 핵산과 관련하여 사용될 때 용어 "이종"은 단백질 또는 핵산이 당해 단백질 또는 핵산이 자연에서 서로 동일한 관계에서 발견되지 않는 또 다른 서열 또는 하위서열과 함께 존재함을 나타낸다.

[0138] "포함하는"은 다른 구성요소 또는 방법 단계를 포함하는 것을 의미하는 용어이다. "포함하는"이 사용될 때, 관련된 실시형태는 다른 구성요소 또는 방법 단계를 배제하는 "~로 이루어진" 용어, 및 실시형태 또는 발명의 성질을 실질적으로 변화시키는 임의의 구성요소 또는 방법 단계를 배제하는 "~로 본질적으로 이루어진" 용어를 사용하는 설명을 포함하는 것으로 이해되어야 한다. 본 명세서에 다양한 실시형태가 "포함하는" 용어를 사용하여 제시되어 있지만, 다른 상황 하에서, 관련 실시형태는 또한 "~로 이루어진" 또는 "~로 본질적으로 이루어진" 용어를 사용하여 기재되는 것으로 이해되어야 한다.

[0139] 본 명세서에서 사용되는 바와 같이, 숫자(nn) 값이 뒤따르는 용어 "e"는 지수를 지칭하며, 이 용어는 " $\times 10^{nn}$ "과 호환 가능하게 사용된다. 예를 들어, 3e13은 3×10^{13} 과 같다.

[0140] 용어 "하나"는 하나 이상을 지칭하며, 예를 들어, "벡터"는 하나 이상의 벡터(들)를 나타내는 것으로 이해된다는 점에 유의해야 한다. 이와 같이, 용어 "하나", "하나 이상", 및 "적어도 하나"는 본 명세서에서 호환 가능하게 사용된다.

[0141] 본 명세서에서 사용되는 바와 같이, 용어 "약"은 달리 명시되지 않는 한, 주어진 참조로부터 $\pm 10\%$ 의 변동성을 의미한다.

[0142] 실시예

[0143] 이제 하기 실시예를 참조하여 본 발명을 설명한다. 이들 실시예는 단지 예시의 목적으로 제공되며, 본 발명은 결코 이들 실시예로 제한되는 것으로 해석되어서는 안되고 본 명세서에서 제공되는 교시내용의 결과로서 명백해지는 임의의 그리고 모든 변형을 포괄하는 것으로 해석되어야 한다.

[0144] 실시예 1: 물질 및 방법

[0145] 벡터 생성

[0146] 780번 위치에 Val이 있는 참조 GAA 서열, 및 V780I 돌연변이가 있는 서열을 역번역하고 뉴클레오타이드 서열을 조작하여 CAG 프로모터 하에 발현 카세트를 이용하여 AAV 생성을 위한 시스-플라스미드를 생성하였다. 추가적인

로, 비-조작된 서열과의 비교를 위해 자연 hGAA(참조 서열)에 대한 cDNA 서열을 동일한 AAV-시스 백본 내로 클로닝하였다. 앞서 기재된 바와 같이 Penn Vector Core에 의해 AAVhu68 벡터를 생성하고 적정하였다. (Lock, et al. 2010, Hum Gene Ther 21(10): 1259-1271). 간단히 말해서, HEK293 세포를 삼중 형질감염시킨 다음, 배양액 상청액을 수확하고, 농축시킨 다음, 이오디사놀 구배를 이용하여 정제하였다. 정제된 벡터를 앞서 기재된 바와 같이 토끼 베타-글로빈 폴리A 서열을 표적화하는 프라이머를 사용하여 액적 디지털 PCR로 적정하였다(Lock, et al. (2014). Hum Gene Ther Methods 25(2): 115-125).

[0147] 동물

[0148] 마우스

[0149] 품페 마우스(Gaa 녹아웃(-/-), C57BL/6/129 배경) 원조를 Jackson Labs에서 구입하였다(스톡#004154, 6^{네오} 마우스라고도 함). 동일한 한배 새끼 내에서 무효(null) 및 WT 대조군을 생성하기 위해 이형접합체 대 이형접합체 교배를 사용하여, 유전자 요법 프로그램(Gene Therapy Program) AAALAC의 공인된 장벽 마우스 시설에서 번식 콜로니를 유지하였다. Gaa 녹아웃 마우스는 품페병에 널리 사용되는 모델이다. 이는 심장, 중추신경계, 골격근, 및 횡경막에 리소좀 글리코겐의 점진적인 축적을 나타내며, 이동성이 감소하고 점진적인 근육 약화가 있다. 작은 크기, 재현가능한 표현형, 및 효율적인 교배는 전임상 후보 생체내 스크리닝에 최적인 신속한 연구를 가능하게 한다.

[0150] 동물 수용실을 64 내지 79°F(18 내지 26°C)의 온도 범위에서 30 내지 70%의 습도 범위로 유지하였다.

[0151] 젓을 뱉 때까지 동물을 부모 및 한배 새끼와 함께 수용하였고, 그 다음 Translational Research Laboratories(TLR) GTP 동물 사육장에서 케이지당 2 내지 5마리의 동물의 표준 케이지에 수용하였다. 모든 케이지 크기 및 수용 조건은 실험 동물의 관리 및 사용에 대한 지침(Guide for the Care and Use of Laboratory Animals)을 따랐다. 케이지, 물병, 및 침구 기재를 차단 시설로 고압멸균처리하였다.

[0152] 자동으로 제어되는 12시간 명/암 주기를 유지하였다. 각각의 암기는 1900시간(± 30분)에 시작되었다. 음식은 임의로 제공하였다(Purina, LabDiet®, 5053, Irradiated, PicoLab®, Rodent Diet 20, 251b). 물은 각각의 수용 케이지에서 개별적으로 배치된 물병을 통해 모든 동물이 임의로 접근 가능하였다. 최소한, 매주 케이지를 바꾸는 동안 일주일에 1번 물병을 교체하였다. 물 공급은 필라델피아 시에서 가져와 Getinge 정수기를 사용하여 염소처리하였다. ULAR에 의해 염소화 수준을 매일 테스트하였고 2 내지 4 ppm(parts per million: 백만분율)으로 유지하였다. 각각의 수용 케이지에 농축물로서 Nestlets™를 제공하였다.

[0153] 생체내 연구 및 조직학

[0154] 측면 꼬리 정맥(IV)을 통해 0.1ml로 5×10^{11} 개 GC(대략 2.5×10^{13} 개 GC/kg) 용량 또는 5×10^{10} 개 GC(대략 2.5×10^{12} 개 GC/kg) 용량의 AAVhu68.CAG.hGAA(다양한 hGAA 작제물)를 마우스에게 투여하고, 혈청 분리를 위해 벡터 투약 후 제7일 및 제21일에 채혈한 다음, (혈장 분리를 위해) 말기에 채혈하고, 주사하고 28일 후에 방혈에 의해 안락사시켰다. 뇌로 시작하여 조직을 즉시 수집하였다.

장기 목록, 부검

조직	급속 냉동 (단백질 추출용)	포말린 침지 (조직학용)
혈장	×	
좌뇌		×
우뇌	×	
경추		×
흉추 + 요추	×	
심장	×	×
간	×	×
횡경막	×	×
삼두근	×	×
대퇴사두근	×	×
비장근	×	×
전경골근	×	×

[0155]

[0156]

조직학용 조직을 표준 방법을 사용하여 포말린으로 고정시키고 파라핀에 포매시켰다. 표준 방법을 사용하여 뇌 및 척수 절편을 록솔 패스트 블루(록솔 패스트 블루 염색 키트, Abcam ab150675)로 염색하고, 말초 장기를 PAS (파요오드산-슈프)로 염색하여 조직에서 글리코젠과 같은 다당류를 검출하였다. hGAA에 대한 면역염색을 포말린-고정 파라핀-포매 샘플에서 수행하였다. 절편을 탈파라핀화하고, 항원 회수를 위해 10mM 시트레이트 완충제(pH 6.0)에서 끓이며, PBS 중 1% 당나귀 혈청 + 0.2% 트리톤으로 15분 동안 차단한 다음, 1차 항체(Sigma HPA029126 항-hGAA 항체) 및 차단 완충액에 희석된 비오티닌화된 2차 항체와 함께 순차적으로 인큐베이션하였고; HRP 기반 비색 반응을 사용하여 신호를 검출하였다.

[0157]

공인 수의 병리학자에 의해 맹검 방식으로 슬라이드를 검토하였다. 반-정량적 스코어링 시스템을 확립하여 저장 및/또는 액포를 제시하는 세포의 총 백분율에 의해 결정되는, 근육의 펌프-관련 조직학적 병변(글리코젠 저장 및 자가포식 추적)의 중증도를 측정하였다:

히스토 스코어링 저장

0	0 %
1	1 내지 9 %
2	10 내지 49 %
3	50 내지 74 %
4	75 내지 100 %

[0158]

[0159]

또한 적용 가능한 경우 벡터 관련 조직병리학적 병변을 추정하였다.

[0160]

시험관내 hGAA780I 효소 성능 특성화

[0161]

GAA 활성

[0162]

균질화된 조직의 혈장 또는 상청액을 5.6mM 4-MU- α -글루코피라노사이드(pH 4.0)와 혼합하고 37°C에서 3시간 동안 인큐베이션하였다. 0.4M 탄산나트륨(pH 11.5)을 이용하여 반응을 정지시킨다. Victor3 형광측정기(355nm에서 여기 및 460nm에서 발광)를 사용하여 상대 형광 단위(RFU)를 측정한다. 4-MU의 표준 곡선으로부터 보간법에 의해 nmol/ml/시간 단위의 활성을 계산한다. 개별 조직 샘플에서 활성 수준을 호모제네이트 상청액의 총 단백질 함량에 대해 정규화한다. 혈장 샘플에 동일한 부피를 사용한다.

- [0163] 실시예 2: 폼페 마우스에서 rAAVhu68.hGAA 벡터의 평가
- [0164] 폼페 마우스에 IV 전달하기 위해 AAV 벡터를 멸균 PBS에 희석하였다. 테스트 물품은 AAVhu68.CAG.hGAAco.rBG, AAVhu68.CAG.hGAAcoV780I.rBG, AAVhu68.CAG.BiP-vIGF2.hGAAco.rBG, AAVhu68.CAG.BiP-vIGF2.hGAAcoV780I.rBG, 및 AAVhu68.CAG.sp7co.Δ8.hGAAcoV780I.rBG를 포함하였다. hGAAcoV780I에 대한 조작된 서열은 서열번호 4에 제공되어 있다. 야생형 및 비히클 대조군을 연구에 포함하였다.
- [0165] 간(도 1A, 도 1B), 심장(도 2A, 도 2B), 대퇴사두근(도 3A, 도 3B), 뇌(도 4A, 도 4B), 혈장(도 9)을 포함하여, 처리된 마우스로부터 수집된 다양한 조직에서 hGAA 단백질 발현 및 활성을 측정하였다. 모든 프로모터는 저용량 및 고용량 둘 다에서 간에서 동등하게 잘 수행하였다. UbC 프로모터 하에서 발현하는 벡터의 투여는 두 가지 용량 모두에서 골격근에서 더 낮은 활성을 초래하였고, CAG 프로모터가 있는 벡터는 최고의 전체 활성을 가졌다. UbC 프로모터가 있는 벡터는 또한 두 가지 용량 모두에서 심장에서 더 낮은 활성을 가졌다.
- [0166] 폼페 마우스 비히클(PBS) 대조군(도 5D)은 심장에서 뚜렷한 글리코겐 저장(PAS로 염색된 절편에서 어두운 염색)을 나타내었다. 야생형 마우스 및 모든 벡터로 처리된 마우스는 저장의 거의 완전하거나 완전한 클리어런스를 가졌다. 그러나, hGAA 참조 서열(V780)을 인코딩하는 벡터를 받은 2개 그룹은 중등도 내지 뚜렷한 섬유화 림프구성 심근염을 나타내었으며(도 5B 및 도 5C), 상기 섬유화 림프구성 심근염은 hGAA 천연 이식유전자를 받은 8마리 동물 중 7마리 및 BiP 및 vIGF2 변형이 있는 조작된 hGAA를 받은 8마리 동물 중 3마리에 존재하였다. hGAAcoV780I 효소를 받은 마우스 중 어느 것도 심근염을 앓지 않았기 때문에(도 5E, 도 5F, 및 도 5G), 이러한 병변은 벡터 관련된 것, 보다 구체적으로는 hGAA 참조 서열 특이적인 것으로 간주되었다.
- [0167] 대퇴사두근 조직의 분석은, 야생형 마우스 및 추가 변형이 있거나 없이 V780I 변이체를 인코딩하는 벡터로 처리된 모든 마우스가 저장 및 자가포식 축적의 거의 완전하거나 완전한 클리어런스를 가졌다는 것을 보였다(도 6A 내지 도 6G). 그러나 hGAAV780의 참조 서열을 인코딩하는 벡터를 받은 2개의 그룹은 남아있는 최소 내지 중간 정도의 글리코겐 저장뿐만 아니라 자가포식 축적을 나타내었으며(도 10), 이와 함께 폼페병의 2가지 주요 특징의 차선의 교정을 입증하였다. 최상의 결과는 천연 형태로 또는 BiP-vIGF2 변형이 있는 V780I 변이체를 인코딩하는 2개의 벡터의 전달로부터 관찰되었다. sp7-델타8 변형은 폼페병에 기인한 조직학적 병변의 일관되지 않는 교정을 유발하는 것으로 나타났다. 참조 hGAAV780 서열을 인코딩하는 2가지 작제물은 글리코겐 저장 및 축적을 소거하는 데 차선택이었다.
- [0168] 고용량 IV 투여($5 \times 10^{11} = 2.5 \times 10^{13}$ GC/kg)에서, hGAAcoV780I 및 BiP-vIGF2.hGAAcoV780I는 대퇴사두근에서 거의 정상 수준의 글리코겐 수준을 나타내었고, 세포 내로의 hGAA 흡수가 현저하게 더 우수하였다. 전경골근(TA) 및 비복근을 포함하는 다른 골격근의 평가는 유사한 결과를 나타내었다(V780I 및 글리코겐 및 중심 자가포식 액포 둘 다의 소거가 있는 변이체). 모든 작제물은 심장에서 글리코겐 저장을 감소시켰으며, 이 때 BiP-vIGF2.hGAAcoV780I 투여가 최저 수준을 초래하였다. 대퇴사두근의 글리코겐 수준이 거의 정상이었지만, PAS 염색은 약간의 차이를 나타내었으며, 이 때 hGAAcoV780I 및 BiP-vIGF2.hGAAcoV780I가 최상의 결과를 나타내었다.
- [0169] 저용량 IV 투여($5 \times 10^{10} = 2.5 \times 10^{12}$ GC/kg)에서, BiP-vIGF2.hGAAcoV780I는 hGAAcoV780I보다 심장 및 대퇴사두근에서 더 나은 글리코겐 감소를 보였다. 뇌 및 척수의 글리코겐 수준은, 아마도 더 나은 표적화로 인하여, BiP-vIGF2.hGAAcoV780I에 의해 거의 정상이었고, 심지어 약 15%의 조직 수준이었다. CNS에서, 조직된 작제물과 V780I 변이체 사이의 강력한 상승 효과가 관찰되었다. BiP-vIGF2.hGAAcoV780I만이 CNS 글리코겐을 소거하였다.
- [0170] 도 8에 나타난 바와 같이, 척수 조직학의 평가는 AAVhu68.BiP-vIGF2.hGAAcoV780I로 처리된 마우스가 글리코겐 저장의 거의 완전하거나 완전한 소거를 가졌던 반면, 참조 hGAAV780 효소를 인코딩하는 벡터로 처리된 마우스는 남아있는 글리코겐 저장을 가졌음을 나타내었다.
- [0171] (서열목록 자유 텍스트)

[0172] 숫자 식별자 <223> 하에 자유 텍스트를 포함하는 서열에 대해 하기 정보가 제공된다.

서열번호 (자유 텍스트를 포함)	<223> 하의 자유 텍스트
3	<223> synthetic construct <220> <221> MISC_FEATURE <222> (1)..(27) <223> Signal peptide <220> <221> MISC_FEATURE <222> (70)..(952) <220> <221> MISC_FEATURE

[0173]

서열번호 (자유 텍스트를 포함)	<223> 하의 자유 텍스트
	<222> (123)..(952) <223> 76kD GAA Protein with V780I <220> <221> MISC_FEATURE <222> (204)..(952) <223> 70 kD GAA Protein with V780I
4	<223> Engineered hGAAI Coding sequence
6	<223> Fusion Protein comprising hGAA780I
7	<223> Engineered sequence encoding fusion protein comprising GAAV780I <220> <221> misc_feature <222> (810)..(810) <223> V810I
8	<223> CAG promoter <220> <221> misc_feature <222> (1)..(243) <223> CMV early enhancer element <220> <221> misc_feature <222> (244)..(525) <223> Chicken Beta actin promoter

[0174]

서열번호 (자유 텍스트를 포함)	<223> 하의 자유 텍스트
	<220> <221> misc_feature <222> (526)..(934) <223> hybrid intron in CAG (b-actin intron and rabbit beta 1-globin <220> <221> misc_feature <222> (526)..(934) <223> hybrid intron
9	<223> Rabbit globin polyA
12	<223> Engineered hGAAV780I signal peptide <220> <221> sig_peptide <222> (1)..(81) <220> <221> CDS <222> (1)..(81)
13	<223> Synthetic Construct
14	<223> engineered hGAAV780I mature protein
15	<223> Synthetic Construct
16	<223> Engineered DNA for hGAA780I 123-890 <220>

[0175]

서열번호 (자유 텍스트를 포함)	<223> 하의 자유 텍스트
	<221> CDS <222> (1)..(2304)
17	<223> Synthetic Construct
18	<223> Engineered hGAA 70kD cDNA <220> <221> CDS <222> (1)..(2247)
19	<223> Synthetic Construct
20	<223> Engineered DNA for hGAAV780I 76 kD protein <220> <221> CDS <222> (1)..(2490)
21	<223> Synthetic Construct
22	<223> synthetic construct <220> <221> CDS <222> (1)..(2952) <220> <221> misc_feature <222> (1)..(270) <223> BiP signal peptide + vIGF2 + 2GS extension

[0176]

서열번호 (자유 텍스트를 포함)	<223> 하의 자유 텍스트
	<220> <221> misc_feature <222> (271)..(2952) <223> engineered DNA for hGAA 61 - 952 780I <220> <221> misc_feature <222> (2428)..(2430) <223> Ile codon
23	<223> Synthetic Construct
24	<223> synthetic construct <220> <221> CDS <222> (1)..(2952) <220> <221> misc_feature <222> (1)..(270) <223> BiP-vIGF peptide <220> <221> misc_feature <222> (1)..(270) <223> BiP signal peptide + vIGF2+2GS extension <220>

[0177]

서열번호 (자유 텍스트를 포함)	<223> 하의 자유 텍스트
	<221> misc_feature <222> (271)..(2952) <223> hGAA 61-952 V780 DNA <220> <221> misc_feature <222> (2428)..(2430) <223> codon for hGAA 780 Valine
25	<223> Synthetic Construct
26	<223> miRNA target sequence
27	<223> miRNA target sequence

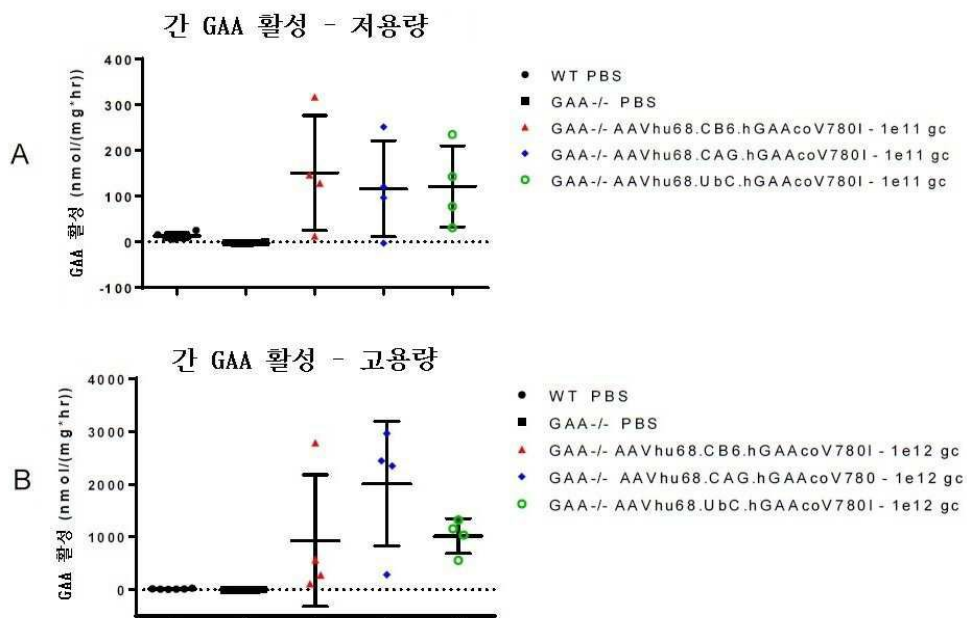
[0178]

[0179]

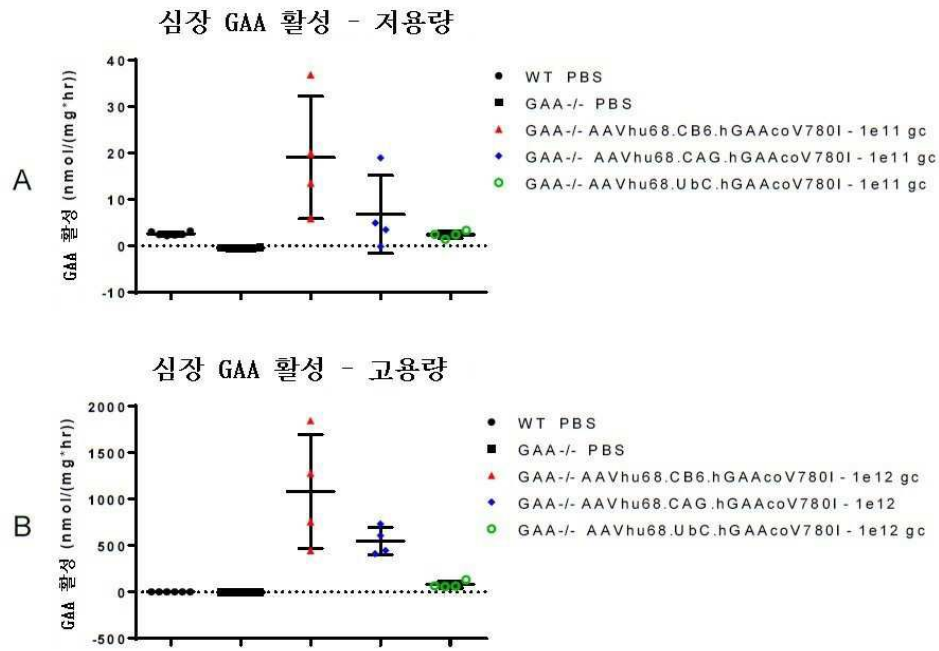
본 명세서에 인용된 모든 문헌은 본 명세서에 참조에 의해 원용된다. 2019년 10월 10일자로 출원된 미국 가특허 출원 제62/913,401호, 및 2019년 4월 30일자로 출원된 미국 가특허 출원 제62/840,911호는 서열목록과 함께 전문이 참조에 의해 원용된다. 명칭이 "19-8856PCT2_ST25.txt"인 본 명세서와 함께 제출된 서열목록 및 그 안의 서열 및 텍스트는 참조에 의해 원용된다. 본 발명이 특정 실시형태를 참조하여 기재되었지만, 본 발명의 사상을 벗어나지 않으면서 변형이 이루어질 수 있음이 이해될 것이다. 이와 같은 변형은 첨부된 청구범위의 범주 내에 속하는 것으로 의도된다.

