



공개특허 10-2020-0095462



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2020-0095462  
(43) 공개일자 2020년08월10일

- (51) 국제특허분류(Int. Cl.)  
*C12N 15/86* (2006.01) *A61K 35/12* (2020.01)  
*A61K 48/00* (2006.01) *A61P 7/06* (2006.01)  
*C07K 14/805* (2006.01) *C12N 15/90* (2006.01)
- (52) CPC특허분류  
*C12N 15/86* (2013.01)  
*A61K 35/12* (2013.01)
- (21) 출원번호 10-2020-7012240
- (22) 출원일자(국제) 2018년10월17일  
심사청구일자 없음
- (85) 번역문제출일자 2020년04월27일
- (86) 국제출원번호 PCT/US2018/056271
- (87) 국제공개번호 WO 2019/079437  
국제공개일자 2019년04월25일
- (30) 우선권주장  
62/574,163 2017년10월18일 미국(US)  
62/621,102 2018년01월24일 미국(US)

- (71) 출원인  
시티 오브 호프  
미국 91010-3000 캘리포니아주 두아르테 이스트  
두아르테 로드 1500  
호플로지 메디슨, 임크.  
미국 매사추세츠, 베드포드 원 패트리어트 파크  
(우편번호 01730)
- (72) 발명자  
채터지 사스와티  
미국, 캘리포니아 91001, 엘더디나, 614 처치 캐  
니언 플레이스  
웡 카메하메하 케이.  
미국, 캘리포니아 91001, 엘터디나, 614 처치 캐  
니언 플레이스  
(뒷면에 계속)
- (74) 대리인  
특허법인한얼

전체 청구항 수 : 총 134 항

(54) 발명의 명칭 HBB 유전자 기능 회복을 위한 아데노-연관 바이러스 조성물 및 이의 사용 방법

### (57) 요 약

베타 글로빈 유전자(HBB) 유전자에서 돌연변이를 교정하기 위한 아데노-연관 바이러스(AAV: adeno-associated virus) 조성물 및 이를 사용하여 세포에서 HBB 유전자 돌연변이를 교정하는 방법이 본원에 제공된다. 또한 아데노-연관 바이러스 조성물 제조용 패키징 시스템이 제공된다.

(52) CPC특허분류

*A61K 48/005* (2013.01)

*A61P 7/06* (2018.01)

*C07K 14/805* (2013.01)

*C12N 15/907* (2013.01)

*C12N 2750/14143* (2013.01)

(72) 발명자

**벤하지살라 마르와**

미국, 캘리포니아 91010, 두아르테, 아파트먼트  
씨, 1412 두아르테 로드

**스미스 로라 제인**

미국, 메사추세츠 01886, 웨스트포드, 14131 베이  
드라이브

**시무어 알버트 반스**

미국, 메사추세츠 01581, 웨스트보로우, 1 브릭야  
드 레인

**라이트 제이슨 보케**

미국, 메사추세츠 01742, 콩코드, 57 위츠 앤드 로  
드

**맥위겐 제임스 안토니**

미국, 메사추세츠 02474, 알링턴, 유닛 에이, 15  
비컨 스트리트

**돌라이브 세례나 니콜**

미국, 메사추세츠 02453, 월섬, 유닛 1, 19 에버렛  
스트리트

**세인트 마틴 티아 바보발**

미국, 메사추세츠 01462, 루넨버그, 488 홀만 스트  
리트

**프라우트 제이미**

미국, 뉴 햄프셔 03051, 허드슨, 3 레이븐 드라이  
브

## 명세서

### 청구범위

#### 청구항 1

세포 내의 베타 글로빈(HBB) 유전자에서 돌연변이의 교정 방법으로서, 상기 방법은 상기 세포에 하기를 포함하는 복제-결합 아데노-연관 바이러스(AAV: adeno-associated virus)로 형질도입하는 단계를 포함하며:

- a) AAV Clade F 캡시드 단백질을 포함하는 AAV 캡시드; 및
- b) (i) 표적 유전자에서 표적 좌위 편집용 편집 요소; (ii) 상기 표적 좌위에 대한 제1 게놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암 뉴클레오티드 서열 5'; 및 (iii) 상기 표적 좌위에 대한 제2 게놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3'을 포함하는 교정 게놈,

상기 세포는 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 공동-형질도입하거나 공동-투여하지 않고 형질도입되는, 방법.

#### 청구항 2

제1항에 있어서, 상기 세포는 다능성 줄기 세포인 방법.

#### 청구항 3

제1항에 있어서, 상기 세포는 조혈 줄기 세포인 방법.

#### 청구항 4

제1항에 있어서, 상기 세포는 CD34<sup>+</sup> 조혈 줄기 세포인 방법.

#### 청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 상기 세포는 포유류 대상체 내에 있으며, 상기 AAV는 상기 대상체에서 상기 세포를 형질도입하는 데 유효한 양으로 상기 대상체에게 투여되는, 방법.

#### 청구항 6

HBB 유전자 돌연변이와 연관된 질환 또는 장애를 갖는 대상체의 치료 방법으로서, 상기 방법은:

- a) 생체 외에서 상기 대상체로부터 적혈구 전구 세포에 하기를 포함하는 복제-결합 AAV로 형질도입하여, 정확한 HBB 유전자를 갖는 형질도입된 세포를 수득하는 단계

(i) AAV Clade F 캡시드 단백질을 포함하는 AAV 캡시드; 및

(ii) 하기를 포함하는 교정 게놈: (A) 표적 유전자에서 표적 좌위 편집용 편집 요소; (B) 상기 표적 좌위에 대한 제1 게놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암 뉴클레오티드 서열 5'; 및 (C) 상기 표적 좌위에 대한 제2 게놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3', 및;

b) 상기 형질도입된 세포를 상기 대상체에게 투여하는 단계를 포함하며,

상기 세포는 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 공동-형질도입하지 않고 형질도입되는, 방법.

#### 청구항 7

제6항에 있어서, 상기 적혈구 전구세포는 다능성 줄기 세포인 방법.

#### 청구항 8

제6항에 있어서, 상기 적혈구 전구 세포는 다능성 줄기 세포인 방법.

#### 청구항 9

제6항에 있어서, 상기 적혈구 전구 세포는 CD34<sup>+</sup> 조혈 줄기 세포인 방법.

#### 청구항 10

HBB 유전자 돌연변이와 연관된 질환 또는 장애를 갖는 대상체의 치료 방법으로서, 상기 방법은 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 공동-투여하지 않고, 하기를 포함하는 복제-결합 AAV의 유효량을 대상체에게 투여하는 단계를 포함하는, 방법

a) AAV Clade F 캡시드 단백질을 포함하는 AAV 캡시드; 및

b) (i) 표적 유전자에서 표적 좌위 편집용 편집 요소; (ii) 상기 표적 좌위에 대한 제1 게놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암 뉴클레오티드 서열 5'; 및 (iii) 상기 표적 좌위에 대한 제2 게놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3'을 포함하는 교정 게놈.

#### 청구항 11

제6항 내지 제10항 중 어느 한 항에 있어서, 상기 질환 또는 장애는 지중해빈혈증 또는 겹상세포질환인 방법.

#### 청구항 12

제6항 내지 제10항 중 어느 한 항에 있어서, 상기 대상체는 인간 대상체인 방법.

#### 청구항 13

선행하는 항들 중 어느 한 항에 있어서, 상기 표적 유전자는 상기 HBB 유전자인 방법.

#### 청구항 14

선행하는 항들 중 어느 한 항에 있어서, 상기 표적 좌위는 상기 HBB 유전자에서 뉴클레오티드 점 돌연변이, 삽입, 또는 결실에 있는 방법.

#### 청구항 15

제14항에 있어서, 상기 HBB 유전자에서 상기 뉴클레오티드 점 돌연변이 또는 결실은 위치 -87에서 G, 위치 -31에서 G, 위치 -30에서 A, 위치 -29에서 G, 위치 -28에서 G, 위치 -10에서 T, 위치 1에서 C, 위치 1에서 A, 위치 2에서 G, 위치 17 및 18에서 C 및 T의 결실, 위치 19에서 A, 위치 20에서 A의 결실, 위치 20에서 T, 위치 25 및 26에서 A 및 A의 결실, 위치 26 이후 G의 추가, 위치 47에서 A, 위치 48에서 A, 위치 51에서 C의 결실, 위치 52에서 A, 위치 58에서 G, 위치 59에서 G, 위치 79에서 A, 위치 82에서 T, 위치 84 이후 C의 추가, 위치 93에서 T, 위치 93에서 A, 위치 97에서 C, 위치 98에서 C, 위치 202에서 G, 위치 208에서 G, 위치 222에서 C, 위치 241 또는 242에서 T의 결실, 위치 254 내지 257에서 T 및 T 및 C 및 T의 결실, 위치 260에서 T, 위치 264 또는 265에서 C의 결실, 위치 343 이후 A의 추가, 위치 399 및 400에서 G 및 T의 결실, 위치 401에서 T, 위치 417 이후 A의 추가, 위치 446에서 A, 위치 1099에서 T, 위치 1293에서 A, 1344에서 T로 이루어진 군으로부터 선택되는 방법.

#### 청구항 16

제14항 또는 제15항에 있어서, 상기 편집 요소는 돌연변이에 상응하는 야생형 HBB 유전자의 일부를 포함하는 방법.

#### 청구항 17

제1항 내지 제16항 중 어느 한 항에 있어서, 상기 편집 요소는 HBB 유전자의 하나 이상의 엑손의 코딩 영역을 포함하는 방법.

#### 청구항 18

제1항 내지 제16항 중 어느 한 항에 있어서, 상기 편집 요소는 엑손 1의 코딩 영역, 전체 인트론 1, 전체 엑손 2, 전체 인트론 2, 및 엑손 3의 코딩 영역을 포함하는 HBB 유전자의 일부를 포함하는 방법.

**청구항 19**

제17항 또는 제18항에 있어서, 상기 코딩 영역은 상기 코딩 영역은 상기 야생형 HBB 유전자의 상응하는 엑손에 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만으로 동일하게 침묵(silently) 변경되는 방법.

**청구항 20**

제19항에 있어서, 상기 편집 요소는 서열 번호: 43 내지 46 및 105 내지 107로 이루어진 군으로부터 선택된 뉴클레오티드 서열 중 적어도 하나를 포함하는 방법.

**청구항 21**

제1항 내지 제12항 중 어느 한 항에 있어서, 상기 표적 좌위는 AAVS1인 방법.

**청구항 22**

제1항 내지 제16항 및 제21항 중 어느 한 항에 있어서, 상기 편집 요소는 서열 번호: 48 또는 이의 일부를 인코딩하는 뉴클레오티드 서열을 포함하는 방법.

**청구항 23**

제22항에 있어서, 상기 서열 번호: 48을 인코딩하는 서열은 서열 번호: 27의 뉴클레오티드 4 내지 444로 구성되는 방법.

**청구항 24**

제22항에 있어서, 상기 서열 번호: 48을 인코딩하는 서열은 서열 번호: 27의 뉴클레오티드 4 내지 444에 70%, 75%, 80%, 85%, 또는 90% 미만으로 동일하게 침묵 변경되는 방법.

**청구항 25**

제24항에 있어서, 상기 서열 번호: 48을 인코딩하는 서열은 서열 번호: 47 또는 100의 서열로 구성되는 방법.

**청구항 26**

제22항에 있어서, 상기 서열 번호: 48을 인코딩하는 서열은 상기 HBB 유전자의 스타퍼(stuffer)-삽입된 코딩 서열로 구성되는 방법.

**청구항 27**

제22항 내지 제26항 중 어느 한 항에 있어서, 상기 표적 좌위는 상기 표적 유전자의 뉴클레오티드 3 및 뉴클레오티드 4 사이의 뉴클레오티드간 결합이며, 이에 의하여 상기 표적 좌위에의 상기 편집 요소의 통합으로 인해 상기 표적 유전자의 시작 코돈으로 시작하는 HBB 코딩 서열 또는 스타퍼-삽입된 코딩 서열을 포함하는 상기 표적 좌위를 생성하게 하는 방법.

**청구항 28**

제22항 내지 제26항 중 어느 한 항에 있어서, 상기 편집 요소는 5' 내지 3' 시작 코돈 및 서열 번호: 48을 인코딩하는 뉴클레오티드 서열로 이루어진 HBB 코딩 서열 또는 스타퍼-삽입된 코딩 서열, 또는 상기 HBB 코딩 서열 또는 스타퍼-삽입된 코딩 서열의 일부를 포함하는 방법.

**청구항 29**

제28항에 있어서, 상기 표적 좌위는 상기 표적 유전자의 인트론 내에 있으며, 상기 편집 요소는 5' 내지 3' 스플라이스 수용체 부위, 리보솜 스키핑 요소(ribosomal skipping element), 및 HBB 코딩 서열 또는 스타퍼-삽입된 코딩 서열을 포함하는 방법.

**청구항 30**

제29항에 있어서, 상기 표적 좌위는 상기 HBB 유전자의 인트론 1 내에 있는 방법.

#### 청구항 31

제28항에 있어서, 상기 표적 좌위는 상기 표적 유전자의 코딩 뉴클레오티드에 인접하여 3'이고, 상기 편집 요소는 5' 내지 3' 리보솜 스키핑 요소 및 HBB 코딩 서열 또는 스터파-삽입된 코딩 서열을 포함하는 방법.

#### 청구항 32

제31항에 있어서, 상기 표적 좌위는 야생형 표적 유전자 또는 돌연변이체 표적 유전자의 상응하는 뉴클레오티드의 정지 코돈인 방법.

#### 청구항 33

제22항 내지 제32항 중 어느 한 항에 있어서, 상기 편집 요소는 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열에 외인성 폴리아데닐화 서열 3'을 더 포함하는 방법.

#### 청구항 34

선행하는 항들 중 어느 한 항에 있어서, 상기 편집 요소는 상기 표적 유전자에 존재하지 않는 제한 엔도뉴클레아제 부위를 더 포함하는 방법.

#### 청구항 35

선행하는 항들 중 어느 한 항에 있어서, 상기 편집 요소는 서열 번호: 23 내지 28 중 어느 하나에 기재된 뉴클레오티드 서열을 포함하는 방법.

#### 청구항 36

선행하는 항들 중 어느 한 항에 있어서, 상기 5' 상동성 암(arm) 뉴클레오티드 서열은 상기 제1 게놈 영역과 적어도 90%, 95%, 96%, 97%, 98%, 또는 99% 동일한 방법.

#### 청구항 37

선행하는 항들 중 어느 한 항에 있어서, 상기 3' 상동성 암 뉴클레오티드 서열은 상기 제2 게놈 영역과 적어도 90%, 95%, 96%, 97%, 98%, 또는 99% 동일한 방법.

#### 청구항 38

선행하는 항들 중 어느 한 항에 있어서, 상기 제1 게놈 영역은 제1 편집창(editing window)에 위치되며, 상기 제2 게놈 영역은 제2 편집창에 위치되는 방법.

#### 청구항 39

제38항에 있어서, 상기 제1 및 제2 편집창들은 상이한 방법.

#### 청구항 40

제38항에 있어서, 상기 제1 및 제2 편집창들은 동일한 방법.

#### 청구항 41

제38항 내지 제40항 중 어느 한 항에 있어서, 상기 제1 편집창은 서열 번호: 101, 102, 또는 103에 기재된 뉴클레오티드 서열로 구성되는 방법.

#### 청구항 42

제38항 내지 제41항 중 어느 한 항에 있어서, 상기 제2 편집창은 서열 번호: 101, 102, 또는 103에 기재된 뉴클레오티드 서열로 구성되는 방법.

#### 청구항 43

제38항 내지 제42항 중 어느 한 항에 있어서, 상기 제1 게놈 영역은 서열 번호: 101에 기재된 뉴클레오티드 서열로 구성되는 방법.

#### 청구항 44

제38항 내지 제43항 중 어느 한 항에 있어서, 상기 제2 게놈 영역은 서열 번호: 102에 기재된 뉴클레오티드 서열로 구성되는 방법.

#### 청구항 45

선행하는 항들 중 어느 한 항에 있어서, 상기 5' 상동성 암은 서열 번호: 101에 기재된 뉴클레오티드 서열로 구성되는 방법.

#### 청구항 46

선행하는 항들 중 어느 한 항에 있어서, 상기 3' 상동성 암은 서열 번호: 102에 기재된 뉴클레오티드 서열로 구성되는 방법.

#### 청구항 47

선행하는 항들 중 어느 한 항에 있어서, 상기 5' 및 3' 상동성 암 뉴클레오티드 서열의 각각은 독립적으로 약 100 내지 약 2000 뉴클레오티드의 길이를 갖는 방법.

#### 청구항 48

선행하는 항들 중 어느 한 항에 있어서, 상기 교정 게놈은 상기 5' 상동성 암 뉴클레오티드 서열의 5' 반전된 (inverted) 말단 반복(5' ITR) 뉴클레오티드 서열 5', 및 상기 3' 상동성 암 뉴클레오티드 서열의 3' 반전된 말단 반복(3' ITR) 뉴클레오티드 서열 3'을 더 포함하는 방법.

#### 청구항 49

제48항에 있어서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 18과 적어도 95% 서열 동일성을 가지며, 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 19와 적어도 95% 서열 동일성을 갖는 방법.

#### 청구항 50

제48항에 있어서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 20과 적어도 95% 서열 동일성을 가지며, 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 21과 적어도 95% 서열 동일성을 갖는 방법.

#### 청구항 51

선행하는 항들 중 어느 한 항에 있어서, 상기 교정 게놈은 서열 번호: 29 내지 42 및 104 중 어느 하나에 기재된 뉴클레오티드 서열을 포함하는 방법.

#### 청구항 52

선행하는 항들 중 어느 한 항에 있어서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 203 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며: 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는

서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G인 방법.

### 청구항 53

제52항에 있어서,

- (a) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며;
- (b) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며;
- (c) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이며;
- (d) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는
- (e) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C인 방법.

### 청구항 54

제52항에 있어서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 203 내지 736의 아미노산 서열을 포함하는 방법.

### 청구항 55

선행하는 항들 중 어느 한 항에 있어서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 138 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며: 서열 번호: 2의 아미노산 151에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 160에 상응하는 캡시드 단백질에서 아미노산은 D이고; 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G인 방법.

### 청구항 56

제55항에 있어서,

- (a) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며;
- (b) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며;
- (c) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이며;

(d) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는

(e) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C인 방법.

### 청구항 57

제55항에 있어서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 138 내지 736의 아미노산 서열을 포함하는 방법.

### 청구항 58

선행하는 항들 중 어느 한 항에 있어서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 1 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 서열 번호: 2의 아미노산 2에 상응하는 캡시드 단백질에서 아미노산은 T이고; 서열 번호: 2의 아미노산 65에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 68에 상응하는 캡시드 단백질에서 아미노산은 V이고; 서열 번호: 2의 아미노산 77에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 119에 상응하는 캡시드 단백질에서 아미노산은 L이고; 서열 번호: 2의 아미노산 151에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 160에 상응하는 캡시드 단백질에서 아미노산은 D이고; 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G인 방법.

### 청구항 59

제58항에 있어서,

(a) 서열 번호: 2의 아미노산 2에 상응하는 캡시드 단백질에서 아미노산은 T이고, 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이며;

(b) 서열 번호: 2의 아미노산 65에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 Y이며;

(c) 서열 번호: 2의 아미노산 77에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이며;

(d) 서열 번호: 2의 아미노산 119에 상응하는 캡시드 단백질에서 아미노산은 L이고, 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이며;

(e) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며;

(f) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며;

(g) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이며;

(h) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는

(i) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C인 방법.

#### 청구항 60

제58항에 있어서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 1 내지 736의 아미노산 서열을 포함하는 방법.

#### 청구항 61

선행하는 항들 중 어느 한 항에 있어서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합 효율은 적어도 1%인 방법.

#### 청구항 62

선행하는 항들 중 어느 한 항에 있어서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합의 대립 유전자 빈도는 적어도 0.5%인 방법.

#### 청구항 63

하기를 포함하는 복제-결합 아데노-연관 바이러스(AAV: adeno-associated virus):

- a) AAV Clade F 캡시드 단백질을 포함하는 AAV 캡시드; 및
- b) (i) 표적 유전자에서 표적 좌위 편집용 편집 요소, (ii) 상기 표적 좌위에 대한 제1 게놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암 뉴클레오티드 서열 5', 및 (iii) 상기 표적 좌위에 대한 제2 게놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3'을 포함하는 교정 게놈.

#### 청구항 64

제63항에 있어서, 상기 표적 유전자는 상기 HBB 유전자인 AAV.

#### 청구항 65

제63항 또는 제64항에 있어서, 상기 표적 좌위는 상기 HBB 유전자에서 뉴클레오티드 점 돌연변이 또는 결실에 있는 AAV.

#### 청구항 66

제65항에 있어서, 상기 HBB 유전자에서 상기 뉴클레오티드 점 돌연변이 또는 결실은 위치 -87에서 G, 위치 -31에서 G, 위치 -30에서 A, 위치 -29에서 G, 위치 -28에서 G, 위치 -10에서 T, 위치 1에서 C, 위치 1에서 A, 위치 2에서 G, 위치 17 및 18에서 C 및 T의 결실, 위치 19에서 A, 위치 20에서 A의 결실, 위치 20에서 T, 위치 25 및 26에서 A 및 A의 결실, 위치 26 이후 G의 추가, 위치 47에서 A, 위치 48에서 A, 위치 51에서 C의 결실, 위치 52에서 A, 위치 58에서 G, 위치 59에서 G, 위치 79에서 A, 위치 82에서 T, 위치 84 이후 C의 추가, 위치 93에서 T, 위치 93에서 A, 위치 97에서 C, 위치 98에서 C, 위치 202에서 G, 위치 208에서 G, 위치 222에서 C, 위치 241 또는 242에서 T의 결실, 위치 254 내지 257에서 T 및 T 및 C 및 T의 결실, 위치 260에서 T, 위치 264 또는 265에서 C의 결실, 위치 343 이후 A의 추가, 위치 399 및 400에서 G 및 T의 결실, 위치 401에서 T, 위치 417 이후 A의 추가, 위치 446에서 A, 위치 1099에서 T, 위치 1293에서 A, 1344에서 T로 이루어진 군으로부터 선택되는 AAV.

#### 청구항 67

제65항 또는 제66항에 있어서, 상기 편집 요소는 돌연변이에 상응하는 야생형 HBB 유전자의 일부를 포함하는

AAV.

### 청구항 68

제63항 내지 제67항 중 어느 한 항에 있어서, 상기 편집 요소는 HBB 유전자의 하나 이상의 엑손의 코딩 영역을 포함하는 AAV.

### 청구항 69

제63항 내지 제68항 중 어느 한 항에 있어서, 상기 편집 요소는 엑손 1의 코딩 영역, 전체 인트론 1, 전체 엑손 2, 전체 인트론 2, 및 엑손 3의 코딩 영역을 포함하는 HBB 유전자의 일부를 포함하는 AAV.

### 청구항 70

제68항 또는 제69항에 있어서, 상기 코딩 영역은 상기 야생형 HBB 유전자의 상응하는 엑손에 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만으로 동일하게 침묵 변경되는 AAV.

### 청구항 71

제70항에 있어서, 상기 편집 요소는 서열 번호: 43 내지 46 및 105 내지 107로 이루어진 군으로부터 선택된 뉴클레오티드 서열 중 적어도 하나를 포함하는 AAV.

### 청구항 72

제63항에 있어서, 상기 표적 좌위는 AAVS1인 AAV.

### 청구항 73

제63항 내지 제66 및 제72항 중 어느 한 항에 있어서, 상기 편집 요소는 서열 번호: 48 또는 이의 일부를 인코딩하는 뉴클레오티드 서열을 포함하는 AAV.

### 청구항 74

제73항에 있어서, 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열은 서열 번호: 27의 뉴클레오티드 4 내지 444로 구성되는 AAV.

### 청구항 75

제73항에 있어서, 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열은 서열 번호: 27의 뉴클레오티드 4 내지 444에 70%, 75%, 80%, 85%, 또는 90% 미만으로 동일하게 침묵 변경되는 AAV.

### 청구항 76

제75항에 있어서, 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열은 서열 번호: 47 또는 100의 서열로 구성되는 AAV.

### 청구항 77

제73항에 있어서, 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열은 상기 HBB 유전자의 스터퍼-삽입된 코딩 서열로 구성되는 AAV.

### 청구항 78

제73항 내지 제77항 중 어느 한 항에 있어서, 상기 표적 좌위는 상기 표적 유전자의 뉴클레오티드 3 및 뉴클레오티드 4 사이의 뉴클레오티드간 결합이며, 이에 의하여 상기 표적 좌위에의 상기 편집 요소의 통합으로 인해 상기 표적 유전자의 시작 코돈으로 시작하는 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열을 포함하는 상기 표적 좌위를 생성하게 하는 AAV.

### 청구항 79

제73항 내지 제77항 중 어느 한 항에 있어서, 상기 편집 요소는 5' 내지 3' 시작 코돈 및 서열 번호: 48을 인코

딩하는 뉴클레오티드 서열로 이루어진 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열, 또는 상기 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열의 일부를 포함하는 AAV.

#### **청구항 80**

제79항에 있어서, 상기 표적 좌위는 상기 표적 유전자의 인트론 내에 있으며, 상기 편집 요소는 5' 내지 3' 스플라이스 수용체 부위, 리보솜 스키핑 요소, 및 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열을 포함하는 AAV.

#### **청구항 81**

제80항에 있어서, 상기 표적 좌위는 상기 HBB 유전자의 인트론 1 내에 있는 AAV.

#### **청구항 82**

제79항에 있어서, 상기 표적 좌위는 상기 표적 유전자의 코딩 뉴클레오티드에 인접하여 3'이고, 상기 편집 요소는 5' 내지 3' 리보솜 스키핑 요소 및 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열을 포함하는 AAV.

#### **청구항 83**

제82항에 있어서, 상기 표적 좌위는 야생형 표적 유전자 또는 돌연변이체 표적 유전자의 상응하는 뉴클레오티드의 정지 코돈인 AAV.

#### **청구항 84**

제73항 내지 제83항 중 어느 한 항에 있어서, 상기 편집 요소는 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열에 외인성 폴리아데닐화 서열 3'을 더 포함하는 AAV.

#### **청구항 85**

제63항 내지 제84항 중 어느 한 항에 있어서, 상기 편집 요소는 상기 표적 유전자에 존재하지 않는 제한 엔도뉴클레아제 부위를 더 포함하는 AAV.

#### **청구항 86**

제63항 내지 제85항 중 어느 한 항에 있어서, 상기 편집 요소는 서열 번호: 23 내지 28 중 어느 하나에 기재된 뉴클레오티드 서열을 포함하는 방법.

#### **청구항 87**

제63항 내지 제86항 중 어느 한 항에 있어서, 상기 5' 상동성 암 뉴클레오티드 서열은 상기 제1 게놈 영역과 적어도 90%, 95%, 96%, 97%, 98%, 또는 99% 동일한 AAV.

#### **청구항 88**

제63항 내지 제87항 중 어느 한 항에 있어서, 상기 3' 상동성 암 뉴클레오티드 서열은 상기 제2 게놈 영역과 적어도 90%, 95%, 96%, 97%, 98%, 또는 99% 동일한 AAV.

#### **청구항 89**

제62항 내지 제88항 중 어느 한 항에 있어서, 상기 제1 게놈 영역은 제1 편집창에 위치되며, 상기 제2 게놈 영역은 제2 편집창에 위치되는 AAV.

#### **청구항 90**

제89항에 있어서, 상기 제1 및 제2 편집창들은 상이한 AAV.

#### **청구항 91**

제89항에 있어서, 상기 제1 및 제2 편집창들은 동일한 AAV.

#### **청구항 92**

제89항 내지 제91항 중 어느 한 항에 있어서, 상기 제1 편집창은 서열 번호: 101, 102, 또는 103에 기재된 뉴클레오티드 서열로 구성되는 AAV.

### 청구항 93

제89항 내지 제92항 중 어느 한 항에 있어서, 상기 제2 편집창은 서열 번호: 101, 102, 또는 103에 기재된 뉴클레오티드 서열로 구성되는 AAV.

### 청구항 94

제89항 내지 제93항 중 어느 한 항에 있어서, 상기 제1 계놈 영역은 서열 번호: 101에 기재된 뉴클레오티드 서열로 구성되는 AAV.

### 청구항 95

제89항 내지 제94항 중 어느 한 항에 있어서, 상기 제2 계놈 영역은 서열 번호: 102에 기재된 뉴클레오티드 서열로 구성되는 AAV.

### 청구항 96

제63항 내지 제95항 중 어느 한 항에 있어서, 상기 5' 상동성 암은 서열 번호: 101에 기재된 뉴클레오티드 서열로 구성되는 AAV.

### 청구항 97

제63항 내지 제96항 중 어느 한 항에 있어서, 상기 3' 상동성 암은 서열 번호: 102에 기재된 뉴클레오티드 서열로 구성되는 AAV.

### 청구항 98

제63항 내지 제97항 중 어느 한 항에 있어서, 상기 5' 및 3' 상동성 암 뉴클레오티드 서열의 각각은 독립적으로 약 100 내지 약 2000 뉴클레오티드의 길이를 갖는 AAV.

### 청구항 99

제63항 내지 제98항 중 어느 한 항에 있어서, 상기 교정 계놈은 상기 5' 상동성 암 뉴클레오티드 서열의 5' 반전된 말단 반복(5' ITR) 뉴클레오티드 서열 5', 및 상기 3' 상동성 암 뉴클레오티드 서열의 3' 반전된 말단 반복(3' ITR) 뉴클레오티드 서열 3'을 더 포함하는 AAV.

### 청구항 100

제99항에 있어서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 18과 적어도 95% 서열 동일성을 가지며, 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 19와 적어도 95% 서열 동일성을 갖는 AAV.

### 청구항 101

제99항에 있어서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 20과 적어도 95% 서열 동일성을 가지며, 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 21과 적어도 95% 서열 동일성을 갖는 AAV.

### 청구항 102

제63항 내지 제101항 중 어느 한 항에 있어서, 상기 교정 계놈은 서열 번호: 29 내지 42 및 104 중 어느 하나에 기재된 뉴클레오티드 서열을 포함하는 AAV.

### 청구항 103

제63항 내지 제102항 중 어느 한 항에 있어서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 203 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며: 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열

번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G인 AAV.

#### 청구항 104

제103항에 있어서,

- (a) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며;
- (b) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며;
- (c) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이며;
- (d) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는
- (e) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C인 AAV.

#### 청구항 105

제103항에 있어서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 203 내지 736의 아미노산 서열을 포함하는 AAV.

#### 청구항 106

제63항 내지 제105항 중 어느 한 항에 있어서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 138 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며: 서열 번호: 2의 아미노산 151에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 160에 상응하는 캡시드 단백질에서 아미노산은 D이고; 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G인 AAV.

#### 청구항 107

제106항에 있어서,

- (a) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며;
- (b) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며;
- (c) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이며;
- (d) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는
- (e) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C인 AAV.

### 청구항 108

제106항에 있어서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 138 내지 736의 아미노산 서열을 포함하는 AAV.

### 청구항 109

제63항 내지 제108항 중 어느 한 항에 있어서T, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 1 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 서열 번호: 2의 아미노산 2에 상응하는 캡시드 단백질에서 아미노산은 T이고; 서열 번호: 2의 아미노산 65에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 68에 상응하는 캡시드 단백질에서 아미노산은 V이고; 서열 번호: 2의 아미노산 77에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 119에 상응하는 캡시드 단백질에서 아미노산은 L이고; 서열 번호: 2의 아미노산 151에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 160에 상응하는 캡시드 단백질에서 아미노산은 D이고; 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G인 AAV.

### 청구항 110

제109항에 있어서,

- (a) 서열 번호: 2의 아미노산 2에 상응하는 캡시드 단백질에서 아미노산은 T이고, 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이며;
- (b) 서열 번호: 2의 아미노산 65에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 Y이며;
- (c) 서열 번호: 2의 아미노산 77에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이며;
- (d) 서열 번호: 2의 아미노산 119에 상응하는 캡시드 단백질에서 아미노산은 L이고, 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이며;

(e) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며;

(f) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며;

(g) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이며;

(h) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는

(i) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C인 AAV.

### 청구항 111

제109항에 있어서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 1 내지 736의 아미노산 서열을 포함하는 AAV.

### 청구항 112

제63항 내지 제111항 중 어느 한 항에 있어서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합 효율은 적어도 1%인 AAV.

### 청구항 113

제63항 내지 제112항 중 어느 한 항에 있어서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합의 대립 유전자 빈도는 적어도 0.5%인 AAV.

### 청구항 114

제62항 내지 제111항 중 어느 한 항의 AAV를 포함하는 약학적 조성물.

### 청구항 115

AAV의 재조합 제조용 패키징 시스템으로서, 상기 패키징 시스템은

- a) 하나 이상의 AAV Rep 단백질을 인코딩하는 Rep 뉴클레오티드 서열;
- b) 제103항 내지 제111항 중 어느 하나에 기재된 바와 같이 하나 이상의 AAV Clade F 캡시드 단백질을 인코딩하는 Cap 뉴클레오티드 서열; 및
- c) 제63항 내지 제102항 중 어느 하나에 기재된 바와 같이 교정 게놈을 포함하며, 상기 패키징 시스템은 상기 캡시드 내에 상기 교정 게놈을 봉입하기 위한 세포 내에서 작용하여 상기 AAV를 형성하는 것인, 패키징 시스템.

### 청구항 116

제115항에 있어서, 상기 패키징 시스템은 상기 Rep 뉴클레오티드 서열 및 상기 Cap 뉴클레오티드 서열을 포함하는 제1 벡터, 및 상기 교정 게놈을 포함하는 제2 벡터를 포함하는 패키징 시스템.

### 청구항 117

제115항 또는 제116항에 있어서, 상기 Rep 뉴클레오티드 서열은 AAV2 Rep 단백질을 인코딩하는 패키징 시스템.

### 청구항 118

제117항에 있어서, 상기 AAV2 Rep 단백질은 78/68 또는 Rep 68/52인 패키징 시스템.

#### 청구항 119

제117항 또는 제118항에 있어서, 상기 AAV2 Rep 단백질은 서열 번호: 22의 AAV2 Rep 아미노산 서열과 최소 퍼센트 서열 동일성을 갖는 아미노산 서열을 포함하며, 상기 최소 퍼센트 서열 동일성은 상기 AAV2 Rep 단백질을 인코딩하는 아미노산 서열의 길이에 걸쳐 적어도 70%인 패키징 시스템.

#### 청구항 120

제115항 내지 제119항 중 어느 한 항에 있어서, 제3 벡터를 더 포함하며, 상기 제3 벡터는 헬퍼 바이러스 벡터인 패키징 시스템.

#### 청구항 121

제120항에 있어서, 상기 헬퍼 바이러스 벡터는 독립적 제3 벡터인 패키징 시스템.

#### 청구항 122

제120항에 있어서, 상기 헬퍼 바이러스 벡터는 상기 제1 벡터와 통합되는 패키징 시스템.

#### 청구항 123

제120항에 있어서, 상기 헬퍼 바이러스 벡터는 상기 제2 벡터와 통합되는 패키징 시스템.

#### 청구항 124

제120항 내지 제123항 중 어느 한 항에 있어서, 상기 제3 벡터는 헬퍼 바이러스 단백질을 인코딩하는 유전자를 포함하는 패키징 시스템.

#### 청구항 125

제120항 내지 제124항 중 어느 한 항에 있어서, 상기 헬퍼 바이러스는 아데노바이러스, 헤르페스 바이러스, 백시니아 바이러스, 및 거대세포바이러스(CMV: cytomegalovirus)로 이루어진 군으로부터 선택되는 패키징 시스템.

#### 청구항 126

제125항에 있어서, 상기 헬퍼 바이러스는 아데노바이러스인 패키징 시스템.

#### 청구항 127

제126항에 있어서, 상기 아데노바이러스 계놈은 E1, E2, E4 및 VA로 이루어진 군으로부터 선택된 하나 이상의 아데노바이러스 RNA 유전자를 포함하는 패키징 시스템.

#### 청구항 128

제125항에 있어서, 상기 헬퍼 바이러스는 헤르페스 단순 바이러스(HSV: herpes simplex virus)인 패키징 시스템.

#### 청구항 129

제128항에 있어서, 상기 HSV 계놈은 UL5/8/52, ICP0, ICP4, ICP22 및 UL30/UL42로 이루어진 군으로부터 선택된 하나 이상의 HSV 유전자를 포함하는 패키징 시스템.

#### 청구항 130

제120항 내지 제129항 중 어느 한 항에 있어서, 상기 제1 벡터 및 상기 제3 벡터는 제1 형질감염 플라스미드 내에 함유되는 패키징 시스템.

#### 청구항 131

제120항 내지 제129항 중 어느 한 항에 있어서, 상기 제2 벡터 및 상기 제3 벡터의 뉴클레오티드는 제2 형질감

염 플라스미드 내에 함유되는 패키징 시스템.

### 청구항 132

제120항 내지 제129항 중 어느 한 항에 있어서, 상기 제1 벡터 및 상기 제3 벡터의 뉴클레오티드는 재조합 헬퍼 바이러스로 클로닝되는 패키징 시스템.

### 청구항 133

제120항 내지 제129항 중 어느 한 항에 있어서, 상기 제2 벡터 및 상기 제3 벡터의 뉴클레오티드는 재조합 헬퍼 바이러스로 클로닝되는 패키징 시스템.

### 청구항 134

AAV의 재조합 제조방법으로서, 상기 방법은 본원에 제115항 내지 제133항 중 어느 한 항의 패키징 시스템을 캡 시드 내에 교정 게놈을 봉입하기 위해 작용하는 조건 하에서 세포 내로 도입하여 상기 AAV를 형성하는 단계를 포함하는, 방법.

## 발명의 설명

### 기술 분야

[0001]

관련 출원의 교차 참조

[0002]

이 출원은 2017년 10월 18일자로 출원된 미국 임시 출원 제62/574,163호; 및 2018년 1월 24일자로 출원된 미국 임시 출원 제62/621, 102호의 이익을 주장하며, 이를 각각은 그 전체 내용이 본원에 인용되어 포함된다.

[0003]

서열 목록

[0004]

본 출원은 606107\_HMT-023PC\_Sequence\_Listing\_ST25.txt라는 ASCII 형식으로 2018년 10월 12일자로 전자적으로 제출된 서열 목록을 포함하며, 크기는 200,630 바이트이다. 서열 목록은 그 전체 내용이 본원에 인용되어 포함된다.

[0005]

정부 이익의 진술

[0006]

본 발명은 국립 보건원(National Institutes of Health)에 의해 수여된 승인 번호 제P30CA033572호 하에서 정부 지원으로 이루어졌다. 정부는 본 발명에서 특정 권리를 갖는다.

### 배경 기술

[0007]

혈색소병증은 헤모글로빈 단백질의 생산, 구조 및/또는 기능이 비정상인 유전적 장애의 패밀리(family)를 포함한다. 헤모글로빈 단백질은 적혈구의 건조 중량의 대략 97 %를 구성하며 혈액의 산소 운반 능력을 약 70 배 증가시킨다. 주요 성인 헤모글로빈 단백질은 2개의 알파 글로빈(HBA) 서브유닛, 2개의 베타 글로빈(HBB) 서브유닛, 및 각 서브유닛과 연관된 헴(heme) 그룹으로 구성된다. 염색체 11에 대한 HBB의 유전적 결함은 겹상 세포질환(SCD: sickle cell disease) 및 베타 지중해빈혈증과 같은 특정 혈색소병증을 유발할 수 있다.

[0008]

겹상세포빈혈이라고도 하는 겹상세포질환은 대략 100,000 명의 미국인에게 영향을 미치는 상염색체 열성 질환이다. 이는 아프리카계 미국인들 사이에서 널리 퍼져 있지만 다른 인종 그룹에도 존재한다. 서부 및 중앙 아프리카에서는 모든 아기의 1 내지 2%가 SCD에 걸린 채 태어난다. SCD는 HBB의 코딩 서열에서 뉴클레오티드 20에서 동형 접합 돌연변이에 의해 야기된다. 이 돌연변이는 음으로 하전된 아미노산 글루타메이트(GAG에 의해 코딩됨)를 성숙한 베타 글로빈의 6 번째 아미노산에서 중성, 소수성 잔기 발린(GUG에 의해 인코딩됨)으로 대체한다. SCD 돌연변이를 갖는 베타 글로빈 사슬을 함유하는 헤모글로빈은 적혈구의 모양을 왜곡하는 다가닥 중합체로 응집되는 경향이 있어서, 이를 세포를 부서지기 쉽고 초승달 또는 겹상같은 형상이 되도록 한다. 이러한 비정상적 적혈구는 용혈을 위해 더 많이 배치되며 조직과 기관에 더 적은 산소를 전달한다. 더욱이, 헤모글로빈 응집은 적혈구를 강성이 되도록 하고 작은 혈관에 용이하게 유지시켜 혈류를 감소시키고 혈관 폐색을 야기한다. 그 결과, SCD 환자는 빈혈과 "위기"라고 불리는 통증 단계(episodic)를 겪으며, 위기 동안의 기관 손상은 SCD와 연관된 사망률 및 이환율의 주요 원인이다. 특히, 뼈, 비장, 신장 및 폐의 경색(즉, 부족한 혈액 공급으로 인한 조직의 괴사)이 특히 일반적이다. 이에 반하여, 겹상세포 돌연변이에 대해 이형 접합 성인들은 대체로

무증상이다.

[0009] 베타 지중해빈혈증은 전 세계 10 만 명 중 약 1 명과 유럽 연합의 1 만 명 중 약 1 명에 영향을 미친다. 베타 지중해빈혈증은 HBB에서 베타 글로빈의 발현을 감소시키는 다양한 돌연변이에 의해 유발된 장애 그룹이다. 지금 까지, 치환, 삽입 및 결실을 포함한 884 개의 상이한 돌연변이가 베타 지중해빈혈증(HbVar 데이터베이스)에서 확인되었다. 이러한 돌연변이는 HBB의 게놈 유전자와 전체에 위치된다. 치환, 삽입 및 작은 결실 중에서, 병원성 변이체는 5' UTR의 상류 및 3' UTR로 발견되었다. 이들 공지된 884 개 돌연변이 이외에, 추가의 변이체는 HBB 서열의 특정 변이와 연계할 때만 병원성일 수 있다. 베타 지중해빈혈증 돌연변이는 유전자 전사, RNA 가공, 전사 후 변형, mRNA의 번역 등에 영향을 줄 수 있다. 베타 지중해빈혈증은 중증도에서 매우 가변적이며, 일부 HBB 돌연변이는 베타 글로빈 생산의 완전한 손실을 초래하며 기타 HBB 돌연변이는 베타 글로빈 양의 감소만 초래한다. 두 대립 유전자 모두에서 종종 HBB 돌연변이를 갖는 중증의 베타 지중해빈혈증(즉, 중증성 지중해빈혈증)을 갖는 환자는 빈혈, 성장 지연 및 비정상적인 기관 발달을 겪는다. 경증 내지 중등도의 베타 지중해빈혈증(즉, 경증성 지중해빈혈증 또는 중간성 지중해빈혈증)이 있는 환자는 덜 심각한 증상을 나타낸다.

[0010] 혈색소병증은 만성 수혈을 필요로 하는 SCD 및 베타 지중해빈혈증으로 수혈 및 지지 치료로 관리될 수 있다. 그러나, 수혈을 반복하면 철분 과부하가 발생하고 합병증의 발생률을 줄이기 위해 철 키클레이션 요법이 필요하다. SCD 및 베타 지중해빈혈증의 이환율 및 사망률은 SCD에 대해 현재까지 FDA 승인된 유일한 약물인 하이드록시우레아에 의해 약화될 수 있다. 그러나, 이 치료는 처방률이 낮고 규정 준수가 좋지 않기 때문에 널리 사용되지 않는다.

[0011] SCD 또는 베타 지중해빈혈증을 치료하기 위해, 환자는 적절하게 발현될 수 있는 기능적 베타 헤모글로빈의 적어도 하나의 카페를 보유하는 조혈 줄기 세포를 받을 필요가 있다. 하나의 접근법은 골수 이식을 통해 동종이계 공여체로부터 야생형 조혈 줄기 세포를 얻는 것이다. 그러나, 일치하는 공여체의 이용 가능성은 주요 제한 요소이며, 골수 이식은 종종 5 내지 10%의 사망률을 초래하는 심각한 합병증과 연관이 있다. 보다 최근에, 유전자 요법 접근법은 생체 외(ex vivo)에서 베타 글로빈 발현 폴리 뉴클레오티드를 환자로부터 단리된 돌연변이 조혈 줄기 세포 내로 도입하기 위해 사용되어 왔다.

[0012] 현재까지, 모든 HBB 유전자 요법 임상 시험은 렌티바이러스 벡터와 같은 레트로바이러스 벡터의 사용을 수반하였다. 그러나, 레트로바이러스-기반 유전자 요법은 많은 안전성 및 효능 우려를 제기한다. 예를 들어, 인간 게놈 내로의 레트로바이러스 벡터의 삽입은 표적화되지 않기 때문에, 벡터가 종양 억제 유전자를 파괴하거나 종양 유전자를 활성화시켜 악성 종양을 유발할 위험이 있다. 실제로, 감마레트로바이러스 벡터로 CD34<sup>+</sup> 골수 전구체를 형질도입함으로써 X-연결된 중증 복합 면역 결핍(SCID)을 치료하기 위한 임상 시험에서, 10 명의 환자 중 4 명은 백혈병으로 발전했다(문헌[Hacein-Bey-Abina et al., J Clin Invest. (2008) 118(9):3132-42]). 더욱이, 이러한 안전성 우려로 인해, 렌티바이러스 유전자 요법은 생체 외에서만 수행될 수 있다. 생체 외 치료를 위해 대상체로부터 추출될 수 있는 조혈 줄기 세포의 수는 대상체에 존재하는 것의 작은 분율에 불과하고 생체 외에서 조혈 줄기 세포를 확장시키기 위해 현재 임상적으로 사용되는 신뢰할 수 있는 방법이 없기 때문에, 이러한 생체 외 사용은 요법의 효능을 감소시킨다.

[0013] 또한 메가뉴클레아제, 징크 핑거 뉴클레아제(ZFN: zinc finger nuclease), 전사 활성화제-유사 이펙터 뉴클레아제(TALEN: transcription activator-like effector nuclease), 및 클러스터된, 규칙적으로 간격을 둔 짧은 회문 반복(CRISPR: short palindromic repeat) 기술과 같은 뉴클레아제 기반 유전자 편집 기술은 SCD 및 베타 지중해빈혈증 환자에서 HBB 유전자의 결함을 교정하는 데 사용된다고 추측되었다. 그러나, 이들 기술의 각각은 의도된 표적 부위와 순차적으로 유사한 인간 게놈에서 부위의 비표적 돌연변이(off-target mutation)에 대한 가능성으로 인해 안전성 우려를 제기한다.

[0014] 따라서, 당업계에는 SCD 및 베타 지중해빈혈증 환자에서 HBB 유전자 기능을 효율적이고 안전하게 회복할 수 있는 개선된 유전자 요법 조성물 및 방법에 대한 요구가 존재한다.

### 발명의 내용

[0015] HBB 유전자에서 돌연변이를 교정하기 위한 아데노-연관 바이러스(AAV: adeno-associated virus) 조성물 및 이를 사용하여 세포에서 HBB 유전자 돌연변이를 교정하는 방법이 본원에 제공된다. 또한 아데노-연관 바이러스 조성물 제조용 패키징(packaging) 시스템이 제공된다.

[0016] 본원에 개시된 AAV 조성물 및 방법은 외인성 뉴클레아제(예를 들어, 메가뉴클레아제, 징크 핑거 뉴클레아제, 전

사 활성화제-유사 뉴클레아제(TALEN: transcriptional activator-like nuclease), 또는 Cas9와 같은 RNA-가이드 뉴클레아제(RNA-guided nuclease))를 사용하여 계놈 DNA를 절단할 필요없이 생체 내 HBB 유전자에서 돌연변이의 매우 효율적인 교정을 허용한다는 점에서 특히 유리하다.

