



(19) 대한민국특허청(KR)  
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0161459  
(43) 공개일자 2023년11월27일

- |   |   |
|---|---|
| <p>(51) 국제특허분류(Int. Cl.)<br/>C12N 15/63 (2006.01) C07K 14/47 (2006.01)<br/>C07K 14/705 (2006.01) C07K 14/78 (2006.01)<br/>C12N 15/86 (2006.01)</p> <p>(52) CPC특허분류<br/>C12N 15/63 (2013.01)<br/>C07K 14/472 (2013.01)</p> <p>(21) 출원번호 10-2023-7034581</p> <p>(22) 출원일자(국제) 2022년03월14일<br/>심사청구일자 없음</p> <p>(85) 번역문제출일자 2023년10월10일</p> <p>(86) 국제출원번호 PCT/GB2022/050649</p> <p>(87) 국제공개번호 WO 2022/189811<br/>국제공개일자 2022년09월15일</p> <p>(30) 우선권주장<br/>2103470.7 2021년03월12일 영국(GB)<br/>(뒷면에 계속)</p> | <p>(71) 출원인<br/>더 유니버시티 오브 브리스톨<br/>영국 브리스톨 퀸즈 로드 비콘 하우스 (우: 비에스8 1큐유)</p> <p>신코나 아이피 홀드코 (3) 리미티드<br/>영국 런던 블룸스버리 스트리트 8 (우: 더블유씨 1비 3에스알)</p> <p>(72) 발명자<br/>살렘-우딘, 모인<br/>영국 비에스8 1큐유 브리스톨 퀸즈 로드 비콘 하우스 더 유니버시티 오브 브리스톨 (내)</p> <p>웰시, 개빈<br/>영국 비에스8 1큐유 브리스톨 퀸즈 로드 비콘 하우스 더 유니버시티 오브 브리스톨 (내)<br/>(뒷면에 계속)</p> <p>(74) 대리인<br/>특허법인 남앤남</p> |
|---|---|

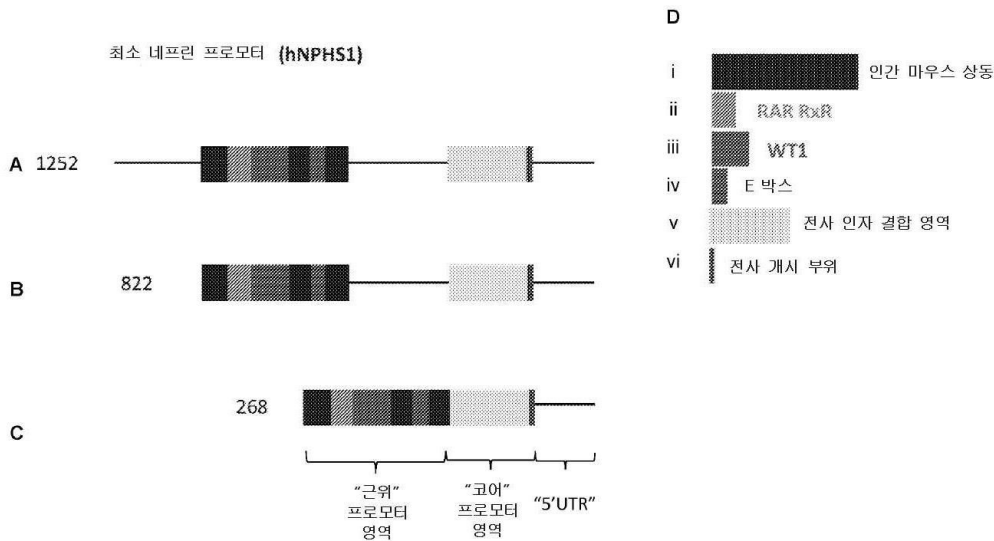
전체 청구항 수 : 총 51 항

(54) 발명의 명칭 최소 네프린 프로모터

(57) 요약

(i) SEQ ID NO: 4 또는 17에 대해 적어도 70%의 동일성을 갖는 뉴클레오티드 서열을 포함하는 프로모터로서, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.

대표도



(52) CPC특허분류

*C07K 14/705* (2013.01)  
*C07K 14/70503* (2013.01)  
*C07K 14/78* (2013.01)  
*C12N 15/86* (2013.01)  
*C07K 2319/41* (2013.01)  
*C07K 2319/42* (2013.01)  
*C12N 2750/14143* (2013.01)  
*C12N 2830/008* (2013.01)  
*C12N 2830/15* (2013.01)

(72) 발명자

**쿠즈뚝, 발렐리아**

영국 더블유씨1비 3에스알 런던 블룸스버리 스트리트 8 세컨드 플로어 퓨어스프링 테라퓨틱스 (내)

**그리피스, 알란 윌리엄**

영국 더블유씨1비 3에스알 런던 블룸스버리 스트리트 8 세컨드 플로어 퓨어스프링 테라퓨틱스 (내)

(30) 우선권주장

PCT/GB2021/050633 2021년03월12일 영국(GB)  
PCT/GB2021/051668 2021년06월30일 영국(GB)

## 명세서

### 청구범위

#### 청구항 1

(i) SEQ ID NO: 4 또는 17에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하는 프로모터로서, 프로모터의 길이가 약 1.1 kb 이하인, 프로모터.

#### 청구항 2

SEQ ID NO: 1에 따른 뉴클레오티드 서열로서, 여기서

(i) SEQ ID NO: 1의 위치 1 내지 위치 n1 (여기서 n1은 100 내지 430의 정수임)이 결실되고/거나;

(ii) SEQ ID NO: 1의 위치 n2 내지 위치 n3 (여기서  $n3 \geq n2$ 이고, n2는 508 내지 1061의 정수이고, n3은 508 내지 1061의 정수임)이 결실되는, 뉴클레오티드 서열, 또는

이에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는 프로모터.

#### 청구항 3

제2항에 있어서,

(iii) SEQ ID NO: 1의 위치 493에 상응하는 위치가 G이고, SEQ ID NO: 1의 위치 1080에 상응하는 위치가 T이고, SEQ ID NO: 1의 위치 1169에 상응하는 위치가 C이고, SEQ ID NO: 1의 위치 1249에 상응하는 위치는 G인, 프로모터.

#### 청구항 4

제2항 또는 제3항에 있어서, 프로모터가 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하고/거나 프로모터의 길이는 약 1.1 kb 이하인, 프로모터.

#### 청구항 5

제1항 내지 제4항 중 어느 한 항에 있어서, 프로모터가 신장-특이적 프로모터인, 프로모터.

#### 청구항 6

제1항 내지 제5항 중 어느 한 항에 있어서, 프로모터가 (ii) SEQ ID NO: 5 또는 18에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열 및/또는 SEQ ID NO: 7 또는 20에 대해 적어도 70%의 동일성을 갖는 뉴클레오티드 서열을 추가로 포함하는, 프로모터.

#### 청구항 7

제1항 내지 제6항 중 어느 한 항에 있어서, 프로모터가 (iii) SEQ ID NO: 8에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열 또는 이의 하나 이상의 단편을 추가로 포함하는, 프로모터.

#### 청구항 8

제1항 내지 제7항 중 어느 한 항에 있어서, 프로모터가 5'에서 3'으로 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열; (iii) 임의적으로 SEQ ID NO: 8에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열 또는 이의 하나 이상의 단편; 및 (ii) SEQ ID NO: 5 또는 18에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열 및/또는 SEQ ID NO: 7 또는 20에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하는, 프로모터.

#### 청구항 9

제1항 내지 제8항 중 어느 한 항에 있어서, 프로모터의 길이가 약 1.0 kb 이하, 약 0.9 kb 이하, 약 0.8 kb 이

하, 약 0.7 kb 이하, 약 0.6 kb 이하, 약 0.5 kb 이하, 약 0.4 kb 이하 또는 약 0.3 kb 이하인, 프로모터.

**청구항 10**

제1항 내지 제9항 중 어느 한 항에 있어서, 프로모터의 길이가 0.265 kb 내지 1.0 kb, 0.265 kb 내지 0.9 kb, 0.265 kb 내지 0.8 kb, 0.265 kb 내지 0.7 kb, 0.265 kb 내지 0.6 kb, 0.265 kb 내지 0.5 kb, 0.265 kb 내지 0.4 kb 또는 0.265 kb 내지 0.3 kb인, 프로모터.

**청구항 11**

제1항 내지 제10항 중 어느 한 항에 있어서, 프로모터가

- (a) 레티노산 수용체 결합 부위;
- (b) WT1 결합 부위;
- (c) 인헨서 박스;
- (d) 전사 인자 결합 영역; 및/또는
- (e) 전사 개시 부위를 포함하는, 프로모터.

**청구항 12**

제11항에 있어서, 레티노산 수용체 결합 부위가 SEQ ID NO: 10으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 10과 비교하여 1 또는 2개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 13**

제11항 또는 제12항에 있어서, WT1 결합 부위가 SEQ ID NO: 11로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 11과 비교하여 1, 2 또는 3개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 14**

제11항 내지 제13항 중 어느 한 항에 있어서, 인헨서 박스가 SEQ ID NO: 12로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 12와 비교하여 1 또는 2개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 15**

제1항 또는 제4항 내지 제14항 중 어느 한 항에 있어서, (i) SEQ ID NO: 4 또는 17과 적어도 70% 동일성을 갖는 뉴클레오티드 서열이 (a) 레티노산 수용체 결합 부위; (b) WT1 결합 부위; 및 (c) 인헨서 박스 중 하나 이상을 포함하는, 프로모터.

**청구항 16**

제1항 또는 제4항 내지 제15항 중 어느 한 항에 있어서, 하기 뉴클레오티드 서열 중 하나 이상이 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열에 존재하는, 프로모터:

- (a) SEQ ID NO: 4 또는 17의 위치 7 내지 위치 13에 상응하는 위치에서 GGGGTCA;
- (b) SEQ ID NO: 4 또는 17의 위치 14 내지 위치 30에 상응하는 위치에서 CGGAGGCTGGGGAGGCA; 및
- (c) SEQ ID NO: 4 또는 17의 위치 49 내지 위치 53에 상응하는 위치에서 ATGTG.

**청구항 17**

제11항 내지 제16항 중 어느 한 항에 있어서, 전사 인자 결합 영역이 SEQ ID NO: 13 또는 43으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 13 또는 43과 비교하여 1, 2, 3, 4 또는 5개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 18**

제11항 내지 제17항 중 어느 한 항에 있어서, 전사 개시 부위가 "AG" 디뉴클레오티드를 포함하거나 이로 구성되는, 프로모터.

**청구항 19**

제11항 내지 제18항 중 어느 한 항에 있어서, 전사 인자 결합 부위가 전사 개시 부위에 작동가능하게 연결되며, 임의적으로 전사 인자 결합 부위는 전사 개시 부위의 바로 업스트림에 있는, 프로모터.

**청구항 20**

제1항 내지 제19항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 2 또는 SEQ ID NO: 3에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 21**

제1항 내지 제20항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 22**

제1항 내지 제21항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 23**

제1항 내지 제22항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 95% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 24**

제1항 내지 제23항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 98% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 25**

제1항 내지 제24항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.

**청구항 26**

SEQ ID NO: 2에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열로 구성되는 프로모터.

**청구항 27**

SEQ ID NO: 3에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열로 구성되는 프로모터.

**청구항 28**

SEQ ID NO: 15 또는 59에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열로 구성되는 프로모터.

**청구항 29**

SEQ ID NO: 47에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열로 구성되는 프로모터.

**청구항 30**

제29항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열로 구성되는, 프로모터.

**청구항 31**

제29항 또는 제30항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 95% 동일성을 갖는 뉴클레오티드 서열로 구성되는, 프로모터.

**청구항 32**

제29항 내지 제31항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 98% 동일성을 갖는 뉴클레오티드 서열로 구성되는, 프로모터.

**청구항 33**

제29항 내지 제32항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 47에 대해 적어도 99% 동일성을 갖는 뉴클레오티드 서열로 구성되는, 프로모터.

**청구항 34**

제29항 내지 제32항 중 어느 한 항에 있어서, 프로모터가 SEQ ID NO: 47의 뉴클레오티드 서열로 구성되는, 프로모터.

**청구항 35**

제1항 내지 제34항 중 어느 한 항에 따른 프로모터를 포함하는 폴리뉴클레오티드.

**청구항 36**

제35항에 있어서, 프로모터가 단백질 코딩 서열에 작동가능하게 연결되는, 폴리뉴클레오티드.

**청구항 37**

제35항 또는 제36항에 있어서, 프로모터가 NPHS2 또는 이의 단편 및/또는 변이체; COL4A3, COL4A4 또는 COL4A5 폴리펩티드, 또는 이의 단편 또는 유도체; 또는 CFI, CFH 또는 FHL-1, 또는 이의 단편 및/또는 변이체를 인코딩하는 단백질 코딩 서열에 작동가능하게 연결되는, 폴리뉴클레오티드.

**청구항 38**

제35항 내지 제37항 중 어느 한 항에 있어서, 프로모터가 NPHS2 또는 이의 단편 및/또는 변이체를 인코딩하는 단백질 코딩 서열에 작동가능하게 연결되고, 바람직하게는 단백질 코딩 서열은 SEQ ID NO: 55와 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 인코딩하는, 폴리뉴클레오티드.

**청구항 39**

제35항 내지 제37항 중 어느 한 항에 있어서, 프로모터가 COL4A3, COL4A4 또는 COL4A5 폴리펩티드, 또는 이의 단편 또는 유도체를 인코딩하는 단백질 코딩 서열에 작동가능하게 연결되고, 바람직하게는 단백질 코딩 서열은 SEQ ID NO: 21-23 중 하나와 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 인코딩하는, 폴리뉴클레오티드.

**청구항 40**

제35항 내지 제37항 중 어느 한 항에 있어서, 프로모터가 CFI, CFH 또는 FHL-1, 또는 이의 단편 및/또는 변이체를 인코딩하는 단백질 코딩 서열에 작동가능하게 연결되고, 바람직하게는 단백질 코딩 서열은 SEQ ID NO: 49, 51 또는 53과 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 인코딩하는, 폴리뉴클레오티드.

**청구항 41**

제35항 내지 제40항 중 어느 한 항에 있어서, 단백질 코딩 서열이 전사-후 조절 요소 및/또는 폴리아데닐화 서열과 같은 하나 이상의 추가 조절 요소에 작동가능하게 연결되는, 폴리뉴클레오티드.

**청구항 42**

제35항 내지 제41항 중 어느 한 항에 따른 폴리뉴클레오티드를 포함하는 벡터.

**청구항 43**

제42항에 있어서, 벡터가 발세포를 형질도입할 수 있고, 임의적으로 벡터는 발세포를 특이적으로 형질도입할 수 있는, 벡터.

**청구항 44**

제42항 또는 제43항에 있어서, 벡터가 바이러스 벡터, 예컨대, 아데노-관련 바이러스(AAV) 벡터, 렌티바이러스 벡터, 레트로바이러스 벡터, 아데노바이러스 벡터, 아데노-관련 바이러스 벡터, 단순 포진 바이러스 벡터, 알파 바이러스 벡터, 플라비바이러스 벡터, 람다바이러스 벡터, 홍역 바이러스 벡터, 뉴캐슬병 바이러스 벡터, 폭스 바이러스 벡터 및 피코르나바이러스 벡터이고, 바람직하게는 벡터는 AAV 벡터인, 벡터.

**청구항 45**

제44항에 있어서, 바이러스 벡터가 바이러스 벡터 입자의 형태이고, 바람직하게는 바이러스 벡터는 AAV 벡터 입자의 형태인, 벡터.

**청구항 46**

제42항 내지 제45항 중 어느 한 항에 있어서, 벡터가 AAV3B, LK03 또는 AAV9 캡시드 단백질에 의해 캡시드화된 AAV 벡터 입자의 형태인, 벡터.

**청구항 47**

제35항 내지 제41항 중 어느 한 항에 따른 폴리뉴클레오티드 또는 제42항 내지 제46항 중 어느 한 항에 따른 벡터를 포함하는 세포.

**청구항 48**

제35항 내지 제41항 중 어느 한 항에 따른 폴리뉴클레오티드, 제42항 내지 제46항 중 어느 한 항에 따른 벡터 또는 제47항에 따른 세포를 포함하는 약학적 조성물.

**청구항 49**

의약에 사용하기 위한 제35항 내지 제41항 중 어느 한 항에 따른 폴리뉴클레오티드, 제42항 내지 제46항 중 어느 한 항에 따른 벡터 또는 47항에 따른 세포.

**청구항 50**

코딩 서열의 발현을 유도하기 위한 제1항 내지 제34항 중 어느 한 항에 따른 프로모터의 용도.

**청구항 51**

제50항에 있어서, 발현이 신장-특이적인, 용도.

**발명의 설명**

**기술 분야**

**발명의 분야**

본 발명은 발세포에서 전이유전자 발현을 유도할 수 있는 프로모터에 관한 것이다. 본 발명은 또한 상기 프로모터를 포함하는 폴리뉴클레오티드 및 벡터에 관한 것이다.

**배경 기술**

[0001]

[0002]

**[0003] 본 발명에 따른 배경**

**[0004]** 사구체를 공격함으로써 신장 기능에 영향을 미치는 많은 질병이 있다. 사구체는 매일 약 180리터의 혈장을 여과하고, 건강한 사구체 여과 장벽은 막힘 없이 우리의 일생에 걸쳐 알부민을 포함하는 큰 단백질의 약 99.9%를 보유하는 놀라운 능력을 갖는다. 사구체 여과 장벽(GFB)은 3개의 주요 층: 사구체 내피 세포, 사구체 기저막(GBM) 및 발세포를 포함한다.

**[0005]** GBM은 유형 IV 콜라겐, 프로테오글리칸 및 라미닌의 고도로 가교된 거대분자 메쉬워크로 제조된다. 사구체 질환의 유전적 형태는 이러한 분자 구조의 유전적 결함에 의해 야기될 수 있다. 예를 들어, 알포트 증후군은 COL4A3, COL4A4 및 COL4A5 유전자의 병원성 변이체에 의해 유발되며, 이는 기저막의 콜라겐 IV α345 네트워크의 이상을 초래한다. 알포트 증후군은 유럽 대륙 및 미국에서 모든 개인의 대략 5,000-10,000명 중 약 1명에게 영향을 미친다. 이 질환은 일반적으로 어린 시절에 나타나며, 신장 기능의 점진적인 상실을 포함하는 표현형의 스펙트럼과 관련이 있으며, 또한 청력 상실 및 눈 이상을 포함할 수 있다. 다른 GBM-관련 질환은 피어슨 증후군 및 네일-슬개골 증후군을 포함한다(Chiang, C.K. and Inagi, R., 2010. Nature Reviews Nephrology, 6(9), p.539).

**[0006]** 발세포는 또한 사구체 질환의 진행에서 핵심 세포로서 연루되어 있다. 발세포는 고도로 전문화되고 신장 사구체에서만 발견되는 중배엽 유래된 세포이다. 이들은 사구체 여과에 중요한 족돌기 및 세극막과 같은 독특한 특성을 나타낸다. 발세포-관련 유전 사구체 질환은 콩팥 증후군, 프레이저 증후군 및 데니스-드라시 증후군, 쉬미케면역-골이형성증, 및 엡스타인 및 페히트너 증후군을 포함한다. (Chiang, C.K. and Inagi, R., 2010. Nature Reviews Nephrology, 6(9), p.539).

**[0007]** 따라서, 발세포와 같은 사구체 세포는 유전자 요법 접근법에 대한 잠재적인 표적을 나타낸다.

**[0008]** 유전자 치료 가능성을 최대화하기 위해, 발세포와 같은 사구체 세포에서 전이유전자 발현을 유도할 수 있는 프로모터 서열이 필요하다. 왕(Wong) 등(American Journal of Physiology Renal Physiology; 2000; 279(6); F1027-32)은 발세포-특이적 발현을 지시할 수 있는 인간 네프린 프로모터 및 5'-측접 영역으로부터의 1.25-kb DNA 단편을 기술하였다. 이 프로모터는 AAV9 벡터를 사용하여 GFP의 신장-특이적 발현을 달성하는데 사용되었다 (Picconi et al.; 2014; Molecular Therapy - Methods & Clinical Development; 1, 14014).

**[0009]** 현재, 상당한 양의 벡터 카고 용량(예를 들어, AAV 카고 용량)은 비-코딩 요소, 예컨대, 네프린 또는 포도신 프로모터, WPRE 요소 및 폴리아데닐화 서열로 흡수된다. 따라서, 사구체 질환에 대한 유전자 요법 접근법에 최대 유연성을 제공하기 위해, 발세포와 같은 사구체 세포에서 전이유전자 발현을 유도할 수 있는 최소 프로모터가 당 분야에 필요하다.

**발명의 내용**

**[0010] 발명의 개요**

**[0011]** 본 발명은 신장 세포, 특히 사구체 세포, 예를 들어, 발세포에서 전이유전자 발현을 유도할 수 있는 최소 프로모터의 본 발명자들의 놀라운 제공에 기초한다.

**[0012]** 일 양태에서, 본 발명은 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하는 프로모터를 제공하며, 여기서 프로모터는 약 1.1 kb 이하의 길이를 갖는다.

**[0013]** 일부 구체예에서, 본 발명은 SEQ ID NO: 1에 따른 뉴클레오티드 서열로서, 여기서

**[0014]** (i) SEQ ID NO: 1의 위치 1 내지 위치 n1 (여기서 n1은 1 내지 430의 정수임)이 결실되고/거나;

**[0015]** (ii) SEQ ID NO: 1의 위치 n2 내지 위치 n3 (여기서  $n3 \geq n2$ 이고, n2는 508 내지 1061의 정수이고, n3은 508 내지 1061의 정수임)이 결실되는, 뉴클레오티드 서열 또는 이에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된 프로모터를 제공한다.

**[0016]** 일부 구체예에서,

**[0017]** (iii) SEQ ID NO: 1의 위치 493에 상응하는 위치가 G이고, SEQ ID NO: 1의 위치 1080에 상응하는 위치가 T이고, SEQ ID NO: 1의 위치 1169에 상응하는 위치가 C이고, SEQ ID NO: 1의 위치 1249에 상응하는 위치는 G이다.

**[0018]** 일 양태에서, 본 발명은 SEQ ID NO: 1에 따른 뉴클레오티드 서열에 대해 적어도 70%, 적어도 80%, 적어도 90%,

적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된 프로모터를 제공하며, 여기서

- [0019] (i) SEQ ID NO: 1의 위치 1 내지 위치 n1 (여기서 n1은 100 내지 430의 정수임)이 결실되고/거나;
- [0020] (ii) SEQ ID NO: 1의 위치 n2 내지 위치 n3 (여기서  $n3 \geq n2$ 이고, n2는 508 내지 1061의 정수이고, n3은 508 내지 1061의 정수임)이 결실된다.
- [0021] 일부 구체예에서,
- [0022] (iii) SEQ ID NO: 1의 위치 493이 G로 치환되고, SEQ ID NO: 1의 위치 1080이 T로 치환되고, SEQ ID NO: 1의 위치 1169가 C로 치환되고, SEQ ID NO: 1의 위치 1249가 G로 치환된다.
- [0023] 프로모터는 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함할 수 있으며/거나 프로모터는 약 1.1 kb 이하의 길이를 가질 수 있다.
- [0024] 프로모터는 신장 세포에서 작동가능할 수 있다. 프로모터는 신장에서 전이유전자 발현을 유도할 수 있다. 프로모터는 신장-특이적 프로모터일 수 있다. 프로모터는 발세포에서 작동가능할 수 있다. 프로모터는 발세포에서 전이유전자 발현을 유도할 수 있다. 프로모터는 발세포-특이적 프로모터일 수 있다.
- [0025] 적합하게는, 프로모터는 (i) SEQ ID NO: 5 또는 18에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 서열 및/또는 SEQ ID NO: 7 또는 20에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 추가로 포함한다. 일부 구체예에서, 프로모터는 (ii) SEQ ID NO: 6 또는 19에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 추가로 포함한다.
- [0026] 적합하게는, 프로모터는 (iii) SEQ ID NO: 8에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열, 또는 이의 하나 이상의 단편을 추가로 포함한다.
- [0027] 적합하게는, 임의의 상기 청구항에 있어서, 프로모터는 5'에서 3'으로: (i) SEQ ID NO: 4 또는 17에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열; (iii) 임의적으로, SEQ ID NO: 8에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열, 또는 이의 하나 이상의 단편; 및 (ii) SEQ ID NO: 5 또는 18에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열 및/또는 SEQ ID NO: 7 또는 20에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함한다.
- [0028] 일부 구체예에서, 프로모터는 5'에서 3'으로 (i) SEQ ID NO: 4에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열; 및 (ii) SEQ ID NO: 7에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함한다. 일부 구체예에서, 프로모터는 5'으로부터 3'으로 (i) SEQ ID NO: 4에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열; 및 (ii) SEQ ID NO: 7에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열로 구성되며, 여기서 프로모터는 약 1.1 kb 이하의 길이를 갖는다.
- [0029] 일부 구체예에서, 프로모터는 5'에서 3'으로 (i) SEQ ID NO: 17에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열; 및 (ii) SEQ ID NO: 20에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열로 구성된다. 일부 구체예에서, 프로모터는 5'에서 3'으로 (i) SEQ ID NO: 17의 뉴클레오티드 서열; 및 (ii) SEQ ID NO: 20의 뉴클레오티드 서열로 구성된다.
- [0030] 일부 구체예에서, 프로모터는 5'에서 3'으로 (i) SEQ ID NO: 17에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열; 및 (ii) SEQ ID NO: 20에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%

또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하며, 여기서 프로모터는 약 1.1 kb 이하의 길이를 갖는다. 일부 구체예에서, 프로모터는 5'에서 3'으로: (i) SEQ ID NO: 17의 뉴클레오티드 서열; 및 (ii) SEQ ID NO: 20의 뉴클레오티드 서열을 포함하고, 여기서 프로모터는 약 1.1 kb 이하의 길이를 갖는다.

- [0031] 프로모터는 약 1.0 kb 이하, 약 0.9 kb 이하, 약 0.8 kb 이하, 약 0.7 kb 이하, 약 0.6 kb 이하, 약 0.5 kb 이하, 약 0.4 kb 이하 또는 약 0.3 kb 이하의 길이를 가질 수 있다. 프로모터는 0.265 kb 내지 1.0 kb, 0.265 kb 내지 0.9 kb, 0.265 kb 내지 0.8 kb, 0.265 kb 내지 0.7 kb, 0.265 kb 내지 0.6 kb, 0.265 kb 내지 0.5 kb, 0.265 kb 내지 0.4 kb 또는 0.265 kb 내지 0.3 kb의 길이를 가질 수 있다.
- [0032] 프로모터는 (a) 레티노산 수용체 결합 부위; (b) WT1 결합 부위; (c) 인헨서 박스; (d) 전사 인자 결합 영역; 및/또는 (e) 전사 개시 부위를 포함할 수 있다. 적합하게는, 레티노산 수용체 결합 부위는 SEQ ID NO: 10으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 10과 비교하여 1 또는 2개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다. 적합하게는, WT1 결합 부위는 SEQ ID NO: 11로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 11과 비교하여 1, 2 또는 3개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다. 적합하게는, 인헨서 박스는 SEQ ID NO: 12로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 12와 비교하여 1 또는 2개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다. 적합하게는, 전사 인자 결합 영역은 SEQ ID NO: 13 또는 43으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 13 또는 43과 비교하여 1, 2, 3, 4 또는 5개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다. 적합하게는, 전사 개시 부위는 "AG" 디뉴클레오티드를 포함하거나 이로 구성된다. 적합하게는, 전사 인자 결합 부위는 전사 개시 부위에 작동가능하게 연결되며, 임의적으로 전사 인자 결합 부위는 전사 개시 부위의 바로 업스트림에 있다.
- [0033] SEQ ID NO: 4 또는 17에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열은 (a) 레티노산 수용체 결합 부위; (b) WT1 결합 부위; 및 (c) 인헨서 박스 중 하나 이상을 포함할 수 있다. 하기 뉴클레오티드 서열 중 하나 이상은 SEQ ID NO: 4 또는 17에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열에 존재할 수 있다: (a) SEQ ID NO: 4 또는 17의 위치 7 내지 위치 13에 상응하는 위치에서 GGGGTCA; (b) SEQ ID NO: 4 또는 17의 위치 14 내지 위치 30에 상응하는 위치에서 CGGAGGCTGGGGAGGCA; 및 (c) SEQ ID NO: 4 또는 17의 위치 49 내지 위치 53에 상응하는 위치에서 ATGTG.
- [0034] 일부 구체예에서, 프로모터는 SEQ ID NO: 2 또는 3에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다.
- [0035] 일부 구체예에서, 프로모터는 SEQ ID NO: 15, SEQ ID NO: 59 또는 SEQ ID NO: 47에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다.
- [0036] 바람직한 구체예에서, 프로모터는 SEQ ID NO: 47에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다. 더욱 바람직한 구체예에서, 프로모터는 SEQ ID NO: 47의 뉴클레오티드 서열을 포함하거나 이로 구성된다.
- [0037] 일 양태에서, 본 발명은 SEQ ID NO: 2에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열로 구성된 프로모터를 제공한다.
- [0038] 일 양태에서, 본 발명은 SEQ ID NO: 3에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열로 구성된 프로모터를 제공한다.
- [0039] 일 양태에서, 본 발명은 SEQ ID NO: 15에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열로 구성된 프로모터를 제공한다.
- [0040] 일 양태에서, 본 발명은 SEQ ID NO: 59에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열로 구성된 프로모터를 제공한다.
- [0041] 일 양태에서, 본 발명은 SEQ ID NO: 47에 대해 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열로 구성된 프로모터를 제공한다.
- [0042] 일 양태에서, 본 발명은 SEQ ID NO: 47의 뉴클레오티드 서열로 구성된 프로모터를 제공한다.
- [0043] 일 양태에서, 본 발명은 본 발명에 따른 프로모터를 포함하는 폴리뉴클레오티드를 제공한다.