도면

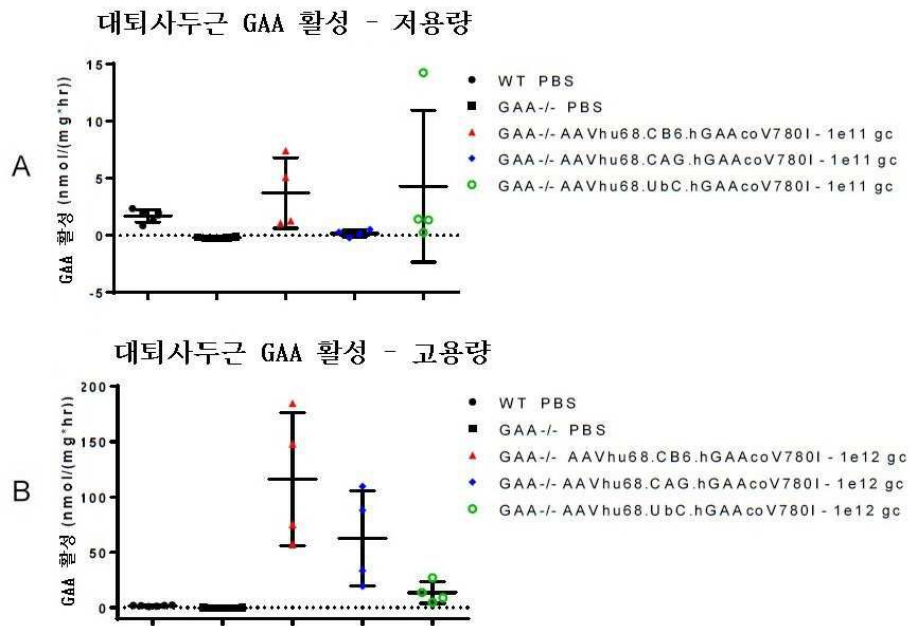
도면1



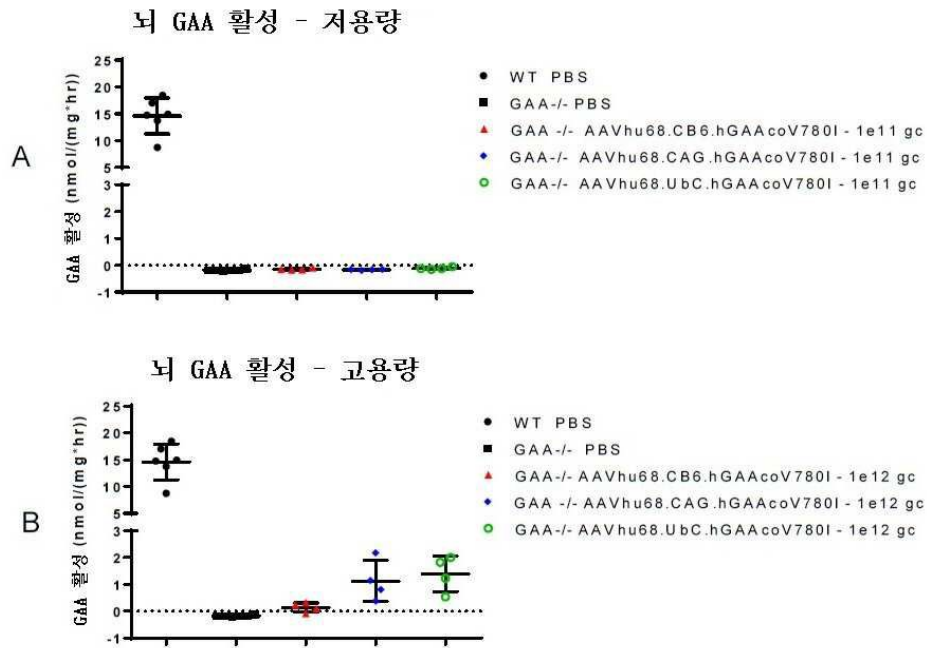
도면2



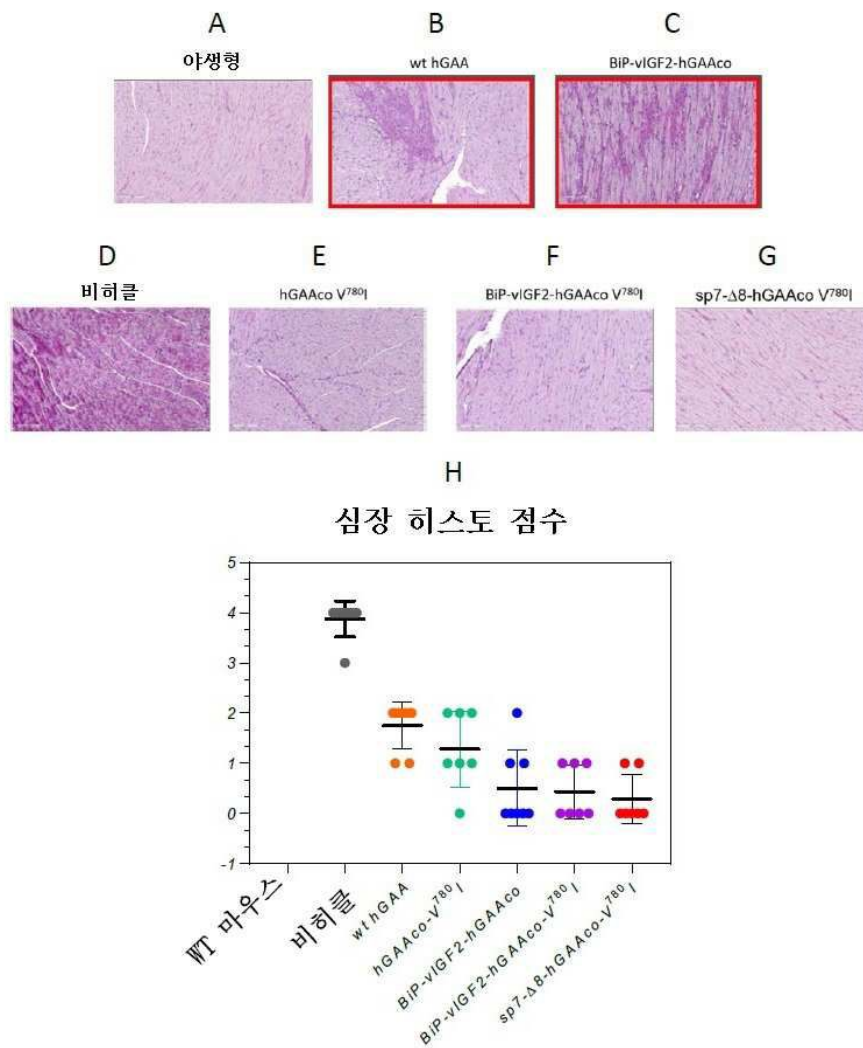
도면3



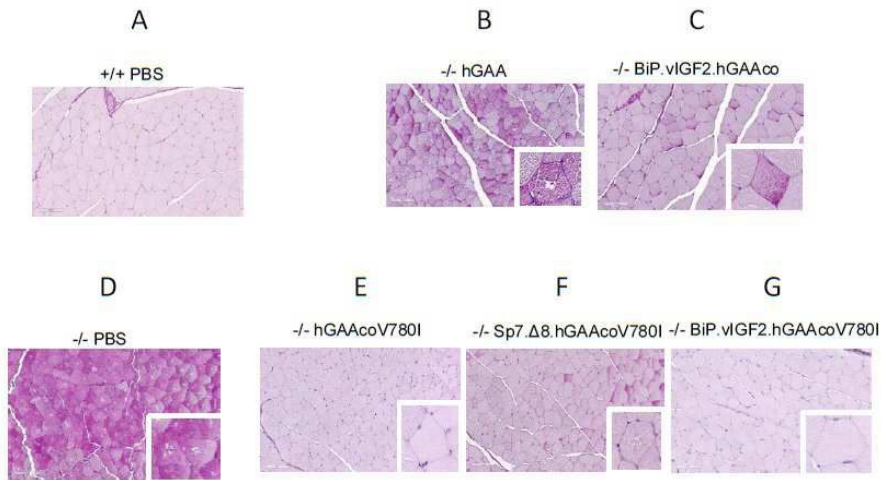
도면4



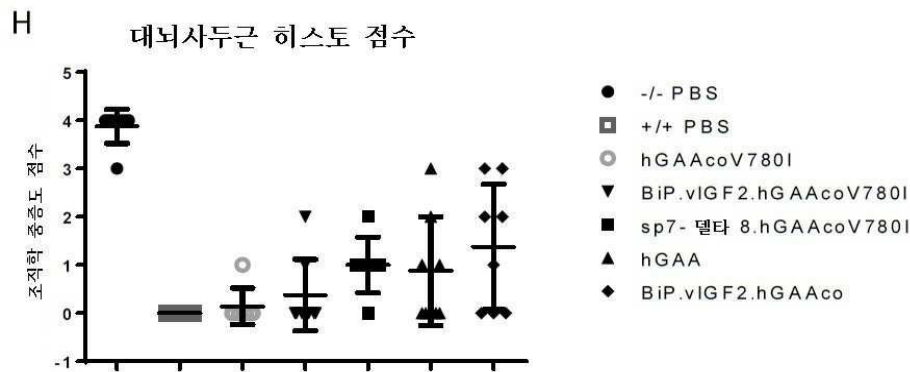
도면5



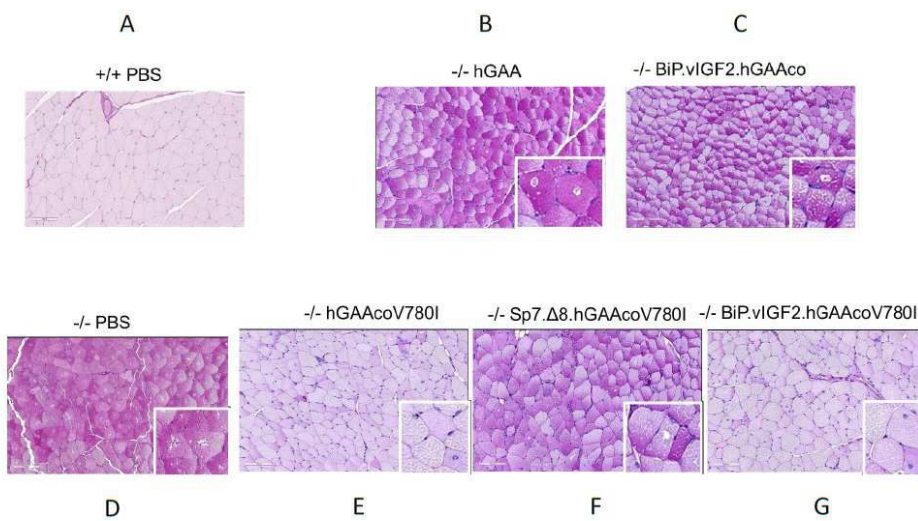
도면6ag



도면6h

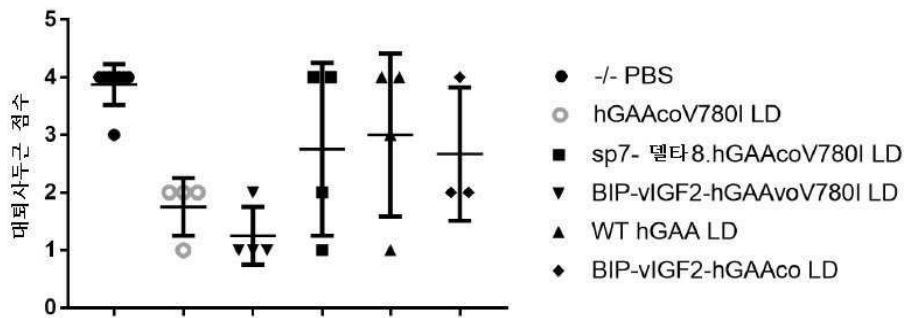


도면7ag

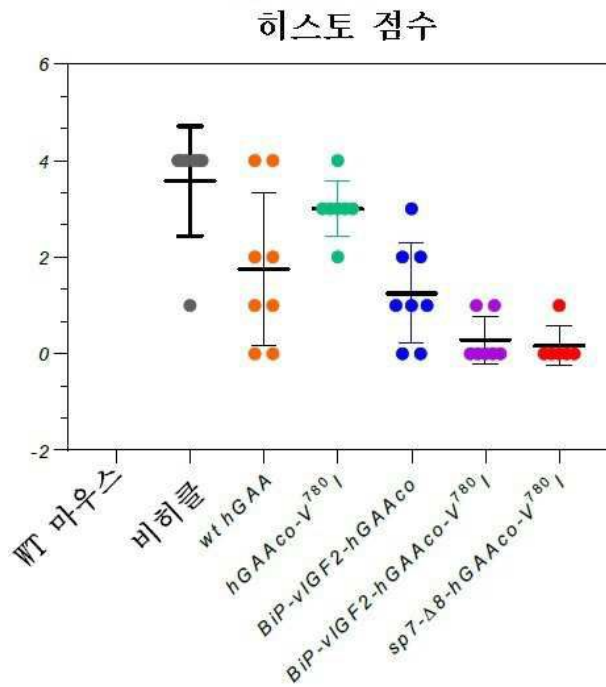


도면7h

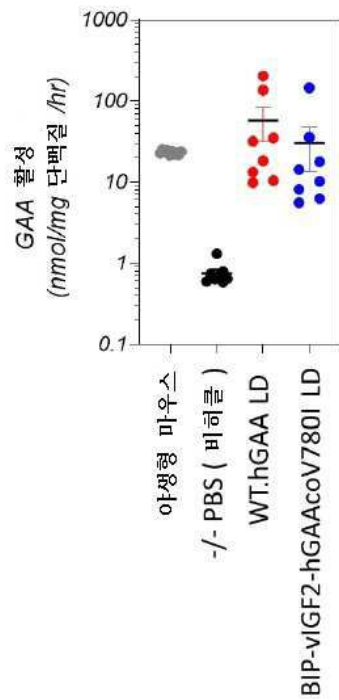
H 대퇴사두근 히스토 점수, 저용량, 수컷



도면8



도면9



서열목록

SEQUENCE LISTING

<110> THE TRUSTEES OF THE UNIVERSITY OF PENNSYLVANIA

<120> COMPOSITIONS USEFUL FOR TREATING POMPE DISEASE

<130> 19-8856PCT2

<140> PCT/US2020/030493

<141> 2020-04-29

<150> US 62/913,401

<151> 2019-10-10

<150> US 62/840,911

<151> 2019-04-30

<160> 31

<170> PatentIn version 3.5

<210> 1

<211> 2211

<212> DNA

<213> adeno-associated virus hu.68

<220><221> CDS

<222> (1)..(2211)

<400> 1

atg gct gcc gat ggt tat ctt cca gat tgg ctc gag gac aac ctc agt 48

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser

1 5 10 15

gaa ggc att cgc gag tgg tgg gct ttg aaa cct gga gcc cct caa ccc 96

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro

20 25 30

aag gca aat caa caa cat caa gac aac gct cgg ggt ctt gtg ctt ccg 144

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro

35 40 45

ggt tac aaa tac ctt gga ccc ggc aac gga ctc gac aag ggg gag ccg 192

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro

50 55 60

gtc aac gaa gca gac gcg gcg gcc ctc gag cac gac aag gcc tac gac 240

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

65 70 75 80

cag cag ctc aag gcc gga gac aac ccg tac ctc aag tac aac cac gcc 288

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala

85 90 95

gac gcc gag ttc cag gag cgg ctc aaa gaa gat acg tct ttt ggg ggc 336

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

100 105 110

aac ctc ggg cga gca gtc ttc cag gcc aaa aag agg ctt ctt gaa cct 384

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro

115 120 125

ctt ggt ctg gtt gag gaa gcg gct aag acg gct cct gga aag aag agg 432

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg

130 135 140

cct gta gag cag tct cct cag gaa ccg gac tcc tcc gtg ggt att ggc 480

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Val Gly Ile Gly

145 150 155 160

aaa tcg ggt gca cag ccc gct aaa aag aga ctc aat ttc ggt cag act 528

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr

 165 170 175

ggc gac aca gag tca gtc ccc gac cct caa cca atc gga gaa cct ccc 576

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro

 180 185 190

gca gcc ccc tca ggt gtg gga tct ctt aca atg gct tca ggt ggt ggc 624

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly

 195 200 205

gca cca gtg gca gac aat aac gaa ggt gcc gat gga gtg ggt agt tcc 672

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser

 210 215 220

tcg gga aat tgg cat tgc gat tcc caa tgg ctg ggg gac aga gtc atc 720

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile

225 230 235 240

acc acc agc acc cga acc tgg gcc ctg ccc acc tac aac aat cac ctc 768

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu

 245 250 255

tac aag caa atc tcc aac agc aca tct gga gga tct tca aat gac aac 816

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn

 260 265 270

gcc tac ttc ggc tac agc acc ccc tgg ggg tat ttt gac ttc aac aga 864

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg

 275 280 285

ttc cac tgc cac ttc tca cca cgt gac tgg caa aga ctc atc aac aac 912

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn

 290 295 300

aac tgg gga ttc cgg cct aag cga ctc aac ttc aag ctc ttc aac att	960
Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile	
305 310 315 320	
cag gtc aaa gag gtt acg gac aac aat gga gtc aag acc atc gct aat	1008
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn	
325 330 335	
aac ctt acc agc acg gtc cag gtc ttc acg gac tca gac tat cag ctc	1056
Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu	
340 345 350	
ccg tac gtg ctc ggg tgc gct cac gag ggc tgc ctc ccg ccg ttc cca	1104
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro	
355 360 365	
gcg gac gtt ttc atg att cct cag tac ggg tat cta acg ctt aat gat	1152
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp	
370 375 380	
gga agc caa gcc gtg ggt cgt tgc tcc ttt tac tgc ctg gaa tat ttc	1200
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe	
385 390 395 400	
ccg tgc caa atg cta aga acg ggt aac aac ttc cag ttc agc tac gag	1248
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu	
405 410 415	
ttt gag aac gta cct ttc cat agc agc tat gct cac agc caa agc ctg	1296
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu	
420 425 430	
gac cga ctc atg aat cca ctc atc gac caa tac ttg tac tat ctc tca	1344
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser	
435 440 445	
aag act att aac ggt tct gga cag aat caa caa acg cta aaa ttc agt	1392
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser	
450 455 460	
gtg gcc gga ccc agc aac atg gct gtc cag gga aga aac tac ata cct	1440

625	630	635	640	
aag cac ccg cct cct cag atc ctc atc aaa aac aca cct gta cct gcg				1968
Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala				
	645	650	655	
gat cct cca acg gct ttc aac aag gac aag ctg aac tct ttc atc acc				2016
Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr				
	660	665	670	
cag tat tct act ggc caa gtc agc gtg gag att gag tgg gag ctg cag				2064
Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln				
	675	680	685	
aag gaa aac agc aag cgc tgg aac ccg gag atc cag tac act tcc aac				2112
Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn				
	690	695	700	
tat tac aag tct aat aat gtt gaa ttt gct gtt aat act gaa ggt gtt				2160
Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val				
705	710	715	720	
tat tct gaa ccc cgc ccc att ggc acc aga tac ctg act cgt aat ctg				2208
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu				
	725	730	735	
taa				2211
<210> 2				
<211> 736				
<212> PRT				
<213> adeno-associated virus hu.68				
<400> 2				
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser				
1	5	10	15	
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro				
	20	25	30	
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro				
	35	40	45	

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

290 295 300
 Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
 305 310 315 320
 Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn

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

 385 390 395 400
 Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
 405 410 415
 Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
 420 425 430
 Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
 435 440 445
 Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser

 450 455 460
 Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
 465 470 475 480
 Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
 485 490 495
 Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
 500 505 510
 Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys

 515 520 525
 Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
 530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
 545 550 555 560
 Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
 565 570 575
 Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln

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

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

705 710 715 720
 Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
 725 730 735

<210> 3

<211> 952

<212> PRT

<213> Artificial Sequence

<220><223> synthetic construct

<220><221> MISC_FEATURE

<222> (1)..(27)

<223> Signal peptide

<220><221> MISC_FEATURE

<222> (70)..(952)

<220><221> MISC_FEATURE

<222> (123)..(952)

<223> 76kD GAA Protein with V780I

<220><221>

MISC_FEATURE

<222> (204)..(952)