[0017] 따라서, 일 양태에서, 본 개시내용은 (a) AAV Clade F 캡시드 단백질을 포함하는 AAV 캡시드, 및 (b) (i) 표적 유전자에서 표적 좌위 편집용 편집 요소, (ii) 상기 표적 좌위에 대한 제1 계놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암(arm) 뉴클레오티드 서열 5', 및 (iii) 상기 표적 좌위에 대한 제2 계놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3'을 포함하는 교정 계놈을 포함하는 복제-결합 아데노-연관 바이러스(AAV: adeno-associated virus)를 제공한다.

[0018] 다른 양태에서, 본 개시내용은 세포 내의 베타 글로빈(HBB) 유전자에서 돌연변이의 교정 방법으로서, 상기 방법은 상기 세포를 (a) AAV Clade F 캡시드 단백질을 포함하는 AAV 캡시드, 및 (b) (i) 표적 유전자에서 표적 좌위 편집용 편집 요소; (ii) 상기 표적 좌위에 대한 제1 계놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암 뉴클레오티드 서열 5'; 및 (iii) 상기 표적 좌위에 대한 제2 계놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3'을 포함하는 교정 계놈을 포함하는 복제-결합 아데노-연관 바이러스(AAV: adeno-associated virus)로 형질도입하는 단계를 포함하며, 상기 세포는 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 공동-형질도입하거나 공동-투여하지 않고 형질도입되는, 방법을 제공한다.

[0019] 특정 실시형태에서, 상기 세포는 다능성 줄기 세포이다. 특정 실시형태에서, 상기 세포는 조혈 줄기 세포이다. 특정 실시형태에서, 상기 세포는 CD34<sup>+</sup> 조혈 줄기 세포이다. 특정 실시형태에서, 상기 세포는 포유류 대상체 내에 있으며, 상기 AAV는 상기 대상체에서 상기 세포를 형질도입하는 데 유효한 양으로 상기 대상체에게 투여된다.

[0020] 다른 양태에서, 본 개시내용은 HBB 유전자 돌연변이와 연관된 질환 또는 장애를 갖는 대상체의 치료 방법에서, 상기 방법은 (a) 생체 외에서 상기 대상체로부터 적혈구 전구 세포에 AAV Clade F 캡시드 단백질; 및 (i) 표적 유전자에서 표적 좌위 편집용 편집 요소; (ii) 상기 표적 좌위에 대한 제1 계놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암 뉴클레오티드 서열 5'; 및 (iii) 상기 표적 좌위에 대한 제2 계놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3'을 포함하는 교정 계놈을 포함하는 복제-결합 AAV로 형질도입하는 단계; 및 (b) 상기 형질도입된 세포를 상기 대상체에게 투여하는 단계를 포함하며, 상기 세포는 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 공동-형질도입하거나 공동-투여하지 않고 형질도입되는, 방법을 제공한다.

[0021] 특정 실시형태에서, 상기 적혈구 전구 세포는 다능성 줄기 세포이다. 특정 실시형태에서, 상기 적혈구 전구 세포는 조혈 줄기 세포이다. 특정 실시형태에서, 상기 적혈구 전구 세포는 CD34<sup>+</sup> 조혈 줄기 세포이다.

[0022] 다른 양태에서, 본 발명은 HBB 유전자 돌연변이와 연관된 질환 또는 장애를 갖는 대상체의 치료 방법에서, 상기 방법은 상기 대상체에게 (a) AAV Clade F 캡시드 단백질을 포함하는 AAV 캡시드, 및 (b) (i) 표적 유전자에서 표적 좌위 편집용 편집 요소; (ii) 상기 표적 좌위에 대한 제1 계놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암 뉴클레오티드 서열 5'; 및 (iii) 상기 표적 좌위에 대한 제2 계놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3'을 포함하는 교정 계놈을 포함하는 복제-결합 AAV의 유효량을 투여하는 단계를 포함하며, 상기 세포는 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 공동-형질도입하거나 공동-투여하지 않고 형질도입되는, 방법을 제공한다.

[0023] 특정 실시형태에서, 상기 질환 또는 장애는 지중해빈혈증 또는 겹상세포질환이다. 특정 실시형태에서, 상기 대상체는 인간 대상체이다.

[0024] 하기 실시형태는 상기 양태들 각각에 적용된다.

[0025] 특정 실시형태에서, 상기 표적 유전자는 상기 HBB 유전자이다. 특정 실시형태에서, 상기 표적 좌위는 상기 HBB 유전자에서 뉴클레오티드 점 돌연변이 또는 결실에 있다. 특정 실시형태에서, 상기 HBB 유전자에서 상기 뉴클레오티드 점 돌연변이 또는 결실은 위치 -87에서 G, 위치 -31에서 G, 위치 -30에서 A, 위치 -29에서 G, 위치 -28에서 G, 위치 -10에서 T, 위치 1에서 C, 위치 1에서 A, 위치 2에서 G, 위치 17 및 18에서 C 및 T의 결실, 위치 19에서 A, 위치 20에서 A의 결실, 위치 20에서 T, 위치 25 및 26에서 A 및 A의 결실, 위치 26 이후 G의 추가, 위치 47에서 A, 위치 48에서 A, 위치 51에서 C의 결실, 위치 52에서 A, 위치 58에서 G, 위치 59에서 G, 위치

79에서 A, 위치 82에서 T, 위치 84 이후 C의 추가, 위치 93에서 T, 위치 97에서 C, 위치 98에서 C, 위치 202에서 G, 위치 208에서 G, 위치 222에서 C, 위치 241 또는 242에서 T의 결실, 위치 254 내지 257에서 T 및 T 및 C 및 T의 결실, 위치 260에서 T, 위치 264 또는 265에서 C의 결실, 위치 343 이후 A의 추가, 위치 399 및 400에서 G 및 T의 결실, 위치 401에서 T, 위치 417 이후 A의 추가, 위치 446에서 A, 위치 1099에서 T, 위치 1293에서 A, 1344에서 T로 이루어진 군으로부터 선택된다. 특정 실시형태에서, 상기 편집 요소는 돌연변이에 상응하는 야생형 HBB 유전자의 일부를 포함한다.

[0026] 특정 실시형태에서, 상기 편집 요소는 HBB 유전자의 하나 이상의 엑손의 코딩 영역을 포함한다. 특정 실시형태에서, 상기 편집 요소는 HBB 유전자의 하나 이상의 엑손의 코딩 영역으로 구성된다.

[0027] 특정 실시형태에서, 상기 편집 요소는 엑손 1의 코딩 영역, 전체 인트론 1, 전체 엑손 2, 전체 인트론 2, 및 엑손 3의 코딩 영역을 포함하는 HBB 유전자의 일부를 포함한다. 특정 실시형태에서, 상기 코딩 영역은 상기 야생형 HBB 유전자의 상응하는 엑손에 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만으로 동일하게 침묵(silently) 변경된다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 43 내지 46 및 105 내지 107로 이루어진 군으로부터 선택된 뉴클레오티드 서열 중 적어도 하나를 포함한다.

[0028] 특정 실시형태에서, 상기 표적 좌위는 AAVS1이다.

[0029] 특정 실시형태에서, 상기 편집 요소는 상기 HBB 유전자 또는 이의 일부의 코딩 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 48을 인코딩하는 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열은 서열 번호: 27의 뉴클레오티드 4 내지 444로 구성된다. 특정 실시형태에서, 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열은 서열 번호: 27의 뉴클레오티드 4 내지 444에 70%, 75%, 80%, 85%, 또는 90% 미만으로 동일하게 침묵 변경된다. 특정 실시형태에서, 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열은 서열 번호: 47 또는 100의 서열로 구성된다. 특정 실시형태에서, 상기 편집 요소는 상기 HBB 유전자의 스터퍼(stuffer)-삽입된 코딩 서열을 포함한다.

[0030] 특정 실시형태에서, 상기 표적 좌위는 상기 표적 유전자의 뉴클레오티드 3 및 뉴클레오티드 4 사이의 뉴클레오티드간 결합이며, 이에 의하여 상기 표적 좌위에의 상기 편집 요소의 통합으로 인해 상기 표적 유전자의 시작 코돈으로 시작하는 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열을 포함하는 상기 표적 좌위를 생성하게 된다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3' 시작 코돈 및 서열 번호: 48을 인코딩하는 뉴클레오티드 서열로 이루어진 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열, 또는 상기 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열의 일부를 포함한다. 특정 실시형태에서, 상기 표적 좌위는 상기 표적 유전자의 인트론 내에 있으며, 상기 편집 요소는 5' 내지 3' 스플라이스 수용체 부위, 리보솜 스키핑 요소(ribosomal skipping element), 및 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열을 포함한다. 특정 실시형태에서, 상기 표적 좌위는 상기 HBB 유전자의 인트론 1 내에 있다. 특정 실시형태에서, 상기 표적 좌위는 상기 표적 유전자의 코딩 뉴클레오티드에 인접하여 3'이고, 상기 편집 요소는 5' 내지 3' 리보솜 스키핑 요소 및 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열을 포함한다. 특정 실시형태에서, 상기 표적 좌위는 야생형 표적 유전자(예를 들어, HBB 유전자) 또는 돌연변이체 표적 유전자(예를 들어, HBB 유전자)의 상응하는 뉴클레오티드의 정지 코돈이다. 특정 실시형태에서, 상기 표적 좌위는 야생형 표적 유전자(예를 들어, HBB 유전자) 또는 돌연변이체 표적 유전자(예를 들어, HBB 유전자)의 상응하는 뉴클레오티드의 정지 코돈에 인접하여 5'이다.

[0031] 특정 실시형태에서, 상기 5' 상동성 암 뉴클레오티드 서열은 상기 제1 게놈 영역과 적어도 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하다. 특정 실시형태에서, 상기 3' 상동성 암 뉴클레오티드 서열은 상기 제2 게놈 영역과 적어도 90%, 95%, 96%, 97%, 98%, 또는 99% 동일하다. 특정 실시형태에서, 상기 제1 게놈 영역은 제1 편집창(editing window)에 위치되며, 상기 제2 게놈 영역은 제2 편집창에 위치된다. 특정 실시형태에서, 상기 제1 및 제2 편집창들은 상이하다. 특정 실시형태에서, 상기 제1 및 제2 편집창들은 동일하다. 특정 실시형태에서, 상기 제1 편집창은 서열 번호: 101, 102, 또는 103에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 제2 편집창은 서열 번호: 101, 102, 또는 103에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 제1 게놈 영역은 서열 번호: 101에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 제2 게놈 영역은 서열 번호: 102에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 5' 상동성 암은 서열 번호: 101에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 3' 상동성 암은 서열 번호: 102에 기재된 뉴클레오티드 서열로 구성된다.

[0032] 특정 실시형태에서, 상기 편집 요소는 서열 번호: 48을 인코딩하는 뉴클레오티드 서열에 외인성 폴리아데닐화 서열 3'을 더 포함한다. 특정 실시형태에서, 상기 편집 요소는 상기 표적 유전자에 존재하지 않는 제한 엔도뉴

클레이어제 부위를 더 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 23 내지 28 중 어느 하나에 기재된 뉴클레오티드 서열을 포함한다.

[0033] 특정 실시형태에서, 상기 5' 및 3' 상동성 암 뉴클레오티드 서열의 각각은 독립적으로 약 100 내지 약 2000 뉴클레오티드의 길이를 갖는다.

[0034] 특정 실시형태에서, 상기 교정 계획은 상기 5' 상동성 암 뉴클레오티드 서열의 5' 반전된(inverted) 말단 반복(5' ITR) 뉴클레오티드 서열 5', 및 상기 3' 상동성 암 뉴클레오티드 서열의 3' 반전된 말단 반복(3' ITR) 뉴클레오티드 서열 3'을 더 포함한다. 특정 실시형태에서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 18과 적어도 95% 서열 동일성을 가지며, 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 19와 적어도 95% 서열 동일성을 갖는다. 특정 실시형태에서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 20과 적어도 95% 서열 동일성을 가지며, 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 21과 적어도 95% 서열 동일성을 갖는다. 특정 실시형태에서, 상기 교정 계획은 서열 번호: 29 내지 42 및 104 중 어느 하나에 기재된 뉴클레오티드 서열을 포함한다.

[0035] 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 203 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 선택적으로: 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이다.

[0036] 특정 실시형태에서, (a) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며; (b) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며; (c) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이며; (d) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는 (e) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이다.

[0037] 특정 실시형태에서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 203 내지 736의 아미노산의 아미노산 서열을 포함한다.

[0038] 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 138 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 선택적으로: 서열 번호: 2의 아미노산 151에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 160에 상응하는 캡시드 단백질에서 아미노산은 D이고; 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호:

2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이다.

[0039] 특정 실시형태에서, (a) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며; (b) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며; (c) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이며; (d) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는 (e) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이다.

[0040] 특정 실시형태에서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 138 내지 736의 아미노산 서열을 포함한다.

[0041] 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 1 내지 736의 아미노산 서열과 적어도 95% 서열 동일성을 갖는 아미노산 서열을 포함하며, 선택적으로: 서열 번호: 2의 아미노산 2에 상응하는 캡시드 단백질에서 아미노산은 T이고; 서열 번호: 2의 아미노산 65에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 68에 상응하는 캡시드 단백질에서 아미노산은 V이고; 서열 번호: 2의 아미노산 77에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 119에 상응하는 캡시드 단백질에서 아미노산은 L이고; 서열 번호: 2의 아미노산 151에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 160에 상응하는 캡시드 단백질에서 아미노산은 D이고; 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이고; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이다.

[0042] 특정 실시형태에서, (a) 서열 번호: 2의 아미노산 2에 상응하는 캡시드 단백질에서 아미노산은 T이고, 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이며; (b) 서열 번호: 2의 아미노산 65에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 Y이며; (c) 서열 번호: 2의 아미노산 77에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이며; (d) 서열 번호: 2의 아미노산 119에 상응하는 캡시드 단백질에서 아미노산은 L이고, 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이며; (e) 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이고, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이며; (f) 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이며; (g) 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; (h) 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며; 또는 (i) 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 R이고, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이다.

- [0043] 특정 실시형태에서, 상기 캡시드 단백질은 서열 번호: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 1 내지 736의 아미노산 서열을 포함한다.
- [0044] 특정 실시형태에서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합 효율은 적어도 1%이다. 특정 실시형태에서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합의 대립 유전자 빈도는 적어도 0.5%이다.
- [0045] 다른 양태에서, 본 개시내용은 본원에 개시된 바와 같이 AAV를 포함하는 약학적 조성물을 제공한다.
- [0046] 다른 양태에서, 본 개시내용은 AAV의 재조합 제조용 패키징 시스템으로서, 상기 패키징 시스템은: (a) 하나 이상의 AAV Rep 단백질을 인코딩하는 Rep 뉴클레오티드 서열, (b) 본원에 기재된 바와 같이 하나 이상의 AAV Clade F 캡시드 단백질을 인코딩하는 Cap 뉴클레오티드 서열, 및 (c) 본원에 개시된 바와 같이 교정 게놈을 포함하며, 상기 패키징 시스템은 상기 캡시드 내에 상기 교정 게놈을 봉입하기 위한 세포 내에서 작용하여 상기 AAV를 형성하는 것인, 패키징 시스템을 제공한다.
- [0047] 특정 실시형태에서, 상기 패키징 시스템은 상기 Rep 뉴클레오티드 서열 및 상기 Cap 뉴클레오티드 서열을 포함하는 제1 백터, 및 상기 교정 게놈을 포함하는 제2 백터를 포함한다. 특정 실시형태에서, 상기 Rep 뉴클레오티드 서열은 AAV2 Rep 단백질을 인코딩한다. 특정 실시형태에서, 상기 AAV2 Rep 단백질은 78/68 또는 Rep 68/52이다. 특정 실시형태에서, 상기 AAV2 Rep 단백질은 서열 번호: 22의 AAV2 Rep 아미노산 서열과 최소 퍼센트 서열 동일성을 갖는 아미노산 서열을 포함하며, 상기 최소 퍼센트 서열 동일성은 상기 AAV2 Rep 단백질을 인코딩하는 아미노산 서열의 길이에 걸쳐 적어도 70%이다.
- [0048] 특정 실시형태에서, 상기 패키징 시스템은 제3 백터를 더 포함하며, 상기 제3 백터는 헬퍼(helper) 바이러스 백터이다. 특정 실시형태에서, 상기 헬퍼 바이러스 백터는 독립적 제3 백터이다. 특정 실시형태에서, 상기 헬퍼 바이러스 백터는 상기 제1 백터와 통합된다. 특정 실시형태에서, 상기 헬퍼 바이러스 백터는 상기 제2 백터와 통합된다. 특정 실시형태에서, 상기 제3 백터는 헬퍼 바이러스 단백질을 인코딩하는 유전자를 포함한다.
- [0049] 특정 실시형태에서, 상기 헬퍼 바이러스는 아데노바이러스, 헤르페스 바이러스, 백시니아 바이러스, 및 거대세포바이러스(CMV: cytomegalovirus)로 이루어진 군으로부터 선택된다. 특정 실시형태에서, 상기 헬퍼 바이러스는 아데노바이러스이다. 특정 실시형태에서, 상기 아데노바이러스 게놈은 E1, E2, E4 및 VA로 이루어진 군으로부터 선택된 하나 이상의 아데노바이러스 RNA 유전자를 포함한다. 특정 실시형태에서, 상기 헬퍼 바이러스는 헤르페스 단순 바이러스(HSV: herpes simplex virus)이다. 특정 실시형태에서, 상기 HSV 게놈은 UL5/8/52, ICP0, ICP4, ICP22 및 UL30/UL42로 이루어진 군으로부터 선택된 하나 이상의 HSV 유전자를 포함한다.
- [0050] 특정 실시형태에서, 상기 제1 백터 및 상기 제3 백터는 제1 형질감염 플라스미드 내에 함유된다. 특정 실시형태에서, 상기 제2 백터 및 상기 제3 백터의 뉴클레오티드는 제2 형질감염 플라스미드 내에 함유된다. 특정 실시형태에서, 상기 제1 백터 및 상기 제3 백터의 뉴클레오티드는 재조합 헬퍼 바이러스로 클로닝된다. 특정 실시형태에서, 상기 제2 백터 및 상기 제3 백터의 뉴클레오티드는 재조합 헬퍼 바이러스로 클로닝된다.
- [0051] 다른 양태에서, 본 개시내용은 AAV의 재조합 제조방법으로서, 상기 방법은 본원에 기재된 바와 같은 패키징 시스템을 캡시드 내에 교정 게놈을 봉입하기 위해 작용하는 조건 하에서 세포 내로 도입하여 상기 AAV를 형성하는 단계를 포함하는, 방법을 제공한다.

### 도면의 간단한 설명

- [0052] 도 1a 및 1b는 AAVS1-FP가 GM16265 세포의 게놈으로 통합된 유세포 분석 결과를 나타내는 그래프이며, AAVS1-FP 백터는 AAHVSC7, AAHVSC15, 및 AAHVSC17에 패키지되어 있다.
- 도 1c는 AAHVSC17 캡시드에 패키지된 AAVS1-FP 백터로 형질도입된 1차 인간 CD34<sup>+</sup> 조혈 줄기 세포(HSC: hematopoietic stem cell)에서 FP 코딩 서열의 통합을 갖는 대립 유전자의 백분율을 나타내는 그래프이다.
- 도 2는 12 bp 링커를 함유하는 HBB 교정 백터 hHBB-hL-014의 플라스미드 맵을 도시하고 있다. 이 도면에서, HBB의 흑색 영역은 인간 HBB 단백질을 인코딩하는 엑손에서의 뉴클레오티드 서열을 나타내고, 흑색 영역 사이의 점선은 엑손 사이의 인트론을 나타낸다.
- 도 3a는 AAHVSC15 및 AAHVSC17 캡시드에 패키지된 hHBB-hL-014 백터로 형질도입된 GM16265 세포의 게놈 DNA로

부터 증폭된 DNA의 편집-특이적 크기를 나타내는 DNA 전기영동의 이미지이다.

도 3b는 AAVHSC17 캡시드에 패키지된 hHBB-hL-014 백터로 형질도입된 GM16265, GM16266 및 GM16267 LCL의 게놈 DNA로부터 증폭된 DNA의 편집-특이적 크기를 보여주는 DNA 전기영동의 이미지이다.

도 4a, 4b, 4c, 및 4d는 각각 HBB 교정 백터 hHBB-hL-001, hHBB-h1W-013, hHBB-hL-011 및 hHBB-h1W-012의 2개의 AAV ITR 사이의 유전적 요소를 보여주는 백터 맵이다. 이들 도면에서, HBB의 흑색 영역은 인간 HBB 단백질을 인코딩하는 엑손에서의 뉴클레오티드 서열을 나타내고, 흑색 영역 사이의 점선은 엑손 사이의 인트론을 나타낸다.

도 5a, 5b 및 5c는 AAVHSC17 캡시드에 패키지된 도면에 표시된 백터로 형질도입된 1차 인간 CD34<sup>+</sup> HSC의 게놈 DNA로부터 증폭된 DNA의 편집-특이적 크기를 나타내는 DNA 전기영동의 이미지이다.

도 6은 표시된 바와 같이 샘플에 걸쳐 편집된 CD34+ 세포의 분율을 나타내는 그래프이다.

도 7a, 7b, 7c, 7d, 및 7e는 각각 HBB 교정 백터 hHBB-hA-009, hHBB-hAW-002, hHBB-h1-010, hHBB-h1W-008, 및 hHBB-hE3C-001의 2개의 AAV ITR 사이의 유전적 요소를 보여주는 백터 맵이다. 도 7a 내지 7d에서, "HBB 코딩 영역" 또는 "HBB 코딩 영역(66%)"으로 표지된 흑색 영역은 시작 코돈에서 정지 코돈까지(도 7c 및 7d) 또는 제2 코돈에서 정지 코돈까지(도 7 및 7b) 인간 HBB 단백질을 인코딩하는 뉴클레오티드 서열을 나타낸다.

도 8은 AAVHSC7 캡시드에 패키지된 도면에 표시된 백터로 형질도입된 RKO 및 GM16265 LCL 세포의 게놈 DNA로부터 증폭된 DNA의 편집-특이적 크기(1,874 bp) 및 비특이적 크기(1,180 bp)를 보여주는 DNA 전기영동의 이미지이다.

도 9는 AAVHSC7 및 AAVHSC17 캡시드에 패키지된 AAVS1-FP 백터의 투여 후 인간 HSC로 이종 이식된 NSG 마우스로부터의 혈액, 골수("BM"), 및 비장 세포에서 FP 코딩 서열의 통합을 갖는 대립 유전자의 백분율을 나타내는 그래프이다.

### 발명을 실시하기 위한 구체적인 내용

[0053]

본 개시내용은 HBB 유전자에서 돌연변이 교정용 아데노-연관 바이러스(AAV: adeno-associated virus) 조성물 및 이를 사용하여 세포에서 HBB 유전자 돌연변이를 교정하는 방법을 제공한다. 또한 아데노-연관 바이러스 조성물 제조용 패키징 시스템이 제공된다.

[0054]

#### I. 정의

[0055]

본원에 사용된 용어 "복제-결합 아데노-연관 바이러스"는 Rep 및 Cap 유전자가 없는 게놈을 포함하는 AAV를 지칭한다.

[0056]

본원에 사용된 용어 "HBB 유전자"는 HBB 유전자의 코딩 영역, 엑손, 인트론, 5' UTR, 3' UTR, 및 전사 조절 영역을 포함하지만, 이에 한정되지 않는 야생형 또는 돌연변이 인간 베타 글로빈 유전자를 지칭한다.

[0057]

본원에 사용된 용어 "HBB 유전자에서 돌연변이를 교정하는 것"은 야생형 HBB 단백질 또는 이의 기능적 등가물을 발현할 수 있는 좌위를 생성하기 위하여 표적 유전자(예를 들어, 돌연변이체 HBB 유전자)의 표적 좌위에서 하나 이상의 뉴클레오티드의 삽입, 결실, 또는 치환을 지칭한다. 특정 실시형태에서, "HBB 유전자에서 돌연변이를 교정하는 것"은 HBB 유전자에서 돌연변이를 야생형 서열로 되돌리는 것을 포함한다. 특정 실시형태에서, "HBB 유전자에서 돌연변이를 교정하는 것"은 야생형 베타 글로빈 단백질 또는 이의 기능적 등가물이 선택적으로는 외인성 표적 유전자 프로모터(예를 들어, HBB 유전자 프로모터)의 제어 하에서 표적 유전자의 좌위(예를 들어, 돌연변이체 HBB 유전자 좌위)로부터 발현되도록, 야생형 베타 글로빈 단백질 또는 이의 기능적 등가물의 적어도 일부를 인코딩하는 뉴클레오티드 서열을 표적 유전자(예를 들어, 돌연변이체 HBB 유전자)에 삽입하는 것을 포함한다. 본원에 사용된 "기능적 등가물"은 야생형 베타 글로빈으로서 기능할 수 있는 유전자 또는 이의 단편의 생성물을 지칭한다. 특정 실시형태에서, HBB의 기능적 등가물은 엡실론 글로빈(HBE), 델타 글로빈(HBD), 감마 글로빈 1(HBG1), 감마 글로빈 2(HBG2) 및 HBB 유사 유전자(pseudogene) HBBP와 같은 다른 글로빈 유전자 또는 유사 유전자를 포함할 수 있다. 특정 실시형태에서, HBB의 기능적 등가물은 변형된 베타 글로빈 단백질일 수 있으며, 변형은 야생형 베타 글로빈에서 발견되지 않은 적어도 하나의 특성, 예를 들어 SCD 돌연변이를 보유한 베타 글로빈의 응집을 억제하는 능력을 부여한다.

[0058]

본원에 사용된 용어 "교정 게놈"은 상동 재조합을 통해 편집 요소(예를 들어, 하나 이상의 뉴클레오티드 또는

뉴클레오티드간 결합)를 표적 좌위에 통합하여 HBB 유전자에서 유전적 결함을 교정할 수 있는 재조합 AAV 계놈을 지칭한다. 특정 실시형태에서, 상기 표적 좌위는 인간 HBB 유전자에 있다. 당업자는 5' 상동성 암, 편집 요소, 및 3' 상동성 암을 포함하는 교정 계놈의 일부가 표적 좌위(예를 들어, 인간 HBB 유전자)에 대해 센스 또는 앤티센스 배향일 수 있음을 이해할 것이다.

[0059] 본원에 사용된 용어 "편집 요소"는 표적 좌위에 통합될 때 표적 좌위를 변형시키는 교정 계놈의 일부를 지칭한다. 편집 요소는 표적 좌위에서 하나 이상의 뉴클레오티드의 삽입, 결실 또는 치환을 매개할 수 있다.

[0060] 본원에 사용된 용어 "표적 좌위"는 편집 요소에 의해 변형된 염색체 또는 뉴클레오티드 간 결합의 영역을 지칭한다. 특정 실시형태에서, 상기 표적 좌위는 HBB 유전자에서 영역 또는 뉴클레오티드 간 결합이며, 선택적으로, 상기 표적 좌위는 베타 글로빈 단백질의 발현 또는 기능을 손상시키는 적어도 하나의 유전적 돌연변이를 포함한다. 특정 실시형태에서, 상기 표적 좌위는 AAVS1이다. 상기 AAVS1 좌위는 문헌[Giraud et al., Proc Natl Acad Sci U S A. (1994) 91(21): 10039-43]; 문헌[Linden et al., Proc Natl Acad Sci U S A. (1996) 93(21): 11288-94]; 및 문헌[Linden et al., Proc Natl Acad Sci U S A. (1996) 93(15):7966-72]에 기재된 바와 같이, NCBI 참조 서열 번호 NC\_000019.10의 뉴클레오티드 55,090,913 내지 55,117,600인 염색체 19 qter13.3-13.4 상에 있으며, 이들 각각은 그 전체 내용이 본원에 인용되어 포함된다. 특정 실시형태에서, 상기 표적 좌위는 안전한 항구(safe harbor) 좌위이다. 안전한 항구 좌위는 새로 삽입된 유전자 물질이: (1) 예측 가능하게 기능하고; 그리고 (2) 숙주 세포 또는 유기체에 위험을 초래할 수 있는 숙주 계놈의 변경을 유발하지 않아야 하도록 보장하는 방식으로 새로운 유전자 물질의 통합을 수용할 수 있는 계놈의 부위이다. 따라서, 특정 실시형태에서, 표적 좌위는 숙주 계놈과의 원하지 않는 상호 작용의 위험을 최소화하면서 예측 가능한 전이유전자(transgene) 발현을 지원할 수 있는 당업계에 공지된 임의의 안전한 항구 좌위일 수 있다.

[0061] 본원에 사용된 용어 "표적 유전자"는 표적 좌위 또는 이의 일부가 위치하는 유전자를 지칭한다. 특정 실시형태에서, 상기 표적 좌위는 표적 유전자에 완전히 존재한다. 특정 실시형태에서, 상기 표적 유전자는 HBB이다. 특정 실시형태에서, 상기 표적 유전자는 인간 PPP1R12C이다. 특정 실시형태에서, 상기 표적 유전자는 적혈구 전구 세포에서 발현된다.

[0062] 본원에 사용된 용어 "상동성 암"은 표적 좌위 측면에 있는 계놈과 실질적으로 동일한 편집 요소의 5' 또는 3'에 위치된 교정 계놈의 일부를 지칭한다. 특정 실시형태에서, 상기 표적 좌위는 인간 HBB 유전자에 있고, 상기 상동성 암은 표적 좌위 측면에 있는 계놈과 실질적으로 동일한 서열을 포함한다.

[0063] 본원에 사용된 용어 "Clade F 캡시드 단백질"은 본원에서 서열 번호: 1의 아미노산 1 내지 736, 138 내지 736, 및 203 내지 736에서 각각 기재된 VP1, VP2, 또는 VP3 아미노산 서열과 적어도 90% 동일성을 갖는 아미노산 서열을 포함하는 AAV VP1, VP2, 또는 VP3 캡시드 단백질을 지칭한다. 본원에서 사용된 바와 같이, 2개의 뉴클레오티드 서열 사이 또는 2개의 아미노산 서열 사이의 동일성은 정렬된 동일한 뉴클레오티드 또는 아미노산의 수를 길이가 더 긴 뉴클레오티드 또는 아미노산 서열의 전장으로 나눈 것에 의해 결정된다.

[0064] 본원에 사용된 용어 "HBB 유전자 돌연변이와 연관된 질환 또는 장애"는 HBB 유전자의 변이에 의해 야기되거나, 악화되거나, 또는 유전적으로 연결된 임의의 질환 또는 장애를 지칭한다. 특정 실시형태에서, 상기 HBB 유전자 돌연변이와 연관된 질환 또는 장애는 겹상세포질환 또는 베타 지중해빈혈증과 같은 혈색소병증이다.

[0065] 본원에서 사용되는 용어 "침묵 변경된"은 코딩 서열 또는 스태퍼(staffer)-삽입된 코딩 서열에 의해 인코딩된 폴리펩티드의 아미노산 서열을 변경시키지 않고 유전자의 코딩 서열 또는 스태퍼-삽입된 코딩 서열의 변형(예를 들어, 뉴클레오티드 치환에 의한)을 지칭한다. 이러한 침묵(silent) 변경은 표적 유전자와 파라성(paralogous)인 다른 유전자 또는 유사 유전자의 좌위(예를 들어, 다른 글로빈 유전자 좌위 또는 베타 글로빈 유사 유전자 좌위) 내로의 교정 계놈의 통합 가능성을 감소시킨다는 점에서 유리하다. 이러한 침묵 변경은 또한 편집 요소와 표적 유전자 사이의 상동성을 감소시켜, 상동성 암보다는 편집 요소에 의해 매개되는 바람직하지 않은 통합을 감소시킨다.

[0066] 본원에 사용된 용어 "코딩 서열"은 시작 코돈에서 시작하여 정지 코돈에서 끝나는 폴리펩티드를 인코딩하는 상보적 DNA(cDNA)의 부분 또는 이의 침묵 변경된 서열을 지칭한다. 유전자는 대안적인 스플라이싱 및/또는 대안적인 번역 개시로 인해 하나 이상의 야생형 코딩 서열을 가질 수 있다. 예시적인 야생형 HBB 코딩 서열은 NCBI 참조 서열: NM\_000518.4의 뉴클레오티드 51 내지 494에 기재되어 있다.

[0067] 본원에 사용된 용어 "코딩 뉴클레오티드"는 정지 코돈의 3' 뉴클레오티드를 제외하고 유전자의 코딩 서열에서 뉴클레오티드에 상응하는 유전자의 뉴클레오티드를 지칭한다. 따라서, 특정 실시형태에서, 상기 HBB 유전자의

코딩 뉴클레오티드는 상기 HBB 유전자의 뉴클레오티드 1 내지 443 중 어느 하나이다.

[0068]

본원에 사용된 유전자의 "스터퍼-삽입된 코딩 서열"이라는 용어는 유전자의 코딩 서열에 삽입된 하나 이상의 인트론을 포함하는 뉴클레오티드 서열을 지칭한다. 특정 실시형태에서, 인트론 중 적어도 하나는 비 천연 인트론, 즉, 유전자의 천연 인트론과 상이한 서열을 갖는 것이다. 특정 실시형태에서, 스터퍼-삽입된 코딩 서열에서 모든 인트론은 비 천연 인트론이다. 비 천연 인트론은 다른 종으로부터의 인트론의 서열 또는 동일한 종과 상이한 유전자에서 인트론의 서열을 가질 수 있다. 대안적으로 또는 추가적으로, 비 천연 인트론 서열의 적어도 일부는 합성일 수 있다. 당업자는 비 천연 인트론 서열이 당업계에 공지된 임의의 컨센서스 스플라이싱 모티프를 도입함으로써 RNA 스플라이싱을 매개하도록 설계될 수 있음을 이해할 것이다. 예시적인 컨센서스 스플라이싱 모티프는 문헌[Sibley et al., (2016) *Nature Reviews Genetics*, 17, 407-21]에 제공되며, 이는 그 전체 내용이 본원에 인용되어 포함된다. 비 천연 인트론의 삽입은 스태퍼 서열이 백터의 조정이 최적 크기(예를 들어, 4.5 내지 4.8 kb)에 도달할 수 있게 함에 따라 백터 패키징의 효율성 및 견고성을 촉진시킬 수 있다. 특정 실시형태에서, 인트론 중 적어도 하나는 유전자의 천연 인트론이다. 특정 실시형태에서, 스터퍼-삽입된 코딩 서열에서 모든 인트론은 유전자의 천연 인트론이다. 비 천연 또는 천연 인트론은 코딩 서열에서 임의의 뉴클레오티드 간 결합에서 삽입될 수 있다. 특정 실시형태에서, 효율적인 스플라이싱을 촉진할 것으로 예측되는 뉴클레오티드 간 결합에서 하나 이상의 비 천연 또는 천연 인트론이 삽입된다(예를 들어, 문헌[Zhang (1998) *Human Molecular Genetics*, 7(5):919-32] 참조, 이는 그 전체 내용이 본원에 인용되어 포함된다). 특정 실시형태에서, 하나 이상의 비 천연 또는 천연 인트론이 2개의 내인성 엑손을 연결하는 뉴클레오티드 간 결합에서 삽입된다.

[0069]

본원에 사용된 용어 "리보솜 스키핑 요소"는 하나의 mRNA 분자의 번역으로부터 2개의 웨티드 사슬을 생성시킬 수 있는 짧은 웨티드 서열을 인코딩하는 뉴클레오티드 서열을 지칭한다. 본원에 사용된 용어 "리보솜 스키핑 웨티드"는 리보솜 스키핑 요소에 의해 인코딩된 웨티드를 지칭한다. 특정 실시형태에서, 상기 리보솜 스키핑 웨티드는  $X_1X_2EX_3NPGP$ 의 컨센서스 모티프를 포함하고, 상기 식에서  $X_1$ 은 D 또는 G이고,  $X_2$ 는 V 또는 I이며,  $X_3$ 는 임의의 아미노산(서열 번호: 49)이다. 특정 실시형태에서, 상기 리보솜 스키핑 웨티드는 토세아-아사인아(those-asigna) 바이러스 2A 웨티드(T2A), 돼지 테스코바이러스(porcine teschovirus)-1 2A 웨티드(P2A), 구제역 바이러스 2A 웨티드(F2A), 말 비염 A 바이러스 2A 웨티드(E2A), 세포질 다각체병 바이러스 2A 웨티드(BmCPV 2A), 및 flacherie virus of *B. mori* 2A 웨티드의 무름병 바이러스(BmIFV 2A)로 이루어진 군으로부터 선택된다. T2A 웨티드 및 P2A 웨티드의 예시적인 아미노산 서열은 각각 서열 번호: 71 및 73에 기재되어 있다. T2A 요소 및 P2A 요소의 예시적인 뉴클레오티드 서열은 각각 서열 번호: 72 및 74에 기재되어 있다. 특정 실시형태에서, 상기 리보솜 스키핑 요소는 N 말단에서 Gly-Ser-Gly의 서열을 더 포함하는 웨티드를 인코딩하며, 선택적으로 N 말단에서 Gly-Ser-Gly의 서열은 GGCAGCGGA(서열 번호: 75)의 뉴클레오티드 서열에 의해 인코딩된다. 이론에 구속되고 싶지는 않지만, 리보솜 스키핑 요소는 제1 웨티드 사슬의 번역을 종결시키고 제2 웨티드 사슬의 번역을 재개시하고; 또는 인코딩된 웨티드의 고유 프로테아제 활성에 의해 또는 환경(예를 들어, 사이토졸) 내의 다른 프로테아제에 의해 리보솜 스키핑 웨티드에서의 웨티드 결합의 절단에 의해 기능한다고 추측된다.

[0070]

본원에 사용된 용어 "폴리아데닐화 서열"은 RNA로 전사될 때 폴리아데닐화 신호 서열을 구성하는 DNA 서열을 지칭한다.

[0071]

본 개시내용에서, 유전자 내에서 뉴클레오티드 위치는 시작 코돈의 제1 뉴클레오티드에 대해 특정된다. 시작 코돈의 제1 뉴클레오티드는 위치 1이고, 시작 코돈의 제1 뉴클레오티드에 대한 뉴클레오티드 5'는 음수를 가지며, 시작 코돈의 제1 뉴클레오티드에 대한 뉴클레오티드 3'은 양수를 갖는다. 예를 들어, 본원에 사용된 HBB 유전자의 뉴클레오티드 1은 NCBI 참조 서열: NG\_000007.3의 뉴클레오티드 70,595이다. 시작 코돈에 인접하여 5'인 뉴클레오티드는 뉴클레오티드-1이다.

[0072]

본 개시내용에서, 유전자 내의 엑손 및 인트론은 시작 코돈의 제1 뉴클레오티드를 포함하는 엑손에 대해 특정된다. 시작 코돈의 제1 뉴클레오티드를 포함하는 엑손은 엑손 1이다. 엑손 3' 내지 엑손 1은 5' 내지 3': 엑손 2, 엑손 3 등이다. 인트론 3' 내지 엑손 1은 5' 내지 3': 인트론 1, 인트론 2 등이다. 따라서, 유전자는 5' 내지 3': 엑손 1, 인트론 1, 엑손 2, 인트론 2, 엑손 3 등을 포함한다. 인간 HBB 유전자의 예시적인 엑손 1은 NCBI 참조 서열: NG\_000007.3이의 뉴클레오티드 70,545 내지 70,686이다. 인간 HBB 유전자의 예시적인 인트론 1은 NCBI 참조 서열: Ng\_000007.3의 뉴클레오티드 70,687 내지 70,816이다. 당업자는 유전자가 다수의 상이한 mRNA로 전사될 수 있음을 이해할 것이다. 이와 같이, 유전자(예를 들어, HBB)는 다수의 상이한 엑손 및 인트론 세트를 가질 수 있다.

[0073]

본원에 사용된 용어 "통합"은 교정 게놈과 표적 유전자 사이의 상동 재조합에 의해 편집 요소가 표적 좌위에 도

입되는 것을 지칭한다. 편집 요소의 통합은 표적 유전자에서 하나 이상의 뉴클레오티드의 치환, 삽입 및/또는 결실을 초래할 수 있다.

[0074] 본원에 사용된 용어 "편집 요소의 표적 좌위로의 통합 효율"은 편집 요소의 표적 좌위로의 통합이 발생한 형질도입된 집단에서 세포의 백분율을 지칭한다.

[0075] 본원에 사용된 용어 "편집 요소의 표적 좌위로의 통합의 대립 유전자 빈도"는 편집 요소의 표적 좌위로의 통합이 발생한 형질도입된 세포의 집단에서 대립 유전자의 백분율을 지칭한다.

[0076] 본원에 사용된 용어 "표준 AAV 형질도입 조건"은  $2 \times 10^5$  CD34<sup>+</sup> 인간 줄기 세포에  $1.5 \times 10^5$ 의 감염 다중도(MOI: multiplicity of infection)에서 AAV의 형질도입을 지칭하며, 상기 세포는 5% 이산화탄소(CO<sub>2</sub>)의 인큐베이션 환경에서 37°C에서 20% 우태혈청(FCS: Fetal Calf Serum), 100 μg/mL 스트렙토마이신, 100 U/mL 폐니실린, 2mmol/L L-글루타민, 10 ng/mL 인간 IL-3, 10 ng/mL 인간 IL-6, 및 1 ng/mL 인간 SCF로 보충된 이스코브 변형된 둘베코 배지(IMDM: Iscove's Modified Dulbecco's Medium)로 배양되고, 상기 AAV는 인산염 완충 식염수(PBS: phosphate buffered saline)에서 제형화되며, 상기 AAV는 배양 배지의 부피의 9분의 1 이하의 부피로 CD34<sup>+</sup> 세포를 함유하는 세포 배양 배지에 첨가된다.

[0077] 본원에 사용된 바와 같이, 대상체에게 AAV의 투여와 관련하여 용어 "유효량"은 원하는 예방 또는 치료 효과를 달성하는 AAV의 양을 지칭한다.

[0078] 본원에 사용된 용어 "적혈구 전구세포"는 적혈구로 분화될 수 있는 세포를 지칭한다. 특정 실시형태에서, 상기 적혈구 전구세포는 다능성 줄기 세포이다. 특정 실시형태에서, 상기 적혈구 전구세포는 유도된 다능성 줄기 세포이다. 특정 실시형태에서, 상기 적혈구 전구세포는 조혈 줄기 세포이다. 특정 실시형태에서, 상기 적혈구 전구세포는 CD34<sup>+</sup> 조혈 줄기 세포이다. 특정 실시형태에서, 상기 적혈구 전구세포는 골수성 전구세포이다. 특정 실시형태에서, 상기 적혈구 전구세포는 거핵구 적혈구 전구세포이다. 특정 실시형태에서, 상기 적혈구 전구세포는 적혈구 전구체 세포이다.

## II. 아데노-연관 바이러스 조성물

[0080] 일 양태에서, 본 개시내용은 HBB 유전자에서 돌연변이를 교정하는 데 유용한 신규한 복제-결합 AAV 조성물을 제공한다. 본원에 개시된 AAV는 일반적으로 AAV Clade F 캡시드 단백질을 포함하는 AAV 캡시드; 및 HBB 유전자에서 표적 좌위 편집용 교정 게놈을 포함한다.

[0081] 임의의 AAV Clade F 캡시드 단백질 또는 이의 유도체가 본원에 개시된 AAV 조성물에 사용될 수 있다. 예를 들어, 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 203 내지 736의 아미노산 서열과 적어도 80%, 85%, 90%, 95%, 또는 99%의 서열 동일성을 갖는 아미노산 서열을 포함한다. 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 203 내지 736의 아미노산 서열과 적어도 80%, 85%, 90%, 95%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하며, 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이며; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이며, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단

백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이며, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이다. 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 203 내지 736의 아미노산 서열을 포함한다.

예를 들어, 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 138 내지 736의 아미노산 서열과 적어도 80%, 85%, 90%, 95%, 또는 99%의 서열 동일성을 갖는 아미노산 서열을 포함한다. 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2의 아미노산 138 내지 736의 아미노산 서열과 적어도 80%, 85%, 90%, 95%, 또는 99% 서열 동일성을 갖는 아미노산 서열을 포함하며, 서열 번호: 2의 아미노산 151에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 160에 상응하는 캡시드 단백질에서 아미노산은 D이고; 서열 번호: 2의 아미노산 206에 상응하는 캡시드 단백질에서 아미노산은 C이고; 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고; 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이고; 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이고; 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고; 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이고; 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고; 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이며; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이며, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 296에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이며, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이다. 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 138 내지 736의 아미노산 서열을 포함한다.

시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 590에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G 또는 Y이고; 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이고; 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이고; 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이고; 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이며; 또는 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 G이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 2에 상응하는 캡시드 단백질에서 아미노산은 T이며, 서열 번호: 2의 아미노산 312에 상응하는 캡시드 단백질에서 아미노산은 Q이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 65에 상응하는 캡시드 단백질에서 아미노산은 I이며, 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 Y이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 77에 상응하는 캡시드 단백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 690에 상응하는 캡시드 단백질에서 아미노산은 K이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 119에 상응하는 캡시드 단백질에서 아미노산은 L이며, 서열 번호: 2의 아미노산 468에 상응하는 캡시드 단백질에서 아미노산은 S이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 626에 상응하는 캡시드 단백질에서 아미노산은 G이며, 서열 번호: 2의 아미노산 718에 상응하는 캡시드 단백질에서 아미노산은 H이고, 서열 번호: 2의 아미노산 464에 상응하는 캡시드 단백질에서 아미노산은 N이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 681에 상응하는 캡시드 단백질에서 아미노산은 M이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 687에 상응하는 캡시드 단백질에서 아미노산은 R이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 346에 상응하는 캡시드 단백질에서 아미노산은 A이며, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이다. 특정 실시형태에서, 서열 번호: 2의 아미노산 501에 상응하는 캡시드 단백질에서 아미노산은 I이고, 서열 번호: 2의 아미노산 505에 상응하는 캡시드 단백질에서 아미노산은 R이며, 서열 번호: 2의 아미노산 706에 상응하는 캡시드 단백질에서 아미노산은 C이다. 특정 실시형태에서, 상기 AAV Clade F 캡시드 단백질은 서열 번호: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 1 내지 736의 아미노산 서열을 포함한다.

[0084]

특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 203 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; (b) 서열 번호: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 138 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 1 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질 중 둘 이상을 포함한다. 특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 서열 203 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질; (b) 서열 번호: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 서열 138 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 또는 17의 아미노산 1 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질을 포함한다.

[0085]

특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 8의 아미노산 203 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; (b) 서열 번호: 8의 아미노산 138 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 8의 아미노산 1 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질 중 하나 이상을 포함한다. 특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 8의 아미노산 203 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; (b) 서열 번호: 8의 아미노산 138 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 8의 아미노산 1 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질 중 둘 이상을 포함한다. 특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 8의 아미노산 203 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질; (b) 서열 번호: 8의 아미노산 138 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 8의 아미노산 1 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질을 포함한다.

[0086]

특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 13의 아미노산 203 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; (b) 서열 번호: 13의 아미노산 138 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 13의 아미노산 1 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질 중 하나 이상을 포함한다. 특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 13의 아미노산 203 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; (b) 서열 번호: 13의 아미노산 138 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질을 포함한다.

산 서열을 포함하는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 13의 아미노산 1 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질 중 둘 이상을 포함한다. 특정 실시형태에서, 상기 AAV 캡시드: (a) 서열 번호: 13의 아미노산 203 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질; (b) 서열 번호: 13의 아미노산 138 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 13의 아미노산 1 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질을 포함한다.