- [0044] 프로모터는 단백질 코딩 서열에 작동가능하게 연결될 수 있다. 일부 구체예에서, 단백질 코딩 서열은 NPHS2 또는 이의 단편 및/또는 변이체; COL4A3, COL4A4 또는 COL4A5 폴리펩티드, 또는 이의 단편 또는 유도체; 또는 CFI, CFH 또는 FHL-1, 또는 이의 단편 및/또는 변이체를 인코딩한다.
- [0045] 일부 구체예에서, 단백질 코딩 서열은 NPHS2, 또는 이의 단편 및/또는 변이체를 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 55와 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 56과 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 57과 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다.
- [0046] 일부 구체예에서, 단백질 코딩 서열은 COL4A3, COL4A4 또는 COL4A5 폴리펩티드, 또는 이의 단편 또는 유도체를 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 21-23 중 하나와 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 24-26 중 하나와 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다.
- [0047] 일부 구체예에서, 단백질 코딩 서열은 CFI, 또는 이의 단편 및/또는 변이체를 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 49와 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 50과 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다.
- [0048] 일부 구체예에서, 단백질 코딩 서열은 CFH, 또는 이의 단편 및/또는 변이체를 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 51과 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 52와 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다.
- [0049] 일부 구체예에서, 단백질 코딩 서열은 FHL-1, 또는 이의 단편 및/또는 변이체를 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 53과 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 인코딩한다. 일부 구체예에서, 단백질 코딩 서열은 SEQ ID NO: 54와 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95%, 적어도 98% 또는 적어도 99% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다.
- [0050] 일부 구체예에서, 단백질 코딩 서열은 전사-후 조절 요소 및/또는 폴리아데닐화 서열과 같은 하나 이상의 추가 조절 요소에 작동가능하게 연결된다.
- [0051] 일 양태에서, 본 발명은 본 발명에 따른 폴리뉴클레오티드를 포함하는 벡터를 제공한다. 적합하게는, 벡터는 발세포를 형질도입할 수 있고, 임의적으로 벡터는 발세포를 특이적으로 형질도입할 수 있다.
- [0052] 적합하게는, 벡터는 바이러스 벡터, 예컨대, 아데노-관련 바이러스(AAV) 벡터, 렌티바이러스 벡터, 레트로바이러스 벡터, 아데노바이러스 벡터, 아데노-관련 바이러스 벡터, 단순 포진 바이러스 벡터, 알파바이러스 벡터, 플라비바이러스 벡터, 람다바이러스 벡터, 홍역 바이러스 벡터, 뉴캐슬병 바이러스 벡터, 폭스바이러스 벡터 및 피코르나바이러스 벡터이고, 바람직하게는 벡터는 AAV 벡터이다. 적합하게는, 바이러스 벡터는 바이러스 벡터 입자의 형태이고, 바람직하게는 바이러스 벡터는 AAV 벡터 입자의 형태이다.
- [0053] 일부 구체예에서, 벡터는 AAV3B, LK03 또는 AAV9 캡시드 단백질에 의해 캡시드화된 AAV 벡터 입자의 형태이다.
- [0054] 일 양태에서, 본 발명은 본 발명에 따른 폴리뉴클레오티드 또는 본 발명에 따른 벡터를 포함하는 세포를 제공한다.
- [0055] 일 양태에서, 본 발명은 본 발명에 따른 폴리뉴클레오티드, 본 발명에 따른 벡터 또는 본 발명에 따른 세포를 포함하는 약학적 조성물을 제공한다.
- [0056] 일 양태에서, 본 발명은 의약에 사용하기 위한 본 발명에 따른 폴리뉴클레오티드, 본 발명에 따른 벡터 또는 본

발명에 따른 세포를 제공한다.

[0057] 일 양태에서, 본 발명은 코딩 서열의 발현을 유도하기 위한 본 발명에 따른 프로모터의 용도를 제공한다. 발현은 신장-특이적일 수 있다. 적합하게는, 발현은 발세포-특이적이다.

**도면의 간단한 설명**

[0058] 도면의 설명

**도 1 - 최소 네프린 프로모터의 개략도**

(A) 예시적인 전장 네프린 프로모터는 길이가 1249 bp일 수 있고(시작 코돈 제외), 이하 "FL" 네프린 프로모터로 지칭된다. (B) 5' 영역이 결실된 예시적인 최소 네프린 프로모터는 길이가 819 bp일 수 있으며(시작 코돈 제외), 이하 "미디" 네프린 프로모터로 지칭된다. (C) 5' 영역이 결실되고, 중앙 영역이 결실된 예시적인 최소 네프린 프로모터는 길이가 265 bp일 수 있으며(시작 코돈 제외), 이하 "미니" 네프린 프로모터로 지칭된다. (D) 네프린 프로모터의 하기 영역이 표시된다: (i) 인간 마우스 상동성 영역, (ii) 레티노산 수용체(RAR) 결합 부위, (iii) WT1 결합 부위, (iv) 인핸서 박스, (v) 전사 인자 결합 영역 및 (vi) 전사 개시 부위.

**도 2 - 미디 네프린 프로모터에 작동가능하게 커플링된 GFP를 포함하는 렌티바이러스 벡터의 개략도**

(A) pACE\_hNPHS1 프로모터를 주형으로 사용하여 BamHI 및 ClaI 제한 부위를 도입하였다. (B) 미디 네프린 프로모터에 작동가능하게 커플링된 GFP를 포함하는 최종 작제물 벡터.

**도 3 - 미니 네프린 프로모터에 작동가능하게 커플링된 GFP를 포함하는 렌티바이러스 벡터의 개략도**

(A) pACE\_hNPHS1 프로모터를 PCR에 대한 주형으로 사용하고 프로모터의 두 섹션을 겔 추출하였다. (B) 미니 네프린 프로모터에 작동가능하게 커플링된 GFP를 포함하는 최종 작제물 벡터.

**도 4 - 렌티바이러스 벡터로 형질도입한 후 Ci발세포에서 GFP의 발현**

GFP-태깅된 네프린 프로모터를 안정적으로 발현하는 인간 Ci발세포를 렌티바이러스 접근법을 사용하여 생성하였다. GFP 발현을 형광 현미경으로 관찰하였다. (A) 형질도입되지 않은 Ci발세포. (B) GFP-태깅된 미니 네프린 프로모터를 안정적으로 발현하는 Ci발세포. (C) GFP-태깅된 FL 네프린 프로모터를 안정적으로 발현하는 CiPodocyte.

**도 5 - 렌티바이러스 벡터로의 형질도입 후 분화된 ci발세포에서 GFP의 발현**

네프린 프로모터에 작동가능하게 연결된 GFP를 포함하는 렌티바이러스 벡터를 분화된 조건부 불멸화 발세포(ci 발세포)로 형질도입하였다. 면역침전(IP)을 사용하여 GFP 발현을 검출하였다.

**도 6 - 렌티바이러스로 형질도입된 인간 사구체 세포 - GFP.네프린 프로모터(FL 또는 265)**

Novocyte Analyser를 사용하여 조건부 불멸화된 인간 발세포(LY) 및 사구체 내피 세포(GEnC)의 모든 살아있는 싱글렛의 중앙값 GFP 형광(AFU)을 나타내는 FACS 분석. 형질도입되지 않은 세포(세포 대조군)를 전장 인간 네프린 프로모터(hNPHS1.GFP) 또는 미니 인간 네프린 프로모터(265.GFP)에 의해 제어되는 GFP 발현 카세트를 보유하는 렌티바이러스 작제물로 형질도입된 세포와 비교하였다. 모든 세포를 10일 동안 분화시키고, 트립신 처리하고 (100 uL), PBS, 2% FBS, 1:1000 DRAQ7(150uL)에 희석하였다. 데이터 및 오차 막대는 3개의 기술적 반복(100 uL, >2500개 세포) ±SEM을 나타낸다.

**도 7 - 미니 네프린 프로모터에 커플링된 COL4A3, COL4A4 및 COL4A5를 포함하는 AAV 전달 플라스미드**

(A) 미니 네프린 프로모터에 커플링된 COL4A3을 포함하는 AAV 플라스미드인 pAAV.265.Col14a3.3flag.sv40에 대한 도식. SmaI 부위가 제시되고, SmaI에 의한 제한 후 다음 단편이 예상된다: 1. 6238 bp, 2. 2753 bp, 3. 56 bp, 4. 11 bp, 5. 11 bp. (B) 미니 네프린 프로모터에 커플링된 COL4A4을 포함하는 AAV 플라스미드인 pAAV.265.Col14a4.3flag.sv40에 대한 도식. SmaI 부위가 제시되고, SmaI에 의한 제한 후 다음 단편이 예상된다: 1. 4052 bp, 2. 2753 bp, 3. 2224 bp, 4. 56 bp, 5. 11 bp, 6. 11 bp. (C) 미니 네프린 프로모터에 커플링된 COL4A5을 포함하는 AAV 플라스미드인 pAAV.265.Col14a5.3flag.sv40에 대한 도식. SmaI 부위가 제시되고, SmaI에 의한 제한 후 다음 단편이 예상된다: 1. 4272 bp, 2. 2753 bp, 3. 2032 bp, 4. 56 bp, 5. 11 bp, 6. 11 bp. (D) SmaI를 이용한 제한 분해. MW = 1 Kb DNA 래더, 레인 1, 2 및 3은 각각 (A), (B) 및 (C)에 제시된 플라스미드에 대한 분해물에 상응한다. (E) 제한 분해를 나타내는 도식.

**도 8 - AAV.COL4.nephrin265 Sv40 바이러스로 형질도입된 발세포.**

(A) 항-FLAG 항체로 풀다운된 인간 분화된 Ci 발세포(조건부 불멸화)에서 전장 FLAG-태깅된 Col4a3(LK03) 또는 Col4a5(LK03)의 면역침전 실험. 항-FLAG 항체는 Col4a3 및 Col4a5 둘 모두를 침전시켰다. 인간-FLAG IgG를 대조군으로 사용하였다. (B) 인간 또는 마우스 분화된 Ci 발세포에서 Col4a3(LK03) 캡시드 혈청형), Col4a5(LK03) 및 Col4a5(2/9) 캡시드 혈청형)의 발현 수준을 나타내는 단백질 용해물의 웨스턴 블롯. 감염되지 않은 인간 및 마우스 Ci 발세포를 대조군으로 사용하였다. (C) F-액틴을 갖는 인간 야생형 Ci 발세포/Col4a5 3xFlag AAV Ci 발세포에서 형질도입된 Col4a5의 면역형광 염색을 나타내는 공초점 이미지. Col4a5는 야생형 대응물과 비교하여 Col4a5 3xFlag AAV 바이러스로 감염된 인간 분화된 발세포에서 세포질 수준으로 존재한다.

**도 9 - HEK 세포에서 CFH, CFI 및 CFHL1의 발현**

293T 인간 배아 신장 세포에서 웨스턴 블롯에 의한 단백질 발현 분석. 세포를 pAAV-265-CFH, pAAV-FL-CFI 또는 pAAV-FL-CFHL1로 형질감염시켰다. NT(형질감염되지 않은 293T HEK 세포). 각각의 전이유전자의 발현은 단백질-특이적 항체 또는 항-FLAG-태그 또는 항-MYC-태그 항체를 사용하여 세포 용해물 및/또는 배지에서 입증된다. 265bp 최소 네프린 프로모터의 제어하에 CFH 전이유전자를 발현하는 플라스미드로 형질감염된 HEK293T 세포는 인간 인자 H를 발현하고 분비하였다.

**도 10 - AAV2/9 265-CFH에 의한 인자 H 돌연변이된 발세포("인간 초기 질환(ED) 발세포")의 형질도입 또는 CFH를 인코딩하는 플라스미드로의 형질감염**

(A) ELISA 검정을 이용한 인간 인자 H 농도의 분석. CFH 전이유전자를 함유하는 AAV2/9 바이러스로 형질도입된 발세포는 형질도입되지 않은 대조군보다 배양 배지에서 더 높은 농도의 인간 인자 H를 나타내었다. (B) (A)로부터의 평균 인간 인자 H 농도. (C) ELISA 검정을 이용한 인간 인자 H 농도의 분석. 265bp 최소 네프린 프로모터의 제어 하에 CFH 전이유전자를 발현하는 플라스미드로 형질감염된 발세포는 형질감염되지 않은 대조군보다 더 높은 농도의 인간 인자 H를 나타내었다.

**도 11 - WT 마우스의 신장에서 CFH의 발현**

염수 대조군으로 주사된 마우스와 비교하여 pAAV.NPHS1(265).hCFH.WPRE.bGH가 주사된 야생형 마우스에서 (A) 바이러스 ITR, (B) 인간 CFH cDNA 및 (C) 바이러스 입자의 증가를 나타내는 유전자 발현 프로파일. (D) 식염수 대조군(m1) 또는 AAV 유전자 요법 생성물(pAAV.NPHS1(265).hCFH.WPRE.bGH, n=3)이 주사된 야생형 C57BL6 마우스로부터 유래된 신장 절편의 면역형광 영상화(m2/ m3/m4). 화살표는 마우스 사구체에서 네프린과 CFH의 공동-국소화를 나타낸다. 줌 = 20x.

**도 12 - 신장 세포에서 265bp v 818bp v FL 최소 네프린 프로모터(플라스미드 형질감염)를 사용한 GFP의 발현**

eGFP를 유도하는 전장(PS0281), 818bp(PS0301) 및 265bp(PS0282) 최소 네프린 프로모터를 함유하는 플라스미드로 형질감염된 세포에서의 eGFP의 발현. (A) 형질감염은 HEK293T 세포를 발생시킨다. (B) 형질감염은 근위 튜브 상피 세포(PTEC)를 발생시킨다. (C) 형질감염은 인간 발세포를 발생시킨다. (D) 형질감염은 사구체 내피 세포(GENC)를 발생시킨다.

**도 13 - 신장 세포에서 265bp v FL 최소 프로모터(AAV 형질도입)를 사용한 GFP의 발현**

eGFP를 유도하는 전장(PS0281) 및 265bp(PS0282) 최소 네프린 프로모터를 함유하는 전이유전자 카세트를 혼입하는 AAV 입자로 형질도입된 세포에서의 eGFP의 발현. (A) 형질도입은 인간 발세포를 발생시킨다. (B) 형질도입은 근위 튜브 상피 세포(PTEC)를 발생시킨다. (C) 형질도입은 사구체 내피 세포(GENC)를 발생시킨다. (D) 형질도입은 혈관사이 세포를 발생시킨다.

**도 14 - 265bp v FL 최소 네프린 프로모터를 사용한 인간 포도신의 발현**

(A) Lenti-X 293T 인간 배아 신장 세포에서 FAC에 의한 단백질 발현 분석. 세포를 1:3의 비율로 PEI를 갖는 3ug의 플라스미드로 형질감염시켰다. 각각의 플라스미드는 전장(FL.NPHS1-Podocin-HA) 또는 265bp(265.hNPHS1-Podocin-HA) 최소 네프린 프로모터에 의해 유도되는 인간 포도신-HA를 발현한다. 인간 포도신의 발현은 형질감염 후 2일째에 항-HA 항체에 의해 입증된다. 평균 값 ± SD로 표현된 N=3 데이터. (B) 인간 발세포에서 FAC에 의한 단백질 발현 분석. 세포를 전장(FL.NPHS1-Podocin-HA) 또는 265bp(265.hNPHS1-Podocin-HA) 최소 네프린 프로모터에 의해 유도된 인간 포도신-HA 전이유전자 카세트가 통합된 AAV 입자로 형질도입시켰다. 인간 포도신의 발현은 형질도입 및 분화 후 9일에 항-HA 항체에 의해 입증된다.

**도 15 - AAV 형질도입을 통한 265bp v FL 최소 네프린 프로모터를 사용한 GFP의 생체내 발현**

WT 마우스(C57/BI6)에서 생체내 265- 및 FL- 네프린 프로모터의 평가. 전장(FL-GFP) 또는 265bp(265-GFP) 최소 네프린 프로모터의 제어 하에 GFP를 발현하는 AAV2/9 입자를 생성하였다. 마우스 당 ~1x10e13 AAV2/9의 꼬리 정맥 주사(처리군 당 2x 마우스 및 1x 미처리 마우스). 신장의 조직 수확은 주사 후 4주에 발생하였다. qPCR 분석을 수행하고 GFP 복사체 수를 18S(하우스키핑 유전자)로 표준화하였다.

**발명을 실시하기 위한 구체적인 내용**

- [0059] **상세한 설명**
- [0060] 본 발명의 다양한 바람직한 특징 및 구체에는 이제 비제한적인 실시예에 의해 기술될 것이다.
- [0061] 본원 및 첨부된 청구범위에서 사용되는 바와 같은 단수 형태("a", "an" 및 "the")는 문맥에서 명백히 달리 지시하지 않는 한 복수의 지시대상을 포함한다는 점에 유의해야 한다.
- [0062] 본원에서 사용되는 용어 "포함하는", "포함한다" 및 "포함하다"은 "포함하는", "포함한다" 또는 "함유하는", "함유한다"와 동의어이고, 포괄적이거나 개방적이며, 추가적인 비-인용된 구성원, 요소 또는 방법 단계를 배제하지 않는다. 용어 "포함하는", "포함한다" 및 "포함하다"는 또한 용어 "로 구성된"을 포함한다.
- [0063] 본원에서 논의된 간행물은 본 출원의 출원일 이전에 이들의 개시를 위해서만 제공된다. 본원의 어떠한 것도 그러한 간행물이 본원에 첨부된 청구범위에 대한 선행 기술을 구성한다는 것을 인정하는 것으로 해석되어서는 안 된다.
- [0064] 본 개시는 본원에 개시된 예시적인 방법 및 물질에 의해 제한되지 않으며, 본원에 기재된 것과 유사하거나 등가인 임의의 방법 및 물질이 본 개시의 구체예의 실시 또는 시험에서 사용될 수 있다. 숫자 범위는 범위를 규정하는 숫자를 포함한다. 달리 지시되지 않는 한, 임의의 핵산 서열은 5'에서 3' 배향으로 왼쪽에서 오른쪽으로 기록되고; 아미노산 서열은 각각 아미노에서 카르복시 배향으로 왼쪽에서 오른쪽으로 기록된다.
- [0065] 본 발명의 실시는 달리 지시되지 않는 한, 당업자의 능력 내에 있는 화학, 생화학, 분자 생물학, 미생물학 및 면역학의 통상적인 기술을 이용할 것이다. 이러한 기술은 문헌에 설명되어 있다. 예를 들어, 하기 문헌 참조: Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd Edition, Cold Spring Harbor Laboratory Press; Ausubel, F.M. et al. (1995 and periodic supplements) *Current Protocols in Molecular Biology*, Ch. 9, 13 and 16, John Wiley & Sons; Roe, B., Crabtree, J. and Kahn, A. (1996) *DNA Isolation and Sequencing: Essential Techniques*, John Wiley & Sons; Polak, J.M. and McGee, J.O'D. (1990) *In Situ Hybridization: Principles and Practice*, Oxford University Press; Gait, M.J. (1984) *Oligonucleotide Synthesis: A Practical Approach*, IRL Press; and Lilley, D.M. and Dahlberg, J.E. (1992) *Methods in Enzymology: DNA Structures Part A: Synthesis and Physical Analysis of DNA*, Academic Press. 이러한 일반적인 텍스트 각각은 본원에 참조로 포함된다.
- [0066] **프로모터**
- [0067] 본 발명은 사구체 유전자 요법을 위한 프로모터를 제공한다.
- [0068] "프로모터"는 DNA 서열로부터 RNA의 전사를 조절 및/또는 개시하는 DNA의 서열을 지칭하는 통상적인 의미에 따라 본원에서 사용된다.
- [0069] 적합하게는, 프로모터는 포유동물 세포, 예를 들어, 인간 세포에서 작동가능하다. 프로모터는 포유동물 세포, 예를 들어, 인간 세포에서 전이유전자 발현을 유도할 수 있다. 프로모터는 포유동물 프로모터, 예를 들어, 인간 프로모터일 수 있다.
- [0070] 프로모터는 신장 세포에서 작동가능할 수 있다. 본 발명의 프로모터는 신장에서 전이유전자 발현을 유도할 수 있다. 프로모터가 작동할 수 있는 신장 세포의 예는 사구체 세포를 포함하나, 이에 제한되지 않는다.
- [0071] 프로모터는 사구체 세포에서 작동가능할 수 있다. 본 발명의 프로모터는 사구체에서 전이유전자 발현을 유도할 수 있다. 성숙한 사구체는 4개의 세포 유형을 함유한다: 보우만 캡슐을 형성하는 정수리 상피 세포, 사구체 여과 장벽의 최외층을 덮는 발세포, 혈액과 직접 접촉하는 글리코칼릭스-코팅된 창내피세포, 및 모세관 루프 사이에 위치하는 혈관사이 세포 (Vaughan, M.R. and Quaggin, S.E., 2008. *Journal of the American Society of Nephrology*, 19(1), pp.24-33).

- [0072] 프로모터는 발세포에서 작동가능할 수 있다. 본 발명의 프로모터는 발세포에서 전이유전자 발현을 유도할 수 있다.
- [0073] 본 발명의 프로모터는 조직-특이적 프로모터일 수 있다. 본원에서 사용되는 "조직-특이적 프로모터"는 특정 유형의 세포 또는 조직에서 전이유전자의 발현을 우선적으로 촉진하는 프로모터이다. 적합하게는, 조직-특이적 프로모터는 다른 세포-유형과 비교하여 한 세포-유형에서 전이유전자의 더 높은 발현을 촉진할 수 있다. 예를 들어, 조직-특이적 프로모터는 다른 세포-유형에서의 발현 수준과 비교하여 한 세포-유형에서 적어도 10% 더 높거나, 적어도 20% 더 높거나, 적어도 30% 더 높거나, 적어도 40% 더 높거나, 적어도 50% 더 높거나, 적어도 100% 더 높거나, 적어도 200% 더 높거나, 적어도 300% 더 높거나, 적어도 400% 더 높거나, 적어도 500% 더 높거나, 적어도 1000% 더 높은 전이유전자 발현 수준을 조장하는 프로모터일 수 있다.
- [0074] 적합하게는, 프로모터는 신장-특이적 프로모터이다. 일부 구체예에서, 프로모터는 사구체-특이적 프로모터이다. 일부 구체예에서, 프로모터는 발세포-특이적 프로모터이다.
- [0075] 전이유전자 발현은 당 분야에 공지된 임의의 적합한 방법에 의해 측정될 수 있다. 예를 들어, 프로모터에 작동가능하게 연결된 리포터 전이유전자, 예를 들어, 녹색 형광 단백질(GFP)의 발현을 측정함으로써 측정될 수 있는데, 여기서 리포터 전이유전자의 발현은 유전자의 발현을 조장하는 프로모터의 능력과 상관관계가 있다. 리포터 전이유전자, 예를 들어, GFP의 발현은 임의의 적합한 방법, 예를 들어, FACS에 의해 결정될 수 있다. 예를 들어, 신장-특이적 프로모터는 다른 세포 유형(예를 들어, CNS, 망막, 폐, 췌장, 심장 또는 근육 세포)과 비교하여 신장 세포에서 리포터 전이유전자의 더 높은 발현을 조장할 수 있다. 예를 들어, 발세포-특이적 프로모터는 다른 세포-유형과 비교하여 조건부 불멸화 발세포에서 리포터 전이유전자의 더 높은 발현을 조장할 수 있다. 적합한 발세포 세포주 예를 들어, CIHP-1은 당업자에게 잘 알려져 있을 것이다. 불멸화 발세포를 생성하는 방법은 당업자에게 잘 알려져 있을 것이다. 적합한 방법은 문헌[Ni, L., et al., 2012. Nephrology, 17(6), pp.525-531]에 기재되어 있다.
- [0076] 적합하게는, 프로모터는 최소의 신장-특이적 프로모터이다. 일부 구체예에서, 프로모터는 최소의 사구체-특이적 프로모터이다. 일부 구체예에서, 프로모터는 최소의 발세포-특이적 프로모터이다.
- [0077] 프로모터는 약 1.2 kb 이하의 길이를 가질 수 있다. 적합하게는, 프로모터는 약 1.18 kb 이하, 약 1.17 kb 이하, 약 1.16 kb 이하, 약 1.15 kb 이하, 약 1.14 kb 이하, 약 1.13 kb 이하, 약 1.12 kb 이하, 약 1.11 kb 이하 또는 약 1.10 kb 이하의 길이를 갖는다. 적합하게는, 프로모터는 약 1.15 kb 이하의 길이를 갖는다.
- [0078] 프로모터는 약 1.1 kb 이하의 길이를 가질 수 있다. 일부 구체예에서, 프로모터는 약 1.1 kb 이하, 1.0 kb 이하, 약 0.9 kb 이하, 약 0.8 kb 이하, 약 0.7 kb 이하, 약 0.6 kb 이하, 약 0.5 kb 이하, 약 0.4 kb 이하 또는 약 0.3 kb 이하의 길이를 갖는다.
- [0079] 일부 구체예에서, 프로모터는 약 0.8 kb 이하, 약 0.7 kb 이하, 약 0.6 kb 이하, 약 0.5 kb 이하, 약 0.4 kb 이하 또는 약 0.3 kb 이하의 길이를 갖는다. 일부 구체예에서, 프로모터는 818 bp 이하의 길이를 갖는다. 일부 구체예에서, 프로모터는 800 bp 이하의 길이를 갖는다.
- [0080] 일부 구체예에서, 프로모터는 약 0.5 kb 이하, 약 0.4 kb 이하 또는 약 0.3 kb 이하의 길이를 갖는다. 일부 구체예에서, 프로모터는 약 0.3 kb 이하의 길이를 갖는다.
- [0081] 프로모터는 약 250 bp 이상의 길이를 가질 수 있다. 일부 구체예에서, 프로모터는 약 250-1100 bp, 250-1000 bp, 250-900 bp, 250-800 bp, 250-700 bp, 250-600 bp, 250-500 bp, 250-400 bp, 250-300 bp의 길이를 갖는다.
- [0082] 프로모터는 약 265 bp 이상의 길이를 가질 수 있다. 일부 구체예에서, 프로모터는 약 265-1100 bp, 265-1000 bp, 265-900 bp, 265-800 bp, 265-700 bp, 265-600 bp, 265-500 bp, 265-400 bp, 265-300 bp의 길이를 갖는다.
- [0083] 일 구체예에서, 프로모터는 250-300 bp, 250-280 bp, 255-275 bp, 260-270 bp 또는 약 265 bp의 길이를 갖는다. 일 구체예에서, 프로모터는 800-850 bp, 800-840 bp, 810-830 bp, 815-825 bp, 약 819 bp 또는 약 818 bp의 길이를 갖는다.
- [0084] 네프린 프로모터
- [0085] 프로모터는 합성 프로모터일 수 있다. 본원에서 사용되는 "합성 프로모터"는 자연적으로 발생하지 않는 프로모



CTCTAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGTTTTTTAAAGAGACAGAGAAAGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGA  
GACACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGGAGGAGTAACGGAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGAC  
CCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGTTCACAGTAGGGGACCTGTC

[0095] 예시적인 변이 네프린 프로모터 - 1,192 bp (SEQ ID NO: 45)

[0096] CACCTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGATGAAACCCGCTCTAGTAAAAATACAAAAATTAGCCAGGCATGGTGCTATATACCTGTAG  
CACCAGCTACTTGGGAGACAGAGGTGGGAGAATTACTTGAACCTGGGAGGTTCAAGCCATGGGAGGTGGAAGTTGCAGTGAGCCGAGATGCCACTGCACCTCC  
AGCCTGAGCAACAGAGCAAGACTATCTCAAGAAAAGAAAAGAAAAGAAAGAGACTTGCCAAGGTCATGTATCAGGCAAGGAAGAGCTGGGGGCCAGCT  
GGCTGCTCCCTGCTGAGCTGGGAGACCACCTTGATCTGACTTCTCCCATCTTCCAGCCTAAGCCAGGCCCTGGGGTTCAGGAGGCTGGGGAGCCAGGAG  
GAACGCGCCTGGCATGTGCTGACAGGGGATTTATGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGCTCTGGAAGGTAC  
CTGGGGGAGGTTGCACTGTGAGAATGAGCTCAAGCTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGCTCTCCTAGGCTCTGG  
CCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAG  
AGAGAGAGAGAGACAGAGAGAGAGGAAGACAGAGACAAAAGGAGAGAAAGAGAAAGAGACTTAGACAAGGAGAGAAAAGATGGAAGATAAAGAGACTGGGGCA  
GTGGCTCACGCTGTAATCCCAACACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCCCGT  
CTCTAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGTTTTTTAAAGAGACAGAGAAAGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGA  
GATACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGGAGGAGTAACGGAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGAC  
CCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGTTCACAGTAGGGGACCTGTC

[0097] 적합하게는, 본 발명의 프로모터는 SEQ ID NO: 1 또는 SEQ ID NO: 1에 대해 적어도 70% 동일성을 갖는 변이체로 부터 유래된다. 적합하게는, 본 발명의 프로모터는 SEQ ID NO: 1 또는 SEQ ID NO: 1에 대해 적어도 70% 동일성을 갖는 변이체와 비교하여 하나 이상의 결실을 갖는다. 적합하게는, 변이체는 SEQ ID NO: 1에 대해 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다. 예시적인 변이체는 SEQ ID NO: 14, SEQ ID NO: 44 또는 SEQ ID NO: 45이다.

[0098] 적합하게는, SEQ ID NO: 1의 변이체는 하기 치환 중 하나 이상을 포함한다: A493G, C1080T, G1169C 및 C1249G. 적합하게는, SEQ ID NO: 1의 변이체는 치환 A493G, C1080T, G1169C 및 C1249G 각각을 포함한다(SEQ ID NO: 14 참조).

[0099] 적합하게는, 본 발명의 프로모터는 SEQ ID NO: 14, SEQ ID NO: 44 또는 SEQ ID NO: 45, 또는 SEQ ID NO: 14, SEQ ID NO: 44 또는 SEQ ID NO: 45에 대해 적어도 70% 동일성을 갖는 변이체로부터 유래된다. 적합하게는, 본 발명의 프로모터는 SEQ ID NO: 14, SEQ ID NO: 44 또는 SEQ ID NO: 45, 또는 SEQ ID NO: 14, SEQ ID NO: 44 또는 SEQ ID NO: 45에 대해 적어도 70% 동일성을 갖는 변이체와 비교하여 하나 이상의 결실을 갖는다. 적합하게는, 변이체는 SEQ ID NO: 14, SEQ ID NO: 44 또는 SEQ ID NO: 45에 대해 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0100] 적합하게는, 본 발명의 프로모터는 SEQ ID NO: 14 또는 SEQ ID NO: 14에 대해 적어도 70% 동일성을 갖는 변이체로부터 유래된다. 적합하게는, 본 발명의 프로모터는 SEQ ID NO: 14 또는 SEQ ID NO: 14에 대해 적어도 70% 동일성을 갖는 변이체와 비교하여 하나 이상의 결실을 갖는다. 적합하게는, 변이체는 SEQ ID NO: 14에 대해 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0101] 적합하게는, 본 발명의 프로모터는 SEQ ID NO: 1에 따르면 하나 이상의 결실, 예를 들어, 1 또는 2개의 결실을 갖는 뉴클레오티드 서열, 또는 이에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다. 적합하게는, 본 발명의 프로모터는 SEQ ID NO: 1에 따르면 2개 이상의 결실을 갖는 뉴클레오티드 서열, 또는 이에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다. 삭제는 임의의 크기일 수 있다. 적합하게는, 결실은 각각 적어도 50 bp, 적어도 100 bp, 적어도 150 bp, 적어도 200 bp, 적어도 250 bp, 적어도 300 bp, 적어도 350 bp, 또는 적어도 400 bp 크기이다. 적합하게는, 결실은 각각 50 내지 500 bp, 100 내지 500 bp, 150 내지 500 bp, 200 내지 500 bp, 250 내지 500 bp, 300 내지 500 bp, 350 내지 500 bp, 또는 400 내지 500 bp 크기이다.