<223> 70 kD GAA Protein with V780I

<400> 3

Met Gly Val Arg His Pro Pro Cys Ser His Arg Leu Leu Ala Val Cys

1 5 10 15

Ala Leu Val Ser Leu Ala Thr Ala Ala Leu Leu Gly His Ile Leu Leu

20 25 30

His Asp Phe Leu Leu Val Pro Arg Glu Leu Ser Gly Ser Ser Pro Val

35 40 45

Leu Glu Glu Thr His Pro Ala His Gln Gln Gly Ala Ser Arg Pro Gly

50 55 60

Pro Arg Asp Ala Gln Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr

65 70 75 80

Gln Cys Asp Val Pro Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys

85 90 95

Ala Ile Thr Gln Glu Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro

100 105 110

Ala Lys Gln Gly Leu Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe

115 120 125

Phe Pro Pro Ser Tyr Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser

130 135 140

Glu Met Gly Tyr Thr Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe

145 150 155 160

Pro Lys Asp Ile Leu Thr Leu Arg Leu Asp Val Met Met Glu Thr Glu

165 170 175

Asn Arg Leu His Phe Thr Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu

180 185 190

Val Pro Leu Glu Thr Pro His Val His Ser Arg Ala Pro Ser Pro Leu

195 200 205

Tyr Ser Val Glu Phe Ser Glu Glu Pro Phe Gly Val Ile Val Arg Arg

210 215 220

Gln Leu Asp Gly Arg Val Leu Leu Asn Thr Thr Val Ala Pro Leu Phe

225 230 235 240

Phe Ala Asp Gln Phe Leu Gln Leu Ser Thr Ser Leu Pro Ser Gln Tyr

245 250 255

Ile Thr Gly Leu Ala Glu His Leu Ser Pro Leu Met Leu Ser Thr Ser

260 265 270

Trp Thr Arg Ile Thr Leu Trp Asn Arg Asp Leu Ala Pro Thr Pro Gly

275 280 285

Ala Asn Leu Tyr Gly Ser His Pro Phe Tyr Leu Ala Leu Glu Asp Gly

290 295 300

Gly Ser Ala His Gly Val Phe Leu Leu Asn Ser Asn Ala Met Asp Val

305 310 315 320

Val Leu Gln Pro Ser Pro Ala Leu Ser Trp Arg Ser Thr Gly Gly Ile

325 330 335

Leu Asp Val Tyr Ile Phe Leu Gly Pro Glu Pro Lys Ser Val Val Gln

340 345 350

Gln Tyr Leu Asp Val Val Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly

355 360 365

Leu Gly Phe His Leu Cys Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr

370 375 380

Arg Gln Val Val Glu Asn Met Thr Arg Ala His Phe Pro Leu Asp Val

385 390 395 400

Gln Trp Asn Asp Leu Asp Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe

405 410 415

Asn Lys Asp Gly Phe Arg Asp Phe Pro Ala Met Val Gln Glu Leu His

420 425 430
Gln Gly Gly Arg Arg Tyr Met Met Ile Val Asp Pro Ala Ile Ser Ser

435 440 445
Ser Gly Pro Ala Gly Ser Tyr Arg Pro Tyr Asp Glu Gly Leu Arg Arg

450 455 460

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

465 470 475 480

Trp Pro Gly Ser Thr Ala Phe Pro Asp Phe Thr Asn Pro Thr Ala Leu

485 490 495

Ala Trp Trp Glu Asp Met Val Ala Glu Phe His Asp Gln Val Pro Phe

500 505 510
Asp Gly Met Trp Ile Asp Met Asn Glu Pro Ser Asn Phe Ile Arg Gly

515 520 525

Ser Glu Asp Gly Cys Pro Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val

530 535 540

Pro Gly Val Val Gly Gly Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser

545 550 555 560

Ser His Gln Phe Leu Ser Thr His Tyr Asn Leu His Asn Leu Tyr Gly

565 570 575
Leu Thr Glu Ala Ile Ala Ser His Arg Ala Leu Val Lys Ala Arg Gly

580 585 590

Thr Arg Pro Phe Val Ile Ser Arg Ser Thr Phe Ala Gly His Gly Arg

595 600 605

Tyr Ala Gly His Trp Thr Gly Asp Val Trp Ser Ser Trp Glu Gln Leu

610 615 620

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

625 630 635 640

Leu Val Gly Ala Asp Val Cys Gly Phe Leu Gly Asn Thr Ser Glu Glu

645 650 655

Leu Cys Val Arg Trp Thr Gln Leu Gly Ala Phe Tyr Pro Phe Met Arg

660 665 670

Asn His Asn Ser Leu Leu Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser
 675 680 685
 Glu Pro Ala Gln Gln Ala Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala

 690 695 700
 Leu Leu Pro His Leu Tyr Thr Leu Phe His Gln Ala His Val Ala Gly
 705 710 715 720
 Glu Thr Val Ala Arg Pro Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser
 725 730 735
 Thr Trp Thr Val Asp His Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile
 740 745 750
 Thr Pro Val Leu Gln Ala Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro

 755 760 765
 Leu Gly Thr Trp Tyr Asp Leu Gln Thr Val Pro Ile Glu Ala Leu Gly
 770 775 780
 Ser Leu Pro Pro Pro Pro Ala Ala Pro Arg Glu Pro Ala Ile His Ser
 785 790 795 800
 Glu Gly Gln Trp Val Thr Leu Pro Ala Pro Leu Asp Thr Ile Asn Val
 805 810 815
 His Leu Arg Ala Gly Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr

 820 825 830
 Thr Thr Glu Ser Arg Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr
 835 840 845
 Lys Gly Gly Glu Ala Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser
 850 855 860
 Leu Glu Val Leu Glu Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala
 865 870 875 880
 Arg Asn Asn Thr Ile Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly

 885 890 895
 Ala Gly Leu Gln Leu Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala
 900 905 910
 Pro Gln Gln Val Leu Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr

915	920	925
Ser Pro Asp Thr Lys Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly		
930	935	940
Glu Gln Phe Leu Val Ser Trp Cys		
945	950	
<210> 4		
<211> 2856		
<212> DNA		
<213> Artificial sequence		
<220><223> Engineered hGAAI Coding sequence		
<400> 4		
atgggcgtta gacaccctcc ttgctctcac agactgctgg ccgtgtgtgc tctgggtgtct	60	
ctggctacag ctgccctgct gggacatatc ctgctgcacg actttctgct ggtgcccaga	120	
gagctgtctg gcagctctcc tgtgctggaa gagacacacc ctgcacatca gcagggcgcc	180	
tctagacctg gacctagaga cgctcaggcc catcctggca gacctagagc tgtgcccaca	240	
cagtgtgacg tgccacctaa cagcagattc gactgcgccc ctgacaaggc catcacacaa	300	
gagcagtgtg aagccagagg ctgctgctac atccctgcc aacaaggact gcagggcgct	360	
cagatgggac agccctggtg cttcttccca ccattctacc ccagctacaa gctggaaaac	420	
ctgagcagca gcgagatggg ctacaccgcc aactgacca gaaccacacc tacattcttc	480	
ccgaaggaca tctgacact gcggctggac gtgatgatgg aaaccgagaa ccggctgcac	540	
ttcaccatca aggacccgc caatcggaga tacgaggtgc cactggaaac ccctcacgtg	600	
cactctagag ccccatctcc actgtacagc gtggaatfff ccgaggaacc cttcggcgtg	660	
atcgtgcgga gacagctgga tggaagagtg ctgctgaaca ccacagtggc ccctctgttc	720	
ttcgccgacc agttttctga gctgagcacc agcctgccta gccagtatat cacaggcctg	780	
gccgagcacc tgtctccact gatgetgagc acatcctgga ccagaatcac cctgtggaac	840	
agagatctgg cccctacacc tggcgccaac ctgtacggct ctacccctt ttatctggcc	900	
ctggaagacg gcgcatctgc ccacggtgtt tttctgtga actccaacgc catggacgtg	960	
gtgctgcagc catctcctgc tctgtcttgg agaagcacag gcggcatcct ggacgtgtac	1020	
atctttctgg gccccgagcc taagagcgtg gtgcagcagt atctggacgt cgtgggctac	1080	
cccttcacgc ctctttatgg gggcctgggc ttccacctgt gcaggtgggg atacagcagc	1140	
accgcatca ccagacaggt ggtggaaaac atgaccggg ctcacttccc actggacgtg	1200	

cagtggaaacg acctggacta catggacagc agacgggact tcaccttcaa caaggacggc 1260
 ttcagagact tccccgccat ggtgcaagaa ctgcaccaag gcggcagacg gtacatgatg 1320
 attgtggacc cagccatcag ctctagcggc cctgccgga gctacagacc ttacgatgag 1380
 ggccctgagaa gaggcgtgtt catcaccaac gagacaggcc agcctctgat cggcaaagtg 1440
 tggcctggca gcacagcctt tccagacttc acaaacccca ccgctctggc ttggtgggaa 1500
 gatattggtg ccgagttcca cgatcagggt ccttcgatg gcatgtggat cgacatgaac 1560
 gagcccagca acttcatccg gggcagcgag gacggctgcc ccaacaacga actggaaaat 1620

cctccttacg tgcccgcgct tgtcggcgga acacttcagg ccgctacaat ctgtgccagc 1680
 agccaccagt tctcagcac ccactacaac ctgcacaacc tgtacggcct gaccgaggcc 1740
 attgcctctc atagagccct ggttaaggcc agaggcacc cggcctttgt gatcagcaga 1800
 agcacattcg ccggccacgg cagatagccc ggacattgga caggcgacgt gtggtctagt 1860
 tgggagcagc tggctagcag cgtgccagag atcctgcagt tcaatctgct gggcgtgcca 1920
 ctgctgggag ccgacgtttg tggcttctg ggcaacacca gcgaggaaact gtgtgtgcgt 1980
 tggacacagc tgggcgcctt ctatcccttc atgagaaacc acaacagcct gctgagcctg 2040

cctcaagagc cctacagctt tagcgagcct gcacagcagg ccatgagaaa ggccctgact 2100
 ctgagatacg ctctgtgcc ccacctgtac accctgtttc accaggtca cgtggccggg 2160
 gagacagtgg ctagacctt gtctctggaa tttcccaagg acagctccac ctggaccgtg 2220
 gatcatcagc tgctgtgggg agaagccctg ctcatcacac ctgttctgca ggccggaaag 2280
 gccgaagtga ccggtatatt tctctcggc acttggtacg acctgcagac cgtgcctatt 2340
 gaggccctgg gatctcttcc tccacctcct gctgtccta gagagcctgc catccactct 2400
 gaaggccagt gggttacact gcccgtcct ctggacacca tcaacgtgca cctgagagct 2460

ggctacatca tccactgca aggccttggc ctgaccacaa ccgaatctag acagcagccc 2520
 atggctctgg ccgtggctct taaaaaggc ggagaggcta gaggcgagct gttctgggac 2580
 gacggcgagt ctctggaagt gctggaacgg ggcgcttata ccaagtgat cttcctggcc 2640
 agaaacaaca ccatcgtgaa cgaactcgtg cgcgtgacca gtgaagggtg tggactgcaa 2700
 ctgcagaaag tgaccgtgct cggagtggcc acagctctc aacagggtgt gtctaacggc 2760
 gtgccgtgt ccaacttcac atacagcccc gacaccaagg tctggacat ctgtgtgtca 2820
 ctgctgatgg gcgagcagtt cctggtgtcc tgggtg 2856

<210> 5

<211> 2859

<212> DNA

<213> Homo sapiens

<400> 5

atgggagtga ggcacccgcc ctgctccac cggtcctgg ccgtctgcgc cctcgtgtcc	60
ttggcaaccg ctgcactcct ggggcacatc ctactccatg atttcctgct ggttccccga	120
gagctgagtg gtcctcccc agtcctggag gagactcacc cagctcacca gcagggagcc	180
agtagaccag gggcccggga tgcccaggca cccccggcc gtcccagagc agtgcccaca	240
cagtgcgacg tccccccaa cagccgttc gattgcgccc ctgacaaggc catcacccag	300
gaacagtgcg agggccgcgg ctgttgctac atccctgcaa agcaggggct gcagggagcc	360
cagatggggc agccctggtg cttcttccca cccagctacc ccagctacaa gctggagaac	420
ctgagctcct ctgaaatggg ctacacggcc accctgacct gtaccacccc caccttcttc	480
cccaaggaca tctgacctt gcggctggac gtgatgatgg agactgagaa ccgcctccac	540
ttcacgatca aagatccagc taacaggcgc tacgaggtgc ccttgagac cccgcatgtc	600
cacagccggg caccgtcccc actctacagc gtggagtctt ccgaggagcc cttcggggtg	660
atcgtgcgcc ggcagctgga cggccgcgtg ctgctgaaca cgacggtggc gcccctgttc	720
tttgcggacc agttccttca gctgtccacc tcgctgcctt cgcagtatat cacaggcctc	780
gccgagcacc tcagtcacct gatgtcagc accagctgga ccaggatcac cctgtggaac	840
cgggaccttg cggccacgcc cgggtgcgaac ctctacgggt ctacccttt ctacctggcg	900
ctggaggacg gggggtcgcc acacggggtg ttctgtctaa acagcaatgc catggatgtg	960
gtcttcagc cgagccctgc ccttagctgg aggtcgacag gtgggacctt ggatgtctac	1020
atcttcttgg gccagagcc caagagcgtg gtgcagcagt acctggacgt tgtgggatac	1080
ccgttatgc cgccatactg gggcctgggc ttccacctgt gccgtgggg ctactcctcc	1140
accgtatca cccgccaggt ggtggagaac atgaccaggg cccacttccc cctggacgtc	1200
cagtggaacg acctggacta catggactcc cggagggact tcacgttcaa caaggatggc	1260
ttccgggact tccggccat ggtgcaggag ctgcaccagg gcggccggcg ctacatgatg	1320
atcgtggatc ctgccatcag cagctcgggc cctgccggga gctacaggcc ctacgacgag	1380
ggtctgcgga ggggggtttt catcaccaac gagaccggcc agccgctgat tgggaaggta	1440
tggccccggg ccaactgcctt ccccgacttc accaacccca cagccctggc ctggtgggag	1500
gacatggtgg ctgagttcca tgaccagggt cccttcgacg gcatgtggat tgacatgaac	1560
gagccttcca acttcatcag gggctctgag gacggctgcc ccaacaatga gctggagaac	1620
ccacctacg tgcttgggtt ggttgggggg accctccagg cggccacat ctgtgcctcc	1680
agccaccagt ttctctccac acactacaac ctgcacaacc tctacggcct gaccgaagcc	1740

atgcctcccc acagggcgct ggtgaaggct cgggggacac gccatttgt gatctccgc 1800
 tcgacctttg ctggccacgg ccgatacgcc ggccactgga cgggggacgt gtggagctcc 1860
 tgggagcagc tcgcctcctc cgtgccagaa atcctgcagt ttaacctgct gggggtgcct 1920
 ctggtcgggg ccgacgtctg cggcttcctg ggcaacacct cagaggagct gtgtgtgcgc 1980
 tggaccacgc tgggggcctt ctacccttc atgcggaacc acaacagcct gctcagctcg 2040

ccccaggagc cgtacagctt cagcgagccg gccagcagg ccatgaggaa ggccctcacc 2100
 ctgcgtacg cactcctccc ccacctctac aactgttcc accaggccca cgtcgcgggg 2160
 gagaccgtgg cccggccctt ctctctggag ttccccaagg actctagcac ctggactgtg 2220
 gaccaccagc tctgtgggg ggaggccctg ctcataccc cagtgtcca ggccgggaag 2280
 gccgaagtga ctggctactt ccccttgggc acatggtacg acctgcagac ggtgccaata 2340
 gagggccctt gcagcctccc acccccacct gcagctcccc gtgagccagc catccacagc 2400
 gaggggcagt ggggtgacgt gccggccccc ctggacacca tcaacgtcca cctccgggct 2460

gggtacatca tccccctgca gggccctggc ctcaaacca cagagtcccg ccagcagccc 2520
 atggccctgg ctgtggccct gaccaagggt ggggaggccc gaggggagct tttctgggac 2580
 gatggagaga gcctggaagt gctggagcga ggggcctaca cacaggtcat ctctctggcc 2640
 aggaataaca cgatcgtgaa tgagctggta cgtgtgacca gtgaggagc tggcctgcag 2700
 ctgcagaagg tgactgtcct gggcgtggcc acggcgcccc agcaggtcct ctccaacggt 2760
 gtccctgtct ccaacttcac ctacagcccc gacaccaagg tcctggacat ctgtgtctcg 2820
 ctgttgatgg gagagcagtt tctcgtcagc tgggtttag 2859

<210> 6

<211> 982

<212> PRT

<213> Artificial sequence

<220><223> Fusion Protein comprising hGAA780I

<400> 6

Met Lys Leu Ser Leu Val Ala Ala Met Leu Leu Leu Ser Ala Ala

1 5 10 15

Arg Ala Ser Arg Thr Leu Cys Gly Gly Glu Leu Val Asp Thr Leu Gln

20 25 30

Phe Val Cys Gly Asp Arg Gly Phe Leu Phe Ser Arg Pro Ala Ser Arg

35 40 45

Val Ser Arg Arg Ser Arg Gly Ile Val Glu Glu Cys Cys Phe Arg Ser
50 55 60
Cys Asp Leu Ala Leu Leu Glu Thr Tyr Cys Ala Thr Pro Ala Arg Ser
65 70 75 80
Glu Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Arg Pro Gly Pro Arg
85 90 95
Asp Ala Gln Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys
100 105 110

Asp Val Pro Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile
115 120 125
Thr Gln Glu Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys
130 135 140
Gln Gly Leu Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro
145 150 155 160
Pro Ser Tyr Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met
165 170 175

Gly Tyr Thr Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys
180 185 190
Asp Ile Leu Thr Leu Arg Leu Asp Val Met Met Glu Thr Glu Asn Arg
195 200 205
Leu His Phe Thr Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu Val Pro
210 215 220
Leu Glu Thr Pro His Val His Ser Arg Ala Pro Ser Pro Leu Tyr Ser
225 230 235 240

Val Glu Phe Ser Glu Glu Pro Phe Gly Val Ile Val Arg Arg Gln Leu
245 250 255
Asp Gly Arg Val Leu Leu Asn Thr Thr Val Ala Pro Leu Phe Phe Ala
260 265 270
Asp Gln Phe Leu Gln Leu Ser Thr Ser Leu Pro Ser Gln Tyr Ile Thr
275 280 285
Gly Leu Ala Glu His Leu Ser Pro Leu Met Leu Ser Thr Ser Trp Thr

290					295					300									
Arg	Ile	Thr	Leu	Trp	Asn	Arg	Asp	Leu	Ala	Pro	Thr	Pro	Gly	Ala	Asn				
305					310					315					320				
Leu	Tyr	Gly	Ser	His	Pro	Phe	Tyr	Leu	Ala	Leu	Glu	Asp	Gly	Gly	Ser				
					325					330					335				
Ala	His	Gly	Val	Phe	Leu	Leu	Asn	Ser	Asn	Ala	Met	Asp	Val	Val	Leu				
					340					345					350				
Gln	Pro	Ser	Pro	Ala	Leu	Ser	Trp	Arg	Ser	Thr	Gly	Gly	Ile	Leu	Asp				
355					360					365									
Val	Tyr	Ile	Phe	Leu	Gly	Pro	Glu	Pro	Lys	Ser	Val	Val	Gln	Gln	Tyr				
370					375					380									
Leu	Asp	Val	Val	Gly	Tyr	Pro	Phe	Met	Pro	Pro	Tyr	Trp	Gly	Leu	Gly				
385					390					395					400				
Phe	His	Leu	Cys	Arg	Trp	Gly	Tyr	Ser	Ser	Thr	Ala	Ile	Thr	Arg	Gln				
					405					410					415				
Val	Val	Glu	Asn	Met	Thr	Arg	Ala	His	Phe	Pro	Leu	Asp	Val	Gln	Trp				
420					425					430									
Asn	Asp	Leu	Asp	Tyr	Met	Asp	Ser	Arg	Arg	Asp	Phe	Thr	Phe	Asn	Lys				
435					440					445									
Asp	Gly	Phe	Arg	Asp	Phe	Pro	Ala	Met	Val	Gln	Glu	Leu	His	Gln	Gly				
450					455					460									
Gly	Arg	Arg	Tyr	Met	Met	Ile	Val	Asp	Pro	Ala	Ile	Ser	Ser	Ser	Gly				
465					470					475					480				
Pro	Ala	Gly	Ser	Tyr	Arg	Pro	Tyr	Asp	Glu	Gly	Leu	Arg	Arg	Gly	Val				
					485					490					495				
Phe	Ile	Thr	Asn	Glu	Thr	Gly	Gln	Pro	Leu	Ile	Gly	Lys	Val	Trp	Pro				
500					505					510									
Gly	Ser	Thr	Ala	Phe	Pro	Asp	Phe	Thr	Asn	Pro	Thr	Ala	Leu	Ala	Trp				
515					520					525									
Trp	Glu	Asp	Met	Val	Ala	Glu	Phe	His	Asp	Gln	Val	Pro	Phe	Asp	Gly				
530					535					540									

Met Trp Ile Asp Met Asn Glu Pro Ser Asn Phe Ile Arg Gly Ser Glu
545 550 555 560

Asp Gly Cys Pro Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val Pro Gly
565 570 575

Val Val Gly Gly Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser Ser His
580 585 590

Gln Phe Leu Ser Thr His Tyr Asn Leu His Asn Leu Tyr Gly Leu Thr
595 600 605

Glu Ala Ile Ala Ser His Arg Ala Leu Val Lys Ala Arg Gly Thr Arg
610 615 620

Pro Phe Val Ile Ser Arg Ser Thr Phe Ala Gly His Gly Arg Tyr Ala
625 630 635 640

Gly His Trp Thr Gly Asp Val Trp Ser Ser Trp Glu Gln Leu Ala Ser
645 650 655

Ser Val Pro Glu Ile Leu Gln Phe Asn Leu Leu Gly Val Pro Leu Val
660 665 670

Gly Ala Asp Val Cys Gly Phe Leu Gly Asn Thr Ser Glu Glu Leu Cys
675 680 685

Val Arg Trp Thr Gln Leu Gly Ala Phe Tyr Pro Phe Met Arg Asn His
690 695 700

Asn Ser Leu Leu Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser Glu Pro
705 710 715 720

Ala Gln Gln Ala Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala Leu Leu
725 730 735

Pro His Leu Tyr Thr Leu Phe His Gln Ala His Val Ala Gly Glu Thr
740 745 750

Val Ala Arg Pro Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser Thr Trp
755 760 765

Thr Val Asp His Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile Thr Pro
770 775 780

Val Leu Gln Ala Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro Leu Gly

785 790 795 800
 Thr Trp Tyr Asp Leu Gln Thr Val Pro Ile Glu Ala Leu Gly Ser Leu
 805 810 815

Pro Pro Pro Pro Ala Ala Pro Arg Glu Pro Ala Ile His Ser Glu Gly
 820 825 830
 Gln Trp Val Thr Leu Pro Ala Pro Leu Asp Thr Ile Asn Val His Leu
 835 840 845
 Arg Ala Gly Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr
 850 855 860
 Glu Ser Arg Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly
 865 870 875 880