[0087] 특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 16의 아미노산 203 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; (b) 서열 번호: 16의 아미노산 138 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질 중 하나 이상을 포함한다. 특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 16의 아미노산 203 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질; (b) 서열 번호: 16의 아미노산 138 내지 736을 포함하는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 16의 아미노산 1 내지 736의 아미노산 서열을 포함하는 Clade F 캡시드 단백질 중 둘 이상을 포함한다. 특정 실시형태에서, 상기 AAV 캡시드는: (a) 서열 번호: 16의 아미노산 203 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질; (b) 서열 번호: 16의 아미노산 138 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질; 및 (c) 서열 번호: 16의 아미노산 1 내지 736으로 이루어진 아미노산 서열을 갖는 Clade F 캡시드 단백질을 포함한다.

[0088] 본원에 개시된 AAV 조성물에서 유용한 교정 계놈은: (i) 표적 유전자에서 표적 좌위의 편집용 편집 요소, (ii) 상기 표적 좌위에 대한 제1 계놈 영역 5'과 상동성을 갖는 상기 편집 요소의 5' 상동성 암 뉴클레오티드 서열 5', 및 (iii) 상기 표적 좌위에 대한 제2 계놈 영역 3'과 상동성을 갖는 상기 편집 요소의 3' 상동성 암 뉴클레오티드 서열 3'을 포함한다. 특정 실시형태에서, 상기 교정 계놈은 상기 5' 상동성 암 뉴클레오티드 서열의 5' 반전된 말단 반복(5' ITR) 뉴클레오티드 서열 5', 및 상기 3' 상동성 암 뉴클레오티드 서열의 3' 반전된 말단 반복(3' ITR) 뉴클레오티드 서열 3'을 포함한다.

[0089] 본원에 개시된 교정 계놈에서 유용한 교정 요소는 표적 좌위에서 하나 이상의 뉴클레오티드의 삽입, 결실, 또는 치환을 매개할 수 있다. 상기 표적 좌위는 적혈구 전구세포에서 발현된 HBB 유전자 또는 다른 유전자일 수 있는 표적 유전자 내에 완전히 또는 부분적으로 위치할 수 있다.

[0090] 특정 실시형태에서, 표적 좌위에서 상동 재조합에 의해 정확하게 통합될 때, 편집 요소는 HBB 유전자에서 돌연변이를 야생형 HBB 서열 또는 야생형 HBB 단백질 또는 이의 기능적 등가물을 인코딩하는 침묵 변경된 서열로 되돌려 교정한다. HBB 유전자에서 대부분의 돌연변이는 본원에 개시된 편집 요소에 의해 교정될 수 있다. 특정 실시형태에서, 상기 편집 요소는 HBB 유전자에서 치환 또는 결실 돌연변이를 교정하는 하나 이상의 뉴클레오티드이다. 특정 실시형태에서, 상기 편집 요소는 HBB 유전자에서 삽입 돌연변이를 결실하는 뉴클레오티드 간 결합이다. 특정 실시형태에서, 상기 편집 요소는 HBB 유전자의 하나 이상의 코딩 엑손을 포함한다. 예를 들어, 상기 편집 요소는 엑손 1의 코딩 영역, 전체 인트론 1, 전체 엑손 2, 전체 인트론 2, 및 엑손 3의 코딩 영역을 포함하는 HBB 유전자의 일부를 포함할 수 있다. 상기 엑손은 본원에 개시된 바와 같이 야생형일 수 있거나 침묵 변경될 수 있다.

[0091] 특정 실시형태에서, 상기 편집 요소는 HBB 유전자의 코딩 서열 또는 스터퍼-삽입된 코딩 서열의 적어도 일부를 포함한다. 특정 실시형태에서, 상기 편집 요소는 HBB 유전자의 코딩 서열 또는 스터퍼-삽입된 코딩 서열의 전부 또는 실질적으로 전부를 포함한다. 예를 들어, 특정 실시형태에서 상기 편집 요소는 HBB 코딩 서열의 뉴클레오티드 4 내지 444, 또는 뉴클레오티드 4로부터 정지 코돈까지의 HBB 스터퍼-삽입된 코딩 서열의 일부를 포함하고, HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열의 일부에 외인성 폴리아데닐화 서열 3'을 선택적으로 더 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 48을 인코딩하는 뉴클레오티드 서열을 포함하고, 상기 서열 번호: 48을 인코딩하는 뉴클레오티드 서열에 외인성 폴리아데닐화 서열 3'을 선택적으로 더 포함할 수 있다. 특정 실시형태에서, 이러한 편집 요소는 상동 재조합에 의해 HBB 유전자의 내인성 시작 코돈 직후(예를 들어, HBB 유전자의 뉴클레오티드 3과 뉴클레오티드 4 사이) 엑손 1에 통합될 수 있으며, 이에 의하여 편집 요소의 통합은 내인성 HBB 유전자의 시작 코돈과 프레임 내(in-frame)에서 완전한 HBB 코딩 서열을 생성시킨다. 특정 실시형태에서, 이러한 편집 요소는 상동 재조합에 의해 비-HBB 표적 유전자의 내인성 시작 코돈 직후(예를 들어, 표적 유전자의 뉴클레오티드 3과 뉴클레오티드 4 사이) 엑손 1에 통합될 수 있으며, 이에 의하여 편집 요소의 통합은 내인성 표적 유전자의 시작 코돈과 프레임 내에서 완전한 HBB 코딩 서열을 생성시킨다. 편집 요소에서 HBB 코딩 서열의 일부는 본원에 개시된 바와 같이 야생형이거나 침묵 돌연변이될 수 있다. 편집 요소에 의해 인코딩된 HBB 아미노산 서열의 일부는 야생형 또는 이의 기능적 등가물일 수 있다.

[0092]

특정 실시형태에서, 상기 편집 요소는 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열(예를 들어, 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열)의 적어도 일부, 및 리보솜 스키핑 요소 또는 외인성 폴리아데닐화 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3' 리보솜 스키핑 요소, 및 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열(예를 들어, 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열)의 적어도 일부를 포함한다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 리보솜 스키핑 요소; 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열; 및 외인성 폴리아데닐화 서열을 포함한다. 특정 실시형태에서, 상기 언급된 편집 요소는 상동 재조합에 의해 표적 유전자의 코딩 뉴클레오티드에 인접하게 3'(예를 들어, 천연 HBB 유전자의 정지 코돈에 대해 인접하게 5')에 통합되어 5' 내지 3': 코딩 뉴클레오티드에서 끝나는 표적 유전자의 5' 부분, 리보솜 스키핑 요소, 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열, 및 외인성 폴리아데닐화 서열을 포함하는 재조합 표적 유전자를 생성할 수 있으며, 상기 리보솜 스키핑 요소는 표적 유전자의 코딩 영역 및 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열과 프레임 내가 되도록 위치된다. 이 재조합 표적 유전자의 발현은 인코딩된 리보솜 스키핑 웨티드의 제1 부분에 융합된 표적 유전자의 N-말단 부분에 의해 인코딩된 아미노산 서열을 포함하는 제1 폴리웨티드, 및 완전한 HBB 아미노산 서열에 융합된 인코딩된 리보솜 스키핑 웨티드의 제2 부분(예를 들어, 단일 프롤린 잔기)을 포함하는 제2 폴리웨티드를 생성한다. 편집 요소에서 HBB 스터퍼-삽입된 코딩 서열의 HBB 코딩 서열 또는 코딩 영역은 본원에 개시된 바와 같이 야생형이거나 침묵 돌연변이될 수 있다. 편집 요소에 의해 인코딩된 HBB 아미노산 서열은 야생형 또는 이의 기능적 등가물(예를 들어, 제1 메티오닌이 없음)일 수 있다. 상기 표적 좌위는 뉴클레오티드 간 결합 또는 표적 유전자의 코딩 뉴클레오티드에 인접하게 3'인 뉴클레오티드 서열일 수 있다. 특정 실시형태에서, 상기 표적 좌위는 HBB 유전자의 천연 정지 코돈으로 구성된다.

[0093]

특정 실시형태에서, 상기 편집 요소는 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열(예를 들어, 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열)의 적어도 일부, 및 스플라이스 수용체 부위; 스플라이스 공여체 부위; 리보솜 스키핑 요소; 및 외인성 폴리아데닐화 서열 중 하나 이상을 포함한다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 스플라이스 수용체 부위; 리보솜 스키핑 요소; 및 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열(예를 들어, 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열)의 적어도 일부를 포함한다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 스플라이스 수용체 부위; 리보솜 스키핑 요소; 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열; 및 외인성 폴리아데닐화 서열을 포함한다. 특정 실시형태에서, 상기 언급된 편집 요소는 상동 재조합에 의해 표적 유전자의 인트론(예를 들어, 내인성 HBB 유전자의 인트론 1)에 통합되어 5' 내지 3': 표적 유전자의 하나 이상의 엑손 5' 내지 인트론; 외인성 스플라이스 공여체 부위를 포함하는 인트론의 5' 부분; 스플라이스 수용체 부위; 리보솜 스키핑 요소, 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열; 및 외인성 폴리아데닐화 서열을 포함하는 재조합 HBB 유전자를 생성할 수 있으며, 리보솜 스키핑 요소는 완전한 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열과 프레임 내에 있도록, 그리고 표적 유전자의 내인성 스플라이스 공여체 부위에 대한 스플라이스 수용체 부위의 스플라이싱은 리보솜 스키핑 요소를 표적 유전자의 코딩 영역과 프레임 내에 놓이도록 위치된다. 이 재조합 표적 유전자의 발현은 인코딩된 리보솜 스키핑 웨티드의 제1 부분에 융합된 삽입 부위에 내인성 엑손(들) 5'에 의해 인코딩된 표적 유전자 아미노산 서열을 포함하는 제1 폴리웨티드, 및 완전한 HBB 아미노산 서열에 융합된 인코딩된 리보솜 스키핑 웨티드(예를 들어, 단일 프롤린 잔기)의 제2 부분을 포함하는 제2 폴리웨티드를 생성한다. 편집 요소에서 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열은 본원에 개시된 바와 같이 야생형이거나 침묵 돌연변이될 수 있다. 편집 요소에 의해 인코딩된 HBB 아미노산 서열은 야생형 또는 이의 기능적 등가물(예를 들어, 제1 메티오닌이 없음)일 수 있다. 상기 표적 좌위는 뉴클레오티드 간 결합 또는 표적 유전자의 인트론에서 뉴클레오티드에 인접하게 3'인 뉴클레오티드 서열일 수 있다.

[0094]

특정 실시형태에서, 편집 요소 내의 HBB 코딩 서열 또는 스터퍼-삽입된 코딩 서열의 하나 이상의 부분은 야생형 HBB 유전자의 상응하는 엑손과 동일하지 않도록 침묵 변경될 수 있다. 이러한 침묵 변경은 다른 글로빈 유전자 또는 유사 유전자의 좌위, 예를 들어, 베타 글로빈 유사 유전자 좌위로 교정 계획의 통합 가능성을 감소시킨다는 점에서 유리하다. 이러한 침묵 변경은 또한 편집 요소와 표적 유전자 사이의 상동성을 감소시켜, 상동성 암 보다는 편집 요소에 의해 매개되는 바람직하지 않은 통합을 감소시킨다.

[0095]

따라서, 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 유전자의 엑손의 상응하는 코딩 영역에 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된 HBB 유전자의 하나 이상의 엑손의 코딩 영역을 포함한다. 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 유전자의 엑손의 상응하는 코딩 영역에 70% 미만으로 동일하게 침묵 변경된 HBB 유전자의 하나 이상의 엑손의 코딩 영역을 포함한다. 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 유전자의 엑손의 상응하는 코딩 영역에 85% 미만으

로 동일하게 침묵 변경된 HBB 유전자의 하나 이상의 엑손의 코딩 영역을 포함한다.

[0096] 특정 실시형태에서, 상기 편집 요소는 엑손 1의 코딩 영역, 전체 인트론 1, 전체 엑손 2, 전체 인트론 2, 및 엑손 3의 코딩 영역을 포함하는 HBB 유전자의 일부를 포함하며, 엑손 1의 코딩 영역, 전체 엑손 2, 및 엑손 3의 코딩 영역 중 하나 이상은 야생형 HBB 유전자의 엑손의 상응하는 영역에 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된다. 특정 실시형태에서, 상기 편집 요소는 엑손 1의 코딩 영역, 전체 인트론 1, 전체 엑손 2, 전체 인트론 2, 및 엑손 3의 코딩 영역을 포함하는 HBB 유전자의 일부를 포함하며, 엑손 1의 코딩 영역, 전체 엑손 2, 및 엑손 3의 코딩 영역의 하나 이상은 야생형 HBB 유전자의 엑손의 상응하는 영역에 70% 미만으로 동일하게 침묵 변경된다.

[0097] 특정 실시형태에서, 상기 편집 요소는 엑손 1의 코딩 영역, 전체 인트론 1, 전체 엑손 2, 전체 인트론 2, 및 엑손 3의 코딩 영역을 포함하는 HBB 유전자의 일부를 포함하며, 엑손 1의 코딩 영역, 전체 엑손 2, 및 엑손 3의 코딩 영역의 각각의 하나는 야생형 HBB 유전자의 엑손의 상응하는 영역에 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된다. 특정 실시형태에서, 상기 편집 요소는 엑손 1의 코딩 영역, 전체 인트론 1, 전체 엑손 2, 전체 인트론 2, 및 엑손 3의 코딩 영역을 포함하는 HBB 유전자의 일부를 포함하며, 엑손 1의 코딩 영역, 전체 엑손 2, 및 엑손 3의 코딩 영역의 각각의 하나는 야생형 HBB 유전자의 엑손의 상응하는 영역에 70% 미만으로 동일하게 침묵 변경된다.

[0098] 특정 실시형태에서, 상기 편집 요소는 서열 번호: 43 내지 46 및 105 내지 107로 이루어진 군으로부터 선택된 뉴클레오티드 서열의 하나 이상을 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 43 내지 46 및 105 내지 107로 이루어진 군으로부터 선택된 뉴클레오티드 서열의 둘 이상을 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 46에 기재된 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 43, 44, 및 45에 기재된 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 105, 106, 및 107에 기재된 뉴클레오티드 서열을 포함한다.

[0099] 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열의 상응하는 부분에 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된 HBB 코딩 서열의 적어도 일부를 포함한다. 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열의 상응하는 부분에 70% 미만으로 동일하게 침묵 변경된 HBB 코딩 서열의 적어도 일부를 포함한다. 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열의 상응하는 부분에 85% 미만으로 동일하게 침묵 변경된 HBB 코딩 서열의 적어도 일부를 포함한다.

[0100] 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열의 적어도 일부가 야생형 HBB 코딩 서열의 상응하는 부분에 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일한 코딩 서열로 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열의 적어도 일부를 포함한다. 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열의 적어도 일부가 야생형 HBB 코딩 서열의 상응하는 부분과 70%미만으로 동일한 코딩 서열로 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열의 적어도 일부를 포함한다. 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열의 적어도 일부가 야생형 HBB 코딩 서열의 상응하는 부분과 85% 미만으로 동일한 코딩 서열로 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열의 적어도 일부를 포함한다.

[0101] 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열의 상응하는 부분과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된 HBB 코딩 서열의 뉴클레오티드 4 내지 444를 포함한다. 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열의 상응하는 부분과 70% 미만으로 동일하게 침묵 변경된 HBB 코딩 서열의 뉴클레오티드 4 내지 444를 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 47에 기재된 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열의 상응하는 부분과 85% 미만으로 동일하게 침묵 변경된 HBB 코딩 서열의 뉴클레오티드 4 내지 444를 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 100에 기재된 뉴클레오티드 서열을 포함한다. 이러한 편집 요소는 HBB 유전자 코딩 서열에 외인성 폴리아데닐화 서열 3'을 더 포함할 수 있다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': HBB 코딩 서열의 뉴클레오티드 4 내지 444; 및 외인성 폴리아데닐화 서열을 포함하며, HBB 코딩 서열의 뉴클레오티드 4 내지 444는 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된다.

[0102] 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열의 일부가 야생형 HBB 코딩 서열의 상응하는

부분과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일한 HBB 코딩 서열의 뉴클레오티드 4 내지 444에 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열의 일부를 포함한다. 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열의 일부가 야생형 HBB 코딩 서열의 상응하는 부분과 70% 미만으로 동일한 HBB 코딩 서열의 뉴클레오티드 4 내지 444로 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열의 일부를 포함한다. 특정 실시형태에서, HBB 스터퍼-삽입된 코딩 서열의 일부는 서열 번호: 47에 기재된 뉴클레오티드 서열로 전사 및 스플라이싱될 수 있다. 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열의 일부가 야생형 HBB 코딩 서열의 상응하는 부분과 85% 미만으로 동일한 HBB 코딩 서열의 뉴클레오티드 4 내지 444로 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열의 일부를 포함한다. 특정 실시형태에서, HBB 스터퍼-삽입된 코딩 서열의 일부는 서열 번호: 100에 기재된 뉴클레오티드 서열로 전사 및 스플라이싱될 수 있다. 이러한 편집 요소는 HBB 유전자 코딩 서열에 외인성 폴리아데닐화 서열 3'을 더 포함할 수 있다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열; 및 외인성 폴리아데닐화 서열을 포함하며, HBB 스터퍼-삽입된 코딩 서열의 일부는 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일한 HBB 코딩 서열의 뉴클레오티드 4 내지 444로 전사 및 스플라이싱될 수 있다.

[0103] 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된 완전한 HBB 코딩 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열과 70% 미만으로 동일하게 침묵 변경된 완전한 HBB 코딩 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3' 시작 코돈 및 서열 번호: 47에 기재된 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 28에 기재된 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 야생형 HBB 코딩 서열과 85% 미만으로 동일하게 침묵 변경된 HBB 유전자의 완전한 HBB 코딩 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3' 시작 코돈 및 서열 번호: 100에 기재된 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 99에 기재된 뉴클레오티드 서열을 포함한다. 이러한 편집 요소는 HBB 유전자 코딩 서열에 스플라이스 수용체 부위; 리보솜 스키핑 요소; 및 외인성 폴리아데닐화 서열 3'의 하나 이상을 더 포함할 수 있다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 리보솜 스키핑 요소; 및 선택적으로 시작 코돈이 없는 완전한 HBB 코딩 서열을 포함하며, 상기 완전한 HBB 코딩 서열은 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 리보솜 스키핑 요소; 및 선택적으로 시작 코돈이 없는 완전한 HBB 코딩 서열을 포함하며, 상기 완전한 HBB 코딩 서열은 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 스플라이스 수용체 부위; 리보솜 스키핑 요소; 및 선택적으로 시작 코돈이 없는 완전한 HBB 코딩 서열을 포함하며, 상기 완전한 HBB 코딩 서열은 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 스플라이스 수용체 부위; 리보솜 스키핑 요소; 선택적으로 시작 코돈이 없는 완전한 HBB 코딩 서열; 및 외인성 폴리아데닐화 서열을 포함하며, 상기 완전한 HBB 코딩 서열은 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일하게 침묵 변경된다.

[0104] 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열이 야생형 HBB 코딩 서열과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일한 완전한 HBB 코딩 서열로 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열이 야생형 HBB 코딩 서열과 70% 미만으로 동일한 완전한 HBB 코딩 서열로 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열을 포함한다. 특정 실시형태에서, 상기 완전한 HBB 코딩 서열은 5' 내지 3' 시작 코돈 및 서열 번호: 47에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 완전한 HBB 코딩 서열은 서열 번호: 28에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 HBB 스터퍼-삽입된 코딩 서열은 서열 번호: 43 내지 45에 기재된 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 HBB 스터퍼-삽입된 코딩 서열이 야생형 HBB 코딩 서열과 85% 미만으로 동일한 완전한 HBB 코딩 서열로 전사 및 스플라이싱될 수 있도록, 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열을 포함한다. 특정 실시형태에서, 상기 HBB 스터퍼-삽입된 코딩 서열은 서열 번호: 105, 106, 및 107에 기재된 5' 내지 3' 뉴클레오티드 서열을 포함한다. 이러한 편집 요소는 HBB 유전자 코딩 서열에 스플라이스

수용체 부위; 리보솜 스키핑 요소; 및 외인성 폴리아데닐화 서열 3'의 하나 이상을 더 포함할 수 있다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 리보솜 스키핑 요소; 및 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열을 포함하며, 상기 HBB 스터퍼-삽입된 코딩 서열은 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일한 완전한 HBB 코딩 서열로 전사 및 스플라이싱될 수 있다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 리보솜 스키핑 요소; 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열; 및 외인성 폴리아데닐화 서열을 포함하며, 상기 HBB 스터퍼-삽입된 코딩 서열은 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일한 완전한 HBB 코딩 서열로 전사 및 스플라이싱될 수 있다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 스플라이스 수용체 부위; 리보솜 스키핑 요소; 및 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열을 포함하며, 상기 HBB 스터퍼-삽입된 코딩 서열은 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일한 완전한 HBB 코딩 서열로 전사 및 스플라이싱될 수 있다. 특정 실시형태에서, 상기 편집 요소는 5' 내지 3': 스플라이스 수용체 부위; 리보솜 스키핑 요소; 침묵 변경된 HBB 스터퍼-삽입된 코딩 서열; 및 외인성 폴리아데닐화 서열을 포함하며, 상기 HBB 스터퍼-삽입된 코딩 서열은 야생형 HBB 코딩 서열의 상응하는 영역과 100% 미만(예를 들어, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만)으로 동일한 완전한 HBB 코딩 서열로 전사 및 스플라이싱될 수 있다.

[0105]

본원에 개시된 임의의 및 모든 편집 요소는 표적 유전자에 존재하지 않는 독특한 서열(예를 들어, 야생형 또는 돌연변이체 HBB 유전자, 또는 그의 기능적 등가물을 인코딩하는 유전자)을 더 포함할 수 있고, 따라서 표적 좌위에서 편집 요소의 통합을 갖는 세포의 동정을 허용하도록 한다. 이러한 독특한 서열은 표적 좌위 및 그의 측면 영역 또는 이로부터 증폭된 핵산의 핵산 시퀀싱 분석(예를 들어, PCR 또는 차세대 시퀀싱)에 적합한 서열일 수 있다. 이러한 독특한 서열은 또한 표적 좌위 및 그의 측면 영역 또는 이로부터 증폭된 핵산의 제한 단편 길이 다형성 분석에 기초하여 표적 좌위에서 편집 요소의 통합을 갖는 세포의 동정을 허용하는 제한 엔도뉴클레아제 부위일 수 있다.

[0106]

본원에 개시된 임의의 및 모든 편집 요소는 표적 좌위에 통합될 때 베타 글로빈 단백질에서 하나 이상의 아미노산 변형(예를 들어, 치환, 삽입 또는 결실)을 야기하는 하나 이상의 뉴클레오티드 변경을 포함할 수 있다. 특정 실시형태에서, 상기 변형된 베타 글로빈 단백질은 야생형 베타 글로빈의 기능적 등가물이며, 즉 야생형 베타 글로빈으로서 기능할 수 있다. 특정 실시형태에서, 기능적으로 등가인 베타 글로빈은 야생형 베타 글로빈에서 발견되지 않은 적어도 하나의 특성, 예를 들어 SCD 돌연변이를 보유한 베타 글로빈의 응집을 억제하는 능력을 더 포함한다.

[0107]

특정 실시형태에서, 본원에 기재된 바와 같은 편집 요소는 적어도 0, 1, 2, 10, 100, 200, 500, 1000, 1500, 2000, 3000, 4000 또는 5000개의 뉴클레오티드를 포함한다. 특정 실시형태에서, 상기 편집 요소는 1 내지 5000, 1 내지 4500, 1 내지 4000, 1 내지 3000, 1 내지 2000, 1 내지 1000, 1 내지 500, 1 내지 200, 1 내지 100, 1 내지 50, 또는 1 내지 10개의 뉴클레오티드를 포함하거나 이로 구성된다.

[0108]

특정 실시 형태에서, 본원에 기재된 바와 같은 편집 요소는 엑손, 인트론, 5' 비번역된 영역(UTR), 3' UTR, 프로모터, 스플라이스 공여체, 스플라이스 수용체, 리보솜 스키핑 요소, 비코딩 RNA를 인코딩하는 서열, 인슐레이터(insulator), 유전자, 또는 이들의 조합을 포함하거나 이로 구성된다.

[0109]

특정 실시형태에서, 상기 편집 요소는 서열 번호: 23 내지 28 중 어느 하나에 기재된 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 편집 요소는 서열 번호: 23 내지 26 중 어느 하나에 기재된 뉴클레오티드 서열로 구성된다.

[0110]

본원에 개시된 교정 게놈에 사용되는 상동성 암은 표적 유전자(예를 들어, HBB 유전자) 또는 게놈 상 부근의 유전자의 임의의 영역으로 향할 수 있다. 상동성 암의 정확한 동일성 및 위치는 편집 요소 및/또는 표적 좌위의 동일성에 의해 결정된다.

[0111]

본원에 개시된 교정 게놈에 사용된 상동성 암은 표적 좌위(예를 들어, HBB 유전자에서 표적 좌위) 측면에 있는 게놈과 실질적으로 동일하다. 특정 실시형태에서, 상기 5' 상동성 암은 표적 좌위에 대한 제1 게놈 영역 5'과 적어도 약 90%(예를 들어, 적어도 약 95%, 96%, 97%, 98%, 99%, 또는 99.5%) 뉴클레오티드 서열 동일성을 갖는다. 특정 실시형태에서, 상기 5' 상동성 암은 상기 제1 게놈 영역과 100% 뉴클레오티드 서열 동일성을 갖는다. 특정 실시형태에서, 상기 3' 상동성 암은 표적 좌위에 대한 제2 게놈 영역 3'과 적어도 약 90%(예를 들어, 적어도 약 95%, 96%, 97%, 98%, 99%, 또는 99.5%) 뉴클레오티드 서열 동일성을 갖는다. 특정 실시형태에서, 상기

3' 상동성 암은 상기 제2 게놈 영역과 100% 뉴클레오티드 서열 동일성을 갖는다. 특정 실시형태에서, 상기 5' 및 3' 상동성 암은 표적 좌위에 대한 제1 게놈 영역 5' 및 표적 좌위에 대한 제2 게놈 영역 3' 각각에 각각 적어도 약 90%(예를 들어, 적어도 약 95%, 96%, 97%, 98%, 99%, 또는 99.5%) 동일하다. 특정 실시형태에서, 상기 5' 및 3' 상동성 암은 상기 제1 및 제2 게놈 영역 각각에 각각 100% 동일하다. 특정 실시형태에서, 상기 5' 상동성 암 및 상기 제1 게놈 영역의 뉴클레오티드 서열에서 차이 및/또는 상기 3' 상동성 암 및 상기 제2 게놈 영역의 뉴클레오티드 서열에서 차이는 뉴클레오티드 서열에서 비-코딩 영역 차이를 포함하고, 필수적으로 구성되거나 구성된다.

[0112] 당업자는 상동성 암이 상동 재조합에 의해 편집 요소의 해당 표적 부위로의 통합을 매개할 수 있도록 표적 좌위 측면에 있는 게놈 서열에 100% 동일할 필요는 없음을 이해할 것이다. 당업자는 상동성 암이 표적 좌위 측면에 있는 게놈 서열과 100% 동일하지 않은 상황에서, 상동성 아암과 게놈 사이의 상동 재조합은 사용된 상동성 암의 서열과 동일하도록 표적 좌위의 측면에 있는 게놈 서열을 변경시킬 수 있다는 것을 더 이해할 것이다.

[0113] 특정 실시형태에서, 표적 좌위에 대한 상기 제1 게놈 영역 5'는 제1 편집창에 위치하며, 상기 제1 편집창의 뉴클레오티드 서열은 서열 번호: 101 내지 103으로 이루어진 군으로부터 선택된 서열로 구성된다. 특정 실시형태에서, 표적 좌위에 대한 상기 제2 게놈 영역 3'은 제2 편집창에 위치하며, 상기 제2 편집창의 뉴클레오티드 서열은 서열 번호: 101 내지 103으로 이루어진 군으로부터 선택된 서열로 구성된다. 특정 실시형태에서, 표적 좌위에 대한 상기 제1 게놈 영역 5'는 제1 편집창에 위치하며, 상기 제1 편집창의 뉴클레오티드 서열은 서열 번호: 101 내지 103으로 이루어진 군으로부터 선택된 서열로 구성되며; 그리고 표적 좌위에 대한 상기 제2 게놈 영역 3'은 제2 편집창에 위치하며, 상기 제2 편집창의 뉴클레오티드 서열은 서열 번호: 101, 102, 또는 103에 기재된 서열로 구성된다.

[0114] 특정 실시형태에서, 상기 제1 및 제2 편집창들은 상이하다. 특정 실시형태에서, 상기 제1 편집창은 상기 제2 편집창에 대해 5'에 위치된다. 특정 실시형태에서, 상기 제1 편집창은 서열 번호: 101에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 제2 편집창은 서열 번호: 102에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 제1 편집창은 서열 번호: 101에 기재된 뉴클레오티드 서열로 구성되며, 상기 제2 편집창은 서열 번호: 102에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 제1 게놈 영역은 상기 제1 편집창의 서열보다 더 짧은 서열로 구성된다. 특정 실시형태에서, 상기 제1 게놈 영역은 상기 제1 편집창의 서열로 구성된다. 특정 실시형태에서, 상기 제2 게놈 영역은 상기 제2 편집창의 서열보다 더 짧은 서열로 구성된다. 특정 실시형태에서, 상기 제2 게놈 영역은 상기 제2 편집창의 서열로 구성된다.

[0115] 특정 실시형태에서, 상기 제1 및 제2 편집창들은 동일하다. 특정 실시형태에서, 상기 좌위는 편집창에서의 뉴클레오티드 간 결합 또는 뉴클레오티드 서열이며, 상기 제1 게놈 좌위는 표적 좌위에 대한 편집창 5'의 제1 부분으로 구성되며, 상기 제2 게놈 좌위는 표적 좌위에 대한 편집창 3'의 제2 부분으로 구성된다. 특정 실시형태에서, 상기 편집창의 제1 부분은 편집창의 5' 말단부터 표적 좌위에 5' 인접한 뉴클레오티드까지의 서열로 구성된다. 특정 실시형태에서, 상기 편집창의 제2 부분은 표적 좌위에 3' 인접한 뉴클레오티드로부터 편집창의 3' 말단까지의 서열로 구성된다. 특정 실시형태에서, 상기 편집창의 제1 부분은 편집창의 5' 말단부터 표적 좌위에 5' 인접한 뉴클레오티드까지의 서열로 구성되며, 그리고 상기 편집창의 제2 부분은 표적 좌위에 3' 인접한 뉴클레오티드로부터 편집창의 3' 말단까지의 서열로 구성된다. 특정 실시형태에서, 상기 편집창은 서열 번호: 103에 기재된 뉴클레오티드 서열로 구성된다. 특정 실시형태에서, 상기 편집창의 제1 및 제2 부분은 실질적으로 동일한 길이를 갖는다(예를 들어, 짧은 부분의 길이 대 긴 부분의 길이의 비가 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, 또는 0.99를 초과한다).

[0116] 특정 실시형태에서, 상기 5' 상동성 암은 약 50 내지 약 4000 뉴클레오티드(예를 들어, 약 100 내지 약 3000, 약 200 내지 약 2000, 약 500 내지 약 1000 뉴클레오티드)의 길이를 갖는다. 특정 실시형태에서, 상기 5' 상동성 암은 약 800 뉴클레오티드의 길이를 갖는다. 특정 실시형태에서, 상기 5' 상동성 암은 약 100 뉴클레오티드의 길이를 갖는다. 특정 실시형태에서, 상기 3' 상동성 암은 약 50 내지 약 4000 뉴클레오티드(예를 들어, 약 100 내지 약 3000, 약 200 내지 약 2000, 약 500 내지 약 1000 뉴클레오티드)의 길이를 갖는다. 특정 실시형태에서, 상기 3' 상동성 암은 약 100 뉴클레오티드의 길이를 갖는다. 특정 실시형태에서, 상기 5' 및 3' 상동성 암의 각각은 독립적으로 약 50 내지 약 4000 뉴클레오티드(예를 들어, 약 100 내지 약 3000, 약 200 내지 약 2000, 약 500 내지 약 1000 뉴클레오티드)의 길이를 갖는다. 특정 실시형태에서, 상기 5' 및 3' 상동성 암은 약 800 뉴클레오티드의 길이를 갖는다.

- [0117] 특정 실시형태에서, 상기 5' 및 3' 상동성 암은 실질적으로 동일한 뉴클레오티드 길이를 갖는다. 특정 실시형태에서, 상기 5' 및 3' 상동성 암은 비대칭 뉴클레오티드 길이를 갖는다. 특정 실시형태에서, 뉴클레오티드의 비대칭성은 길이가 최대 80%, 70%, 60%, 50%, 40%, 30%, 20%, 또는 10% 차이와 같이, 길이가 최대 90%의 5' 및 3' 상동성 암 사이의 차이에 의해 정의된다.
- [0118] 특정 실시형태에서, 상기 교정 게놈은 서열 번호: 29 내지 42 중 어느 하나에 기재된 뉴클레오티드 서열을 포함한다.
- [0119] 특정 실시형태에서, 상기 교정 게놈은 5' 내지 3': 서열 번호: 101에 기재된 서열; 리보솜 스키핑 요소; 야생형 HBB 코딩 영역에 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만으로 동일하게 침묵 변경된 HBB 코딩 서열; 외인성 폴리아데닐화 서열(예를 들어, 서열 번호: 76, 77, 78, 또는 79에 기재된 바와 같이); 및 서열 번호: 102에 기재된 서열을 포함한다. 특정 실시형태에서, 상기 교정 게놈은 5' 내지 3': 서열 번호: 101에 기재된 서열; 리보솜 스키핑 요소; 서열 번호: 99에 기재된 서열; 외인성 폴리아데닐화 서열(예를 들어, 서열 번호: 76, 77, 78, 또는 79에 기재된 바와 같이); 및 서열 번호: 102에 기재된 서열을 포함한다. 특정 실시형태에서, 상기 교정 게놈은 5' 내지 3': 서열 번호: 101에 기재된 서열; 리보솜 스키핑 요소; HBB의 제1 엑손의 야생형 코딩 영역에 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만으로 동일하게 침묵 변경된 HBB의 제1 엑손의 코딩 영역; 선택적 제1 비천연 인트론; HBB의 야생형 제2 엑손에 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만으로 동일하게 침묵 변경된 HBB의 제2 엑손; HBB의 제3 엑손의 야생형 코딩 영역에 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, 또는 50% 미만으로 동일하게 침묵 변경된 HBB의 제3 엑손의 코딩 영역; 외인성 폴리아데닐화 서열(예를 들어, 서열 번호: 76, 77, 78, 또는 79에 기재된 바와 같이); 및 서열 번호: 102에 기재된 서열을 포함한다. 특정 실시형태에서, 상기 교정 게놈은 5' 내지 3': 서열 번호: 101에 기재된 서열; 리보솜 스키핑 요소; 서열 번호: 105에 기재된 서열; 선택적 제1 비천연 인트론; 서열 번호: 106에 기재된 서열; 선택적 제2 인트론; 서열 번호: 107에 기재된 서열; 외인성 폴리아데닐화 서열(예를 들어, 서열 번호: 76, 77, 78, 또는 79에 기재된 바와 같이); 및 서열 번호: 102에 기재된 서열을 포함한다. 특정 실시형태에서, 상기 교정 게놈은 서열 번호: 104에 기재된 뉴클레오티드 서열을 포함한다.
- [0120] 특정 실시형태에서, 본원에 개시된 교정 게놈은 상기 5' 상동성 암 뉴클레오티드 서열의 5' 반전된 말단 반복(5' ITR) 뉴클레오티드 서열 5', 및 상기 3' 상동성 암 뉴클레오티드 서열의 3' 반전된 말단 반복(3' ITR) 뉴클레오티드 서열 3'을 더 포함한다. 임의의 AAV 혈청형 또는 이의 변이체로부터의 ITR 서열은 본원에 개시된 교정 게놈에 사용될 수 있다. 5' 및 3' ITR은 동일한 혈청형의 AAV 또는 상이한 혈청형의 AAV에서 유래할 수 있다. 본원에 개시된 교정 게놈에 사용하기 위한 예시적인 ITR은 본원의 서열 번호: 18 내지 21에 기재되어 있다. 특정 실시형태에서, 상기 5' ITR 뉴클레오티드 서열 및 상기 3' ITR 뉴클레오티드 서열은 실질적으로 서로 상보적이다(예를 들어, 상기 5' 또는 3' ITR에서 1, 2, 3, 4, 또는 5 뉴클레오티드 위치에서 미스매치(mismatch)를 제외하고 서로 상보적이다).
- [0121] 특정 실시형태에서, 상기 5' ITR 또는 3' ITR는 AAV2로부터 유래된다. 특정 실시형태에서, 상기 5' ITR 및 상기 3' ITR 양쪽 모두는 AAV2로부터 유래된다. 특정 실시형태에서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 18과 적어도 95%(예를 들어, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%) 서열 동일성을 가지며, 또는 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 19와 적어도 95%(예를 들어, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%) 서열 동일성을 갖는다. 특정 실시형태에서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 18과 적어도 95%(예를 들어, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%) 서열을 가지며, 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 19와 적어도 95%(예를 들어, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%) 서열 동일성을 갖는다. 특정 실시형태에서, 상기 교정 게놈은 서열 번호: 23 내지 28 중 어느 하나에 기재된 뉴클레오티드 서열, 서열 번호: 18의 서열을 갖는 5' ITR 뉴클레오티드 서열, 및 서열 번호: 19의 서열을 갖는 3' ITR 뉴클레오티드 서열을 갖는 편집 요소를 포함한다. 특정 실시형태에서, 상기 교정 게놈은 서열 번호: 29 내지 42 중 어느 하나에 기재된 뉴클레오티드 서열, 서열 번호: 18의 서열을 갖는 5' ITR 뉴클레오티드 서열, 및 서열 번호: 19의 서열을 갖는 3' ITR 뉴클레오티드 서열을 포함한다. 특정 실시형태에서, 상기 교정 게놈은 5' 내지 3' 서열 번호: 18의 서열을 갖는 5' ITR 뉴클레오티드 서열, 서열 번호: 29 내지 42 중 어느 하나에 기재된 뉴클레오티드 서열, 및 서열 번호: 19의 서열을 갖는 3' ITR 뉴클레오티드 서열로 구성된다.
- [0122] 특정 실시형태에서, 상기 5' ITR 또는 3' ITR는 AAV5로부터 유래된다. 특정 실시형태에서, 상기 5' ITR 및 3' ITR 양쪽 모두는 AAV5로부터 유래된다. 특정 실시형태에서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 20과 적어도 95%(예를 들어, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%) 서열 동일성을 가지며,

또는 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 21과 적어도 95% 서열 동일성을 갖는다. 특정 실시형태에서, 상기 5' ITR 뉴클레오티드 서열은 서열 번호: 20과 적어도 95%(예를 들어, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%) 서열 동일성을 가지며, 상기 3' ITR 뉴클레오티드 서열은 서열 번호: 21과 적어도 95%(예를 들어, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100%) 서열 동일성을 갖는다. 특정 실시형태에서, 상기 교정 게놈은 서열 번호: 23 내지 28 중 어느 하나에 기재된 뉴클레오티드 서열, 서열 번호: 20의 서열을 갖는 5' ITR 뉴클레오티드 서열, 및 서열 번호: 21의 서열을 갖는 3' ITR 뉴클레오티드 서열을 갖는 편집 요소를 포함한다. 특정 실시형태에서, 상기 교정 게놈은 서열 번호: 29 내지 42 중 어느 하나에 기재된 뉴클레오티드 서열, 서열 번호: 20의 서열을 갖는 5' ITR 뉴클레오티드 서열, 및 서열 번호: 21의 서열을 갖는 3' ITR 뉴클레오티드 서열을 갖는 편집 요소를 포함한다. 특정 실시형태에서, 상기 교정 게놈은 5' 내지 3' 서열 번호: 20의 서열을 갖는 5' ITR 뉴클레오티드 서열, 서열 번호: 29 내지 42 중 어느 하나에 기재된 뉴클레오티드 서열, 및 서열 번호: 21의 서열을 갖는 3' ITR 뉴클레오티드 서열로 구성된다.

[0123] 특정 실시형태에서, 본원에 개시된 교정 게놈은 약 0.5 내지 약 8 kb, 및 이들 사이에 둘러싸인 임의의 범위(예를 들어, 약 1 내지 약 5, 약 2 내지 약 5, 약 3 내지 약 5, 약 4 내지 약 5, 약 4.5 내지 약 4.8 또는 약 4.7 kb)의 길이를 갖는다.

[0124] 본원에 개시된 교정 게놈은 편집 요소를 HBB 유전자의 임의의 원하는 표적 좌위에 통합시키도록 구성될 수 있다. 특정 실시형태에서, 상기 표적 좌위는 상응하는 야생형 HBB 유전자 서열에 대한 HBB 유전자 서열에서의 돌연변이(예를 들어, 하나 이상의 뉴클레오티드의 삽입, 결실 또는 치환)이다. 특정 실시형태에서, 상기 표적 좌위는 상기 HBB 유전자에서 뉴클레오티드 점 돌연변이 또는 결실에 있다. 예시적인 HBB 점 돌연변이 또는 결실은 제한없이, 위치 -87에서 G, 위치 -31에서 G, 위치 -30에서 A, 위치 -29에서 G, 위치 -28에서 G, 위치 -10에서 T, 위치 1에서 C, 위치 1에서 A, 위치 2에서 G, 위치 17 및 18에서 C 및 T의 결실, 위치 19에서 A, 위치 20에서 A의 결실, 위치 20에서 T, 위치 25 및 26에서 A 및 A의 결실, 위치 26 이후 G의 추가, 위치 47에서 A, 위치 48에서 A, 위치 51에서 C의 결실, 위치 52에서 A, 위치 58에서 G, 위치 59에서 G, 위치 79에서 A, 위치 82에서 T, 위치 84 이후 C의 추가, 위치 93에서 T, 위치 93에서 A, 위치 97에서 C, 위치 98에서 C, 위치 202에서 G, 위치 208에서 G, 위치 222에서 C, 위치 241 또는 242에서 T의 결실, 위치 254 내지 257에서 T 및 T 및 C 및 T의 결실, 위치 260에서 T, 위치 264 또는 265에서 C의 결실, 위치 343 이후 A의 추가, 위치 399 및 400에서 G 및 T의 결실, 위치 401에서 T, 위치 417 이후 A의 추가, 446에서 A, 위치 1099에서 T, 위치 1293에서 A, 위치 1344에서 T를 포함한다. 특정 실시형태에서, 상기 표적 좌위는 겹상세포질환 돌연변이이다(즉, HBB 유전자의 위치 20에서 T). 특정 실시형태에서, 상기 표적 좌위는 HBB 유전자의 엑손 1에서 염색체 또는 뉴클레오티드 간 결합의 영역, 예를 들어 내인성 시작 코돈 직후(예를 들어, HBB 유전자의 뉴클레오티드 3과 뉴클레오티드 4 사이의 뉴클레오티드 간 결합)이다. 특정 실시형태에서, 상기 표적 좌위는 HBB 유전자의 인트론 1에서 염색체 또는 뉴클레오티드 간 결합의 영역이다. 특정 실시형태에서, 상기 표적 좌위는 야생형 HBB 유전자의 천연 정지 코돈 또는 돌연변이체 HBB 유전자의 상응하는 뉴클레오티드로 구성된다. 특정 실시형태에서, 상기 표적 좌위는 야생형 HBB 유전자의 정지 코돈에 인접한 5' 뉴클레오티드 간 결합 또는 돌연변이체 HBB 유전자의 상응하는 뉴클레오티드 간 결합으로 구성된다.

[0125] 본원에 개시된 AAV 조성물은 생체 내 및 시험관 내 모두에서 고효율로 세포에서 HBB 유전자의 돌연변이를 교정 할 수 있다는 점에서 특히 유리하다. 특정 실시형태에서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 조혈 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합 효율은 적어도 0.1%(예를 들어, 적어도 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95%)이다. 특정 실시형태에서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 조혈 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합의 대립 유전자 빈도는 적어도 0.05%(예를 들어, 적어도 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95%)이다.

[0126] 유전자 편집의 효율을 결정하는 임의의 방법이 사용될 수 있다. 특정 실시형태에서, 개별 세포는 형질도입된 세포의 집단으로부터 분리되고 표적 좌위에 올바르게 통합된 편집 요소의 존재를 동정할 수 있는 PCR 프라이머를 사용하여 단일 세포 PCR을 받는다. 이러한 방법은 변형되지 않은 표적 좌위를 선택적으로 증폭시키는 PCR 프라이머를 사용하여 동일한 세포의 단일 세포 PCR을 더 포함할 수 있다. 이러한 방식으로, 세포의 유전자형이 결정

될 수 있다. 예를 들어, 단일 세포 PCR이 세포가 편집된 표적 좌위 및 변형되지 않은 표적 좌위 양쪽 모두를 가지고 있음을 보이면, 세포는 편집된 HBB 유전자에 대한 이형 접합체로 간주될 것이다.

[0127] 추가적으로 또는 대안적으로, 특정 실시형태에서, 형질도입된 세포의 집단(예를 들어, 조직 또는 기관)으로부터 추출된 DNA에 대해 선형 증폭 매개 PCR(LAM-PCR), 정량적 PCR(qPCR) 또는 디지털 액적 PCR(ddPCR)을 수행하여 통합의 대립 유전자 빈도를 평가할 수 있다. 특정 실시형태에서, 상기 추출된 DNA는 상이한 서열을 검출하는 적어도 2개의 프라이머 쌍을 사용하여 디지털 액적 PCR(ddPCR)에 의해 분석된다. 예를 들어, ddPCR은 비통합 및 통합된 벡터뿐만 아니라, 비편집 및 편집된 표적 좌위를 검출하는 제1 프라이머 쌍, 비통합 및 통합된 벡터에 존재하는 서열을 검출하는 제2 프라이머 쌍, 및 선택적으로 비편집 및 편집된 표적 좌위에 존재하는 상동성 암에서 서열을 검출하는 제3 프라이머 쌍을 이용할 수 있다. 비편집 계놈 DNA 및 비통합 벡터의 공동-파티셔닝(co-partitioning) 가능성의 교정 후, 제1 및 제2 프라이머 쌍 양쪽 모두에 양성인 액적의 백분율은 통합의 대립 유전자 빈도에 해당한다. 본 방법의 일예는 본원의 실시예 1에 기재되어 있다.

[0128] 추가적으로 또는 대안적으로, 특정 실시형태에서, 상기 HBB 좌위는 형질도입된 세포의 집단(예를 들어, 조직 또는 기관)으로부터 추출된 DNA로부터 교정 계놈에 의해 포함된 계놈 영역의 측면에 HBB 유전자의 영역과 결합하는 프라이머를 사용하는 PCR 또는 교정 계놈 내의 영역(예를 들어, 좌위에 비천연인 외인성 서열을 포함하는 영역)과 결합하는 프라이머를 사용하는 선형 증폭 매개 PCR(LAM-PCR)에 의해 증폭될 수 있다. 생성된 PCR 앰플리콘은 형질도입된 세포의 집단에 존재하는 편집 및 비편집된 HBB 대립 유전자의 상대적인 수를 결정하기 위해 단일 문자 차세대 시퀀싱(NGS: next generation sequencing) 기법을 사용하여 개별적으로 시퀀싱될 수 있다. 이 숫자는 편집 요소의 표적 좌위로의 통합의 대립 유전자 빈도를 결정하는 데 사용될 수 있다.

[0129] 다른 양태에서, 본 개시내용은 본원에 개시된 바와 같은 AAV를 약학적으로 허용 가능한 부형제, 보조제, 희석제, 운반체 또는 담체, 또는 이들의 조합과 함께 포함하는 약학적 조성물을 제공한다. "약학적으로 허용 가능한 담체"는 조성물의 활성 성분과 조합될 때 성분이 생물학적 활성을 유지하게 하고, 의도하지 않은 면역 반응과 같은 파괴적인 생리학적 반응을 유발하지 않으면서 임의의 물질을 포함한다. 약학적으로 허용 가능한 담체는 물, 인산염 완충 식염수, 오일/물 혼탁액과 같은 혼탁액 및 습윤제를 포함한다. 이러한 담체를 포함하는 조성물은 문헌[Remington's Pharmaceutical Sciences, current Ed., Mack Publishing Co., Easton Pa. 18042, USA; A. Gennaro (2000) "Remington: The Science and Practice of Pharmacy", 20th edition, Lippincott, Williams, & Wilkins]; 문헌[Pharmaceutical Dosage Forms and Drug Delivery Systems (1999) H. C. Ansel et al, 7th ed., Lippincott, Williams, & Wilkins]; 및 문헌[Handbook of Pharmaceutical Excipients (2000) A. H. Kibbe et al, 3rd ed. Amer. Pharmaceutical Assoc]에 기재된 것과 같은 주지된 종래 방법에 의해 제형화된다.