[0102] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 1에 따른 뉴클레오티드 서열로서,

[0103] (i) SEQ ID NO: 1의 위치 1 내지 위치 n1 (여기서 n1은 1 내지 430의 정수임)이 결실되고/거나;

- [0104] (ii) SEQ ID NO: 1의 위치 n2 내지 위치 n3 (여기서  $n3 \geq n2$ 이고, n2는 508 내지 1061의 정수이고, n3은 508 내지 1061의 정수임)이 결실된 뉴클레오티드 서열
- [0105] 또는 이에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된 프로모터를 제공한다.
- [0106] 일부 구체예에서,
- [0107] (iii) SEQ ID NO: 1의 위치 493에 상응하는 위치가 G이고/거나 SEQ ID NO: 1의 위치 1080에 상응하는 위치가 T이고/거나 SEQ ID NO: 1의 위치 1169에 상응하는 위치가 C이고/거나 SEQ ID NO: 1의 위치 1249에 상응하는 위치는 G이다.
- [0108] 바람직한 구체예에서:
- [0109] (iii) SEQ ID NO: 1의 위치 493에 상응하는 위치가 G이고, SEQ ID NO: 1의 위치 1080에 상응하는 위치가 T이고, SEQ ID NO: 1의 위치 1169에 상응하는 위치가 C이고, SEQ ID NO: 1의 위치 1249에 상응하는 위치는 G이다.
- [0110] 상기 뉴클레오티드 위치는 SEQ ID NO: 1의 위치를 참고로 언급되지만, 당업자는 예를 들어, SEQ ID NO: 1과 변이체 서열을 정렬함으로써 이의 변이체에서 상응하는 뉴클레오티드 위치를 쉽게 확인할 수 있을 것이다. 예를 들어, SEQ ID NO: 1의 위치 493은 SEQ ID NO: 14의 위치 436에 상응하고; SEQ ID NO: 1의 위치 1080은 SEQ ID NO: 14의 위치 1023에 상응하고; SEQ ID NO: 1의 위치 1169는 SEQ ID NO: 14의 위치 1112에 상응하고; SEQ ID NO: 1의 위치 1249는 SEQ ID NO: 14의 위치 1192에 상응한다.
- [0111] 예를 들어, 본 발명의 프로모터는 SEQ ID NO: 1에 따른 뉴클레오티드 서열에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성될 수 있으며, 여기서
- [0112] (i) SEQ ID NO: 1의 위치 1 내지 위치 n1 (여기서 n1은 1 내지 430의 정수임)이 결실되고/거나;
- [0113] (ii) SEQ ID NO: 1의 위치 n2 내지 위치 n3 (여기서  $n3 \geq n2$ 이고, n2는 508 내지 1061의 정수이고, n3은 508 내지 1061의 정수임)이 결실된다.
- [0114] 일부 구체예에서,
- [0115] (iii) SEQ ID NO: 1의 위치 493이 G로 치환되고/거나 SEQ ID NO: 1의 위치 1080이 T로 치환되고/거나 SEQ ID NO: 1의 위치 1169가 C로 치환되고/거나 SEQ ID NO: 1의 위치 1249가 G로 치환된다.
- [0116] 바람직한 구체예에서,
- [0117] (iii) SEQ ID NO: 1의 위치 493이 G로 치환되고, SEQ ID NO: 1의 위치 1080이 T로 치환되고, SEQ ID NO: 1의 위치 1169가 C로 치환되고, SEQ ID NO: 1의 위치 1249가 G로 치환된다.
- [0118] 적합하게는, n1은 50 내지 430, 100 내지 430, 150 내지 430, 200 내지 430, 250 내지 430, 300 내지 430, 350 내지 430, 또는 400 내지 430의 정수이다. 일부 구체예에서, n1은 100 내지 430의 정수이다. 일부 구체예에서, n1 = 430, 즉, SEQ ID NO: 1의 위치 1 내지 위치 430은 본 발명의 프로모터에서 결실된다.
- [0119] n3과 n2의 차이는 결실 크기를 지정한다. 적합하게는,  $n3 \geq n2 + 49$ ,  $n3 \geq n2 + 99$ ,  $n3 \geq n2 + 149$ ,  $n3 \geq n2 + 199$ ,  $n3 \geq n2 + 249$ ,  $n3 \geq n2 + 299$ ,  $n3 \geq n2 + 349$ ,  $n3 \geq n2 + 399$ ,  $n3 \geq n2 + 449$ ,  $n3 \geq n2 + 499$  또는  $n3 \geq n2 + 549$ 이다. 일부 구체예에서,  $n3 \geq n2 + 49$ 이다.
- [0120] n2 및 n3이 취하는 값은 결실 위치를 결정한다. 적합하게는, n2 및 n3은 각각 550 내지 1050의 정수이고, n2 및 n3은 각각 600 내지 1000의 정수이고, n2 및 n3은 각각 650 내지 950의 정수이고, n2 및 n3은 각각 700 내지 900의 정수이고, n2 및 n3은 750 내지 850의 정수이다.
- [0121] 일부 구체예에서, n2 = 508 및 n3 = 1061, 즉, SEQ ID NO: 1의 위치 508 내지 1061이 본 발명의 프로모터에서 결실된다.
- [0122] 프로모터 영역
- [0123] 본 발명자들은 전이유전자 발현을 유도하는 네프린 프로모터의 영역을 결정하였다.

- [0124] 프로모터는 전형적으로 "코어" 및 "근위" 영역을 포함한다. "코어 프로모터 영역"은 전사 시작 부위, RNA 폴리머라제 결합 부위 및 일반적인 전사 인자 결합 부위를 포함할 수 있다. "근위 프로모터 영역"은 예를 들어, 효과적이고 제어가능한 전사를 조장하는데 필요한 일차 조절 요소 및 특정 전사 인자 결합 부위를 포함할 수 있다. 코어 및 근위 프로모터 영역 둘 모두의 크기 및 성분은 전형적으로 유전자 특이적 방식으로 다양하다. 프로모터는 또한 코어 프로모터 영역의 다운스트림 및 개시 코돈의 업스트림에 5' 비번역 영역(5' UTR)(리더 서열로도 공지됨)을 포함할 수 있다. (예를 들어, 도 1 참조).
- [0125] 프로모터는 하이브리드 프로모터일 수 있다. 본원에서 사용되는 "하이브리드 프로모터"는 상이한 프로모터로부터 유래된 요소의 조합을 포함한다. 예를 들어, 하이브리드 프로모터는 요망되는 전이유전자 발현을 달성하기 위해 하나의 기존 프로모터로부터 유래된 근위 프로모터 영역 및 또 다른 기존 프로모터로부터의 코어 프로모터를 포함할 수 있다. 근위 하이브리드 프로모터는 문헌[Piekarowicz, K., et al. (2019). *Methods & clinical development*, 15, 157-169]에 기술되어 있다.
- [0126] 일부 구체예에서, 본 발명의 프로모터는 (i) SEQ ID NO: 4로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 4와 적어도 70% 동일한 변이체를 포함한다. 이론으로 국한시키려는 것은 아니지만, SEQ ID NO: 4에 대해 적어도 약 70%의 동일성을 갖는 뉴클레오티드 서열은 근위 프로모터 영역을 제공할 수 있는 것으로 고려된다.
- [0127] *예시적인 근위 프로모터 영역(SEQ ID NO: 4)*
- [0128] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGAATTTTATGCTCCAG
- [0129] 적합하게는, 변이체는 SEQ ID NO: 4와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다. 프로모터는 SEQ ID NO: 17로 제시된 SEQ ID NO: 4의 변이체를 포함할 수 있다. 적합하게는, SEQ ID NO: 4의 변이체는 치환 A63G를 포함한다(SEQ ID NO: 17 참조).
- [0130] 일부 구체예에서, 본 발명의 프로모터는 (i) SEQ ID NO: 17로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 17과 적어도 70% 동일한 변이체를 포함한다. 적합하게는, 변이체는 SEQ ID NO: 17과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0131] *예시적인 근위 프로모터 변이 영역(SEQ ID NO: 17)*
- [0132] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTATGCTCCAG
- [0133] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 5 또는 18로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 5 또는 18과 적어도 70% 동일한 변이체; SEQ ID NO: 6 또는 19로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 6 또는 19와 적어도 70% 동일한 변이체; 및/또는 SEQ ID NO: 7 또는 20으로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 7 또는 20과 적어도 70% 동일한 변이체를 포함한다.
- [0134] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 5로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 5와 적어도 70% 동일한 변이체를 포함한다. 이론으로 국한시키려는 것은 아니지만, SEQ ID NO: 5에 대해 적어도 약 70%의 동일성을 갖는 뉴클레오티드 서열은 코어 프로모터 영역을 제공할 수 있는 것으로 고려된다.
- [0135] *예시적인 코어 프로모터 영역(SEQ ID NO: 5)*
- [0136] GAGCAAGACAGAGAGACTCACAGGGAAG
- [0137] 적합하게는, 변이체는 SEQ ID NO: 5와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다. 프로모터는 SEQ ID NO: 18로 제시된 SEQ ID NO: 5의 변이체를 포함할 수 있다. 적합하게는, SEQ ID NO: 5의 변이체는 치환 C19T를 포함한다(SEQ ID NO: 18 참조).
- [0138] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 18로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 18과 적어도 70% 동일한 변이체를 포함한다. 적합하게는, 변이체는 SEQ ID NO: 18과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0139] *예시적인 코어 프로모터 변이 영역(SEQ ID NO: 18)*
- [0140] GAGCAAGACAGAGAGATACTCACAGGGAAG

[0141] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 6으로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 6과 적어도 70% 동일한 변이체를 포함한다. 이론으로 국한시키려는 것은 아니지만, SEQ ID NO: 6에 대해 적어도 약 70%의 동일성을 갖는 뉴클레오티드 서열은 5'UTR을 제공할 수 있는 것으로 고려된다.

[0142] 예시적인 5'UTR(SEQ ID NO: 6)

[0143] AGGGGAAGAGGAAAACGAGAAAGGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGTCACAGTAGGGGACCTGTC

[0144] 적합하게는, 변이체는 SEQ ID NO: 6과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다. 프로모터는 SEQ ID NO: 19로 제시된 SEQ ID NO: 6의 변이체를 포함할 수 있다. 적합하게는, SEQ ID NO: 6의 변이체는 치환 G76C 및/또는 C156G를 포함한다(SEQ ID NO: 19 참조).

[0145] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 19로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 19와 적어도 70% 동일한 변이체를 포함한다. 적합하게는, 변이체는 SEQ ID NO: 19와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0146] 예시적인 변이 5'UTR(SEQ ID NO: 19)

[0147] AGGGGAAGAGGAAAACGAGAAAGGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGTCACAGTAGGGGACCTGTG

[0148] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 7로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 7과 적어도 70% 동일한 변이체를 포함한다. 이론으로 국한시키려는 것은 아니지만, SEQ ID NO: 7에 대해 적어도 약 70%의 동일성을 갖는 뉴클레오티드 서열은 코어 프로모터 영역 및 5'UTR을 제공할 수 있는 것으로 고려된다.

[0149] 예시적인 코어 프로모터 영역 및 5'UTR(SEQ ID NO: 7)

[0150] GAGCAAGACAGAGAGAGACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAAGGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGTCACAGTAGGGGACCTGTC

[0151] 적합하게는, 변이체는 SEQ ID NO: 7과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다. 프로모터는 SEQ ID NO: 20으로 제시된 SEQ ID NO: 7의 변이체를 포함할 수 있다. 적합하게는, SEQ ID NO: 7의 변이체는 치환 C19T, G108C 및/또는 C188G를 포함한다(SEQ ID NO: 20 참조). 적합하게는, SEQ ID NO: 7의 변이체는 치환 C19T, G108C 및 C188G를 포함한다(SEQ ID NO: 20 참조).

[0152] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 20으로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 20과 적어도 70% 동일한 변이체를 포함한다. 적합하게는, 변이체는 SEQ ID NO: 20에 대해 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0153] 예시적인 코어 프로모터 변이 영역 및 5'UTR(SEQ ID NO: 20)

[0154] GAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAAGGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGTCACAGTAGGGGACCTGTG

[0155] 일부 구체예에서, 본 발명의 프로모터는 (iii) SEQ ID NO: 8에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열, 또는 이의 하나 이상의 단편을 포함한다. 적합하게는, 프로모터는 근위 프로모터 영역의 바로 다운스트림 및/또는 코어 프로모터 영역의 바로 업스트림에 SEQ ID NO: 8에 대해 적어도 약 70% 동일성을 갖는 뉴클레오티드 서열 또는 이의 하나 이상의 단편을 포함한다.

[0156] 프로모터는 SEQ ID NO: 8에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일성을 갖는 뉴클레오티드 서열 또는 이의 하나 이상의 단편을 포함할 수 있다. 프로모터는 SEQ ID NO: 8의 뉴클레오티드 서열, 또는 이의 하나 이상의 단편을 포함할 수 있다.

[0157] 예시적인 선택적 프로모터 영역(SEQ ID NO: 8)

[0158] CTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGAGGTTGCACTGTGAGAAATGAGCTCAAGCTGGGTCA



도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다.

- [0176] 일부 구체예에서, 본 발명의 프로모터는 5'에서 3'으로
- [0177] (i) SEQ ID NO: 4 또는 17에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열;
- [0178] (iii) 임의적으로 SEQ ID NO: 8에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열 또는 이의 하나 이상의 단편; 및
- [0179] (ii) SEQ ID NO: 5 또는 18에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열 및/또는 SEQ ID NO: 7 또는 20에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다.
- [0180] 일부 구체예에서, 본 발명의 프로모터는 5'에서 3'으로
- [0181] (i) SEQ ID NO: 4 또는 17에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열;
- [0182] (iii) 임의적으로 (a) SEQ ID NO: 8의 5' 말단 단편에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열; 및/또는 (b) SEQ ID NO: 8의 3' 말단 단편에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열; 및
- [0183] (ii) SEQ ID NO: 5 또는 18에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열 및/또는 SEQ ID NO: 7 또는 20에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다.
- [0184] 일부 구체예에서, 본 발명의 프로모터는 5'에서 3'으로
- [0185] (i) SEQ ID NO: 4 또는 17에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열;
- [0186] (iii) 임의적으로 (a) SEQ ID NO: 8의 5' 말단 단편에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열; 및/또는 (b) SEQ ID NO: 8의 3' 말단 단편에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열; 및
- [0187] (ii) SEQ ID NO: 5 또는 18에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열 및/또는 SEQ ID NO: 7 또는 20에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다.
- [0188] 프로모터 요소
- [0189] 본 발명자들은 전이유전자 발현을 유도하는 네프린 프로모터의 기능성 요소를 결정하였다.
- [0190] 본 발명의 프로모터는 (a) 레티노산 수용체 결합 부위; (b) WT1 결합 부위; (c) 인핸서 박스; (d) 전사 인자 결합 영역; 및 (e) 전사 개시 부위 중 하나 이상을 포함할 수 있다.
- [0191] 적합하게는, 본 발명의 프로모터는 (a) 레티노산 수용체 결합 부위; (b) WT1 결합 부위; (c) 인핸서 박스; (d) 전사 인자 결합 영역; 및 (e) 전사 개시 부위 전부를 포함할 수 있다.
- [0192] 레티노산 수용체(RAR) 결합 부위는 RAR 알파, RAR 베타 및/또는 RAR 감마에 결합할 수 있는 폴리뉴클레오티드 서열을 지칭한다. RAR 결합 부위는 SEQ ID NO: 10으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 10과 비교하여 1 또는 2개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성될 수 있다. 치환, 결실 또는 삽입은 RAR 결합 부위가 이의 내인성 기능 중 적어도 하나를 유지하도록 하는 단일 뉴클레오티드의 임의의 치환, 결실 또는 삽입일 수 있다.
- [0193] 예시적인 RAR 결합 부위(SEQ ID NO: 10)
- [0194] GGGGTCA

- [0195] WT1 결합 부위는 윌름스 증양 억제인자 유전자, WT1에 의해 인코딩된 징크 핑거 폴리펩티드에 결합할 수 있는 폴리뉴클레오티드 서열을 지칭한다. WT1 결합 부위는 SEQ ID NO: 11로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 11과 비교하여 1, 2 또는 3개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성될 수 있다. 치환, 결실 또는 삽입은 WT1 결합 영역이 이의 내인성 기능 중 적어도 하나를 유지하도록 하는 단일 뉴클레오티드의 임의의 치환, 결실 또는 삽입일 수 있다.
- [0196] *예시적인 WT1 결합 부위(SEQ ID NO: 11)*
- [0197] CGGAGGCTGGGGAGGCA
- [0198] 인헨서 박스는 단백질-결합 부위로서 작용하는 일부 진핵생물에서 발견되는 DNA 반응 요소를 지칭한다. 인헨서 박스는 SEQ ID NO: 12로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 12와 비교하여 1 또는 2개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성될 수 있다. 치환, 결실 또는 삽입은 인헨서 박스가 이의 내인성 기능 중 적어도 하나를 유지하도록 하는 단일 뉴클레오티드의 임의의 치환, 결실 또는 삽입일 수 있다.
- [0199] *예시적인 인헨서 박스(SEQ ID NO: 12)*
- [0200] ATGTG
- [0201] (a) 레티노산 수용체 결합 부위; (b) WT1 결합 부위; 및 (c) 인헨서 박스 중 하나 이상은 근위 프로모터 영역에 존재할 수 있다. 적합하게는, (a) 레티노산 수용체 결합 부위; (b) WT1 결합 부위; 및 (c) 인헨서 박스 각각은 근위 프로모터 영역에 존재한다.
- [0202] 일부 구체예에서, 하기 요소 중 하나 이상은 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열에 존재한다: (a) SEQ ID NO: 4 또는 17의 위치 7 내지 위치 13에 대략적으로 상응하는 위치에서 RAR 결합 부위; (b) SEQ ID NO: 4 또는 17의 위치 14 내지 위치 30에 대략적으로 상응하는 위치에서 WT1 결합 부위; 및 (c) SEQ ID NO: 4 또는 17의 위치 49 내지 위치 53에 대략적으로 상응하는 위치에 인헨서 박스. 일부 구체예에서, 요소 각각은 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열에 존재한다.
- [0203] 일부 구체예에서, 하기 뉴클레오티드 서열 중 하나 이상은 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열에 존재한다: (a) SEQ ID NO: 4 또는 17의 위치 7 내지 위치 13에 상응하는 위치에서 GGGGTCA; (b) SEQ ID NO: 4 또는 17의 위치 14 내지 위치 30에 상응하는 위치에서 CGGAGGCTGGGGAGGCA; 및 (c) SEQ ID NO: 4 또는 17의 위치 49 내지 위치 53에 상응하는 위치에서 ATGTG. 일부 구체예에서, 뉴클레오티드 서열 각각은 (i) SEQ ID NO: 4 또는 17에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열에 존재한다.
- [0204] 적합하게는, 프로모터는 SEQ ID NO: 13으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 13과 비교하여 1, 2, 3, 4 또는 5개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는 전사 인자 결합 영역을 포함할 수 있다. 치환, 결실 또는 삽입은 전사 인자 결합 영역이 이의 내인성 기능 중 적어도 하나를 유지하도록 하는 단일 뉴클레오티드의 임의의 치환, 결실 또는 삽입일 수 있다. 예시적인 변이체는 SEQ ID NO: 43에 제시되어 있다. 적합하게는, SEQ ID NO: 13의 변이체는 치환 C19T를 포함한다(SEQ ID NO: 43 참조).
- [0205] *예시적인 전사 인자 결합 영역(SEQ ID NO: 13)*
- [0206] GAGCAAGACAGAGAGAGACTCACAGGGA
- [0207] 적합하게는, 프로모터는 SEQ ID NO: 43으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 43과 비교하여 1, 2, 3, 4 또는 5개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는 전사 인자 결합 영역을 포함할 수 있다. 치환, 결실 또는 삽입은 전사 인자 결합 영역이 이의 내인성 기능 중 적어도 하나를 유지하도록 하는 단일 뉴클레오티드의 임의의 치환, 결실 또는 삽입일 수 있다.
- [0208] *예시적인 전사 인자 결합 영역(SEQ ID NO: 43)*
- [0209] GAGCAAGACAGAGAGATACTCACAGGGA
- [0210] 다른 적합한 전사 인자 결합 영역은 당업자에게 잘 알려져 있을 것이다. 예를 들어, 다른 적합한 전사 인자 결합 영역은 TACGAT(SEQ ID NO: 37), TATAAT(SEQ ID NO: 38), GATACT(SEQ ID NO: 39), TATGAT(SEQ ID NO: 40) 및 TATGTT(SEQ ID NO: 41)를 포함한다.

- [0211] 적합하게는, 프로모터는 "AG" 디뉴클레오티드를 포함하거나 이로 구성된 전사 개시 부위를 포함할 수 있다.
- [0212] 적합하게는, 전사 인자 결합 부위는 전사 개시 부위에 작동가능하게 연결된다. 적합하게는, 전사 인자 결합 부위는 전사 개시 부위의 바로 업스트림에 있을 수 있다. 이론으로 국한시키려는 것은 아니지만, 전사 인자 결합 부위 및 전사 개시 부위는 코어 프로모터 영역을 제공할 수 있는 것으로 고려된다.
- [0213] 적합하게는 프로모터는 5' 비번역 영역을 포함할 수 있다. 5' 비번역 영역은 SEQ ID NO: 6 또는 19에 대해 적어도 약 70%, 80%, 90%, 95% 또는 99% 서열 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성될 수 있다. 5' 비번역 영역은 SEQ ID NO: 6 또는 19를 포함하거나 이로 구성될 수 있다.
- [0214] 적합하게는, 5' 비번역 영역은 전사 개시 부위에 작동가능하게 연결된다. 적합하게는, 5' 비번역 영역은 전사 개시 부위의 바로 다운스트림에 있을 수 있다.
- [0215] 일부 구체예에서, 본 발명의 프로모터는 5'에서 3'으로
- [0216] (i) SEQ ID NO: 4 또는 17에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열로서, 하기 요소 각각이 존재하는, 뉴클레오티드 서열: (a) SEQ ID NO: 4 또는 17의 위치 7 내지 위치 13에 대략적으로 상응하는 위치에서 RAR 결합 부위; (b) SEQ ID NO: 4 또는 17의 위치 14 내지 위치 30에 대략적으로 상응하는 위치에서 WT1 결합 부위; 및 (c) SEQ ID NO: 4 또는 17의 위치 49 내지 위치 53에 대략적으로 상응하는 위치에서 인헨서 박스;
- [0217] (iii) 임의적으로 SEQ ID NO: 8에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열, 또는 이의 하나 이상의 단편; 및
- [0218] (ii) SEQ ID NO: 5 또는 18에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99%, 또는 100% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열, 또는 SEQ ID NO: 7 또는 20에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다.
- [0219] 일부 구체예에서, 본 발명의 프로모터는 5'에서 3'으로
- [0220] (i) SEQ ID NO: 4 또는 17에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열로서, 여기에 하기 요소 각각이 존재하는 뉴클레오티드 서열: (a) SEQ ID NO: 4 또는 17의 위치 7 내지 위치 13에 대략적으로 상응하는 위치에서 RAR 결합 부위; (b) SEQ ID NO: 4 또는 17의 위치 14 내지 위치 30에 대략적으로 상응하는 위치에서 WT1 결합 부위; 및 (c) SEQ ID NO: 4 또는 17의 위치 49 내지 위치 53에 대략적으로 상응하는 위치에 인헨서 박스;
- [0221] (iii) 임의적으로 SEQ ID NO: 8에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열 또는 이의 하나 이상의 단편; 및
- [0222] (ii) SEQ ID NO: 5 또는 18에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6 또는 19에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열 또는 SEQ ID NO: 7 또는 20에 대해 적어도 90% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된다.
- [0223] 예시적인 프로모터
- [0224] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 2로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 2와 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.
- [0225] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 2로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 2와 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.
- [0226] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 2로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 2와 적어도 70% 동일한 변이체로 구성된다.
- [0227] 예시적인 최소 네프린 프로모터 - 819 bp (SEQ ID NO: 2)

[0228] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGAATTTTATGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGAGGTTGCACTGTGAGAATGAGCTCAAGTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTGCTGCATCAGCCTCATCGCTCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGTGGCAGGAGAGAATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGAGAAAGAGACAGAGACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAGATGAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCTGTAATCCCAACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCCCGTCTCTAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAAAGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGACCCAGAGAAAGCCAGACAGACGAGGTGGCTGGCAGCGGGCGCTGTGGGGTACAGTAGGGGGACCTGTC

[0229] 적합하게는, 변이체는 SEQ ID NO: 2에 대해 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다. 프로모터는 SEQ ID NO: 15 또는 SEQ ID NO: 46으로 제시된 SEQ ID NO: 2의 변이체를 포함하거나 이로 구성될 수 있다.

[0230] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 15로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 15와 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.

[0231] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 15로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 15와 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.

[0232] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 15로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 15와 적어도 70% 동일한 변이체로 구성된다.

[0233] 적합하게는, 변이체는 SEQ ID NO: 15와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0234] 예시적인 최소 네프린 프로모터 변이체 - 819 bp (SEQ ID NO: 15)

[0235] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTATGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGAGGTTGCACTGTGAGAATGAGCTCAAGTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTGCTGCATCAGCCTCATCGCTCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGTGGCAGGAGAGAATAGAGATTGAGAGAGAGAGAGAGAGAGACAGAGAGAGAGAAAGAGACAGAGACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAGATGAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCTGTAATCCCAACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCCCGTCTCTAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAAAGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGGGACCCAGAGAAAGCCAGACAGACGAGGTGGCTGGCAGCGGGCGCTGTGGGGTACAGTAGGGGGACCTGTG

[0236] 일부 구체예에서, 본 발명의 프로모터는 (ii) SEQ ID NO: 46으로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 46과 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.

[0237] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 46으로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 46과 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.

[0238] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 46으로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 46과 적어도 70% 동일한 변이체로 구성된다.

[0239] 적합하게는, 변이체는 SEQ ID NO: 46과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0240] 예시적인 최소 네프린 프로모터 변이체 - 819 bp (SEQ ID NO: 46)

[0241] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTATGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGAGGTTGCACTGTGAGAATGAGCTCAAGTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTGCTGCATCAGCCTCATCGCTCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGTGGCAGGAGAGAATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGAGAAAGAGACAGAGACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAGATGAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCTGTAATCCCAACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCCCGTCTCTAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAAAGAG

ACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGAAAACGAGAAAGGGAGAGAGTAACGAAAAGAGATAAAAA  
AGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAAGCCAGACAGACGCGAGGTGGCTGGCAGCGGGCGCTGTGGGGGTACAGTAGGGGGACCT  
GTC

[0242] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 58로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 58과 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.

[0243] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 58로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 58과 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.

[0244] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 58로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 58과 적어도 70% 동일한 변이체로 구성된다.

[0245] 예시적인 최소 네프린 프로모터 - 818 bp (SEQ ID NO: 58)

[0246] GGCCCTGGGGTACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGATTTTATGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCT  
GGGCAGAGGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGAGGTTGCACTGTGAGAATGAGCTCAAGCTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTG  
CCTGCATCAGCCTCATCGCTCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTG  
AGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGGAAGAGACAGAGACAAAAGGAGAGAGAACGGCTTAGACAAGGA  
GAGAAAGATGAAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCTGTAATCCCAACTTGGGGAGGCAAGGTGGGAGGATGGCTTGAAGGAAAAGAGTCT  
GAGATCAACCTGGCCAACATAGTGAGACCCGCTCTCTAAAAAAGAAAAAAGAAAAAAGAAAAAAGTTTTTTAAAGAGACAGAGAAAAGAG  
ACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGACTCACAGGGAAGAGGGGAAGAGAAAACGAGAAAGGGAGGAGTAACGAAAAGAGATAAAAA  
AGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAAGCCAGACAGACGCGAGGTGGCTGGCAGCGGGCGCTGTGGGGGTACAGTAGGGGGACCT  
GT

[0247] 적합하게는, 변이체는 SEQ ID NO: 58과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다. 프로모터는 SEQ ID NO: 59 또는 SEQ ID NO: 60으로 제시된 SEQ ID NO: 58의 변이체를 포함하거나 이로 구성될 수 있다.

[0248] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 59로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 59와 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.

[0249] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 59로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 59와 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.

[0250] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 59로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 59와 적어도 70% 동일한 변이체로 구성된다.

[0251] 적합하게는, 변이체는 SEQ ID NO: 59와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0252] 예시적인 최소 네프린 프로모터 변이체 - 818 bp (SEQ ID NO: 59)

[0253] GGCCCTGGGGTACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGATTTTATGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCT  
GGGCAGAGGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGAGGTTGCACTGTGAGAATGAGCTCAAGCTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTG  
CCTGCATCAGCCTCATCGCTCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTG  
AGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGGAAGAGACAGAGACAAAAGGAGAGAGAACGGCTTAGACAAG  
GAGAGAAAGATGAAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCTGTAATCCCAACTTGGGGAGGCAAGGTGGGAGGATGGCTTGAAGGAAAAGAGT  
CTGAGATCAACCTGGCCAACATAGTGAGACCCGCTCTCTAAAAAAGAAAAAAGAAAAAAGAAAAAAGTTTTTTAAAGAGACAGAGAAAAGAG  
ACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGAAAACGAGAAAGGGAGGAGTAACGAAAAGAGATAAAAA  
AGAAAAGCAGGTGGCAGAGACACAGAGAGGGACCCAGAGAAAAGCCAGACAGACGCGAGGTGGCTGGCAGCGGGCGCTGTGGGGGTACAGTAGGGGGACCT  
GT

[0254] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 60으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 60과 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.

[0255] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 60으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 60과 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.