Gly Glu Ala Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu
 885 890 895
 Val Leu Glu Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn
 900 905 910
 Asn Thr Ile Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly
 915 920 925
 Leu Gln Leu Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln
 930 935 940

Gln Val Leu Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro
 945 950 955 960
 Asp Thr Lys Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln
 965 970 975
 Phe Leu Val Ser Trp Cys
 980

<210> 7

<211> 2952

<212> DNA

<213> Artificial sequence

<220><223> Engineered sequence encoding fusion protein comprising GAAV780I

<220><221> misc_feature

<222> (810)..(810)

<

223> V810I

<400> 7

atgaagctgt ctctggtggc tgctatgctg ctgctcctgt ctgccgccag agccagcaga	60
acactttgtg gcggagagct ggtggacacc ctgcagtttg tgtgtggcga cagaggcttc	120
ctgttcagca gacctgccag ccgggtttcc agacggctta gaggaatcgt ggaagagtgc	180
tgcttcagaa gctgcgatct ggccctgctg gaaacctact gtgccacacc agccagatct	240
gaaggcggcg gaggatctgg cggaggcgga tctagacctg gacctagaga cggccaggct	300
caccttgga gacctagagc tgtgcctaca cagtgcgacg tgccacctaa cagcagattc	360
gactgcgccc ctgacaaggc catcacacaa gagcagtgtg aagccagagg ctgctgctac	420
atccctgcc aacaaggact gcagggcgcc cagatgggac agccttggtg cttcttccca	480
ccatcttacc ccagctacaa gctgaaaaac ctgagcagct ccgagatggg ctacaccgcc	540
acactgacca gaaccacacc tacattcttc ccgaaggaca tctgacact gcggctggac	600
gtgatgatgg aaaccgagaa ccggctgcac ttaccatca aggacccgc caatcggaga	660
tacgaggtag ccttgaaac accccactg cactctagag cacctctcc actgtacagc	720
gtggaattti ccgaggaacc cttcgcgctg atcgtgcgga gacagctgga tggcagagt	780
ctctgaata ccacagtggc cctctgttc ttcgccgacc agtttctgca gctgagcacc	840
agcctgccta gccagtatat cacaggcctg gccgagcatc tgagccctct gatgctgagc	900
acatcctgga ccagaatcac cctgtggaac cgcgacctgg ctctacacc tggcgccaat	960
ctgtacggct ctacccctt ctacctggca ctggaagacg gtggatctgc ccacggtgtc	1020
tttctgctga atagcaacgc catggacgtg gtgctgcagc cctctcctgc actgtcttgg	1080
agatctacag gcggcatcct ggacgtgtac atctttctgg gccccgagcc taagagcgtg	1140
gtgcagcagt atctggacgt cgtgggctac ccttcatgc ctcttattg gggcctgggc	1200
ttccacctgt gtaggtgggg ctacagcagc accgcatca ccagacaggt ggtggaaaac	1260
atgaccggg ctcaattccc actggacgtg cagtggaaac acctggacta catggacagc	1320
agacgggact tcacattcaa caaggacggc ttcagagact tccccgcat ggtgcaagag	1380
ctgcatcaag gcggacggcg gtacatgatg attgtggacc ctgccatcag cagctctgga	1440
ccagccggca gctacagacc ttacgatgag ggactgagaa gaggcgtgtt catcaccaac	1500
gagacaggcc agcctctgat cggcaaagtg tggcctggca gcacagcctt tccagacttc	1560
acaaacccca ccgctctggc ttggtgggaa gatatggtgg ccgagttcca cgatcaggtg	1620

cccttcgatg gcatgtggat cgacatgaac gagcccagca attcatccg gggcagcgag 1680

gatggctgcc ccaacaacga actagaaaat cctccttacg tgcccggcgt tgtcggcgga 1740

acacttcagg ccgtacaat ctgtgccagc agccatcagt ttctgagcac ccactacaac 1800

ctgcacaacc tgiacggcct gaccgaggcc attgcctctc atagagccct ggttaaggcc 1860

agaggcaccg ggccttttgt gatcagcaga agcacattcg ccggccacgg cagatacgca 1920

ggacattgga caggcgacgt gtggtctagt tgggagcagc tggctagcag cgtgccagag 1980

atcctgcagt tcaatctgct gggcgtgcc a ctcgtgggag ccgacgtttg tggcttctg 2040

ggcaacacca gcgaggaact gtgtgtgctg tggacacagc tgggcgcctt ctatcccttc 2100

atgagaaacc acaacagcct gctgagccct cctcaagagc cctacagctt tagcgagcct 2160

gcacagcagg ccatgagaaa ggccctgact ctgagatagc ccctgctgcc tcacctgtac 2220

accctgtttc atcaggccca cgtggcaggc gagacagtgg ctagacctct gttcctggaa 2280

tttccaagg acagctccac ctggaccgtg gatcatcagc tgctgtgggg agaagccctg 2340

ctgattacac cagtgtgca ggccggaaaag gccgaagtga caggctatct ccctctcggc 2400

acttggtagc acctgcagac cgtgcctatc gaggtcttgg gatctcttcc tccacctcct 2460

gccgctccta gagagcctgc cattactct gaaggccagt gggttaccct gcctgtcct 2520

ctggacacca tcaacgtgca cctgagagcc ggctacatca tccctctgca aggccctggc 2580

ctgaccacaa ccgaatctag acagcagccc atggcactgg ccgtggctct taaaaaggc 2640

ggagaggcta gaggcgagct gttctgggat gatggcgaga gcctagaagt gctggaacgg 2700

ggcgcttata cccaagtgat ctctctggcc agaacaaca ccatcgtgaa cgaactcgtg 2760

cgcgtagaca gtgaaggctg tggactgcaa ctgcagaaag tgaccgtgct cggagtggcc 2820

acagctcttc agcaggttct gtctaattgc gtccccgtgt ccaacttcac atacagcccc 2880

gacaccaagg tcttgacat ctgtgtgtcc ctgcttatgg gcgagcagtt cctggtgtcc 2940

tggtgctgat aa 2952

<210> 8

<211> 934

<212> DNA

<213> Artificial sequence

<220><223> CAG promoter

<220><221> misc_feature

<222> (1)..(243)

<223> CMV early enhancer element

<220><221> misc_feature

<222> (244)..(525)

<223> Chicken Beta actin promoter

<220><221> misc_feature

<222> (526)..(934)

<223> hybrid intron in CAG (b-actin intron and rabbit beta 1-globin

<220><221> misc_feature

<222> (526)..(934)

<223>

hybrid intron

<400> 8

attgacgtca ataatacgt atgttcccat agtaacgcca atagggactt tccattgacg	60
tcaatgggtg gagtatttac ggtaaaactgc ccacttggca gtacatcaag tgtatcatat	120
gccaaagtacg cccctatttg acgtcaatga cggtaaatgg cccgcctggc attatgccca	180
gtacatgacc ttatgggact ttctacttgg gcagtacatc tacgtattag tcatacgtat	240
taccatgggtc gaggtgagcc ccacgttctg cttcactctc cccatctccc cccctcccc	300
acccccaatt ttgtatttat ttatttttta attattttgt gcagcgatgg gggcgggggg	360
gggggggggg cgcgcgccag gcggggcggg gcggggcgag gggcggggag gggcgaggcg	420
gagaggtgcg gcggcagcca atcagagcgg cgcgctccga aagtttcctt ttatggcgag	480
gcggcgggcg cgcgggccct ataaaaagcg aagcgcgcg cgggcgggag tcgctgcgcg	540
ctgccttcgc cccgtgcccc gctccgccc gcgctcgcg cgcccgcccc ggctctgact	600
gaccgcgtta ctccacagg tgagcgggcg ggacggccct tctcctcgg gctgtaatta	660
gcgcttggtt taatgacggc ttgtttcttt tctgtggctg cgtgaaagcc ttgaggggct	720
ccgggagggc ctttgtgcg gggggagcgg ctcggggctg tccgcggggg gacggctgcc	780
ttcggggggg acggggcagg gcgggggtcg gcttctggcg tgtgaccggc ggctctagag	840
cctctgctaa ccatgttcat gccttcttct ttttctaca gctcctgggc aacgtgctgg	900
ttattgtgct gtctcatcat tttggcaaag aatt	934

<210> 9

<211> 127

<212> DNA

<213> Artificial sequence

<220><223> Rabbit globin polyA

<400> 9

gatctttttc cctctgccaa aaattatggg gacatcatga agcccttga gcatctgact 60
tctggctaataaaggaatttattttcattgcaatagtgtgttggaattttttgtgtctc 120
tcactcg 127
<210> 10
<211>
188
<212> DNA
<213> adeno-associated virus 2
<400> 10
gtgctgcgctcgtctgctcactgaggccgc ccgggcaaagcccggcgctcggcgacctt 60
tggctgccccggcctcagtgagcgagcgcgcagagaggagtggccactccatcac 120
taggggttcc tttagttaa tgattaaccgccatgctac ttatctacgt agccatgctc 180
taggaaga 188
<210> 11
<211> 188
<212> DNA
<213> adeno-associated virus 2
<400> 11
tcttcctaga gcatggctac gtagataagt agcatggcgggttaatcatt aactacaagg 60
aaccctagt gatggagtggccactccctctctgcgcgc tgcctcgtc actgaggccg 120
ggcgacaaaa ggtcggccga cgccggggt ttgccgggc ggcctcagtg agcgagcgag 180
cgcgcagc 188
<210> 12
<211> 81
<212> DNA
<213> Artificial sequence
<220><223> Engineered hGAAV780I signal peptide
<220><221> sig_peptide
<222> (1)..(81)
<220><221> CDS
<222> (1)..(81)
<400> 12
atg ggc gtt aga cac cct cct tgc tct cac aga ctg ctg gcc gtg tgt 48

Met Gly Val Arg His Pro Pro Cys Ser His Arg Leu Leu Ala Val Cys

1 5 10 15

gct ctg gtg tct ctg gct aca gct gcc ctg ctg 81

Ala Leu Val Ser Leu Ala Thr Ala Ala Leu Leu

20 25

<210> 13

<211> 27

<212> PRT

<213> Artificial sequence

<220><223> Synthetic Construct

<400> 13

Met Gly Val Arg His Pro Pro Cys Ser His Arg Leu Leu Ala Val Cys

1 5 10 15

Ala Leu Val Ser Leu Ala Thr Ala Ala Leu Leu

20 25

<210> 14

<211> 2649

<212> DNA

<213> Artificial Sequence

<220><223> engineered hGAHV780I mature protein

<220><221> CDS

<222> (1)..(2649)

<400> 14

gcc cat cct ggc aga cct aga gct gtg ccc aca cag tgt gac gtg cca 48

Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys Asp Val Pro

1 5 10 15

cct aac agc aga ttc gac tgc gcc cct gac aag gcc atc aca caa gag 96

Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile Thr Gln Glu

20 25 30

cag tgt gaa gcc aga ggc tgc tgc tac atc cct gcc aaa caa gga ctg 144

Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys Gln Gly Leu

35 40 45

cag ggc gct cag atg gga cag ccc tgg tgc ttc ttc cca cca tct tac 192

Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro Pro Ser Tyr
50 55 60

ccc agc tac aag ctg gaa aac ctg agc agc agc gag atg ggc tac acc 240
Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met Gly Tyr Thr
65 70 75 80

gcc aca ctg acc aga acc aca cct aca ttc ttc ccg aag gac atc ctg 288
Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys Asp Ile Leu
85 90 95

aca ctg cgg ctg gac gtg atg atg gaa acc gag aac cgg ctg cac ttc 336
Thr Leu Arg Leu Asp Val Met Met Glu Thr Glu Asn Arg Leu His Phe
100 105 110

acc atc aag gac ccc gcc aat cgg aga tac gag gtg cca ctg gaa acc 384
Thr Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu Val Pro Leu Glu Thr
115 120 125

cct cac gtg cac tct aga gcc cca tct cca ctg tac agc gtg gaa ttt 432
Pro His Val His Ser Arg Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe
130 135 140

tcc gag gaa ccc ttc ggc gtg atc gtg cgg aga cag ctg gat gga aga 480
Ser Glu Glu Pro Phe Gly Val Ile Val Arg Arg Gln Leu Asp Gly Arg
145 150 155 160

gtg ctg ctg aac acc aca gtg gcc cct ctg ttc ttc gcc gac cag ttt 528

Val Leu Leu Asn Thr Thr Val Ala Pro Leu Phe Phe Ala Asp Gln Phe
165 170 175

ctg cag ctg agc acc agc ctg cct agc cag tat atc aca ggc ctg gcc 576
Leu Gln Leu Ser Thr Ser Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala
180 185 190

gag cac ctg tct cca ctg atg ctg agc aca tcc tgg acc aga atc acc 624
Glu His Leu Ser Pro Leu Met Leu Ser Thr Ser Trp Thr Arg Ile Thr

195	200	205	
ctg tgg aac aga gat ctg gcc cct aca cct ggc gcc aac ctg tac ggc			672
Leu Trp Asn Arg Asp Leu Ala Pro Thr Pro Gly Ala Asn Leu Tyr Gly			
210	215	220	
tct cac ccc ttt tat ctg gcc ctg gaa gac ggc gga tct gcc cac ggt			720
Ser His Pro Phe Tyr Leu Ala Leu Glu Asp Gly Gly Ser Ala His Gly			
225	230	235	240
gtt ttt ctg ctg aac tcc aac gcc atg gac gtg gtg ctg cag cca tct			768
Val Phe Leu Leu Asn Ser Asn Ala Met Asp Val Val Leu Gln Pro Ser			
245	250	255	
cct gct ctg tct tgg aga agc aca ggc ggc atc ctg gac gtg tac atc			816
Pro Ala Leu Ser Trp Arg Ser Thr Gly Gly Ile Leu Asp Val Tyr Ile			
260	265	270	
ttt ctg ggc ccc gag cct aag agc gtg gtg cag cag tat ctg gac gtc			864
Phe Leu Gly Pro Glu Pro Lys Ser Val Val Gln Gln Tyr Leu Asp Val			
275	280	285	
gtg ggc tac ccc ttc atg cct cct tat tgg ggc ctg ggc ttc cac ctg			912
Val Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly Leu Gly Phe His Leu			
290	295	300	
tgc agg tgg gga tac agc agc acc gcc atc acc aga cag gtg gtg gaa			960
Cys Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr Arg Gln Val Val Glu			
305	310	315	320
aac atg acc cgg gct cac ttc cca ctg gac gtg cag tgg aac gac ctg			1008
Asn Met Thr Arg Ala His Phe Pro Leu Asp Val Gln Trp Asn Asp Leu			
325	330	335	
gac tac atg gac agc aga cgg gac ttc acc ttc aac aag gac ggc ttc			1056
Asp Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe			
340	345	350	
aga gac ttc ccc gcc atg gtg caa gaa ctg cac caa ggc ggc aga cgg			1104
Arg Asp Phe Pro Ala Met Val Gln Glu Leu His Gln Gly Gly Arg Arg			
355	360	365	

tac atg atg att gtg gac cca gcc atc agc tct agc ggc cct gcc gga	1152
Tyr Met Met Ile Val Asp Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly	
370 375 380	
agc tac aga cct tac gat gag ggc ctg aga aga ggc gtg ttc atc acc	1200
Ser Tyr Arg Pro Tyr Asp Glu Gly Leu Arg Arg Gly Val Phe Ile Thr	
385 390 395 400	
aac gag aca ggc cag cct ctg atc ggc aaa gtg tgg cct ggc agc aca	1248
Asn Glu Thr Gly Gln Pro Leu Ile Gly Lys Val Trp Pro Gly Ser Thr	
405 410 415	
gcc ttt cca gac ttc aca aac ccc acc gct ctg gct tgg tgg gaa gat	1296
Ala Phe Pro Asp Phe Thr Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp	
420 425 430	
atg gtg gcc gag ttc cac gat cag gtg ccc ttc gat ggc atg tgg atc	1344
Met Val Ala Glu Phe His Asp Gln Val Pro Phe Asp Gly Met Trp Ile	
435 440 445	
gac atg aac gag ccc agc aac ttc atc cgg ggc agc gag gac ggc tgc	1392
Asp Met Asn Glu Pro Ser Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys	
450 455 460	
ccc aac aac gaa ctg gaa aat cct cct tac gtg ccc ggc gtt gtc ggc	1440
Pro Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val Pro Gly Val Val Gly	
465 470 475 480	
gga aca ctt cag gcc gct aca atc tgt gcc agc agc cac cag ttc etc	1488
Gly Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser Ser His Gln Phe Leu	
485 490 495	
agc acc cac tac aac ctg cac aac ctg tac ggc ctg acc gag gcc att	1536
Ser Thr His Tyr Asn Leu His Asn Leu Tyr Gly Leu Thr Glu Ala Ile	
500 505 510	
gcc tct cat aga gcc ctg gtt aag gcc aga ggc acc cgg cct ttt gtg	1584
Ala Ser His Arg Ala Leu Val Lys Ala Arg Gly Thr Arg Pro Phe Val	
515 520 525	
atc agc aga agc aca ttc gcc ggc cac ggc aga tac gcc gga cat tgg	1632

Ile Ser Arg Ser Thr Phe Ala Gly His Gly Arg Tyr Ala Gly His Trp

530	535	540	
aca ggc gac gtg tgg tct agt tgg gag cag ctg gct agc agc gtg cca			1680
Thr Gly Asp Val Trp Ser Ser Trp Glu Gln Leu Ala Ser Ser Val Pro			
545	550	555	560
gag atc ctg cag ttc aat ctg ctg ggc gtg cca ctc gtg gga gcc gac			1728
Glu Ile Leu Gln Phe Asn Leu Leu Gly Val Pro Leu Val Gly Ala Asp			
565	570	575	
gtt tgt ggc ttc ctg ggc aac acc agc gag gaa ctg tgt gtg cgt tgg			1776
Val Cys Gly Phe Leu Gly Asn Thr Ser Glu Glu Leu Cys Val Arg Trp			
580	585	590	
aca cag ctg ggc gcc ttc tat ccc ttc atg aga aac cac aac agc ctg			1824
Thr Gln Leu Gly Ala Phe Tyr Pro Phe Met Arg Asn His Asn Ser Leu			
595	600	605	
ctg agc ctg cct caa gag ccc tac agc ttt agc gag cct gca cag cag			1872
Leu Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln			
610	615	620	
gcc atg aga aag gcc ctg act ctg aga tac gct ctg ctg ccc cac ctg			1920
Ala Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala Leu Leu Pro His Leu			
625	630	635	640
tac acc ctg ttt cac cag gct cac gtg gcc ggg gag aca gtg gct aga			1968
Tyr Thr Leu Phe His Gln Ala His Val Ala Gly Glu Thr Val Ala Arg			
645	650	655	
cct ctg ttc ctg gaa ttt ccc aag gac agc tcc acc tgg acc gtg gat			2016
Pro Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser Thr Trp Thr Val Asp			
660	665	670	
cat cag ctg ctg tgg gga gaa gcc ctg ctc atc aca cct gtt ctg cag			2064
His Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile Thr Pro Val Leu Gln			
675	680	685	

gcc gga aag gcc gaa gtg acc ggc tat ttt cct ctc ggc act tgg tac	2112
Ala Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr	
690 695 700	
gac ctg cag acc gtg cct att gag gcc ctg gga tct ctt cct cca cct	2160
Asp Leu Gln Thr Val Pro Ile Glu Ala Leu Gly Ser Leu Pro Pro Pro	
705 710 715 720	
cct gct gct cct aga gag cct gcc atc cac tct gaa ggc cag tgg gtt	2208
Pro Ala Ala Pro Arg Glu Pro Ala Ile His Ser Glu Gly Gln Trp Val	
725 730 735	
aca ctg ccc gct cct ctg gac acc atc aac gtg cac ctg aga gct ggc	2256
Thr Leu Pro Ala Pro Leu Asp Thr Ile Asn Val His Leu Arg Ala Gly	
740 745 750	
tac atc atc cca ctg caa ggc cct ggc ctg acc aca acc gaa tct aga	2304
Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg	
755 760 765	
cag cag ccc atg gct ctg gcc gtg gct ctt aca aaa ggc gga gag gct	2352
Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly Gly Glu Ala	
770 775 780	
aga ggc gag ctg ttc tgg gac gac ggc gag tct ctg gaa gtg ctg gaa	2400
Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu Val Leu Glu	
785 790 795 800	
cgg ggc gct tat acc caa gtg atc ttc ctg gcc aga aac aac acc atc	2448
Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn Asn Thr Ile	
805 810 815	
gtg aac gaa ctc gtg cgc gtg acc agt gaa ggt gct gga ctg caa ctg	2496
Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly Leu Gln Leu	
820 825 830	
cag aaa gtg acc gtg ctc gga gtg gcc aca gct cct caa cag gtg ctg	2544
Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln Gln Val Leu	
835 840 845	
tct aac ggc gtg ccc gtg tcc aac ttc aca tac agc ccc gac acc aag	2592

Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro Asp Thr Lys
 850 855 860
 gtc ctg gac atc tgt gtg tca ctg ctg atg ggc gag cag ttc ctg gtg 2640
 Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln Phe Leu Val

865 870 875 880
 tcc tgg tgc 2649
 Ser Trp Cys

<210> 15

<211> 883

<212> PRT

<213> Artificial Sequence

<220><223> Synthetic Construct

<400> 15

Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys Asp Val Pro
 1 5 10 15

Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile Thr Gln Glu
 20 25 30
 Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys Gln Gly Leu
 35 40 45
 Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro Pro Ser Tyr
 50 55 60
 Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met Gly Tyr Thr
 65 70 75 80

Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys Asp Ile Leu
 85 90 95
 Thr Leu Arg Leu Asp Val Met Met Glu Thr Glu Asn Arg Leu His Phe
 100 105 110
 Thr Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu Val Pro Leu Glu Thr
 115 120 125
 Pro His Val His Ser Arg Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe

130	135	140	
Ser Glu Glu Pro Phe Gly Val Ile Val Arg Arg Gln Leu Asp Gly Arg			
145	150	155	160
Val Leu Leu Asn Thr Thr Val Ala Pro Leu Phe Phe Ala Asp Gln Phe			
	165	170	175
Leu Gln Leu Ser Thr Ser Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala			
	180	185	190
Glu His Leu Ser Pro Leu Met Leu Ser Thr Ser Trp Thr Arg Ile Thr			
195	200	205	
Leu Trp Asn Arg Asp Leu Ala Pro Thr Pro Gly Ala Asn Leu Tyr Gly			
210	215	220	
Ser His Pro Phe Tyr Leu Ala Leu Glu Asp Gly Gly Ser Ala His Gly			
225	230	235	240
Val Phe Leu Leu Asn Ser Asn Ala Met Asp Val Val Leu Gln Pro Ser			
	245	250	255
Pro Ala Leu Ser Trp Arg Ser Thr Gly Gly Ile Leu Asp Val Tyr Ile			
260	265	270	
Phe Leu Gly Pro Glu Pro Lys Ser Val Val Gln Gln Tyr Leu Asp Val			
275	280	285	
Val Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly Leu Gly Phe His Leu			
290	295	300	
Cys Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr Arg Gln Val Val Glu			
305	310	315	320
Asn Met Thr Arg Ala His Phe Pro Leu Asp Val Gln Trp Asn Asp Leu			
	325	330	335
Asp Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe			
340	345	350	
Arg Asp Phe Pro Ala Met Val Gln Glu Leu His Gln Gly Gly Arg Arg			
355	360	365	
Tyr Met Met Ile Val Asp Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly			
370	375	380	

Ser Tyr Arg Pro Tyr Asp Glu Gly Leu Arg Arg Gly Val Phe Ile Thr
385 390 395 400

Asn Glu Thr Gly Gln Pro Leu Ile Gly Lys Val Trp Pro Gly Ser Thr
405 410 415

Ala Phe Pro Asp Phe Thr Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp
420 425 430

Met Val Ala Glu Phe His Asp Gln Val Pro Phe Asp Gly Met Trp Ile
435 440 445

Asp Met Asn Glu Pro Ser Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys
450 455 460

Pro Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val Pro Gly Val Val Gly
465 470 475 480

Gly Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser Ser His Gln Phe Leu
485 490 495

Ser Thr His Tyr Asn Leu His Asn Leu Tyr Gly Leu Thr Glu Ala Ile
500 505 510

Ala Ser His Arg Ala Leu Val Lys Ala Arg Gly Thr Arg Pro Phe Val
515 520 525

Ile Ser Arg Ser Thr Phe Ala Gly His Gly Arg Tyr Ala Gly His Trp
530 535 540

Thr Gly Asp Val Trp Ser Ser Trp Glu Gln Leu Ala Ser Ser Val Pro
545 550 555 560

Glu Ile Leu Gln Phe Asn Leu Leu Gly Val Pro Leu Val Gly Ala Asp
565 570 575

Val Cys Gly Phe Leu Gly Asn Thr Ser Glu Glu Leu Cys Val Arg Trp
580 585 590

Thr Gln Leu Gly Ala Phe Tyr Pro Phe Met Arg Asn His Asn Ser Leu
595 600 605

Leu Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln
610 615 620

Ala Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala Leu Leu Pro His Leu

625 630 635 640
 Tyr Thr Leu Phe His Gln Ala His Val Ala Gly Glu Thr Val Ala Arg
 645 650 655

 Pro Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser Thr Trp Thr Val Asp
 660 665 670
 His Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile Thr Pro Val Leu Gln
 675 680 685
 Ala Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr
 690 695 700
 Asp Leu Gln Thr Val Pro Ile Glu Ala Leu Gly Ser Leu Pro Pro Pro
 705 710 715 720

 Pro Ala Ala Pro Arg Glu Pro Ala Ile His Ser Glu Gly Gln Trp Val
 725 730 735
 Thr Leu Pro Ala Pro Leu Asp Thr Ile Asn Val His Leu Arg Ala Gly
 740 745 750
 Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg
 755 760 765
 Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly Gly Glu Ala
 770 775 780

 Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu Val Leu Glu
 785 790 795 800
 Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn Asn Thr Ile
 805 810 815
 Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly Leu Gln Leu
 820 825 830
 Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln Gln Val Leu
 835 840 845

 Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro Asp Thr Lys
 850 855 860
 Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln Phe Leu Val
 865 870 875 880

Ser Trp Cys

<210> 16

<211> 2304

<212> DNA

<213> Artificial sequence

<220><223> Engineered DNA for hGAA780I 123-890

<220><221> CDS

<222> (1)..(2304)

<400> 16

gga cag ccc tgg tgc ttc ttc cca cca tct tac ccc agc tac aag ctg 48

Gly Gln Pro Trp Cys Phe Phe Pro Pro Ser Tyr Pro Ser Tyr Lys Leu

1 5 10 15

gaa aac ctg agc agc agc gag atg ggc tac acc gcc aca ctg acc aga 96

Glu Asn Leu Ser Ser Ser Glu Met Gly Tyr Thr Ala Thr Leu Thr Arg

20 25 30

acc aca cct aca ttc ttc ccg aag gac atc ctg aca ctg cgg ctg gac 144

Thr Thr Pro Thr Phe Phe Pro Lys Asp Ile Leu Thr Leu Arg Leu Asp

35 40 45

gtg atg atg gaa acc gag aac cgg ctg cac ttc acc atc aag gac ccc 192

Val Met Met Glu Thr Glu Asn Arg Leu His Phe Thr Ile Lys Asp Pro

50 55 60

gcc aat cgg aga tac gag gtg cca ctg gaa acc cct cac gtg cac tct 240

Ala Asn Arg Arg Tyr Glu Val Pro Leu Glu Thr Pro His Val His Ser

65 70 75 80

aga gcc cca tct cca ctg tac agc gtg gaa ttt tcc gag gaa ccc ttc 288

Arg Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe Ser Glu Glu Pro Phe

85 90 95

ggc gtg atc gtg cgg aga cag ctg gat gga aga gtg ctg ctg aac acc 336

Gly Val Ile Val Arg Arg Gln Leu Asp Gly Arg Val Leu Leu Asn Thr

100 105 110

aca gtg gcc cct ctg ttc ttc gcc gac cag ttt ctg cag ctg agc acc 384

Thr Val Ala Pro Leu Phe Phe Ala Asp Gln Phe Leu Gln Leu Ser Thr
115 120 125

agc ctg cct agc cag tat atc aca ggc ctg gcc gag cac ctg tct cca 432

Ser Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala Glu His Leu Ser Pro
130 135 140

ctg atg ctg agc aca tcc tgg acc aga atc acc ctg tgg aac aga gat 480

Leu Met Leu Ser Thr Ser Trp Thr Arg Ile Thr Leu Trp Asn Arg Asp

145 150 155 160

ctg gcc cct aca cct ggc gcc aac ctg tac ggc tct cac ccc ttt tat 528

Leu Ala Pro Thr Pro Gly Ala Asn Leu Tyr Gly Ser His Pro Phe Tyr
165 170 175

ctg gcc ctg gaa gac ggc gga tct gcc cac ggt gtt ttt ctg ctg aac 576

Leu Ala Leu Glu Asp Gly Gly Ser Ala His Gly Val Phe Leu Leu Asn
180 185 190

tcc aac gcc atg gac gtg gtg ctg cag cca tct cct gct ctg tct tgg 624

Ser Asn Ala Met Asp Val Val Leu Gln Pro Ser Pro Ala Leu Ser Trp
195 200 205

aga agc aca ggc ggc atc ctg gac gtg tac atc ttt ctg ggc ccc gag 672

Arg Ser Thr Gly Gly Ile Leu Asp Val Tyr Ile Phe Leu Gly Pro Glu
210 215 220

cct aag agc gtg gtg cag cag tat ctg gac gtc gtg ggc tac ccc ttc 720

Pro Lys Ser Val Val Gln Gln Tyr Leu Asp Val Val Gly Tyr Pro Phe
225 230 235 240

atg cct cct tat tgg ggc ctg ggc ttc cac ctg tgc agg tgg gga tac 768

Met Pro Pro Tyr Trp Gly Leu Gly Phe His Leu Cys Arg Trp Gly Tyr
245 250 255

agc agc acc gcc atc acc aga cag gtg gtg gaa aac atg acc cgg gct 816

Ser Ser Thr Ala Ile Thr Arg Gln Val Val Glu Asn Met Thr Arg Ala

260	265	270	
cac ttc cca ctg gac gtg cag tgg aac gac ctg gac tac atg gac agc			864
His Phe Pro Leu Asp Val Gln Trp Asn Asp Leu Asp Tyr Met Asp Ser			
275	280	285	
aga cgg gac ttc acc ttc aac aag gac ggc ttc aga gac ttc ccc gcc			912
Arg Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe Arg Asp Phe Pro Ala			
290	295	300	
atg gtg caa gaa ctg cac caa ggc ggc aga cgg tac atg atg att gtg			960
Met Val Gln Glu Leu His Gln Gly Gly Arg Arg Tyr Met Met Ile Val			
305	310	315	320
gac cca gcc atc agc tct agc ggc cct gcc gga agc tac aga cct tac			1008
Asp Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly Ser Tyr Arg Pro Tyr			
325	330	335	
gat gag ggc ctg aga aga ggc gtg ttc atc acc aac gag aca ggc cag			1056
Asp Glu Gly Leu Arg Arg Gly Val Phe Ile Thr Asn Glu Thr Gly Gln			
340	345	350	
cct ctg atc ggc aaa gtg tgg cct ggc agc aca gcc ttt cca gac ttc			1104
Pro Leu Ile Gly Lys Val Trp Pro Gly Ser Thr Ala Phe Pro Asp Phe			
355	360	365	
aca aac ccc acc gct ctg gct tgg tgg gaa gat atg gtg gcc gag ttc			1152
Thr Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp Met Val Ala Glu Phe			
370	375	380	
cac gat cag gtg ccc ttc gat ggc atg tgg atc gac atg aac gag ccc			1200
His Asp Gln Val Pro Phe Asp Gly Met Trp Ile Asp Met Asn Glu Pro			
385	390	395	400
agc aac ttc atc cgg ggc agc gag gac ggc tgc ccc aac aac gaa ctg			1248
Ser Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys Pro Asn Asn Glu Leu			
405	410	415	
gaa aat cct cct tac gtg ccc ggc gtt gtc ggc gga aca ctt cag gcc			1296
Glu Asn Pro Pro Tyr Val Pro Gly Val Val Gly Gly Thr Leu Gln Ala			
420	425	430	

gct aca atc tgt gcc agc agc cac cag ttc ctc agc acc cac tac aac	1344
Ala Thr Ile Cys Ala Ser Ser His Gln Phe Leu Ser Thr His Tyr Asn	
435 440 445	
ctg cac aac ctg tac ggc ctg acc gag gcc att gcc tct cat aga gcc	1392
Leu His Asn Leu Tyr Gly Leu Thr Glu Ala Ile Ala Ser His Arg Ala	
450 455 460	
ctg gtt aag gcc aga ggc acc cgg cct ttt gtg atc agc aga agc aca	1440
Leu Val Lys Ala Arg Gly Thr Arg Pro Phe Val Ile Ser Arg Ser Thr	
465 470 475 480	
ttc gcc ggc cac ggc aga tac gcc gga cat tgg aca ggc gac gtg tgg	1488
Phe Ala Gly His Gly Arg Tyr Ala Gly His Trp Thr Gly Asp Val Trp	
485 490 495	
tct agt tgg gag cag ctg gct agc agc gtg cca gag atc ctg cag ttc	1536
Ser Ser Trp Glu Gln Leu Ala Ser Ser Val Pro Glu Ile Leu Gln Phe	
500 505 510	
aat ctg ctg ggc gtg cca ctc gtg gga gcc gac gtt tgt ggc ttc ctg	1584
Asn Leu Leu Gly Val Pro Leu Val Gly Ala Asp Val Cys Gly Phe Leu	
515 520 525	
ggc aac acc agc gag gaa ctg tgt gtg cgt tgg aca cag ctg ggc gcc	1632
Gly Asn Thr Ser Glu Glu Leu Cys Val Arg Trp Thr Gln Leu Gly Ala	
530 535 540	
ttc tat ccc ttc atg aga aac cac aac agc ctg ctg agc ctg cct caa	1680
Phe Tyr Pro Phe Met Arg Asn His Asn Ser Leu Leu Ser Leu Pro Gln	
545 550 555 560	
gag ccc tac agc ttt agc gag cct gca cag cag gcc atg aga aag gcc	1728
Glu Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln Ala Met Arg Lys Ala	
565 570 575	
ctg act ctg aga tac gct ctg ctg ccc cac ctg tac acc ctg ttt cac	1776
Leu Thr Leu Arg Tyr Ala Leu Leu Pro His Leu Tyr Thr Leu Phe His	
580 585 590	
cag gct cac gtg gcc ggg gag aca gtg gct aga cct ctg ttc ctg gaa	1824

Gln Ala His Val Ala Gly Glu Thr Val Ala Arg Pro Leu Phe Leu Glu

595 600 605
 ttt ccc aag gac agc tcc acc tgg acc gtg gat cat cag ctg ctg tgg 1872
 Phe Pro Lys Asp Ser Ser Thr Trp Thr Val Asp His Gln Leu Leu Trp

610 615 620
 gga gaa gcc ctg ctc atc aca cct gtt ctg cag gcc gga aag gcc gaa 1920
 Gly Glu Ala Leu Leu Ile Thr Pro Val Leu Gln Ala Gly Lys Ala Glu
 625 630 635 640

gtg acc ggc tat ttt cct ctc ggc act tgg tac gac ctg cag acc gtg 1968
 Val Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr Asp Leu Gln Thr Val
 645 650 655

cct att gag gcc ctg gga tct ctt cct cca cct cct gct gct cct aga 2016
 Pro Ile Glu Ala Leu Gly Ser Leu Pro Pro Pro Pro Ala Ala Pro Arg
 660 665 670

gag cct gcc atc cac tct gaa ggc cag tgg gtt aca ctg ccc gct cct 2064

Glu Pro Ala Ile His Ser Glu Gly Gln Trp Val Thr Leu Pro Ala Pro
 675 680 685

ctg gac acc atc aac gtg cac ctg aga gct ggc tac atc atc cca ctg 2112
 Leu Asp Thr Ile Asn Val His Leu Arg Ala Gly Tyr Ile Ile Pro Leu
 690 695 700

caa ggc cct ggc ctg acc aca acc gaa tct aga cag cag ccc atg gct 2160
 Gln Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg Gln Gln Pro Met Ala

705 710 715 720
 ctg gcc gtg gct ctt aca aaa ggc gga gag gct aga ggc gag ctg ttc 2208
 Leu Ala Val Ala Leu Thr Lys Gly Gly Glu Ala Arg Gly Glu Leu Phe

725 730 735
 tgg gac gac ggc gag tct ctg gaa gtg ctg gaa cgg ggc gct tat acc 2256
 Trp Asp Asp Gly Glu Ser Leu Glu Val Leu Glu Arg Gly Ala Tyr Thr

740 745 750

caa gtg atc ttc ctg gcc aga aac aac acc atc gtg aac gaa ctc gtg 2304

Gln Val Ile Phe Leu Ala Arg Asn Asn Thr Ile Val Asn Glu Leu Val

755 760 765

<210> 17

<211> 768

<212> PRT

<213> Artificial sequence

<220><223> Synthetic Construct

<400> 17

Gly Gln Pro Trp Cys Phe Phe Pro Pro Ser Tyr Pro Ser Tyr Lys Leu

1 5 10 15

Glu Asn Leu Ser Ser Ser Glu Met Gly Tyr Thr Ala Thr Leu Thr Arg

20 25 30

Thr Thr Pro Thr Phe Phe Pro Lys Asp Ile Leu Thr Leu Arg Leu Asp

35 40 45

Val Met Met Glu Thr Glu Asn Arg Leu His Phe Thr Ile Lys Asp Pro

50 55 60

Ala Asn Arg Arg Tyr Glu Val Pro Leu Glu Thr Pro His Val His Ser

65 70 75 80

Arg Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe Ser Glu Glu Pro Phe

85 90 95

Gly Val Ile Val Arg Arg Gln Leu Asp Gly Arg Val Leu Leu Asn Thr

100 105 110

Thr Val Ala Pro Leu Phe Phe Ala Asp Gln Phe Leu Gln Leu Ser Thr

115 120 125

Ser Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala Glu His Leu Ser Pro

130 135 140

Leu Met Leu Ser Thr Ser Trp Thr Arg Ile Thr Leu Trp Asn Arg Asp

145 150 155 160

Leu Ala Pro Thr Pro Gly Ala Asn Leu Tyr Gly Ser His Pro Phe Tyr

165 170 175

Leu Ala Leu Glu Asp Gly Gly Ser Ala His Gly Val Phe Leu Leu Asn

180 185 190
Ser Asn Ala Met Asp Val Val Leu Gln Pro Ser Pro Ala Leu Ser Trp

195 200 205
Arg Ser Thr Gly Gly Ile Leu Asp Val Tyr Ile Phe Leu Gly Pro Glu

210 215 220
Pro Lys Ser Val Val Gln Gln Tyr Leu Asp Val Val Gly Tyr Pro Phe

225 230 235 240
Met Pro Pro Tyr Trp Gly Leu Gly Phe His Leu Cys Arg Trp Gly Tyr

245 250 255
Ser Ser Thr Ala Ile Thr Arg Gln Val Val Glu Asn Met Thr Arg Ala

260 265 270
His Phe Pro Leu Asp Val Gln Trp Asn Asp Leu Asp Tyr Met Asp Ser

275 280 285
Arg Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe Arg Asp Phe Pro Ala

290 295 300
Met Val Gln Glu Leu His Gln Gly Gly Arg Arg Tyr Met Met Ile Val

305 310 315 320
Asp Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly Ser Tyr Arg Pro Tyr

325 330 335
Asp Glu Gly Leu Arg Arg Gly Val Phe Ile Thr Asn Glu Thr Gly Gln

340 345 350
Pro Leu Ile Gly Lys Val Trp Pro Gly Ser Thr Ala Phe Pro Asp Phe

355 360 365
Thr Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp Met Val Ala Glu Phe

370 375 380
His Asp Gln Val Pro Phe Asp Gly Met Trp Ile Asp Met Asn Glu Pro

385 390 395 400
Ser Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys Pro Asn Asn Glu Leu

405 410 415
Glu Asn Pro Pro Tyr Val Pro Gly Val Val Gly Gly Thr Leu Gln Ala

420 425 430

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

675 680 685
 Leu Asp Thr Ile Asn Val His Leu Arg Ala Gly Tyr Ile Ile Pro Leu
 690 695 700
 Gln Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg Gln Gln Pro Met Ala
 705 710 715 720
 Leu Ala Val Ala Leu Thr Lys Gly Gly Glu Ala Arg Gly Glu Leu Phe

725 730 735
 Trp Asp Asp Gly Glu Ser Leu Glu Val Leu Glu Arg Gly Ala Tyr Thr
 740 745 750
 Gln Val Ile Phe Leu Ala Arg Asn Asn Thr Ile Val Asn Glu Leu Val
 755 760 765

<210> 18

<211> 2247

<212> DNA

<213> Artificial sequence

<220><223> Engineered hGAA 70kD cDNA

<220><221> CDS

<222> (1)..(2247)

<400> 18

gcc cca tct cca ctg tac agc gtg gaa ttt tcc gag gaa ccc ttc ggc 48

Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe Ser Glu Glu Pro Phe Gly
 1 5 10 15

gtg atc gtg cgg aga cag ctg gat gga aga gtg ctg ctg aac acc aca 96

Val Ile Val Arg Arg Gln Leu Asp Gly Arg Val Leu Leu Asn Thr Thr

20 25 30

gtg gcc cct ctg ttc ttc gcc gac cag ttt ctg cag ctg agc acc agc 144

Val Ala Pro Leu Phe Phe Ala Asp Gln Phe Leu Gln Leu Ser Thr Ser

35 40 45

ctg cct agc cag tat atc aca ggc ctg gcc gag cac ctg tct cca ctg 192

Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala Glu His Leu Ser Pro Leu

50 55 60

atg ctg agc aca tcc tgg acc aga atc acc ctg tgg aac aga gat ctg 240

Met	Leu	Ser	Thr	Ser	Trp	Thr	Arg	Ile	Thr	Leu	Trp	Asn	Arg	Asp	Leu	
65				70				75				80				
gcc	cct	aca	cct	ggc	gcc	aac	ctg	tac	ggc	tct	cac	ccc	ttt	tat	ctg	288
Ala	Pro	Thr	Pro	Gly	Ala	Asn	Leu	Tyr	Gly	Ser	His	Pro	Phe	Tyr	Leu	
85				90				95								
gcc	ctg	gaa	gac	ggc	gga	tct	gcc	cac	ggg	gtt	ttt	ctg	ctg	aac	tcc	336
Ala	Leu	Glu	Asp	Gly	Gly	Ser	Ala	His	Gly	Val	Phe	Leu	Leu	Asn	Ser	
100				105				110								
aac	gcc	atg	gac	gtg	gtg	ctg	cag	cca	tct	cct	gct	ctg	tct	tgg	aga	384
Asn	Ala	Met	Asp	Val	Val	Leu	Gln	Pro	Ser	Pro	Ala	Leu	Ser	Trp	Arg	
115				120				125								
agc	aca	ggc	ggc	atc	ctg	gac	gtg	tac	atc	ttt	ctg	ggc	ccc	gag	cct	432
Ser	Thr	Gly	Gly	Ile	Leu	Asp	Val	Tyr	Ile	Phe	Leu	Gly	Pro	Glu	Pro	
130				135				140								
aag	agc	gtg	gtg	cag	cag	tat	ctg	gac	gtc	gtg	ggc	tac	ccc	ttc	atg	480
Lys	Ser	Val	Val	Gln	Gln	Tyr	Leu	Asp	Val	Val	Gly	Tyr	Pro	Phe	Met	
145				150				155				160				
cct	cct	tat	tgg	ggc	ctg	ggc	ttc	cac	ctg	tgc	agg	tgg	gga	tac	agc	528
Pro	Pro	Tyr	Trp	Gly	Leu	Gly	Phe	His	Leu	Cys	Arg	Trp	Gly	Tyr	Ser	
165				170				175								
agc	acc	gcc	atc	acc	aga	cag	gtg	gtg	gaa	aac	atg	acc	cgg	gct	cac	576
Ser	Thr	Ala	Ile	Thr	Arg	Gln	Val	Val	Glu	Asn	Met	Thr	Arg	Ala	His	
180				185				190								
ttc	cca	ctg	gac	gtg	cag	tgg	aac	gac	ctg	gac	tac	atg	gac	agc	aga	624
Phe	Pro	Leu	Asp	Val	Gln	Trp	Asn	Asp	Leu	Asp	Tyr	Met	Asp	Ser	Arg	
195				200				205								
cgg	gac	ttc	acc	ttc	aac	aag	gac	ggc	ttc	aga	gac	ttc	ccc	gcc	atg	672
Arg	Asp	Phe	Thr	Phe	Asn	Lys	Asp	Gly	Phe	Arg	Asp	Phe	Pro	Ala	Met	
210				215				220								

gtg caa gaa ctg cac caa ggc ggc aga cgg tac atg atg att gtg gac 720

Val Gln Glu Leu His Gln Gly Gly Arg Arg Tyr Met Met Ile Val Asp
225 230 235 240

cca gcc atc agc tct agc ggc cct gcc gga agc tac aga cct tac gat 768
Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly Ser Tyr Arg Pro Tyr Asp
245 250 255

gag ggc ctg aga aga ggc gtg ttc atc acc aac gag aca ggc cag cct 816
Glu Gly Leu Arg Arg Gly Val Phe Ile Thr Asn Glu Thr Gly Gln Pro
260 265 270

ctg atc ggc aaa gtg tgg cct ggc agc aca gcc ttt cca gac ttc aca 864
Leu Ile Gly Lys Val Trp Pro Gly Ser Thr Ala Phe Pro Asp Phe Thr
275 280 285

aac ccc acc gct ctg gct tgg tgg gaa gat atg gtg gcc gag ttc cac 912
Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp Met Val Ala Glu Phe His
290 295 300

gat cag gtg ccc ttc gat ggc atg tgg atc gac atg aac gag ccc agc 960
Asp Gln Val Pro Phe Asp Gly Met Trp Ile Asp Met Asn Glu Pro Ser
305 310 315 320

aac ttc atc cgg ggc agc gag gac ggc tgc ccc aac aac gaa ctg gaa 1008
Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys Pro Asn Asn Glu Leu Glu
325 330 335

aat cct cct tac gtg ccc ggc gtt gtc ggc gga aca ctt cag gcc gct 1056

Asn Pro Pro Tyr Val Pro Gly Val Val Gly Gly Thr Leu Gln Ala Ala
340 345 350

aca atc tgt gcc agc agc cac cag ttc ctc agc acc cac tac aac ctg 1104
Thr Ile Cys Ala Ser Ser His Gln Phe Leu Ser Thr His Tyr Asn Leu
355 360 365

cac aac ctg tac ggc ctg acc gag gcc att gcc tct cat aga gcc ctg 1152
His Asn Leu Tyr Gly Leu Thr Glu Ala Ile Ala Ser His Arg Ala Leu