### III. 사용법

[0130] 다른 양태에서, 본 개시내용은 세포 내 HBB 유전자에서 돌연변이를 교정하는 방법을 제공한다. 상기 방법은 일반적으로 세포에 본원에 개시된 바와 같은 복제-결합 AAV를 형질도입하는 단계를 포함한다. 이러한 방법은 HBB 유전자에서 돌연변이를 교정하는 데 매우 효율적이며 그러한 교정을 용이하게 하기 위하여 외인성 뉴클레아제(예를 들어, 거대뉴클레아제, 징크 핑거 뉴클레아제, 전사 활성화제-유사 뉴클레아제(TALEN: transcriptional activator-like nuclelease), 또는 Cas9와 같은 RNA-가이드 뉴클레아제)의 작용에 의해 표적 좌위에서 계놈의 절단을 필요로 하지 않는다. 따라서, 특정 실시형태에서, 본원에 개시된 방법은 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오타이드 서열을 공동-형질도입하거나 공동-투여하지 않고 세포에 본원에 개시된 바와 같은 복제-결합 AAV를 형질도입하는 단계를 포함한다.

[0132] 본원에 개시된 방법은 HBB 유전자에서 돌연변이를 보유하는 임의의 세포에 적용될 수 있다. 당업자는 적혈구로 분화할 수 있는 세포가 특히 중요하다는 것을 이해할 것이다. 따라서, 특정 실시형태에서, 상기 방법은 다능성 줄기 세포, 유도된 다능성 줄기 세포(iPSC: induced pluripotent stem cell), 및 조혈 줄기 세포(HSC: hematopoietic stem cell)를 포함하나 이에 한정되지 않는 줄기 세포에 적용된다. 상기 방법이 적용될 수 있는 예시적인 HSC는 제한없이 CD34<sup>+</sup> HSC를 포함한다.

[0133] 본원에 개시된 방법은 연구 목적으로 시험 관내에서 수행될 수 있거나 치료 목적으로 생체 외 또는 생체 내에서 수행될 수 있다.

[0134] 특정 실시형태에서, 형질도입될 세포는 대상체로부터 취해지고 본원에 개시된 방법에 따라 생체 외 HBB 유전자에서 돌연변이를 교정하도록 형질도입되며, 이후, 형질도입된 세포는 대상체에게 다시 투여된다. 따라서, 특정

실시형태에서, 본 개시내용은 HBB 유전자 돌연변이와 연관된 질환 또는 장애를 갖는 대상체의 치료 방법에서, 상기 방법은 생체 외 줄기 세포(예를 들어, CD34<sup>+</sup> 조혈 줄기 세포)에 본원에 개시된 바와 같은 복제-결합 AAV를 형질도입하여 형질도입된 세포를 수득하는 단계; 및 상기 형질도입된 세포를 대상체에게 투여하는 단계를 포함하는, 방법을 제공한다. 상기 형질도입된 세포는 대상체에게 투여되기 전에 정확한 유전자 통합을 위해 선택되고/되거나 클론 확장을 위해 배양될 수 있다. 특정 실시형태에서, 형질도입될 줄기 세포는 콜수, 제대혈, 또는 말초 혈액으로부터 수득되며, 상기 줄기 세포는 하나 이상의 세포 마커(예를 들어, 세포 크기, 세포 밀도, 및 CD34와 같은 표면 마커)에 기반한 방법에 의해 선택적으로 선택된다. 특정 실시형태에서, 상기 줄기 세포는 자가성(autologous), 즉 AAV 형질도입 후 세포가 투여될 대상체로부터 유래된다. 특정 실시형태에서, 상기 줄기 세포는 이를 필요로 하는 대상체와 동종이계적이며, 즉 상기 줄기 세포는 수령자 대상체와 유전적으로 동일하지 않은 공여체로부터 수득된다. 따라서, 특정 실시형태에서, 본 개시내용은 HBB 유전자 돌연변이와 연관된 질환 또는 장애를 갖는 대상체의 치료 방법에서, 상기 방법은 생체 외 동종이계 줄기 세포(예를 들어, CD34<sup>+</sup> 조혈 줄기 세포)에 본원에 개시된 바와 같은 복제-결합 AAV를 형질도입하여 형질도입된 세포를 수득하는 단계; 및 상기 형질도입된 세포를 대상체에게 투여하는 단계를 포함하는, 방법을 제공한다. 특정 실시형태에서, 상기 동종이계 줄기 세포는 매칭된(matched) 공여체로부터 유래된다. 당업자는 동종이계 적용을 위해, 형질도입된 세포가 투여 전에 추가적인 변형, 예를 들어 이식편 대 숙주 질환(GVHD: graft-versus-host disease)의 발생을 예방 및/또는 감소시키기 위한 유전적 변형을 요구할 수 있음을 인식할 것이다. 상기 대상체는 인간 대상체 또는 인간 적혈구 전구세포를 함유하는 설치류 대상체(예를 들어, 마우스)일 수 있다. 적합한 마우스 대상체는 제한적으로 인간 줄기 세포(예를 들어, 인간 CD34<sup>+</sup> HSC)가 이식된 마우스를 포함한다. HBB 유전자 돌연변이와 연관된 임의의 질환 또는 장애는 본원에 개시된 방법을 사용하여 치료될 수 있다. 적합한 질환 또는 장애는 제한없이 베타 지중해빈혈증 또는 겸상세포질환을 포함한다. 특정 실시형태에서, 상기 세포는 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 공동-형질도입하지 않고 형질도입된다.

[0135]

특정 실시형태에서, 형질도입될 세포는 포유 동물 대상체에 있으며, AAV는 대상체에서 세포를 형질도입시키기에 효과적인 양으로 대상체에게 투여된다. 따라서, 특정 실시형태에서, 본 개시내용은 HBB 유전자 돌연변이와 연관된 질환 또는 장애를 갖는 대상체의 치료 방법에서, 상기 방법은 일반적으로 유효량의 본원에 개시된 바와 같은 복제-결합 AAV를 대상체에게 투여하는 단계를 포함하는, 방법을 제공한다. 상기 대상체는 인간 대상체, 비인간 영장류 대상체(예를 들어, 마카카 파시콜라리스(*Macaca fascicularis*)), 또는 인간 적혈구 전구세포를 함유하는 설치류 대상체(예를 들어, 마우스)일 수 있다. 적합한 마우스 대상체는 제한없이 인간 줄기 세포(예를 들어, 인간 CD34<sup>+</sup> HSC)가 이식된 마우스를 포함한다. HBB 유전자 돌연변이와 연관된 임의의 질환 또는 장애는 본원에 개시된 방법을 사용하여 치료될 수 있다. 적합한 질환 또는 장애는 제한없이 베타 지중해빈혈증 또는 겸상세포질환을 포함한다. 특정 실시형태에서, 상기 세포는 외인성 뉴클레아제 또는 외인성 뉴클레아제를 인코딩하는 뉴클레오티드 서열을 공동-형질도입하거나 공동-투여하지 않고 형질도입된다.

[0136]

본원에 개시된 방법은 생체 내 및 시험 관내 양쪽 모두에서 고효율로 세포에서 HBB 유전자를 교정할 수 있다는 점에서 특히 유리하다. 특정 실시형태에서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합 효율은 적어도 0.1%(예를 들어, 적어도 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95%)이다. 특정 실시형태에서, AAV가 표준 AAV 형질도입 조건 하에서 CD34<sup>+</sup> 인간 줄기 세포의 집단과 외인성 뉴클레아제의 부재 하에 시험관 내에서 접촉될 때, 편집 요소의 표적 좌위로의 통합의 대립 유전자 빈도는 적어도 0.05%(예를 들어, 적어도 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 또는 95%)이다. 유전자 편집의 효율을 결정하는 임의의 방법이 본원에 기재된 것을 제한없이 포함하여 이용될 수 있다.

[0137]

특정 실시형태에서, 세포에 본원에 개시된 AAV 조성물의 형질도입은 본원에 제공된 바와 같이 또는 당업자에게 공지된 임의의 형질도입 방법에 의해 수행될 수 있다. 특정 실시형태에서, 상기 세포는 50,000; 100,000; 150,000; 200,000; 250,000; 300,000; 350,000; 400,000; 450,000; 또는 500,000의 감염 다중도(MOI: multiplicity of infection)에서 또는 세포의 최적 형질도입을 제공하는 임의의 MOI에서 AAV와 접촉될 수 있다. 특정 실시형태에서, 상기 대상체는 체중 kg 당 약 10<sup>11</sup>, 10<sup>12</sup>, 10<sup>13</sup>, 10<sup>14</sup>, 또는 10<sup>15</sup>개의 벡터 게놈의 용량으로

AAV를 투여받을 수 있다.

[0138] 본원에 개시된 AAV 조성물은 정맥 내, 복장 내, 피하, 근육 내, 비내, 국소 또는 피내 경로를 포함하지만 이에 제한되지 않는 임의의 적절한 경로에 의해 대상체에게 투여될 수 있다. 특정 실시형태에서, 상기 조성물은 정맥 내 주사 또는 피하 주사를 통한 투여를 위해 제형화된다.

#### IV. AAV 패키징 시스템

[0140] 다른 양태에서, 본 개시내용은 본원에 개시된 복제-결합 AAV의 재조합 제조용 패키징 시스템을 제공한다. 이러한 패키징 시스템은 일반적으로: 하나 이상의 AAV Rep 단백질을 인코딩하는 Rep 뉴클레오티드 서열; 본원에 개시된 바와 같은 하나 이상의 AAV Clade F 캡시드 단백질을 인코딩하는 Cap 뉴클레오티드 서열; 및 본원에 개시된 바와 같은 HBB 유전자에서 돌연변이의 교정을 위한 교정 게놈을 포함하며, 상기 패키징 시스템은 상기 캡시드 내에 상기 교정 게놈을 봉입하기 위한 세포 내에서 작용하여 상기 AAV를 형성한다.

[0141] 특정 실시형태에서, 상기 패키징 시스템은 상기 Rep 뉴클레오티드 서열 및 상기 Cap 뉴클레오티드 서열을 포함하는 제1 벡터, 및 상기 교정 게놈을 포함하는 제2 벡터를 포함한다. 본원에 기재된 바와 같은 패키징 시스템과 관련하여 사용된 "벡터"는 핵산을 세포 내로 도입하기 위한 운반체인 핵산 분자(예를 들어, 플라스미드, 바이러스, 코스미드, 인공 염색체 등)를 지칭한다.

[0142] 임의의 AAV Rep 단백질이 본원에 개시된 패키징 시스템에 사용될 수 있다. 패키징 시스템의 특정 실시형태에서, 상기 Rep 뉴클레오티드 서열은 AAV2 Rep 단백질을 인코딩한다. 적합한 AAV2 Rep 단백질은 제한없이 Rep 78/68 또는 Rep 68/52를 포함한다. 패키징 시스템의 특정 실시형태에서, AAV2 Rep 단백질을 인코딩하는 뉴클레오티드 서열은 서열 번호: 22의 AAV2 Rep 아미노산 서열과 최소 퍼센트 서열 동일성을 갖는 단백질을 인코딩하는 뉴클레오티드 서열을 포함하며, 상기 최소 퍼센트 서열 동일성은 상기 AAV2 Rep 단백질의 아미노산 서열의 길이에 걸쳐 적어도 70%(예를 들어, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 98%, 적어도 99%, 또는 100%)이다. 패키징 시스템의 특정 실시형태에서, 상기 AAV2 Rep 단백질은 서열 번호: 22에 기재된 아미노산 서열을 갖는다.

[0143] 패키징 시스템의 특정 실시형태에서, 상기 패키징 시스템은 제3 벡터, 예를 들어, 헬퍼 바이러스 벡터를 더 포함한다. 상기 제3 벡터는 상기 제1 벡터와 통합되거나, 상기 제2 벡터와 통합된 독립 제3 벡터일 수 있다. 특정 실시형태에서, 상기 제3 벡터는 헬퍼 바이러스 단백질을 인코딩하는 유전자를 포함한다.

[0144] 패키징 시스템의 특정 실시형태에서, 상기 헬퍼 바이러스는 아데노바이러스, 헤르페스 바이러스(헤르페스 단순 바이러스(HSV: herpes simplex virus) 포함), 폭스바이러스(예를 들어, 백시니아 바이러스), 거대세포바이러스(CMV: cytomegalovirus), 및 배콜로바이러스로 이루어진 군으로부터 선택된다. 상기 헬퍼 바이러스가 아데노바이러스인 패키징 시스템의 특정 실시형태에서, 아데노바이러스 게놈은 E1, E2, E4, 및 VA로 이루어진 군으로부터 선택된 하나 이상의 아데노바이러스 RNA 유전자를 포함한다. 상기 헬퍼 바이러스가 HSV인 패키징 시스템의 특정 실시형태에서, HSV 게놈은 UL5/8/52, ICPO, ICP4, ICP22, 및 UL30/UL42로 이루어진 군으로부터 선택된 HSV 유전자 중 하나 이상을 포함한다.

[0145] 패키징 시스템의 특정 실시형태에서, 상기 제1, 제2, 및/또는 제3 벡터는 하나 이상의 형질감염 플라스미드 내에 함유된다. 특정 실시형태에서, 상기 제1 벡터 및 상기 제3 벡터는 제1 형질감염 플라스미드 내에 함유된다. 특정 실시형태에서, 상기 제2 벡터 및 상기 제3 벡터는 제2 형질감염 플라스미드 내에 함유된다.

[0146] 패키징 시스템의 특정 실시형태에서, 상기 제1, 제2, 및/또는 제3 벡터는 하나 이상의 재조합 헬퍼 바이러스 내에 함유된다. 특정 실시형태에서, 상기 제1 벡터 및 상기 제3 벡터는 재조합 헬퍼 바이러스 내에 함유된다. 특정 실시형태에서, 상기 제2 벡터 및 상기 제3 벡터는 재조합 헬퍼 바이러스 내에 함유된다.

[0147] 또 다른 양태에서, 본 개시내용은 본원에 기재된 바와 같은 AAV의 재조합 제조방법으로서, 상기 방법은 세포에 캡시드 내에 교정 게놈을 봉입하기 위해 작용하는 조건 하에서 기재된 바와 같은 패키징 시스템으로 형질감염 또는 형질도입하는 단계를 포함하는, 방법을 제공한다. 예시적인 AAV의 재조합 제조 방법은 일시적 형질 감염(예를 들어, 본원에 기재된 바와 같이 제1, 및 제2, 및 선택적으로 제3 벡터를 함유하는 하나 이상의 형질감염 플라스미드를 사용함), 바이러스 감염(예를 들어, 본원에 기재된 바와 같이 제1, 및 제2, 및 선택적으로 제3 벡터를 함유하는, 아데노바이러스, 폭스바이러스(예를 들어, 백시니아 바이러스), 헤르페스 바이러스(HSV 포함), 거대세포바이러스, 또는 배콜로바이러스와 같은 하나 이상의 재조합 헬퍼 바이러스를 사용함), 및 안정한 생산자 세포주 형질감염 또는 감염(예를 들어, 본원에 기재된 바와 같이 하나 이상의 AAV Rep 단백질을 인코딩하는 Rep 뉴클레오티드 서열 및/또는 하나 이상의 AAV Clade F 캡시드 단백질을 인코딩하는 Cap 뉴클레오티드 서열을 함

유하는, 포유 동물 또는 곤충 세포와 같은 안정한 생산자 세포를 사용하며, 형질감염 플라스미드 또는 재조합 헬퍼 바이러스의 형태로 전달되는 본원에 기재된 바와 같이 교정 게놈을 사용함)을 포함한다.

## [0148] V. 실시예

[0149] 본원에 개시된 재조합 AAV 벡터는 뉴클레아제가 없는 상동성 의존 복구 기반 메커니즘을 통해 시험 관내 및 생체 내에서 매우 효율적인 유전자 편집을 매개한다. 하기 실시예는 본원에 개시된 바와 같은 AAV-기반 벡터를 사용하여 겹상세포질환(SCD: sickle cell disease) 및 베타 지중해빈혈증과 같은 특정 인간 질환에서 돌연변이된 HBB 유전자의 효율적인 교정을 입증한다. 이들 실시예는 제한이 아니라 예시로서 제공된다.

### [0150] 실시예 1 : HBB-돌연변이체 세포의 유전자 편집을 위한 AAV 캡시드의 선택

[0151] 본 실시예는 HBB 돌연변이체 세포에서 AAVHSC7, AAVHSC15, 및 AAVHSC17과 같은 clade F AAV 캡시드에 패키지된 유전자 편집 AAV 벡터, AAVS1-FP의 통합 효율을 특징으로 한다. 각각 AAVF7, AAVF15, 및 AAVF17로도 공지된 AAVHSC7, AAVHSC15, 및 AAVHSC17은 WO2016049230A1에 완전히 기재되어 있으며, 이는 그 전체 내용이 본원에 인용되어 포함된다.

[0152] 본원에 사용된 유전자 편집 벡터인 AAVS1-FP는 WO2016049230A1에 완전히 기재되어 있다. 이는 5' 내지 3': AAV2 5' 반전된 말단 반복(ITS: inverted terminal repeat), 표적 좌위로부터 상류 DNA의 서열을 갖는 800개의 뉴클레오티드로 구성된 5' 상동성 암, 업스트림, 스플라이스 수용체, 2A 요소, 형광 단백질(FP: fluorescent protein)의 코딩 서열, 표적 좌위로부터 하류 DNA의 서열을 갖는 800개의 뉴클레오티드로 구성된 3' 상동성 암, 및 AAV2 3' ITR을 포함하며, 상기 표적 좌위는 염색체 19 상의 AAVS1에서 인간 PPP1R12C의 인트론 1에 존재하며, AAVS1-FP 벡터와 인간 게놈 사이의 상동 재조합 후, PPP1R12C의 엑손 1, 2A 요소, 및 FP 코딩 서열이 프레임 내에 있다. 벡터에서 FP 코딩 서열은 프로모터가 없기 때문에, 이 벡터로 형질도입된 세포는 벡터가 게놈에 통합된 경우에만 FP를 발현할 것이다. AAV2 ITR 및 향상된 녹색 형광 단백질(EGFP: enhanced green fluorescent protein)에 작동 가능하게 연결된 프로모터를 포함하는 자가-상보성 AAV 벡터인 AAVHSC-scEGFP는 형질도입 효율의 제어로서 역할하였다(예를 들어, 그 전체 내용이 본원에 인용되어 포함된 미국 특허 제 8,628,966호 참조).

[0153] 겹상세포질환(SCD: sickle cell disease)의 특징인 HBB의 인트론 1에서 위치 20에서 A에서 T로 돌연변이를 갖는 림프 모세포 세포주(LCL: lymphoblastoid cell line)인 GM16265 세포를 코리엘 의료 연구소(Coriell Institute for Medical Research)(뉴저지 주 캠든 소재)로부터 얻었다. LCL을 15% FCS 및 2 mM L-글루타민이 보충된 RPMI에서 배양하였다. 세포를 m1 당 대략 200,000개의 세포에 시딩(seed)하고 m1 당 500,000 내지 1,000,000개의 세포에 도달할 때 분열시켰다. 세포를 도말하기 전에, 필요한 바이러스 양을 각각의 형질도입에 대해 계산하였다. 바이러스의 형질도입 부피는 웰의 총 부피의 10%를 초과하지 않았다. 형질도입 당일, 로그 상 세포를 계수하고 도말하였다.  $1.5 \times 10^5$ 의 감염 다중도(MOI: multiplicity of infection)에서 바이러스로 세포에 형질도입하였다. 패키지된 AAV 입자를 형질도입 전에 얼음 상에서 해동시키고 필요한 경우 얼음 상에서 초음파 처리하고, 각 웰에 개별적으로 첨가하였다. 형질도입 48 시간 후 세포를 수확하였다.

[0154] GM16265 세포는 AAVHSC7, AAVHSC15 또는 AAVHSC17 캡시드에 패키지된 AAVS1-FP 벡터로 형질도입되었다. 통합 효율은 하기 방법을 사용하여 유세포 분석에 의해 평가되었다: 세포를 FACS 완충액(IX PBS, 2% FCS, 0.1% 나트륨 아지드)을 사용하여 수확하고, 1200 RPM에서 10 분 동안 원심분리하였다. 여분의 상청액을 따라내어 대략 200  $\mu$ l가 남았다. 4',6-디아미디노-2-페닐리놀레(DAPI)를 유세포 분석 직전에 100  $\mu$ M 위킹 스톡(working stock)으로부터 3  $\mu$ M의 최종 농도로 첨가하였다.

[0155] 도 1a에 도시된 바와 같이, AAVHSC7-AAVS1-FP 및 AAVHSC17-AAVS1-FP에 의해 형질도입된 모든 활물 세포 중 FP-양성 세포의 백분율은 각각 24.3%(34.0% - 배경 레벨 9.7%) 및 7.8%(17.5% - 배경 레벨 9.7%)였다. 도 1b에 도시된 바와 같이, 다른 실험에서, AAVHSC15-AAVS1-FP 및 AAVHSC17-AAVS1-FP에 의해 형질도입된 모든 활물 세포 중 FP-양성 세포의 백분율은 각각 25.1%(29.8 % - 배경 레벨 4.7%) 및 37.6%(42.3% - 배경 레벨 4.7%)였다. 이 데이터는 GM16265 세포가 AAVHSC7, AAVHSC15 및 AAVHSC17 캡시드에 패키지된 AAVS1-FP에 의해 효율적으로 형질도입될 수 있음을 보여준다.

[0156] AAVHSC17 캡시드에 패키지된 이 AAVS1-FP 벡터의 통합 효율은 또한 1차 인간 CD34<sup>+</sup> 조혈 줄기 세포(HSC)에서 검사되었다. 1차 인간 CD34<sup>+</sup> HSC를 Miltenyi CD34 MicroBeads를 2회 농축함으로써 SCD를 갖는 공여체로부터의 인간 말초 혈액 세포로부터 정제하거나, 유사한 이중 농축 절차에 따라 ReachBio Inc.로부터 수득하였다. 세포를

20% 우태혈청(FCS: Fetal Calf Serum), 100  $\mu\text{g/mL}$  스트렙토마이신, 100 U/mL 페니실린, 2mmol/L L-글루타민, 10 ng/mL 인간 IL-3, 10 ng/mL 인간 IL-6, 및 1 ng/mL 인간 SCF으로 보충된 이스코브 변형된 둘베코 배지(IMDM: Iscove's Modified Dulbecco's Medium)에서 배양하였다. 약 200,000개의 세포를 500  $\mu\text{l}$  배지에 도말하였다. AAV 입자를 배지에 직접 첨가하였다. 형질도입 48 시간 후 세포를 수확하였다.

[0157]

BioRad QX200™ Droplet Digital™ PCR 시스템을 사용하는 디지털 액적 PCR에 의해 편집 효율을 측정하였다. 표 1에 나타낸 바와 같이, 두 세트의 프라이머 및 프로브를 ddPCR에 의한 통합을 정량화하도록 설계하였다. AAVS1\_Genomic 세트는 게놈의 AAVS1 좌위에 FP 코딩 서열의 표적화된 통합 후 비편집된 게놈 및 편집된 게놈에서 상동성 암 외부에 존재하는 AAVS1 좌위에서 서열을 검출하였다. AAVS1\_FP 세트는 게놈의 AAVS1 좌위에 FP 코딩 서열의 표적화된 통합 후 편집된 게놈에서만 존재하는 FP 코딩 영역에서 서열을 검출하였다. 2개의 프로브를 상이한 과장의 형광 모이어티에 접합시켰다.

표 1: ddPCR 분석용 프라이머

프라이머 명칭	서열 번호:	뉴클레오티드 서열
AAVS1_Genomic, 정방향 프라이머	90	GCGTTAGAGGGCAGAGTTC
AAVS1_Genomic, 역방향 프라이머	91	AGCTCCCATAGCTCAGTCT
AAVS1_Genomic, 프로브	92	CATTGTCACTTGCCTGCCCTC
AAVS1_FP 정방향 프라이머	93	GCAATAGCATCACAAATTTCAC
AAVS1_FP, 역방향 프라이머	94	GATCCAGACATGATAAGATAACATTG
AAVS1_FP, 프로브	95	TCACTGCATTCTAGTTGTGGTTGTCCA

[0158]

100 pg/ $\mu\text{l}$ 의 DNA를 갖는 샘플을 오일 액적으로 분할하였다. 대부분의 오일 액적은 DNA 분자를 함유하지 않거나 단일 DNA 분자-AAVS1\_Genomic 세트에만 양성인 비편집된 게놈; AAVS1\_FP 세트에만 양성인 비통합된 벡터; 또는 프라이머/프로브 세트 양쪽 모두에 양성인 편집된 게놈을 함유하였다. 공동-파티션의 가능성을 여러 표준 샘플에 의해 결정하였다(그 전체 내용이 본원에 인용되어 포함되어 있는, 문현[Regan et al, A rapid molecular approach for chromosomal phasing, PLoS One. (2015) 10(3):e0118270] 참조). 표준 샘플은  $\mu\text{l}$  당 100개의 비편집된 게놈,  $\mu\text{l}$  당 1000개의 에피솜 벡터, 및 각각  $\mu\text{l}$  당 1, 5, 10, 15, 20, 및 25개의 편집된 대립 유전자에서 클로닝된 양성 대립 유전자의 범위를 함유하였다. 비편집 대 편집된 대립 유전자의 비율에 대한 공동-분할의 표준 곡선을 플롯팅하였다( $R^2 = 0.972$ , 피어슨(Pearson) 상관  $p < 0.001$ ).

[0160]

각각의 샘플을 적어도 3회의 실험에서 ddPCR로 분석하였고, 각각의 샘플에서 AAVS1\_Genomic 양성, AAVS1\_FP 양성, 및 이중 양성 액적의 양을 측정하였으며 각각의 샘플에서 비편집 대 편집된 대립 유전자의 알려진 비율에 대해 플롯팅하였다. 도 1c에 도시된 바와 같이, 게놈으로의 FP 코딩 서열의 통합을 1차 인간 CD34<sup>+</sup> HSC로부터의 모든 대립 유전자의 약 30%에서 검출하였다. 따라서, AAVHSC17 캡시드에 패키지된 AAVS1-FP는 1차 인간 Cd34<sup>+</sup>를 효율적으로 형질도입하였다.

[0161]

## 실시예 2: HBB 돌연변이의 시험관 내 교정

[0162]

도 2에 도시된 바와 같이, hHBB-hL-014라는 명칭의 AAV-기반 HBB 교정 벡터를 생성하였다. 이 교정 벡터는 겸상 세포질환에서 HBB 유전자의 엑손 1의 코딩 영역(시작 코돈으로부터 시작)에서 뉴클레오티드 20에서 HBB 돌연변이, 예를 들어 A에서 T로의 돌연변이를 교정하도록 설계되었다. hHBB-hL-014 벡터는 HBB의 일부 및 그 주변 게놈 서열 측면의 5' 및 3' AAV2 ITR을 포함하였으며, 엑손 1에서 뉴클레오티드 20에서 A에서 T로의 돌연변이는 역전되었다. HBB 게놈 서열의 일부를 NCBI Primer Blast([www.ncbi.nlm.nih.gov/tools/primer-blast/](http://www.ncbi.nlm.nih.gov/tools/primer-blast/))를 사용하여 설계된 표 2에 나타낸 바와 같이 증폭 프라이머를 사용하여 야생형 HBB 및 그의 주변 좌위로부터 수득 하였다. PCR 산물은 HBB의 모든 엑손 및 인트론을 커버하였고, HBB 전사 개시 부위로부터 상류에 1678개의 뉴클레오티드 및 HBB 폴리아데닐화 서열로부터 하류에 234개의 뉴클레오티드를 더 포함하였다. 겸상세포질환에서 HBB 돌연변이와 관련하여, 이 벡터는 각각 약 1.7 kb 길이의 상동성 암(돌연변이 역전 부위에 대한 게놈 서열 5' 및 3')을 함유하였다. ITR의 무결성(integrity)을 *Bgl*I, *Msc*I, 및 *Sma*I를 사용한 제한 다이제스트 스크리닝 및

ITR 특이적 시퀀싱 프로토콜을 사용한 시퀀싱에 의해 확인하였다(문헌[Mroske et al., Hum Gene Ther Methods (2012) 23 (2) : 128-36]). 인서트(insert)는 본원의 표 3에 나타낸 프라이머를 사용한 제한 다이제스트 및 생거(Sanger) 시퀀싱에 의해 확인되었다. 이 벡터는 HBB 엑손 및 인트론에서의 돌연변이뿐만 아니라 베타 지중해 빈혈증에서 관찰된 바와 같이 HBB 발현에 영향을 미치는 5' 및 3' 비번역된 영역에서의 돌연변이를 교정할 수 있었다.

표 2: 인간 베타 글로빈 게놈 영역 증폭 프라이머

프라이머 명칭	서열 번호:	뉴클레오티드 서열
HBB2M Gib 5' fwd	50	AGGGGTGGAGTCGTGACGTGCCAAATCAAGCCTCTACTTGAATCC
HBB2M Gib 3' rev	51	AATGATTAACCCGCCATGCTACTTATCTACGTAAA CCTAGGCTCCAGATAGCCA

[0163]

표 3: 인간 베타 글로빈 시퀀싱 프라이머

프라이머 명칭	서열 번호:	뉴클레오티드 서열
Fwd3 seq HBB2	52	GGAAGCAGAACTCTGCAC
Fwd4 seq HBB2	53	GCATTAAGAGGTCTCTAGTTTTTATC
Fwd5 seq HBB2	54	GATGGTATGGGGCCAAGAGATATATC
Fwd6 seq HBB2	55	GTCTACCCTTGGACCCAGAG
Fwd7 seq HBB2	56	CAGTCTGCCTAGTACATTACTATTG
Fwd8 seq HBB2	57	CATGTTCATACCTCTTATCTTCC
Fwd9 seq HBB2	58	GCAAACAGCTAATGCACATTGG
Rev3 seq HBB2	59	CAGAACATCCAGATGCTCAAGGCC
Rev4 seq HBB2	60	CCCTGATTGGTCAATATGT
Rev5 seq HBB2	61	CATCAAGCGTCCCATAAGACTCAC
Rev6 seq HBB2	62	GCAGACTTCTCCTCAGGAGTC
Rev7 seq HBB2	63	CTTACAGGACAGAATGGATGAAAAC
Rev8 seq HBB2	64	GAAAAGGTCTTCTACTTGGCTC
Rev9 seq HBB2	65	GGTTAACCAAAAGAAACTGG

[0164]

교정된 유전자의 검출을 용이하게 하기 위해, *Cla*I 제한 부위 및 *Spe*I 제한 부위를 함유하는 ACTAGTATCGAT(서열 번호: 80)의 서열을 갖는 12-bp 링커를 HBB 유전자에 삽입하였다. 이 링커 서열은 시작 코돈으로부터 117 bp 및 돌연변이 역전 부위로부터 97 bp인 인트론 1에 위치하여, 링커(Linker)와 원하는 유전자 교정 사이에 강력한 유전적 연결을 확립하였다. 인트론 1에서 주요 공여체 및 수용체 부위의 파괴를 교정된 HBB의 mRNA 스플라이싱을 유지하기 위해 회피하였다.

[0166]

hHBB-hL-014 벡터를 그 전체 내용이 본원에 인용되어 포함되어 있는 문헌[Chatterjee et al., (1993) Methods 5:51-59]에 기재된 패키징 방법을 사용하여 AAVHSC15 또는 AAVHSC17 캡시드 단백질로 패키징하였다. 패키지된 바이러스의 역가를 표 4에 나타낸 프라이머 및 프로브를 사용하여 qPCR에 의해 결정하였다.

표 4: HBB 교정 벡터 qPCR 프라이머 및 프로브

프라이머 명칭	서열 번호:	뉴클레오티드 서열
qHBB2M 정방향 프라이머	66	TGCAGATTAGTCCAGGCAGAAA
qHBB2M 역방향 프라이머	67	GGGTAATCAGTGGTGTCAAATAGGA
qHBB2M 프로브	68	AGTTAGATGTCCCCAGTTAA

[0167]

AAVHSC15-hHBB-hL-014 및 AAVHSC17-hHBB-hL-014 바이러스를 표적된 통합(TI) 분석법을 사용하여 GM16265 세포에서 HBB 유전자를 편집하는 능력에 대해 테스트하였다. 이 분석에서, 세포를 4000 RPM에서 10분 동안 원심분리

하고, IX PBS로 세척하였고, 펠릿을 후속 사용을 위해 80°C에서 동결시켰다. 냉동된 세포 펠릿을 100,000개 세포에 대해 200 μl에서 재현탁시켰다. 무 DNase(DNase-free) RNase 1 μl를 첨가하고 37°C에서 1 시간 동안 배양하였다. 10% SDS 10 μl 및 프로테나제 K 1.2 μl를 첨가하고 56°C에서 하룻밤동안 배양하였다. 고 분자량 DNA를 표준 페놀 및 클로로포름 추출에 의해 추출하였다. 높은 DNA 수율을 위해, 0.5X 부피의 Tris-EDTA 완충액 (TE: Tris-EDTA buffer) (pH 8.0)으로 역 추출을 수행하고 최종 튜브에 첨가하였다. DNA를 2.5 M의 최종 농도에서 10 M 암모늄 아세테이트로 침전시켰다. 대략 4X 부피의 빙냉 100% 에탄올을 첨가하였다. DNA를 -80°C에서 적어도 1 시간 동안 침전시켰다. DNA를 70% 에탄올로 세척하고, 건조시키고, 대략 30 내지 50 μl의 TE에 재현탁시키고, 나노드롭(Nanodrop)에 의해 정량화하였다. 정량화 후, hHBB-hL-014 벡터와 HBB 유전자 사이의 정확한 재조합을 확인하기 위해 DNA를 적절한 프라이머를 사용하여 PCR 기반의 "표적된 통합"(TI) 분석을 거쳤다.

[0169] 표 5에 기재된 서열을 갖는 프라이머를 본 실시예에서 TI 검정에 사용하였다. HBB2MTI100 프라이머는 링커 및 그 주변 영역을 표적하였고, HBB350 프라이머는 상동성 암 외부의 계놈 서열을 표적하였다. PCR 반응은 편집된 계놈으로부터 단리된 DNA로부터 2,219 bp 앰플리콘을 생성하지만 형질도입되지 않은 세포 또는 hHBB-hL-014 벡터 단독으로부터 단리된 DNA를 실질적으로 증폭하지 않을 것이다.

[0170] PCR 반응을 다음과 같이 설정하였다: 최대 50 μl의 PCR 물; 10 μl의 5X Q5 완충액; 5 μl의 베타인; 1 μl의 10 mM dNTP; 1 μl의 HBB2MTI100 정방향 프라이머(25 μM 농도); 1 μl의 HBB 350 역방향 프라이머(25 μM 농도); 100 ng 내지 1 μg의 계놈 DNA; 1 μl의 NEB Q5 고온실도 폴리머라제. PCR 기계를 다음과 같이 설정하였다: 95°C에서 5 분 동안 초기 변성; 95°C에서 10 초 동안 변성 15 사이클, 70°C에서 30 초 동안 어닐링, 사이클마다 0.5 도 감소, 및 72°C에서 2 분 동안의 연장; 95°C에서 10 초 동안 변성 20 사이클; 65°C에서 30 초 동안 어닐링, 및 72°C에서 2 분 동안 연장; 및 72°C에서 5 분 동안의 최종 연장. PCR 산물을 겔 전기 영동에 의해 분석하였다.

[0171] 2.2 kb의 겉보기 크기를 갖는 앰플리콘을 단리하고 무딘 말단(blunt-end)을 pUC18 백본(backbone)에 결찰시켰다. 생성된 플라스미드를 *Cla*I 또는 *Spe*I 엔도뉴클레아제를 사용하여 제한 소화에 의해 분석하여 링커의 정확한 삽입으로 클론을 동정하였다. 제한 소화에 의해 동정된 양성 클론을 M13F 및 M13R 올리고뉴클레오티드 프라이머를 사용한 DNA 시퀀싱에 의해 추가로 분석하였다.

**표 5: HBB 표적된 통합 검정용 프라이머**

프라이머 명칭	서열 번호:	뉴클레오티드 서열
HBB2MTI100 정방향 프라이머 (TM 54.1°C)	69	CTATTGGTCTCCTTAAAATCGATACTAGT
HBB350 역방향 프라이머 (TM 54.8°C)	70	ATATTCAAACTTCCGCAGAACACT

[0172] [0173] 도 3a에 도시된 바와 같이, AAVHSC15-hHBB-hL-014 및 AAVHSC17-hHBB-hL-014 바이러스로 형질도입된 GM16265 세포에서 2.2 kb PCR 밴드(hHBB-hL-014 벡터에 의한 정확한 편집을 나타냄)가 TI 검정에서 검출된 반면에, 이 PCR 산물은 형질도입되지 않은 GM16265 세포에서 검출되지 않았다.

[0174] GM16265 세포에서 HBB 유전자의 위치 20에서 A에서 T 돌연변이의 교정은 TI 검정에서 동정된 양성 클론의 시퀀싱에 의해 검증되었다. 서열 분석은 AAVHSC15-hHBB-hL-014 또는 AAVHSC17-hHBB-hL-014 바이러스로 형질도입한 후, T 돌연변이가 A로 교정되었고, 근처 무증상 돌연변이도 야생형으로 교정되었다. 반대쪽 PCR 가닥의 시퀀싱은 돌연변이의 교정을 확인하였다. HBB 좌위에 링커의 삽입은 모든 클론에서 검출되었다. 또한, 시험된 클론 중 어느 것도 상동성 암의 말단에 상응하는 계놈 영역에서 바람직하지 않은 돌연변이(예를 들어, 추가 삽입, 결실 또는 역전)를 나타내지 않았다.

[0175] HBB 유전자를 편집하는 hHBB-hL-014 벡터의 능력을 더 확인하기 위해, 2개의 추가 LCL인 GM16266 및 GM16267을 사용하였다. 이들 LCL은 또한 코리엘 의료 연구소(Coriell Institute for Medical Research) (뉴저지 주 캠든 소재)로부터 입수하였고, GM16265와 다른 중여체로부터 수집되었다. 두 LCL 양쪽 모두 HBB의 인트론 1에서 위치 20에서 A에서 T로의 돌연변이를 가졌다. GM16265의 배양 및 형질도입을 위한 동일한 방법에 따라, 세포는 AAVHSC17 캡시드에 패키지된 hHBB-hL-014로 형질도입되었다.

[0176] 도 3b에 도시된 바와 같이, 약 2.2 kb의 PCR 앰플리콘이 AAVHSC17-hHBB-hL-014로 형질도입된 세포로부터는 검출

되었지만, 형질도입되지 않은 세포로부터는 검출되지 않았다. 시퀀싱 결과는 SCD 돌연변이가 3 개의 모든 LCL(GM16265 세포 포함)로부터 형질도입된 세포에서 교정되었고, 바람직하지 않은 돌연변이(예를 들어, 추가 삽입, 결실, 또는 역전)는 상동성 암의 말단에 상응하는 게놈 영역에서 검출되지 않았음을 보여주었다. 시퀀싱 결과는 또한 상동성 암으로부터 게놈으로의 적절한 링커 삽입 및 끊김없는(seamless) 전이를 확인하였다.

[0177] 상기 결과는 hHBB-hL-014 벡터가 돌연변이체 HBB 유전자를 다수의 SCD 세포주에서 야생형 서열로 되돌릴 수 있음을 보여준다. 따라서, 베타 지중해빈혈증과 같은 다른 유전 질환에서 HBB 유전자의 엑손, 인트론 또는 조절 서열에서의 돌연변이는 또한 hHBB-hL-014 벡터를 사용하여 교정할 수 있어야 한다.

#### 실시예 3: HBB의 게놈 서열을 포함하는 HBB 교정 벡터

[0179] 본 실시예는 HBB 유전자에서 돌연변이를 교정할 수 있는 HBB의 게놈 서열을 포함하는 AAV-기반 HBB 교정 벡터를 제공한다.

##### a) HBB 교정 벡터 hHBB-hL-001

[0181] 도 4a에 도시된 바와 같이, HBB 교정 벡터 hHBB-hL-001은 모든 엑손, 모든 인트론 및 폴리아데닐화 서열을 포함하는 HBB 게놈 서열을 함유한다. 이 벡터는 HBB 전사 개시 부위(도 4a에서 "HBB HAL"로 지칭됨)로부터 상류에서 800 bp를 포함하는 5' 영역, 및 HBB 폴리아데닐화 서열(도 4a에서 "HBB HAR")로부터 하류에서 800 bp를 포함하는 3' 영역을 추가로 함유한다. hHBB-hL-001 벡터는 서열 번호: 31(TI RE 링커 포함) 또는 서열 번호: 32 (TI RE 링커 제외)에 기재된 뉴클레오티드 서열을 포함하고, 5' ITR(예를 들어, 서열 번호: 18의 서열을 가짐) 및 3' ITR(예를 들어, 서열 번호: 19의 서열을 가짐)을 더 포함한다. 이 벡터는 HBB 엑손 및 인트론에서의 돌연변이뿐만 아니라 베타 지중해빈혈증에서 관찰된 바와 같이 HBB 발현에 영향을 미치는 5' 및 3' 비번역된 영역에서의 돌연변이를 교정한다.

##### b) HBB 교정 벡터 hHBB-h1W-013

[0183] 도 4b에 도시된 바와 같이, HBB 교정 벡터 hHBB-h1W-013은 엑손 1, 2 및 3의 코딩 영역의 DNA 서열이 상응하는 야생형 서열에 완전히 동일하기보다는 약 67% 동일하게 침묵 변경되는 것을 제외하고는 HBB 교정 벡터 hHBB-hL-001에서와 동일한 유전적 요소를 함유한다. 감소된 서열 동일성은 코돈 변경에 기인하며, 여기서 변성 코돈은 인코딩된 아미노산을 변화시키지 않고 오리지널(original) 코돈으로 치환된다. 이 침묵 코돈 변경은 HBB의 발현 수준을 크게 변화시킬 것으로 예상되지 않는다. 대신, 엠실론 글로빈(HBE), 엘타 글로빈(HBD), 감마 글로빈 1(HBG1), 감마 글로빈 2(HBG2), 및 HBB 유사 유전자 HBBP와 같은 다른 글로빈 유전자 또는 유사 유전자의 HBB 엑손의 상동성을 감소시켜 다른 게놈 좌위에서 이 벡터의 바람직하지 않은 재조합 가능성을 감소시킨다. 특정 실시예에서, 엑손 1, 엑손 2, 및 엑손 3의 코딩 영역의 침묵 변경된 서열은 각각 서열 번호: 43, 44, 및 45에 기재되어 있다. hHBB-h1W-013 벡터는 서열 번호: 33(TI RE 링커 포함) 또는 서열 번호: 34(TI RE 링커 제외)에 기재된 뉴클레오티드 서열을 포함하고, 5' ITR(예를 들어, 서열 번호: 18의 서열을 가짐) 및 3' ITR(예를 들어, 서열 번호: 19의 서열을 가짐)을 더 포함한다.

##### c) HBB 교정 벡터 hHBB-hL-011

[0185] 도 4c에 도시된 바와 같이, HBB 교정 벡터 hHBB-hL-011은 HBB 폴리아데닐화 서열(도 4c에서 "HBB HAR"로 지칭됨)로부터 하류의 3' 영역이 길이가 약 100 bp인 것을 제외하고는 HBB 교정 벡터 hHBB-hL-001에서와 동일한 유전적 요소를 함유한다. 이러한 변형은 프로모터 서열이 전사 인자 및 보조 인자를 모집할 수 있기 때문에 HBB 폴리아데닐화 서열로부터 약 100 bp 하류에 위치한 다른 유전자(예를 들어, GATA 1, MYC 등)의 전사 프로모터 서열의 포함을 최소화하도록 하여, 이에 의해 상동 재조합의 효율을 감소시킨다. 추가적으로, 전사 프로모터 서열의 포함은 벡터로부터 HBB의 비정상적인 발현을 증가시킬 수 있다. hHBB-hL-011 벡터는 서열 번호: 35(TI RE 링커 포함) 또는 서열 번호: 36(TI RE 링커 제외)에 기재된 뉴클레오티드 서열을 포함하고, 5' ITR(예를 들어, 서열 번호: 18의 서열을 가짐) 및 3' ITR(예를 들어, 서열 번호: 19의 서열을 가짐)을 더 포함한다.

##### d) HBB 교정 벡터 hHBB-h1W-012

[0187] 도 4d에 도시된 바와 같이, HBB 교정 벡터 hHBB-h1W-012는 엑손 1, 2 및 3의 코딩 영역의 DNA 서열이 상응하는 야생형 서열에 67% 동일하게 침묵 변경되는 것을 제외하고는 HBB 교정 벡터 hHBB-hL-011에서와 동일한 유전적 요소를 함유한다. 특정 실시예에서, 엑손 1, 엑손 2, 및 엑손 3의 코딩 영역의 침묵 변경된 서열은 각각 서열 번호: 43, 44, 및 45에 기재되어 있다. hHBB-h1W-012 벡터는 서열 번호: 37(TI RE 링커 포함) 또는 38(TI RE 링커 제외)에 기재된 뉴클레오티드 서열을 포함하고, 5' ITR(예를 들어, 서열 번호: 18의 서열을 가짐) 및 3'

ITR(예를 들어, 서열 번호: 19의 서열을 가짐)을 더 포함한다.

[0188]

본 실시예에서 4개의 HBB 교정 벡터 각각은 교정 유전자의 검출을 용이하게 하기 위해 독특한 제한 엔도뉴클레아제를 위한 인식 및 절단 부위를 포함하는 링커 서열을 함유한다. 이 링커 서열은 시작 코돈으로부터 117 bp인 인트론 1에 위치한다. 인트론 1에서 주요 공여체 및 수용체 부위의 파괴를 교정된 HBB의 mRNA 스플라이싱을 유지하기 위해 회피하였다.

[0189]

상기 기재된 4개의 HBB 교정 벡터 각각은 AAVHSC17에서 생성되고 패키지되었다. 1차 인간 CD34<sup>+</sup> HSC를 실시예 1에 기재된 방법을 사용하여 바이러스로 형질도입되고, 실시예 2에 기재된 TI 검정에 의해 통합을 평가하였다. 도 5a 및 5b에 도시된 바와 같이, hHBB-hL-001, hHBB-hL-011, 및 hHBB-h1W-012는 모두 HBB 유전자를 편집할 수 있었다.

[0190]

계놈 편집의 효율은 차세대 시퀀싱(NGS: next generation sequencing)에 의해 정량적으로 측정되었다. 표 6에 나타낸 바와 같이, 상동성 암 외부의 계놈의 영역에 특이적인 프라이머를 사용하여 PCR 반응을 수행하였다. 이들 프라이머는 비편집 및 편집된 대립 유전자로부터 2,342 bp의 산물을 증폭시킬 것이지만, AAV 벡터를 증폭시키지 않을 것이다.

**표 6: NGS 샘플 제조용 프라이머**

프라이머 명칭	서열 번호:	뉴클레오티드 서열
HBB350 역방향 프라이머	70	ATATTCAAACCTCCGCAGAACACT
HBB L NGS S1	81	CCTCTGGGTCCAAGGGTAGA

[0191]

PCR 반응을 다음과 같이 설정하였다: 최대 50 μl의 PCR 물; 10 μl의 5X Q5 완충액; 5 μl의 베타인; 1 μl의 10 mM dNTP; 1 μl의 HBB350 역방향 프라이머(25 μM 농도); 1 μl의 HBB L NGS S1 프라이머(25 μM 농도); 200 ng의 계놈 DNA; 및 1 μl의 Q5 고총실도 폴리머라제. PCR 기계를 다음과 같이 설정하였다: 98°C에서 30 분 동안 초기 변성; 95°C에서 10 초 동안 변성 30 사이클, 65°C에서 30 초 동안 어닐링, 및 72°C에서 2 분 동안의 연장; 및 72°C에서 5 분 동안의 최종 연장.