- [0256] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 60으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 60과 적어도 70% 동일한 변이체로 구성된다.
- [0257] 적합하게는, 변이체는 SEQ ID NO: 60과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0258] *예시적인 최소 네프린 프로모터 변이체 - 818 bp (SEQ ID NO: 60)*
- [0259] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTATGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGGAGGTTGCACTGTGAGAATGAGCTCAAGTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTGCTGCATCAGCCTCATCGCTCTCTAGGCTCTGGCCTGTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGTGGCAGGAGAGAATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGAGAAAGAGACAGACAAAAGGAGAGACAACGGCTTAGACAAGGAGAGAAAGATGAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCTGTAATCCCAACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCCGCTCTCTAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAAGTTTTTTAAAGAGACAGAGAAAGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGATACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGAGGAGAGTAACGGAAAGAGATAAAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAAGCCAGACAGACGCGAGGTGGCTGGCAGCGGGCGCTGTGGGGTTCACAGTAGGGGGACCTGT
- [0260] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 3으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 3과 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.
- [0261] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 3으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 3과 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.
- [0262] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 3으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 3과 적어도 70% 동일한 변이체로 구성된다.
- [0263] *예시적인 최소 네프린 프로모터 - 265 bp (SEQ ID NO: 3)*
- [0264] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGAATTTTATGCTCCAGGAGCAAGACAGAGAGAGACTCACAGGGAAGAGGGAAAGAGAAAACGAGAAAGGGAGGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCGAGGTGGCTGGCAGCGGGCGCTGTGGGGTTCACAGTAGGGGGACCTGTC
- [0265] 적합하게는, 변이체는 SEQ ID NO: 3과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다. 적합하게는, SEQ ID NO: 3의 변이체는 치환 A63G, C96T, G185C 및 C265G 중 하나 이상을 포함한다. 적합하게는, SEQ ID NO: 3의 변이체는 치환 A63G, C96T, G185C 및 C265G를 포함한다(SEQ ID NO: 17-20 참조). 프로모터는 SEQ ID NO: 16 또는 SEQ ID NO: 47로 제시된 SEQ ID NO: 3의 변이체를 포함하거나 이로 구성될 수 있다.
- [0266] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 16으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 16과 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.
- [0267] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 16으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 16과 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.
- [0268] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 16으로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 16과 적어도 70% 동일한 변이체로 구성된다.
- [0269] 적합하게는, 변이체는 SEQ ID NO: 16과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0270] *예시적인 최소 네프린 프로모터 변이체 - 265 bp (SEQ ID NO: 16)*
- [0271] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTATGCTCCAGGAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGAAAGAGAAAACGAGAAAGGGAGGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCGAGGTGGCTGGCAGCGGGCGCTGTGGGGTTCACAGTAGGGGGACCTGTC
- [0272] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 47로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 47과 적어도 70% 동일한 변이체를 포함하거나 이로 구성된다.

- [0273] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 47로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 47과 적어도 70% 동일한 변이체를 포함하고, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는다.
- [0274] 일부 구체예에서, 본 발명의 프로모터는 SEQ ID NO: 47로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 47과 적어도 70% 동일한 변이체로 구성된다.
- [0275] 적합하게는, 변이체는 SEQ ID NO: 47과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0276] *예시적인 최소 네프린 프로모터 변이체 - 265 bp (SEQ ID NO: 47)*
- [0277] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTATGCTCCAGGAGCAAGACAGAGAGATACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGGAGGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAA GCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGTG
- [0278] **폴리뉴클레오티드**
- [0279] 본 발명은 본 발명의 프로모터를 포함하는 폴리뉴클레오티드를 제공한다.
- [0280] 본 발명의 폴리뉴클레오티드는 DNA 또는 RNA, 바람직하게는 DNA를 포함할 수 있다. 이것들은 단일-가닥 또는 이중-가닥일 수 있다. DNA 폴리뉴클레오티드와 같은 폴리뉴클레오티드는 제조합적으로, 합성적으로 또는 당업자에게 이용 가능한 임의의 수단에 의해 생성될 수 있다. 이들은 또한 표준 기술에 의해 클로닝될 수 있다. 폴리뉴클레오티드는 분리된 폴리뉴클레오티드일 수 있다.
- [0281] 더 긴 폴리뉴클레오티드는 일반적으로 제조합 수단을 사용하여, 예를 들어, 폴리머라제 연쇄 반응(PCR) 클로닝 기술을 사용하여 생성될 것이다. 이는 클로닝하고자 하는 표적 서열에 축적된 한 쌍의 프라이머(예를 들어, 약 15 내지 30개의 뉴클레오티드)를 제조하고, 프라이머를 동물 또는 인간 세포로부터 수득된 mRNA 또는 cDNA와 접촉시키고, 요망되는 영역의 증폭을 야기하는 조건 하에 폴리머라제 연쇄 반응을 수행하고, (예를 들어, 반응 혼합물을 아가로스 겔로 정제함으로써) 증폭된 단편을 분리하고, 증폭된 DNA를 회수하는 것을 포함할 것이다. 프라이머는 증폭된 DNA가 적합한 벡터로 클로닝될 수 있도록 적합한 제한 효소 인식 부위를 함유하도록 설계될 수 있다.
- [0282] 폴리뉴클레오티드는 당업계에서 이용가능한 임의의 방법에 의해 변형될 수 있다. 이러한 변형은 본 발명의 폴리뉴클레오티드의 생체내 활성 또는 수명을 향상시키기 위해 수행될 수 있다.
- [0283] 단백질-코딩 서열
- [0284] 프로모터는 하나 이상의 단백질-코딩 서열에 작동가능하게 연결될 수 있다.
- [0285] 단백질-코딩 서열은 임의의 관심 폴리펩티드를 인코딩할 수 있다. 예를 들어, 단백질-코딩 서열은 사구체 질환과 관련된 임의의 폴리펩티드를 인코딩할 수 있다. 단백질-코딩 서열은 알포트 증후군과 같은 GBM-관련 유전적 사구체 질환과 관련된 폴리펩티드를 인코딩할 수 있다. 단백질-코딩 서열은 발세포-관련 유전적 사구체 질환에 관여하는 폴리펩티드를 인코딩할 수 있다.
- [0286] 적합하게는, 단백질-코딩 서열은 COL4A3, COL4A4, COL4A5, NPHS2, CFH, CFL, CFHL1, C1INH, C4BP, MASP2, C3, C5aR1, C5, C5a, CD55, CD35, CD46, CD59, 비트로넥틴, 클루스테인, ADCK4, ALG1, ARHGAP24, ARGHDIA, CD151, CD2AP, COQ2, COQ6, DGKE, E2F3, EMP2, KANK2, LAGE3, LMNA, LMX1B, MAF B, NUP85, NUP93, NXF5, OSGEP, PAX2, PDSS2, PMM2, PODXL, SCARB2, SGPL1, Smad7, TP53RK, TPRKB, VDR, WDR73, WT1, ZMPSTE24, APOL1, NPHS1, TRPC6, NUP107, NUP133, NUP160, ACTN4, INF2, ANKFY1, ANLN, CRB2, ITGA3, KANK1, KANK4, MAGI2, MYO1E, OCRL, PTPRO, SMARCAL1, SYNPO, TBC1D8B, XPO5, TNS2, NLRP3 또는 VEGFC 폴리펩티드를 인코딩할 수 있다.
- [0287] 적합하게는, 단백질-코딩 서열은 1450개 이상의 아미노산, 1500개 이상의 아미노산, 1550개 이상의 아미노산, 1600개 이상의 아미노산 또는 1650개 이상의 아미노산 길이를 갖는 폴리펩티드를 인코딩한다.
- [0288] 단백질-코딩 폴리뉴클레오티드에 있어서, 유전자 코드의 축퇴의 결과로서 다수의 상이한 폴리뉴클레오티드가 동일한 폴리펩티드를 인코딩할 수 있음이 당업자에 의해 이해될 것이다. 또한, 당업자는 관례적인 기술을 사용하여 본 발명의 폴리펩티드가 발현되는 임의의 특정 숙주 유기체의 코돈 사용을 반영하기 위해 본 발명의 폴리뉴클레오티드에 의해 인코딩된 폴리펩티드 서열에 영향을 미치지 않는 뉴클레오티드 치환을 수행할 수 있음이 이

해되어야 한다.

- [0289] 단백질-코딩 서열은 코돈-최적화될 수 있다. 상이한 세포는 특정 코돈의 사용이 상이하다. 이러한 코돈 편향은 세포 유형에서 특정 tRNA의 상대적 풍부도에서의 편향에 상응한다. 상응하는 tRNA의 상대적 존재비와 일치하도록 맞춤화되게 서열의 코돈을 변경함으로써, 발현을 증가시키는 것이 가능하다. 동일한 토근으로, 상응하는 tRNA가 특정 세포 유형에서 드문 것으로 알려진 코돈을 의도적으로 선택함으로써 발현을 감소시키는 것이 가능하다. 따라서, 추가적인 정도의 번역 제어가 가능하다. 코돈 사용 표는 포유동물 세포(예를 들어, 인간) 뿐만 아니라 다양한 다른 유기체에 대해 당 분야에 공지되어 있다.
- [0290] 본원에 개시된 단백질 코딩 뉴클레오티드 서열은 이들의 3' 말단에 정지 코돈을 포함하거나 없을 수 있다. 따라서, 본 개시는 종결 코돈이 존재하거나 부재하는 본원에 개시된 SEQ ID NO를 포함한다.
- [0291] *COL4A3*, *COL4A4* 및 *COL4A5* 폴리펩티드
- [0292] 단백질 코딩 서열은 *COL4A3*, *COL4A4* 또는 *COL4A5* 폴리펩티드, 또는 이의 단편 또는 유도체를 인코딩할 수 있다.
- [0293] *COL4A3*, *COL4A4* 및 *COL4A5* 단백질은 비-콜라겐성 서열에 의해 빈번하게 중단되고 삼중 나선 반복을 형성하는 콜라겐성 Gly-X-Y 반복 서열을 함유하는 대략 170-185 kDa 상동성 폴리펩티드이다. 각각의 폴리펩티드는 또한 카르복실-종결 말단에 큰 구형 비-콜라겐성 도메인을 함유한다.
- [0294] 알포트 증후군(AS)은 *COL4A3*, *COL4A4* 및 *COL4A5* 유전자의 병원성 변이체에 의해 유발되며, 이는 기저막의 콜라겐 IV α345 네트워크의 이상을 초래한다.
- [0295] *COL4A3*, *COL4A4* 또는 *COL4A5* 폴리펩티드 또는 이의 단편 또는 유도체는 콜라겐 IV α345 네트워크를 형성할 수 있다.
- [0296] 약 200-300개의 아미노산은 각각의 *COL4A3*, *COL4A4* 및 *COL4A5* 폴리펩티드로부터 제거되어 미니-유전자 접근법에 적합한 절두된 전이유전자를 생성할 수 있다. 아미노산은 삼중 나선 반복으로부터 제거될 수 있다. 바람직하게는, 아미노산은 비-콜라겐성 영역으로부터 제거되지 않는다.
- [0297] 일부 구체예에서, *COL4A3*, *COL4A4* 및 *COL4A5* 폴리펩티드는 전장 폴리펩티드이다.
- [0298] 바람직하게는, *COL4A3*, *COL4A4* 또는 *COL4A5* 폴리펩티드는 인간이다. 예시적인 인간 *COL4A3*은 UniProtKB 수탁 번호 Q01955를 갖는 *COL4A3*이다. 예시적인 인간 *COL4A4*은 UniProtKB 수탁 번호 P53420를 갖는 *COL4A3*이다. 예시적인 인간 *COL4A5*은 UniProtKB 수탁 번호 P29400를 갖는 *COL4A5*이다.
- [0299] 적합하게는, *COL4A3* 펩티드는 SEQ ID NO: 21로 제시된 폴리펩티드 서열 또는 SEQ ID NO: 21과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 21과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0300] 적합하게는, *COL4A4* 펩티드는 SEQ ID NO: 22로 제시된 폴리펩티드 서열 또는 SEQ ID NO: 22와 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 22와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0301] 적합하게는, *COL4A5* 펩티드는 SEQ ID NO: 23으로 제시된 폴리펩티드 서열 또는 SEQ ID NO: 23과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 23과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0302] 예시적인 *COL4A3* 아미노산 서열 - Uniprot 참조번호 Q01955 (SEQ ID NO: 21)
- [0303] MSARTAPRPQVLLLPLLLVLLAAAPAASKGCVCKDKGQCFCDGAKGKGEKGFPGPPGSPGQKGFITGPEGLPGPQPKGFPLPGLTGSKGVIRGISGLPGFSGSPGLPGTPGNTGYPYGLVGVPGCSGSKGEQGFPLPGTLGYPGIPGAAGLKGQKGAPEKEEDIELDAKGDPLPGAPGQGLPGPPGFPVPGPPGPFVGFPGFPGFPGAMGRPRGPKGHMGERVIGHKGERGVKGLTGPPGPPGTVIVTLTGPDNRTDLKGEKGDKGAMGEPGPPGSPGLPGESYGSEKGAPEGLQGGKPKGDKGVPGFPGSEGVKGNRGPFLMGEDGKIKQKGDIGPPGFRGPTEYYDYQEKGDEGTPGPPGPRGARGPQGPSGPPGVPSPGSSRPLRGLRAGWPGLKGSKGERGRPGKDMGTGSPGCAGSPGLPGSPGPPGPPGDIIVFRKGPDPHGLPGYLGSPGIPGVDGPKGEPGLLCTQCPYIPGPPGLPGLPGLHGVKIPGRQGAAGLKGSPGSPGNTGLPFGFPGAQDPLKGEKGETLQPEQGVGVPDPLRGRQGRKGLDIPGTPGVKGLPGPKGELALSGEKDQGGPPGDPGSPGSPGAGPAGPPYGPQGEPLGQGTQGVGAPGPPGEAGPRGELSVSTVPVPGPPGPPGPPGHPGPPGPPGIPGSLGKCGDPLPGPDGEPGIPGIFGPPGPPGKGDQGFPGTKGSLGCPGKMGEPGLPGKPLPGAKGEPVAVMPGGGTPGPFPERGNSGEHGEIGLPLPGLPGTPGNEGLDGRGDPGQGPPEQGGPPGRCIEGPRGAQGLPGLNGLKQGGRRGKTGPKGDPGIPGLDRSGFPGETGSPGIPGHQGEMLGQRRGYPNPGILGPPGEDGVI GMMGFPGAIGPPGPPGNPTGPPGQGRGSP

GIPGVKQQRGTPGAKGEQDKGNPSPSEI SHVIGDKGEPGLKGFAGNPGEKGNRVPMPGLKGLKGLPGPAGPPGPRGDLGSTGNPGEPLRGI PGSMGNM  
 GMPGSKGKRGTGLGFPGRAGRPLPGIHLQGDKEGPGYSEGTRPGPPGTGDPGLPGDMGKKGEMQPGPPGHLGPAGPEGAPGSPGSPGLPGKPGPHGDLG  
 FKGIKGLLPGPIRGPPPLPGFPGSPGPMIRGDQGRDIPGPAGEKGETGLLRAPPGRPNPGAQGAQKDRGAPGFPPLPGRKGMAGDAGPRGPTGIEGFP  
 GPPGLPGAIIPGQTNRGPSPGSRGSPGAPGPPGPPGSHVIGIKGDKGSMGHPGPKGPPGTAGDMGPPGRLGAPGTPGLPGRGDPGFQGFPGVKGKGNPGF  
 LGSIGPPGPIGPKGPPGVRGDPGLTKIISLPGSPGPPGTPGEPGMQGEFPPGPPGNLPGCGPRGKPKGDKGKPGTGPAGEKGNKSKGEPGAGSDGLPGL  
 KGKRGDSGSPATWTRGFVTRHSQTTAIPSCPEGTVPLYSGFSFLVQGNQRAHQDQLGTLGSCLQRFTTMPFLFCNVNDVCFASRNDYSYWLSTPALMP  
 MNMAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPPCPHGWISLWKGFSFIMFTSAGSEGTQALASPGSCLEEFRASPFLECHGRGTCNYNSYNSYFW  
 LASLNPERMFRKPI PSTVKAGELEKII SRCQVCMKKRH

[0304] 예시적인 COL4A4 아미노산 서열 - Uniprot 참조번호 P53420 (SEQ ID NO: 22)

[0305] MWSLHIVLMRCSFRLTKSLATGPWSLILILFSVQYVYVSGSKKYIGPCGGRDCSVCHCVPEKGSRGPPGPPGQGP IGPLGAPGPIGLSGEKGMGRDRGPPGA  
 AGDKGDKGPTGVPGFPLDGI PHGPPGPRGKPGMSHNGSRGDPGFPGGRGALPGGGLGHPGEKGEKGNVSVILGAVKGIQGDGRDPLPGLPGSWGAG  
 GPAGTGYGEPGLVGPQPGPRPGLKGNPVGKGMQMDPGEVQOQSPGPTLLVEPPDFCLYKGEKGIKGI PMVGLPGRPGRKGESGIGAKGEKGI PGF  
 PGRGDPGSGSPGFPGLKGLGLVGDPLGFLIGPKGDPNRRGHPGPPGVLVTPPLPLKGGPPGDPGFPGRYGETGDVGPVPPGGLGRPEACAGMIGPPG  
 PQGFPLPGLPEAGIPGRPDSAPGPKGKPSGLPGAPLQGLPGSSVIYCSVGNPQPQIKGKVGPPGGRGPKGEKNEGLCACEPGMGPVPPGPPGLPGR  
 QGSKGDLGLPGWLGTKDPPGAEPPGLPGKHGASGPPGNKGAKGMVSVRVKHKGERGPDGPPGFPQPGSHRDRGHAGEKDPGPPGDHEDATPGGK  
 GPPGLGPPKAGVGPPLGFPGPPGERGHPGVPHPGVRGPDGLKQKQDITISCNVTYPRHRGPPGDFGPPGPKGFPQGAPGLSGSDGHKGRPGTPT  
 AEIPGPPFRGMDGDPGFGEKSSVPPGPPGSPGVNQKGI PGDPAFHLGPPGKRGLSGVPIKGRGDPGCPGAEAGIPGFLGLKPKGREGHAG  
 FPGVPPGHSCERAGPIPGQPLPGYPSGAPGGKQPGDVGPPGAGMKGLPGLPGRPGAHPGGLPIPGPFDDGLPVPKGRPLPGFPGFP  
 ERGKPGAEGCPGAKGEPGEKMSGLPGRDLRGAAGAI GPPGDEGEMAIISQKGTGEPGPPGDDGFPGERGDKGTPGMQRRRGEPRYGPFGHRGEPGEK  
 GQPPGPPGPPGSTGLRGI GFPLPGDQGEPSGPPGFSGIDGARGPKGNKDPASHFPPGPKGEPGSPGCPGHFGASGEQGLPIQGPRGSPRGP  
 PGSSGPPGCPGDHMPGLRGQPGEKMDPGRGLQGDPIGPPGPKGSPGSLNHLGLKQKGTGASGLHDVGPVPIGLKGERGDPGSPGISPPG  
 PRGKGPVPPGSSGPPGATGRAPKIDPDPGPDQPPGDPGRGAPGPPGLPGVDLLRGEPCDGLPVPVPPGPPGPPYKGFPGCDKDGKQKGP  
 VGFPGQPHGFPVPPGEKGLPVPGRKGTGLPGRGEPGPPADVDDCPRIPLPGAPMRGPEGAMGLPMRGPSPGCKGEPGLDGRRGVDVGPVSPGP  
 PGRKGTGEGDYPGPPGPPGIGDPPKGFPGYLGGFLVLSHTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPAFCNIHQV  
 CHYQRNDRSYWLASAAPLMMPLSEEAIRPVYSRCAVCEAPAQAVAVHSQDQSI PCPQTWRSLWIGYSFLMHTGAGDQGGGQALMSPGSCLEDFRAAPFL  
 ECQGRGTCHFFANKYSFWLTTVKADLQFSSAPADTLKESQAQRQKISRCQVCVKYS

[0306] 예시적인 COL4A5 아미노산 서열 - Uniprot 참조번호 P29400 (SEQ ID NO: 23)

[0307] MKLRGVSLAAGLFLALSLWQPAEAAAACYGCSGSKDCSCSIKGEKGERGFPLEGHPGLPGFPGPEGPPGPRGQKDDGIPGPPGPKGIRGPPPLPGFPG  
 TPGLPGMPGHDGAPGQIPGCGNGTKGERGFPSPGFPGLQGGPPPGIPGMKGEPSIIMSSLPGKGNPGYPPGPIQLPGPTIPGPIGPPGPPGLMG  
 PPGPGLPGPKGNMGLNFQGPKGEKGEQGLQGGPPGQISEQKRPIDVEFKQDQGLPDRGPPGPPGIRGPPGPPGKEKGEKEQGEKGRKPKGDGEN  
 GQPGIPLPGDGYGEPGRDGEKQKQDGTGPPGPPGLVIRPRGTGITIGEKNIPLPGLPEKGERGFPIQGPPPLPGPPAAVMGPPGPPGPFGERGQK  
 GDEGPPGISIPGPPLDGQPGAPGLPGGPAGPHIPPSDEICEPGPPGPPSPGDKLQGEQVKGDKGDTFCNICITGISGPPGQPLPGLPGGPSLGF  
 PGQKGEKQAGATGPKLPGIPGAPGAPGFPGSKGEPGDIITFPGMKDKGELGSPGAPGLPGLPTPGQDGLPGLPKGEPGGITFKGERGPPGNPLPG  
 LPGNI GMPGPPGFPVGEKGIQGVAGNPQPIPGPKGDPQITITQPKPGLPNPGRDGDVGLPGDPLPGQPLPIPGSKGEPGIPGILGPPGPP  
 KGFPGIPGPPGAPGTPGRI GLEGGPPGPPGPKGEPGFALGPPGPPGLPGFKGALGPKGDRGFPGPPGPPGRTGLDGLPGPKGDVGNQPGPMGPPGLP  
 GIGVQPPGPPGIPGPIGQPLHGIPGEKDPGPPGLDVPGPPGERGSPGIPGAPGPIGPPGSPGLPGKAGASGFPGTGKEMGMMGPPGPPGLPIGRSGV  
 PGLKGDGLQGGPLPGPTGEKSKGEPGLPVPMPDNLGSKGKGEKGLPGIPGVSGPKYQGLPGDPPGQPLSGQPLPVPKGNPGLPGQPLI  
 GPPGLKGTIGDMGFPQGVGPPGSPVPGQPSPLPGQKDKGDPGIISSIGLPLGPKGEPGLPGYGNPPIKGSVGDPLPGLPTGAKGQPLPG  
 FPGTGPVPPKGISGPPGNPLGPEGPPVGGGHPGQPPGPKGKQDQIPGAPGQKGEPPGPFGNPGLPGLSGQKGDGGLPIGNPLPGPKGE  
 PGFHGFPVQGGPPGSPGPALEPKGNPQGPGRPLGPEGPPPLGNGGIKGEKGNPQPLPGLPGLKGDQPPGLQGNPGRPLNGMKDPLP  
 GVPFPGMKGSPGVPAGPEGEPGLIGPPGPPGLPGSPQSI IIKGDAGPPGIPGQPLKGLPGPQGPGLPGTTPGDPGRNPLPGFDGAGGRKDPGL  
 PGQPTRGLDGGPPDGLQGGPPGPPGTTSSVAHFLLITRHSQTTDAPQCQGLQVYEGFSLLYVQGNKRAHQDQLGTAGSCLRRFSTMPFMFCNINNVCF  
 SRNDYSYWLSTPEPMPMSQPLKQSIQPFISRCVCEAPAVVIAVHSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEGSQALASPGSCLEEFRASPFIEC  
 HGRGTCNYNSYNSYFWLATVDVSDMFSKQSETLKAGDLRTRISRCQVCMKRT

[0308] 단백질-코딩 서열은 COL4A3, COL4A4 또는 COL4A5 전이유전자를 포함하거나 이로 구성될 수 있다.

[0309] COL4A3을 인코딩하는 예시적인 전이유전자는 NM\_000091.5로 제공된다. COL4A4을 인코딩하는 예시적인 전이유전자는 NM\_000092.5로 제공된다. COL4A5을 인코딩하는 예시적인 전이유전자는 NM\_000495.5로 제공된다.

[0310] 적합하게는, COL4A3 전이유전자는 SEQ ID NO: 24로 제시된 폴리뉴클레오티드 서열 또는 SEQ ID NO: 24와 적어도

70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 24와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0311] 적합하게는, COL4A4 전이유전자는 SEQ ID NO: 25로 제시된 폴리뉴클레오티드 서열 또는 SEQ ID NO: 25와 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 25에 대해 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0312] 적합하게는, COL4A5 전이유전자는 SEQ ID NO: 26으로 제시된 폴리뉴클레오티드 서열 또는 SEQ ID NO: 26과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 26과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0313] 예시적인 COL4A3 전이유전자 서열 (SEQ ID NO: 24)

[0314] atgagcgcgccgaccgccccaggccgcaggtgctcctgctgcccctcctgctggctcctggcggcgcccgagccagcaagggttgtgtctgtaaa  
gacaaaggccagtgcttctgtgacggggccaaaggggagaagggggagaagggttttctggacccccgggttctcctggccagaaaggatcacaggtcct  
gaaggcttgctggaccgcagggaccaagggtttccaggacttccaggactcacgggttccaaaggtgaaggggaataagtggtatgcccaggttttct  
gggttctcctggacttccaggcaccaggaataaccgggcttaccggacttgtcggtgtaccaggatgcagtggttctaaagggtgagcagggtttccagga  
ctcccaggacactgggctaccaggatcccgggtgctgctggtttgaaaggacaaaagggtgctcctgctaaagaagaagatagaacttgatgcaaaa  
ggcgacccccgggttgccagggttccaggacccagggtttgccaggcctccagggttttctggcctgttgcccacctggtcctccgggtatctttggc  
tttccaggagccatgggacctagaggacctaaagggtcacatgggtgaaagagtgataggacataaaggagagcgggtgtgaaagggttaacaggacccccg  
ggaccaccaggaacagttatgtgacctaaactggcccagataacagaacggacctcaaggggaaaaggagacaaggagcaatgggcgagcctggacct  
cctggacctcaggactgctggagaatcatatggatctgaaaagggtgctcctggagacctggcctgcagggaaaaccggaaaagatggtgttctggc  
ttcctggaagtgagggagtcaagggcaacagggtttcctgggttaatgggtgaagatggcataaaggacagaaaaggggacatggcctccaggat  
cgtggtccaacagaatattatgacacataccaggaagggagatgaaggcactccaggcccaccagggccagaggagctcgtggcccacaaggtcccagt  
ggtccccggagtctcctggaagtctggatcatcaaggcctggcctcagaggagccctggatggccaggcctgaaaggaagtaaaggggaacaggccgc  
ccaggaaaggatgcatggggactcctgggtccccagggtgtgctggttccagggtctccaggatcaccgggacctccaggaccagggtgacatcgtt  
tttgcgaagggtccacctggagatcacggactgccaggctatctagggtctccaggaatccaggagttgatgggcccagggagaaccaggcctcctgtgt  
acacagtgcccttatatcccaggcctcccgggtctcccaggattgccagggttacctgggtgtaaaaggaatccagggaagacaaggccagctggctgaaa  
ggaagcccagggtcccaggaaatcacaggtcttccaggatttccagggttccagggtgccagggtgaccaggacttaaggagaaaaaggtgaaacactt  
cagcctgaggggcaagtgggtgtcccagggtgacccgggctcagaggccaacctgggagaaaagggttggatggaattcctggaactccgggagtgaagga  
ttaccaggacctaaaggcgaactggctctgagtggtgagaaaggggaccaagggtcctccaggggatcctggctccccgggtccccaggacctgcaggacca  
gctggaccacctggctacggacccaaggagaacctgggtctccagggcacgcaaggagtctcctggagccccggaccaccggagaagccggcctagggga  
gagctcagttttcaacaccagttccaggcccaccaggacctccaggccccctggccatcctggccccaaaggtccacctggtatcctggatcctgggg  
aaatgtggagatcctggtctccaggcctgatggtgaaccaggaatccaggaatggatttctgggctcctggacctaaaggagaccaaggttttcca  
ggtacaaaaggatcactgggtgtcctggaaaaatgggagagcctgggttacctggaagccaggcctccaggagccaagggagaaccagcagtagccatg  
cctggaggaccaggaacaccaggttttccaggagaagaggcaattctggggaacatggagaaatggactcctggacttccagggtctcctggaactcca  
ggaaatgaagggttgatggaccacgaggagatccagggcacctggaccacctggagaacaaggacccccaggaaaggtgcatagagggtcccaggggagcc  
caaggacttccaggcttaaatggatgaaagggcaacaaggcagaagggtaaaacggggccaaaggagaccaggaatccaggcttggatagatcagga  
tttctggagaaactggatcaccaggaatccagggtcatcaagggtgaaatgggacctgggtcaagaggatccaggaaatccgggaattttagggcca  
ccagggtgaagatggagtgatgggatgatgggtttcctggagccattggcctccagggccccctgggaaccaggcacaccagggcagagggggagccct  
ggaatccaggagtaaggggccagagaggaacccaggagccaagggggaacaaggagataaaggaaatccgggctcctcagagatcccacgtaataggg  
gacaaaggagaaccaggtctcaaggatctgcaggaaatccagggtgagaagaaacagaggcgttccagggtgccaaggtttaaaggcctcaaggacta  
cccggaccagcaggaccaccaggccccagaggatttgggcagcactgggaatcctggagaaccaggactgcgtggtataccaggaagcatggggaacatg  
ggcatgccaggttctaaaggaaaaaggggaactttgggatcccaggctcagcaggaagaccaggcctccagggtatctatggtctccaggagataaagga  
gagccaggttatcagaaggtacaaggccaggaccaccgggaccaacgggggatccaggactgccgggtgatgggaaagaaaggagaaatggggcaacct  
ggccccctggacatttggggctgctggacctgagggagccccctggaagtctggaagtctggcctccaggaaagccaggtcctcatggtgatttgggt  
tttaagggaatcaaggcctcctggcctccaggaatcagaggcctccagggtctccaggatccaggatctcctggaccaatgggtatagagggtgac  
caaggacgtgatggaattcctggtccagccggagaaaagggagaaacgggtttatgaggccccctccaggcccaagagggaacctggtgctcaaggacc  
aaaggagacaggggagccccaggtttcctggcctcccgggcagaaaaggggcatgggagatgctggacctcagggaccacaggcatagaaggatccca  
gggccaccaggtctccccgtgcaatattcctggccagacaggaaatcgtggtccaccaggctcaagaggaagccaggtgcgctgggtccccctggacct

ccaggagtcattgataataggcataaaaggagacaaagggtctatggccaccctggccaaaagggtccacctggaactgcaggagacatgggaccaccaggt  
 cgtctgggagcaccaggtactccaggtctccaggaccagaggtgatcctggattccagggttccaggcgtgaaaggagaaaagggtaatcctggattt  
 ctaggatccattggacctccaggaccaatgggcaaaaaggaccacctggtgtacgtggagaccctggcacactaagattatctccctccaggaagccca  
 gggccacctggcacacctggagaaccagggatgcaggagaaacctgggcccaccaggccacctggaacctaggacctgtgggcaagaggt aagccaggc  
 aaggatgaaaccaggaaactcctggaccagctggagaaaaaggcaacaagggttctaaaggagagccaggaccagctggatcagatggattgccaggtttg  
 aaaggaaaactgtggagacagtgatcacctgcaacctggacaacgagaggtttgtcttcccccacacagctcaaacacagcaatccttcatgtccagag  
 gggacagtgccactctacagtggttttctttcttttgtacaaggaaatcaacgagcccaccggacaagaccttggaaactctggcagctgcctgcagcga  
 ttaccacaatgccattcttattctgcaatgtcaatgatgtatgtaattttgcatctcgaaatgatattcatactggctgtcaaccagctctgatgcca  
 atgaacatggctcccatfactggcagagcccttgagccttataaagcagatgcactgtttgtgaaggtcctgcgatcgccatagccgttcacagccaaacc  
 actgacattcctccatgtcctcacggctggatttctctctggaaaggatttctattcatcatgttcacaagtgcaggttctgaggggcaccgggcaagcactg  
 gcctccctggctcctgcctggaagaattccgagccagccatttctagaatgtcatggaagaggaactgcaactactattcaaatcctacagtttctgg  
 ctggcttcataaaccagaaaagatttccagaaagcctattccatcaactgtgaaagctggggaattagaaaaataataagtcgctgtcaggtgtgcatg  
 aagaaaagacactga