370	375	380	
gtt aag gcc aga ggc acc cgg cct ttt gtg atc agc aga agc aca ttc			1200
Val Lys Ala Arg Gly Thr Arg Pro Phe Val Ile Ser Arg Ser Thr Phe			
385	390	395	400
gcc ggc cac ggc aga tac gcc gga cat tgg aca ggc gac gtg tgg tct			1248
Ala Gly His Gly Arg Tyr Ala Gly His Trp Thr Gly Asp Val Trp Ser			
405	410	415	
agt tgg gag cag ctg gct agc agc gtg cca gag atc ctg cag ttc aat			1296
Ser Trp Glu Gln Leu Ala Ser Ser Val Pro Glu Ile Leu Gln Phe Asn			
420	425	430	
ctg ctg ggc gtg cca ctc gtg gga gcc gac gtt tgt ggc ttc ctg ggc			1344
Leu Leu Gly Val Pro Leu Val Gly Ala Asp Val Cys Gly Phe Leu Gly			
435	440	445	
aac acc agc gag gaa ctg tgt gtg cgt tgg aca cag ctg ggc gcc ttc			1392
Asn Thr Ser Glu Glu Leu Cys Val Arg Trp Thr Gln Leu Gly Ala Phe			
450	455	460	
tat ccc ttc atg aga aac cac aac agc ctg ctg agc ctg cct caa gag			1440
Tyr Pro Phe Met Arg Asn His Asn Ser Leu Leu Ser Leu Pro Gln Glu			
465	470	475	480
ccc tac agc ttt agc gag cct gca cag cag gcc atg aga aag gcc ctg			1488
Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln Ala Met Arg Lys Ala Leu			
485	490	495	
act ctg aga tac gct ctg ctg ccc cac ctg tac acc ctg ttt cac cag			1536
Thr Leu Arg Tyr Ala Leu Leu Pro His Leu Tyr Thr Leu Phe His Gln			
500	505	510	
gct cac gtg gcc ggg gag aca gtg gct aga cct ctg ttc ctg gaa ttt			1584
Ala His Val Ala Gly Glu Thr Val Ala Arg Pro Leu Phe Leu Glu Phe			
515	520	525	
ccc aag gac agc tcc acc tgg acc gtg gat cat cag ctg ctg tgg gga			1632
Pro Lys Asp Ser Ser Thr Trp Thr Val Asp His Gln Leu Leu Trp Gly			
530	535	540	

gaa gcc ctg ctc atc aca cct gtt ctg cag gcc gga aag gcc gaa gtg	1680
Glu Ala Leu Leu Ile Thr Pro Val Leu Gln Ala Gly Lys Ala Glu Val	
545 550 555 560	
acc ggc tat ttt cct ctc ggc act tgg tac gac ctg cag acc gtg cct	1728
Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr Asp Leu Gln Thr Val Pro	
565 570 575	
att gag gcc ctg gga tct ctt cct cca cct cct gct gct cct aga gag	1776
Ile Glu Ala Leu Gly Ser Leu Pro Pro Pro Pro Ala Ala Pro Arg Glu	
580 585 590	
cct gcc atc cac tct gaa ggc cag tgg gtt aca ctg ccc gct cct ctg	1824
Pro Ala Ile His Ser Glu Gly Gln Trp Val Thr Leu Pro Ala Pro Leu	
595 600 605	
gac acc atc aac gtg cac ctg aga gct ggc tac atc atc cca ctg caa	1872
Asp Thr Ile Asn Val His Leu Arg Ala Gly Tyr Ile Ile Pro Leu Gln	
610 615 620	
ggc cct ggc ctg acc aca acc gaa tct aga cag cag ccc atg gct ctg	1920
Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg Gln Gln Pro Met Ala Leu	
625 630 635 640	
gcc gtg gct ctt aca aaa ggc gga gag gct aga ggc gag ctg ttc tgg	1968
Ala Val Ala Leu Thr Lys Gly Gly Glu Ala Arg Gly Glu Leu Phe Trp	
645 650 655	
gac gac ggc gag tct ctg gaa gtg ctg gaa cgg ggc gct tat acc caa	2016
Asp Asp Gly Glu Ser Leu Glu Val Leu Glu Arg Gly Ala Tyr Thr Gln	
660 665 670	
gtg atc ttc ctg gcc aga aac aac acc atc gtg aac gaa ctc gtg cgc	2064
Val Ile Phe Leu Ala Arg Asn Asn Thr Ile Val Asn Glu Leu Val Arg	
675 680 685	
gtg acc agt gaa ggt gct gga ctg caa ctg cag aaa gtg acc gtg ctc	2112
Val Thr Ser Glu Gly Ala Gly Leu Gln Leu Gln Lys Val Thr Val Leu	
690 695 700	
gga gtg gcc aca gct cct caa cag gtg ctg tct aac ggc gtg ccc gtg	2160

Gly Val Ala Thr Ala Pro Gln Gln Val Leu Ser Asn Gly Val Pro Val

705 710 715 720
tcc aac ttc aca tac agc ccc gac acc aag gtc ctg gac atc tgt gtg 2208

Ser Asn Phe Thr Tyr Ser Pro Asp Thr Lys Val Leu Asp Ile Cys Val

 725 730 735
tca ctg ctg atg ggc gag cag ttc ctg gtg tcc tgg tgc 2247

Ser Leu Leu Met Gly Glu Gln Phe Leu Val Ser Trp Cys

 740 745

<210> 19

<211> 749

<212> PRT

<213> Artificial sequence

<220><223> Synthetic Construct

<400> 19

Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe Ser Glu Glu Pro Phe Gly

1 5 10 15

Val Ile Val Arg Arg Gln Leu Asp Gly Arg Val Leu Leu Asn Thr Thr

 20 25 30

Val Ala Pro Leu Phe Phe Ala Asp Gln Phe Leu Gln Leu Ser Thr Ser

 35 40 45

Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala Glu His Leu Ser Pro Leu

 50 55 60

Met Leu Ser Thr Ser Trp Thr Arg Ile Thr Leu Trp Asn Arg Asp Leu

65 70 75 80

Ala Pro Thr Pro Gly Ala Asn Leu Tyr Gly Ser His Pro Phe Tyr Leu

 85 90 95

Ala Leu Glu Asp Gly Gly Ser Ala His Gly Val Phe Leu Leu Asn Ser

 100 105 110

Asn Ala Met Asp Val Val Leu Gln Pro Ser Pro Ala Leu Ser Trp Arg

 115 120 125

Ser Thr Gly Gly Ile Leu Asp Val Tyr Ile Phe Leu Gly Pro Glu Pro

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

 180 185 190
 Phe Pro Leu Asp Val Gln Trp Asn Asp Leu Asp Tyr Met Asp Ser Arg
 195 200 205
 Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe Arg Asp Phe Pro Ala Met
 210 215 220
 Val Gln Glu Leu His Gln Gly Gly Arg Arg Tyr Met Met Ile Val Asp
 225 230 235 240
 Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly Ser Tyr Arg Pro Tyr Asp

 245 250 255
 Glu Gly Leu Arg Arg Gly Val Phe Ile Thr Asn Glu Thr Gly Gln Pro
 260 265 270
 Leu Ile Gly Lys Val Trp Pro Gly Ser Thr Ala Phe Pro Asp Phe Thr
 275 280 285
 Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp Met Val Ala Glu Phe His
 290 295 300
 Asp Gln Val Pro Phe Asp Gly Met Trp Ile Asp Met Asn Glu Pro Ser

 305 310 315 320
 Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys Pro Asn Asn Glu Leu Glu
 325 330 335
 Asn Pro Pro Tyr Val Pro Gly Val Val Gly Gly Thr Leu Gln Ala Ala
 340 345 350
 Thr Ile Cys Ala Ser Ser His Gln Phe Leu Ser Thr His Tyr Asn Leu
 355 360 365
 His Asn Leu Tyr Gly Leu Thr Glu Ala Ile Ala Ser His Arg Ala Leu

 370 375 380

Val Lys Ala Arg Gly Thr Arg Pro Phe Val Ile Ser Arg Ser Thr Phe
385 390 395 400
Ala Gly His Gly Arg Tyr Ala Gly His Trp Thr Gly Asp Val Trp Ser
405 410 415
Ser Trp Glu Gln Leu Ala Ser Ser Val Pro Glu Ile Leu Gln Phe Asn
420 425 430
Leu Leu Gly Val Pro Leu Val Gly Ala Asp Val Cys Gly Phe Leu Gly
435 440 445
Asn Thr Ser Glu Glu Leu Cys Val Arg Trp Thr Gln Leu Gly Ala Phe
450 455 460
Tyr Pro Phe Met Arg Asn His Asn Ser Leu Leu Ser Leu Pro Gln Glu
465 470 475 480
Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln Ala Met Arg Lys Ala Leu
485 490 495
Thr Leu Arg Tyr Ala Leu Leu Pro His Leu Tyr Thr Leu Phe His Gln
500 505 510
Ala His Val Ala Gly Glu Thr Val Ala Arg Pro Leu Phe Leu Glu Phe
515 520 525
Pro Lys Asp Ser Ser Thr Trp Thr Val Asp His Gln Leu Leu Trp Gly
530 535 540
Glu Ala Leu Leu Ile Thr Pro Val Leu Gln Ala Gly Lys Ala Glu Val
545 550 555 560
Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr Asp Leu Gln Thr Val Pro
565 570 575
Ile Glu Ala Leu Gly Ser Leu Pro Pro Pro Pro Ala Ala Pro Arg Glu
580 585 590
Pro Ala Ile His Ser Glu Gly Gln Trp Val Thr Leu Pro Ala Pro Leu
595 600 605
Asp Thr Ile Asn Val His Leu Arg Ala Gly Tyr Ile Ile Pro Leu Gln
610 615 620
Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg Gln Gln Pro Met Ala Leu

625 630 635 640
 Ala Val Ala Leu Thr Lys Gly Gly Glu Ala Arg Gly Glu Leu Phe Trp
 645 650 655
 Asp Asp Gly Glu Ser Leu Glu Val Leu Glu Arg Gly Ala Tyr Thr Gln
 660 665 670
 Val Ile Phe Leu Ala Arg Asn Asn Thr Ile Val Asn Glu Leu Val Arg
 675 680 685
 Val Thr Ser Glu Gly Ala Gly Leu Gln Leu Gln Lys Val Thr Val Leu

690 695 700
 Gly Val Ala Thr Ala Pro Gln Gln Val Leu Ser Asn Gly Val Pro Val
 705 710 715 720
 Ser Asn Phe Thr Tyr Ser Pro Asp Thr Lys Val Leu Asp Ile Cys Val
 725 730 735
 Ser Leu Leu Met Gly Glu Gln Phe Leu Val Ser Trp Cys
 740 745

<210> 20

<211> 2490

<212> DNA

<213> Artificial Sequence

<220

><223> Engineered DNA for hGAHV780I 76 kD protein

<220><221> CDS

<222> (1)..(2490)

<400> 20

gga cag ccc tgg tgc ttc ttc cca cca tct tac ccc agc tac aag ctg 48

Gly Gln Pro Trp Cys Phe Phe Pro Pro Ser Tyr Pro Ser Tyr Lys Leu

1 5 10 15

gaa aac ctg agc agc agc gag atg ggc tac acc gcc aca ctg acc aga 96

Glu Asn Leu Ser Ser Ser Glu Met Gly Tyr Thr Ala Thr Leu Thr Arg

20 25 30

acc aca cct aca ttc ttc ccg aag gac atc ctg aca ctg cgg ctg gac 144

Thr Thr Pro Thr Phe Phe Pro Lys Asp Ile Leu Thr Leu Arg Leu Asp

35 40 45

gtg atg atg gaa acc gag aac cgg ctg cac ttc acc atc aag gac ccc	192
Val Met Met Glu Thr Glu Asn Arg Leu His Phe Thr Ile Lys Asp Pro	
50 55 60	
gcc aat cgg aga tac gag gtg cca ctg gaa acc cct cac gtg cac tct	240
Ala Asn Arg Arg Tyr Glu Val Pro Leu Glu Thr Pro His Val His Ser	
65 70 75 80	
aga gcc cca tct cca ctg tac agc gtg gaa ttt tcc gag gaa ccc ttc	288
Arg Ala Pro Ser Pro Leu Tyr Ser Val Glu Phe Ser Glu Glu Pro Phe	
85 90 95	
ggc gtg atc gtg cgg aga cag ctg gat gga aga gtg ctg ctg aac acc	336
Gly Val Ile Val Arg Arg Gln Leu Asp Gly Arg Val Leu Leu Asn Thr	
100 105 110	
aca gtg gcc cct ctg ttc ttc gcc gac cag ttt ctg cag ctg agc acc	384
Thr Val Ala Pro Leu Phe Phe Ala Asp Gln Phe Leu Gln Leu Ser Thr	
115 120 125	
agc ctg cct agc cag tat atc aca ggc ctg gcc gag cac ctg tct cca	432
Ser Leu Pro Ser Gln Tyr Ile Thr Gly Leu Ala Glu His Leu Ser Pro	
130 135 140	
ctg atg ctg agc aca tcc tgg acc aga atc acc ctg tgg aac aga gat	480
Leu Met Leu Ser Thr Ser Trp Thr Arg Ile Thr Leu Trp Asn Arg Asp	
145 150 155 160	
ctg gcc cct aca cct ggc gcc aac ctg tac ggc tct cac ccc ttt tat	528
Leu Ala Pro Thr Pro Gly Ala Asn Leu Tyr Gly Ser His Pro Phe Tyr	
165 170 175	
ctg gcc ctg gaa gac ggc gga tct gcc cac ggt gtt ttt ctg ctg aac	576
Leu Ala Leu Glu Asp Gly Gly Ser Ala His Gly Val Phe Leu Leu Asn	
180 185 190	
tcc aac gcc atg gac gtg gtg ctg cag cca tct cct gct ctg tct tgg	624
Ser Asn Ala Met Asp Val Val Leu Gln Pro Ser Pro Ala Leu Ser Trp	
195 200 205	
aga agc aca ggc ggc atc ctg gac gtg tac atc ttt ctg ggc ccc gag	672

Arg Ser Thr Gly Gly Ile Leu Asp Val Tyr Ile Phe Leu Gly Pro Glu

210	215	220	
cct aag agc gtg gtg cag cag tat ctg gac gtc gtg ggc tac ccc ttc			720
Pro Lys Ser Val Val Gln Gln Tyr Leu Asp Val Val Gly Tyr Pro Phe			
225	230	235	240
atg cct cct tat tgg ggc ctg ggc ttc cac ctg tgc agg tgg gga tac			768
Met Pro Pro Tyr Trp Gly Leu Gly Phe His Leu Cys Arg Trp Gly Tyr			
245	250	255	
agc agc acc gcc atc acc aga cag gtg gtg gaa aac atg acc cgg gct			816
Ser Ser Thr Ala Ile Thr Arg Gln Val Val Glu Asn Met Thr Arg Ala			
260	265	270	
cac ttc cca ctg gac gtg cag tgg aac gac ctg gac tac atg gac agc			864
His Phe Pro Leu Asp Val Gln Trp Asn Asp Leu Asp Tyr Met Asp Ser			
275	280	285	
aga cgg gac ttc acc ttc aac aag gac ggc ttc aga gac ttc ccc gcc			912
Arg Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe Arg Asp Phe Pro Ala			
290	295	300	
atg gtg caa gaa ctg cac caa ggc ggc aga cgg tac atg atg att gtg			960
Met Val Gln Glu Leu His Gln Gly Gly Arg Arg Tyr Met Met Ile Val			
305	310	315	320
gac cca gcc atc agc tct agc ggc cct gcc gga agc tac aga cct tac			1008
Asp Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly Ser Tyr Arg Pro Tyr			
325	330	335	
gat gag ggc ctg aga aga ggc gtg ttc atc acc aac gag aca ggc cag			1056
Asp Glu Gly Leu Arg Arg Gly Val Phe Ile Thr Asn Glu Thr Gly Gln			
340	345	350	
cct ctg atc ggc aaa gtg tgg cct ggc agc aca gcc ttt cca gac ttc			1104
Pro Leu Ile Gly Lys Val Trp Pro Gly Ser Thr Ala Phe Pro Asp Phe			
355	360	365	

aca aac ccc acc gct ctg gct tgg tgg gaa gat atg gtg gcc gag ttc	1152
Thr Asn Pro Thr Ala Leu Ala Trp Trp Glu Asp Met Val Ala Glu Phe	
370 375 380	
cac gat cag gtg ccc ttc gat ggc atg tgg atc gac atg aac gag ccc	1200
His Asp Gln Val Pro Phe Asp Gly Met Trp Ile Asp Met Asn Glu Pro	
385 390 395 400	
agc aac ttc atc cgg ggc agc gag gac ggc tgc ccc aac aac gaa ctg	1248
Ser Asn Phe Ile Arg Gly Ser Glu Asp Gly Cys Pro Asn Asn Glu Leu	
405 410 415	
gaa aat cct cct tac gtg ccc ggc gtt gtc ggc gga aca ctt cag gcc	1296
Glu Asn Pro Pro Tyr Val Pro Gly Val Val Gly Gly Thr Leu Gln Ala	
420 425 430	
gct aca atc tgt gcc agc agc cac cag ttc ctc agc acc cac tac aac	1344
Ala Thr Ile Cys Ala Ser Ser His Gln Phe Leu Ser Thr His Tyr Asn	
435 440 445	
ctg cac aac ctg tac ggc ctg acc gag gcc att gcc tct cat aga gcc	1392
Leu His Asn Leu Tyr Gly Leu Thr Glu Ala Ile Ala Ser His Arg Ala	
450 455 460	
ctg gtt aag gcc aga ggc acc cgg cct ttt gtg atc agc aga agc aca	1440
Leu Val Lys Ala Arg Gly Thr Arg Pro Phe Val Ile Ser Arg Ser Thr	
465 470 475 480	
ttc gcc ggc cac ggc aga tac gcc gga cat tgg aca ggc gac gtg tgg	1488
Phe Ala Gly His Gly Arg Tyr Ala Gly His Trp Thr Gly Asp Val Trp	
485 490 495	
tct agt tgg gag cag ctg gct agc agc gtg cca gag atc ctg cag ttc	1536
Ser Ser Trp Glu Gln Leu Ala Ser Ser Val Pro Glu Ile Leu Gln Phe	
500 505 510	
aat ctg ctg ggc gtg cca ctc gtg gga gcc gac gtt tgt ggc ttc ctg	1584
Asn Leu Leu Gly Val Pro Leu Val Gly Ala Asp Val Cys Gly Phe Leu	
515 520 525	
ggc aac acc agc gag gaa ctg tgt gtg cgt tgg aca cag ctg ggc gcc	1632

Gly Asn Thr Ser Glu Glu Leu Cys Val Arg Trp Thr Gln Leu Gly Ala	
530 535 540	
ttc tat ccc ttc atg aga aac cac aac agc ctg ctg agc ctg cct caa	1680
Phe Tyr Pro Phe Met Arg Asn His Asn Ser Leu Leu Ser Leu Pro Gln	
545 550 555 560	
gag ccc tac agc ttt agc gag cct gca cag cag gcc atg aga aag gcc	1728
Glu Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln Ala Met Arg Lys Ala	
565 570 575	
ctg act ctg aga tac gct ctg ctg ccc cac ctg tac acc ctg ttt cac	1776
Leu Thr Leu Arg Tyr Ala Leu Leu Pro His Leu Tyr Thr Leu Phe His	
580 585 590	
cag gct cac gtg gcc ggg gag aca gtg gct aga cct ctg ttc ctg gaa	1824
Gln Ala His Val Ala Gly Glu Thr Val Ala Arg Pro Leu Phe Leu Glu	
595 600 605	
ttt ccc aag gac agc tcc acc tgg acc gtg gat cat cag ctg ctg tgg	1872
Phe Pro Lys Asp Ser Ser Thr Trp Thr Val Asp His Gln Leu Leu Trp	
610 615 620	
gga gaa gcc ctg ctc atc aca cct gtt ctg cag gcc gga aag gcc gaa	1920
Gly Glu Ala Leu Leu Ile Thr Pro Val Leu Gln Ala Gly Lys Ala Glu	
625 630 635 640	
gtg acc ggc tat ttt cct ctc ggc act tgg tac gac ctg cag acc gtg	1968
Val Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr Asp Leu Gln Thr Val	
645 650 655	
cct att gag gcc ctg gga tct ctt cct cca cct cct gct gct cct aga	2016
Pro Ile Glu Ala Leu Gly Ser Leu Pro Pro Pro Pro Ala Ala Pro Arg	
660 665 670	
gag cct gcc atc cac tct gaa ggc cag tgg gtt aca ctg ccc gct cct	2064
Glu Pro Ala Ile His Ser Glu Gly Gln Trp Val Thr Leu Pro Ala Pro	
675 680 685	
ctg gac acc atc aac gtg cac ctg aga gct ggc tac atc atc cca ctg	2112
Leu Asp Thr Ile Asn Val His Leu Arg Ala Gly Tyr Ile Ile Pro Leu	