[0192]

정확한 크기의 PCR 산물을 겔 전기영동에 의해 단리하고, 표준 프로토콜에 따라 Qiagen Qiaquick Gel 추출 키트를 사용하여 추출하였다. 겔 추출된 엠플리콘에서 벡터 계놈의 부재를 벡터 특이적 프라이머 및 공지된 수의 벡터 특이적 계놈 주형을 양성 대조군으로 사용하여 PCR에 의해 확인하였다. 벡터 계놈의 부재를 확인하기 위해, 하기 PCR 조건을 이용하였다: 98°C에서 30 분 동안 초기 변성; 98°C에서 10 초 동안 변성 30 사이클, 66°C에서 30 초 동안 어닐링, 및 72°C에서 1 분 동안의 연장; 및 72°C에서 2 분 동안의 최종 연장. 사용된 정방향 프라이머는 AAAGTCAGGGCAGAGCCATC(서열 번호: 108)이고, 사용된 역방향 프라이머는 AATGATTAACCGCCATGCT(서열 번호: 109)였으며, 1,797개의 염기쌍의 엠플리콘을 증폭시킬 것이다.

[0193]

추출된 PCR 산물을 NGS 시퀀싱 및/또는 하기에 기재된 바와 같은 디지털 PCR 정량화에 사용하였다. NGS 시퀀싱의 경우, 추출된 PCR 산물을 표 7에 나타낸 프라이머를 사용하여 중첩된 PCR 라운드에 적용하였다. 각각의 샘플은 순방향 및 역방향 프라이머의 독특한 조합을 가졌으며, 각각의 정확한 PCR 산물의 크기는 약 388 bp였다. PCR 기계를 다음과 같이 설정하였다: 98°C에서 30 분 동안 초기 변성; 98°C에서 10 초 동안 변성 30 사이클, 72°C에서 30 초 동안 어닐링, 및 72°C에서 30 초 동안의 연장; 및 72°C에서 2 분 동안의 최종 연장.

표 7: NGS 샘플 제조용 프라이머

프라이머 명칭	서열 번호:	뉴클레오티드 서열
HBB NGS Nest 5 정방향 어댑터 1	82	AATGATAACGGCGACCACCGAGATCTACACAAGTAGA GTCTTCCCTACACGACGCTCTCCGATCTGGGCATA AAAGTCAGGGCAGA
HBB NGS Nest 5 정방향 어댑터 2	83	AATGATAACGGCGACCACCGAGATCTACACCATGCTTA TCTTCCCTACACGACGCTCTCCGATCTGGGCATAA AAGTCAGGGCAGA
HBB NGS Nest 5 정방향 어댑터 3	84	AATGATAACGGCGACCACCGAGATCTACACGCACATCT TCTTCCCTACACGACGCTCTCCGATCTGGGCATAA AAAGTCAGGGCAGA
HBB NGS Nest 5 정방향 어댑터 4	85	AATGATAACGGCGACCACCGAGATCTACACTGCTCGAC TCTTCCCTACACGACGCTCTCCGATCTGGGCATAA AAAGTCAGGGCAGA
HBB NGS Nest 5 역방향 어댑터 1	86	CAAGCAGAACGGCATACGAGATCATGATCGGTGA CTGGAGTTCAAGACGTGTGCTCTCCGATCTGTCTCCA CATGCCAGTTCTA
HBB NGS Nest 5 역방향 어댑터 2	87	CAAGCAGAACGGCATACGAGATAGGATCTAGTGA CTGGAGTTCAAGACGTGTGCTCTCCGATCTGTCTCCA CATGCCAGTTCTA
HBB NGS Nest 5 역방향 어댑터 3	88	CAAGCAGAACGGCATACGAGATGACAGTAAGTGA CTGGAGTTCAAGACGTGTGCTCTCCGATCTGTCTCCA ACATGCCAGTTCTA
HBB NGS Nest 5 역방향 어댑터 4	89	CAAGCAGAACGGCATACGAGATCCTATGCCGTGA CTGGAGTTCAAGACGTGTGCTCTCCGATCTGA GTCTCC ACATGCCAGTTCTA

[0195]

[0196] 앰플리콘 크기를 겔 전기영동에 의해 확인하고, PCR 산물을 표준 프로토콜에 따라 Qiagen Qiaquick PCR 정제 키트를 사용하여 정제하였다. 모든 시험된 샘플로부터의 앰플리콘을 동일한 몰 농도로 혼합하고, 농도를 Advanced Analytical 생물분석기(bioanalyzer)로 확인하였다. 샘플을 MiSeq V2 300 사이클 키트를 사용하여 시퀀싱하였다.

[0197]

표 8은 NGS 분석으로부터 결정된 바와 같이, 링커 서열의 존재 또는 부재 하에 수용 가능한 판독 수 뿐만 아니라 HBB 서열을 갖는 판독의 수를 나타낸다. 링커 서열의 존재는 AAV 벡터 통합을 나타내기 때문에, 링커를 갖는 대립 유전자의 백분율은 이들 벡터의 통합의 대립 유전자 빈도를 나타내며, 이는 일반적으로 0.1% 내지 1%이다.

표 8: HBB 교정 벡터의 통합의 대립 유전자 빈도

샘플	링커 존재 수	링커 부재 수	링커 인접 서열 수	링커 존재의 %	캡처된 판독의 %*
HBB-hL-014 #1	4	2703	2710	0.14	99.88
HBB-hL-011 #1	3101	475533	479142	0.64	99.89
HBB-hL-011 #2	5	1917	1925	0.26	99.84
HBB-hL-001 #1	3	449	455	0.66	99.34
HBB-hL-011 #3	2	235	235	0.84	100.85
음성 대조군	0	1	1	0	1
형질도입되지 않은	0	0	0	0	0

\*캡처된 판독의 %는 링커 서열에 인접한 서열을 포함하는 품질 통과 판독의 백분율을 링커 서열을 포함하는 판독 및 링커 서열을 포함하지 않은 판독에 대한 합계로 나눈 값을 나타내는 품질 관리 지표이다.

[0198]

[0199] 디지털 PCR의 경우, 추출된 PCR 산물에 BioRad QX200™ Droplet Digital™ PCR 시스템을 사용하여 디지털 PCR 분석을 수행하였다. 게놈 표적과 삽입된 벡터 폐이로드(링커) 사이의 연결을 계산하여 편집을 결정하고, 변이체 간의 유전적 연결에 사용되는 방법을 이용하는 우연의 일치(coincidence by chance)의 가능성과 관련하여 벡터 및 게놈 양쪽 모두를 함유하는 분할된 액적의 양을 검출함으로써 측정하였다(예를 들어, 그 전체 내용이 본원에 인용되어 포함되어 있는 문헌[Regan et al., A rapid molecular approach for chromosomal phasing, PLoS One.

(2015) 10(3):e0118270] 참조). 0.1 ng/ul의 게놈 DNA의 농도를 연결용 샘플 당 최소 3회의 실험에 걸쳐 분석하고, 벡터 특이적 프로브 세트 및 게놈 특이적 프로브 세트를 갖는 다중화된 ddPCR을 사용하여 측정하였다. 프라이머 및 프로브 세트는 다음과 같다:

**표 9: 디지털 PCR에 사용되는 프라이머 및 프로브 세트**

		서열 번호: 뉴클레오티드 서열
<u>링커 특이적 프라이머 및 프로브 세트</u>		
HBB 프로브	110	AACTGGGCATGTGGAGACAGAGAA
HBB 링커 F	111	GTTACAAGACAGGACTAGTATCGAT
HBB R	112	TAGACCAATAGGCAGAGAGAGT
<u>게노머(Genomer) 특이적 프라이머 및 프로브 세트</u>		
HBB0101GDNA F	113	CTGAGCCAAGTAGAAGACCTTT
HBB0101GDNA R	114	CTGTTTCTGCCTGGACTAATCT
HBB0101GDNA 프로브	115	CCCTACTTTCTAAGTCACAGAGGCT

[0200] 알려진 양의 편집된 물질에 대해 상기 방법에 따라 게놈 편집/연결을 측정하기 위해, 표준 DNA 시리즈가 생성되었다. 표준은 ul 당 100개의 비편집된 게놈, ul 당 1000개의 에피솜 벡터, 및 각각 ul 당 1, ul 당 5, ul 당 10, ul 당 15, ul 당 20, 및 ul 당 25개의 편집된 대립 유전자에서 클로닝된 양성 대립 유전자의 범위로 구성된다. 각각의 샘플에서 유전적 연결의 양을 측정하고 각각의 샘플에서 비편집 대 편집된 대립 유전자의 알려진 비율에 대해 플롯팅하였다( $R^2 = 0.972$ , 피어슨 상관  $p < 0.001$ ).

[0202]  $1.5 \times 10^5$ 의 MOI에서 다양한 AAVHSC7 HBB 편집 벡터로 형질도입된 1차 CD34<sup>+</sup> 혼합 제대혈 세포에서 HBB 좌위의 편집을 디지털 PCR에 의해 측정하였다. 형질도입 후 48 시간에 세포를 수확하고, PCR 산물의 아웃/아웃 PCR 및 디지털 PCR 분석에 의해 편집된 대립 유전자의 백분율에 대해 분석하였다. 도 6은 표시된대로 샘플에서 편집된 좌위의 분율을 도시한다.

#### 실시예 4: HBB 코딩 서열 또는 이의 일부를 포함하는 HBB 교정 벡터

[0203] 본 실시예는 HBB 코딩 서열 또는 이의 일부를 HBB 유전자 내로 예를 들어, 시작 코돈 이후 또는 인트론 1 내로 삽입할 수 있는 HBB 교정 벡터를 제공한다. 삽입된 서열은 천연 전사 조절 요소의 제어 하에 천연 좌위로부터 전사 및 번역될 수 있고, 이에 의해 기능적 HBB 단백질의 발현을 회복시킨다.

[0204] HBB 교정 벡터의 각각은 제2 코돈에서 정지 코돈까지 HBB 코딩 서열 또는 이의 일부를 함유하였다. HBB 코딩 서열 또는 이의 일부는 SV40 폴리아데닐화 서열이 뒤따르는데, 이는 적절한 발현을 지지하고 나머지 내인성 HBB 유전자의 추가 전사를 상당히 감소시키기에 충분히 강하다. 독특한 제한 엔도뉴클레아제에 대한 인식 및 절단 부위를 포함하는 표적화된 통합 제한 카세트("TI RE 카세트")는 폴리아데닐화 서열의 하류에 선택적으로 삽입되어 원하는 상동 재조합의 검출을 용이하게 한다.

##### a) HBB 교정 벡터 hHBB-hA-009

[0205] 도 7a에 도시된 바와 같이, HBB 교정 벡터 hHBB-hA-009는 제2 코돈에서 정지 코돈까지의 야생형 HBB 코딩 서열의 일부(서열 번호: 27의 뉴클레오티드 4 내지 444), 이후 상기 기재된 바와 같이 SV40 폴리아데닐화 서열을 포함한다. 벡터는 HBB 시작 코돈으로부터 상류이고 이를 포함하는 야생형 게놈 서열을 포함하는 5' 상동성 암(도 7a에서 "HBB HAL"로 지칭됨), 및 HBB 시작 코돈으로부터 하류이고 이를 포함하지 않는 3' 상동성 암(도 7a에서 "HBB HAR"로 지칭됨)을 더 포함한다. 상기 5' 상동성 암 및 상기 3' 상동성 암은 길이가 약 800 bp이다. hHBB-hA-009 벡터는 서열 번호: 39에 기재된 뉴클레오티드 서열을 포함하고, 5' ITR(예를 들어, 서열 번호: 18의 서열을 가짐) 및 3' ITR(예를 들어, 서열 번호: 19의 서열을 가짐)을 더 포함한다. 5' 상동성 암은 일부 베타 지중해성빈혈 환자에서 관찰된 바와 같이 HBB 발현에 영향을 주는 시작 코돈 및/또는 5' 비번역된 영역(UTR)에서 돌연변이를 교정하는 능력을 갖는다. 그 결과, HBB 교정 벡터 hHBB-hA-009의 통합은 5' UTR, 코딩 서열 또는 3' UTR에서 돌연변이에 의해 손상된 야생형 HBB의 발현을 회복시킬 수 있다.

##### b) HBB 교정 벡터 hHBB-hAW-002

[0206] 도 7b에 도시된 바와 같이, HBB 교정 벡터 hHBB-hAW-002는 HBB 코딩 서열의 부분이 야생형 cDNA 서열의 상응하는 영역에 서열 번호: 47, 67% 동일하게 침묵 변경되는 것을 제외하고는 HBB 교정 벡터 hHBB-hA-009에서와 동일

한 유전적 요소를 함유한다. 실시예 3에서 설명된 바와 같이, 이 코돈 변경은 HBB의 발현 수준을 상당히 변경시킬 것으로 예상되지는 않는다. 대신, 다른 글로빈 유전자 및 유사 유전자와 HBB 엑손의 상동성을 감소시켜 다른 계놈 좌위에서 이 벡터의 바람직하지 않은 재조합을 감소시킨다. hHBB-hAW-002 벡터는 서열 번호: 40에 기재된 뉴클레오티드 서열을 포함하고, 5' ITR(예를 들어, 서열 번호: 18의 서열을 가짐) 및 3' ITR(예를 들어, 서열 번호: 19의 서열)을 더 포함한다.

[0210] c) HBB 교정 벡터 hHBB-h1-010

도 7c에 도시된 바와 같이, HBB 교정 벡터 hHBB-h1-010은 상동 재조합에 의해 인트론 1에 야생형 HBB 코딩 서열을 삽입하도록 설계된다. 구체적으로, 삽입 부위는 HBB 유전자의 뉴클레오티드 160 및 161 사이에 있으며, 삽입은 인트론 1에서 주요 스플라이스 공여체 부위의 파괴를 회피한다. 이 벡터의 편집 요소(삽입된 영역)는 5' 내지 3' 스플라이스 수용체 부위(도 7c에서 "SA", 예를 들어, 서열 번호: 14), 통합시 HBB 시작 코돈과 프레임 내에 있는 리보솜 스키핑 요소(도 7c에서 "T2A", 예를 들어, 서열 번호: 72), 야생형 HBB 코딩 서열(서열 번호: 27), 및 SV40 폴리아데닐화 서열을 함유한다. 통합시, HBB 좌위로부터 전사된 전-mRNA는 5' 내지 3': 내인성 HBB의 엑손 1; 인트론 1의 5' 말단에 내인성 스플라이스 공여체를 포함하여 인트론 1의 제1 68개 뉴클레오티드; hHBB-h1-010 벡터에 의해 도입된 스플라이스 수용체; 리보솜 스키핑 요소; HBB 코딩 서열; 및 poly(A) 테일을 함유한다. 스플라이싱 후, mRNA는 5' 내지 3': 내인성 HBB의 엑손 1; 인-프레임 리보솜 스키핑 요소; HBB 코딩 서열; 및 poly(A) 테일을 함유한다. 리보솜 스키핑 요소는 2개의 폴리펩티드: 부분 리보솜 스키핑 웨이드와 융합된 엑손 1의 말단에서 종결된 절두된 HBB 웨이드 및 전장 HBB의 N- 말단에 융합된 리보솜 건너 뛰기 웨이드의 프롤린 폴리펩티드를 생성하게 된다.

[0212] hHBB-h1-010 벡터는 삽입 부위로부터 상류에 야생형 계놈 서열을 포함하는 5' 상동성 암(도 7c에서 "HBB HAL"로 지칭됨) 및 삽입 부위로부터 하류에 야생형 계놈 서열을 포함하는 3' 상동성 암(도 7c에서 "HBB HAR"로 지칭됨)을 더 포함한다. 상기 5' 상동성 암 및 상기 3' 상동성 암은 길이가 약 800 bp이다. hHBB-h1-010 벡터는 서열 번호: 41에 기재된 뉴클레오티드 서열을 포함하며, 5' ITR(예를 들어, 서열 번호: 18를 가짐) 및 3' ITR(예를 들어, 서열 번호: 19를 가짐)을 더 포함한다. 5' 상동성 암은 일부 베타 지중해성빈혈 환자에서 관찰된 바와 같이 HBB 발현에 영향을 주는 5' UTR에서 돌연변이를 교정하며, 5' UTR, 코딩 서열, 또는 3' UTR에서 돌연변이에 의해 손상된 야생형 HBB의 발현을 회복하는 능력을 갖는다.

[0213] d) HBB 교정 벡터 hHBB-h1W-008

도 7d에 도시된 바와 같이, HBB 교정 벡터 hHBB-h1W-008은 HBB 코딩 서열이 야생형 cDNA 서열의 상응하는 영역과 67% 동일하게 침묵 변경되는 것을 제외하고, HBB 교정 벡터 hHBB-h1-010에서와 동일한 유전적 요소를 함유한다. 실시예 3에 기재된 바와 같이, 이 서열 변형은 HBB의 발현 수준을 상당히 변화시킬 것으로 예상되지는 않는다. 대신에, 다른 글로빈 유전자 및 유사 유전자와 HBB 엑손의 상동성을 감소시켜, 다른 계놈 좌위에서 이 벡터의 바람직하지 않은 재조합을 감소시킨다. 특정 실시예에서, 침묵 변경된 HBB 코딩 서열은 서열 번호: 47에 기재되어 있다. hHBB-h1W-008 벡터는 서열 번호: 42에 기재된 뉴클레오티드 서열을 포함하며, 5' ITR(예를 들어, 서열 번호: 18을 가짐) 및 3' ITR(예를 들어, 서열 번호: 19를 가짐)을 더 포함한다.

[0215] e) HBB 교정 벡터 hHBB-hE3C-001

도 7e에 도시된 바와 같이, HBB 교정 벡터 hHBB-hE3C-001은 상동 재조합에 의해 정지 코돈 직후에 HBB 유전자의 엑손 3에 HBB 코딩 서열을 삽입하도록 설계된다. 이 벡터의 편집 요소(삽입된 영역)는 침묵 변경된 HBB 코딩 서열(서열 번호: 99, 야생형 HBB 코딩 서열에 85% 동일)과 프레임 내에 있는 5' 내지 3' 리보솜 스키핑 요소(도 7e에서 "P2A", 예를 들어 서열 번호 74), 및 SV40 폴리아데닐화 서열(서열 번호: 77)을 함유한다. HBB 코딩 서열의 침묵 변경은 코딩 서열로부터 발현된 단백질의 수준을 증가시키고, 바람직하지 않은 계놈 좌위로의 벡터의 오프-타겟팅으로 초래할 수 있는 저-복잡성(low-complexity) 서열을 제거하고/하거나 편집 요소 및 계놈 사이의 상동성을 감소시켜, 상동성 암보다는 편집 요소에 의해 매개되는 바람직하지 않은 통합을 감소시키도록 설계된다.

[0217] 통합시, HBB 좌위로부터 전사된 mRNA는 5' 내지 3': 정지 코돈에 인접하여 5'의 천연 HBB mRNA의 일부, 리보솜 스키핑 요소, 침묵 변경된 HBB 코딩 서열, 및 SV40 폴리아데닐화 서열을 함유한다. 리보솜 스키핑 요소는 2개의 폴리펩티드: 리보솜 스키핑 웨이드의 N-말단 부분과 융합된 천연 전장 HBB 웨이드, 및 전장 야생형 HBB 폴리펩티드의 N-말단에 융합된 리보솜 스키핑 웨이드로부터의 프롤린 잔기를 생성시킨다.

[0218] hHBB-hE3C-001 벡터는 서열 번호: 101을 포함하는 5' 상동성 암(도 7e에서 "HAL"로 지칭됨), 서열 부위로부터 상

류의 야생형 게놈 서열, 및 삽입 부위로부터 하류의 야생형 게놈서열인 서열 번호: 102의 서열을 포함하는 3' 상동성 암(도 7e에서 "HAR"로 지칭됨)을 더 포함한다.

[0219] hHBB-hE3C-001의 뉴클레오티드 서열은 서열 번호: 104에 기재되어 있다. 상기 벡터는 5' ITR(예를 들어, 서열 번호: 18) 및 3' ITR(예를 들어, 서열 번호: 19의 서열을 가짐)을 더 포함한다.

[0220] hHBB-h1-010 및 hHBB-h1W-008의 통합 효율을 RKO 및 LCL 세포에서 평가하였다. GM16265 LCL 세포를 실시예 1에 기재된 방법을 사용하여 AAVHSC7에 패키지된 HBB 교정 벡터로 형질도입하였다.

[0221] RKO 세포를 ATCC로부터 수득하였다. 세포를 10% FCS 및 2 mM L-글루타민이 보충된 DMEM에서 배양하였다. 세포를 6 웨 플레이트에서 웨당 750,000개의 세포 밀도로 도말하였다. 세포를 하기 방법을 사용하여 형질감염시켰다: 도말 24 시간 후, (a) 250  $\mu$ l OptiMEM에 희석된 2  $\mu$ g의 HBB 편집 플라스미드 및 (b) 250  $\mu$ l OptiMEM에 희석된 5  $\mu$ l Lipofectamine 2000을 15 분 동안 혼합 및 배양함으로써 세포를 OptiMEM 배지에서 형질감염시켰다. 형질감염 24 시간 후 세포를 수확하였다.

[0222] 표 9에 기재된 서열을 갖는 프라이머를 사용하여 TI 검정에 의해 통합을 평가하였다. SA-2A-FM1 및 SA-2A-FM2 프라이머는 hHBB-h1-010 또는 hHBB-h1W-008 벡터에 의해 게놈에 삽입된 스플라이스 수용체 및 T2A 요소에 특이적이었으며, HBB-Out-RM2 프라이머는 3' 상동성 암으로부터 하류 게놈의 영역에 특이적이었다. SA-2A-FM1 또는 SA-2A-FM2과 HBB-Out-RM2의 프라이머 쌍은 형질도입되지 않은 세포에서의 생성물 또는 교정 벡터 단독으로부터의 생성물을 증폭시키지 않는다. SA-2A-FM1 및 HBB-Out-RM2를 사용한 PCR 반응은 hHBB-h1-010 또는 hHBB-h1W-008 벡터가 5' 및 3' 상동성 암을 통한 상동 재조합에 의해 통합되면 1,881 bp 앰플리콘을 생성할 것이며, 상기 벡터가 편집 요소에서 5' 상동성 암 및 엑손 2 서열을 통한 상동 재조합에 의해 통합되면 1,188 bp 앰플리콘을 생성할 것이다. SA-2A-FM2 및 HBB-Out-RM2를 사용한 PCR 반응은 또한 이러한 두 가지 통합 방식에서 다른 크기의 앰플리콘을 생성할 것이다.

[0223] PCR 반응을 다음과 같이 설정하였다: 최대 50  $\mu$ l의 PCR 률; 5  $\mu$ l의 10X PCR 완충액; 5  $\mu$ l의 베타인; 1  $\mu$ l의 10 mM dNTP; 1  $\mu$ l의 50 mM MgCl<sub>2</sub>; 10  $\mu$ l의 5X Q 시약; 2.5  $\mu$ l의 TI 정방향 프라이머(5  $\mu$ M 농도); 2.5  $\mu$ l의 TI 역방향 프라이머(5  $\mu$ M 농도); 100 ng 게놈 DNA; 및 0.5  $\mu$ l의 HotStarTaq 폴리머라제. PCR 기계를 다음과 같이 설정하였다: 95°C에서 15 분 동안 초기 변성; 94°C에서 10 초 동안 변성 40 사이클, 58°C에서 30 초 동안 어닐링, 및 72°C에서 3 분 동안의 연장; 및 68°C에서 10 분 동안의 최종 연장. PCR 산물을 겔 전기영동에 의해 분석하였다.

**표 9: HBB 표적된 통합 검정용 프라이머**

프라이머 명칭	서열 번호:	뉴클레오티드 서열
SA-2A-FM1	96	GCTTCTGACCTCTCTCTCCCTCCC
SA-2A-FM2	97	GCGGTGACGTGGAGGAGAACATC
HBB-Out-RM2	98	GCAGAATGGTAGCTGGATTGTAGC

[0224]

[0225] 도 8에 도시된 바와 같이, SA-2A-FM1 및 HBB-Out-RM2 프라이머를 사용하여, 1,874 뉴클레오티드 길이를 갖는 표적화된 통합 PCR 산물이 hHBB-h1W-002로 형질도입된 RKO 세포에서 검출되어 벡터의 성공적인 통합을 원하는 방식으로 나타내었다. 이와 대조적으로, HBB 좌위와 편집 요소(3' 상동성 암에서가 아닌)에서 엑손 2의 재조합에 의해 생성된 단축된 PCR 산물을 hHBB-h1-010으로 형질도입된 RKO 및 LCL 세포에서 검출하였다. SA-2A-FM2 및 HBB-Out-RM2 프라이머를 사용하여 PCR로부터 유사한 결과를 얻었다. 이 결과는 편집 요소에서의 침묵 코돈 변경이 바람직하지 않은 재조합을 감소시키거나 제거함으로써 HBB 유전자의 정확한 편집을 보장한다는 것을 나타낸다.

#### 실시예 5: HBB 돌연변이의 생체 내 교정

[0226]

[0227] 본 실시예는 이전 실시예에서 기재된 것과 같은 HBB 교정 벡터를 검사하기 위한 동물 모델을 제공한다. NOD.Cg-Prkdc<sup>scid</sup> I12rg<sup>tmtWj1</sup>/SzJ의 유전자형을 갖는 NSG 마우스에 치사량에 가깝게 조사하고 1차 인간 야생형 CD34<sup>+</sup> HSC를 이식하여 조혈 재구성을 실시하였다. NSG 마우스에서 인간 CD34<sup>+</sup> HSC의 생착(engraftment) 수준을 유세포 분석에 의해 인간 및 뮤린 CD45<sup>+</sup> 세포의 존재에 대해 말초 혈액을 분석함으로써 이식 12 주 후에 결정하였다. 말초 혈액에서 25% 초과 순환 인간 세포를 갖는 마우스를 사용하여 생체 내에서 1차 인간 CD34<sup>+</sup> HSC를 형질도입시

키기 위한 특정 AAV 벡터의 통합 효율을 평가하였다.

[0228] AAVS1-FP 벡터를 AAVHSC7 및 AAVHSC17에 패키지하고, 바이러스 입자를 kg 당  $1.22 \times 10^{13}$  내지  $1.54 \times 10^{13}$  벡터 계놈의 용량으로 재구성된 NSG 마우스에 정맥 내 투여하였다. 혈액, 골수 및 비장 샘플을 투여 4 주 후 수집하였다. 당업계에 공지된 폐놀/클로로포름 추출법에 의해 샘플로부터 DNA를 정제하고, 추출된 DNA를 실시예 1에 기재된 바와 같은 방법을 사용하여 ddPCR에 의해 분석하였다.

[0229] AAVHSC7 및 AAVHSC17 그룹으로부터의 데이터를 모았다. 도 9에 도시된 바와 같이, AAVS1-FP 벡터가 투여된 이식된 NSG 마우스에서, 통합의 대립 유전자 빈도는 혈액에서 약 3%, 골수에서 약 1%였다. 이 결과는 AAVHSC7 및 AAVHSC17 캡시드가 AAVS1 좌위에 통합하기 위한 벡터를 효율적으로 전달하고, HBB 편집 치료 벡터를 전달하는데 잠재적으로 사용될 수 있음을 시사하였다.

[0230] HBB 결합 1차 인간 CD34<sup>+</sup> HSC로 재구성된 변형된 동물 모델은 교정 벡터 또는 상기 실시예에 기재된 바와 같은 교정 벡터를 사용하여 HBB 돌연변이의 교정을 테스트하는 데 유용하다. 예를 들어, AAVHSC7, AAVHSC15 또는 AAVHSC17 캡시드와 같은 AAV clade F 캡시드에 패키지된 교정 벡터는 재구성된 동물에게 투여될 수 있다. 통합 효율은 혈액 또는 골수 샘플을 수집하고, 원하는 상동 재조합이 광범위한 집단 또는 적혈구의 전구세포와 같은 특정 유형의 세포에서 발생한 세포의 백분율을 정량함으로써 측정될 수 있다.

[0231] HBB 결합 1차 인간 CD34<sup>+</sup> HSC로 재구성된 동물은 HBB 유전자의 부족으로 인해 혈색소병증이 나타날 것으로 예상되며, 다양한 AAV 캡시드에 패키지된 HBB 교정 벡터의 효능 및 안전성을 결정하는 데 사용될 수 있다. 효능은 망상 적혈구 수, 완전 혈구 수(CBC), 혈액 도말 및 벡터 서열의 표적된 통합을 측정하여 평가된다. 안전성은 아스파테이트 트랜스아미나제(AST) 및 알라닌 트랜스아미나제(ALT)와 같은 간 트랜스아미나제 수준을 측정함으로써 평가된다.

[0232] 이 모델은 또한 각각의 투여 후 HBB 교정의 수명을 평가하여 투여 요법을 최적화하는 데 사용될 수 있다.

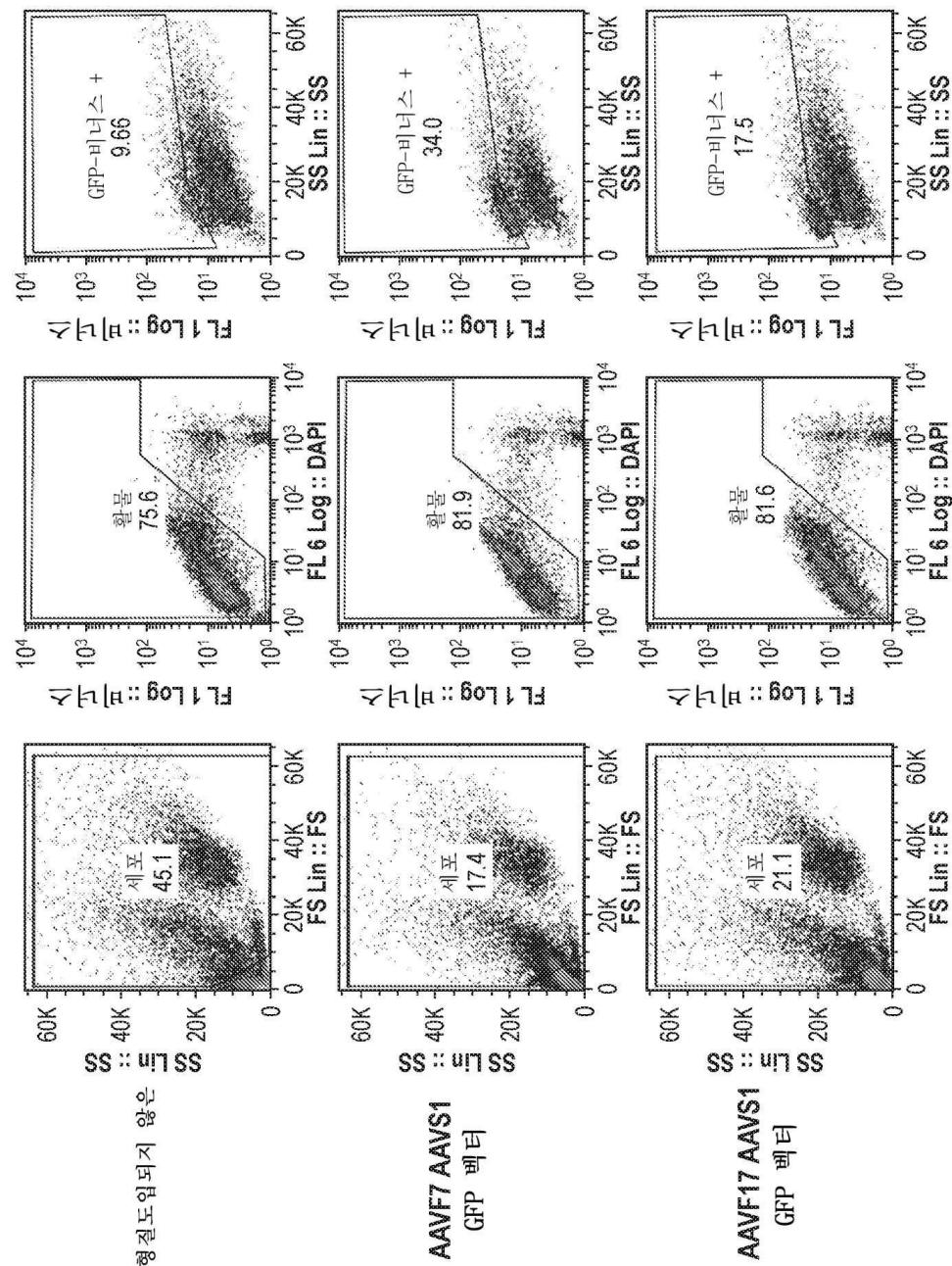
[0233] \* \* \*

[0234] 본 발명은 본원에 기재된 특정 실시형태에 의해 범위가 제한되지 않아야 한다. 실제로, 설명된 것들에 추가하여 본 발명의 다양한 변형이 전술한 설명 및 첨부 도면으로부터 당업자에게 명백해질 것이다. 이러한 변형은 첨부된 청구 범위의 범주 내에 속한다.

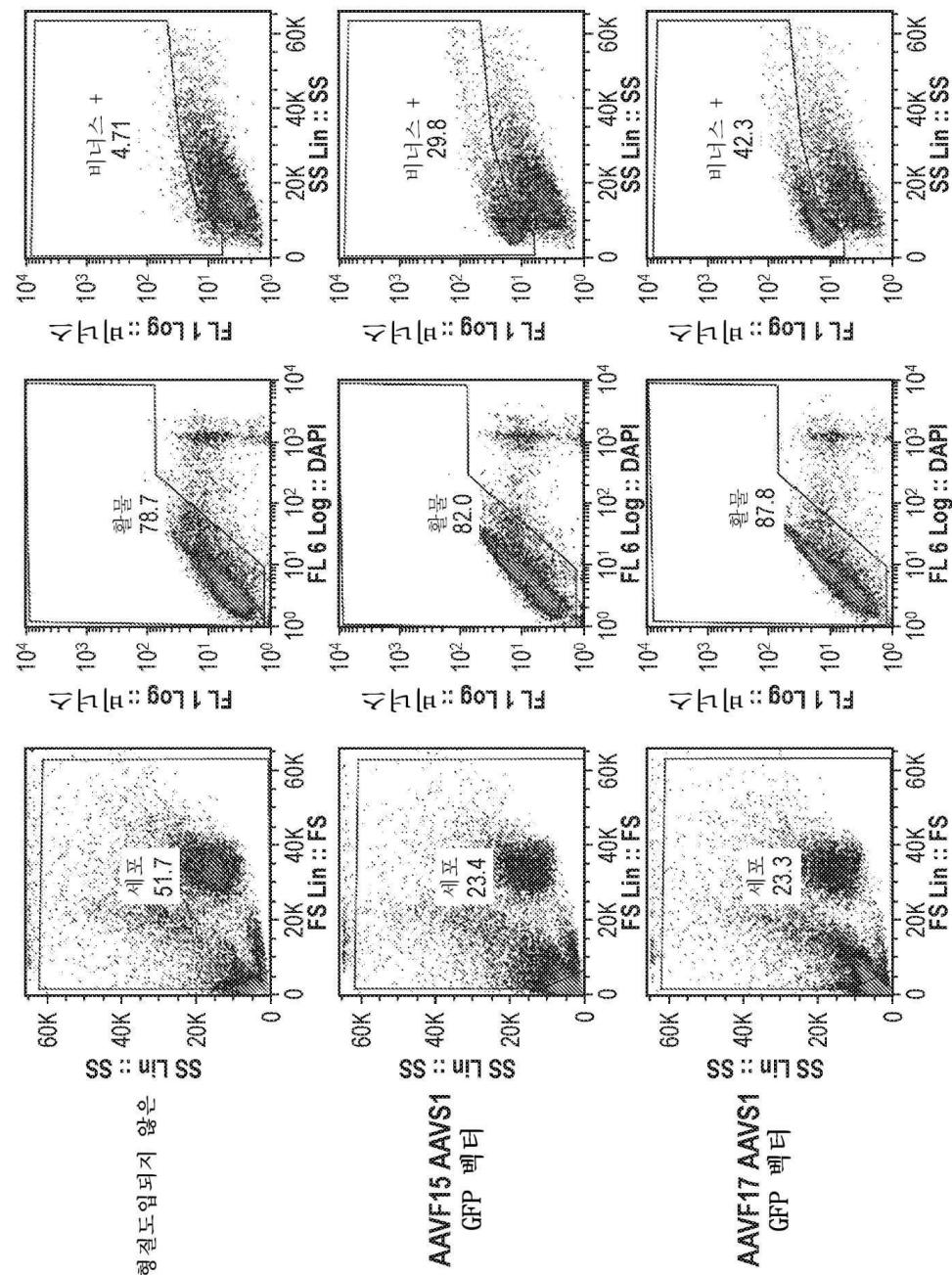
[0235] 본원에 인용된 모든 참고 문헌(예를 들어, 공개 또는 특허 또는 특허 출원)은 각각의 개별 참조(예를 들어, 공개 또는 특허 또는 특허 출원)가 모든 목적을 위해 그 전체 내용이 본원에 인용되어 포함되도록 구체적이고 개별적으로 지시된 것과 동일한 정도로 그 전체 내용이 본원에 인용되어 그리고 모든 목적을 위해 포함된다. 다른 실시형태는 하기 청구범위 내에 있다.

## 도면

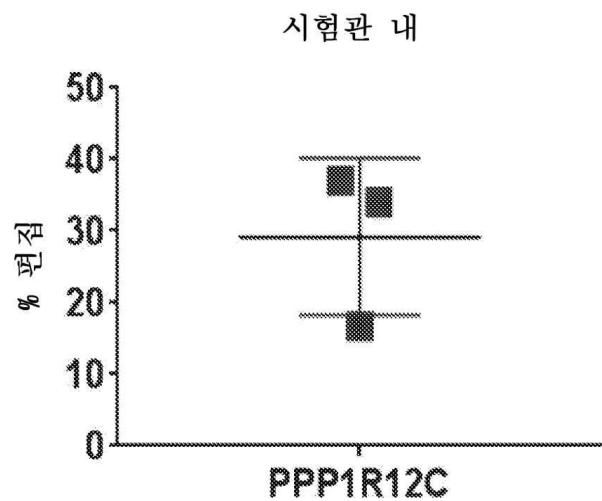
## 도면 1a



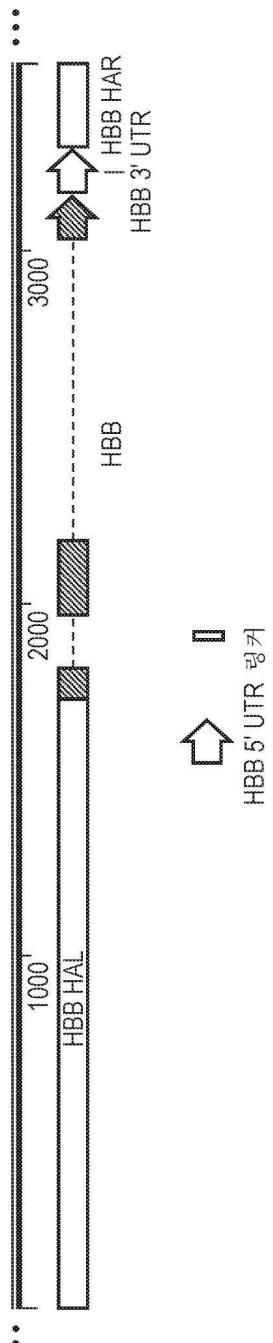
## 도면1b



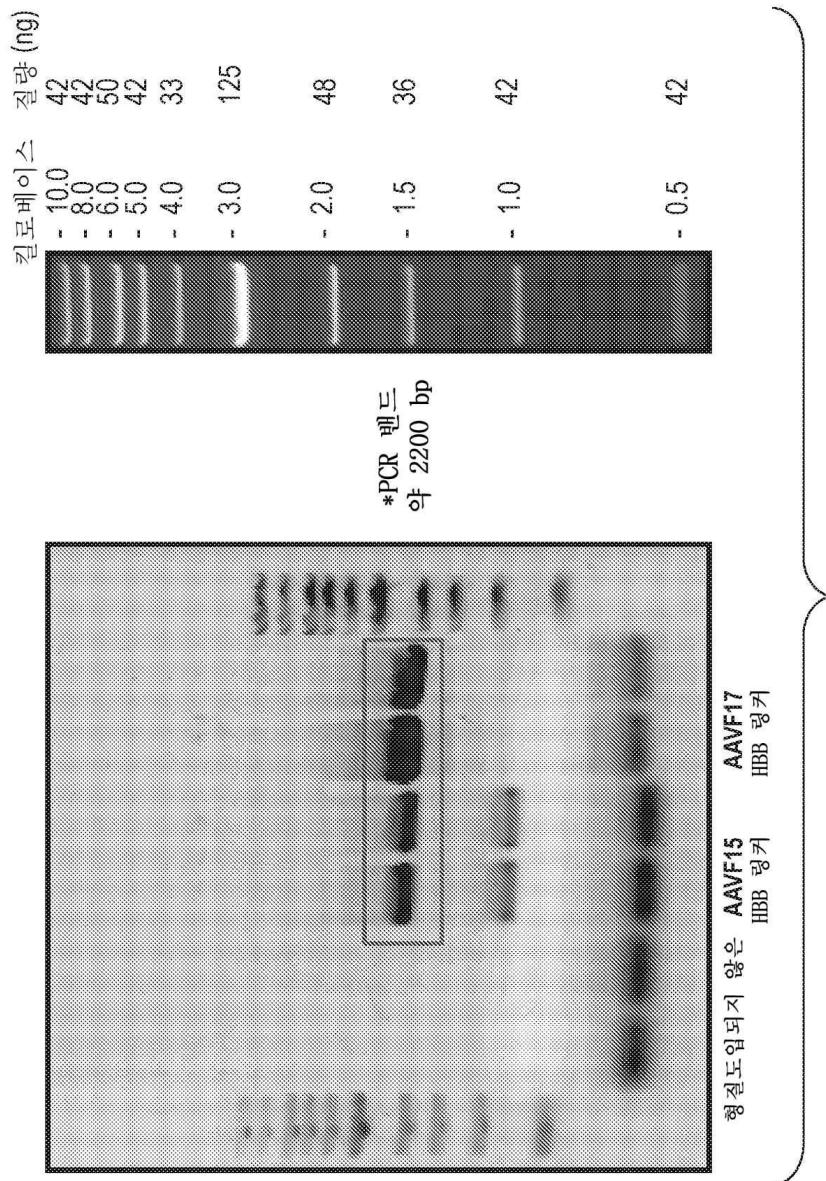
도면 1c



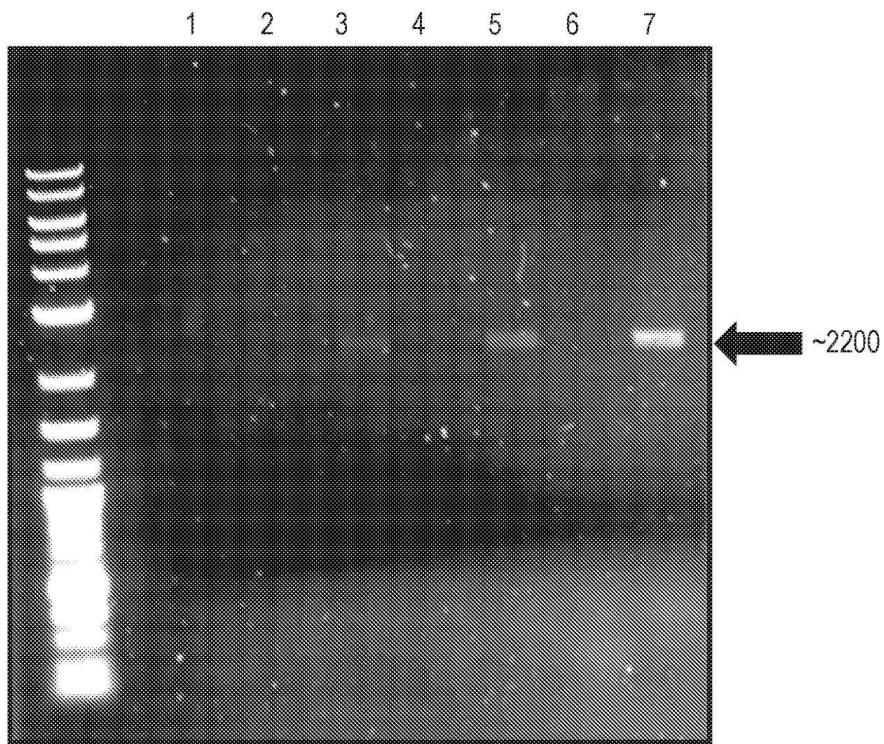
도면2



도면3a



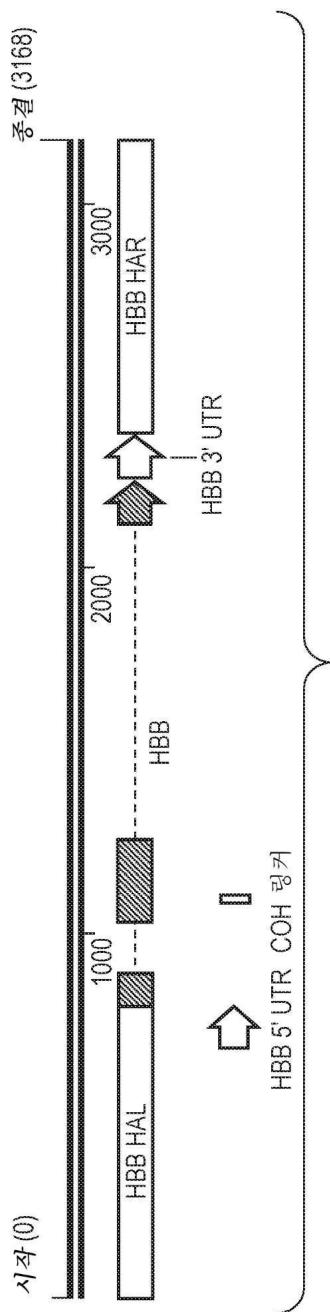
도면3b



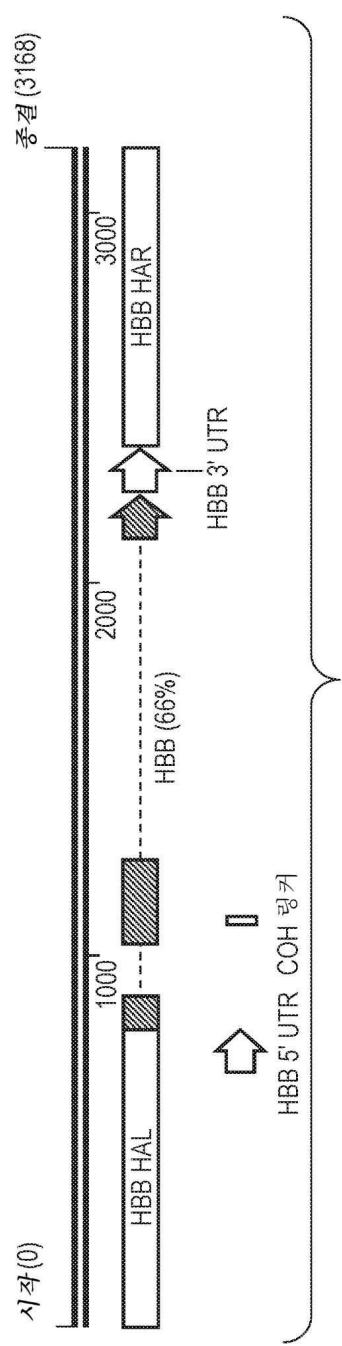
레인:

1. 음성 대조군
2. 형질도입되지 않은 16265 SCD LCL
3. AAVHSC17-hHBB-ht-014 벡터로 형질도입된 16265 SCD LCL
4. 형질도입되지 않은 16266 SCD LCL
5. AAVHSC17-hHBB-hL-014 벡터로 형질도입된 16266 SCD LCL
6. 형질도입되지 않은 16267 SCD LCL
7. AAVHSC17-hHBB-hL-014 벡터로 형질도입된 16267 SCD LCL

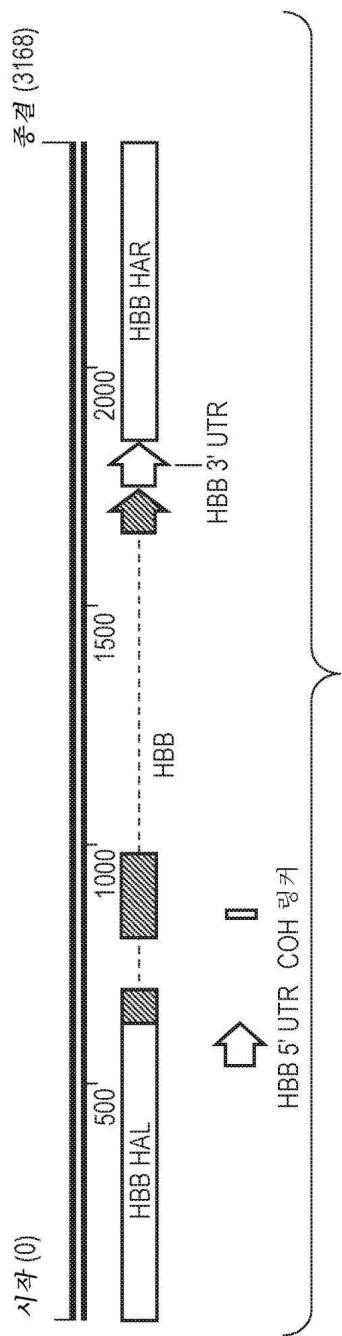
도면4a



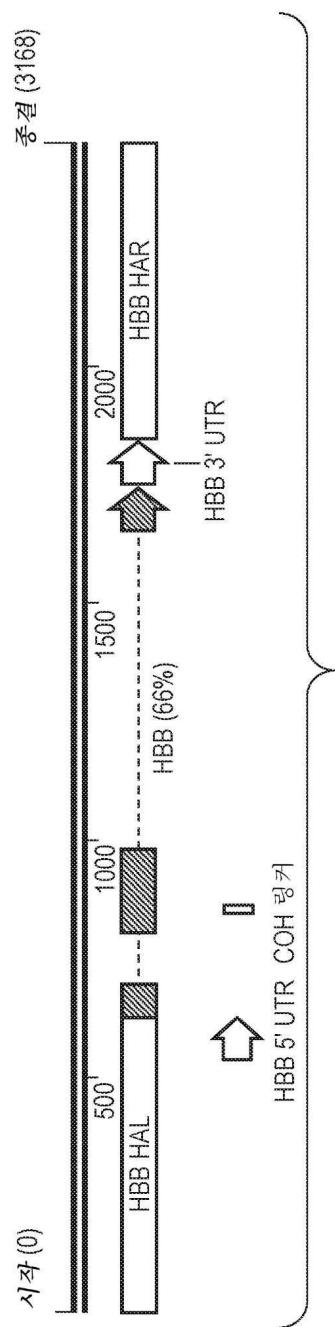
도면4b



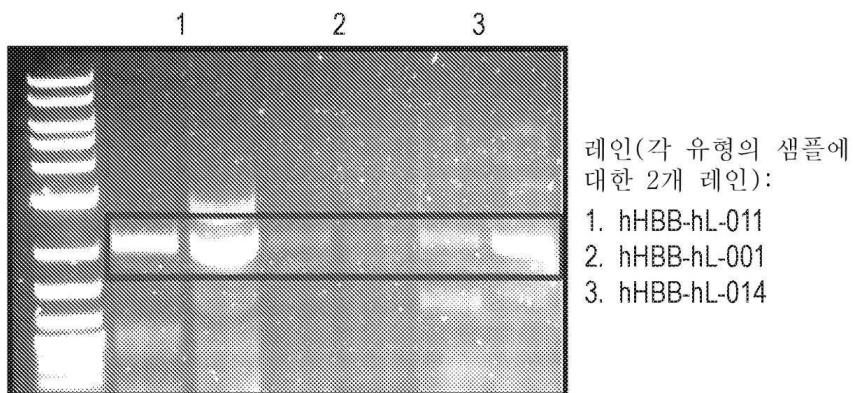
## 도면4c



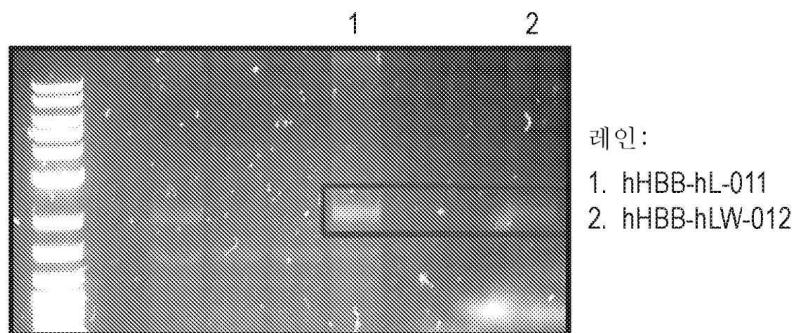
## 도면 4d



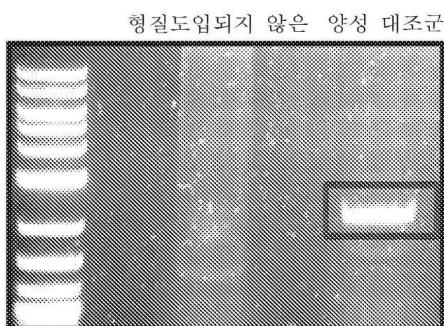
도면5a



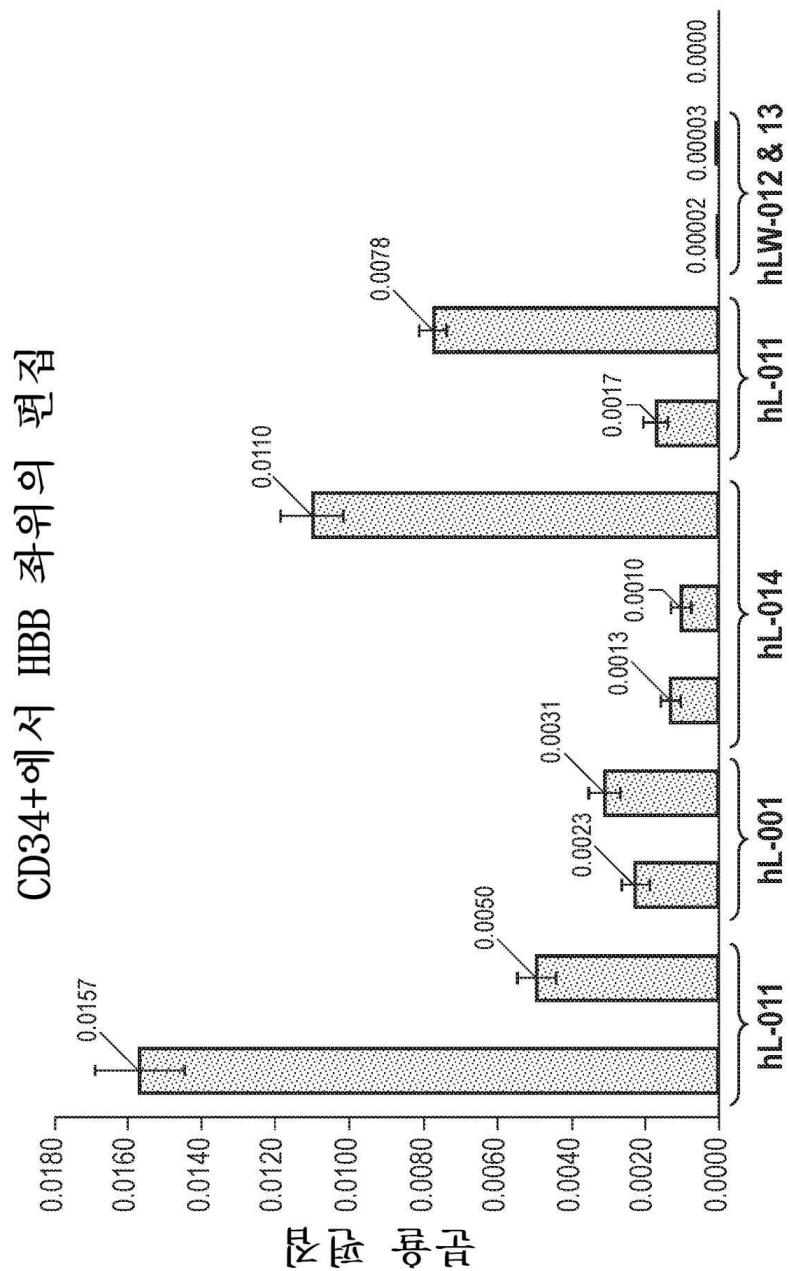
도면5b



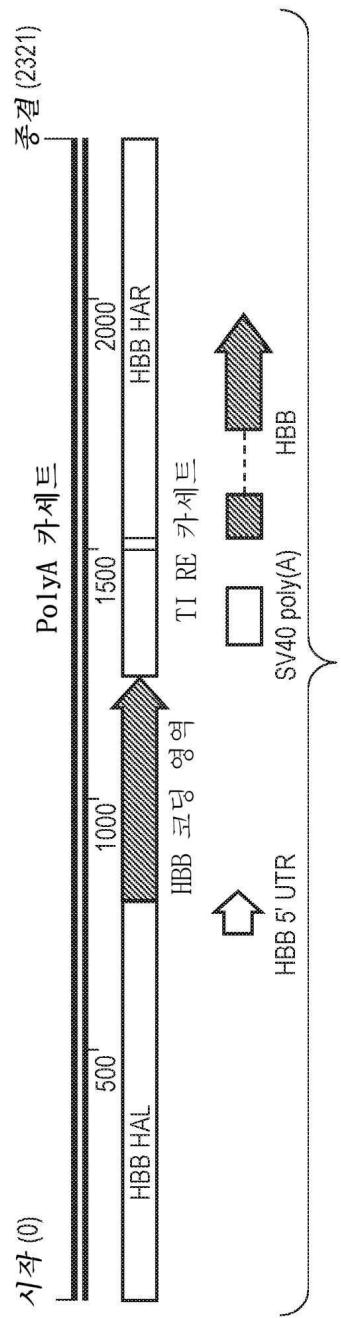
도면5c



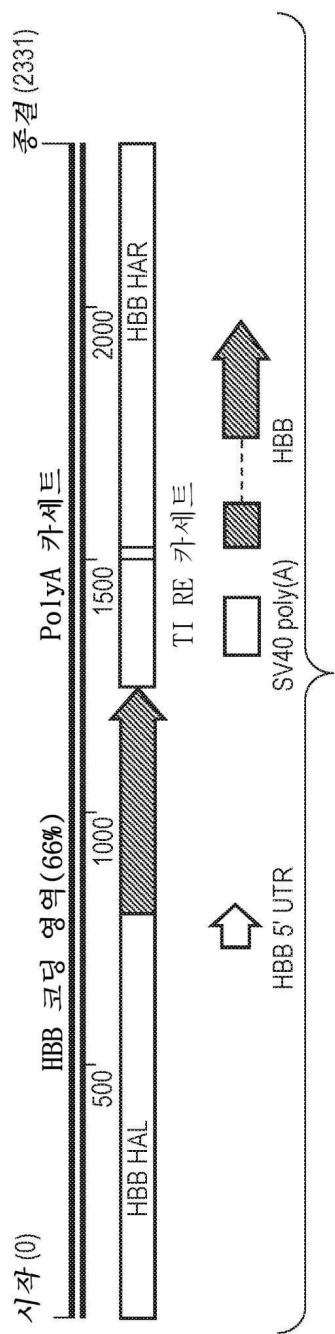
## 도면6



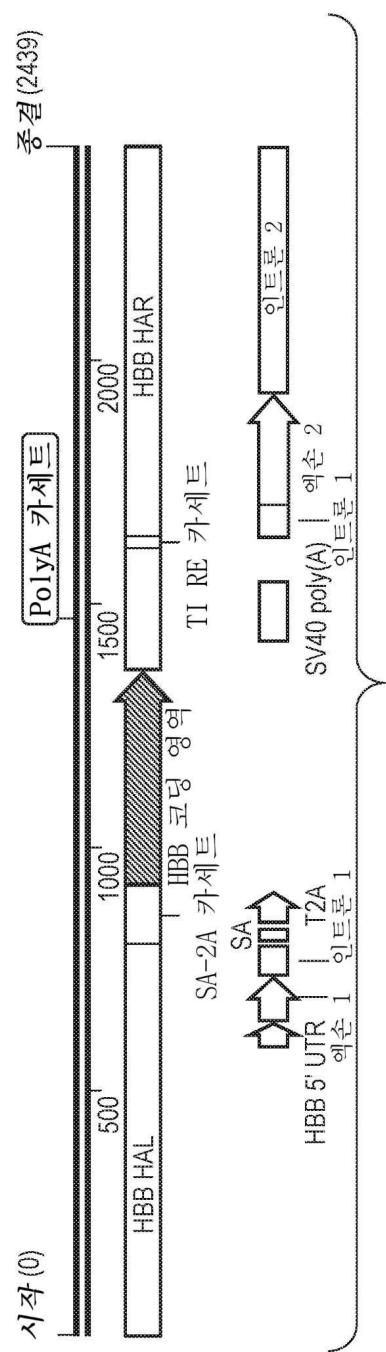
### 도면7a



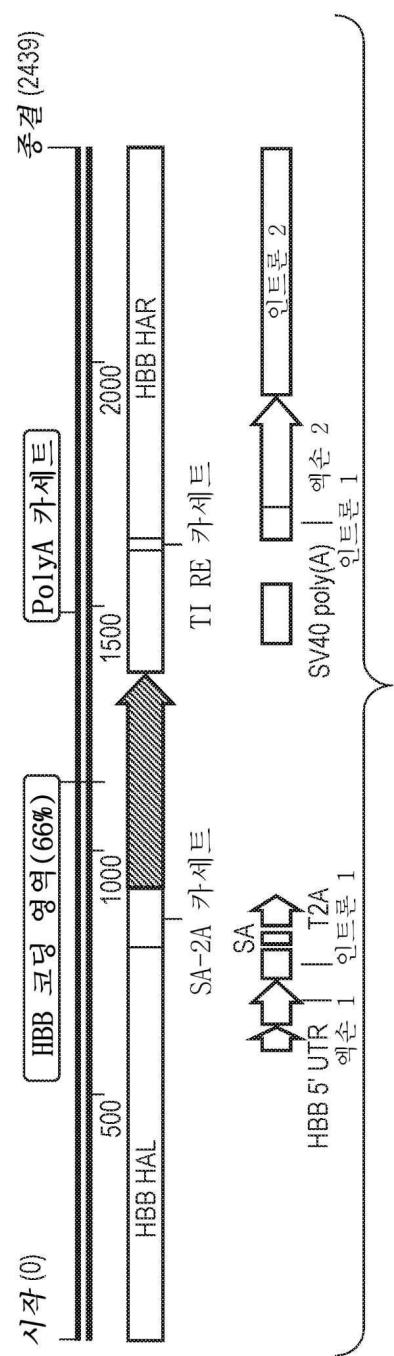
도면7b



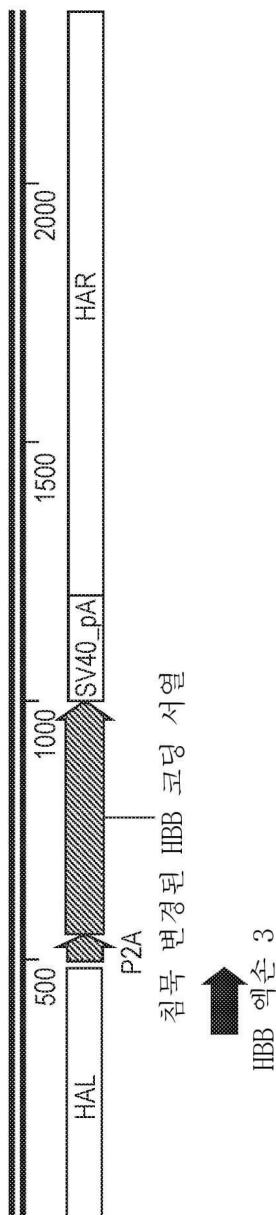
## 도면7c



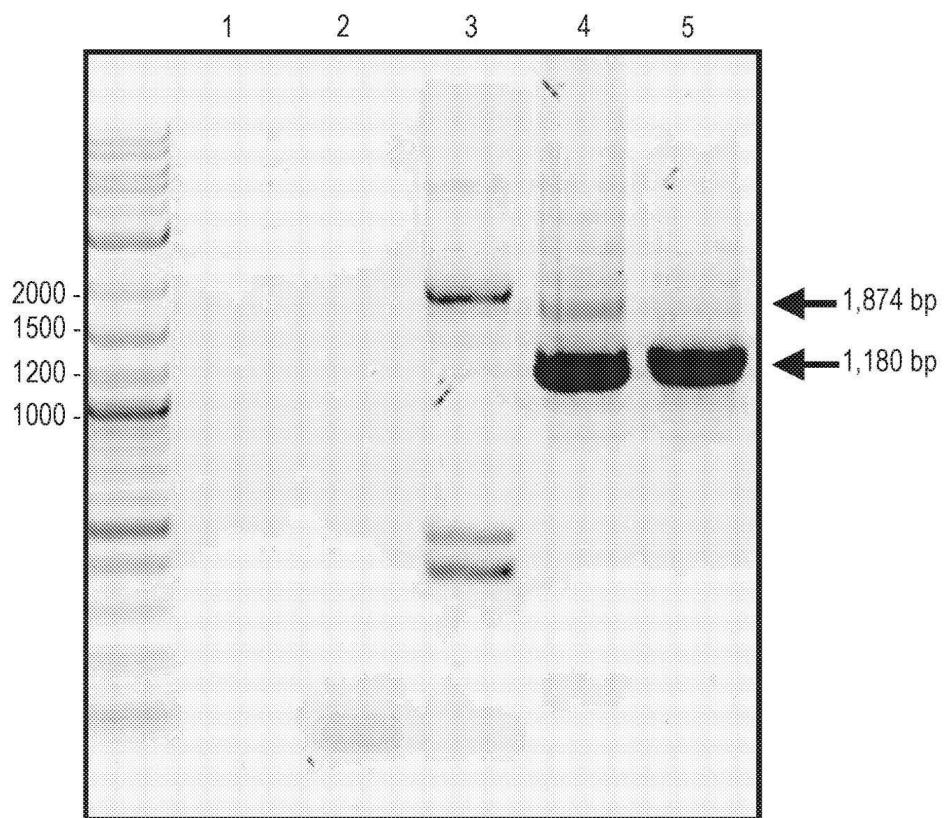
도면7d



도면 7e



## 도면8

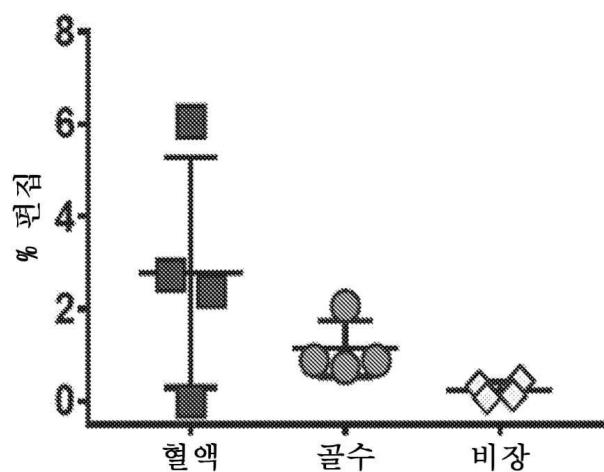


레인:

1. 음성 대조군
2. 형질도입되지 않은 세포
3. hHBB-h1W-002로 형질도입된 RKO 세포
4. hHBB-h1-010으로 형질도입된 RKO 세포
5. hHBB-h1-010으로 형질도입된 LCL 세포

## 도면9

## PPP1R12C



## 서 열 목록

## SEQUENCE LISTING

&lt;110&gt; CITY OF HOPE

HOMOLOGY MEDICINES, INC.

&lt;120&gt; ADENO-ASSOCIATED VIRUS COMPOSITIONS FOR RESTORING HBB GENE

FUNCTION AND METHODS OF USE THEREOF

&lt;130&gt; IPA200361-US

&lt;150&gt; US 62/574,163

&lt;151&gt; 2017-10-18

&lt;150&gt; US 62/621,102

&lt;151&gt; 2018-01-24

&lt;160&gt; 115

&lt;170&gt; PatentIn version 3.5

&lt;210&gt; 1

&lt;211&gt; 736

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; adeno-associated AAV9

&lt;400&gt; 1

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 Ala Ala Asp 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 Ala 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

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
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
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
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		
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 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> 2

<211> 736

<212> PRT

<213> Artificial Sequence

<220><223> novel AAV isolate

<400> 2

Met Thr 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 Ala Ala Asp 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 Ala 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 Gln 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
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu		
400	405	410
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu		
415	420	425
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser		
430	435	440
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser		
445	450	455
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro		
460	465	470
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn		
475	480	485
Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn		
490	495	500
Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys		
505	510	515
Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly		
520	525	530
Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile		
535	540	545
Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser		
555	560	565
570		
575		

Tyr Gly Gln Val Ala Thr Asn His Gln Ser 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

&lt;210&gt; 3

&lt;211&gt; 736

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; novel AAV isolate

&lt;400&gt; 3

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 Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp		
65	70	75
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 Ala Gly Ile Gly		
145	150	155
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		
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
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
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
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
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 Gly 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 Gly 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> 4  
 <211> 736  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> novel AAV isolate  
 <400> 4  
 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											
Ile	Asn	Ala	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	Ala	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
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 Tyr Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met		
625	630	635
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
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735
<210> 5		
<211> 736		

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; novel AAV isolate

&lt;400&gt; 5

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu 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 Ala Ala Asp 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 Ala Gly Ile Asp

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> 6

<211> 736

<212> PRT

<213> Artificial Sequence

<220><223> novel AAV isolate

<400> 6

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 Ala 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 Leu 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 Ala 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		
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
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
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
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 Ser 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> 7

<211> 736

<212> PRT

<213> Artificial Sequence

<220><223> novel AAV isolate

<400> 7

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 Ala 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 Ala 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 Arg 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 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> 8  
 <211> 736  
 <212> PRT  
 <213> Artificial Sequence  
 <220><223> novel AAV isolate  
 <400> 8  
 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 Ala Val 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 Ala Gly Ile Gly		
145	150	155
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
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
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735
<210> 9		
<211> 736		
<212> PRT		
<213> Artificial Sequence		
<220><223> novel AAV isolate		
<400> 9		
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser		
1	5	10
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 Ala 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 Arg Glu Pro Asp Ser Ser Ala 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> 10

<211> 736

<212> PRT

<213> Artificial Sequence

<220><223> novel AAV isolate

<400> 10

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 Ala 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 Ala 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 Cys 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> 11

<211> 736

<212> PRT

<213> Artificial Sequence

<220><223> novel AAV isolate

<400> 11

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu 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 Ala Ala Asp Ala Ala Ala Leu Glu His Asp Arg 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 Ala 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
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
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 Lys 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
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735

&lt;210&gt; 12

&lt;211&gt; 736

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; novel AAV isolate

&lt;400&gt; 12

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 Ala Ala Asp 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 Arg

130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala 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 His 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 Asn

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 Arg 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 Met 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> 13			
<211> 736			
<212> PRT			
<213> Artificial Sequence			
<220><223> novel AAV isolate			
<400> 13			
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	Ala	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 Ala 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 Arg 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

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> 14

<211> 26

<212> DNA

<213> Artificial Sequence

<220><223> Splice acceptor

<400> 14

ctgacacctt ctttcctcc cacagg 26

<210> 15

<211> 736

<212> PRT

<213> Artificial Sequence

<220><223> novel AAV isolate

<400> 15

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 Ala Ala Asp 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 Ala 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			
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 Arg 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		
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
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 Arg 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
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735
<210> 16		
<211> 736		
<212> PRT		
<213> Artificial Sequence		
<220><223> novel AAV isolate		
<400> 16		
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 Ala 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 Ala 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

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 Ala 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 Arg 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
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
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
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735
<210> 17		
<211> 736		
<212> PRT		
<213> Artificial Sequence		
<220><223> novel AAV isolate		
<400> 17		
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 Ala Ala Asp 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 Ala 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 Ile Ala Trp Pro Arg 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
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
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 Cys Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val		
705	710	715
Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu		
725	730	735
<210> 18		
<211> 145		

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; AAV2 5' ITR

&lt;400&gt; 18

ttggccactc cctctctgctcg cgctcgctcg ctcactgagg ccgggcgacc aaaggtcgcc	60
cgacgcccgg gcttgcggccg ggccgcctca gtgagcgagc gagcgcgcag agagggagtg	120
gccaaactcca tcacttagggg ttctt	145

&lt;210&gt; 19

&lt;211&gt; 145

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; AAV2 3' ITR

&lt;400&gt; 19

aggaaccctt agtgatggag ttggccactc cctctctgctcg cgctcgctcg ctcactgagg	60
ccgggcgacc aaaggtcgcc cgacgcccgg gcttgcggccg ggccgcctca gtgagcgagc	120
gagcgcgcag agagggagtg gccaa	145

&lt;210&gt; 20

&lt;211&gt; 167

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; AAV5 5' ITR

&lt;400&gt; 20

ctctcccccc tgcgcgttc gtcgctcg tggctcggtt ggggggggtgg cagctcaaag	60
agctgccaga cgacggccct ctggccgtcg cccccccaaa cgagccagcg agcgagcga	120
cgcgcacagg gggagagtgc cacactctca agcaaggggg tttgtta	167

&lt;210&gt; 21

&lt;211&gt; 167

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; AAV5 3' ITR

&lt;400&gt; 21

tacaaaacctt cttgtgttga gagtgtggca ctctcccccc tgcgcgttc gtcgctcg	60
tggctcggtt ggggggggtgg cagctcaaag agctgccaga cgacggccct ctggccgtcg	120

cccccccaaa cgagccagcg agcgagcgaa cgcgacaggg gggagag 167  
<210> 22  
<211> 621  
<212> PRT  
<213> Artificial Sequence  
<220><223> AAV2 Rep  
<400> 22  
Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp  
1 5 10 15

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

Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu  
85 90 95  
Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile  
100 105 110  
Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu  
115 120 125  
Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly  
130 135 140

Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys  
145 150 155 160  
Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu  
165 170 175  
Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His  
180 185 190  
Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn

195	200	205
-----	-----	-----

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr

210	215	220
-----	-----	-----

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys

225	230	235	240
-----	-----	-----	-----

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala

245	250	255
-----	-----	-----

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys

260	265	270
-----	-----	-----

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln

275	280	285
-----	-----	-----

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

290	295	300
-----	-----	-----

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala

305	310	315	320
-----	-----	-----	-----

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala

325	330	335
-----	-----	-----

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro

340	345	350
-----	-----	-----

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp

355	360	365
-----	-----	-----

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala

370	375	380
-----	-----	-----

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg

385	390	395	400
-----	-----	-----	-----

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val

405	410	415
-----	-----	-----

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser

420	425	430
-----	-----	-----

Thr Thr Phe Glu His Gln Gln Pro Leu Gln Asp Arg Met Phe Lys Phe

435	440	445
-----	-----	-----

Glu Leu Thr Arg Arg Leu Asp His Asp Phe Gly Lys Val Thr Lys Gln

450 455 460

Glu Val Lys Asp Phe Phe Arg Trp Ala Lys Asp His Val Val Glu Val

465 470 475 480

Glu His Glu Phe Tyr Val Lys Lys Gly Gly Ala Lys Lys Arg Pro Ala

485 490 495

Pro Ser Asp Ala Asp Ile Ser Glu Pro Lys Arg Val Arg Glu Ser Val

500 505 510

Ala Gln Pro Ser Thr Ser Asp Ala Glu Ala Ser Ile Asn Tyr Ala Asp

515 520 525

Arg Tyr Gln Asn Lys Cys Ser Arg His Val Gly Met Asn Leu Met Leu

530 535 540

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

545 550 555 560

Phe Thr His Gly Gln Lys Asp Cys Leu Glu Cys Phe Pro Val Ser Glu

565 570 575

Ser Gln Pro Val Ser Val Val Lys Lys Ala Tyr Gln Lys Leu Cys Tyr

580 585 590

Ile His His Ile Met Gly Lys Val Pro Asp Ala Cys Thr Ala Cys Asp

595 600 605

Leu Val Asn Val Asp Leu Asp Asp Cys Ile Phe Glu Gln

610 615 620

<210> 23

<211> 1833

<212> DNA

<213> Homo sapiens

<400> 23

cggctgtcat cacttagacc tcaccctgtg gagccacacc ctagggttg ccaatctact 60

cccaggagca gggagggcag gagccaggc tgcccataaa agtcaggca gagccatcta 120

ttgccttacat ttgcttctga cacaactgtg ttcactagca acctcaaaca gacaccatgg 180

tgcatctgac tcctgaggag aagtctgccg ttactgcct gtggggcaag gtgaacgtgg 240

atgaagtgg tggtgaggcc ctggcgagg tggtatcaag gttacaagac aggtttaagg	300
agaccaata gaaactggca tgtggagaca gagaagactc ttgggttct gataggcact	360
gactctctc gcctatttgtt ctatttccc acccttaggc tgctggtgtt ctacccttgg	420
acccagagg tctttgagtc ctggggat ctgtccactc ctgtatgtt tatggcaac	480
cctaaggta aggctcatgg caagaaagtgc ctgggtgcct ttagtgtatgg cctggctcac	540
ctggacaacc tcaagggcac ctggccaca ctgagtgagc tgcaatgttca aagctgtcac	600

gtggatccctg agaacttcag ggtgagtcta tgggacgctt gatgtttct ttcccctct	660
tttctatggtaaagttcatgtcataaggaaggggataagtaacagggtaca gtttagaaatg	720
ggaaacagac gaatgattgc atcagtgtgg aagtctcagg atcgtttag tttcttttat	780
ttgcgtttca taacaattgtttctttgtttaattcttg ctttcttttttttcttctc	840
cgcatttttactattatacttaatgcctt aacattgtgtataacaaaag gaaatatctc	900
ttagatacat taagtaactt aaaaaaaaaac ttacacagt ctgcctagta cattactatt	960
tggaaatataatgtgtgttat ttgcataatcataatctcc tactttattttcttttattt	1020

ttaattgata cataatcatt atacatattt atgggttaaa gtgtatgtt ttaatatgtg	1080
tacacatatt gaccaaatca ggtaatttt gcatttgtaa tttaaaaaa tgcttccttc	1140
tttaaatata ctttttgtt tatcttattt ctaatacttt ccctaattctc tttctttcag	1200
ggcaataatg atacaatgta tcatgcctct ttgcaccatt ctaaagaata acagtgataa	1260
tttctgggtt aaggcaatag caatatctct gcatataaat atttctgcat ataaattgtt	1320
actgatgtaa gaggttcat attgctaata gcagctacaa tccagctacc attctgttt	1380
tatTTtatgg ttgggataag gctggattat tctgagtcga agctaggcccc ttttgctaat	1440

catgttcata cctcttatct tcctccaca gctcctggc aacgtgctgg tctgtgtgct	1500
ggcccatcac ttggcaaag aattcacccc accagtgcag gctgcctatc agaaaagtgg	1560
ggctggtgtg gctaattgcc tggcccacaa gtatcaactaa gctcgcttc ttgctgtcca	1620
atttctatta aaggttcctt tgttccctaa gtccaaactac taaaactgggg gatattatga	1680
agggccttga gcatctggat tctgcctaat aaaaaacatt tatttcatt gcaatgatgt	1740
atttaaatata ttctgtataa tttactaaa aaggaaatgt gggaggtcag tgcatttaaa	1800
acataaagaa atgaagagct agttcaaaacc ttg	1833

<210> 24

1833

<212> DNA

<213> Artificial Sequence

<220><223> HBB sequence from 5' mutation to 3' mutation with silent codon  
alterations

<400> 24

cggttgtcat cacttagacc tcaccctgtg gagccacacc ctagggttgg ccaatctact	60
cccaggagca gggagggcag gagccaggc tggtcataaa agtcaggca gagccatcta	120
ttgtttacat ttgttctga cacaactgtg ttcactagca acctcaaaca gacaccatgg	180
tccacactac accagaagaa aaaagtgtg taacagctt ctggggaaaa gtcaatgtcg	240
acgggttgg gggggaaagct ctggaaagggt tggtatcaag gttacaagac aggttaagg	300
agaccaatag aaactggca tgtggagaca gagaagactc ttgggttct gataggcact	360
gactctctt gcctatttgtt ctatccc acccttagac tcctcgctgt gtatccatgg	420
acacaagat ttgcgaaag cttcgagac ctcagcacac cagacgcagt aatggaaat	480
ccaaaagtca aagcacacgg aaaaagggtc ctggggctt tctctgacgg actcgacat	540
ctcgataatc tggaaaggaaat attcgctacc ctctctgaac tccattgcga taaactccat	600
gtcgaccacag aaaattttag agtgagtcta tgggacgctt gatgtttct ttcccttct	660
tttctatggtaatgtt taatgttcatg tcataggaag gggataagta acagggtaca gtttagaatg	720
ggaaacagac gaatgattgc atcagtgtgg aagtctcagg atcggttttag tttcttttat	780
ttgtgttca taacaattgtt tttctttgtt ttaattctgtt ctttcttttt ttttcttctc	840
cgcattttt actattatac ttaatgcctt aacattgtt ataacaaaag gaaatatctc	900
ttagatacat taatgttactt aaaaaaaaaac tttacacagt ctgcctagta cattactatt	960
tggaaatataat gtgtgtttat ttgcattttc ataatctcc tacttttattt tctttttttt	1020
ttaattgtataatcattt atacatattt atgggttaaa gtgtatgtt ttaatatgtg	1080
tacacatattt gaccaatca gggtaatttt gcattgtaa ttttaaaaaa tgcttttttc	1140
ttttaaatataat cttttttgtt tatcttattt ctaatacttt ccataatctc tttctttcag	1200
ggcaataatg atacaatgtt tcatgcctt ttgcaccatt ctaaagaata acagtgtataa	1260
tttctgggtt aaggcaatag caatatctct gcatataat attctgtcat ataaattgtt	1320
actgtatgtttaa gaggtttcat attgctaata gcagctacaa tccagctacc attctgtttt	1380
tatTTTATGG ttggataag gctggattat tctgagtcca agctaggccc ttttgcataat	1440
catgttccata ccttttatct tctcccaca gctgctcgaa aatgtcctcg tgtgcgtct	1500
cgctcaccat ttggaaagg agtttacacc tcctgtccaa gcagcttacc aaaagggtgt	1560
cgcaacgcgtc tcgctcataa ataccattag gctcgcttcc ttgtgttca	1620
atttcttataa aagtttcattt tggccctaa gtccaaactac taaactgggg gatattatgt	1680

agggccttga gcatctggat tctgccta ataaaaacatt tattttcatt gcaatgatgt	1740
atttaaatta ttctcgataa ttttactaaa aagggaatgt gggaggtcag tgcatttaaa	1800
acataaaagaa atgaagagct agttcaaacc ttg	1833
<210> 25	
<211> 1845	
<212> DNA	
<213> Artificial Sequence	
<220><223> HBB sequence from 5' mutation to 3' mutation with Linker	
<400> 25	
cggttgtcat cacttagacc tcaccctgtg gagccacacc ctagggttgg ccaatctact	60
cccaggagca gggagggcag gagccaggc tggtcataaa agtcaggcag gagccatcta	120
ttgcttacat ttgcttctga cacaactgtg ttcaactagca acctcaaaaca gacaccatgg	180
tgcacatgtac tcctgaggag aagtctgccc ttactgcct gtggggcaag gtgaacgtgg	240
atgaagttgg tggtgaggcc ctggcaggt tggtatcaag gttacaagac aggactagta	300
tcgatttaa ggagaccaat agaaactggg catgtggaga cagagaagac tcttgggtt	360
ctgataggca ctgactctct ctgcctatttgc tctatatttcc accccttag gctgtgg	420
gtctaccctt ggacccagag gttcttttag tcctttgggg atctgtccac tcctgtatgt	480
gttatggca accctaaggt gaaggctcat ggcaagaaag tgctcggtgc ctttagtgtat	540
ggcctggctc accttggacaa cctcaaggc acctttgcca cactgagtga gctgcactgt	600
gacaagctgc acgtggatcc tgagaacttc agggtagtc tatggacgc ttgtatgttt	660
ctttccctt ctttctatg gttaagttca tgtcatagga agggataag taacaggta	720
cagtttagaa tggaaacag acgaatgatt gcatcgtgt ggaagttca ggatcg	780
agtttctttt attgtgtt cataacaatt gttttttt gttaattct tgctttttt	840
ttttttcttc tccgcaattt ttactattt acttaatgcc ttaacattgt gtataacaaa	900
aggaaatatac tctgagatac attaagtaac ttaaaaaaaaaa actttacaca gtctgcctag	960
tacattacta ttggaaatat atgtgtgtt atttgcatat tcataatctc cctactttat	1020
tttctttat tttaattga tacataatca ttatacatat ttatgggtta aagtgtatg	1080
tttaaatatg tgtacacata ttgaccaat cagggtaatt ttgcattgt aattttaaaa	1140
aatgtttct tcttttaata tacttttttgc ttatcttat ttctaatact ttccctaatac	1200
tcttttttc agggcaataa tgatacaatg tatcatgcct cttgcacca ttctaaagaa	1260
taacagtgtat aatttctggg ttaaggcaat agcaatatct ctgcataataa atatttctgc	1320

atataaattg taactgatgt aagagggttc atattgctaa tagcagctac aatccagcta	1380
ccattctgct ttattttat ggttggata aggctggatt attctgagtc caagctaggc	1440
cctttgcta atcatgttca tacctttat cttccccc aagctcctgg gcaacgtgct	1500
gtcttgttg ctggccatc actttggcaa agaattcacc ccaccagtgc aggctgccta	1560
tcagaaaatg gtggctggtg tggctaatgc cctggccac aagtatcact aagctcgctt	1620
tcttgctgta caatttctat taaagggttcc tttgtccct aagtccaaact actaaactgg	1680
gggatattat gaagggcctt gagcatctgg attctgccta ataaaaaaca tttatttca	1740
ttgcaatgtat gtatTTaaat tatttctgaa tattttacta aaaagggaaat gtgggaggtc	1800
agtgcattta aaacataaag aatgaagag ctagttcaaa ctttg	1845
<210> 26	
<211> 1845	
<212> DNA	
<213> Artificial Sequence	
<220><223> HBB sequence from 5' mutation to 3' mutation with silent codon alterations, with Linker	
<400> 26	
cggtgtgtcat cacttagacc tcaccctgtg gagccacacc ctagggttgg ccaatctact	60
cccaggagca gggagggcag gagccaggc tggtcataaaa agtcaggc gagccatcta	120
ttgcttacat ttgttctga cacaactgtg ttcaactagca acctcaaaaca gacaccatgg	180
tccacctcac accagaagaa aaaagtgtct taacagctct ctggggaaaa gtcaatgtcg	240
acgaggtagg gggggaaagct ctcggaagggt tggtatcaag gttacaagac aggactagta	300
tgcattttaa ggagaccaat agaaactggg catgtggaga cagagaagac tcttgggtt	360
ctgataggca ctgactctct ctgcctattt gtctatTTcc accccttag actcctcgctc	420
gtgtatccat ggacacaaag attttcgaa agcttcggag acctcagcac accagacgca	480
gtaatggaa atccaaaatg caaagcacac ggaaaaaagg tcctggggc tttctctgac	540
ggactcgcac atctcgataa tctgaaagga acattcgcta ccctctctga actccattgc	600
gataaaactcc atgtcgaccc agaaaatttt agagtgagtc tatggacgc ttgatgttt	660
ctttccccctt ctttctatg gtttaagtca tgtcatagga agggataag taacaggta	720
cagtttagaa tggaaacag acgaatgatt gcatcagtgt ggaagtcata ggatcgttt	780
agtttctttt attgctgtt cataacaatt gttttctttt gtttaattct tgctttttt	840
ttttttcttc tccgcaattt ttactattat acttaatgcc ttaacattgt gtataacaaa	900

aggaaatatac tctgagatac attaagtaac ttaaaaaaaaa actttacaca gtctgcctag	960
tacattacta ttggatat atgtgtgctt atttgcatat tcataatctc cctactttat	1020
tttctttat tttaattga tacataatca ttatacatat ttatgggta aagtgtaatg	1080
tttaatatg tgtacacata ttgaccaaatt caggtaatt ttgcattgt aattttaaaa	1140
aatgcttct tctttata tactttttg tttatcttat ttctaatact ttccctaatac	1200
tcttccttc aggcaataa tgatacaatg tatcatgcct ttgcacca ttctaaagaa	1260
taacagtatg aatttctggg ttaaggcaat agcaataatct ctgcataataa atatttctgc	1320
atataaatttga taactgatgt aagaggttc atattgctaa tagcagctac aatccagctaa	1380
ccattctgct ttatTTTat ggTTGGGata aggctggatt attctgagtc caagctaggc	1440
cctttgcta atcatgttca tacctttat cttccccc cagctgctcg gaaatgtcct	1500
cgtgtgcgtc ctgcgtcacc atttcggaaa ggagtttaca cttctgtcc aagcagctta	1560
ccaaaaggctc gtcgcagggg tcgcaaacgc tctcgctcat aaataccatt aggctcgctt	1620
tcttgctgtc caatttctat taaaggttcc ttgttccct aagtccaact actaaactgg	1680
gggatattat gaagggcctt gagcatctgg attctgccta ataaaaaaca ttatTTTca	1740
ttgcaatgtat gtatTTTat tatttctgaa tattttacta aaaaggaaat gtggagggtc	1800
agtgcattta aaacataaag aatgaagag ctagttcaaa ctttg	1845
<210> 27	
<211> 444	
<212> DNA	
<213> Homo sapiens	
<400> 27	
atggtgcatc tgactcctga ggagaagtct gccgttactg ccctgtgggg caagggtgaac	60
gtggatgaag ttgggtgtga ggccctgggc aggctgctgg tggctaccc ttggaccagg	120
agttctttg agtccttgg ggatctgtcc actcctgatg ctgttatggg caaccctaag	180
gtgaaggctc atggcaagaa agtgctcggt gccttagtgc atggcctggc tcacctggac	240
aacctcaagg gcaccttgc cacactgagt gagctgcact gtgacaagct gcacgtggat	300
cctgagaact tcaggctcct gggcaacgtg ctggctgtg tgctggccca tcactttggc	360
aaagaattca cccccaccagt gcaggctgcc tatcagaaag tggtggtgg tggctaat	420
gccctggccc acaagtatca ctaa	444
<210> 28	
<211> 444	

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; cDNA sequences with silent codon alterations

&lt;400&gt; 28

atggtccacc tcacaccaga agaaaaaaagt gctgtaacag ctctctgggg aaaagtcaat 60

gtcgacgagg taggggggga agctctcgga agactcctcg tcgtgtatcc atggacacaa 120

agattttcg aaagcttcgg agacacctcgc acaccagacg cagtaatggg aaatccaaaa 180

gtcaaaggcac acggaaaaaa ggtcctgggg gccttctctg acggactcgc acatctcgat 240

aatctgaaag gaacattcgc taccctctt gaactccatt gcgataaact ccatgtcgac 300

ccagaaaatt ttagactgct cgaaatgtc ctctgtgcg tcctcgctca ccatttcgga 360

aaggagttt cacccctgtt ccaaggcaget taccaaaagg tcgtcgagg ggtcgcaaac 420

gctctcgctc ataaatacca ttag 444

&lt;210&gt; 29

&lt;211&gt; 3533

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; hHBB-hL-014 vector, ITRs not included