[0315] 예시적인 COL4A4 전이유전자 서열 (SEQ ID NO: 25)

[0316] atgtggtctctgcacatagtactaatgaggtgctccttcagattgaccaagtccttggccacaggtccctggtcacttatactcatctcttttctgtacaa  
 tatgtatgggagtggaagaataacattggctccttgtggaggaagagattgctctgtttgccactgtgttctgaaaaggggtctcggggtccaccagga  
 ccaccaggccaccaggtccaattggaccttgggagcccaggaccattgggcttccaggagagaaaggaatgagaggggaccgcccctcctggagca  
 gcaggggacaaaggagataaggggtccaactggtgttctggatttccaggtttagatggcatacctgggcaccaccaggccctcctggaccagaggcaaacct  
 ggtatgagtggccacaatggctcaagaggtgaccagggttccaggaggaagaggagctctggcccaggaggccccctaggccatcctggggaaaaggga  
 gaaaaaggaattcagtttctatttaggtgccgttaaaggtatccaggagacagaggggaccaggactgcctggcttaccaggatctgggggtgcagga  
 ggaccggcaggtcccacaggatctcctggagagccagggttagtgggacctccgggccaaccaggggcctccaggtttagagggaaatcccgggtgtgggagt  
 aaggggcaaatgggagaccgggtgaggttggtcagcaaggttctcctggaccacctgttggtagagccacctgacttttctctataaaggagaaaaag  
 ggtataaaaggaattcctggaatggtggactgccaggaccaccaggacgcaagggagaatctggtatggggcaaaaaggagaaaaaggtatcctggattt  
 ccagggcctcggggggatcctggttctatggatctccaggttttccaggataaagggagaactaggactggtggagatctgggctatttggatfaat  
 ggcccaagggggatcctggaatcgagggcaccaggaccaccagggtgtttggtgactccacctctccactcaaggcccaccaggggaccagggttcc  
 cctggccgtatggagaaacaggggatgtggaccacctggtccccaggctcttgggcagaccaggggaagcctgtgcaggcatgataggaccacctggg  
 ccacaaggatttctggcttctgggcttccaggagaagctggtattcctgggagacctgatctgctccaggaaaaccagggaagccaggatcacctggc  
 ttgcttggagcaccaggcctgcagggcctccaggatcaagtgtgatatactgtagtgtgggaacccggaccacaaggaataaaaggcaaatgtgtccc  
 ccaggaggaagagggccaaaaggagaaaaaggaatgaaggactctgtgctgtgagcctggacctgggccccctggcctccaggacttctgggagg  
 caggggagt aagggagacttgggctcctggcttggacttggaaacaaagggtgaccaggacctcctggtgctgaaggacctccagggtaccaggaaagcat  
 ggtgcctctggaccacctggcaacaaagggcgagggtgacatggttgtatcaagagttaaaggccacaaggagaaagaggtcctgatgggccccagga  
 tttccagggcagccaggatcacatggtcgggatggacatgctggagaaaaaggggatccaggacctccaggggatcatgaagatgcgacccaggtggtaaa  
 ggatttctggacctctgggccccccaggcaaaagcaggacctgtgggccccaggactgggatttctggtccaccaggagagcaggccaccaggagt  
 ccaggccaccagggtgtgagggccctgatggcttgaagggtcagaaaggtgacacaatttcttgcaactaacctacctgggaggcatggcctccaggt  
 tttgatggacctccaggtccgaaggatttccaggtccccagggtgccctgggctgagtggttcagatgggcataaaggcagacctggcacaccaggaaca  
 gcggaataaccaggtccacctggttttctggtgacatgggagatccgggttttggaggtgaaaaggggtcctccccgttgggccccaggccctcccggc  
 taccaggagtgaatggtcagaaaggaatcccgggagacctgcatttggtcacctgggaccccggaagaggggtcttccaggagtccaggataaaa  
 ggaccagaggtgatccgggatgtccagggtggaagggccagctggcatcctggattcctaggctcaaggtcccaaggcagagagggacatgctggg  
 tttccaggtgtccaggtccacctggcattcctgtgaaagaggtgctccaggataccaggcaaccgggactccctgggtatccaggtagcccaggtgct  
 ccaggtgggaaaggacagccgggagatgtgggctcccgggaccagctggaatgaaaggcctcccggactccaggacggcctggggcacatggtccccca  
 ggctcccaggatcccaggtcctttggagatgatgggctacctggtcctccaggtccaaagggacccgggggctgcctggtttccagggttttccgga  
 gaaagaggaaagcctggtgcagagggatgtcctggcgcaaaagggagaacctggagagaagggcatgtctggccttctggagaccgggactgagagggcc  
 aaaggagccataggacctcccggagatgaaggagaaatggctatcattcacaaaagggaacacctggggaacctggacctcctggagatgatggattccca  
 ggagaaagaggtgataaaggaactcccgggatgcaaggggagaagaggagagccgggaagatccggaccacctggatttccagaggggaaacctggtgagaaa  
 ggtcagccaggccctcctggacccccaggccctccaggctcaactggtctaaaggggttcatgggtttccaggactccaggtgaccagggtgagccaggt  
 tctccaggtccccctggattttcaggaatgatggagcaagaggacctaaaggaaacaaagggtgacctgccagtcactttggtccacctggtccaaagggt  
 gagccaggtagccctggatgtccaggcattttggagcatccggagagcagggttgcctggtattcaagggccagaggatcacctggaagggcagggcca  
 cctggctcctctggaccaccagggtgcccaggtgatcacgggatgcctgggctgaggggacagccaggagaaatgggagacctgggccaagaggcctccag  
 ggggatccaggataccaggtcctccgggaataaaagggtcctccggatcacctggcctgaaccgcttgcattgaaaggtcagaaaggaactaaaggt  
 gctcaggtttgcatgatgtgggcccacctggtccagtggaatacctgggctaaaaggggagagaggagacctgggagcccaggatctcctccaggt

cctcgtgaaagaaaggtccccaggacccccaggagttcaggaccacctggtcctgcaggtgccacaggaagacctctaaggacatctctgacccgggtccacctggagatcagggacctctggtcctgatggcccaagaggagcacctgggacctccaggacctcctgggagtggtgaccttctgagaggggagccaggtgactgtggtctaccagggccaccaggtccccctggcccaccaggccctccaggatacaaaggcttccaggatgtgatggaaaagatggccagaaaaggaccagtgggatccccgggaccgcagggaccacatggatttctgggccacctggagagaagggttacctggacctccagggagaaaaggggcccactggtcttccgggtccagaggtgaaccggggccacctcgagatgtggatgactgtccccgaatcccaggaccttctggggcgccaggcatgagaggaccagaaggagccatggggtcctggaatgagaggccctcaggaccagggtgcaaaggagagcctgggctggatggcaggaggggtgtggtggcgtcctgggtctctgggacctccccggacctaaaggtgacacaggagaagacggctacctggaggaccaggacctcctggctccatggggatctggggccaaaagggttggccctggatacctcgggtggcttctcctggttctccacagtacacggaccaggagcccacctgccccctgggcatgcccaggctctggactgggtatagtctgttatactggaaggcaagagaaagctcacaatcaagaccttggctctggcagggtcttgccttcccgtatttagcacgtgcccttgcctactgcaacatccaccaggttgccactatccccagagaaacgacagatcctactggctggccagcgtgcgcccctcccctgatgccactctctgaagagccgatccgcccctatgtcagccgtgtcgggtatgcagggccccggcccaggcgggtggcgtgcacagccaggaccagttccatccccctgtccgcagacctggaggagcctctggatcgggtattcatctctgatgcacacaggagctggggaccaaggaggaggcaggcccttatgtcacctggcagctgcctggaagattcagagcagccacctcttgaatgccaggggccgaggaacttgcacttttgcgaaataagtagctctggctcacaacggtgaaagcagacttgcagtttctctgctccagcacagacacctaaaagaaagccaggccaacccagaaaatcagccggtgccaggtctgcgtgaagtatagctag

[0317] 예시적인 COL4A5 전이유전자 서열 (SEQ ID NO: 26)

[0318] atgaaactgcgtggagt cagcctggctgcccgttgttcttactggccctgagctcttgggggcagcctgcagaggctgaggcttgctatgggtgttctccaggtcaaagtgtgactgcagtgccataaaaggggaaaaggagagagagggttccagggttggaaaggacaccaggatggctggatttccagggtccagaaaggcctccggggcctcggggacaaaagggtgatgatggaattccagggccaccaggacaaaaggaatcagaggtcctcctggacttctggatttccagggtacaccaggtcttctggaatgccaggccacgatggggccccaggacctcaaggtatccccggatgcaatggaaccaaggagaaactggatttccaggcagttcccgttttctggtttacagggtcctccaggacccccgggatcccagggtatgaagggtgaaccaggtagtataatgatgcatcactgccaggaccaagggtaatccaggatccagggtcctcctggaatacaaggcctacctggctcccactggataccaggggcaattgggtccccaggaccaccaggtttgatgggccccctcctggtccaccaggacttccaggacctaaagggaatgggcttaatttccagggacccaagggtgaaaagggtgagcaaggtcttccaggcccacctggggccacctgggcagatcagtgaaacagaaaagaccaatgtatgtagagtttcagaaaaggagatcaggacttctggtgaccgaggccctcctggacctccaaggatcgtggtcctccagggtccccagggtgagaaaagggtgagaagggtgagcaaggagagccaggcaaaagggtaaaccaggcaaatggagaaaatggccaaccaggaaatcctggtttgcctggtgatcctggttacctggtgaaaccggaagggtggtaaaagggcaaaaagggtgacactggcccacctggacctcctggacttgtaatctctagacctgggactggtataactataggagaaaaggaaacatgggttgctgggttgctggagaaaaggagagcaggaatttctggaatccagggtccacctggccttctggacctccaggggctgcagttatgggtcctcctggcctcctggatttctggagaaaagggtcagaaaggatgaaggaccacctggaatttccatctctggacctcctggacttgacggacagcctggggctcctgggttccagggcctcctggcctgctggcctcacatctctcctagtgatgagatgtgaaccaggccctccaggccccaggatctccagggtgataaaggactccaaggagaaacagggtgaaagggtgacaaagggtgacacttgcctcaactgcatggaaactggtatctcagggcctccagggtcaacctggtttgccagggtctcccagggtcctccaggatctctggtttcctggacagaaaggggaaaaaggacaagctggtgcaactggtccaaaggataccaggcatccaggagctccagggtgctccaggcttctggatctaaagggtgaaactggtgatcctcactttccaggatgaagggtgacaaaggagagttgggttccccggagctccagggtcctggtttacctggcactcctggacaggatggattgccagggttctggcccgaaggagagcctggtggaattacttttaagggtgaaagaggtccccctgggaaccagggttaccaggctcccagggaataggggctatgggtccccctggttccggccctccaggcccagtaggtgaaaaaggcatacaagggtggcaggaaatccaggccagccaaggaatccagggtcctaaagggtatccagggtcagactataaccagccggggaaagcctggcttgcctggtaaaccaggcagagatggtgatgtaggtcttccagggtgacctggacttccagggcaaccaggcttgccagggtacctggtagcaaaaggagaaccagggtatcctggaattgggcttctggaccacctggtccc aaaggcttctggaatccaggacctccaggagacctgggacacctggaagaatggctctagaaggccctcctggccaccggcttccaggaccaagggtgaaaccaggatggatctacctggggccacctgggcccaccaggacttccagggttcaaaaggagcactgggtccaaaagggtgatcgtggttccaggacctccgggtcctccaggacgcactggcttagatgggctcctggacaaaagggtgatgtggacaaaatggacaacctggaccaatgggacctcctgggtgccaaggaataggtgtcagggaccaccaggaccaccagggttctgggccaataggtcaacctggtttacatggaataccaggagagaaggggatccaggacctcctggacttgatgttccaggacccccagggtgaaagagcagctccagggtatccccggagcactggtcctataggacctccaggatccagggttccaggaaagcaggtgcctctggatttccagggtaccaaaagggtgaaatgggtatgatgggacctccaggcccaccaggaccttgggaattcctggcaggagtggtgta cctggtcttaagggtgatgatggcttgcagggtcagccaggacttctggcctacaggagaaaaagggtagt aaaggagagcctggccttccaggccctcctggaccaatggatccaaatctctgggctcaaaaggagagaagggggaacctggcttaccagggtatacctggagtttcaggggcaaaaagggtatcagggtttg cctggagaccaggggcaacctggactgagtggacaacctggataccaggaccaccagggtccaaaggtaacctggtctcctggacagccaggctctataggacctcctggacttaaaaggaacctcgggtgatgggtttccagggcctcagggtgtggaaggcctcctggaccttctggagtctggacaacctggctccccaggatctacctggacagaaaggcgacaaagggtgatcctggtatcaagcatggcttccagggtctcctggtccaaagggtgagcctggtctgctggataccagggaacctggtatcaaaagggtctgtgggagatcctggtttgcccggataccaggaaacctggagcaaaaggacaaccaggcctcctggatttccaggaaaccaggccctcctggacaaaagggtatagtgccctcctgggaacccccggccttccaggagaaacctggtcctgtagggtggagggtcatcctgggcaaccaggccctccaggcgaaaaggcaaacccgggtcaagatggtatcctggaccagctggacagaagggtgaaccaggtcaaccaggcttggaa

aaccaggaccccctggacttcaggactttctggccaaaagggtgatggaggattacctgggatccaggaaatcctggcctccagggtccaaagggcgaa  
ccaggctttcacggtttccctggtgtcagggtccccaggccctcctggttctccgggtccagctctggaaggacctaaaggcaaccctgggccccagggt  
cctcctgggagaccaggctaccagggtccagaaggtcctccagggtcctcctggaaatggaggattaaaggagagaagggaaatccaggccaacctgggcta  
cctggcttgctggtttgaaaggagatcaaggaccaccaggactccagggtaatcctggccggccgggtctcaatggaatgaaaggagatcctggtctcct  
ggtgttcaggattcccaggcatgaaaggaccagtgagtagctggatcagctggccctgagggggaaccgggacttatggctcctccagggtcctcctgga  
ttacctggtccttcaggacagatataaattaaaggagatgctggtcctccaggaatcctggccagcctgggctaaagggtctaccaggaccccaagga  
cctcaaggcttaccagggtccaactggcctccaggagatcctggacgcaatggactcctggctttgatggtgcaggaggcgcaaggagaccagggtctg  
ccaggacagccagggtaccctggtttggatggtccccctggtccagatggattgcaagggtccccagggtccccctggaacctcctctgttgacatggatt  
cttatfacacgccacagccagacaacggatgcaccacaatgccacagggaaccttcagggtctatgaaggcttttctcctctgtatgtacaaggaaataaa  
agagcccacgggtcaagacttggggacgggtggcagctgccttcgctcgttttagtaccatgcctttcatgttctgcaacatcaataatgtttgcaactttgct  
tcaagaaatgactattcttactggctcttaccacagagcccatgccaatgagcatgcaaccctaaaggccagagcatccagccattcattagtcgatgt  
gcagtagtgaagctccagctgtggtgtagcagttcacagtcagacatccagatccccattgctcctcagggtgggattctctgtggattggttatcc  
ttcatgatgatacaagtgaggggcagaaggctcagggtcaagccctagcctccccctggttctgcttggagggtttcgttcagctcccttcacgaatgt  
catgggggggtacctgtaactactatgccaactcctacagcttttggtggcaactgtagatgtgtcagacatgttcagtaaacctcagtcagaacgctg  
aaagcaggagacttgaggacacgaattagccgatgtcaagtgtgcatgaagaggacataa

[0319] COL4A3, COL4A4 또는 COL4A5 전이유전자는 유전자 발현을 개선시키기 위해 사용될 수 있는 인트론 또는 인트론 서열을 포함할 수 있다. COL4A3, COL4A4 또는 COL4A5 전이유전자는 헤마글루티닌(HA) 태그와 같은 단백질 태그를 포함할 수 있다. HA는 에피토프 태그로서 사용될 수 있고, 이것이 첨가된 단백질의 생물활성 또는 생물분포를 방해하지 않는 것으로 나타났다. 단백질 태그는 전이유전자의 검출, 분리 및 정제를 용이하게 할 수 있다. 다른 적합한 단백질 태그는 Myc 태그, 폴리히스티딘 태그 및 플래그 태그를 포함할 수 있다.

[0320] 일부 구체예에서, SEQ ID NO: 47에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는 프로모터는 COL4A3, COL4A4 또는 COL4A5 폴리펩티드 또는 이의 단편 또는 유도체를 인코딩하는 단백질-코딩 서열에 작동가능하게 연결된다.

[0321] 일부 구체예에서, SEQ ID NO: 47의 뉴클레오티드 서열로 구성된 프로모터는 COL4A3, COL4A4 또는 COL4A5 폴리펩티드, 또는 이의 단편 또는 유도체를 인코딩하는 단백질-코딩 서열에 작동가능하게 연결된다.

[0322] *콩팥 증후군(NS)-관련 전이유전자*

[0323] 단백질-코딩 서열은 NS-관련 전이유전자를 포함하거나 이로 구성될 수 있다.

[0324] *콩팥 증후군(NS)*은 유의한 단백뇨, 저알부민혈증, 부종 및 고지혈증을 특징으로 하는 만성 신장 질환이다. NS-관련 전이유전자는 단일유전자 형태의 NS와 관련되고 발세포에서 발현되는 유전자일 수 있다.

[0325] 적합한 NS-관련 전이유전자는 *NPHS2, ADCK4, ALG1, ARHGAP24, ARGHDIA, CD151, CD2AP, COQ2, COQ6, DGKE, E2F3, EMP2, KANK2, LAGE3, LMNA, LMX1B, MAFB, NUP85, NUP93, NXF5, OSGEP, PAX2, PDSS2, PMM2, PODXL, SCARB2, SGPL1, Smad7, TP53RK, TPRKB, VDR, WDR73, WT1, ZMPSTE24, APOL1, NPHS1, TRPC6, NUP107, NUP133, NUP160, ACTN4, INF2, ANKFY1, ANLN, CRB2, ITGA3, KANK1, ANK1, ANK KANK4, MAGI2, MYO1E, OCRL, PTPRO, SMARCAL1, SYNPO, TBC1D8B, XPO5, TNS2* 및 *NLRP3*을 포함한다.

[0326] 일부 구체예에서, SEQ ID NO: 47에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는 프로모터는 NS-관련 전이유전자에 작동가능하게 연결된다.

[0327] 일부 구체예에서, SEQ ID NO: 47의 뉴클레오티드 서열로 구성된 프로모터는 NS-관련 전이유전자에 작동가능하게 연결된다.

[0328] *NPHS2*

[0329] 단백질-코딩 서열은 *NPHS2* 또는 이의 단편 및/또는 변이체를 인코딩할 수 있다.

[0330] "NPHS2"는 *NPHS2* 유전자에 의해 인코딩되는 폴리펩티드의 약어이며, 포도신으로도 알려져 있다. *NPHS2*는 세극막에서 단백질 복합체의 핵심 성분인 42kDa 헤어핀 유사 막-관련 발세포-특이적 단백질이다; 인접한 발세포 발달 사이의 세포-세포 접합. 이는 지질 래프트에 국소화되고 네프린, CD2AP 및 TRPC6과 같은 다른 중요한 세극막

단백질과 상호작용한다. 이는 세극막의 유지, 및 결과적으로 사구체 여과 장벽의 완전성에 필수적이다.

- [0331] NPHS2의 단편 및/또는 변이체는 NPHS2 활성을 보유할 수 있다. 예를 들어, 포도신의 단편 및/또는 변이체는 사구체 투과성을 조절할 수 있다. 적합하게는, NPHS2의 단편 및/또는 변이체는 NPHS2와 동일하거나 유사한 활성을 가질 수 있고, 예를 들어, NPHS2의 활성의 적어도 40%, 적어도 50%, 적어도 60%, 적어도 70%, 적어도 80%, 적어도 90%, 적어도 95% 또는 적어도 100%를 가질 수 있다.
- [0332] 당업자는 NPHS2의 공지된 구조적 및 기능적 특징에 기반하여 (예를 들어, 문헌 [Tabassum, A., et al., 2014. Interdisciplinary Sciences: Computational Life Sciences, 6(1), pp.32-39] 참조) 및/또는 공지된 변이체에 기반하여 (예를 들어, NCBI Gene ID: 7827 및 NCBI HomoloGene: 22826 참조) 보존적 치환을 사용하여 단편 및/또는 변이체를 생성할 수 있을 것이다. 적합하게는, NPHS2의 단편 및/또는 변이체는 N- 및 C-말단에 2개의 세포질 도메인을 갖는 막횡단 도메인을 포함한다.
- [0333] NPHS2 유전자는 침팬지, 붉은털 원숭이, 개, 소, 마우스 및 래트에서 보존된다. NPHS2는 인간 NPHS2일 수 있다. 적합하게는, NPHS2는 UniProtKB 수탁 Q9NP85의 폴리펩티드 서열, 또는 이의 단편 및/또는 변이체를 포함하거나 이로 구성될 수 있다.
- [0334] 일부 구체예에서, NPHS2는 SEQ ID NO: 55와 적어도 70% 동일한 아미노산 서열 또는 이의 단편을 포함하거나 이로 구성된다. 적합하게는, NPHS2는 SEQ ID NO: 55와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일한 아미노산 서열 또는 이의 단편을 포함하거나 이로 구성된다.
- [0335] 일부 구체예에서, NPHS2는 SEQ ID NO: 55 또는 이의 단편을 포함하거나 이로 구성된다.
- [0336] MERRARSSSRESRGRGRTRPHKENKRAKAERSGGGRGRQEAPEPSGSRAGTPGEPRAPAATVVDVDEVRGSGEETVVALLESERPPEGTKSSGLGACEWLLVLI SLLFI IMTFPFSIWFCVKVVQEYERVI IFRLGHLLPGRAKGPLFFFLPCLDTHYHKVDLRLQTLIEIPFHEIVTKDMFIMEIDAICYRMENASLLLSSLAHVSKAVQFLVQTTMKRLLAHRSLTEILLERKSI AQDAKVALDSVTCIWGIKVERIEIKDVRLPAGLQHS LAVEAEAQRQAKVRMIAAEAEKAASESLRMAAEILSGTAAVQLRYLHQLSLSTEKSTVVLPLPFDLLNCLSSPSNRTQGSLPFPSPSKPVEPLNPKKKDSPML
- [0337] 예시적인 NPHS2 아미노산 서열(SEQ ID NO: 55)
- [0338] 일부 구체예에서, NPHS2 전이유전자는 SEQ ID NO: 56과 적어도 70% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다. 적합하게는, NPHS2 전이유전자는 SEQ ID NO: 56과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다.
- [0339] 일부 구체예에서, NPHS2 전이유전자는 뉴클레오티드 서열 SEQ ID NO: 56 또는 이의 단편을 포함하거나 이로 구성된다.
- [0340] ATGGAGAGGAGGGCGCGGAGCTCCTCCAGGGAGTCCCGCGGGCGAGGCGGCAGGACTCCGCACAAGGAGAACAAGAGGGCAAAGGCCGAGAGGAGCGGGCGGGGCGCCGCGGGCGCCAGGAGGCTGGGCCCGAGCCGTCGGGCTCGGACGGGCGGGGACCCCGGGGAGCCCCGAGCGCCCGCCACGGTGTGGACGTGGATGAGGTCCGAGGCTCCGGCGAGGAGGGCACCGAGGTGGTGGCGCTGTTGAGAGCGAGCGGCCGAGGAAGGTACCAAATCCTCCGGCTTAGGGCCTGTGAGTGGCTTCTTGTCTCATTCCCTGTCTTTCATCATCATGACCTTCCCTTTTTCCATCTGGTTCTGCGTAAAGGTTGTACAAGAGTATGAAAGAGTAATTATA TTCCGACTGGGACATCTGCTTCTGGAAGAGCCAAAGGCCCTGGTCTTTTCTTTTTTTGTCCTGCTGGATACTACCACAAGGTTGACCTTCGTCTCCAAACTCTGGAGATACCTTTTCATGAGATCGTGACCAAGACATGTTATAATGGAGATAGATGCCATTTGCTACTACCGAATGAAAAATGCCTCTCTTCTCCTAAGCAGTCTTGCTCATGTATCTAAAGCTGTGCAATTCCTTGTGCAAAACCACTATGAAGCGTCTCCTAGCACATCGATCCCTCACTGAAATTTCTTCTAGAGAGGAAGAGCATCGCCAAAGATGCAAAGGTTGCTTGGATTGAGTACCTGTATTGGGAATCAAAGTGGAGAGAATAGAAATTAAGATGTGAGGTTGCCAGCTGGGCTTCAGCACTACTGGCTGTGGAGGCTGAAGCGCAAAGACAAGCAAAGTGGGATGATTGCTGCAGAAAGCGAAAGGCTGCTTCTGAGTCCCTGAGGATGGCAGCTGAGATTCTGTGAGGACCCCTGCTGTGTTGACGTTTCGATACCTCCACACCCCTCAGTCTCTGTCCACAGAGAAGCCTTCCACTGTGGTTTACCTTTGCCATTTGACCTACTGAATTGCTGTCTTCTCCAGCAACAGAACTCAGGAAGCCTCCCTTCCCAAGTCTTCCAAACCTGTTGAGCCACTAAATCCTAAAAAGAAAGACTCTCCCATGTTA
- [0341] 예시적인 NPHS2 전이유전자 서열(SEQ ID NO: 56)
- [0342] 일부 구체예에서, NPHS2 전이유전자는 SEQ ID NO: 57과 적어도 70% 동일한 뉴클레오티드 서열 또는 이의 단편을 포함하거나 이로 구성된다. 적합하게는, NPHS2 전이유전자는 SEQ ID NO: 57과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일한 뉴클레오티드 서열

또는 이의 단편을 포함하거나 이로 구성된다.

[0343] 일부 구체예에서, NPHS2 전이유전자는 뉴클레오티드 서열 SEQ ID NO: 57 또는 이의 단편을 포함하거나 이로 구성된다.

[0344] ATGGAGAGGAGGGCGGGAGCTCCTCCAGGGAGTCCCAGGGGCGAGGCGGCAGGACTCCGCACAAGGAGAACAAGAGGGCAAAGGCCGAGAGGAGCGGCGGA  
GGCCCGGGCGCCAGGAGGCTGGGCCGAGCCGTCGGGCTCCGGACGGGCGGGGACCCCGGGGGAGCCCCGAGCGCCCGCCACGGTGGTGGACGTGGAT  
GAGGTCGAGGCTCCGGCGAGGAGGGCACCGAGGTGGTGGCGCTGTTGGAGAGCGAGCGGCCGAGGAAGGTACCAAAATCCTCCGGCTTAGGGGCTGTGAG  
TGGCTTCTTGTCTCATTCCCTGCTCTTCATCATCATGACCTTCCCTTTTTCCATCTGGTTCTGCGTAAAGGTTGTACAAGAGTATGAAAGAGTAATTATA  
TTCCGACTGGGACATCTGCTTCTGGAAGAGCCAAAGGCCCTGGTCTTTTCTTTTTTTTGCCTGCTGGATACCTACCACAAGTTGACCTTCGTCTCCAA  
ACTCTGGAGATACCTTTTCATGAGATCGTGACCAAAGACATGTTTATAATGGAGATAGATGCCATTTGCTACTACCGAATGAAAAATGCCTCTCTTCTCCTA  
AGCAGTCTTGTCTATGATCTAAAGCTGTGCAATTCCTTGTGCAAAACCACTATGAAGCGTCTCCTAGCACATCGATCCCTCACTGAAATTTCTTCTAGAGAGG  
AAGAGCATCGCCCAAGATGCAAAGTTGCCTTGGATTTCAGTGCCTGTATTTGGGAATCAAAGTGGAGAGAATAGAAATTAAGATGTGAGGTTGCCAGCT  
GGGCTTCAGCACTACTGGCTGTGGAGGCTGAAGCGCAAAGACAAGCCAAAGTGCGGATGATTGCTGCAGAAGCGGAAAAGGCTGCTTCTGAGTCCCTGAGG  
ATGCGAGTGTGATTCTGTGAGGACCCCTGCTGCCGTTTCAGTCTGCATACCTCCACACCCTTCAGTCTGTGTCACAGAGAAGCCTTCCACTGTGGTTTAA  
CCTTTGCCATTTGACCTACTGAATTGCTGTCTTCTCCAGCAACAGAACTCAGGAAGCCTCCCTTCCCAAGTCTTCCAAACCTGTTGAGCCACTAAAT  
CCTAAAAAGAAAGACTCTCCCATGTTATAG

[0345] 예시적인 NPHS2 전이유전자 서열(SEQ ID NO: 57)

[0346] 일부 구체예에서, SEQ ID NO: 47에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된 프로모터는 NPHS2를 인코딩하는 단백질-코딩 서열 또는 이의 단편 및/또는 변이체에 작동가능하게 연결된다.

[0347] 일부 구체예에서, SEQ ID NO: 47의 뉴클레오티드 서열로 구성된 프로모터는 NPHS2를 인코딩하는 단백질-코딩 서열, 또는 이의 단편 및/또는 변이체에 작동가능하게 연결된다.

[0348] *혈관 내피 성장 인자(VEGF)C 전이유전자*

[0349] 단백질-코딩 서열은 혈관 내피 성장 인자(VEGF)C 전이유전자를 포함하거나 이로 구성될 수 있다.