690	695	700	
caa ggc cct ggc ctg acc aca acc gaa tct aga cag cag ccc atg gct			2160
Gln Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg Gln Gln Pro Met Ala			
705	710	715	720
ctg gcc gtg gct ctt aca aaa ggc gga gag gct aga ggc gag ctg ttc			2208
Leu Ala Val Ala Leu Thr Lys Gly Gly Glu Ala Arg Gly Glu Leu Phe			
	725	730	735
tgg gac gac ggc gag tct ctg gaa gtg ctg gaa cgg ggc gct tat acc			2256
Trp Asp Asp Gly Glu Ser Leu Glu Val Leu Glu Arg Gly Ala Tyr Thr			
	740	745	750
caa gtg atc ttc ctg gcc aga aac aac acc atc gtg aac gaa ctc gtg			2304
Gln Val Ile Phe Leu Ala Arg Asn Asn Thr Ile Val Asn Glu Leu Val			
	755	760	765
cgc gtg acc agt gaa ggt gct gga ctg caa ctg cag aaa gtg acc gtg			2352
Arg Val Thr Ser Glu Gly Ala Gly Leu Gln Leu Gln Lys Val Thr Val			
	770	775	780
ctc gga gtg gcc aca gct cct caa cag gtg ctg tct aac ggc gtg ccc			2400
Leu Gly Val Ala Thr Ala Pro Gln Gln Val Leu Ser Asn Gly Val Pro			
785	790	795	800
gtg tcc aac ttc aca tac agc ccc gac acc aag gtc ctg gac atc tgt			2448
Val Ser Asn Phe Thr Tyr Ser Pro Asp Thr Lys Val Leu Asp Ile Cys			
	805	810	815
gtg tca ctg ctg atg ggc gag cag ttc ctg gtg tcc tgg tgc			2490
Val Ser Leu Leu Met Gly Glu Gln Phe Leu Val Ser Trp Cys			
	820	825	830
<210> 21			
<211> 830			
<212> PRT			
<213> Artificial Sequence			
<220><223> Synthetic Construct			
<400> 21			

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

245 250 255
 Ser Ser Thr Ala Ile Thr Arg Gln Val Val Glu Asn Met Thr Arg Ala
 260 265 270
 His Phe Pro Leu Asp Val Gln Trp Asn Asp Leu Asp Tyr Met Asp Ser

 275 280 285
 Arg Arg Asp Phe Thr Phe Asn Lys Asp Gly Phe Arg Asp Phe Pro Ala
 290 295 300
 Met Val Gln Glu Leu His Gln Gly Gly Arg Arg Tyr Met Met Ile Val
 305 310 315 320
 Asp Pro Ala Ile Ser Ser Ser Gly Pro Ala Gly Ser Tyr Arg Pro Tyr
 325 330 335
 Asp Glu Gly Leu Arg Arg Gly Val Phe Ile Thr Asn Glu Thr Gly Gln

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

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

 465 470 475 480
 Phe Ala Gly His Gly Arg Tyr Ala Gly His Trp Thr Gly Asp Val Trp
 485 490 495

Ser Ser Trp Glu Gln Leu Ala Ser Ser Val Pro Glu Ile Leu Gln Phe
500 505 510

Asn Leu Leu Gly Val Pro Leu Val Gly Ala Asp Val Cys Gly Phe Leu
515 520 525

Gly Asn Thr Ser Glu Glu Leu Cys Val Arg Trp Thr Gln Leu Gly Ala
530 535 540

Phe Tyr Pro Phe Met Arg Asn His Asn Ser Leu Leu Ser Leu Pro Gln
545 550 555 560

Glu Pro Tyr Ser Phe Ser Glu Pro Ala Gln Gln Ala Met Arg Lys Ala
565 570 575

Leu Thr Leu Arg Tyr Ala Leu Leu Pro His Leu Tyr Thr Leu Phe His
580 585 590

Gln Ala His Val Ala Gly Glu Thr Val Ala Arg Pro Leu Phe Leu Glu
595 600 605

Phe Pro Lys Asp Ser Ser Thr Trp Thr Val Asp His Gln Leu Leu Trp
610 615 620

Gly Glu Ala Leu Leu Ile Thr Pro Val Leu Gln Ala Gly Lys Ala Glu
625 630 635 640

Val Thr Gly Tyr Phe Pro Leu Gly Thr Trp Tyr Asp Leu Gln Thr Val
645 650 655

Pro Ile Glu Ala Leu Gly Ser Leu Pro Pro Pro Pro Ala Ala Pro Arg
660 665 670

Glu Pro Ala Ile His Ser Glu Gly Gln Trp Val Thr Leu Pro Ala Pro
675 680 685

Leu Asp Thr Ile Asn Val His Leu Arg Ala Gly Tyr Ile Ile Pro Leu
690 695 700

Gln Gly Pro Gly Leu Thr Thr Thr Glu Ser Arg Gln Gln Pro Met Ala
705 710 715 720

Leu Ala Val Ala Leu Thr Lys Gly Gly Glu Ala Arg Gly Glu Leu Phe
725 730 735

Trp Asp Asp Gly Glu Ser Leu Glu Val Leu Glu Arg Gly Ala Tyr Thr

740 745 750
Gln Val Ile Phe Leu Ala Arg Asn Asn Thr Ile Val Asn Glu Leu Val
755 760 765
Arg Val Thr Ser Glu Gly Ala Gly Leu Gln Leu Gln Lys Val Thr Val
770 775 780
Leu Gly Val Ala Thr Ala Pro Gln Gln Val Leu Ser Asn Gly Val Pro

785 790 795 800
Val Ser Asn Phe Thr Tyr Ser Pro Asp Thr Lys Val Leu Asp Ile Cys
805 810 815
Val Ser Leu Leu Met Gly Glu Gln Phe Leu Val Ser Trp Cys
820 825 830

<210> 22
<211> 2952
<212> DNA
<213> Artificial sequence
<220><223> synthetic construct
<220><221> CDS
<222> (1)..(2952)
<220><221> misc_feature
<222> (1)..(270)
<223> BiP signal peptide + vIGF2 + 2GS extension

<220><221> misc_feature
<222> (271)..(2952)
<223> engineered DNA for hGAA 61 - 952 780I
<220><221> misc_feature
<222> (2428)..(2430)
<223> Ile codon

<400> 22
atg aag ctg tct ctg gtg gct gct atg ctg ctg ctc ctg tct gcc gcc 48
Met Lys Leu Ser Leu Val Ala Ala Met Leu Leu Leu Ser Ala Ala
1 5 10 15
aga gcc agc aga aca ctt tgt ggc gga gag ctg gtg gac acc ctg cag 96

Arg Ala Ser Arg Thr Leu Cys Gly Gly Glu Leu Val Asp Thr Leu Gln

20 25 30
 ttt gtg tgt ggc gac aga ggc ttc ctg ttc agc aga cct gcc agc cgg 144
 Phe Val Cys Gly Asp Arg Gly Phe Leu Phe Ser Arg Pro Ala Ser Arg

35 40 45
 gtt tcc aga cgg tct aga gga atc gtg gaa gag tgc tgc ttc aga agc 192
 Val Ser Arg Arg Ser Arg Gly Ile Val Glu Glu Cys Cys Phe Arg Ser

50 55 60
 tgc gat ctg gcc ctg ctg gaa acc tac tgt gcc aca cca gcc aga tct 240
 Cys Asp Leu Ala Leu Leu Glu Thr Tyr Cys Ala Thr Pro Ala Arg Ser

65 70 75 80
 gaa ggc ggc gga gga tct ggc gga ggc gga tct aga cct gga cct aga 288
 Glu Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Arg Pro Gly Pro Arg

85 90 95
 gac gcc cag gct cac cct ggt aga cct aga gct gtg cct aca cag tgc 336

Asp Ala Gln Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys
 100 105 110

gac gtg cca cct aac agc aga ttc gac tgc gcc cct gac aag gcc atc 384
 Asp Val Pro Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile

115 120 125
 aca caa gag cag tgt gaa gcc aga ggc tgc tgc tac atc cct gcc aaa 432
 Thr Gln Glu Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys

130 135 140
 caa gga ctg cag ggc gcc cag atg gga cag cct tgg tgc ttc ttc cca 480
 Gln Gly Leu Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro

145 150 155 160
 cca tct tac ccc agc tac aag ctg gaa aac ctg agc agc tcc gag atg 528
 Pro Ser Tyr Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met

165 170 175

ggc tac acc gcc aca ctg acc aga acc aca cct aca ttc ttc ccg aag 576
 Gly Tyr Thr Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys
 180 185 190

gac atc ctg aca ctg cgg ctg gac gtg atg atg gaa acc gag aac cgg 624
 Asp Ile Leu Thr Leu Arg Leu Asp Val Met Met Glu Thr Glu Asn Arg
 195 200 205

ctg cac ttc acc atc aag gac ccc gcc aat cgg aga tac gag gtg ccc 672

Leu His Phe Thr Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu Val Pro
 210 215 220

ctg gaa aca ccc cac gtg cac tct aga gca ccc tct cca ctg tac agc 720
 Leu Glu Thr Pro His Val His Ser Arg Ala Pro Ser Pro Leu Tyr Ser
 225 230 235 240

gtg gaa ttt tcc gag gaa ccc ttc ggc gtg atc gtg cgg aga cag ctg 768
 Val Glu Phe Ser Glu Glu Pro Phe Gly Val Ile Val Arg Arg Gln Leu
 245 250 255

gat ggc aga gtg ctc ctg aat acc aca gtg gcc cct ctg ttc ttc gcc 816
 Asp Gly Arg Val Leu Leu Asn Thr Thr Val Ala Pro Leu Phe Phe Ala
 260 265 270

gac cag ttt ctg cag ctg agc acc agc ctg cct agc cag tat atc aca 864
 Asp Gln Phe Leu Gln Leu Ser Thr Ser Leu Pro Ser Gln Tyr Ile Thr
 275 280 285

ggc ctg gcc gag cat ctg agc cct ctg atg ctg agc aca tcc tgg acc 912
 Gly Leu Ala Glu His Leu Ser Pro Leu Met Leu Ser Thr Ser Trp Thr
 290 295 300

aga atc acc ctg tgg aac cgc gac ctg gct cct aca cct ggc gcc aat 960
 Arg Ile Thr Leu Trp Asn Arg Asp Leu Ala Pro Thr Pro Gly Ala Asn
 305 310 315 320

ctg tac ggc tct cac ccc ttc tac ctg gca ctg gaa gac ggt gga tct 1008

Leu Tyr Gly Ser His Pro Phe Tyr Leu Ala Leu Glu Asp Gly Gly Ser
 325 330 335

gcc cac ggt gtc ttt ctg ctg aat agc aac gcc atg gac gtg gtg ctg 1056

Ala His Gly Val Phe Leu Leu Asn Ser Asn Ala Met Asp Val Val Leu	
340 345 350	
cag ccc tct cct gca ctg tct tgg aga tct aca ggc ggc atc ctg gac	1104
Gln Pro Ser Pro Ala Leu Ser Trp Arg Ser Thr Gly Gly Ile Leu Asp	
355 360 365	
gtg tac atc ttt ctg ggc ccc gag cct aag agc gtg gtg cag cag tat	1152
Val Tyr Ile Phe Leu Gly Pro Glu Pro Lys Ser Val Val Gln Gln Tyr	
370 375 380	
ctg gac gtc gtg ggc tac ccc ttc atg cct cct tat tgg ggc ctg ggc	1200
Leu Asp Val Val Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly Leu Gly	
385 390 395 400	
ttc cac ctg tgt agg tgg ggc tac agc agc acc gcc atc acc aga cag	1248
Phe His Leu Cys Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr Arg Gln	
405 410 415	
gtg gtg gaa aac atg acc cgg gct cac ttc cca ctg gac gtg cag tgg	1296
Val Val Glu Asn Met Thr Arg Ala His Phe Pro Leu Asp Val Gln Trp	
420 425 430	
aac gac ctg gac tac atg gac agc aga cgg gac ttc acc ttc aac aag	1344
Asn Asp Leu Asp Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe Asn Lys	
435 440 445	
gac ggc ttc aga gac ttc ccc gcc atg gtg caa gag ctg cat caa ggc	1392
Asp Gly Phe Arg Asp Phe Pro Ala Met Val Gln Glu Leu His Gln Gly	
450 455 460	
gga cgg cgg tac atg atg att gtg gac cct gcc atc agc agc tct gga	1440
Gly Arg Arg Tyr Met Met Ile Val Asp Pro Ala Ile Ser Ser Ser Gly	
465 470 475 480	
cca gcc ggc agc tac aga cct tac gat gag gga ctg aga aga ggc gtg	1488
Pro Ala Gly Ser Tyr Arg Pro Tyr Asp Glu Gly Leu Arg Arg Gly Val	
485 490 495	
ttc atc acc aac gag aca ggc cag cct ctg atc ggc aaa gtg tgg cct	1536
Phe Ile Thr Asn Glu Thr Gly Gln Pro Leu Ile Gly Lys Val Trp Pro	

500	505	510	
ggc agc aca gcc ttt cca gac ttc aca aac ccc acc gct ctg gct tgg			1584
Gly Ser Thr Ala Phe Pro Asp Phe Thr Asn Pro Thr Ala Leu Ala Trp			
515	520	525	
tgg gaa gat atg gtg gcc gag ttc cac gat cag gtg ccc ttc gat ggc			1632
Trp Glu Asp Met Val Ala Glu Phe His Asp Gln Val Pro Phe Asp Gly			
530	535	540	
atg tgg atc gac atg aac gag ccc agc aac ttc atc cgg ggc agc gag			1680
Met Trp Ile Asp Met Asn Glu Pro Ser Asn Phe Ile Arg Gly Ser Glu			
545	550	555	560
gat ggc tgc ccc aac aac gaa cta gaa aat cct cct tac gtg ccc ggc			1728
Asp Gly Cys Pro Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val Pro Gly			
565	570	575	
gtt gtc ggc gga aca ctt cag gcc gct aca atc tgt gcc agc agc cat			1776
Val Val Gly Gly Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser Ser His			
580	585	590	
cag ttt ctg agc acc cac tac aac ctg cac aac ctg tac ggc ctg acc			1824
Gln Phe Leu Ser Thr His Tyr Asn Leu His Asn Leu Tyr Gly Leu Thr			
595	600	605	
gag gcc att gcc tct cat aga gcc ctg gtt aag gcc aga ggc acc cgg			1872
Glu Ala Ile Ala Ser His Arg Ala Leu Val Lys Ala Arg Gly Thr Arg			
610	615	620	
cct ttt gtg atc agc aga agc aca ttc gcc ggc cac ggc aga tac gca			1920
Pro Phe Val Ile Ser Arg Ser Thr Phe Ala Gly His Gly Arg Tyr Ala			
625	630	635	640
gga cat tgg aca ggc gac gtg tgg tct agt tgg gag cag ctg gct agc			1968
Gly His Trp Thr Gly Asp Val Trp Ser Ser Trp Glu Gln Leu Ala Ser			
645	650	655	
agc gtg cca gag atc ctg cag ttc aat ctg ctg ggc gtg cca ctc gtg			2016

Ser Val Pro Glu Ile Leu Gln Phe Asn Leu Leu Gly Val Pro Leu Val	
660 665 670	
gga gcc gac gtt tgt ggc ttc ctg ggc aac acc agc gag gaa ctg tgt	2064
Gly Ala Asp Val Cys Gly Phe Leu Gly Asn Thr Ser Glu Glu Leu Cys	
675 680 685	
gtg cgt tgg aca cag ctg ggc gcc ttc tat ccc ttc atg aga aac cac	2112
Val Arg Trp Thr Gln Leu Gly Ala Phe Tyr Pro Phe Met Arg Asn His	
690 695 700	
aac agc ctg ctg agc ctg cct caa gag ccc tac agc ttt agc gag cct	2160
Asn Ser Leu Leu Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser Glu Pro	
705 710 715 720	
gca cag cag gcc atg aga aag gcc ctg act ctg aga tac gcc ctg ctg	2208
Ala Gln Gln Ala Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala Leu Leu	
725 730 735	
cct cac ctg tac acc ctg ttt cat cag gcc cac gtg gca ggc gag aca	2256
Pro His Leu Tyr Thr Leu Phe His Gln Ala His Val Ala Gly Glu Thr	
740 745 750	
gtg gct aga cct ctg ttc ctg gaa ttt ccc aag gac agc tcc acc tgg	2304
Val Ala Arg Pro Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser Thr Trp	
755 760 765	
acc gtg gat cat cag ctg ctg tgg gga gaa gcc ctg ctg att aca cca	2352
Thr Val Asp His Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile Thr Pro	
770 775 780	
gtg ctg cag gcc gga aag gcc gaa gtg aca ggc tat ttc cct ctc ggc	2400
Val Leu Gln Ala Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro Leu Gly	
785 790 795 800	
act tgg tac gac ctg cag acc gtg cct atc gag gct ctg gga tct ctt	2448
Thr Trp Tyr Asp Leu Gln Thr Val Pro Ile Glu Ala Leu Gly Ser Leu	
805 810 815	
cct cca cct cct gcc gct cct aga gag cct gcc att cac tct gaa ggc	2496
Pro Pro Pro Pro Ala Ala Pro Arg Glu Pro Ala Ile His Ser Glu Gly	

820	825	830	
cag tgg gtt acc ctg cct gct cct ctg gac acc atc aac gtg cac ctg			2544
Gln Trp Val Thr Leu Pro Ala Pro Leu Asp Thr Ile Asn Val His Leu			
835	840	845	
aga gcc ggc tac atc atc cct ctg caa ggc cct ggc ctg acc aca acc			2592
Arg Ala Gly Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr			
850	855	860	
gaa tct aga cag cag ccc atg gca ctg gcc gtg gct ctt aca aaa ggc			2640
Glu Ser Arg Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly			
865	870	875	880
gga gag gct aga ggc gag ctg ttc tgg gat gat ggc gag agc cta gaa			2688
Gly Glu Ala Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu			
885	890	895	
gtg ctg gaa cgg ggc gct tat acc caa gtg atc ttc ctg gcc aga aac			2736
Val Leu Glu Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn			
900	905	910	
aac acc atc gtg aac gaa ctc gtg cgc gtg acc agt gaa ggt gct gga			2784
Asn Thr Ile Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly			
915	920	925	
ctg caa ctg cag aaa gtg acc gtg ctc gga gtg gcc aca gct cct cag			2832
Leu Gln Leu Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln			
930	935	940	
cag gtt ctg tct aat ggc gtg ccc gtg tcc aac ttc aca tac agc ccc			2880
Gln Val Leu Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro			
945	950	955	960
gac acc aag gtc ctg gac atc tgt gtg tcc ctg ctt atg ggc gag cag			2928
Asp Thr Lys Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln			
965	970	975	
ttc ctg gtg tcc tgg tgc tga taa			2952
Phe Leu Val Ser Trp Cys			
980			

<210> 23

<211> 982

<212> PRT

<213> Artificial sequence

<220><223> Synthetic Construct

<400> 23

Met Lys Leu Ser Leu Val Ala Ala Met Leu Leu Leu Ser Ala Ala

1 5 10 15

Arg Ala Ser Arg Thr Leu Cys Gly Gly Glu Leu Val Asp Thr Leu Gln

20 25 30

Phe Val Cys Gly Asp Arg Gly Phe Leu Phe Ser Arg Pro Ala Ser Arg

35 40 45

Val Ser Arg Arg Ser Arg Gly Ile Val Glu Glu Cys Cys Phe Arg Ser

50 55 60

Cys Asp Leu Ala Leu Leu Glu Thr Tyr Cys Ala Thr Pro Ala Arg Ser

65 70 75 80

Glu Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Arg Pro Gly Pro Arg

85 90 95

Asp Ala Gln Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys

100 105 110

Asp Val Pro Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile

115 120 125

Thr Gln Glu Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys

130 135 140

Gln Gly Leu Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro

145 150 155 160

Pro Ser Tyr Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met

165 170 175

Gly Tyr Thr Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys

180 185 190

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

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

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

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

 Leu Asp Val Val Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly Leu Gly
 385 390 395 400
 Phe His Leu Cys Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr Arg Gln
 405 410 415
 Val Val Glu Asn Met Thr Arg Ala His Phe Pro Leu Asp Val Gln Trp
 420 425 430
 Asn Asp Leu Asp Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe Asn Lys
 435 440 445

Asp Gly Phe Arg Asp Phe Pro Ala Met Val Gln Glu Leu His Gln Gly
450 455 460

Gly Arg Arg Tyr Met Met Ile Val Asp Pro Ala Ile Ser Ser Ser Gly
465 470 475 480

Pro Ala Gly Ser Tyr Arg Pro Tyr Asp Glu Gly Leu Arg Arg Gly Val
485 490 495

Phe Ile Thr Asn Glu Thr Gly Gln Pro Leu Ile Gly Lys Val Trp Pro
500 505 510

Gly Ser Thr Ala Phe Pro Asp Phe Thr Asn Pro Thr Ala Leu Ala Trp
515 520 525

Trp Glu Asp Met Val Ala Glu Phe His Asp Gln Val Pro Phe Asp Gly
530 535 540

Met Trp Ile Asp Met Asn Glu Pro Ser Asn Phe Ile Arg Gly Ser Glu
545 550 555 560

Asp Gly Cys Pro Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val Pro Gly
565 570 575

Val Val Gly Gly Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser Ser His
580 585 590