&lt;400&gt; 29

atcaacctag gctccagata gccatagaag aaccaaacac ttctgcgtg tgtgagaata 60

atcagagtga gatTTTCA caagtacctg atgagggttg agacaggtg aaaaagttag 120

agatctctat ttatTTGCA ataATAGAGA aAGCATTAA gAGATAAAAG CAATGGAAAT 180

aAGAAATTG TAAATTCCt TCTGATAACT AGAAATAGAG GATCCAGTT CTTTGGTTA 240

ACCTAAATTt TATTCAtt TATTGTTTA TTTTATTtA TTTTGTGAA 300

TCGTAGTTc AGAGTGTAG AGCTGAAAGG AAGAAGTAGG AGAAACATGC AAAGTAAAAG 360

tataacactt tccttactaa accgacatgg gtttccaggt agggcagga ttcaggatga 420

ctgacaggc ctttagggaa cactgagacc ctacgctgac ctctaaatg cttgctacct 480

ttgcgtttt aattacatct ttaatAGCA ggaAGCAGAA ctctgcactt caaaagtttt 540

tcctcacctg aggagttaat ttagtacaag gggaaaaagt acagggggat gggagaaagg 600

cgatcacgtt ggaaagctat agagaaagaa gagtaatTT tagtaaAGGA gtttaaaca 660

AACAAAAT AACAGAGAAAT AGGAACtGA ATCAAGGAA TGATTTAAA ACgcagtatt 720

cttagtgac tagaggaaaa aaataatctg agccaagtag aagacctttt cccctctac 780

ccctacttgc taagtcacag aggcttttgc ttccccaga cactctgca gattagtcca	840
ggcagaaaca gtttagatgtc cccagttAAC ctcctatttgc acaccactga ttaccccatt	900
gatagtaca cttgggttg taagtactt tttatttttgtt actgcattaa	960
gaggctctta gtttttatc tcttggttcc caaaaccta taagtaacta atgcacagag	1020
cacattgatt tgtatttttattt ctatTTTAC acataattta ttagcatgca tgagcaaatt	1080
aagaaaaaca acaacaaatg aatgcataata tatgtatatg tatgtgtgtatatacaca	1140
catatatata tatatTTTTT ctttcttac cagaaggTTT taatccaaat aaggagaaga	1200
tatgcttaga accgaggttag agttttcatc cattctgtcc tgtaagtattt ttgcatattc	1260
tggagacgca ggaagagatc catctacata tcccaagct gaattatggt agacaaaact	1320
cttccacttt tagtgcattca acttcttattt tgtgtataaa gaaaattggg aaaacgatct	1380
tcaaatatgt taccaagctg tgattccaaa tattacgtaa atacacttgc aaaggaggat	1440
gttttagta gcaatttgc ctgatggat gggccaaga gatatatctt agagggagg	1500
ctgagggttt gaagtccaaac tcctaagcca gtgccagaag agccaaggac aggtacggct	1560
gtcatcactt agacctcacc ctgtggagcc acaccctagg gttggccaat ctactcccag	1620
gaggcaggaggg ggcaggagcc agggctgggc ataaaagtca gggcagagcc atctattgt	1680
tacatttgc tctgacacaa ctgtgttac tagcaaccc tc aaacagacac catggcat	1740
ctgactcctg aggagaagtc tgccgttact gcccgtggg gcaagggtgaa cgtggatgaa	1800
gttgggtgtt aggcctggg caggtggta tcaagggttac aagacaggac tagtatcgat	1860
ttaaggaga ccaatagaaa ctggcatgt ggagacagag aagactctt gtttctgtat	1920
aggcaactgac tctctctgccc tattggtcta tttccacc cttaggctgc tgggtgtcta	1980
cccttggacc cagaggttct ttgagtccctt tggggatctg tccactcctg atgctgttat	2040
ggcaaccctt aaggtgaagg ctcatggcaaa gaaagtgc ggtgcctta gtgtggcct	2100
ggctcacctg gacaacctca agggcacctt tgccacactg agtgagctgc actgtgacaa	2160
gctgcacgtg gatctgaga acttcagggt gagtctatgg gacgctgtat gtttcttcc	2220
cccttctttt ctatggtaa gttcatgtca taggaagggg ataagtaaca gggtacagtt	2280
tagaatggaa aacagacgaa tgattgcattc agtgtggaag tctcaggatc gttttagttt	2340
cttttatttgc ctgttcataaa caattgtttt ctttgttta attcttgctt tctttttttt	2400
tcttctccgc aattttactt attatactta atgccttaac attgtgtata acaaaaggaa	2460
atatctctga gatacattaa gtaacttaaa aaaaaactttt acacagtctg cctagttat	2520
tactattttgg aatataatgtg tgcttatttgc catattcata atctccctac tttatTTTCT	2580
tttattttta attgatacat aatcattata catattatg gtttaaagtg taatgtttta	2640

atatgtgtac acatattgac caaatcaggg taatttgca tttgttaattt taaaaaatgc	2700
tttcttcattt taataactt ttttgttat cttattcta atacttccc taatctttt	2760
ctttcaggcc aataatgata caatgtatca tgcctttg caccattcta aagaataaca	2820
gtgataattt ctgggttaag gcaatagcaa tatctctgca tataaatattt tctgcatata	2880
aattgtaact gatgttaagag gtttcatatt gctaatagca gctacaatcc agtaccatt	2940
ctgcctttat ttatggttg ggataaggct ggattattct gagtccaagc taggccctt	3000
tgctaatcat gttcataacctt cttatcttcc tcccacagct cctggcaac gtgctggct	3060
gtgtgctggc ccatcacttt ggcaaagaat tcacccacc agtgcaggct gcctatcaga	3120
aagtggtgcc tggtgtggct aatgccctgg cccacaagta tcactaagct cgcttcttgc	3180
ctgtccaatt tctattaaag gttcctttgt tccctaagtc caactactaa actggggat	3240
attatgaagg gccttgagca tctggattct gcctaataaa aaacatttat tttcattgca	3300
atgatgtatt taaatttattt ctgaatattt tactaaaag ggaatgtggg aggtcagtgc	3360
atttaaaaca taaagaaatg aagagctagt tcaaacccttg ggaaaataca ctatactta	3420
aactccatga aagaaggtaa ggctgcaaac agctaatgca cattggcaac agccctgat	3480
gcatatgcct tattcatccc tcagaaaagg attcaagtag aggcttgatt tgg	3533
<210> 30	
<211> 3521	
<212> DNA	
<213> Artificial Sequence	
<220><223> hHBB-hL-014 vector, ITRs not included, Linker removed	
<400> 30	
atcaaccttag gctccagata gccatagaag aaccaaacac tttctgcgtg tgtgagaata	60
atcagagtga gatTTTTca caagtacctg atgagggttg agacaggttag aaaaagttag	120
agatctctat ttatTTGca ataatacgata aagcatttaa gagaataaaag caatggaaat	180
aagaaatttg taaatttcct tctgataact agaaatagag gatccagttt ctTTTGGTTA	240
acctaaattt tatttcattt tattgtttt ttttattttt ttttattttt ttttGTGtaa	300
tcgttagttc agagtgttag agctgaaagg aagaagttagg agaaacatgc aaagtaaaag	360
tataacactt tccttactaa accgacatgg gtttccaggt agggcagga ttcaggatga	420
ctgacaggcc ccttagggaa cactgagacc ctacgctgac ctcataaaatg cttgctacct	480
ttgctgtttt aattacatct ttaatagca ggaagcagaa ctctgcactt caaaagtttt	540
tcctcacctg aggagttaat ttagtacaag gggaaaaagt acagggggat gggagaaagg	600

cgatcacgtt gggaaagctat agagaaaagaa gagtaaattt tagtaaagga ggttaaaca	660
aacaaaatat aaagagaaaat aggaacttga atcaaggaaa tgattttaaa acgcagtatt	720
cttagtgac tagagaaaaa aaataatctg agccaagtag aagaccttt cccctctac	780
ccctactttc taagtcacag aggcttttg ttccccaga cactctgca gattagtc	840
ggcagaaaaca gtttagatgtc cccagttac ctcttattt acaccactga ttaccccatt	900
gatagtcaca ctttgggttg taagtgactt tttattttt tgtattttt actgcattaa	960
gaggctctta gtttttatac tcttgtttcc caaaacctaa taagtaacta atgcacagag	1020
cacattgatt tgtattttt ctattttag acataattta ttagcatgca tgagcaaatt	1080
aagaaaaaca acaacaaatg aatgcatata tatgtatatg tattgtgtat tatatacaca	1140
catatatata tatattttt ctttcttac cagaaggttt taatccaaat aaggagaaga	1200
tatgcttaga accgaggttag agttttcattc cattctgtcc ttagtattt ttgcattttc	1260
tggagacgca ggaagagatc catctacata tcccaaagct gaattatggt agacaaaact	1320
cttccacttt tagtgcattca acttcttatt tttgtataaa gaaaattggg aaaacgtct	1380
tcaaatatgct taccaagctg tgattccaaa tattacgtaa atacacttgc aaaggaggat	1440
gttttagta gcaatttgta ctgatggat gggccaaga gatatatctt agagggaggg	1500
ctgagggttt gaagtccaaac tcctaageca gtgccagaag agccaaggac aggtacggct	1560
gtcatcaattt agacctcacc ctgtggagcc acaccctagg gttggccaaat ctactccag	1620
gagcaggggag ggcaggagcc agggctggc ataaaagtc gggcagagcc atctattgct	1680
tacatttgct tctgacacaa ctgtgttac tagcaacccaa aaacagacac catggcat	1740
ctgactcctg aggagaagtc tgccgttact gccctgtggg gcaagggtgaa cgtggatgaa	1800
gttgtggtg aggccctggg caggttgta tcaagggttac aagacagggtt taaggagacc	1860
aatagaaaact gggcatgtgg agacagagaa gactcttggg tttctgtatgc gcactgactc	1920
tctctgccta ttggcttattt ttcccacct taggctgtcg gtggcttacc cttggaccca	1980
gagggttctt gagtccttg gggatctgtc cactctgtat gctgttatgg gcaaccctaa	2040
ggtgaaggct catggcaaga aagtgtcg gtcctttagt gatggcctgg ctcacctgga	2100
caacctcaag ggcacccctt ccacactgag tgagctgcac tttgtacaagc tgcacgtgga	2160
tcctgagaac ttcaagggtga gtctatggga cgcttgcgtt tttttttccc cttttttct	2220
atggtaagt tcatgtcata ggaaggggat aagtaacagg gtacagtttta gaatggaaa	2280
cagacgaatg attgcatcg tttggaaatgc tcaggatgtt tttttttctt ttatgttgc	2340
gttcataaca attgtttctt ttgtttaat tcttgcgtt ttttttttcc ttctccgaa	2400
tttttactat tatacttaat gccttaacat tttgtataac aaaagggaaat atctctgaga	2460

tacattaagt aactaaaaaa aaaactttac acagtctgcc tagtacatta ctatttgaa	2520
tatatgtgtg ctatattgca tattcataat ctccctactt tattttctt tattttaat	2580
tgatacataa tcattataca tatttatggg tttaaagtgt aatgtttaat atgtgtacac	2640
atattgacca aatcagggtt aatttgcatt tgtaattttt aaaaatgtt tcttcttttta	2700
ataatactttt ttgtttatct tatttctaat acttcccta atctttctt ttcagggcaa	2760
taatgataca atgtatcatg cctcttgc a ccattctaaa gaataacagt gataattct	2820
gggttaaggc aatagcaata tctctgcata taaatattt c tgcataataaa ttgttaactga	2880
tgtaagaggt ttcatattgc taatagcagc tacaatccag ctaccattct gcttttattt	2940
tatgttggg ataaggctgg attattctga gtccaagcta ggccttttig ctaatcatgt	3000
tcataccctt tatcttcctc ccacagctcc tggcaacgt gctggctgt gtgcgtggccc	3060
atcactttgg caaaaatttcc accccaccag tgcaggctgc ctatcagaaa gtggggctg	3120
gtgtggctaa tgccctggcc cacaagtatc actaagctcg ctttctgtct gtccaaattt	3180
tattaaaggt tcctttgttc ctaagtcca actactaaac tggggatata tatgaaggc	3240
ctttagcatc tggattctgc ctaataaaaa acatttattt tcattgcaat gatgtattt	3300
aattatttctt gaatatttttta ctaaaaaggg aatgtgggag gtcagtgcat taaaacata	3360
aagaaatgaa gagctagttc aaaccttggg aaaatacact atatctaaa ctccatgaaa	3420
gaagggtgagg ctgcaaacag ctaatgcaca ttggcaacag cccctgatgc atatgcctta	3480
ttcatccctc agaaaaggat tcaagtagag gcttgatttg g	3521
<210> 31	
<211> 3168	
<212> DNA	
<213> Artificial Sequence	
<220><223> hHBB-hL-001 vector, ITRs not included	
<400> 31	
ttattttttt gtattttga ctgcattaa aggtctctag ttttttatct cttgtttccc	60
aaaacctaattt aagtaactaa tgcacagagc acattgattt gtatttattt tattttttaga	120
cataatttat tagcatgcat gagcaaatttta agaaaaacaa caacaatga atgcataat	180
atgtatatgt atgtgtgtat atatacacac atatataat atatttttc ttttcttacc	240
agaagggtttt aatccaaata aggagaagat atgcttagaa ccgaggtaga gttttcatcc	300
attctgtcct gtaagtattt tgcattttct ggagacgcag gaagagatcc atctacat	360
cccaaagctg aattatggta gacaaaactc ttccactttt agtgcataa cttcttattt	420

gtgtaataag aaaattggga aaacgatctt caatatgctt accaagctgt gattccaaat	480
attacgtaaa tacacttgca aaggaggatg ttttagtag caatttgtac tgatggatg	540
ggccaagag atatatctt gagggaggc tgagggttg aagtccaaact cctaagccag	600
tgcagaaga gccaaggaca ggtacggctg tcatactt gacccaccc tgtggagcca	660
caccctagg ttggcaatc tactcccagg agcaggagg gcaggagcca gggctggca	720
taaaagttag ggcagagcca tctattgctt acatttgctt ctgacacaac tgtgttact	780
agcaaccta aacagacacc atggtcattc tgactcctga ggagaagtct gccgttactg	840
ccctgtgggg caaggtgaac gtggatgaag ttggatggta ggcctggc aggttggat	900
caaggtaaca agacaggact agtacgtt ttaaggagac caatagaaac tggcatgtg	960
gagacagaga agactttgg gttctgata ggcactgact ctctctgcattt attgttat	1020
tttccaccc ttaggctgctt ggtggctac cttggaccc agaggttctt tgagtccccc	1080
ggggatctgt ccactcctga tgctgtttag ggcaaccctt aggtgaaggc tcatggcaag	1140
aaagtgtcg gtgcctttag tcatggctt gctcacctgg acaaccta gggcacccccc	1200
gccacactga gtgagctgca ctgtgacaag ctgcacgtgg atcctgagaa cttcagggtg	1260
agtctatggg acgttggatg tttctttcc ctttttttcatggtaag ttcatgtcat	1320
aggaaggggta taagtaacag ggtacagttt agaatggaa acagacgaat gattgcatca	1380
gtgtggaagt ctctggatcg ttttagtttgc ttatgttgc ttttgcataac aattgttttc	1440
ttttgtttaa ttcttgcttt ctttttttt cttctccgca attttacta ttatacttaa	1500
tgccttaaca ttgtgtataa caaaaggaaa tatctctgag atacattaag taacttaaaa	1560
aaaaacttta cacagtctgc ctagtacatt actatttggaa atatatgtgt gcttatttgc	1620
atattcataa tctccctact ttatttttttt ttattttaa ttgatacata atcattatac	1680
atatttattgg gttaaagtgt aatgttttaa tatgtgtaca catattgacc aaatcagggt	1740
aatttttgttcat ttgtatattttt aaaaaatgtt ttctttttt aatatacttt tttttttatc	1800
ttatattctaa tactttccct aatcttttc ttccaggcataatgatac aatgtatcat	1860
gcctttgc accattctaa agaataacag tgataatttc tgggttaagg caatagcaat	1920
atctctgcat ataaatattt ctgcatataa attgtactg atgttaagg tttcatatttgc	1980
ctaatagcag ctacaatcca gctaccattc tgctttatt ttatgttgg gataaggctg	2040
gattattctg agtccaaatgtt aggccctttt gctaatcatg ttcataccc ttatcttctt	2100
cccacagctc ctggcaacg tgctggctg tggctggcc catcaacttg gcaaagaatt	2160
caccccaacca gtgcaggctg cctatcgaa agtggtgctt ggtgtggcta atgcctggc	2220

ccacaaggat cactaagctc gctttcttgc tgtccaattt ctattaaagg ttcctttgtt	2280
ccctaagtcc aactactaaa ctggggata ttatgaaggg ctttgagcat ctggattctg	2340
cctaataaaa aacatttattt ttcattgcaa tgatgtatTTT aaatttttc tgaatatTTT	2400
actaaaaagg gaatgtggga ggtcagtgca tttaaaacat aaagaaatga agagctatTTT	2460
caaacccttgg gaaaatacac tatacttAA actccatgaa agaaggtag gctgaaaca	2520
gctaattgcac attggcaaca gcccctgatg catatgcctt attcatccct cagaaaagga	2580
ttcaagtaga ggcttgattt ggaggttaaa gttttgtat gctgtatTTT acattactta	2640
ttgttttagc tgtcctcatg aatgtctttt cactaccat ttgcttatcc tgcatctc	2700
agccttgact ccactcgatTTT ctcttgctta gagataccac cttccctg aagtgttcc	2760
tccatgtttt acggcgagat ggTTTCTCT cgccctggcca ctcaGCTTA gttgtctctg	2820
ttgtcttata gaggctact tgaagaagga aaaacagggg tcatggTTT actgtcctgt	2880
gagcccttct tccctgcctc ccccactcac agtgaccggg aatctgcagt gctagctcc	2940
cggaactatc actcttcac agtctgtttt ggaaggactg ggcttagtat gaaaagttag	3000
gactgagaag aatttgaag gggctttt gtagctttagt attcactact gtcttattac	3060
cctgtcatag gcccacccca aatggaagtc ccattttcc tcaggatgtt taagattagc	3120
attcaggaag agatcagagg tctgctggct cccttatcat gtccctta	3168
<210> 32	
<211> 3156	
<212> DNA	
<213> Artificial Sequence	
<220><223> hHBB-hL-001 vector, ITRs not included, Linker removed	
<400> 32	
ttatTTTATTt gtatTTTGA ctgcattaAG aggtctctAG tttttatct ctTgtttccc	60
aaaacctaA aagtaactaa tgcacagAGC acattgattt gtatTTTATC tattttAGA	120
cataatttt tagcatgcat gagcaaattt agaaaaacAA caacaaatGA atgcataatat	180
atgtatATGT atgtgtgtat atatacacac atatataatat atatTTTTC ttttcttacc	240
agaaggTTT aatccaaATA aggagaAGAT atgcttagAA ccgaggTAGA gtttcatCC	300
attctgtcCT gtaagtatTTT tgcataTTCT ggagacgCAG gaagagatCC atctacatAT	360
ccccaaAGCTG aattatGGTA gacAAAACtC ttccactTTT agtgcataAA cttcttattt	420
gtgtataAG aaaaATGGGA aaacgatTTT caatATGCTT accaAGCTGT gattccAAAT	480
attacgtAAA tacactTGCA aaggaggatG ttttagtag caatttgtAC tGATGGTATG	540

ggccaagag atatatctta gagggaggc tgagggttg aagtccact cctaagccag	600
tgccagaaga gccaaggaca ggtacggctg tcatacacta gacccaccc tgtggagcca	660
caccctaggg ttggccaatc tactcccagg agcagggagg gcaggagcca gggctggca	720
taaaagttagtgc ggcagagcca tctattgctt acatttgctt ctgacacaac tgtttact	780
agcaaccta aacagacacc atgggcacatc tgactcctga ggagaagtct gccgttact	840
ccctgtgggg caaggtgaac gtggatgaag ttggatggta ggcctggc aggttggat	900
caaggttaca agacaggaaa aaggagacca atagaaactg ggcacatgtgaa gacagagaag	960
actcttgggt ttctgtatagg cactgactct ctctgcctat tggcttattt tcccaccctt	1020
aggctgtgg tggcttaccc ttggacccag aggttcttgg agtccttgg ggatctgtcc	1080
actcctgatg ctgttatggg caaccctaag gtgaaggctc atggcaagaa agtgcgtcggt	1140
gcctttatgt atggcctggc tcacctggac aacctcaagg gcacctttgc cacactgagt	1200
gagctgcact gtgacaagct gcacgtggat cctgagaact tcagggttag tctatggac	1260
gcttgcgtt ttctttcccc ttctttcta tggtaagtt catgtcatag gaagggata	1320
agtaacaggg tacagtttag aatggaaac agacgaatga ttgcacatgt gtggaaagtct	1380
caggatcggtt ttagttctt ttatgtctg ttccataacaa ttgtttctt ttgtttaatt	1440
cttgcgttctt ttttttctt tctccgaat ttttactatt atacttaatg ccttaacatt	1500
gtgtataaca aaaggaaata tctctgagat acattaagta actaaaaaaa aaactttaca	1560
cagtcgtgcct agtacattac tatttggat atatgtgtgc ttatttgcattt attcataatc	1620
tccctacttt atttctttt attttaattt gatacataat cattatacat atttatgggt	1680
taaagtgtaa tggtaataa tggtaacaca tattgaccaa atcagggtaa tttgcattt	1740
gtatattttaa aaaatgtttt cttctttaa tatactttt tggatattttt atttctaataa	1800
ctttccctaa tctctttctt tcaggcaat aatgatacaa tggatcatgc ctcttgcac	1860
cattctaaag aataacagtg ataatttctg ggttaaggca atagcaatata ctctgcata	1920
aaatatttctt gcatataaat tggtaactgat gtaagggtt tcatattgtt aatagcagct	1980
acaatccagc taccattctg cttttattttt atgggtggaa taaggctggaa ttatttgcatt	2040
tccaaagctgtt gccctttgc taatcatgtt catacctttt atcttgcctt cacagctt	2100
ggcaacgtt ctggctgtt tgctggccca tcactttggc aaagaattca cccaccaggat	2160
gcaggctgcc tatcagaag tggatggctgg tggatggtaat ggcctggccc acaagttca	2220
ctaagctcgc ttcttgcgtt tccaaatttctt attaaagggtt cttttgttcc ctaagtccaa	2280

ctactaaact ggggatatt atgaagggcc ttgagcatct ggattctgcc taataaaaaa	2340
--	------

catttattt cattgcaatg atgtatttaa attatttcg aatattttac taaaaaggga	2400
atgtgggagg tcagtgcatt taaaacataa agaaatgaag agctagttca aaccttggga	2460
aaatacacta tatcttaaac tccatgaaag aaggtgaggc tgcaaacagc taatgcacat	2520
tggcaacagc ccctgatgca tatgccttat tcacccctca gaaaaggatt caagtagagg	2580
cttgatttg aggttaagt ttgctatgc tgtatTTAC attacttatt gtttagctg	2640
tcctcatgaa tgtctttca ctacccattt gcttacccctg catctctcg ccttgactcc	2700
actcagttct ctigcttaga gataccacct ttcccctgaa gtgttccttc catgtttac	2760

ggcgagatgg ttctccctcg cctggccact cagccttagt tgtctctgtt gtcttataga	2820
ggctacttg aagaaggaaa aacaggggtc atggtttgc acgttgttgc gcccctttc	2880
cctgcctccc ccactcacag tgacccggaa tctgcagtgc tagtctcccg gaactatcac	2940
tcttcacag tctgctttgg aaggactggg ctttagtatga aaagtttagga ctgagaagaa	3000
tttgaaaaggc ggcttttgt agcttgatat tcactactgt ctattaccc tgtcataggc	3060
ccaccccaa tggaaagtccc attcttcctc aggatgtta agattagcat tcaggaagag	3120
atcagaggc tgcggctcc cttatcatgt ccctta	3156

&lt;210&gt; 33

&lt;211&gt; 3168

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; hHBB-hLW-013 vector, ITRs not included

&lt;400&gt; 33

ttattttattt gtatTTTGA ctgcattaag aggtctctag tttttatct cttgtttccc	60
aaaacctaattt aagtaactaa tgcacagagc acattgattt gtatTTATC tatttttaga	120
cataatttt tagcatgcat gagcaaattt agaaaaacaa caacaatga atgcataatat	180
atgtatatgt atgtgtgtat atatacacac atatataatat atatTTTC ttttcttacc	240
agaaggTTT aatccaaata aggagaagat atgcttagaa ccgaggtaga gttttcatcc	300

attctgtcct gtaagtattt tgcattttct ggagacgcag gaagagatcc atctacat	360
ccaaagctg aattatggta gacaaaactc ttccactttt agtgcattaa cttttatTTT	420
gtgtataaag aaaattggga aaacgatctt caatatgctt accaagctgt gattccaaat	480
attacgtaaa tacacttgca aaggaggatg ttttagtagt caatttgtac tggatgtatg	540
ggccaagag atatatcttta gagggaggc tgagggttg aagtccact cctaagccag	600

tgccagaaga gccaaggaca ggtacggctg tcatacactt gacctcaccc tgtggagcca	660
caccctaggg ttggccaatc tactcccagg agcagggagg gcaggagcca gggctggca	720
taaaagttag cgcagagcca tctattgctt acatttgctt ctgacacaac tgtttcact	780
agcaaccta aacagacacc atggccacc tcacaccaga agaaaaaaagt gctgtacag	840
ctctctgggg aaaagtcaat gtgcacgagg tagggggga agctctcgga aggttggtat	900
caaggttaca agacaggact agtacgtt ttaaggagac caatagaaac tggcatgt	960
gagacagaga agactcttgg gtttctgata ggcactgact ctctctgcct attggctat	1020
tttcccaccc ttagactcct cgtcgtgtat ccatggacac aaagatttt cgaaagcttc	1080
ggagacctca gcacaccaga cgcaatgt ggaaatccaa aagtcaaagc acacggaaaa	1140
aaggccctgg gggcttctc tgacggactc gcacatctcg ataatctgaa aggaacattc	1200
gctaccctct ctgaactcca ttgcataaa ctccatgtcg acccagaaaa ttttagatg	1260
agtctatgg acgcttgatg tttcttcc ccttcttgc tatggtaag ttcatgtcat	1320
aggaagggga taagtaacag ggtacagttt agaatggaa acagacgaat gattgcatca	1380
gtgtgaaatg ctcaggatcg ttttagtttgc ttatgttgc ttttataac aattgtttc	1440
ttttgtttaa ttcttgctt cttttttt cttctccgca attttacta ttatcttaa	1500
tgccttaaca ttgtgtataa caaaaggaaa tatctctgag atacatataa taacttaaaa	1560
aaaaacttta cacagtctgc ctgtacatt actatttggaa atatatgtgt gcttatttgc	1620
atattcataa tctccctact ttatttctt ttattttaa ttgatacata atcattatac	1680
atatttattgg gttaaagtgt aatgttttaa tatgtgtaca catattgacc aaatcagggt	1740
aattttgcat ttgtatattt aaaaaatgtt ttcttctttt aatatacttt tttgtttatc	1800
ttatttctaa tactttccct aatcttttc tttcaggcataa ataatgatac aatgtatcat	1860
gcctcttgc accattctaa agaataaacag tgataatttc tgggttaagg caatagcaat	1920
atctctgcat ataaatattt ctgtatataa attgtactg atgtaaaggagg tttcatatttgc	1980
ctaatacgat ctacaatcca gtcatttttattttt ttatgggtgg gataaggctg	2040
gattattctg agtccaaatc aggcctttt gctaatcatg ttcatacctc ttatcttctt	2100
cccacagctg ctggaaatg tcctcggtg cgtccctcgct caccattcg gaaaggagtt	2160
tacacctctt gtccaaaggcag cttaccaaaa ggtcgtcgca ggggtcgaa acgctctcg	2220
tcataaatac cattaggctc gttttcttgc tgcataattt ctattaaagg ttcccttgc	2280
ccctaagtcc aactactaaa ctggggata ttatgaagg ctttgagcat ctggattctg	2340

cctaataaaa aacatttattt ttcatggcaaa tgatgtattt aaatttttc tgaatatttt	2400
--	------

actaaaaagg gaatgtggga ggtcagtgcata tttaaaacat aaagaaatga agagctagtt	2460
---	------

caaacccttgg gaaaatacac tatacttaa actccatgaa agaaggtag gctgcaaaca	2520
--	------

gctaattgcac attggcaaca gccctgtatg catatgcctt attcatccct cagaaaagga	2580
--	------

ttcaagttaga ggcttgattt ggaggttaaa gtttgctat gctgtatttt acattactta	2640
---	------

ttgttttagc tgtcctcatg aatgtcttt cactaccat ttgcttatcc tgcatctc	2700
---	------

agccttgact ccactcagtt ctcttgctta gagataccac cttccctg aagtgttc	2760
---	------

tccatgtttt acggcgagat ggtttctctt cgccctggcca ctgcgcctta gttgtctcg	2820
---	------

ttgtcttata gaggtctact tgaagaagga aaaacagggg tcatggttt actgtcctgt	2880
--	------

gagcccttct tcctgcctc ccccactcac agtgaccggg aatctgcagt gctagtc	2940
---	------

cggaactatc actctttcac agtctgcttt ggaaggactg ggcttagtat gaaaagtt	3000
---	------

gactgagaag aatttgaag gcccgtttt gtagctgtat attcactact gtcttattac	3060
---	------

cctgtcatag gcccacccca aatggaagtc ccattttcc tcaggatgtt taagattgc	3120
---	------

attcaggaag agatcagagg tctgctggct cccttatcat gtccctta	3168
--	------

&lt;210&gt; 34

&lt;211&gt; 3156

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; hHBB-hLW-013 vector, ITRs not included, Linker removed

&lt;400&gt; 34

ttatTTTTttaa gtatTTTGA ctgcattaag aggtctctag tttttatct ctgtttccc	60
--	----

aaaacctaataa aagtaactaa tgcacagagc acattgattt gtatTTTTC tattttaga	120
---	-----

cataatttat tagcatgcat gagcaaattt agaaaaacaa caacaaatga atgcataat	180
--	-----

atgtatATGT atgtgtgtat atatacacac atatatatat atatTTTTC ttttcttacc	240
--	-----

agaaggTTTT aatccaaata aggagaagat atgcTTAGAA ccgaggtaga gttttcatcc	300
---	-----

attctgtcct gtaagtattt tgcattttt ctgcattttt ggagacgcag gaagagatcc atctacat	360
---	-----

ccccaaagctg aattatggta gacaaaactc ttccactttt agtgcataa cttttatTTT	420
---	-----

gtgttaataag aaaattggaa aacgatctt caaatatgctt accaagctgt gattccaaat	480
--	-----

attacgtaaa tacacttgca aaggaggatg ttttagtag caatttgtac ttaggtatg	540
---	-----

gggccaagag atatatctt aagggaggcc tgagggttg aagtccaaat cctaagccag	600
---	-----

tgcagaaga gccaaggaca ggtacggctg tcatcactta gacctcaccc tgtggagcca	660
--	-----

caccctaggg ttggccaatc tactcccagg agcagggagg gcaggagcca gggctggca	720
taaaagttag ggcagagcca tctattgctt acatttgctt ctgacacaac tgtgttact	780
agcaaccta aacagacacc atggccacc tcacaccaga agaaaaaaagt gctgtacag	840
ctctctgggg aaaagtcaat gtgcacgagg taggggggaa agctctcgga aggttgttat	900
caaggttaca agacaggtt aaggagacca atagaaactg ggcacgtgga gacagagaag	960
actcttgggt ttctgtatgg cactgactct ctctgcctat tggcttattt tcccaccctt	1020
agactcccg tcgtgtatcc atggacacaa agattttcg aaagcttcgg agacccacg	1080
acaccagacg cagtaatggg aaatccaaaa gtcaaagcac acggaaaaaa ggtcctgggg	1140
gccttctcg acggactcgc acatctcgat aatctgaaag gaacatcgc taccctct	1200
gaactccatt gcgataaaact ccatgtcgac ccagaaaatt ttagagttag tctatggac	1260
gcgttgcgtt ttctttcccc ttctttctt tggtaagtt catgtcatag gaagggata	1320
agtaacaggg tacagtttag aatggaaac agacgaatga ttgcacatgt gtggaaagtct	1380
caggatcggtt ttatgttctt ttatgtctt ttcataacaa ttgtttctt ttgtttaatt	1440
tttgctttct ttttttct tctccgcaat ttacttattt atacttaatg ccttaacatt	1500
gtgtataaca aaaggaaata tctctgagat acattaagta actaaaaaaa aaactttaca	1560
cagtcgtcct agtacattttc tatttggat atatgtgtgc ttatgtcat attcataatc	1620
tccctacttt atttctttt attttaattt gatacataat cattatacat atttatgggt	1680
taaagtgtaa tggtaataa tggtaacaca tattgaccaa atcaggtaa tttgcattt	1740
gtaatttaa aaaaatgtttt cttctttaa tatactttt tggttatctt atttctaata	1800
ctttccctaa tctctttctt tcaggcaat aatgatacaa tggatcatgc ctcttcac	1860
cattctaaag aataacagtg ataatttctg ggttaaggca atagcaatat ctctgcata	1920
aaatatttctt gcatataaat tggtaactgat gtaagagggtt tcatattgtt aatagcagct	1980
acaatcccagg taccattctg cttttattttt atgggtggaa taaggctgga ttattctgag	2040
tccaaatgtt gccctttgc taatcatgtt catacctttt atttcttcc cacagctgt	2100
cgaaatgtc ctcgtgtcgc tcctcgctca ccatttcgga aaggagttt cacctctgt	2160
ccaaggcact taccaaaagg tcgtcgagg ggtcgaaac gctctcgctc ataaatcca	2220
tttaggctcgc ttcttgctg tccaaatttctt attaaagggtt cttttgttcc ctaagtccaa	2280
ctactaaact gggggatattt atgaagggcc ttgagcatct ggattctgcc taataaaaaaa	2340
catttattttt cattgcaatg atgtatattaa attatttctg aatattttac taaaaaggaa	2400

atgtgggagg tcagtgcatt taaaacataa agaaatgaag agctagttca aaccttggga 2460

aaatacacta tatcttaaac tccatgaaag aaggtgaggc tgcaaacagc taatgcacat 2520  
 tggcaacagc cccigtatgc tatgccttat tcatccctca gaaaaggatt caagtagagg 2580  
 cttgatttg aggttaagt ttgctatgc tgtatttac attacttatt gtttagctg 2640  
 tcctcatgaa tgtctttca ctaccattt gcttacccctg catctctcag ccttactcc 2700  
 actcagttct ctgcttaga gataccacct ttcccctgaa gtgtcccttc catgtttac 2760  
 ggcgagatgg ttctcctcg cctggccact cagccttagt tgtctctgtt gtcttataga 2820  
 ggtctacttg aagaaggaaa aacaggggtc atggttgac tgtcctgtga gcccttcttc 2880

cctgcctccc ccactcacag tgacccggaa tctgcagtgc tagtctcccg gaactatcac 2940  
 tcttcacag tctgctttgg aaggactggg ctttagtatga aaagtttagga ctgagaagaa 3000  
 tttgaaaggc ggcttttgt agcttgatat tcactactgt ctattaccc tgtcataggc 3060  
 ccaccccaa tggaaagtccc attcttcctc aggatgtta agattagcat tcaggaagag 3120  
 atcagaggc tgcggctcc cttatcatgt cccta 3156

<210> 35

<211> 2468

<212> DNA

<213> Artificial Sequence

<220><223> hHBB-hL-011 vector, ITRs not included

<400> 35

ttattttattt gtattttga ctgcattaag aggtctctag ttttttatct cttgtttccc 60  
 aaaacctaatt aagtaactaa tgcacagagg acattgattt gtatttattt tatttttaga 120  
 cataatttt tagcatgcat gagcaaattt agaaaaacaa caacaatga atgcataat 180  
 atgtatatgt atgtgttat atatacacac atatataat atatttttc ttttcttacc 240  
 agaaggtttt aatccaaata aggagaagat atgcttagaa ccgaggtaga gttttcatcc 300  
 attctgtcct gtaagtattt tgcatattt ggagacgcag gaagagatcc atctacat 360  
 cccaaagctg aattatggta gacaaaactc ttccactttt agtgcataa cttcttattt 420

gtgttaataag aaaattggga aaacgatctt caaatgcattt accaagctgt gattccaaat 480  
 attacgtaaa tacacttgca aaggaggatg ttttttagtag caatttgtac tgatggatg 540  
 gggccaagag atatatctt gaggggggc tgagggttg aagtccaaact cctaagccag 600  
 tgccagaaga gccaaggaca ggtacggctg tcatacactt gacccaccc tgtggagcca 660  
 caccctaggg ttggccaatc tactcccagg agcaggagg gcaggagcca gggctggca 720

taaaagttag ggcagagcca tctattgctt acatttgctt ctgacacaac tgtgttact	780
agcaacctca aacagacacc atggtgcatc tgactcctga ggagaagtct gccgttactg	840
ccctgtgggg caaggtgaac gtggatgaag ttgggttgta ggccctggc aggttgttat	900
caaggttaca agacaggact agtatcgatt ttaaggagac caatagaaac tggcatgtg	960
gagacagaga agactttgg gtttctgata ggcactgact ctctctgcct attggctat	1020
tttcccaccc ttaggctgct ggtggtctac ccttggaccc agaggttctt tgagtccctt	1080
ggggatctgt ccactcctga tgctgtttag ggcaacccta aggtgaaggc tcatggcaag	1140
aaagtgcctg gtgcctttag ttagtgcctg gtcacccctt acaaccccaa gggcacccctt	1200
gccacactga gtgagctgca ctgtgacaag ctgcacgtgg atcctgagaa cttagggtg	1260
agtctatggg acgcttgatg tttctttcc ctttctttc tatggtaag ttcatgtcat	1320
aggaagggga taagtaacag ggtacagttt agaatggaa acagacgaat gattgcatca	1380
gtgttggaaat ctctggatcg ttttagtttcc ttatgttgc tggtcataac aattgttttcc	1440
ttttgtttaa ttcttgctt cttttttttt ctctccgca attttacta ttatacttaa	1500
tgccttaaca ttgtgtataa caaaaaggaaa tatctctgag atacattaag taacttaaaa	1560
aaaaacttta cacagtctgc ctagtacatt actatttggaa atatatgtgt gcttatttgc	1620
atattcataa tctccctact ttattttctt ttattttaa ttgatacata atcattatac	1680
atattttatgg gttaaagtgt aatgttttaa tatgtgtaca catattgacc aaatcagggt	1740
aattttgcattt ttgttaatttt aaaaaatgtt ttcttctttt aatatactttt tttgtttatc	1800
ttatattctaa tactttccctt aatcttttc ttccaggca ataatgatac aatgtatcat	1860
gcctcttgc accattctaa agaataacag tgataatttc tgggttaagg caatagcaat	1920
atctctgcat ataaatattt ctgcatataa attgtactg atgttaagg tttcatatttgc	1980
ctaatacgat ctacaatcca gtcaccattc tgctttattt ttatgtttgg gataaggctg	2040
gattattctg agtccaaatgtt ggtccatgtt gtcataatcatg ttatcttctt	2100
ccccacagctc ctgggcaacg tgctggctg tgggtggcc catcaatttgc gcaaaaggatt	2160
caccccaacca gtgcaggctg cctatcagaa agtgggtggct ggtgtggcta atgcctggc	2220
ccacaaggat cactaagctc gctttcttgc tgggtggccat ttatatttgc ttatatttgc	2280
ccctaagtcc aactactaaa ctggggata ttatgtttgg ctttggatcat ctggattctg	2340
cctaataaaaa aacatttttattt ttatgtttgg ctttggatcat ctggattctg	2400
actaaaaagg gaatgtggaa ggtcagtgcata ttatgtttgg ctttggatcat ctggattctg	2460

caaacctt	2468
<210> 36	
<211> 2456	
<212> DNA	
<213> Artificial Sequence	
<220><223> hHBB-hL-011 vector, ITRs not included, Linker removed	
<400> 36	
ttattttattt gtatTTTga ctgcattaag aggtctctag ttttttatct cttgtttccc	60
aaaacctaattt aagttaactaa tgacacagagc acattgattt gtatTTTattc tatttttaga	120
cataattttat tagcatgcat gagcaaattt agaaaaacaa caacaaatga atgcataatat	180
atgtatatagt atgtgtgtat atatacacac atatataatat atatTTTtc ttttcttacc	240
agaaggTTTT aatccaaata aggagaagat atgcttagaa ccgaggtaga gtttcatcc	300
attctgtcct gtaagtattt tgcattattt ggagacgcag gaagagatcc atctacatat	360
cccaaaagctg aattatggta gacaaaactc ttccactttt agtgcatttcaa ctttttattt	420
gtgttaataag aaaattggga aaacgatctt caatatgctt accaagctgt gattccaaat	480
attacgtaaa tacacttgca aaggaggatg ttttagtag caatttgtac tggatgtatg	540
gggccaagag atatactta gaggggaggc tgagggtttg aagtccaaact cctaagccag	600
tgccagaaga gccaggaca ggtacggctg tcattacttta gacccaccc tggatggcca	660
caccctaggg ttggccaatc tactcccagg agcaggagg gcaggagcca gggctggca	720
taaaagtctggc ggcagagcca tctattgctt acatttgctt ctgacacaac tgtgttca	780
agcaacctca aacagacacc atggtgcatc tgactcctga ggagaagtct gcccgttactg	840
ccctgtgggg caaggtgaac gtggatgaag ttgggtgtga ggcctggc aggttggat	900
caagggttaca agacaggTTTT aaggagacca atagaaactg ggcattgttgc gacagagaag	960
actcttgggt ttctgtatagg cactgactct ctctgcctat tggcttattt tcccaccctt	1020
aggctgctgg tggcttaccc ttggacccag aggttcttgc agtgccttgg ggtatgttcc	1080
actcctgatg ctgttatggg caaccctaag gtgaaggctc atggcaagaa agtgcgttgt	1140
gcctttagtg atggcctggc tcacctggac aacctcaagg gcaccttgc cacactgagt	1200
gagctgcact gtgacaagct gcacgtggat cctgagaact tcagggtgag tctatggac	1260
gctttagtgc ttctttcccc ttctttctat tggtaagtt catgtcatag gaaggggata	1320
agtaacaggg tacagtttag aatgggaaac agacgaatga ttgcattcgt gtggaaagtct	1380
caggatcgat ttagtttctt ttatTTGCTG ttcataacaa ttgttttctt ttgtttaatt	1440

cttgcttct tttttttct tctccgcaat ttttactatt atacttaatg ccttaacatt	1500
gtgtataaca aaaggaaata tctctgagat acattaagta actaaaaaaa aaactttaca	1560
cagctctgcct agtacattac tatttggaat atatgtgtc ttatttgcattt attcataatc	1620
tccctacttt attttctttt attttaattt gatacataat cattatacat atttatgggt	1680
taaagtgtaa tggttaata tgtgtacaca tattgaccaa atcaggtaa ttgcattt	1740
gtaatttaa aaaatgctt cttctttaa tatactttt tggttatctt atttctaata	1800
ctttccctaa tctcttctt tcagggcaat aatgatacaa tgtatcatgc ctcttgcac	1860
cattctaaag aataaacagtg ataattctg ggtaaggca atagcaatat ctctgcataat	1920
aaatatttct gcataataat tgtaacttgt gtaagaggtt tcatattgtc aatagcagct	1980
acaatccagc taccattctg cttttatttt atgggtggga taaggctgga ttattctgag	2040
tccaagctag gccctttgc taatcatgtt cataacctt atcttcctcc cacagctcct	2100
ggccaacgtg ctggctgtg tgctggccca tcactttggc aaagaattca ccccaccagt	2160
gcaggctgcc tatcagaaag tggtggtgg tggtggtaat gccctggccc acaagttatca	2220
ctaagctcg tttcttgctg tccaatttctt attaaagggtt cttttttcc ctaagtccaa	2280
ctactaaact gggggatattt atgaagggcc ttgagcatct ggattctgcc taataaaaaaa	2340
catttatttt cattgcaatg atgtatttaa attatttctg aatattttac taaaaaggga	2400
atgtgggagg tcagtgcatt taaaacataa agaaatgaag agcttagttca aacctt	2456
<210> 37	
<211> 2468	
<212> DNA	
<213> Artificial Sequence	
<220><223> hHBB-hLW-012 vector, ITRs not included	
<400> 37	
ttatTTTTt gtatTTTGA ctgcattaaAG aggtctctAG ttttttatCT cttgtttccc	60
aaaacctaattt aagtaactaa tgacacagac acattgattt gtatTTTTC tatttttagA	120
cataatttat tagcatgcatt gagcaaattt agaaaaacAA caacaaatGA atgcataatAT	180
atgtatATGT atgtgtgtat atatacacac atatataatAT atTTTTTC ttttcttacc	240
agaaggTTTTt aatccaaATA aggagaAGAT atgcttagAA ccgaggTAGA gttttcatCC	300
attctgtcct gtaagtattt tgcatattct ggagacgcAG gaagagatCC atctacatAT	360
ccccaaAGCTG aattatggTA gacAAAactC ttccactttt agtgcataAA ctTCTTATTt	420
gtgtataAGG AAAATTGGGA aaacgatCTT caatatgCTT accaagCTGT gattccAAat	480

attacgtaaa tacacttgca aaggaggatg ttttagtag caattgtac tgatggatg	540
ggccaagag atatatctt gaggaggc tgagggttg aagtccact cctaagccag	600
tgccagaaga gccaaggaca ggtacggctg tcatactta gacctcaccc tgtggagcca	660
caccctaggg ttggccaatc tactcccagg agcagggagg gcaggagcca gggctggca	720
taaaagttag ggcagagcca tctattgctt acatttgctt ctgacacaac tgtttcact	780
agcaaccta aacagacacc atggtccacc tcacaccaga agaaaaaaagt gctgtacag	840
ctctctgggg aaaagtcaat gtgcacgagg tagggggga agctctcgga aggttggat	900
caaggttaca agacaggact agtatcgatt ttaaggagac caatagaaac tggcatgtg	960
gagacagaga agactttgg gtttctgata ggcactgact ctctctgcctt attggctat	1020
tttccccccctt ttagactctt cgtcgtgtat ccatggacac aaagatttt cgaaagcttc	1080
ggagacctca gcacaccaga cgcatgtatggaaatccaa aagtcaaagc acacggaaaa	1140
aaggccctgg gggcttctc tgacggactc gcacatctcg ataatctgaa aggaacattc	1200
gctaccctct ctgaaactcca ttgcataaa ctccatgtcg acccagaaaa tttagatgt	1260
agtctatgg acgcttgatg tttcttcc cttttttc tatggtaag ttcatgtcat	1320
aggaagggga taagtaacag ggtacagttt agaatggaa acagacgaat gattgcatca	1380
gtgtggaagt cttagatcg ttttagtttgc ttatgttgc ttatgttgc aattttttc	1440
ttttgtttaa ttcttgctt ctttttttt ctctccgca attttacta ttatacttaa	1500
tgccttaaca ttgtgtataa caaaaggaaa tatctctgag atacattaag taacttaaaa	1560
aaaaacttta cacagtctgc ctagtacatt actatttggaa atatatgtgt gcttatttgc	1620
atattcataa tctccctact ttatttctt ttattttaa ttgatacata atcattatac	1680
atatttatgg gtaaagtgt aatgtttaa tatgtgtaca catattgacc aaatcagggt	1740
aattttgcat ttgtatttt aaaaaatgtt ttcttctttt aatatacttt tttgtttatc	1800
ttatcttctaa tactttccctt aatcttttc ttccaggcataatgatac aatgtatcat	1860
gcctcttgc accattctaa agaataacag tgataatttc tgggttaagg caatagcaat	1920
atctctgcat ataaatattt ctgcatataa attgtactg atgtaaaggagg tttcatatttgc	1980
ctaatacgat ctaaatcca gctaccattc tgcttttattt ttatgggtgg gataaggctg	2040
gattattctg agtccaaatgtt agggccctttt gctaattcatg ttccatctc ttatcttcttgc	2100
cccaacatgtc tcggaaatgtt ccctcgatg cgtccctcgat caccatttgc gaaaggagtt	2160
tacacccctt gtccaaaggcag cttacaaaaa ggtcgatcgaa ggggtcgca acgctctcg	2220
tcataaatac cattaggctc gcttttttgc tggcaatttt ctatataagg ttcccttgc	2280

ccctaagtcc aactactaaa ctggggata ttatgaaggg cttgagcat ctggattctg	2340
cctaataaaa aacatttatt ttcattgcaa tgatgtattt aaattatttc tgaatatttt	2400
actaaaaagg gaatgtggga ggtcagtgcata ttaaaacat aaagaatga agagcttgtt	2460
caaacctt	2468
<210> 38	
<211> 2456	
<212> DNA	
<213> Artificial Sequence	
<220><223> hB2B-hLW-012 vector, ITRs not included, Linker removed	
<400> 38	
ttatttattt gtatTTGA ctgcattaag aggtctctag ttttatct ctgtttccc	60
aaaacctaata aagtaactaa tgcacagagc acattgattt gtatTTATC tattttaga	120
cataatttat tagcatgcat gagcaaatta agaaaaacaa caacaatga atgcatatat	180
atgtatATGT atgtgtgtat atatacacac atatatatat atatTTTC ttttctacc	240
agaaggTTT aatccaaata aggagaagat atgcttagaa ccgaggtaga gtttcatcc	300
attctgtcct gtaagtattt tgcattattt ggagacgcag gaagagatcc atctacatat	360
cccaaAGCTG aattatggta gacaaaACTC ttccactttt agtgcataa cttttatTT	420
gtgtAATAAG aaaATTGGGA aaACGATCTT caatATGCTT accaAGCTGT gattccAAAT	480
attacgtaaa tacacttgca aaggaggatg ttttagtag caatttgac tGATGGTATG	540
ggccaagag atatactta gagggaggc tgagggttg aagtccact cctaagccag	600
tgcagaaga gccaaggaca ggtacggctg tcattactta gacccaccc tgtggagcca	660
caccctaggg ttggccaatc tactcccagg agcaggagg gcaggagCCA gggctggca	720
taaaAGTCAG ggcagagCCA tctattgctt acatttgctt ctgacacaac tgtttcact	780
agcaacCTCA aacagacacc atggccacc tcacaccaga agaaaaaaAGT gctgtAACAG	840
ctctctgggg aaaAGTCAT gtcgacgagg taggggggAgCtctcgGA aggttggat	900
caaggTTACA agacaggTTT aaggagacca atagaaactg ggcatttgGA gacagagaAG	960
actcttgggt ttctgatagg cactgactct ctctgcctat tggcttattt tccaccctt	1020
agactcctcg tctgttatcc atggcacaaa agatTTTCG aaagctcgg agacccAGC	1080
acaccagacg cagtaatggg aaatccaaaa gtcaaAGCAC acggaaaaaa ggtcctgggg	1140
gctttctctg acggactcgc acatctcgat aatctgaaag gaacattcgc taccctct	1200
gaactccatt gcgataaaact ccatgtcgac ccagaaaatt ttagagttag tctatggac	1260