[0350] VEGFC는 2개의 수용체, VEGFR-3(Flt4) 및 VEGFR-2(Flk4)를 통해 신호전달하는 것으로 알려진 림프관형성 성장 인자이다. VEGFC는 사랑체로 절단되기 전에 이량체화되는 프리프로펩티드 형태로 세포에 의해 생성된다.

[0351] VEGFC 전이유전자는 프리프로펩티드 형태, 사랑체 형태, 중간체 형태 또는 완전히 가공된 성숙 VEGFC와 같은 임의의 형태의 VEGFC를 인코딩하는 폴리뉴클레오티드를 포함할 수 있다.

[0352] 요망되는 경우, 상이한 형태의 VEGFC 폴리펩티드를 인코딩하는 폴리뉴클레오티드는 임의의 조합으로 사용될 수 있다. 바람직하게는, VEGFC 전이유전자는 VEGFC 생물학적 활성을 갖는 하나 이상의 폴리펩티드, 즉, VEGFR-2 및/또는 VEGFR-3에 결합하고 이를 활성화시킬 수 있는 펩티드를 인코딩하는 폴리뉴클레오티드를 포함한다. 더욱 바람직하게는, VEGFC 전이유전자는 VEGFC 상동성 도메인을 포함하고 VEGFC 생물학적 활성을 갖는 폴리펩티드, 즉, VEGFR-2 및/또는 VEGFR-3에 결합하고 이를 활성화시킬 수 있는 폴리펩티드를 인코딩하는 폴리뉴클레오티드를 포함한다. 적합한 VEGFC 폴리뉴클레오티드 및 폴리펩티드의 추가 세부사항은 WO 2015/022447 및 US 2014/0087002에 기재된 것들을 포함한다.

[0353] VEGFC 폴리뉴클레오티드는 SEQ ID NO: 36의 VEGFC 오픈 리딩 프레임(ORF) 서열을 포함할 수 있다. VEGFC 폴리뉴클레오티드는 SEQ ID NO: 36의 VEGFC ORF에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 서열 동일성을 갖는 핵산 서열을 포함할 수 있다. 변이체 서열은 VEGFR-2 및 VEGFR-3에 결합하고 활성화하는 능력을 보유하는 VEGFC 폴리펩티드를 인코딩할 수 있다.

[0354] *예시적인 VEGFC 폴리뉴클레오티드(SEQ ID NO: 36)*

[0355] ATGCACTGTGCTGGGCTTCTTCTGTGGCGTGTCTCTGCTCGCCGCTGCGCTGCTCCCGGTCTCGCGAGGCGCCCGCCGCGCCGCTTCGAGTCC  
GGACTCGACCTCTCGGACGCGGAGCCGACGCGGGCGAGGCCACGGCTTATGAAGCAAAGATCTGGAGGAGCAGTTACGGTCTGTGTCCAGTGTAGATGAA  
CTCATGACTGTACTCTACCCAGAATATTGAAAAATGTACAAGTGTGAGTAAAGGAGGCTGGCAACATAACAGAGAACAGGCCAACCTCACTCAAGG  
ACAGAAGAGACTATAAAAATTTGCTGCGAGCACATTATAATACAGAGATCTGAAAAGTATTGATAATGAGTGGAGAAAGACTCAATGCATGCCACGGGAGGTG

TGTATAGATGTGGGAAGGAGTTGGAGTCGCGACAAACACCTTCTTTAAACCTCCATGTGTGTCGGTCTACAGATGTGGGGTTGCTGCAATAGTGAGGGG  
 CTGCAGTGCATGAACACCAGCACGAGCTACCTCAGCAAGACGTTATTTGAAATTACAGTGCCTCTCTCAAGGCCCAACCAGTAACAATCAGTTTTGCC  
 AATCACACTTCTGCGGATGCATGTCTAAACTGGATGTTTACAGACAAGTTCATTCCATTATTAGACGTTCCCTGCCAGCAACACTACCACAGTGCAGGCA  
 GCGAACAAGACCTGCCCAACAATTACATGTGGAATAATCACATCTGCAGATGCCTGGCTCAGGAAGATTTATGTTTTCTCGGATGCTGGAGATGACTCA  
 ACAGATGGATTCCATGACATCTGTGGACCAACAAGGAGCTGGATGAAGAGACCTGTGAGTGTCTGCAGAGCGGGCTTCGGCCTGCCAGCTGTGGACCC  
 CACAAAGAACTAGACAGAAACTCATGCCAGTGTGTCTGTAACAACTCTTCCCAGCCAATGTGGGGCCAACCGAGAATTGATGAAAACACATGCCAG  
 TGTGTATGTAAGAAGCTGCCAGAAATCAACCCCTAACTCTGGAAAATGTGCCTGTGAATGTACAGAAAGTCCAGAAATGCTTGTAAAAGGAAAG  
 AAGTTCACCACCAACATGCAGCTGTACAGACGGCCATGTACGAACCGCCAGAAGGCTTGTGAGCCAGGATTTTCATATAGTGAAGAAGTGTGTCGTTGT  
 GTCCCTTCATATTGAAAAGACCACAAATGAGC

- [0356] 일부 구체예에서, SEQ ID NO: 47에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된 프로모터는 VEGF(C) 전이유전자에 작동가능하게 연결된다.
- [0357] 일부 구체예에서, SEQ ID NO: 47의 뉴클레오티드 서열로 구성된 프로모터는 VEGF(C) 전이유전자에 작동가능하게 연결된다.
- [0358] *보체 단백질 및 보체 억제제*
- [0359] 단백질-코딩 서열은 보체 단백질 또는 이의 단편 및/또는 변이체를 인코딩한다.
- [0360] 본원에서 사용되는 "보체 단백질"은 보체 시스템의 일부인 단백질이다. 보체 캐스케이드로도 알려진 보체 시스템은 외래 및 변경된 숙주 세포에 대한 1차 방어선으로 작용하는 선천 면역의 중심 부분이다. 보체 시스템은 간 또는 세포 표면에서 발현되는 막 단백질에 의해 주로 생성되는 혈장 단백질로 구성된다. 보체는 혈장, 조직 또는 세포 내에서 작동한다. 보체 단백질은 캐스케이드로서 협력하여 병원체를 옅소닌화하고 면역 세포가 감염과 싸우고 항상성을 유지하도록 돕는 일련의 염증 반응을 유도한다(Merle, N.S., et al., 2015. *Frontiers in immunology*, 6, 262).
- [0361] 보체 활성화의 3가지 경로가 있다: 고전적, 대안적 및 렉틴 경로. 3개의 보체 경로는 표적 인식의 메카니즘에서 상이하지만 중심 성분 C3의 활성화에 수렴한다. 이러한 활성화 후, C5가 절단되고, 막 공격 복합체(MAC)의 조립이 개시된다. C3 및 C5의 효소적 절단은 아나필로톡신 C3a 및 C5a의 생성 및 방출을 초래한다.
- [0362] 적합하게는, 보체 단백질은 CFI, CFH, FHL-1, C1INH, C4BP, MASP2, C3, C5aR1, C5, C5a, CD55, CD35, CD46, CD59, 비트로넥틴 및 클러스터린, 또는 이의 단편 및/또는 변이체로 구성된 목록으로부터 선택된다.
- [0363] 단백질-코딩 서열은 보체 시스템의 억제제, 또는 이의 단편 및/또는 변이체를 인코딩할 수 있다.
- [0364] 본원에서 사용되는 "보체 시스템의 억제제" 또는 "보체 억제제"는 보체 시스템의 활성화를 방지하는 단백질이다. 보체는 원치 않는 보체 활성화로부터 자가 세포 및 조직을 자연적으로 보호하는 이러한 억제제에 의해 엄격하게 제어된다. 보체 억제제는 고전적, 렉틴, 및 대체 경로의 상이한 단계에서 보체 활성화를 조절할 수 있다. 적합하게는, 보체 억제제는 자연-발생 보체 억제제, 또는 이의 단편 및/또는 변이체이다. 바람직하게는, 보체 시스템의 억제제는 인간에서 보체 시스템의 억제제이다.
- [0365] 보체 억제제는 가용성 억제제 및 막-결합 억제제의 2개의 카테고리로 분류된다. 바람직하게는, 보체 시스템의 억제제는 가용성 보체 억제제이다. 가용성 보체 억제제는 C1 억제제(C1INH), 보체 인자 I(CFI), 보체 인자 H(CFH), 보체 인자 H-유사 단백질 1(FHL-1), C4 결합 단백질(C4BP), 클러스터린 및 비트로넥틴을 포함한다. 막-결합 조절제는 CD46, CD55, CD59, CD35 및 CUB 및 Sushi 다중 도메인 1(CSMD1)을 포함한다.
- [0366] 보체 시스템의 억제제는 CFI, CFH, FHL-1, C1INH, C4BP, CD46, CD55, CD59, CD35, 비트로넥틴, 클러스터린 및 CSMD1, 또는 이의 단편 및/또는 변이체로부터 선택될 수 있다.
- [0367] 바람직하게는, 보체 시스템의 억제제는 CFI, CFH 및 FHL-1, 또는 이의 단편 및/또는 변이체로부터 선택된다.
- [0368] *CFI*
- [0369] 단백질-코딩 서열은 CFI, 또는 이의 단편 및/또는 변이체를 인코딩한다.
- [0370] 보체 인자 I(CFI)는 트립신-유사 세린 프로테아제이며, 이는 C3b의 알파-사슬에서 3개의 펩티드 결합 및 C4b의 알파-사슬에서 2개의 결합을 절단하여 이들 단백질을 불활성화시킴으로써 보체 시스템을 억제한다.

- [0371] CFI는 디설파이드 연결된 중쇄 및 경쇄로 구성된 당단백질 이중이량체이다. 중쇄는 4개의 도메인을 갖는다: FI 막 공격 복합체(FIMAC) 도메인, CD5 도메인, 및 저밀도 지단백질 수용체 1 및 2(LDLr1 및 LDLr2) 도메인. 중쇄는 기질(C3b 또는 C4b) 및 보조인자 단백질(인자 H, C4b-결합 단백질, 보체 수용체 1, 및 막 보조인자 단백질)에 의해 형성된 복합체를 만날 때까지 효소를 불활성으로 유지하는데 억제 역할을 한다. 기질:보조인자 복합체에 대한 효소의 결합시, 중쇄:경쇄 계면이 파괴되고, 효소는 알로스테리에 의해 활성화된다. 경쇄는 세린 프로테아제 도메인만을 함유한다. 이 도메인은 C3b 및 C4b의 특정 절단을 담당하는 촉매 트리어드 His-362, Asp-411 및 Ser-507을 함유한다.
- [0372] CFI 또는 이의 단편 및/또는 변이체는 C3b를 C3b로 절단할 수 있고/있거나 iC3b를 C3d,g로 절단할 수 있다.
- [0373] CFI의 단편 및/또는 변이체는 천연 CFI의 C3b-불활성화 및 iC3b-분해 활성의 적어도 50%, 60%, 70%, 80%, 90%, 95% 또는 100%를 유지할 수 있다. CFI 및 천연 CFI의 단편 및/또는 변이체의 C3b-불활성화 및 iC3b-분해 활성은 당업자에게 공지된 임의의 적합한 방법을 사용하여 결정될 수 있다. 예를 들어, 단백질분해 점정을 이용한다.
- [0374] 바람직하게는, CFI는 인간 CFI이다. 예시적인 인간 CFI는 UniProtKB 수탁 번호 P05156을 갖는 CFI이다.
- [0375] 적합하게는, CFI는 SEQ ID NO: 49로서 제시된 폴리펩티드 서열 또는 SEQ ID NO: 49와 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다.
- [0376] 예시적인 CFI 폴리펩티드 서열(SEQ ID NO: 49):
- [0377] M K L L H V F L L F L C F H L R F C K V T Y T S Q E D L V E K K C L A K K Y T H L S C D K V F C Q P W Q R C I E G T C V C K L P Y Q C P K N G T A V C A T N R R S F P T Y C Q Q S L E C L H P G T K F L N  
N G T C T A E G K F S V S L K H G N T D S E G I V E V K L V D Q D K T M F I C K S S W M R E A N V A C L D L G F Q Q G A D T Q R R F K L S D L S I N S T E C L H V H C R G L E T S L A E C T F T K R R T M  
G Y Q D F A D V V C Y T Q K A D S P M D D F F Q C V N G K Y I S Q M K A C D G I N D C G D Q S D E L C C K A C Q G K G F H C K S G V C I P S Q Y Q C N G E V D C I T G E D E V G C A G F A S V T Q E E T E I  
L T A D M D A E R R R I K S L L P K L S C G V K N R M H I R R K R I V G G K R A Q L G D L P W Q V A I K D A S G I T C G G I Y I G G C W I L T A A H C L R A S K T H R Y Q I W T T V V D W I H P D L K R I V  
I E Y V D R I I F H E N Y N A G T Y Q N D I A L I E M K K D G N K K D C E L P R S I P A C V P W S P Y L F Q P N D T C I V S G W G R E K D N E R V F S L Q W G E V K L I S N C S K F Y G N R F Y E K E M E C  
A G T Y D G S I D A C K G D S G G P L V C M D A N N V T Y V W G V S W G E N C G K P E F P G V Y T K V A N Y F D W I S Y H V G R P F I S Q Y N V
- [0378] 적합하게는, 변이체는 SEQ ID NO: 49와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0379] CFI를 인코딩하는 뉴클레오티드 서열의 예는 NM\_000204.5이다. 적합하게는, CFI를 인코딩하는 단백질-코딩 서열은 SEQ ID NO: 50으로 제시된 폴리뉴클레오티드 서열 또는 SEQ ID NO: 50과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다.
- [0380] 예시적인 CFI 폴리뉴클레오티드 서열(SEQ ID NO: 50):
- [0381] A T G A A G C T T C T C A T G T T T T C C T G T T A T T T C T G T G C T T C C A C T T A A G G T T T T G C A A G G T C A C T T A T A C A T C T C A A G A G G A T C T G G T G G A G A A A A A G T G C T T A  
G C A A A A A A T A T A C T C A C C T C T C T G C G A T A A A G T C T T C T G C C A G C C A T G G C A G A G A T G C A T T G A G G G C A C C T G T G T T T G T A A A C T A C C G T A T C A G T G C C C A  
A A G A A T G G C A C T G C A G T G T G T G C A A C T A A C A G G A G A A G C T T C C C A A C A T A C T G T C A A C A A A A G A G T T T G G A A T G T C T T C A T C C A G G G A C A A A G T T T T A A A T  
A A C G G A A C A T G C A C A G C C G A A G G A A A G T T T A G T G T T C C T T G A A G C A T G G A A T A C A G A T T C A G A G G A A T A G T T G A A G T A A A C T T G T G G A C C A A G A T A A G  
A C A A T G T T C A T A T G C A A A G C A G C T G G A G C A T G A G G G A A G C C A A C G T G G C C T G C C T T G A C C T T G G G T T T C A A C A A G G T G C T G A T C T A A A G A A G G T T T A A G  
T T G T C T G A T C T C T A T A A A T T C C A C T G A A T G T C A T A T G C A T T G C C A G G A T T A G A G A C C A G T T T G G C T G A A T G T A C T T T T A C T A A G A G A A A C A G A A T G  
G G T T A C C A G G A T T T C G C T G A T G T G G T T T G T T A T A C A C A G A A A G C A G A T T C C C A A T G G A T G A C T T C T T C A G T G T G T G A A T G G G A A T A C A T T T C T C A G A T G  
A A A G C C T G T G A T G T A T C A A T G A T T G T G G A G A C C A A A G T G A A C T G T G T T G T A A A G C A T G C C A A G G C A A A G G C T T C C A T T G C A A A T C G G G T G T T T G C A T T  
C C A A G C C A G T A T C A A T G C A A T G G T G A G G T G G A C T G C A T T A C A G G G A A G A T G A A G T T G G C T G T G C A G G C T T T G C A T C T G T G A C T C A A G A A G A A C A G A A A T T  
T T G A C T G C T G A C A T G G A T G C A G A A A A G A C G G A T A A A A C A T T A T T A C C T A A A C T A T C T T G T G G A G T T A A A A C A G A A T G C A C A T T C G A A G G A A C G A A T T  
G T G G G A G G A A A G C G A C A A C T G G G A G A C C T C C C A T G G C A G G T G G C A A T T A A G G A T G C C A G T G G A A T C A C C T G T G G G G A A T T A T A T T G G T G G C T G T T G G  
A T T C T G A C T G C T G C A C A T T G T C T C A G A G C C A G T A A A A C T A C T G T T A C C A A A T A T G G A C A C A G T A G T A G A C T G G A T A C A C C C G A C C T T A A C G T A T A G T A  
A T T G A A T A C G T G G A T A G A A T T A T T T C C A T G A A A C T A C A A T G C A G G C A C T T A C C A A A T G A C A T C G C T T T G A T T G A A A T G A A A A A G A C G G A A C A A A A A  
G A T T G T G A G T G C C T G T T C C A T C C C T G C C T G T G T C C C C T G T C C T T A C C A T T C C A C C T A A T G A T A C A T G C A T C G T T T C G G C T G G G A C G A G A A A A  
G A T A A C G A A A G A G T C T T T C A C T T C A G T G G G T G A A G T T A A C T A A T A A G C A A C T G C T A A G T T T A C G A A A T C G T T T C T A T G A A A A G A A A T G G A A T G T  
G C A G G T A C A T A T G A T G T T C C A T C G A T G C C T G T A A A G G G A C T T G G A G G C C C T A G T C T G T A T G G A T G C C A A C A A T G T G A C T T A T G T C T G G G T G T T G T G  
A G T T G G G G G A A A C T G T G G A A A C C A G A G T T C C C A G G T G T T A C A C C A A A G T G G C A A T T A T T T G A C T G G A T T A G C T A C C A T G T A G G A A G G C C T T T A T T  
T C T C A G T A C A A T G T A T A A
- [0382] 적합하게는, 변이체는 SEQ ID NO: 50와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도

96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

- [0383] 일부 구체예에서, SEQ ID NO: 47에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된 프로모터는 CFI를 인코딩하는 단백질-코딩 서열 또는 이의 단편 및/또는 변이체에 작동가능하게 연결된다.
- [0384] 일부 구체예에서, SEQ ID NO: 47의 뉴클레오티드 서열로 구성된 프로모터는 CFI를 인코딩하는 단백질-코딩 서열, 또는 이의 단편 및/또는 변이체에 작동가능하게 연결된다.
- [0385] CFH
- [0386] 단백질-코딩 서열은 CFH 또는 이의 단편 및/또는 변이체를 인코딩한다.
- [0387] 보체 인자 H(CFH)는 자가 세포 및 표면에서 보체 활성화를 조절한다. CFH는 C3b에 대한 보체 인자 B(CFB)의 결합에 대해 경쟁하고, C3b의 CFI-촉매된 단백질분해 절단에 대한 보조인자로서 작용하고, C3bBb 및 C3b2Bb의 개별 성분으로의 비가역적 해리를 가속화한다. 따라서, CFH는 전환효소의 형성을 억제할 뿐만 아니라 형성되는 임의의 전환효소 복합체의 수명을 단축시킨다.
- [0388] CFH는 큰(155 kDa) 가용성 당단백질이다. CFH는 총 20개의 도메인으로 구성되며, 각각은 대략 60개의 아미노산 잔기를 함유하고 3-8개의 잔기로 구성된 짧은 링커에 의해 연결된 보체 제어 단백질 모듈(CCP) 또는 짧은 컨센서스 반복부로 치장된다. CCP 모듈은 (단백질의 N-말단으로부터) 1-20으로 번호가 매겨진다: CCP 1-4 및 CCP 19-20은 C3b와 맞물리는 반면, CCP 7 및 CCP 19-20은 GAG 및 시알산에 결합한다.
- [0389] CFH 또는 이의 단편 및/또는 변이체는 C3b 및/또는 C3d에 결합할 수 있고/거나; C3b의 CFI-촉매된 단백질분해 절단을 위한 보조인자로서 작용하고/거나; C3bBb 및 C3b2Bb의 이들의 개별 성분으로의 비가역적 해리를 증가시킬 수 있다. CFH의 단편 및/또는 변이체는 천연 CFH의 활성의 적어도 50%, 60%, 70%, 80%, 90%, 95% 또는 100%를 유지할 수 있다. CFH 및 천연 CFH의 단편 및/또는 변이체의 활성은 당업자에게 공지된 임의의 적합한 방법을 사용하여 결정될 수 있다.
- [0390] 바람직하게는, CFH는 인간 CFH이다. 예시적인 인간 CFH는 UniProtKB 수탁 번호 P08603을 갖는 CFH이다.
- [0391] 적합하게는, CFH는 SEQ ID NO: 51로 제시된 폴리펩티드 서열 또는 SEQ ID NO: 51과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다.
- [0392] 예시적인 CFH 폴리펩티드 서열(SEQ ID NO: 51):
- [0393] MRLLAKIICMLLWAI CVAEDCNELPPRRNTEILTGSWSQTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLRKCQKRPCGHPGDTFPGTFTLTGGN VFEYGVKAVYTCNEGYQLLGEINYRECDTDGWTNDIPICEVVKCLPVTAPENGIIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFWSKKEPK CVEISCKSPDVINGSPI SQKI IYKENERFQYKCNMGYEYSERGDVCTESGWRPLPSCEEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNT AKCTSTGWIPAPRCTLKPCDYPDIKHGGLYHENMRRPYFPVAVGKYYSYCDHEFETPSGSYWDHIHCTQDQGWSPA VPLCRKCYFPYLENGYNQNYGRKFVQ GKSIDVACHPGYALPKAQTTVTCTMENGWSPTPRCIRVKTCSSKSIDIENGFISESQYTYALKEKAKYQCKLGYVTADGETSGSITCGKDGWSAQPTCIKSCD IPVFMNARTKNDFTWFKLNDTLDYEDHDGYESNTGSTTGSIVCGYNGWSDLPICYERECELPKIDVHLPDRKKDQYKVGELKFSCKPGFTIVGPNSVQCY HFGLSPDLPICKEQVQSCGPPPELLNGNVKEKTKEEYGHSEVVEYCNPRFLMKGPNKIQCVDGWETTLPVCIVEESTCGDIPPELHGWAQLSSPPYYGDS VEFNCSESFMTIGHRSITCIHGVTQLPQCV AIDKLLKCKSSNLIILEEHLKNKKEFDHNSNIRYRCRGKEGW IHTVCINGRWDPEVNCMAQIQLCPPPPQ IPNSHNMTTTLNRYRDEKAVSVLCQENYLIQEGEEITCKDGRWQSIPLCVEKIPCSQPPQIEHGTINSSRSSQESYAHGTKLSYTCGGFRISEENETTCYMG KWSSPPQCEGLPCKSPPEISHGVVAHMSDSYQYGEVTYKCFEGFGIDGPAIAKCLGEKWSHPPSCIKTDCLSLPSFENAI PMGEKKDVYKAGEQVYTYTCAT YYKMDGASNVTICINSRWTGRPTCRDTSVNPPTVQNAVIVSRQMSKYPSGERVRYQCRSPYEMFGDEEVMCLNGNWTEPPQCKDSTGKCGPPPPIDNGDITS FPLSVYAPASSVEYQCQNLQLEGNKRITCRNGQWSEPPKHLPCVISRIMENYNIALRWTAKQLYSRTGESVEFVCKRGYRLSSRSHTLRTTCWDGKLE YPTCAKR
- [0394] 적합하게는, 변이체는 SEQ ID NO: 51과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0395] CFH를 인코딩하는 뉴클레오티드 서열의 예는 NM\_000186.4이다. 적합하게는, CFH를 인코딩하는 단백질-코딩 서열은 SEQ ID NO: 52로 제시된 폴리뉴클레오티드 서열 또는 SEQ ID NO: 52와 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다.
- [0396] 예시적인 CFH 폴리뉴클레오티드 서열(SEQ ID NO: 52):

[0397] ATGAGACTTCTAGCAAAGATTATTTGCCTTATGTTATGGGCTATTTGTGTAGCAGAAGATTGCAATGAACTTCCTCCAAGAAGAAATACAGAAATTTCTGACAGGTTCTGGTCTGACCAAACATATCCAGAAGGCACCCAGGCTATCTATAAATGCCGCCCTGGATATAGATCTCTTGAAATGTAATAATGGTATGCAGGAAGGGAGAATGGGTTGCTCTTAATCCATTAAGGAAATGTCAGAAAAGGCCCTGTGGACATCTCGGAGATACTCCTTTTGGTACTTTTACCCTTACAGGAGGAAATGTGTTGAATATGGTGTAAAAGCTGTGTATACATGTAATGAGGGGTATCAATTGCTAGGTGAGATTAATTACCGTGAATGTGACACAGATGGATGGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTACCAGTGACAGCACCAGAGAATGAAAAAATGTGTCAGTAGTGAATGGAACCAGATCGGGAAATACCATTTTGGACAAGCAGTACGGTTTGTATGTAACCTCAGGCTACAAGATTGAAGGAGATGAAGAAATGCATTGTTTCAGACGATGGTTTTTGGAGTAAAGAGAAACAAAGTGTGTGAAAATTTTCATGCAAATCCCAGATGTTATAAATGGATCTCCTATATCTCAGAAGATTATTTATAAGGAGAATGAACGATTTCATATAAATGTAACATGGGTTATGAATACAGTGAAAGAGGAGATGCTGTATGCACTGAATCTGGATGGCGTCCGTTGCCTTCATGTGAAGAAAAATCATGTGATAATCCTTATATTCCAAATGGTGACTACTCACCTTTAAGGATTAACACAGAAGCTGGAGATGAAATCACGTACCAGTGTAGAAATGGTTTTTATCTGCAACCCGGGAAATACAGCAAATGCACAAGTACTGGCTGGATACCTGCTCCGAGATGTACCTTGAACCTTGTGATTATCCAGACATTAACATGGAGGTCTATATCATGAGAATATGCGTAGACCATACTTTCCAGTAGCTGTAGGAAAATATTACTCCTATTACTGTGATGAACATTTTGGAGACTCCGTCAGGAAGTTACTGGGATCACATTTCATTGCACACAAGATGGATGGTCCGACAGTACCATGCCTCAGAAAATGTTATTTTCCCTATTTGGAAAATGGATATAATCAAATCATGGAAGAAAGTTGTACAGGGTAAATCTATAGACGTTGCCTGCCATCCTGGCTACGCTCTCCAAAAGCGCAGACCACAGTTACATGTATGGAGAATGGCTGGTCTCCTACTCCAGATGCATCCGTGTCAAAACATGTTCCAAATCAAGTATAGATATTGAGAATGGGTTATTTCTGAATCTCAGTATACATATGCCTTAAAAGAAAAAGCGAAATATCAATGCAAACTAGGATATGTAACAGCAGATGGTGAACATCAGGATCAATTACATGTGGGAAAGATGGATGGTCAGCTCAACCCACGTGCATTAATCTTGTGATATCCAGTATTTATGAATGCCAGAATAAAAATGACTTCACATGGTTAAGCTGAATGACACATTGGACTATGAATGCCATGATGGTTATGAAAGCAATACTGGAAGCACCCTGGTTCCATAGTGTGGTTACAATGGTTGGTCTGATTTACCCATATGTTATGAAAGAGAATGCGAACTTCCTAAAATAGATGTACACTTAGTTCTGATCGAAAGAAAGACCAGTATAAAGTTGGAGAGGTGTGAAATTTCTCTGCAAAACCAGGATTTACAATAGTTGGACCTAATTCGGTTCAGTGCTACACTTTGGATTGTCTCCTGACCTCCCAATATGTAAGAGCAAGTACAATCATGTGGTCCACCTCCTGAACTCCTCAATGGGAATGTTAAGGAAAAACGAAAAGAATATGGACACAGTGAAGTGGTGAATATTATTGCAATCCTAGATTTCTAATGAAGGGACCTAATAAAATTCATGTGTTGATGGAGAGTGGACAATTTACCAGTGTGATTGTGGAGGAGAGTACCTGTGGAGATATACCTGAACTTGAACATGGCTGGGCCAGCTTTCTTCCCTCCTTATTACTATGGAGATTCAGTGAATTCAAATGTCTCAGAATCATTACAATGATTGGACACAGATCAATTACGTGATTTCATGGAGTATGGACCAACTTCCCAGTGTGTGGCAATAGATAAACTTAAGAAGTCAAATCATCAAATTTAATTATACTTGAGGAACATTTAAAAACAAGAAGGAATTCGATCATAATTCTAACATAAGGTACAGATGTAGAGGAAAAGAAGGATGGATACACACAGTCTGCATAAATGGAAGATGGGTCCAGAAGTGAACCTGCTCAATGGCACAAATACAATTATGCCACCTCCACCTCAGATTCCCAATTCTCACAATATGACAACCACCTGAATTATCGGGATGGAGAAAAAGTATCTGTTCTTTGCCAAGAAAATATCTAATTCAGGAAGGAGAAGAAATTACATGCAAAAGATGGAAGATGGCAGTCAATACCACCTCTGTGTTGAAAAAATCCATGTTCAACAACCCTCAGATAGAACCAGGAACCAATTAATTCATCCAGGTCTCACAAGAAAGTTATGCACATGGGACTAAATGAGTTATACTGTGAGGGTGGTTTCAGGATATCTGAAGAAAATGAAACAACATGCTACATGGGAAATGGAGTTCTCCACCTCAGTGTGAAGGCCTTCTTGTAAATCTCCACCTGAGATTTCTCATGGTGTGTAGCTCACATGTCAGACAGTTATCAGTATGGAGAAGAAGTTACGTACAAATGTTTTGAAGGTTTTGGAATTGATGGCCTGCAATTGCAAAAATGCTTAGGAGAAAATGGTCTCACCTCCATCATGCATAAAAACAGATTGTCTCAGTTTACCTAGCTTTGAAAATGCCATACCATGGGAGAGAAGAAGGATGTGTATAAGGCGGGTGAAGCAAGTACTTACACTTGTGCAACAATTACAAAATGGATGGAGCCAGTAATGTAACATGCATTAATAGCAGATGGACAGGAAGGCCAACATGCAGAGACACCTCCTGTGTAATCCGCCACAGTACAAAATGCTTATATAGTGTGAGACAGATGAGTAAATATCCATCTGGTGTGAGAGATACGTTATCAATGTAGGAGCCCTTATGAAATGTTTGGGGATGAAGAAAGTGTGTTTAAATGAAAATGGACCGGAACCCTCAATGCAAAAGATTCTACAGGAAAATGTGGCCCCCTCCACCTATTGACAATGGGGACATTACTTCAATCCCGTTGTCAGTATATGCTCCAGCTTCATCAGTTGAGTACCAATGCCAGAAGTTGATCAACTTGAGGGTAACAAGCGAATAACATGTAGAAATGGACAAATGGTCAAGAACCAAAAATGCTTACATCCGTGTGTAATATCCGAGAAAATATGAAAATTTATAACATAGCATTAAAGTGGACAGCAAAACAGAAGCTTTATTCGAGAACAGGTGAATCAGTTGAATTTGTGTGTAACCGGGATATCGTCTTTCATCAGTTCTCACACATGCGAACAACATGTTGGGATGGGAACTGGAGTATCCAATTTGTGCAAAAAGATAG

[0398] 적합하게는, 변이체는 SEQ ID NO: 52와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.