Gln Phe Leu Ser Thr His Tyr Asn Leu His Asn Leu Tyr Gly Leu Thr
595 600 605

Glu Ala Ile Ala Ser His Arg Ala Leu Val Lys Ala Arg Gly Thr Arg
610 615 620

Pro Phe Val Ile Ser Arg Ser Thr Phe Ala Gly His Gly Arg Tyr Ala
625 630 635 640

Gly His Trp Thr Gly Asp Val Trp Ser Ser Trp Glu Gln Leu Ala Ser
645 650 655

Ser Val Pro Glu Ile Leu Gln Phe Asn Leu Leu Gly Val Pro Leu Val
660 665 670

Gly Ala Asp Val Cys Gly Phe Leu Gly Asn Thr Ser Glu Glu Leu Cys
675 680 685

Val Arg Trp Thr Gln Leu Gly Ala Phe Tyr Pro Phe Met Arg Asn His

690	695	700
Asn Ser Leu Leu Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser Glu Pro		
705	710	715
Ala Gln Gln Ala Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala Leu Leu		
	725	730
Pro His Leu Tyr Thr Leu Phe His Gln Ala His Val Ala Gly Glu Thr		
	740	745
Val Ala Arg Pro Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser Thr Trp		
	755	760
		765
Thr Val Asp His Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile Thr Pro		
	770	775
		780
Val Leu Gln Ala Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro Leu Gly		
785	790	795
		800
Thr Trp Tyr Asp Leu Gln Thr Val Pro Ile Glu Ala Leu Gly Ser Leu		
	805	810
		815
Pro Pro Pro Pro Ala Ala Pro Arg Glu Pro Ala Ile His Ser Glu Gly		
	820	825
		830
Gln Trp Val Thr Leu Pro Ala Pro Leu Asp Thr Ile Asn Val His Leu		
	835	840
		845
Arg Ala Gly Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr		
	850	855
		860
Glu Ser Arg Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly		
865	870	875
		880
Gly Glu Ala Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu		
	885	890
		895
Val Leu Glu Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn		
	900	905
		910
Asn Thr Ile Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly		
	915	920
		925
Leu Gln Leu Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln		
	930	935
		940

Gln Val Leu Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro
 945 950 955 960

Asp Thr Lys Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln
 965 970 975

Phe Leu Val Ser Trp Cys
 980

<210> 24

<211> 2952

<212> DNA

<213> Artificial sequence

<220><223> synthetic construct

<220><221> CDS

<222> (1)..(2952)

<220><221> misc_feature

<222> (1)..(270)

<223> BiP-vIGF peptide

<220><221> misc_feature

<222> (1)..(270)

<223> BiP signal peptide + vIGF2+2GS extension

<220><221> misc_feature

<222> (271)..(2952)

<223> hGAA 61-952 V780 DNA

<220><221> misc_feature

<222> (2428)..(2430)

<223> codon for hGAA 780 Valine

<400> 24

atg aag ctg tct ctg gtg gct gct atg ctg ctg ctc ctg tct gcc gcc 48

Met Lys Leu Ser Leu Val Ala Ala Met Leu Leu Leu Leu Ser Ala Ala

1 5 10 15

aga gcc agc aga aca ctt tgt ggc gga gag ctg gtg gac acc ctg cag 96

Arg Ala Ser Arg Thr Leu Cys Gly Gly Glu Leu Val Asp Thr Leu Gln

20 25 30

ttt gtg tgt ggc gac aga ggc ttc ctg ttc agc aga cct gcc agc cgg	144
Phe Val Cys Gly Asp Arg Gly Phe Leu Phe Ser Arg Pro Ala Ser Arg	
35 40 45	
gtt tcc aga cgg tct aga gga atc gtg gaa gag tgc tgc ttc aga agc	192
Val Ser Arg Arg Ser Arg Gly Ile Val Glu Glu Cys Cys Phe Arg Ser	
50 55 60	
tgc gat ctg gcc ctg ctg gaa acc tac tgt gcc aca cca gcc aga tct	240
Cys Asp Leu Ala Leu Leu Glu Thr Tyr Cys Ala Thr Pro Ala Arg Ser	
65 70 75 80	
gaa ggc ggc gga gga tct ggc gga ggc gga tct aga cct gga cct aga	288
Glu Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Arg Pro Gly Pro Arg	
85 90 95	
gac gcc cag gct cac cct ggt aga cct aga gct gtg cct aca cag tgc	336
Asp Ala Gln Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys	
100 105 110	
gac gtg cca cct aac agc aga ttc gac tgc gcc cct gac aag gcc atc	384
Asp Val Pro Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile	
115 120 125	
aca caa gag cag tgt gaa gcc aga ggc tgc tgc tac atc cct gcc aaa	432
Thr Gln Glu Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys	
130 135 140	
caa gga ctg cag ggc gcc cag atg gga cag cct tgg tgc ttc ttc cca	480
Gln Gly Leu Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro	
145 150 155 160	
cca tct tac ccc agc tac aag ctg gaa aac ctg agc agc tcc gag atg	528
Pro Ser Tyr Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met	
165 170 175	
ggc tac acc gcc aca ctg acc aga acc aca cct aca ttc ttc ccg aag	576
Gly Tyr Thr Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys	
180 185 190	
gac atc ctg aca ctg cgg ctg gac gtg atg atg gaa acc gag aac cgg	624

Asp Ile Leu Thr Leu Arg Leu Asp Val Met Met Glu Thr Glu Asn Arg	
195 200 205	
ctg cac ttc acc atc aag gac ccc gcc aat cgg aga tac gag gtg ccc	672
Leu His Phe Thr Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu Val Pro	
210 215 220	
ctg gaa aca ccc cac gtg cac tct aga gca ccc tct cca ctg tac agc	720
Leu Glu Thr Pro His Val His Ser Arg Ala Pro Ser Pro Leu Tyr Ser	
225 230 235 240	
gtg gaa ttt tcc gag gaa ccc ttc ggc gtg atc gtg cgg aga cag ctg	768
Val Glu Phe Ser Glu Glu Pro Phe Gly Val Ile Val Arg Arg Gln Leu	
245 250 255	
gat ggc aga gtg ctc ctg aat acc aca gtg gcc cct ctg ttc ttc gcc	816
Asp Gly Arg Val Leu Leu Asn Thr Thr Val Ala Pro Leu Phe Phe Ala	
260 265 270	
gac cag ttt ctg cag ctg agc acc agc ctg cct agc cag tat atc aca	864
Asp Gln Phe Leu Gln Leu Ser Thr Ser Leu Pro Ser Gln Tyr Ile Thr	
275 280 285	
ggc ctg gcc gag cat ctg agc cct ctg atg ctg agc aca tcc tgg acc	912
Gly Leu Ala Glu His Leu Ser Pro Leu Met Leu Ser Thr Ser Trp Thr	
290 295 300	
aga atc acc ctg tgg aac cgc gac ctg gct cct aca cct ggc gcc aat	960
Arg Ile Thr Leu Trp Asn Arg Asp Leu Ala Pro Thr Pro Gly Ala Asn	
305 310 315 320	
ctg tac ggc tct cac ccc ttc tac ctg gca ctg gaa gac ggt gga tct	1008
Leu Tyr Gly Ser His Pro Phe Tyr Leu Ala Leu Glu Asp Gly Gly Ser	
325 330 335	
gcc cac ggt gtc ttt ctg ctg aat agc aac gcc atg gac gtg gtg ctg	1056
Ala His Gly Val Phe Leu Leu Asn Ser Asn Ala Met Asp Val Val Leu	
340 345 350	
cag ccc tct cct gca ctg tct tgg aga tct aca ggc ggc atc ctg gac	1104

Gln Pro Ser Pro Ala Leu Ser Trp Arg Ser Thr Gly Gly Ile Leu Asp

355	360	365	
gtg tac atc ttt ctg ggc ccc gag cct aag agc gtg gtg cag cag tat	1152		
Val Tyr Ile Phe Leu Gly Pro Glu Pro Lys Ser Val Val Gln Gln Tyr			
370	375	380	
ctg gac gtc gtg ggc tac ccc ttc atg cct cct tat tgg ggc ctg ggc	1200		
Leu Asp Val Val Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly Leu Gly			
385	390	395	400
ttc cac ctg tgt agg tgg ggc tac agc agc acc gcc atc acc aga cag	1248		
Phe His Leu Cys Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr Arg Gln			
405	410	415	
gtg gtg gaa aac atg acc cgg gct cac ttc cca ctg gac gtg cag tgg	1296		
Val Val Glu Asn Met Thr Arg Ala His Phe Pro Leu Asp Val Gln Trp			
420	425	430	
aac gac ctg gac tac atg gac agc aga cgg gac ttc acc ttc aac aag	1344		
Asn Asp Leu Asp Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe Asn Lys			
435	440	445	
gac ggc ttc aga gac ttc ccc gcc atg gtg caa gag ctg cat caa ggc	1392		
Asp Gly Phe Arg Asp Phe Pro Ala Met Val Gln Glu Leu His Gln Gly			
450	455	460	
gga cgg cgg tac atg atg att gtg gac cct gcc atc agc agc tct gga	1440		
Gly Arg Arg Tyr Met Met Ile Val Asp Pro Ala Ile Ser Ser Ser Gly			
465	470	475	480
cca gcc ggc agc tac aga cct tac gat gag gga ctg aga aga ggc gtg	1488		
Pro Ala Gly Ser Tyr Arg Pro Tyr Asp Glu Gly Leu Arg Arg Gly Val			
485	490	495	
ttc atc acc aac gag aca ggc cag cct ctg atc ggc aaa gtg tgg cct	1536		
Phe Ile Thr Asn Glu Thr Gly Gln Pro Leu Ile Gly Lys Val Trp Pro			
500	505	510	

ggc agc aca gcc ttt cca gac ttc aca aac ccc acc gct ctg gct tgg	1584
Gly Ser Thr Ala Phe Pro Asp Phe Thr Asn Pro Thr Ala Leu Ala Trp	
515 520 525	
tgg gaa gat atg gtg gcc gag ttc cac gat cag gtg ccc ttc gat ggc	1632
Trp Glu Asp Met Val Ala Glu Phe His Asp Gln Val Pro Phe Asp Gly	
530 535 540	
atg tgg atc gac atg aac gag ccc agc aac ttc atc cgg ggc agc gag	1680
Met Trp Ile Asp Met Asn Glu Pro Ser Asn Phe Ile Arg Gly Ser Glu	
545 550 555 560	
gat ggc tgc ccc aac aac gaa cta gaa aat cct cct tac gtg ccc ggc	1728
Asp Gly Cys Pro Asn Asn Glu Leu Glu Asn Pro Pro Tyr Val Pro Gly	
565 570 575	
gtt gtc ggc gga aca ctt cag gcc gct aca atc tgt gcc agc agc cat	1776
Val Val Gly Gly Thr Leu Gln Ala Ala Thr Ile Cys Ala Ser Ser His	
580 585 590	
cag ttt ctg agc acc cac tac aac ctg cac aac ctg tac ggc ctg acc	1824
Gln Phe Leu Ser Thr His Tyr Asn Leu His Asn Leu Tyr Gly Leu Thr	
595 600 605	
gag gcc att gcc tct cat aga gcc ctg gtt aag gcc aga ggc acc cgg	1872
Glu Ala Ile Ala Ser His Arg Ala Leu Val Lys Ala Arg Gly Thr Arg	
610 615 620	
cct ttt gtg atc agc aga agc aca ttc gcc ggc cac ggc aga tac gca	1920
Pro Phe Val Ile Ser Arg Ser Thr Phe Ala Gly His Gly Arg Tyr Ala	
625 630 635 640	
gga cat tgg aca ggc gac gtg tgg tct agt tgg gag cag ctg gct agc	1968
Gly His Trp Thr Gly Asp Val Trp Ser Ser Trp Glu Gln Leu Ala Ser	
645 650 655	
agc gtg cca gag atc ctg cag ttc aat ctg ctg ggc gtg cca ctc gtg	2016
Ser Val Pro Glu Ile Leu Gln Phe Asn Leu Leu Gly Val Pro Leu Val	
660 665 670	
gga gcc gac gtt tgt ggc ttc ctg ggc aac acc agc gag gaa ctg tgt	2064

Gly	Ala	Asp	Val	Cys	Gly	Phe	Leu	Gly	Asn	Thr	Ser	Glu	Glu	Leu	Cys	
675				680				685								
gtg	cgt	tgg	aca	cag	ctg	ggc	gcc	ttc	tat	ccc	ttc	atg	aga	aac	cac	2112
Val	Arg	Trp	Thr	Gln	Leu	Gly	Ala	Phe	Tyr	Pro	Phe	Met	Arg	Asn	His	
690				695				700								
aac	agc	ctg	ctg	agc	ctg	cct	caa	gag	ccc	tac	agc	ttt	agc	gag	cct	2160
Asn	Ser	Leu	Leu	Ser	Leu	Pro	Gln	Glu	Pro	Tyr	Ser	Phe	Ser	Glu	Pro	
705				710				715				720				
gca	cag	cag	gcc	atg	aga	aag	gcc	ctg	act	ctg	aga	tac	gcc	ctg	ctg	2208
Ala	Gln	Gln	Ala	Met	Arg	Lys	Ala	Leu	Thr	Leu	Arg	Tyr	Ala	Leu	Leu	
725				730				735								
cct	cac	ctg	tac	acc	ctg	ttt	cat	cag	gcc	cac	gtg	gca	ggc	gag	aca	2256
Pro	His	Leu	Tyr	Thr	Leu	Phe	His	Gln	Ala	His	Val	Ala	Gly	Glu	Thr	
740				745				750								
gtg	gct	aga	cct	ctg	ttc	ctg	gaa	ttt	ccc	aag	gac	agc	tcc	acc	tgg	2304
Val	Ala	Arg	Pro	Leu	Phe	Leu	Glu	Phe	Pro	Lys	Asp	Ser	Ser	Thr	Trp	
755				760				765								
acc	gtg	gat	cat	cag	ctg	ctg	tgg	gga	gaa	gcc	ctg	ctg	att	aca	cca	2352
Thr	Val	Asp	His	Gln	Leu	Leu	Trp	Gly	Glu	Ala	Leu	Leu	Ile	Thr	Pro	
770				775				780								
gtg	ctg	cag	gcc	gga	aag	gcc	gaa	gtg	aca	ggc	tat	ttc	cct	ctc	ggc	2400
Val	Leu	Gln	Ala	Gly	Lys	Ala	Glu	Val	Thr	Gly	Tyr	Phe	Pro	Leu	Gly	
785				790				795				800				
act	tgg	tac	gac	ctg	cag	acc	gtg	cct	atc	gag	gct	ctg	gga	tct	ctt	2448
Thr	Trp	Tyr	Asp	Leu	Gln	Thr	Val	Pro	Ile	Glu	Ala	Leu	Gly	Ser	Leu	
805				810				815								
cct	cca	cct	cct	gcc	gct	cct	aga	gag	cct	gcc	att	cac	tct	gaa	ggc	2496
Pro	Pro	Pro	Pro	Ala	Ala	Pro	Arg	Glu	Pro	Ala	Ile	His	Ser	Glu	Gly	
820				825				830								
cag	tgg	gtt	acc	ctg	cct	gct	cct	ctg	gac	acc	atc	aac	gtg	cac	ctg	2544
Gln	Trp	Val	Thr	Leu	Pro	Ala	Pro	Leu	Asp	Thr	Ile	Asn	Val	His	Leu	

835	840	845	
aga gcc ggc tac atc atc cct ctg caa ggc cct ggc ctg acc aca acc			2592
Arg Ala Gly Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr			
850	855	860	
gaa tct aga cag cag ccc atg gca ctg gcc gtg gct ctt aca aaa ggc			2640
Glu Ser Arg Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly			
865	870	875	880
gga gag gct aga ggc gag ctg ttc tgg gat gat ggc gag agc cta gaa			2688
Gly Glu Ala Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu			
885	890	895	
gtg ctg gaa cgg ggc gct tat acc caa gtg atc ttc ctg gcc aga aac			2736
Val Leu Glu Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn			
900	905	910	
aac acc atc gtg aac gaa ctc gtg cgc gtg acc agt gaa ggt gct gga			2784
Asn Thr Ile Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly			
915	920	925	
ctg caa ctg cag aaa gtg acc gtg ctc gga gtg gcc aca gct cct cag			2832
Leu Gln Leu Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln			
930	935	940	
cag gtt ctg tct aat ggc gtg ccc gtg tcc aac ttc aca tac agc ccc			2880
Gln Val Leu Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro			
945	950	955	960
gac acc aag gtc ctg gac atc tgt gtg tcc ctg ctt atg ggc gag cag			2928
Asp Thr Lys Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln			
965	970	975	
ttc ctg gtg tcc tgg tgc tga taa			2952
Phe Leu Val Ser Trp Cys			
980			
<210> 25			
<211> 982			
<212> PRT			

<213> Artificial sequence

<220><223> Synthetic Construct

<400> 25

Met Lys Leu Ser Leu Val Ala Ala Met Leu Leu Leu Ser Ala Ala

1 5 10 15

Arg Ala Ser Arg Thr Leu Cys Gly Gly Glu Leu Val Asp Thr Leu Gln

20 25 30

Phe Val Cys Gly Asp Arg Gly Phe Leu Phe Ser Arg Pro Ala Ser Arg

35 40 45

Val Ser Arg Arg Ser Arg Gly Ile Val Glu Glu Cys Cys Phe Arg Ser

50 55 60

Cys Asp Leu Ala Leu Leu Glu Thr Tyr Cys Ala Thr Pro Ala Arg Ser

65 70 75 80

Glu Gly Gly Gly Gly Ser Gly Gly Gly Gly Ser Arg Pro Gly Pro Arg

85 90 95

Asp Ala Gln Ala His Pro Gly Arg Pro Arg Ala Val Pro Thr Gln Cys

100 105 110

Asp Val Pro Pro Asn Ser Arg Phe Asp Cys Ala Pro Asp Lys Ala Ile

115 120 125

Thr Gln Glu Gln Cys Glu Ala Arg Gly Cys Cys Tyr Ile Pro Ala Lys

130 135 140

Gln Gly Leu Gln Gly Ala Gln Met Gly Gln Pro Trp Cys Phe Phe Pro

145 150 155 160

Pro Ser Tyr Pro Ser Tyr Lys Leu Glu Asn Leu Ser Ser Ser Glu Met

165 170 175

Gly Tyr Thr Ala Thr Leu Thr Arg Thr Thr Pro Thr Phe Phe Pro Lys

180 185 190

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

195 200 205

Leu His Phe Thr Ile Lys Asp Pro Ala Asn Arg Arg Tyr Glu Val Pro

210 215 220

Leu Glu Thr Pro His Val His Ser Arg Ala Pro Ser Pro Leu Tyr Ser
 225 230 235 240
 Val Glu Phe Ser Glu Glu Pro Phe Gly Val Ile Val Arg Arg Gln Leu
 245 250 255

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

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

 Leu Asp Val Val Gly Tyr Pro Phe Met Pro Pro Tyr Trp Gly Leu Gly
 385 390 395 400
 Phe His Leu Cys Arg Trp Gly Tyr Ser Ser Thr Ala Ile Thr Arg Gln
 405 410 415
 Val Val Glu Asn Met Thr Arg Ala His Phe Pro Leu Asp Val Gln Trp
 420 425 430
 Asn Asp Leu Asp Tyr Met Asp Ser Arg Arg Asp Phe Thr Phe Asn Lys
 435 440 445

 Asp Gly Phe Arg Asp Phe Pro Ala Met Val Gln Glu Leu His Gln Gly
 450 455 460
 Gly Arg Arg Tyr Met Met Ile Val Asp Pro Ala Ile Ser Ser Ser Gly

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

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

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

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

 Asn Ser Leu Leu Ser Leu Pro Gln Glu Pro Tyr Ser Phe Ser Glu Pro
 705 710 715 720

Ala Gln Gln Ala Met Arg Lys Ala Leu Thr Leu Arg Tyr Ala Leu Leu
725 730 735

Pro His Leu Tyr Thr Leu Phe His Gln Ala His Val Ala Gly Glu Thr
740 745 750

Val Ala Arg Pro Leu Phe Leu Glu Phe Pro Lys Asp Ser Ser Thr Trp
755 760 765

Thr Val Asp His Gln Leu Leu Trp Gly Glu Ala Leu Leu Ile Thr Pro
770 775 780

Val Leu Gln Ala Gly Lys Ala Glu Val Thr Gly Tyr Phe Pro Leu Gly
785 790 795 800

Thr Trp Tyr Asp Leu Gln Thr Val Pro Ile Glu Ala Leu Gly Ser Leu
805 810 815

Pro Pro Pro Pro Ala Ala Pro Arg Glu Pro Ala Ile His Ser Glu Gly
820 825 830

Gln Trp Val Thr Leu Pro Ala Pro Leu Asp Thr Ile Asn Val His Leu
835 840 845

Arg Ala Gly Tyr Ile Ile Pro Leu Gln Gly Pro Gly Leu Thr Thr Thr
850 855 860

Glu Ser Arg Gln Gln Pro Met Ala Leu Ala Val Ala Leu Thr Lys Gly
865 870 875 880

Gly Glu Ala Arg Gly Glu Leu Phe Trp Asp Asp Gly Glu Ser Leu Glu
885 890 895

Val Leu Glu Arg Gly Ala Tyr Thr Gln Val Ile Phe Leu Ala Arg Asn
900 905 910

Asn Thr Ile Val Asn Glu Leu Val Arg Val Thr Ser Glu Gly Ala Gly
915 920 925

Leu Gln Leu Gln Lys Val Thr Val Leu Gly Val Ala Thr Ala Pro Gln
930 935 940

Gln Val Leu Ser Asn Gly Val Pro Val Ser Asn Phe Thr Tyr Ser Pro
945 950 955 960

Asp Thr Lys Val Leu Asp Ile Cys Val Ser Leu Leu Met Gly Glu Gln

965	970	975
Phe Leu Val Ser Trp Cys		
980		
<210> 26		
<211> 22		
<212> DNA		
<213> Artificial Sequence		
<220><223> miRNA target sequence		
<400> 26		
agtgaattct accagtgcc		22
<210> 27		
<211> 24		
<212> DNA		
<213> Artificial Sequence		
<220><223> miRNA target sequence		
<400> 27		
agtgtgagtt ctaccattgc caaa		24
<210> 28		
<211> 6		
<212> PRT		
<213> Homo sapiens		
<400> 28		
Trp Ile Asp Met Asn Glu		
1	5	
<210> 29		
<211> 7		
<212> PRT		
<213> Homo sapiens		
<400> 29		
Thr Val Pro Ile Glu Ala Leu		
1	5	
<210> 30		
<211> 9		
<212> PRT		

<213> Homo sapiens

<400> 30

Gln Thr Val Pro Ile Glu Ala Leu Gly

1 5

<210> 31

<211> 25

<212> PRT

<213> Homo sapiens

<400> 31

Pro Leu Gly Thr Trp Tyr Asp Leu Gln Thr Val Pro Ile Glu Ala Leu

1 5 10 15

Gly Ser Leu Pro Pro Pro Pro Ala Ala

20 25