gcttgatgtt ttctttcccc ttctttcta tggtaagtt catgtcatag gaagggata	1320
agtaacagg tacagtttag aatggaaac agacgaatga ttgcatcagt gtgaaagtct	1380
caggatcgtt ttagttctt ttatttgctg ttcataacaa ttgtttctt ttgtttaatt	1440
cttgcttct tttttttct tctccgcaat ttttactatt atacttaatg ccttaacatt	1500
gtgtataaca aaaggaaata tctctgagat acattaagta actaaaaaaa aaacttaca	1560
cagtcgcct agtacattac tatttgaat atatgtgtc ttatttgcattt attcataatc	1620
tccctacttt atttctttt attttaatt gatacataat cattatacat atttatgggt	1680
taaagtgtaa tggttaata tgtgtacaca tattgaccaa atcaggtaa ttgcattt	1740
gtaatttaa aaaaatgcctt ctcttttaa tatactttt tgttatctt atttctaata	1800
ctttccctaa tctcttctt tcagggcaat aatgatacaa tgtatcatgc ctcttgcac	1860
cattctaaag aataacagtg ataattctg ggtaaggca atagcaatat ctctgcataat	1920
aaatatttct gcatataaat tgtaactgat gtaagaggtt tcatattgtct aatagcagct	1980
acaatccagc taccattctg cttttattt atgggtggta taaggctgga ttattctgag	2040
tccaagctag gccctttgc taatcatgtt catacctttt atcttcctcc cacagctgt	2100
cggaaatgtc ctgcgtgcg tcctcgctca ccatttcgga aaggagttt cacctctgt	2160
ccaagcagct taccaaaagg tcgtcgagg ggtcgcaaac gctctcgctc ataaatacca	2220
ttaggctcgc ttcttgctg tccaatttct attaaagggtt cttttgttcc ctaagtccaa	2280
ctactaaact ggggatattt atgaagggcc ttgagcatct ggattctgcc taataaaaaaa	2340
catttatttt cattgcaatg atgtatTTTaa attatttctg aatattttac taaaaaggga	2400
atgtgggagg tcagtgcatt taaaacataa agaaatgaag agctagttca aacctt	2456
<210> 39	
<211> 2321	
<212> DNA	
<213> Artificial Sequence	
<220><223> hHBB-hA-009 vector, ITRs not included	
<400> 39	
ttattttattt gtatTTTGA ctgcattaaAG aggtctctAG tttttatCT ctgtttccc	60
aaaacctaattt aagtaactaa tgcacagagc acattgattt gtattttttc tattttttaga	120
cataattttat tagcatgcat gagcaaatttta agaaaaacaa caacaaatga atgcataat	180
atgtatatgt atgtgtgtat atatacacac atatataat atatTTTTC ttTTTCTTAC	240
agaaggTTTTT aatccaaata aggagaagat atgcttagaa ccgaggtaga gttttcatcc	300

attctgtcct gtaagtattt tgcattttct ggagacgcag gaagagatcc atctacatat	360
cccaaagctg aattatggta gacaaaactc ttccactttt agtgcataa ctttttattt	420
gtgtataaag aaaattggaa aaacgatctt caatatgctt accaagctgt gattccaaat	480
attacgtaaa tacacttgca aaggaggatg ttttagtag caatttgtac tgatggatg	540
ggccaagag atatatctt gagggaggc tgagggttg aagtccact cctaagccag	600
tgcagaaga gccaaggaca ggtacggctg tcatactta gacctcaccc tgtggagcca	660
caccctaggg ttggccatc tactcccagg agcaggagg gcaggagcca gggctggca	720
taaaagttag cgcagagcca tctattgctt acatttgctt ctgacacaac tgtttcact	780
agcaacctca aacagacacc atggtgcatc tgactcctga ggagaagtc gcccattactg	840
ccctgtgggg caaggtgaac gtggatgaag ttgggtgtga ggccctggc aggctgtgg	900
tggctaccc ttggacccag aggttcttg agtccttgg ggatctgtcc actcctgatg	960
ctgttatggg caaccctaag gtgaaggctc atggcaagaa agtgcctcggt gcctttatgt	1020
atggccctggc tcacctggac aacctaagg gcaccttgc cacactgagt gagctgcact	1080
gtgacaagct gcacgtggat cctgagaact tcaggtctt gggcaacgtg ctggctgtg	1140
tgctggccca tcaatttggc aaagaattca cccaccagt gcaggctgcc tatcagaaag	1200
tggggctgg tggctaat gccctggccc acaagtatca ctaagaattc aaggccttc	1260
gagcctctag aactatagtg agtcgtatta cgtagatcca gacatgataa gatacattga	1320
tgagtttggaa caaacacaa ctagaatgca gtgaaaaaaa tgcttttattt gtgaaatttg	1380
tgtatgtatt gcttattttt taaccattat aagctgcaat aaacaagtt acaacaacaa	1440
ttgcattcat ttatgtttc aggttcaggg ggaggtgtgg gaggttttt aagttcagc	1500
tgtgatcata acgttgagct cgtgcatactg actcctgagg agaagtcgtc cgttactgcc	1560
ctgtggggca aggtgaacgt ggtgaagtt ggtggtgagg ccctggcag gttggatca	1620
aggttacaag acaggttaa ggagaccaat agaaactggg catgtggaga cagagaagac	1680
tcttgggttt ctgataggca ctgactctt ctgcctattt gtctattttc ccacccttag	1740
gctgctggtg gtctaccctt ggacccagag gttctttagt tcctttgggg atctgtccac	1800
tcctgatgct gtatggca accctaaggta gaaggctcat ggcaagaaag tgctcggtgc	1860
cttttagtcat ggcctggctc acctggacaa cctcaaggcc acctttgccca cactgagtga	1920
gctgcactgt gacaagctgc acgtggatcc tgagaacttc agggtgagtc tatggacgc	1980
ttgatgtttt ctccccctt ctttctatg gttaagttca tgtcatagga agggataag	2040
taacaggta cagtttagaa tggaaacag acgaatgatt gcatcagtgt ggaagtctca	2100

ggatcgttt agtttctttt atttgctgtt cataacaatt gttttctttt gttaattct	2160
tgcttcctt tttttcttc tcgcattt ttactattat acttaatgcc ttaacattgt	2220
gtataacaaa aggaaatatac tctgagatac attaagtaac taaaaaaa actttacaca	2280
gtctgcctag tacattacta ttggaatat atgtgtgctt a	2321
<210> 40	
<211> 2321	
<212> DNA	
<213> Artificial Sequence	
<220><223> hB2B-hAW-002 vector, ITRs not included	
<400> 40	
ttattttattt gtatTTTga ctgcattaaag aggtctctag tttttatct cttgtttccc	60
aaaacctaattt aagtaactaa tgacacagac acattgattt gtatttattt tattttaga	120
cataatttat tagcatgcattt gagcaatttta agaaaaacaa caacaatgtt atgcataat	180
atgtatatgtt atgtgtgtt atatacacac atatataat atatTTTtc ttttctacc	240
agaaggTTTTt aatccaaata aggagaatgtt atgcTTtagaa ccgaggtaga gtttcatcc	300
attctgtcctt gtaagtattt tgcattattt ggagacgcag gaagagatcc atctacat	360
cccaaagctg aattatggta gacaaaactc ttccactttt agtgcattaa cttttat	420
gtgttaataag aaaattggtaa aacgatctt caatatgtt accaagctgtt gattccaaat	480
attacgtaaa tacacttgca aaggaggatg ttttagtagt caatttgtac tggatgtat	540
ggccaagag atatactta gaggggggc tgagggttg aagtccactt cctaagccag	600
tgcagaaga gccaaggaca ggtacggctg tcattactta gacctcaccc tgtggagcca	660
cacccttaggg ttggccaatc tactcccagg agcaggagg gcaggagcca gggctggca	720
taaaagtctt ggcagagcca tctattgtt acatttgctt ctgcacacaac tgttttact	780
agcaacctca aacagacacc atggccacc tcacaccaga agaaaaaaagt gctgtaacag	840
ctctctgggg aaaagtcaat gtcgacgagg taggggggaa agctctcgga agactcctcg	900
tcgtgtatcc atggacacaa agatTTTcg aaagcttcgg agacctcagc acaccagacg	960
cagtaatggg aaatccaaaa gtcacaaagcac acggaaaaaa ggtcctgggg gctttctctg	1020
acggactcgc acatctcgat aatctgaaag gaacattcgc taccctctt gaactccatt	1080
gcgataaaactt ccatgtcgac ccagaaaattt ttagactgctt cgaaatgtc ctcgtgtgct	1140
tcctcgatca ccatttcggaa aaggagttt cacccctgtt ccaagcagctt accaaaagg	1200
tcgtcgagg ggtcgaaac gctctcgctc ataaatacca ttaggaattt aaggcctctc	1260

gagcctctag aactatagt agtcgttata cgtagatcca gacatgataa gatacattga	1320
tgagtttggaa caaaccacaa ctagaatgca gtaaaaaaa tgcttattt gtgaaatttgc	1380
tgtatgttattttaaccattttaaagctgcaat aaacaaggta acaacaacaa	1440
ttgcattcat ttatgtttc aggttcaggg ggagggttgg gaggtttttt aagcttcagc	1500
tgtgtatata acgttgacgt cgtcatcg actcctgagg agaagtcgtc cgttactgcc	1560
ctgtggggca aggtgaacgt ggtgaagtt ggtgggttgg ccctggcag gttggatca	1620
aggttacaag acaggtaa ggagaccaat agaaactggg catgtggaga cagagaagac	1680
tcttgggtt ctgataggca ctgactctct ctgcctatttgcgtatatttccacccttag	1740
gctgcgttgt gctaccctt ggacccagag gttcttttag tccttgggg atctgtccac	1800
tcctgtatgtcttatggca accctaaggtaaaggctcat ggcaagaaag tgctcggtgc	1860
cttttagtcat ggcctggctc acctggacaa cctcaaggcc acctttgcca cactgagtga	1920
gctgcactgt gacaagctgc acgtggatcc tgagaacttc agggtgagtc tatggacgc	1980
ttgtatgtttt ctcccctt ctttctatg gttaagttca tgtcatagga agggataag	2040
taacaggta cagtttagaa tggaaacag acgaatgatt gcatcgtgt ggaagtctca	2100
ggatcgccccccatttttttt atttgctgtt cataacaatt gttttttttt gtttaattct	2160
tgcttttttttttttttc tccgaatttttactattat acttaatgcc ttaacattgt	2220
gtataacaaa aggaatatac tctgagatatac attaagtaac ttaaaaaaaaaa actttacaca	2280
gtctgccttag tacattacta ttggaaatatacgtgtgc ttatgtgc	2321
<210> 41	
<211> 2439	
<212> DNA	
<213> Artificial Sequence	
<220><223> hHBB-h1-010 vector, ITRs not included	
<400> 41	
caacaacaaa tgaatgcata tatatgtata tgtatgtgt tatatacata cacatataata	60
tatataattttt ttcttttttcc accagaaggat ttaatccaa ataaggagaa gatatgttta	120
gaaccggaggat agagtttca tccattctgt cctgtatgttgcataat tctggagacgt	180
caggaagaga tccatctaca tatccaaag ctgaattatgt gtagacaaaa ctctccact	240
tttagtgcataacttctta ttgtgtat aagaaaatttgg gaaaaacgtat cttcaatatgt	300
cttaccaacgt tttgttccaa atattacgt aaatacattt gcaaaggagg atgttttttag	360
tagcaatttgcattt tactgtatgtt atggggccaa gagatatac ttagagggag ggctgagggt	420

ttgaagtcca actcctaagc cagtgcaga agagccaagg acaggtacgg ctgtcatcac	480
ttagaccta ccctgtggag ccacacccta gggttggca atctactccc aggagcaggg	540
agggcaggag ccagggctgg gcataaaaagt cagggcagag ccatctattg cttacatttgc	600
cttctgacac aactgtgttc actagcaacc tcaaacagac accatggtgc atctgactcc	660
tgaggagaag tctgccgtta ctgccctgtg gggcaaggtg aacgtggatg aagttggtgg	720
tgaggccctg ggcagggttgg tatcaaggtt acaagacagg tttaaaggaga ccaatagaaa	780
ctggccatgt ggagacagag aagcttctga cctcttctct tcctccacaca gggcggtacc	840
agatctggca gcggagaggg cagaggaagt cttctaacat gcggtgacgt ggaggagaat	900
cccgcccta ggggtaccat ggtgcacatg actccitgagg agaagtcgtc cgtaactgcc	960
ctgtggggca aggtgaacgt ggtgaagtt ggtggtgagg ccctggcag gctgctggtg	1020
gtctaccctt ggacctcagag gttcttttag tcctttgggg atctgtccac tcctgatgt	1080
gttatggca accctaaggt gaaggctcat ggcaagaag tgctcggtgc cttagtgat	1140
ggcctggctc acctgtacaa cctcaaggcc accttgcacactgactgt	1200
gacaagtc acgtggatcc tgagaacttc aggctctgg gcaacgtgct ggtctgtgt	1260
ctggccatc actttggcaa agaattcacc ccaccagtgc aggctgccta tcagaaagtgc	1320
gtggctggtg tgctaatgc cctggccac aagtatcaact aagaattcaaa ggcctctcgaa	1380
gcctctagaa ctatagttagtgc tgcttattacg tagatccaga catgataaga tacattgt	1440
agtttggaca aaccacaact agaatgcagt gaaaaaaatg ctttatttgc gaaatttgc	1500
atgttatgc ttatgttgc accattataa gctgaataa acaagttAAC aacaacaatt	1560
gcattcattt tatgtttcag gttcaggggg aggtgtggaa ggtttttaa gtttcagctg	1620
tgatcataac gttgagctca agactcttgg gtttctgata ggcactgact ctctctgcct	1680
attggcttat tttccaccc ttaggctgtc ggtggcttac cttggaccc agaggttctt	1740
tgagtcctt gggatctgt ccactcttgc tgctgttatg ggcaacccta aggtgaaggc	1800
tcatggcaag aaagtgcgtc gtgcctttag tgatggcctg gtcacactgg acaaccta	1860
ggcacctt gccacactga gtgagctgca ctgtgacaag ctgcacgtgg atcctgagaa	1920
ttcagggtg agtctatggg acgcttgcgt ttttcttcc cttcttttc tatgtttag	1980
ttcatgtcat aggaaggggta taagtaacag ggtacagttt agaatggaa acagacgaat	2040
gattgcatca gtgtggaaatg ctcaggatcg ttttagtttc ttttatttgc tggtcataac	2100
aattgttttc ttgttttaa ttcttgctt cttttttt cttctccgca atttttacta	2160
ttatacttaa tgccataaca ttgtgtataa caaaaggaaa tatctctgag atacattaag	2220

taactaaaa aaaaacttta cacagtctgc ctagtacatt actatttgga atatatgtgt	2280
gcttatttgc atattcataa tctccctact ttatttctt ttattttaa ttgatacata	2340
atcattatac atatttatgg gttaaagtgt aatgtttaa tatgtgtaca catattgacc	2400
aaatcagggt aatttgcat ttgtattt aaaaaatgc	2439
<210> 42	
<211> 2439	
<212> DNA	
<213> Artificial Sequence	
<220><223> hBHB-h1W-008 vector, ITRs not included	
<400> 42	
caacaacaaa tgaatgcata tatatgtata tgtatgtgt tatatacata cacatata	60
tatataaaaa ttctttctt accagaaggt tttaatccaa ataaggagaa gatatgctta	120
gaaccgaggt agagtttca tccattctgt cctgtaagta tttgcatac tctggagacg	180
caggaagaga tccatctaca tatccaaag ctgaattatg gtagacaaaa ctcttcact	240
tttagtgcat caacttctta ttgtgtaat aagaaaattg ggaaaacgat cttcaatatg	300
cttaccaagg tgtgattcca aatattacgt aaatacactt gcaaaggagg atgttttag	360
tagcaattt tactgatggt atggggccaa gagatatac tttagagggag ggctgagggt	420
ttgaagtcca actcctaagc cagtgccaga agagccaagg acaggtacgg ctgtcatcac	480
ttagacctca ccctgtggag ccacacccta gggttggcca atctactccc aggagcaggg	540
agggcaggag ccaggctgg gcataaaagt cagggcagag ccatctattt cttacattt	600
cttctgacac aactgtgttc actagcaacc tcaaacagac accatggc atctgactcc	660
tgaggagaag tctgccgtt ctgccctgt gggcaaggtg aacgtggatg aagtgggtgg	720
tgagggccctg ggcaggttgg tatcaagggtt acaagacagg tttaaggaga ccaatagaaa	780
ctggccatgt ggagacagag aagttctga cctttctt tcccccaca gggcggtacc	840
agatctggca gccccgggg cagaggaat cttctaaccat gccgtgacgt ggaggagaat	900
cccgcccta gggtaccat ggtccaccc acaccagaag aaaaaagtgc tgtaacagct	960
ctctgggaa aagtcaatgt cgacgaggta gggggggaaag ctctcgaaag actcctcgac	1020
gtgtatccat ggacacaaag attttcgaa agcttcggag acctcagcac accagacgca	1080
gtatggaa atccaaaagt caaagcacac ggaaaaagg tccctggggc tttctctgac	1140
ggactcgcac atctcgatata tctgaaagga acattcgcta ccctcttga actccattgc	1200
gataaactcc atgtcgaccc agaaaattt agactgctcg gaaatgtcct cgtgtgcgtc	1260

ctcgctcacc attcgaaaa ggagttaca ctcctgtcc aagcagctt ccaaaaggc 1320  
 gtcgcagggg tcgcaaacgc tctcgctcat aaataccatt aggaattcaa ggcctctg 1380

gcctctagaa ctatagttag tcgttattacg tagatccaga catgataaga tacatttgatg 1440  
 agttggaca aaccacaact agaatgcagt gaaaaaaatg ctttatttgt gaaatttg 1500  
 atgttattgc ttatgttta accattataa gctgcaataa acaagttAAC aacaacaatt 1560  
 gcattcatt tatgtttcag gttcaggggg aggtgtggg ggtttttaa gtttcagctg 1620  
 tgatcataac gttgagctca agactcttgg gtttctgata ggcactgact ctctctgc 1680  
 attggtctat ttcccaccc ttaggctgct ggtggctac cttggaccc agaggttct 1740  
 tgagtcctt gggatctgt ccactcctga tgctgttatg ggcaacccta aggtgaaggc 1800

tcatggcaag aaagtgcctcg gtgcctttag tgcgttgcgtt gtcacccgtt acaacctaa 1860  
 gggcacctt gccacactga gtgagctgca ctgtgacaag ctgcacgtgg atcctgagaa 1920  
 cttcagggtg agtctatgg acgcttgcgtt ttttcttcc ctttctttc tatggtaag 1980  
 ttcatgtcat aggaagggga taagtaacag ggtacagttt agaatggaa acagacgaat 2040  
 gattgcatca gtgttggaaatg ctcaggatcg ttttagtttc ttttatttgc tgttcataac 2100  
 aattgttttcc ttgtttaaa ttcttgctt ctttttttt ctctccgca attttacta 2160  
 ttatacttaa tgccttaaca ttgtgtataa caaaaggaaa tatctctgag atacattaag 2220

taacttaaaa aaaaacttta cacagtctgc ctagfacatt actatttgaa atataatgtt 2280  
 gcttatttgc atattcataa tctccctact ttatatttctt ttatttttaa ttgatacata 2340  
 atcattatac atatttatgg gttaaagtgt aatgtttaa tatgtgtaca catattgacc 2400  
 aaatcagggtt aattttgcattt ttgttaatttt aaaaaatgc 2439

&lt;210&gt; 43

&lt;211&gt; 92

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; hHBB coding sequence in exon 1 with silent codon alterations

&lt;400&gt; 43

atggtccacc tcacaccaga agaaaaaaagt gctgtacacag ctctctgggg aaaagtcaat 60

gtcgacgagg tagggggggg agctctcgga ag 92

&lt;210&gt; 44

&lt;211&gt; 223

&lt;212&gt; DNA

<213> Artificial Sequence

<220><223> hHBB coding sequence in exon 2 with silent codon alterations

<400> 44

actcctcgta	gtgtatccat	ggacacaaaag	attttcgaa	agtttgcggag	acctcagcac	60
accagacgca	gtaatggaa	atccaaaagt	caaagcacac	ggaaaaaagg	tcctggggc	120
tttctctgac	ggactcgac	atctcgataa	tctgaaagga	acattcgcta	ccctctgt	180
actccattgc	gataaactcc	atgtcgaccc	agaaaattt	aga		223

<210> 45

<211> 129

<212> DNA

<213> Artificial Sequence

<220><223> hHBB coding sequence in exon 3 with silent codon alterations

<400> 45

ctgctcgaa	atgtcctcg	gtgcgtcctc	gctcaccatt	tcgaaagga	gtttacacct	60
cctgtccaag	cagttacca	aaaggctcg	gcaggggtcg	caaacgtct	cgctcataaa	120
taccattag						129

<210> 46

<211> 352

<212> DNA

<213> Artificial Sequence

<220><223> hHBB coding sequence from second codon to stop codon, with silent

codon alterations

<400> 46

actcctcgta	gtgtatccat	ggacacaaaag	attttcgaa	agtttgcggag	acctcagcac	60
accagacgca	gtaatggaa	atccaaaagt	caaagcacac	ggaaaaaagg	tcctggggc	120
tttctctgac	ggactcgac	atctcgataa	tctgaaagga	acattcgcta	ccctctgt	180
actccattgc	gataaactcc	atgtcgaccc	agaaaattt	agactgctcg	gaaatgtcct	240
cgtgtcgta	ctcgctacc	atttcgaaa	ggagttaca	cctcgttcc	aagcagctta	300
ccaaaaggta	gtcgagggg	tcgcaaacgc	tctcgctcat	aaataccatt	ag	352

<210> 47

<211> 441

<212> DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; hHBB coding sequence from start codon to stop codon, with silent codon alterations

&lt;400&gt; 47

gtccaccta caccagaaga	aaaaagtgtc gtaacagctc	tctgggaaa agtcaatgtc	60
gacgaggtag gggggaaagc	tctcggaaga ctccctcgtcg	tgtatccatg gacacaaaga	120
ttttcgaaa gcttcggaga	cctcagcaca ccagacgcag	taatggaaa tccaaaagtc	180
aaagcacacg gaaaaaaggt	cctggggct ttctctgacg	gactcgacata tcgataat	240
ctgaaaggaa cattcgctac	cctctctgaa ctccattgcg	ataaactcca tgtcgaccca	300

gaaaattta gactgctcg	aatgtcctc gtgtgcgtcc	tcgctcacca tttcgaaag	360
gagtttacac ctctgtcca	agcagcttac caaaagggtcg	tcgcagggt cgcaaacgct	420
ctcgctata aataccatta g			441

&lt;210&gt; 48

&lt;211&gt; 146

&lt;212&gt; PRT

&lt;213&gt; Homo sapiens

&lt;400&gt; 48

Val His Leu Thr Pro Glu Glu Lys Ser Ala Val Thr Ala Leu Trp Gly

1 5 10 15

Lys Val Asn Val Asp Glu Val Gly Gly Glu Ala Leu Gly Arg Leu Leu

20 25 30

Val Val Tyr Pro Trp Thr Gln Arg Phe Phe Glu Ser Phe Gly Asp Leu

35 40 45

Ser Thr Pro Asp Ala Val Met Gly Asn Pro Lys Val Lys Ala His Gly

50 55 60

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

65 70 75 80

Leu Lys Gly Thr Phe Ala Thr Leu Ser Glu Leu His Cys Asp Lys Leu

85 90 95

His Val Asp Pro Glu Asn Phe Arg Leu Leu Gly Asn Val Leu Val Cys

100 105 110

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

115

120

125

Ala Tyr Gln Lys Val Val Ala Gly Val Ala Asn Ala Leu Ala His Lys

130

135

140

Tyr His

145

&lt;210&gt; 49

&lt;211&gt; 8

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; 2A peptide consensus motif

&lt;220&gt;&lt;221&gt; VARIANT

&lt;222&gt; (1)..(1)

&lt;223&gt; X is Asp or Gly.

&lt;220&gt;&lt;221&gt; VARIANT

&lt;222&gt; (2)..(2)

&lt;223&gt; X is Val or Ile.

&lt;220&gt;&lt;221&gt; VARIANT

&lt;222&gt; (4)..(4)

&lt;223&gt; X is any amino acid.

&lt;400&gt; 49

Xaa Xaa Glu Xaa Asn Pro Gly Pro

1 5

&lt;210&gt; 50

&lt;211&gt; 45

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 50

agggtggag tcgtgacgtg ccaaatacg cctctacttg aatcc 45

&lt;210&gt; 51

&lt;211&gt; 54

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 51

aatgattaac ccgccatgct acttatctac gtaaacctag gctccagata gccaa	54
<210> 52	
<211> 18	
<212> DNA	
<213> Homo sapiens	
<400> 52	
ggaagcagaa ctctgcac	18
<210> 53	
<211> 27	
<212> DNA	
<213> Homo sapiens	
<400> 53	
gcattaaagg gtccttagtt ttttac	27
<210> 54	
<211> 26	
<212> DNA	
<213> Homo sapiens	
<400> 54	
gatggatgg ggccaagaga tatatc	26
<210> 55	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 55	
gtctaccctt ggaccaggag	20
<210> 56	
<211> 26	
<212> DNA	
<213> Homo sapiens	
<400> 56	
cagtcctgcct agtacattac tatttg	26
<210> 57	
<211> 23	

<212> DNA	
<213> Homo sapiens	
<400> 57	
catgttcata ccctttatct tcc	23
<210> 58	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 58	
gcaaacagct aatgcacatt gg	22
<210> 59	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 59	
cagaatccag atgctcaagg cc	22
<210> 60	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 60	
ccctgatttg gtcaatatgt	20
<210> 61	
<211> 23	
<212> DNA	
<213> Homo sapiens	
<400> 61	
catcaagcgt cccatagact cac	23
<210> 62	
<211> 21	
<212> DNA	
<213> Homo sapiens	
<400> 62	

gcagacttct cctcaggagt c	21
<210> 63	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 63	
cttacaggac agaatggatg aaaac	25
<210> 64	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 64	
gaaaaggct tctacttggc tc	22
<210> 65	
<211> 20	
<212> DNA	
<213> Homo sapiens	
<400> 65	
ggtaaccaa aagaaactgg	20
<210> 66	
<211> 22	
<212> DNA	
<213> Homo sapiens	
<400> 66	
tgcagattag tccaggcaga aa	22
<210> 67	
<211> 25	
<212> DNA	
<213> Homo sapiens	
<400> 67	
gggttaatcag tggtgtcaaa tagga	25
<210> 68	
<211> 20	
<212> DNA	

<213> Homo sapiens  
<400> 68  
agtttagatgt ccccagttaa 20  
<210> 69  
<211> 29  
<212> DNA  
<213> Artificial Sequence  
<220>  
><223> HBB2MTI100 Forward Primer  
<400> 69  
ctattggtct cctaaaatc gatactagt 29  
<210> 70  
<211> 24  
<212> DNA  
<213> Homo sapiens  
<400> 70  
atattcaaac ttccgcagaa cact 24  
<210> 71  
<211> 18  
<212> PRT  
<213> Artificial Sequence  
<220><223> T2A peptide  
<400> 71  
Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro  
1 5 10 15  
Gly Pro  
<210> 72  
<211> 54  
<212> DNA  
<213> Artificial Sequence  
<220><223> T2A element  
<400> 72  
gagggcagag gaagtcttct aacatgcgg t gacgtggagg agaatcccg ccct 54

<210> 73  
<211> 19  
<212> PRT  
<213> Artificial Sequence  
<220><223> P2A peptide  
<400> 73  
Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn  
1 5 10 15  
Pro Gly Pro

<210> 74  
<211> 57  
<212> DNA  
<213> Artificial Sequence  
<220><223> P2A element  
<400> 74  
gctactaact tcagcctgct gaagcaggct ggagacgtgg aggagaaccc tggacct 57  
<210> 75  
<211> 9  
<212> DNA  
<213> Artificial Sequence  
<220><223> GSG coding sequence  
<400> 75  
ggcagcggaa 9  
<210> 76  
<211> 198  
<212> DNA  
<213> Artificial Sequence  
<220><223> SV40 polyA signal  
<400> 76  
ttaaaaaaccc tccccacacct cccccctgaac ctgaaaacata aatgaatgc aatttgtt 60  
gttaacttgt ttattgcagc ttataatggc tacaataaa gcaatagcat cacaatttc 120  
acaataaag cattttttc actgcattct agtttggtt tgtccaaact catcaatgta 180

tcttatcatg tctggate	198
<210> 77	
<211> 198	
<212> DNA	
<213> Artificial Sequence	
<220><223> SV40 polyA signal	
<400> 77	
gatccagaca tgataagata cattgtatgag tttggacaaa ccacaactag aatgcagtga	60
aaaaaaatgct ttatttgta aatttgtat gctattgctt tattttaac cattataagc	120
tgcaataaac aagttAACaa caacaattgc attcattta tgTTTcaggt tcagggggag	180
gtgtgggagg tttttaa	198
<210> 78	
<211> 122	
<212> DNA	
<213> Artificial Sequence	
<220><223> SV40 polyA signal	
<400> 78	
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca	60
aataaaagcat tttttcact gcattctagt tgtggttgt ccaaactcat caatgtatct	120
ta	122
<210> 79	
<211> 122	
<212> DNA	
<213> Artificial Sequence	
<220><223> SV40 polyA signal	
<400> 79	
taagatacat ttagatgttt ggacaaacca caactagaat gcagtggaaaa aaatgctta	60
tttgtgaaat ttgtgtatgc attgctttat ttgttaaccat tataagctgc aataaacaag	120
tt	122
<210> 80	
<211> 12	
<212> DNA	
<213> Artificial Sequence	

<220><223> TI RE Linker

<400> 80

actagtatcg at 12

<210> 81

<211> 20

<212> DNA

<213> Artificial Sequence

<220><223> HBB L NGS \$1

<400> 81

cctctgggtc caaggtaga 20

<210> 82

<211> 87

<212> DNA

<213> Artificial Sequence

<220><223> HBB NGS Nest 5 Forward Adapter 1

<400> 82

aatgatacgg cgaccaccga gatctacaca agtagagttt ttcctacac gacgctttc 60

cgcattggc ataaaagtca gggcaga 87

<210> 83

<211> 88

<212> DNA

<213> Artificial Sequence

<220><223> HBB NGS Nest 5 Forward Adapter 2

<400> 83

aatgatacgg cgaccaccga gatctacacc atgcttatct ttcctacac gacgctttc 60

cgcattggc ataaaagtca gggcaga 88

<210> 84

<211> 89

<212> DNA

<213> Artificial Sequence

<220><223> HBB NGS Nest 5 Forward Adapter 3

<400> 84

aatgatacgg cgaccaccga gatctacacg cacatttct ttcctacac gacgctttc 60

cgatctatgg gcataaaagt cagggcaga	89
<210> 85	
<211> 90	
<212> DNA	
<213> Artificial Sequence	
<220><223> HBB NGS Nest 5 Forward Adapter 4	
<400> 85	
aatgatacgg cgaccaccga gatctacact gctcgactct ttccctacac gacgctttc	60
cgatctgatg ggcataaaag tcagggcaga	90
<210> 86	
<211> 88	
<212> DNA	
<213> Artificial Sequence	
<220><223> HBB NGS Nest 5 Reverse Adapter 1	
<400> 86	
caaggagaag acggcatacg agatcatgtat cggtgactgg agttcagacg tgtgctttc	60
cgcattgtct ccacatgccccc agtttcta	88
<210> 87	
<211> 89	
<212> DNA	
<213> Artificial Sequence	
<220><223> HBB NGS Nest 5 Reverse Adapter 2	
<400> 87	
caaggagaag acggcatacg agataggatc tagtgactgg agttcagacg tgtgctttc	60
cgcattgttc tccacatgccccc agtttcta	89
<210> 88	
<211> 90	
<212> DNA	
<213> Artificial Sequence	
<220><223> HBB NGS Nest 5 Reverse Adapter 3	
<400> 88	
caaggagaag acggcatacg agatgacagt aagtgactgg agttcagacg tgtgctttc	60
cgcattgtatgc ctccacatgc ccagtttcta	90

<210> 89  
<211> 90  
<212> DNA  
<213> Artificial Sequence  
<220><223> HBB NGS Nest 5 Reverse Adapter 4  
<400> 89  
caagcagaag acggcatacg agatcctatg ccgtgactgg agttcagacg tgtgctttc 60  
cgatctgagt ctccacatgc ccagtttcta 90

<210> 90  
<211> 19  
<212> DNA  
<213> Artificial Sequence  
<220><223> AAVS1\_Genomic, forward primer  
<400> 90  
gcgttagagg gcagagttc 19  
<210> 91  
<211> 19  
<212> DNA  
<213> Artificial Sequence  
<220><223> AAVS1\_Genomic, reverse primer  
<400> 91  
agctccata gctcagttc 19  
<210> 92  
<211> 23  
<212> DNA  
<213> Artificial Sequence  
<220><223> AAVS1\_Genomic, probe  
<400> 92  
cattgtcact ttgcgctgcc ctc 23

<210> 93  
<211> 22  
<212> DNA  
<213> Artificial Sequence

<220><223> AAVS1_FP forward primer	
<400> 93	
gcaatagcat cacaatttc ac	22
<210> 94	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220><223> AAVS1_FP, reverse primer	
<400> 94	
gatccagaca tgataagata cattg	25
<210> 95	
<211> 28	
<212> DNA	
<213> Artificial Sequence	
<220><223> AAVS1_FP, probe	
<400> 95	
tcaactgcatt ctagttgtgg tttgtcca	28
<210> 96	
<211> 25	
<212> DNA	
<213> Artificial Sequence	
<220><223> SA-2A-FM1	
<400> 96	
gcttctgacc tcttcttcc ctccc	25
<210> 97	
<211> 21	
<212> DNA	
<213> Artificial Sequence	
<220><223> SA-2A-FM2	
<400> 97	
gcggtgacgt ggaggagaat c	21
<210> 98	
<211> 24	
<212> DNA	

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; HBB-Out-RM2

&lt;400&gt; 98

gcagaatggc agcggattt tagc

24

&lt;210&gt; 99

&lt;211&gt; 444

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; silently altered HBB coding sequence

&lt;400&gt; 99

atggtgccacc tcaccccaga agagaagtct gctgtgacag ctctctgggg aaaagtaaat	60
gtggatgagg tgggagggga ggccttagga aggcttcttgc ttgtctaccc ctggacacag	120
agattctttg aatccttgg ggacctgagc actcctgatg cagtatggg caacccaaa	180
gtcaaagctc atggaaagaa gttttggga gccttctcag atggctggc tcacctggac	240
aacctgaagg gcacccgtc caccctgtct gagctgact gtgacaagct gcatgtggac	300

cctgagaact tcaggctcct gggcaatgtt ctggatgtg tgtagcaca ccattttggc	360
aaggaattca cccctccagt gcaggctgcc taccagaaag tggggcagg tgtggcaat	420
gccctggccc acaaataatca ctga	444

&lt;210&gt; 100

&lt;211&gt; 441

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; silently altered HBB coding sequence without start codon

&lt;400&gt; 100

gtgcacctca ccccaagaaga gaagtctgt gtgacagctc tctggggaaa agtaaatgtg	60
gatgaggtgg gaggggaggc ccttaggaagg cttcttgttgc tctaccctg gacacagaga	120

ttcttgaat ctttggga cctgagcact cctgatgcag tggatggcaa ccccaaagtcc	180
aaagctcatg ggaagaaggt ttgggagcc ttctcagatg gcctggctca cctggacacac	240
ctgaaggcaca ctttgcac cctgatgtg ctgcactgtg acaagctgca tgtggaccct	300
gagaacttca ggctctggg caatgttctg gtatgtgt tagcacacca tttggcaag	360
gaattcaccc ctccagtgca ggctgcctac cagaaatgg tggcagggtgt ggccatgcc	420
ctggccaca aatatcactg a	441

<210> 101  
<211> 489  
<212> DNA  
<213> Homo sapiens

<400> 101 ttgcatttgt aatttaaaa aatgcttct tcttttaata tactttttg tttatcttat 60  
ttctaatact ttccctaatac tcttttttc agggcaataa tgatacaatg tatcatgcct 120  
cttgcacca ttctaaagaa taacagtgt aatttctggg ttaaggcaat agcaatatct 180  
ctgcatataaa atatttctgc atataaattg taactgtatgt aagaggttc atattgctaa 240  
tagcagctac aatccagcta ccattctgct tttatttat ggttgggata aggctggatt 300  
attctgagtc caagcttaggc ccttttgcta atcatgttca tacctttat ctccctccca 360  
cagctcctgg gcaacgtgtc ggtctgtgtg ctggccccatc actttggcaa agaattcacc 420

ccaccaggatgc aggctgccta tcagaaaatgt gtggctggtg tggctaatgc cctggcccac 480  
aagtatcac 489

<210> 102

<211> 1134

<212> DNA

<213> Homo sapiens

<400> 102

gctcgcttc ttgctgtcca atttctatta aaggttcctt tgttccctaa gtccaaactac 60  
taaactgggg gatattatga agggccttga gcatctggat tctgccta at aaaaaacatt 120  
tat ttcatt gcaatgtatgt atttaaatttta tttctgaata ttttactaaa aagggaatgt 180  
gggaggtcag tgcatttaaa acataaagaa atgaagagct agttcaaaacc ttggggaaaat 240

caactatatc ttcaaactcca tgaaaagaagg tgaggctgca aacagcta at gcacattggc	300
aacagccccct gatgcata tg ctttattcat ccctcagaaa aggattcaag tagaggctg	360
atttgagggt taaagtttt ctatgctgta tttacatta cttattgttt tagctgtcct	420
catgaatgtc tttcactac ccatttgctt atccctgcata tctcagcctt gactccactc	480
agttctcttg cttagagata ccacccttcc cctgaagtgt tccttccatg ttttacggcg	540
agatggtttc tcctcgccctg gccactcagc cttagttgtc tctgttgtct tatagaggtc	600
tacttgaaga agggaaaaaca ggggtcatgg tttgactgtc ctgtgagcc ttcttccctg	660

cctccccac tcacagtgac ccggaaatcg cagtgttagt ctcccggaac tatcacttt 720

tcacagtctg ctttggagg actggccta gtagaaaat ttaggactga gaagaatttg aaaggcggtctttgttagct tgatattcac tactgtctta ttaccctgtc atagggccac cccaaatgga agtcccattc tccctcagga tggttaagat tagcattcag gaagagatca gagggtctgct ggccccccta tcatgtccct tatggcgtt ctggctctgc agttattagc atagtgttac catcaaccac ctttaactca ttttcttat tcaataccta ggttaggtaga tgcttagattc tgaaataaaa atatgagtct caagtggcc ttgtcccttc tcccgatcaa	780 840 900 960 1020 1080
attctgaatc tagttggcaa gattctgaaa tcaaggcata taatcagtaa taag <210> 103 <211> 1626 <212> DNA <213> Homo sapiens <400> 103	1134
ttgcatttgt aattttaaaa aatgctttct tcttttaata tacttttttg tttatcttat ttctaatact ttccctaatac tctttcttc agggcaataa tgatacaatg tatcatgcct ctttgcacca ttctaaagaa taacagtat aatttctggg ttaaggcaat agcaatatct ctgcataataa atatttctgc atataaatttgc taactgtat aagagggttc atattgctaa tagcagctac aatccagcta ccattctgtc tttatcttat ggttggata aggctggatt	60 120 180 240 300
attctgagtc caagctaggc ccttttgcta atcatgttca tacctttat cttccccc cagctccctgg gcaacgtgct ggtctgtgt ctggcccatc actttggcaa agaattcacc ccaccaggatgc aggtgcctc tcaaaaaatgt gtggctgggt tggttaatgc cctggccac aagtatcaatc aagctcgctt tcttgctgtc caatttctat taaagggttcc tttttccct aagtccaaacttcaacttgg gggatattat gaagggccctt gagcatctgg attctgcata ataaaaaaaca ttatcttca ttgcaatgtat gtatcttataat tatttctgaa tattttacta aaaaggaaat gtgggagggtc agtgcattt aacataaaag aaatgaagag ctatgtcaaa	360 420 480 540 600 660 720
ccttggaaa atacactata tcttaaactc catgaaagaa ggtgaggctg caaacagcta atgcacatttgc gcaacagccc ctgatgcata tgcatttttc atccctcaga aaaggattca agttagggct tgatggag gttaaagttt tgctatgtc tattttacat tacttattgt tttagctgtc ctcatgaatg tcttttcaact acccattgc ttatcctgca tctctcagcc ttgactccac tcagttctct tgcttagaga taccacctt cccctgaagt gttccctcca tgtttacgg cgagatgggtt ttcctcgcc tggccactca gccttagttg tctctgttgt ctttagagg tctacttgaa gaaggaaaaa caggggtcat ggtttgactg tcctgtgagc	780 840 900 960 1020 1080 1140

ccttctccc tgcctcccc actcacagt acccgaaatc tgcaatgcta gtctccgga	1200
actatcaact tticacagtc tgcttgaa ggactggct tagtatgaaa agttaggact	1260
gagaagaatt taaaaggcggt cttttgttag ctgtatattc actactgtct tattaccctg	1320
tcataggccc accccaaatg gaagtccat tcttccttag gatgttaag attagcattc	1380
aggaagagat cagaggtctg ctggctccct tatcatgtcc ctatggtgc ttctggctct	1440
gcagttatta gcatacggtt accatcaacc acctaactt cattttctt attcaataacc	1500
tagtaggtta gatgctagat tctggaaata aaatatgagt ctcaagtggc ccttgtcctc	1560
tctccagtc aaattctgaa tctagttggc aagattctga aatcaaggca tataatcagt	1620
aataag	1626
<210> 104	
<211> 2332	
<212> DNA	
<213> Artificial Sequence	
<220><223> hHBB vector	
<400> 104	
ttgcattgt aatttaaaa aatgcattct tcttttaata tactttttg tttatcttat	60
ttcttaatact ttccctaatac tctttcttgc agggcaataa tgatacaatg tatcatgcct	120
ctttgcacca ttctaaagaa taacagtgtat aattctggg ttaaggcaat agcaatatct	180
ctgcataataa atattctgc atataaatttgc taactgtatgt aagagggttc atattgctaa	240
tagcagctac aatccagcta ccattctgt tttatttat ggtgggata aggctggatt	300
attctgagtc caagcttaggc cttttgcta atcatgttca tacctttat cttccccc	360
cagctcctgg gcaacgtgct ggtctgtgt ctggccatc actttggcaa agaattcacc	420
ccaccagtgc aggctgccta tcagaaagtgt gtggctgggt tggtatgc cctggccac	480
aagtatcagc gcagcggagc tactaacttc agcctgctga agcaggctgg agacgtggag	540
gagaaccctg gacctatggt gcacccctacc ccagaagaga agtctgtgt gacagcttc	600
tgggaaaag taaatgtgga tgaggtgggaa ggggaggccc taggaaggct tcttgttgc	660
tacccctgga cacagagatt ctttgaatcc ttggggacc tgagcaatcc tgatgcagt	720
atggcaacc ccaaagtcaa agctcatggg aagaagggtt tggagccctt ctcagatggc	780
ctggctcacc tggacaacctt gaagggcacc ttgccaccc tgcgtgtgt gcactgtgac	840
aagctgcatttggaccctga gaacttcagg ctccctggca atgttctggatgtgtgtta	900
gcacaccatt ttggcaagga attcaccctt ccagtgcagg ctgcctacca gaaagtggtg	960
gcaggtgtgg ccaatgcctt gggccacaaa tatcactgag atccagacat gataagatac	1020

attgatgagt ttggacaaac cacaactaga atgcagtcaa aaaaatgctt tatttgtcaa	1080
---	------

atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca agttaacaac aacaattgca ttcatttat gttcagggtt cagggggagg tgtggaggt ttttaaagc tcgccttctt gctgtccaat ttctttaaa ggttccttg ttccctaagt ccaactacta aactggggta tattatgaag ggccttgagc atctggattc tgcttaataa aaaacattt ttttcattgc aatgatgtat ttaaattatt tctgaatatt ttactaaaaa ggaaatgtgg gaggtcagtg catttaaac ataaagaaat gaagagctag ttcaaacctt gggaaaatac actatatctt aaactccatg aaagaagggtg aggctgcaa cagctaattc acatggcaa	1140 1200 1260 1320 1380 1440 1500
---	--

cagccccgtga tgcataatgcc ttattcatcc ctcagaaaag gattcaagta gaggcttgat ttggaggta aagtttgcgt atgctgtatt ttacattact tattgtttt gctgtccca tgaatgtctt ttcaactaccc atttgcttat cctgcatttc tcagccttga ctccactcag ttctcttgct tagagataacc acctttcccc tgaagtgttc cttccatgtt ttacggcgag atggttctc ctcgcctggc cactcagect tagttgtctc tgggtctta tagaggtcta cttgaagaag gaaaaacagg ggtcatggtt tgactgtctt gtgagccctt cttccctgcc tcccccaactc acagtgaccc ggaatctgca gtgcttagtct cccggaacta tcactttc	1560 1620 1680 1740 1800 1860 1920
--	--

acagtctgct ttggaaaggac tgggcttagt atgaaaagtt aggactgaga agaatttcaa aggccgcttt ttgttagcttg atattcacta ctgtcttatt accctgtcat aggcccaccc caaatggaaag tccattttt cctcaggatg tttaagatta gcattcagga agagatcaga ggctgtctgg ctcccttatac atgtccctta tgggtcttgg ctgtctgcag ttatttagcat agtgttacca tcaaccacct taacttcatt ttcttattc aatacctagg taggttagatg ctagattctg gaaataaaat atgagtctca agtggccctt gtcctctctc ccagtcaat tctgaatcta gttggcaaga ttctgaaatc aaggcatata atcagtaata ag	1980 2040 2100 2160 2220 2280 2332
--	--

<210> 105

<211> 92

<212> DNA

<213> Artificial Sequence

<220><223> Silently altered HBB coding sequence corresponding to HBB exon 1

<400> 105

atgggtgcacc tcacccaga agagaagtct gctgtgacag ctctctgggg aaaagtaat	60
--	----

gtggatgagg tgggagggaa ggccttagga ag	92
-------------------------------------	----

<210> 106

&lt;211&gt; 223

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Silently altered HBB coding sequence corresponding to HBB exon 2

&lt;400&gt; 106

gcttcttgtt gctaccct ggacacagag attcttgaa tccttgggg acctgagcac 60

tcctgatgca gtatggca accccaaagt caaagctcat gggagaagg ttttggagc 120

cttctcagat ggctggctc acctggacaa cctgaaggc acctttgccca ccctgtctga 180

gctgcactgt gacaagctgc atgtggaccc tgagaacttc agg 223

&lt;210&gt; 107

&lt;211&gt; 129

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Silently altered HBB coding sequence corresponding to HBB exon 3

&lt;400&gt; 107

tcctggca atgttctggatgtgttgcacaccatt ttggcaagga attcaccct 60

ccagtgcagg ctgcctacca gaaagtggc gcaggtgtgg ccaatgccct ggcccacaaa 120

tatcactga 129

&lt;210&gt; 108

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Synthetic polynucleotide

&lt;400&gt; 108

aaagt caggg cagagccatc 20

&lt;210&gt; 109

&lt;211&gt; 20

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Synthetic polynucleotide

&lt;400&gt; 109

aatgattaac ccgccatgct 20

&lt;210&gt; 110

&lt;211&gt; 24

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Synthetic polynucleotide

&lt;400&gt; 110

aactggcat gtggagacag agaa 24

&lt;210&gt; 111

&lt;211&gt; 25

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Synthetic polynucleotide

&lt;400&gt; 111

gttacaagac aggactagta tcgat 25

&lt;210&gt; 112

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Synthetic polynucleotide

&lt;400&gt; 112

tagaccaata ggcagagaga gt 22

&lt;210&gt; 113

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Synthetic polynucleotide

&lt;400&gt; 113

ctgagccaag tagaagacct tt 22

&lt;210&gt; 114

&lt;211&gt; 22

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;&lt;223&gt; Synthetic polynucleotide

&lt;400&gt; 114

ctgtttctgc ctggactaat ct 22  
<210> 115  
<211> 25  
<212> DNA  
<213> Artificial Sequence  
<220><223> Synthetic polynucleotide  
<400> 115  
ccctactttc taagtacacag aggct 25