[0399] CFH 단편은 스플라이스 변이체일 수 있다. 예를 들어, 보체 인자 H-유사 단백질 1(FHL-1)은 CFH 유전자 스플라이스 변이체이며, 이는 CFH의 N-말단 7 도메인(CCP 1-7)과 거의 동일하다.

[0400] 일부 구체예에서, SEQ ID NO: 47에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된 프로모터는 CFH를 인코딩하는 단백질-코딩 서열 또는 이의 단편 및/또는 변이체에 작동가능하게 연결된다.

[0401] 일부 구체예에서, SEQ ID NO: 47의 뉴클레오티드 서열로 구성된 프로모터는 CFH를 인코딩하는 단백질-코딩 서열 또는 이의 단편 및/또는 변이체에 작동가능하게 연결된다.

[0402] FHL-1

- [0403] 단백질-코딩 서열은 FHL-1, 또는 이의 단편 및/또는 변이체를 인코딩할 수 있다.
- [0404] FHL-1 또는 이의 단편 및/또는 변이체는 C3b 및/또는 C3d에 결합할 수 있다. FHL-1의 단편 및/또는 변이체는 천연 FHL-1의 활성의 적어도 50%, 60%, 70%, 80%, 90%, 95% 또는 100%를 유지할 수 있다. FHL-1 및 천연 FHL-1의 단편 및/또는 변이체의 활성은 당업자에게 공지된 임의의 적합한 방법을 사용하여 결정될 수 있다.
- [0405] 바람직하게는, FHL-1은 인간 FHL-1이다. 예시적인 인간 FHL-1은 NCBI 참조 서열: NP\_001014975.1을 갖는 FHL-1이다.
- [0406] 적합하게는, FHL-1은 SEQ ID NO: 53으로 제시된 폴리펩티드 서열 또는 SEQ ID NO: 53과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다.
- [0407] 예시적인 FHL-1 폴리펩티드 서열(SEQ ID NO: 53):
- [0408] MRLLAKI ICLMLWAI CVAEDCNELPPRRNTE ILTGSWSWDTYPEGTQAIYKCRPGYRSLGNVIMVCRKGEWVALNPLRKCQKRPCGHPGDTFFGTFTLTGGN VFEYGVKAVYTCNEGYQLLGE INYRECDTDGWTNDIPICEVVKCLPVTAPENKIVSSAMEPDREYHFGQAVRFVCSNGYKIEGDEEMHCSDDGFWFSKEKPK CVEISCKSPDIVNGSPI SQKI IYKENERFQYKCNMGYEYSERGDAVCTESGWRPLPSCEEKSCDNPIYIPNGDYSPLRIKHRTGDEITYQCRNGFYPATRGNT AKCTSTGWIPAPRCTLKPCDYPD IKHGGLYHENMRRPYFPVAVGKYYSYCDHFETPSGSYWDHIHCTQDQGWSPA VPLCRKCYFPYLENGYNQNHGRKFVQ GKSIDVACHPGYALPKAQT TVTCMENGWSPTPRCIRVSFTL
- [0409] 적합하게는, 변이체는 SEQ ID NO: 53과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0410] FHL-1을 인코딩하는 뉴클레오티드 서열의 예는 NM\_001014975.2이다. 적합하게는, FHL-1을 인코딩하는 단백질-코딩 서열은 SEQ ID NO: 54로 제시된 폴리뉴클레오티드 서열 또는 SEQ ID NO: 54와 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다.
- [0411] 예시적인 FHL-1 폴리뉴클레오티드 서열(SEQ ID NO: 54):
- [0412] ATGAGACTTCTAGCAAAGATTATTTGCTTATGTTATGGGCTATTTGTGTAGCAGAAGATTGCAATGAACTTCTCCAAGAAGAAATACAGAAATCTGACA GGTTCCTGGTCTGACCAAACATATCCAGAAGGCACCCAGGCTATCTATAAATGCCGCCCTGGATATAGATCTCTTGAAATGTAATAATGGTATGCAGGAAG GGAGAATGGGTGCTCTTAATCCATTAAGGAAATGTCAGAAAAGGCCCTGTGGACATCTCGGAGATACTCCTTTTGGTACTTTTACCCTTACAGGAGGAAAT GTGTTTGAATATGGTGTAAAAGCTGTGTATACATGTAATGAGGGTATCAATTGCTAGGTGAGATTAATTACCGTGAATGTGACACAGATGGATGGACCAAT GATATTCCTATATGTGAAGTTGTAAGTGTTTACCAGTGACAGCACCAGAGAATGAAAAAATGTCAGTAGTGAATGGAACAGATCGGGAATACCATTTT GGACAAGCAGTACGGTTTGTATGTAACCTCAGGCTACAAGATTGAAGGAGATGAAGAAATGCATTGTTTCAGACGATGGTTTTGGAGTAAAGAGAAACCAAAG TGTGTGGAATTTTATGCAAATCCCCAGATGTTATAAATGGATCTCCTATATCTCAGAAGATTATTTATAAGGAGAATGAACGATTTCATATAAATGTAAC ATGGGTTATGAATACAGTGAAGAGGAGATGCTGTATGCACTGAATCTGGATGGCGTCCGTTGCCTTCATGTGAAGAAAAATCATGTGATAATCCTTATATT CCAAATGGTGACTACTCACCTTTAAGGATTAACACAGAAGTGGAGATGAAATCACGTACCAGTGTAGAAATGGTTTTTATCTGCAACCCGGGGAATACA GCAAATGCACAAGTACTGGCTGGATACCTGCTCCGAGATGTACCTTGAACCTTGTGATTATCCAGACATTAACATGGAGGTCTATATCATGAGAATATG CGTAGACCATACTTTCCAGTAGCTGTAGGAAAATATTACTCTATTACTGTGATGAACATTTTGAGACTCCGTCAGGAAGTTACTGGGATCACATTCATTGC ACACAAGATGGATGGTTCGCCAGCAGTACCATGCCTCAGAAAATGTTATTTTCTTATTTGGAAAATGGATATAATCAAATCATGGAAGAAAGTTGTACAG GGTAAATCTATAGACGTTGCCTGCCATCCTGGCTACGCTCTTCCAAAAGCGCAGACCACAGTTACATGTATGGAGAATGGCTGGTCTCCTACTCCAGATGC ATCCGTGTCAGCTTTACCCTCTGA
- [0413] 적합하게는, 변이체는 SEQ ID NO: 54와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0414] 일부 구체예에서, SEQ ID NO: 47에 대해 적어도 70%, 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 또는 100% 동일성을 갖는 뉴클레오티드 서열을 포함 하거나 이로 구성된 프로모터는 FHL-1을 인코딩하는 단백질-코딩 서열 또는 이의 단편 및/또는 변이체에 작동가능하게 연결된다.
- [0415] 일부 구체예에서, SEQ ID NO: 47의 뉴클레오티드 서열로 구성된 프로모터는 FHL-1을 인코딩하는 단백질-코딩 서열, 또는 이의 단편 및/또는 변이체에 작동가능하게 연결된다.
- [0416] 기타 조절 요소
- [0417] 본 발명의 프로모터 이외에, 폴리뉴클레오티드는 전사-전 또는 후에 작용할 수 있는 하나 이상의 추가 조절 서열을 포함할 수 있다. 적합하게는, 단백질-코딩 서열은 하나 이상의 추가 조절 서열에 작동가능하게 연결될 수

있다. 하나 이상의 추가 조절 서열은 사구체 세포(예를 들어, 발세포)에서 단백질의 발현을 촉진할 수 있다.

[0418] "조절 서열"은 폴리펩티드의 발현을 조장하게 하는, 예를 들어, 전사체의 발현을 증가시키거나 mRNA 안정성을 향상시키는 작용을 하는 임의의 서열이다. 적합한 추가 조절 서열은 예를 들어, 인핸서 요소, 전사-후 조절 요소 및 폴리아데닐화 부위를 포함한다.

[0419] 적합하게는, 본 발명의 폴리뉴클레오티드는 임의의 추가 조절 서열(본 발명의 프로모터 제외)을 포함하지 않는다.

[0420] 인핸서

[0421] 본 발명의 폴리뉴클레오티드는 인핸서를 포함할 수 있다. 적합하게는, 인핸서는 단백질-코딩 서열에 작동가능하게 연결될 수 있다. 인핸서는 사구체 세포(예를 들어, 발세포)에서 단백질의 발현을 조장할 수 있다. 적합하게는, 인핸서는 포유동물 인핸서, 예를 들어, 인간 인핸서이다.

[0422] "인핸서"는 특정 유전자의 전사가 일어날 가능성을 증가시키기 위해 단백질(활성화제)에 의해 결합될 수 있는 DNA의 영역이다. 인핸서는 시스-작용한다. 이들은 유전자로부터 최대 1 Mbp(1,000,000 bp) 떨어져, 출발 부위로부터 업스트림 또는 다운스트림에 위치할 수 있다. 임의의 적합한 인핸서가 사용될 수 있으며, 이들의 선택은 당업자에 의해 용이하게 이루어질 수 있다.

[0423] 본 발명의 폴리뉴클레오티드는 조직 특이적 인핸서를 포함할 수 있다. 본원에서 사용되는 "조직-특이적 인핸서"는 특정 세포 또는 조직에서 유전자의 발현을 우선적으로 촉진하는 인핸서이다. 적합하게는, 조직-특이적 인핸서는 다른 세포-유형과 비교하여 특정 세포-유형에서 유전자의 더 높은 발현을 촉진할 수 있다. 더 높은 발현은 예를 들어, 인핸서에 작동가능하게 연결된 전이유전자, 예를 들어, GFP의 발현을 측정함으로써 측정될 수 있으며, 여기서 전이유전자의 발현은 유전자의 발현을 용이하게 하는 인핸서의 능력과 상관관계가 있다. 예를 들어, 조직-특이적 인핸서는 다른 세포-유형에서의 발현 수준과 비교하여 특정 세포-유형에서 적어도 10% 더 높거나, 적어도 20% 더 높거나, 적어도 30% 더 높거나, 적어도 40% 더 높거나, 적어도 50% 더 높거나, 적어도 100% 더 높거나, 적어도 200% 더 높거나, 적어도 300% 더 높거나, 적어도 400% 더 높거나, 적어도 500% 더 높거나, 적어도 1000% 더 높은 유전자 발현 수준을 조장하는 인핸서일 수 있다.

[0424] 적합한 조직-특이적 인핸서는 당업자에게 잘 알려져 있을 것이다. 인핸서는 신장-특이적 인핸서, 바람직하게는 사구체-특이적 인핸서, 더욱 바람직하게는 발세포-특이적 인핸서일 수 있다. 적합하게는, 인핸서는 단백질-코딩 서열에 작동가능하게 연결될 수 있다.

[0425] 적합하게는, 인핸서는 인간 신장, 사구체 세포 및/또는 발세포에서 선택적 발현을 갖는 유전자와 관련된 인핸서이거나 이로부터 유래될 수 있다. 유전자와 관련된 인핸서 영역을 확인하는 방법은 당업자에게 잘 알려져 있을 것이다.

[0426] 적합하게는, 발세포-특이적 인핸서는 NPHS1 또는 NPHS2 인핸서, 또는 이의 단편 또는 유도체이다. 적합하게는, 발세포-특이적 인핸서는 NPHS1 인핸서, 또는 이의 단편 또는 유도체이다.

[0427] NPHS1 인핸서는 문헌[Guo, G., et al., 2004. Journal of the American Society of Nephrology, 15(11), pp.2851-2856]에 기재되어 있다. 인간 NPHS1 프로모터로부터의 186-bp 단편은 유전자전이 마우스에서 이종성 최소 프로모터 앞에 배치될 때  $\beta$ -갈락토시다제 전이유전자의 발세포-특이적 발현을 지시할 수 있었다.

[0428] 적합하게는, NPHS1 인핸서는 SEQ ID NO: 27로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 27과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다.

[0429] 예시적인 NPHS1 인핸서(SEQ ID NO: 27):

[0430] ctgctgagctgggagaccaccttgatctgacttctcccatctcccagcctaagccaggccctggggtcacggaggctggggaggcaccgaggaacgcgcctggcatgtgctgacaggggattttatgctccagctgggccagctgggaggagcctgctgggcagagccagagctgggggctctg

[0431] 적합하게는, 변이체는 SEQ ID NO: 27과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일할 수 있다.

[0432] 적합하게는, 본 발명의 폴리뉴클레오티드는 인핸서를 포함하지 않는다.

[0433] Kozak 서열

[0434] 본 발명의 폴리뉴클레오티드는 Kozak 서열을 포함할 수 있다. 적합하게는, Kozak 서열은 단백질 코딩 서열에 작

동가능하게 연결될 수 있다. Kozak 서열은 번역의 개시를 개선하기 위해 단백질의 시작 코돈 전에 삽입될 수 있다.

- [0435] 적합한 Kozak 서열은 당업자에게 잘 알려져 있을 것이다.
- [0436] 적합하게는, Kozak 서열은 SEQ ID NO: 28로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 28과 적어도 65% 동일한 변이체를 포함하거나 이로 구성될 수 있다.
- [0437] *예시적인 Kozak 서열(SEQ ID NO: 28):*
- [0438] GCCGCCACCAUGG
- [0439] 적합하게는, 변이체는 SEQ ID NO: 28과 적어도 75%, 적어도 85% 또는 적어도 90% 동일할 수 있다.
- [0440] 적합하게는, 본 발명의 폴리뉴클레오티드는 Kozak 서열을 포함하지 않는다.
- [0441] *전사후 조절 요소*
- [0442] 본 발명의 폴리뉴클레오티드는 전사-후 조절 요소를 포함할 수 있다. 적합하게는, 전사-후 조절 요소는 단백질-코딩 서열에 작동가능하게 연결될 수 있다. 전사후 조절 요소는 유전자 발현을 개선할 수 있다.
- [0443] 폴리뉴클레오티드는 우드척 간염 바이러스 전사후 조절 요소(WPRE)를 포함할 수 있다. 적합하게는, WPRE는 단백질-코딩 서열에 작동가능하게 연결될 수 있다.
- [0444] WPRE 서열은 X-항원 프로모터 및/또는 X-항원의 개시 코돈 내에 치환, 결실 또는 삽입을 가질 수 있다. 이는 기 능성 X-항원의 생성을 방지할 수 있다.
- [0445] 적합하게는, WPRE는 SEQ ID NO: 29로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 29와 적어도 70% 동일한 변 이체를 포함하거나 이로 구성될 수 있다.
- [0446] *예시적인 WPRE(SEQ ID NO: 29):*
- [0447] aatcaacctctggattacaaaatttgtgaaagattgactggtatcttaactatggtgctccttttagcctatgtggatagcctgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttgtataaatcctggtgctgctctttatgaggagtgtggccctgtgtaggcaactggcgtggtgtgcaactgtttgctgacgcaacccccactggtggggcattgccaccactgtcagctcctttccgggactttcgctttccccctccatttgcacggcggaactcatcgccgctgcttggccgctgctggacaggggctcggctgtgggcaactgacaattccgtggtgtgtcggggaaatcatcgtcctttccttggctgctcgctgtgttggcaactggattctgcgcgggacgtccttctgctacgtcccttcggccctcaatccagcggaccttcttcccgcggcctgctgccggctctcgccgctctccgcgtcttcgccttcgcctcagacgagtcggatctcctttggggccgctccccgc
- [0448] 적합하게는, 변이체는 SEQ ID NO: 29와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0449] 적합하게는, 본 발명의 폴리뉴클레오티드는 전사 후 조절 요소를 포함할 수 없다.
- [0450] *폴리아데닐화 신호*
- [0451] 본 발명의 폴리뉴클레오티드는 폴리아데닐화 신호를 포함할 수 있다. 적합하게는, 폴리아데닐화 신호는 단백질-코딩 서열에 작동가능하게 연결될 수 있다. 폴리아데닐화 신호는 유전자 발현을 개선시킬 수 있다.
- [0452] 적합한 폴리아데닐화 신호는 초기 SV40 폴리아데닐화 신호(SV40pA), 닭 베타-글로빈 폴리아데닐화 신호, 소 성장 호르몬 폴리아데닐화 신호(bGH) 또는 가용성 뉴로필린-1 폴리아데닐화 신호를 포함한다. 일부 구체예에서, 폴리아데닐화 신호는 초기 SV40 폴리아데닐화 신호(SV40pA) 또는 닭 베타-글로빈 폴리아데닐화 신호이다. 바람 직하게는, 폴리아데닐화 신호는 초기 SV40 폴리아데닐화 신호(SV40pA)이다.
- [0453] 적합하게는, 폴리아데닐화 신호는 SEQ ID NO: 30으로 제시된 뉴클레오티드 서열, 또는 SEQ ID NO: 30과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 30과 적어도 75%, 적 어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0454] *예시적인 SV40pA 신호 서열(SEQ ID NO: 30):*
- [0455] aacttgtttatgtagcttataatggttacaataaagcaatagcatcacaatattcacaataaagcattttttactgcatctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatc

- [0456] 적합하게는, 폴리아데닐화 신호는 SEQ ID NO: 31로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 31과 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 31과 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0457] 예시적인 bGH 폴리(A) 신호 서열(SEQ ID NO: 31):
- [0458] ctgtgccttctagttgccagccatctgttgtttgccctccccctgccttctctgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcatctgtctgagtaggtgtcattctattctgggggtgggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctgggatgcggtgggctctatgg
- [0459] 적합하게는, 폴리아데닐화 신호는 SEQ ID NO: 32로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 32와 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 32와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0460] 예시적인 가용성 뉴로필린-1 폴리아데닐화 신호(SEQ ID NO: 32):
- [0461] aaataaaatcgaaatg
- [0462] 적합하게는, 폴리아데닐화 신호는 SEQ ID NO: 42로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 42와 적어도 70% 동일한 변이체를 포함하거나 이로 구성될 수 있다. 적합하게는, 변이체는 SEQ ID NO: 42와 적어도 75%, 적어도 80%, 적어도 85%, 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98%, 적어도 99% 동일할 수 있다.
- [0463] 예시적인 닭 베타-글로빈 폴리아데닐화 신호(SEQ ID NO: 42)
- [0464] caataaaagatctttattttcattagatctgtgtgtgtgtttttgtgtg
- [0465] **백터**
- [0466] 본 발명은 본 발명의 폴리뉴클레오티드를 포함하는 백터를 제공한다.
- [0467] 백터는 하나의 환경에서 다른 환경으로의 실체의 이동을 허용하거나 용이하게 하는 도구이다. 4가지 주요 유형의 백터는 플라스미드, 바이러스 백터, 코스미드 및 인공 염색체이다.
- [0468] 바람직하게는, 본 발명의 백터는 바이러스 백터이다. 본 발명의 백터는 바람직하게는 아데노-관련 바이러스(AAV) 백터이지만, 다른 바이러스 백터가 사용될 수 있는 것으로 고려된다.
- [0469] 본 발명의 백터는 바이러스 백터 입자의 형태일 수 있다. 바람직하게는, 본 발명의 바이러스 백터는 AAV 백터 입자의 형태이다.
- [0470] AAV로부터 유래된 것과 같은 바이러스 백터 및 바이러스 백터 입자를 제조하고 변형시키는 방법은 당 분야에 널리 공지되어 있다. 적합한 방법은 문헌 [Ayuso, E., et al., 2010. Current gene therapy, 10(6), pp.423-436, Merten, O.W., et al., 2016. Molecular Therapy-Methods & Clinical Development, 3, p.16017; and Nadeau, I. and Kamen, A., 2003. Biotechnology advances, 20(7-8), pp.475-489]에 기재되어 있다.
- [0471] 본 발명의 백터는 신장 세포를 형질도입할 수 있다. 일부 구체예에서, 본 발명의 백터는 신장 세포를 특이적으로 형질도입할 수 있다.
- [0472] 본 발명의 백터는 바람직하게는 사구체 세포(예를 들어, 발세포)를 형질도입할 수 있다. 일부 구체예에서, 본 발명의 백터는 사구체 세포(예를 들어, 발세포)를 특이적으로 형질도입할 수 있다. 본 발명의 백터는 바람직하게는 발세포를 형질도입할 수 있다. 일부 구체예에서, 본 발명의 백터는 발세포를 특이적으로 형질도입할 수 있다.
- [0473] 아데노-관련 바이러스(AAV) 백터
- [0474] 본 발명의 백터는 아데노-관련 바이러스(AAV) 백터일 수 있다. 본 발명의 백터는 AAV 백터 입자의 형태일 수 있다.
- [0475] AAV 계놈
- [0476] AAV 백터 또는 AAV 백터 입자는 AAV 계놈 또는 이의 단편 또는 유도체를 포함할 수 있다.
- [0477] AAV 계놈은 AAV 입자의 생성에 필요한 기능을 인코딩할 수 있는 폴리뉴클레오티드 서열이다. 이러한 기능은 AAV 입자로의 AAV 계놈의 캡시드화를 포함하여, 숙주 세포에서 AAV의 복제 및 패키징 사이클에서 작동하는 것을 포

함한다. 자연 발생 AAV는 복제-결핍이며, 복제 및 패키징 사이클의 완료를 위해 트랜스로 헬퍼 기능의 제공에 의존한다. 따라서, 본 발명의 AAV 벡터의 AAV 게놈은 전형적으로 복제-결핍이다.

- [0478] AAV 게놈은 양성 또는 음성-센스의 단일-가닥 형태(ssAAV) 또는 대안적으로 이중-가닥 형태(dsAAV)일 수 있다. 이중 가닥 형태의 사용은 표적 세포에서 DNA 복제 단계의 우회를 가능하게 하여 전이유전자 발현을 가속화할 수 있다. 단일 가닥 형태의 최대 패키징 용량은 이중 가닥 형태보다 크다. 적합하게는, AAV 게놈은 단일-가닥 형태이다.
- [0479] 자연에서 발생하는 AAV는 다양한 생물학적 시스템에 따라 분류될 수 있다. AAV 게놈은 AAV의 임의의 천연 유래된 혈청형, 분리물 또는 클레이드로부터 비롯될 수 있다.
- [0480] AAV는 이들의 혈청형과 관련하여 언급될 수 있다. 혈청형은 캡시드 표면 항원의 발현 프로파일로 인해 이를 다른 변이체 아종과 구별하는데 사용될 수 있는 독특한 반응성을 갖는 AAV의 변이체 아종에 상응한다. 전형적으로, 특정 AAV 혈청형을 갖는 AAV 벡터 입자는 임의의 다른 AAV 혈청형에 특이적인 중화 항체와 효율적으로 교차-반응하지 않는다. AAV 혈청형은 AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10 및 AAV11을 포함한다. 일부 구체예에서, 본 발명의 AAV 벡터는 AAV3B, LK03, AAV9, 또는 AAV8 혈청형일 수 있다. 일부 구체예에서, 본 발명의 AAV 벡터는 AAV3B, LK03 또는 AAV9 혈청형일 수 있다.
- [0481] AAV는 또한 클레이드 또는 클론과 관련하여 언급될 수 있다. 이는 자연 유래된 AAV의 계통발생적 관계, 및 전형적으로 공통 조상으로 역추적될 수 있는 AAV의 계통발생적 그룹을 지칭하고, 이의 모든 후손을 포함한다. 또한, AAV는 특정 분리물, 즉, 자연에서 발견되는 특정 AAV의 유전적 분리물과 관련하여 언급될 수 있다. 용어 유전적 분리물은 다른 자연 발생 AAV와 제한된 유전적 혼합을 거친 AAV의 집단을 기술함으로써, 유전자 수준에서 인식 가능하게 별개의 집단을 정의한다.
- [0482] 전형적으로, AAV의 자연 유래된 혈청형, 분리물 또는 클레이드의 AAV 게놈은 적어도 하나의 역 말단 반복 서열(ITR)을 포함한다. ITR 서열은 시스로 작용하여 기능적 복제 기점을 제공하고 세포의 게놈으로부터 벡터의 통합 및 절제를 가능하게 한다. ITR은 시스에서 치료 유전자 옆에 필요한 유일한 서열일 수 있다.
- [0483] AAV 게놈은 또한 AAV 입자에 대한 패키징 기능을 인코딩하는 rep 및/또는 cap 유전자와 같은 패키징 유전자를 포함할 수 있다. 프로모터는 각각의 패키징 유전자에 작동가능하게 연결될 수 있다. 이러한 프로모터의 특정 예는 p5, p19 및 p40 프로모터를 포함한다. 예를 들어, p5 및 p19 프로모터는 일반적으로 rep 유전자를 발현하는데 사용되는 반면, p40 프로모터는 일반적으로 cap 유전자를 발현하는데 사용된다. rep 유전자는 단백질 Rep78, Rep68, Rep52 및 Rep40 중 하나 이상 또는 이의 변이체를 인코딩한다. cap 유전자는 하나 이상의 캡시드 단백질, 예컨대, VP1, VP2 및 VP3 또는 이의 변이체를 인코딩한다. 이러한 단백질은 AAV 혈청형을 결정하는 AAV 입자의 캡시드를 구성한다. VP1, VP2 및 VP3은 교대 mRNA 스플라이싱에 의해 생성될 수 있다(Trempe, J.P. and Carter, B.J., 1988. Journal of virology, 62(9), pp.3356-3363). 따라서, VP1, VP2 및 VP3은 동일한 서열을 가질 수 있지만, VP2는 VP1에 비해 N-말단에서 절두되고, VP3은 VP2에 비해 N-말단에서 절두된다.
- [0484] AAV 게놈은 자연 발생 AAV의 전체 게놈일 수 있다. 예를 들어, 전체 AAV 게놈을 포함하는 벡터는 AAV 벡터 또는 벡터 입자를 제조하는데 사용될 수 있다.
- [0485] 바람직하게는, AAV 게놈은 환자에게 투여할 목적으로 유도체화된다. 이러한 유도체화는 당 분야의 표준이며, 본 발명은 AAV 게놈의 임의의 공지된 유도체, 및 당 분야에 공지된 기술을 적용함으로써 생성될 수 있는 유도체의 사용을 포함한다. AAV 게놈은 임의의 자연 발생 AAV의 유도체일 수 있다. 적합하게는, AAV 게놈은 AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10 또는 AAV11의 유도체이다. 적합하게는, AAV 게놈은 AAV2의 유도체이다.
- [0486] AAV 게놈의 유도체는 생체내에서 본 발명의 AAV 벡터로부터 전이유전자의 발현을 가능하게 하는 임의의 절두된 또는 변형된 형태의 AAV 게놈을 포함한다. 전형적으로, 최소 바이러스 서열을 포함하면서도 상기 기능을 유지하기 위해 AAV 게놈을 유의하게 절두하는 것이 가능하다. 이는 벡터와 야생형 바이러스의 재조합 위험을 감소시키고 또한 표적 세포에서 바이러스 유전자 단백질의 존재에 의해 세포 면역 반응을 촉발시키는 것을 피하기 위한 안전상의 이유로 바람직하다.
- [0487] 전형적으로, 유도체는 적어도 하나의 역 말단 반복 서열(ITR), 바람직하게는 하나 초과 ITR, 예를 들어, 2개 이상의 ITR을 포함할 것이다. 하나 이상의 ITR은 상이한 혈청형을 갖는 AAV 게놈으로부터 유래될 수 있거나, 키메라 또는 돌연변이체 ITR일 수 있다. 바람직한 돌연변이체 ITR은 trs(말단 분해 부위)가 결실된 것이다. 이러한 결실은 코딩 및 상보적 서열 둘 모두를 함유하는 단일-가닥 게놈, 즉, 자가-상보적 AAV(scAAV) 게놈을 생성

하기 위해 게놈의 지속적인 복제를 가능하게 한다. 이는 표적 세포에서 DNA 복제의 우회를 가능하게 하여, 가속화된 전이유전자 발현을 가능하게한다. 그러나, scAAV의 최대 패키징 용량은 감소된다. 적합하게는, AAV 게놈은 scAAV 게놈이 아니다.

- [0488] AAV 게놈은 AAV 또는 이의 변이체의 임의의 자연 유래된 혈청형, 분리물 또는 클레이드로부터의 하나 이상의 ITR 서열을 포함할 수 있다. AAV 게놈은 적어도 하나, 예컨대, 2개의 AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10 또는 AAV11 ITR 또는 이들의 변이체를 포함할 수 있다. 적합하게는, AAV 게놈은 적어도 하나, 예를 들어, 2개의 AAV2 ITR을 포함할 수 있다.
- [0489] 하나 이상의 ITR의 포함은 예를 들어, 숙주 세포 DNA 폴리머라제의 작용에 의해 단일-가닥 벡터 DNA의 이중-가닥 DNA로의 전환 후, 숙주 세포의 핵에서 AAV 벡터의 연쇄동일서열(concatamer) 형성을 돕는데 바람직하다. 이러한 에피솜 연쇄동일서열의 형성은 숙주 세포의 수명 동안 AAV 벡터를 보호하여, 생체내에서 전이유전자의 연장된 발현을 가능하게한다.
- [0490] 적합하게는, ITR 요소는 유도체에서 자연 AAV 게놈으로부터 보유된 유일한 서열일 것이다. 유도체는 바람직하게는 천연 게놈의 rep 및/또는 cap 유전자 및 천연 게놈의 임의의 다른 서열을 포함하지 않을 것이다. 이는 상기 기재된 이유로 및 또한 숙주 세포 게놈으로의 벡터의 통합 가능성을 감소시키는데 바람직하다. 또한, AAV 게놈의 크기를 감소시키는 것은 전이유전자에 추가하여 벡터 내에 다른 서열 요소(예를 들어, 조절 요소)를 혼입하는데 있어서 증가된 유연성을 허용한다.
- [0491] 따라서, 하기 부분이 본 발명의 유도체에서 제거될 수 있다: 하나의 역 말단 반복부(ITR) 서열, 복제(rep) 및 캡시드(cap) 유전자. 그러나, 유도체는 AAV 게놈의 하나 이상의 rep 및/또는 cap 유전자 또는 다른 바이러스 서열을 추가로 포함할 수 있다. 자연 발생 AAV는 인간 염색체 19 상의 특정 부위에서 높은 빈도로 통합되고, AAV 벡터에서 통합 능력의 보유가 치료 환경에서 허용될 수 있도록 무시할 만한 무작위 통합 빈도를 나타낸다.
- [0492] 본 발명은 추가로 천연 AAV 게놈의 것과 상이한 순서 및 구성의 AAV 게놈의 서열의 제공을 포함한다. 본 발명은 또한 하나 이상의 AAV 서열 또는 유전자를 또 다른 바이러스로부터의 서열로 또는 하나 초과의 바이러스로부터의 서열로 구성된 키메라 유전자로 대체하는 것을 포함한다. 이러한 키메라 유전자는 상이한 바이러스 종의 2개 이상의 관련된 바이러스 단백질로부터의 서열로 구성될 수 있다.
- [0493] *AAV 혈청형 및 캡시드 단백질*
- [0494] AAV 벡터 입자는 캡시드 단백질에 의해 캡시드화될 수 있다. 혈청형은 사구체 세포(예를 들어, 발세포)의 형질 도입, 예를 들어, 사구체 세포(예를 들어, 발세포)의 특이적 형질도입을 촉진할 수 있다.
- [0495] AAV 벡터 입자는 신장-특이적 벡터 입자일 수 있다. 바람직하게는, AAV 벡터 입자는 사구체-특이적(예를 들어, 발세포-특이적) 벡터 입자이다. AAV 벡터 입자는 사구체-특이적(예를 들어, 발세포-특이적) 캡시드에 의해 캡시드화될 수 있다. AAV 벡터 입자는 사구체-특이적(예를 들어, 발세포-특이적) 캡시드 단백질을 포함할 수 있다.
- [0496] 적합하게는, AAV 벡터 입자는 트랜스캡시드화된 형태일 수 있고, 여기서 하나의 혈청형의 ITR을 갖는 AAV 게놈 또는 유도체는 상이한 혈청형의 캡시드에 패키징된다. AAV 벡터 입자는 또한 2개 이상의 상이한 혈청형으로부터의 비변형된 캡시드 단백질의 혼합물이 바이러스 캡시드를 구성하는 모자이크 형태를 포함한다. AAV 벡터 입자는 또한 캡시드 표면에 흡착된 리간드를 보유하는 화학적으로 변형된 형태를 포함한다. 예를 들어, 이러한 리간드는 특정 세포 표면 수용체를 표적화하기 위한 항체를 포함할 수 있다.
- [0497] 유도체가 캡시드 단백질, 즉, VP1, VP2 및/또는 VP3을 포함하는 경우, 유도체는 하나 이상의 자연 발생 AAV의 키메라, 서플된 또는 캡시드-변형된 유도체일 수 있다. 특히, 본 발명은 동일한 벡터(즉, 슈도타입화된 벡터) 내의 AAV의 상이한 혈청형, 클레이드, 클론 또는 분리물로부터의 캡시드 단백질 서열의 제공을 포함한다. AAV 벡터는 슈도타입화된 AAV 벡터 입자의 형태일 수 있다.
- [0498] 키메라, 서플된 또는 캡시드-변형된 유도체는 전형적으로 AAV 벡터에 대한 하나 이상의 요망되는 작용기를 제공하도록 선택될 것이다. 따라서, 이러한 유도체는 자연 발생 AAV 게놈을 포함하는 AAV 벡터와 비교하여 증가된 유전자 전달 효율, 감소된 면역원성(체액 또는 세포), 변경된 지향성 범위 및/또는 발세포의 개선된 표적화를 나타낼 수 있다. 유전자 전달의 증가된 효율은 세포 표면에서 개선된 수용체 또는 공동-수용체 결합, 개선된 내재화, 세포 내 및 핵으로의 개선된 트래피킹, 바이러스 입자의 개선된 언코딩 및 단일-가닥 게놈의 이중 가닥 형태로의 개선된 전환에 의해 달성될 수 있다. 증가된 효율은 또한 벡터 용량이 필요하지 않은 조직으로의 투여에 의해 희석되지 않도록 변경된 지향성 범위 또는 발세포의 표적화와 관련될 수 있다.

- [0499] 키메라 캡시드 단백질은 자연 발생 AAV 혈청형의 2개 이상의 캡시드 코딩 서열 사이의 재조합에 의해 생성된 것들을 포함한다. 이는 예를 들어, 한 혈청형의 비-감염성 캡시드 서열이 상이한 혈청형의 캡시드 서열로 공동-형질감염되고, 지시된 선택이 요망되는 특성을 갖는 캡시드 서열을 선택하기 위해 사용되는 마커 구조 접근법에 의해 수행될 수 있다. 상이한 혈청형의 캡시드 서열은 세포 내에서 상동성 재조합에 의해 변경되어 신규한 키메라 캡시드 단백질을 생성할 수 있다.
- [0500] 키메라 캡시드 단백질은 또한 2개 이상의 캡시드 단백질 사이, 예를 들어, 상이한 혈청형의 2개 이상의 캡시드 단백질 사이에 특정 캡시드 단백질 도메인, 표면 루프 또는 특정 아미노산 잔기를 전달하기 위해 캡시드 단백질 서열의 조작에 의해 생성된 것들을 포함한다.
- [0501] 서플링된 또는 키메라 캡시드 단백질은 또한 DNA 서플링 또는 오류-유발 PCR에 의해 생성될 수 있다. 하이브리드 AAV 캡시드 유전자는 관련된 AAV 유전자의 서열, 예를 들어, 다수의 상이한 혈청형의 캡시드 단백질을 인코딩하는 서열을 무작위로 단편화한 다음, 후속하여 서열 상동성 영역에서 크로스오버를 또한 유발할 수 있는 자가-프라이밍 폴리머라제 반응에서 단편을 재조합함으로써 생성될 수 있다. 여러 혈청형의 캡시드 유전자를 서플링함으로써 이러한 방식으로 생성된 하이브리드 AAV 유전자의 라이브러리는 요망되는 기능을 갖는 바이러스 클론을 확인하기 위해 스크리닝될 수 있다. 유사하게, 오류가 발생하기 쉬운 PCR은 AAV 캡시드 유전자를 무작위로 돌연변이시켜 원하는 특성에 대해 선택될 수 있는 다양한 변이체 라이브러리를 생성하는데 사용될 수 있다.
- [0502] 캡시드 유전자의 서열은 또한 천연 야생형 서열과 관련하여 특정 결실, 치환 또는 삽입을 도입하도록 유전적으로 변형될 수 있다. 특히, 캡시드 유전자는 캡시드 코딩 서열의 오픈 리딩 프레임 내에, 또는 캡시드 코딩 서열의 N- 및/또는 C-말단에서 관련되지 않은 단백질 또는 펩티드의 서열의 삽입에 의해 변형될 수 있다. 관련되지 않은 단백질 또는 펩티드는 유리하게는 특정 세포 유형에 대한 리간드로서 작용함으로써 표적 세포에 대한 개선된 결합을 부여하거나 특정 세포 집단에 대한 벡터의 표적화 특이성을 개선시키는 것일 수 있다. 관련되지 않은 단백질은 또한 생성 공정의 일부로서 바이러스 입자의 정제를 돕는 것, 즉, 에피토프 또는 친화성 태그일 수 있다. 삽입 부위는 전형적으로 바이러스 입자의 다른 기능, 예를 들어, 바이러스 입자의 내재화, 트래피킹을 방해하지 않도록 선택될 것이다.
- [0503] 캡시드 단백질은 인공 또는 돌연변이체 캡시드 단백질일 수 있다. 본원에서 사용되는 용어 "인공 캡시드"는 캡시드 입자가 자연에서 발생하지 않거나 자연 발생 캡시드 아미노산 서열로부터 조작된(예를 들어, 변형된) 아미노산 서열을 포함하는 아미노산 서열을 포함하는 것을 의미한다. 다시 말해서, 인공 캡시드 단백질은 인공 캡시드 아미노산 서열과 모 캡시드 아미노산 서열이 정렬되는 경우, 인공 캡시드 단백질이 유래된 모 캡시드의 서열과 비교하여 아미노산 서열의 돌연변이 또는 변이를 포함한다.
- [0504] 캡시드 단백질은 비변형된 또는 야생형 바이러스 입자에 비해 발세포를 형질도입하는 능력을 개선시키는, 야생형 캡시드 단백질에 대한 돌연변이 또는 변형을 포함할 수 있다. 발세포를 형질도입하는 개선된 능력은 예를 들어, AAV 벡터 입자에 의해 운반되는 전이유전자 예를 들어, GFP의 발현을 측정함으로써 측정될 수 있으며, 여기서 발세포에서의 전이유전자의 발현은 발세포를 형질도입시키는 AAV 벡터 입자의 능력과 상관관계가 있다.
- [0505] AAV 벡터 입자는 AAV3B, LK03, AAV9, 또는 AAV8 벡터 입자일 수 있다. 본 발명자들은 AAV3B, LK03, AAV9 및 AAV8 혈청형을 갖는 AAV 벡터 입자가 발세포를 형질도입할 수 있음을 보여주었다. 바람직하게는, AAV 벡터 입자는 AAV3B 벡터 입자 또는 LK03 벡터 입자이다. 더욱 바람직하게는, AAV 벡터 입자는 AAV3B 벡터 입자이다.
- [0506] AAV 벡터 입자는 AAV3B, LK03, AAV9 또는 AAV8 캡시드 단백질을 포함할 수 있다. 바람직하게는, AAV 벡터 입자는 AAV3B 캡시드 단백질 또는 LK03 캡시드 단백질을 포함한다. 더욱 바람직하게는, AAV 벡터 입자는 AAV3B 캡시드 단백질을 포함한다.
- [0507] AAV 벡터 입자는 AAV3B, LK03, AAV9 또는 AAV8 캡시드 단백질 VP1, VP2 및 VP3을 포함할 수 있다. 바람직하게는, AAV 벡터 입자는 AAV3B 또는 LK03 캡시드 단백질 VP1, VP2 및 VP3을 포함한다. 더욱 바람직하게는, AAV 벡터 입자는 AAV3B 캡시드 단백질 VP1, VP2 및 VP3을 포함한다.
- [0508] AAV 벡터 입자는 하나 이상의 AAV2 ITR 서열 및 AAV3B 캡시드 단백질, LK03 캡시드 단백질, AAV9 캡시드 단백질 또는 AAV8 캡시드 단백질을 포함할 수 있다. 바람직하게는, AAV 벡터 입자는 하나 이상의 AAV2 ITR 서열 및 AAV3B 또는 LK03 캡시드 단백질을 포함한다. 더욱 바람직하게는, AAV 벡터 입자는 하나 이상의 AAV2 ITR 서열 및 AAV3B 캡시드 단백질을 포함한다.
- [0509] AAV 벡터 입자는 AAV2 계통 및 AAV3B 캡시드 단백질(AAV2/3B), AAV2 계통 및 LK03 캡시드 단백질, AAV2 계통

및 AAV9 캡시드 단백질(AAV2/9), 또는 AAV2 계통 및 AAV8 캡시드 단백질(AAV2/8)을 가질 수 있다. 바람직하게는, AAV 벡터 입자는 AAV2 계통 및 AAV3B 또는 LK03 캡시드 단백질을 포함한다. 더욱 바람직하게는, AAV 벡터 입자는 AAV2 계통 및 AAV3B 캡시드 단백질을 포함한다.

[0510] 명명법 AAVX/Y는 예를 들어, ITR 서열이 AAVX로부터 유래되고 혈청형 AAVY(즉, AAVY 캡시드 단백질과 함께)로 캡시드화된 페이로드를 보유하는 카세트에 측접하는 슈도타입화된 AAV를 나타낼 수 있다.

[0511] *AAV3B 혈청형*

[0512] AAV 벡터 입자는 AAV3B 캡시드 단백질을 포함할 수 있다. 적합하게는, AAV 벡터 입자는 AAV3B 캡시드 단백질에 의해 캡시드화될 수 있다.

[0513] 2개의 별개의 AAV3 분리물(AAV3A 및 AAV3B)이 클로닝되었다. 다른 AAV 혈청형에 기반한 벡터와 비교하여, AAV3 벡터는 대부분의 세포 유형을 비효율적으로 형질도입하는 것으로 생각된다. 그러나, AAV3B는 발세포를 효율적으로 형질도입할 수 있다. AA3B는 문헌[Rutledge, E.A., et al., 1998. Journal of virology, 72(1), pp.309-319]에 기술되어 있다.

[0514] AAV 벡터 입자는 AAV3B VP1 캡시드 단백질, AAV3B VP2 캡시드 단백질 및/또는 AAV3B VP3 캡시드 단백질을 포함할 수 있다. 적합하게는, AAV 벡터 입자는 AAV3B VP1 캡시드 단백질, AAV3B VP2 캡시드 단백질 및/또는 AAV3B VP3 캡시드 단백질에 의해 캡시드화될 수 있다. 적합하게는, AAV 벡터 입자는 AAV3B VP1, VP2 및 VP3 캡시드 단백질에 의해 캡시드화될 수 있다.

[0515] 적합하게는, AAV3B VP1 캡시드 단백질은 SEQ ID NO: 33으로 제시된 아미노산 서열 또는 SEQ ID NO: 33과 적어도 90% 동일한 변이체를 포함하거나 이로 구성될 수 있다.

[0516] *예시적인 AAV3B VP1 캡시드 단백질(SEQ ID NO: 33):*

[0517] MAADGYLPDWLEDNLSEGIREWALKPGVQPKANQQHQDNRRGLVLPGYKYLPGGNLTKGEPVNEADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLQEDTSFGGNLGRAVFQAKKRILEPLGLVEEAAKTAPGKKRPVDQSPQEPDSSSGVGKSGKQPARKRLNFGQTGDSESVDPDQPLGEPAAPTSLGSGNTMASGGGAPMADNNEGADGVGNSSGNWHCDSQLWGDRIITSTRTWALPTYNNHLYKQISSQSGASNDNHFYGYSTPWGYFDNRFHCHFSPRDWQRLINNNWGF RPKKLSFKLFNIQVKEVTQNDGTTTIANNLSTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNN FQFSYTFEDVPFHSSYAHSQLDRLMNPLIDQYLYLNRTQGTTSQTNSRLLFSQAGPQMSLQARNWLPGPCYRQRLSKTANDNNNSNFPWTAASKYH LNGRDSLVPNPGPAMASHKDEEKFFPMHGNLIFGKEGTTASNAELDNVMI TDEEIRITNPVATEQYGTVANLQSSNTAPTTRTVNDQGalPgmVwQDRDV YLQGPiWAKiPHTDGHFHPSPMLGGFGLKHPPPPQIMiKNTVPANPPTTFSPAKFASFITQYSTGQVSVIEWELQKENSkrwnPEIQYTSnYNKsvNVDFt VDTNGVYSEPRPIGTRYLTRN

[0518] 적합하게는, 변이체는 SEQ ID NO: 33과 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일할 수 있다.

[0519] 적합하게는, AAV3B VP2 및 VP3 캡시드 단백질은 SEQ ID NO: 33의 N-말단 절두, 또는 SEQ ID NO: 33과 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일한 변이체의 N-말단 절두일 수 있다.

[0520] *LK03 혈청형*

[0521] AAV 벡터 입자는 LK03 캡시드 단백질을 포함할 수 있다. 적합하게는, AAV 벡터 입자는 LK03 캡시드 단백질에 의해 캡시드화될 수 있다.

[0522] AAV-LK03 캡 서열은 7개의 상이한 야생형 혈청형(AAV1, 2, 3B, 4, 6, 8, 9)으로부터의 단편으로 구성되며, 문헌 [Lisowski, L., et al., 2014. Nature, 506(7488), pp.382-386]에 기술되어 있다. 본 발명자들은 AAV-LK03 벡터가 시험관내 인간 발세포에서 거의 100%의 높은 형질도입을 달성할 수 있음을 입증하였다.

[0523] AAV 벡터 입자는 LK03 VP1 캡시드 단백질, LK03 VP2 캡시드 단백질 및/또는 LK03 VP3 캡시드 단백질을 포함할 수 있다. 적합하게는, AAV 벡터 입자는 LK03 VP1 캡시드 단백질, LK03 VP2 캡시드 단백질 및/또는 LK03 VP3 캡시드 단백질에 의해 캡시드화될 수 있다. 적합하게는, AAV 벡터 입자는 LK03B VP1, VP2 및 VP3 캡시드 단백질에 의해 캡시드화될 수 있다.

[0524] 적합하게는, LK03B VP1 캡시드 단백질은 SEQ ID NO: 34로 제시된 아미노산 서열, 또는 SEQ ID NO: 34와 적어도 90% 동일한 변이체를 포함하거나 이로 구성될 수 있다.

- [0525] 예시적인 LK03 VP1 캡시드 단백질(SEQ ID NO: 34):
- [0526] MAADGYLPDWLEDNLSEGIREWVALQPGAPKPKANQQHQDNARGLVLPGYKYLGPNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKEDTSFGGNLGRAVFAQKRLLEPLGLVEEAAKTAPGKKRPVDQSPQEPDSSSGVGKSGKQPARKRLNFGQTGDSESVDPDPQPLGEPAAPTSLSGNTMASGGGAPMADNNEGADGVGNSSGNWHCDSQLGDRVITTTSTRTWALPTYNNHLYKQISSQSGASNDNHYFGYSTPWGYFDNRFHCHFSRDPWQRLINNNWGRPKKLSFKLFNIQVKEVTQNDGTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVMPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYLNRTOGTTSGTTNQSRLLFSAQAGPQMSLQARNWLPGPCYRQRLSKTANDNNNSNFPWTAASKYHLNGRDSLVPNPGPAMASHKDDEEKFFPMHGNIIFGKEGTTASNAELDNVMI TDEEEIRTTNPVATEQYGTVANLQSSNTAPTTRTVNDQGALPGMVWQDRDVIYLGQPIWAKIPHDTDGHFHPSPMLGGFGLKHPPPIIMIKNTPVPANPPTTFSPAFAFASITQYSTGQVSVEIEWELQKENSKRWNPEIQYTSNYNKSNNVDFTVDTNGVYSEPRPIGTRYLTRPL
- [0527] 적합하게는, 변이체는 SEQ ID NO: 34와 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일할 수 있다.
- [0528] 적합하게는, LK03B VP2 및 VP3 캡시드 단백질은 SEQ ID NO: 34의 N-말단 절두, 또는 SEQ ID NO: 34와 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일한 변이체의 N-말단 절두일 수 있다.
- [0529] AAV9 혈청형
- [0530] AAV 벡터 입자는 AAV9 캡시드 단백질을 포함할 수 있다. 적합하게는, AAV 벡터 입자는 AAV9 캡시드 단백질에 의해 캡시드화될 수 있다.
- [0531] 본 발명자들은 AAV9 벡터가 시험관내 인간 발세포에서 높은 형질도입을 달성할 수 있음을 입증하였다.
- [0532] AAV 벡터 입자는 AAV9 VP1 캡시드 단백질, AAV9 VP2 캡시드 단백질 및/또는 AAV9 VP3 캡시드 단백질을 포함할 수 있다. 적합하게는, AAV 벡터 입자는 AAV9 VP1 캡시드 단백질, AAV9 VP2 캡시드 단백질 및/또는 AAV9 VP3 캡시드 단백질에 의해 캡시드화될 수 있다. 적합하게는, AAV 벡터 입자는 AAV9 VP1, VP2 및 VP3 캡시드 단백질에 의해 캡시드화될 수 있다.
- [0533] 적합하게는, AAV9 VP1 캡시드 단백질은 SEQ ID NO: 35로 제시된 아미노산 서열 또는 SEQ ID NO: 35와 적어도 90% 동일한 변이체를 포함하거나 이로 구성될 수 있다.
- [0534] 예시적인 AAV9 VP1 캡시드 단백질(SEQ ID NO: 35):
- [0535] MAADGYLPDWLEDNLSEGIREWVALKPGAPKPKANQQHQDNARGLVLPGYKYLGPNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKEDTSFGGNLGRAVFAQKRLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQAQKRLNFGQTGDSESVDPDPQPIGEPPAAPSGVGLTMA SGGGAPVADNNEGADGVGNSSGNWHCDSQLGDRVITTTSTRTWALPTYNNHLYKQISNSTSGSSNDNAYFGYSTPWGYFDNRFHCHFSRDPWQRLINNNW GFRPKRLNFKLFNIQVKEVTDNNGVKTIANNLSTVQVFTDSDYQLPYVLGSAHEGCLPPFPADVFMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTG NNFQFSYEFENVPFHSSYAHSQSLDRLMNPLIDQYLYLSKTINGSGQNQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQNNSEFAWPAGASSWA LNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNDADKVMITNEEEIKTTNPVATESYQVATNHQSAQAQQTGWVQNGILPGMVWQDRDVIYLGQPIWAKIPHDTDGNFHPSPMLGGFGMKHPPPIILIKNTPVPADPPTAFNKDKLNSFITQYSTGQVSVEIEWELQKENSKRWNPEIQYTSNYNKSNNVEFA VNTGEGVYSEPRPIGTRYLTRNL
- [0536] 적합하게는, 변이체는 SEQ ID NO: 35와 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일할 수 있다.
- [0537] 적합하게는, AAV9 VP2 및 VP3 캡시드 단백질은 SEQ ID NO: 35의 N-말단 절두 또는 SEQ ID NO: 35와 적어도 90%, 적어도 95%, 적어도 96%, 적어도 97%, 적어도 98% 또는 적어도 99% 동일한 변이체의 N-말단 절두일 수 있다.
- [0538] 다른 바이러스 벡터
- [0539] 레트로바이러스 및 렌티바이러스 벡터
- [0540] 본 발명의 벡터는 레트로바이러스 벡터 또는 렌티바이러스 벡터일 수 있다. 본 발명의 벡터는 레트로바이러스 벡터 입자 또는 렌티바이러스 벡터 입자일 수 있다.
- [0541] 레트로바이러스 벡터는 임의의 적합한 레트로바이러스로부터 유래되거나 유래 가능할 수 있다. 다수의 상이한 레트로바이러스가 확인되었다. 예는 뮤린 백혈병 바이러스(MLV), 인간 T-세포 백혈병 바이러스(HTLV), 마우스

유방 중앙 바이러스(MMTV), 라우스 육종 바이러스(RSV), 후지나미 육종 바이러스(FuSV), 몰로니 뮤린 백혈병 바이러스(Mo-MLV), FBR 뮤린 골육종 바이러스(FBR MSV), 몰로니 뮤린 육종 바이러스(Mo-MSV), 아벨슨 뮤린 백혈병 바이러스(A-MLV), 조류 골수세포종증 바이러스-29(MC29) 및 조류 적혈구세포종증 바이러스(AEV)를 포함한다.

- [0542] 레트로바이러스는 "단순한" 및 "복합체"의 2개의 카테고리로 크게 나눌 수 있다. 레트로바이러스는 심지어 7개의 그룹으로 추가로 분할될 수 있다. 이들 그룹 중 5개는 발암 가능성이 있는 레트로바이러스를 나타낸다. 나머지 2개의 그룹은 렌티바이러스 및 스푸마바이러스이다.
- [0543] 레트로바이러스 및 렌티바이러스 계통의 기본 구조는 5' LTR 및 3' LTR과 같은 많은 공통 특징을 공유한다. 이들 사이 또는 내부에는 계통이 패키징될 수 있도록 하는 패키징 신호, 프라이머 결합 부위, 숙주 세포 계통으로의 통합을 가능하게 하는 통합 부위, 및 패키징 성분을 인코딩하는 gag, pol 및 env 유전자가 위치하며, 이들은 바이러스 입자의 어셈블리에 필요한 폴리펩티드이다. 렌티바이러스는 HIV에서 rev 및 RRE 서열과 같은 추가 특징을 가지며, 이는 통합된 프로바이러스의 RNA 전사체를 핵으로부터 감염된 표적 세포의 세포질로 효율적으로 내보낼 수 있게 한다.
- [0544] 프로바이러스에서, 이들 유전자는 긴 말단 반복부(LTR)로 불리는 영역에 의해 양쪽 말단에 축적된다. LTR은 프로바이러스 통합 및 전사를 담당한다. LTR은 또한 인핸서-프로모터 서열로서 작용하고 바이러스 유전자의 발현을 제어할 수 있다.
- [0545] LTR 자체는 3개의 요소: U3, R 및 U5로 분할될 수 있는 동일한 서열이다. U3은 RNA의 3' 말단에 고유한 서열로부터 유래된다. R은 RNA의 양 말단에서 반복되는 서열로부터 유래된다. U5은 RNA의 5' 말단에 고유한 서열로부터 유래된다. 3개의 요소의 크기는 상이한 레트로바이러스 사이에서 상당히 다를 수 있다.
- [0546] 결합 있는 레트로바이러스 벡터 계통 gag에서, pol 및 env는 부재하거나 기능하지 않을 수 있다.
- [0547] 전형적인 레트로바이러스 벡터에서, 복제에 필수적인 하나 이상의 단백질 코딩 영역의 적어도 일부는 바이러스로부터 제거될 수 있다. 이는 바이러스 벡터 복제-결합을 만든다. 바이러스 계통의 일부는 또한 표적 숙주 세포를 형질도입할 수 있고/거나 이의 계통을 숙주 계통에 통합할 수 있는 후보 조절 모이어티를 포함하는 벡터를 생성하기 위해 벡터 계통에서 조절 제어 영역 및 리포터 모이어티에 작동가능하게 연결된 후보 조절 모이어티를 인코딩하는 라이브러리로 대체될 수 있다.
- [0548] 렌티바이러스 벡터는 더 큰 그룹의 레트로바이러스 벡터의 일부이다. 요약하면, 렌티바이러스는 영장류 및 비-영장류 그룹으로 나눌 수 있다. 영장류 렌티바이러스의 예는 비제한적으로, 인간 면역결핍 바이러스(HIV), 인간 후천성 면역결핍 증후군(AIDS)의 원인 인자; 및 원숭이 면역결핍 바이러스(SIV)를 포함한다. 비-영장류 렌티바이러스의 예는 프로토타입 "느린 바이러스" 비스나/메디 바이러스(VMV), 뿐만 아니라 관련된 염소 관절염-뇌염 바이러스(CAEV), 말 감염성 빈혈 바이러스(EIAV), 및 보다 최근에 기술된 고양이 면역결핍 바이러스(FIV) 및 소 면역결핍 바이러스(BIV)를 포함한다.
- [0549] 렌티바이러스 과는 렌티바이러스가 분열 세포 및 비분열 세포 둘 모두를 감염시키는 능력을 갖는다는 점에서 레트로바이러스와 상이하다. 대조적으로, MLV와 같은 다른 레트로바이러스는 예를 들어, 근육, 뇌, 폐 및 간 조직을 구성하는 것과 같은 비분열 또는 천천히 분열하는 세포를 감염시킬 수 없다.
- [0550] 본원에서 사용되는 렌티바이러스 벡터는 렌티바이러스로부터 유래될 수 있는 적어도 하나의 성분 부분을 포함하는 벡터이다. 바람직하게는, 상기 성분 부분은 벡터가 세포를 감염시키거나, 유전자를 발현시키거나, 복제되는 생물학적 기전에 관여한다.
- [0551] 렌티바이러스 벡터는 "영장류" 벡터일 수 있다. 렌티바이러스 벡터는 "비-영장류" 벡터(즉, 영장류, 특히 인간을 주로 감염시키지 않는 바이러스로부터 유래됨)일 수 있다. 비-영장류 렌티바이러스의 예는 영장류를 자연적으로 감염시키지 않는 렌티바이러스 과의 임의의 구성원일 수 있다.
- [0552] 렌티바이러스-기반 벡터의 예로서, HIV-1- 및 HIV-2-기반 벡터가 하기에 기재되어 있다.
- [0553] HIV-1 벡터는 단순 레트로바이러스에서도 발견되는 시스-작용 요소를 함유한다. gag 오픈 리딩 프레임으로 연장되는 서열은 HIV-1의 패키징에 중요한 것으로 나타났다. 따라서, HIV-1 벡터는 종종 번역 개시 코돈이 돌연변이된 gag의 관련 부분을 함유한다. 또한, 대부분의 HIV-1 벡터는 또한 RRE를 포함하는 env 유전자의 일부를 함유한다. Rev는 핵으로부터 세포질로 전장 또는 단일 스플라이싱된 mRNA의 수송을 허용하는 RRE에 결합한다. Rev 및/또는 RRE의 부재 하에, 전장 HIV-1 RNA는 핵에 축적된다. 대안적으로, Mason-Pfizer 원숭이 바이러스와 같은 특정 단순 레트로바이러스로부터의 구성적 수송 요소는 Rev 및 RRE에 대한 요건을 완화시키기 위해 사용될 수

있다. HIV-1 LTR 프로모터로부터의 효율적인 전사는 바이러스 단백질 tat를 필요로한다.

- [0554] 대부분의 HIV-2-기반 벡터는 구조적으로 HIV-1 벡터와 매우 유사하다. HIV-1-기반 벡터와 유사하게, HIV-2 벡터는 또한 전장 또는 단일 스플라이싱된 바이러스 RNA의 효율적인 수송을 위해 RRE를 필요로한다.
- [0555] 바람직하게는, 본 발명에서 사용되는 바이러스 벡터는 최소 바이러스 게놈을 갖는다.
- [0556] "최소 바이러스 게놈"에 있어서, 바이러스 벡터는 관심 뉴클레오티드 서열의 표적 숙주 세포로 감염, 형질도입 및 전달하는 데 필요한 기능성을 제공하기 위해 비필수 요소를 제거하고 필수 요소를 보유하도록 조작된 것으로 이해되어야 한다. 이 전략의 추가 세부사항은 WO 1998/017815호에서 찾을 수 있다.
- [0557] 바람직하게는, 숙주 세포/패키징 세포 내에서 바이러스 게놈을 생성하는데 사용되는 플라스미드 벡터는 패키징 성분의 존재 하에 표적 세포를 감염시킬 수 있는 바이러스 입자로 RNA 게놈의 패키징을 허용하기에 충분한 렌티 바이러스 유전자 정보를 가질 것이나, 최종 표적 세포 내에서 감염성 바이러스 입자를 생성하기 위한 독립적인 복제는 불가능하다. 바람직하게는, 벡터는 복제에 필수적인 기능성 gag-pol 및/또는 env 유전자 및/또는 다른 유전자가 결여되어 있다.
- [0558] 그러나, 숙주 세포/패키징 세포 내에서 바이러스 게놈을 생성하는데 사용되는 플라스미드 벡터는 또한 숙주 세포/패키징 세포에서 게놈의 전사를 지시하기 위해 렌티바이러스 게놈에 작동가능하게 연결된 전사 조절 제어 서열을 포함할 것이다. 이러한 조절 서열은 전사된 바이러스 서열(즉, 5' U3 영역)과 회합된 천연 서열일 수 있거나, 이들은 또 다른 바이러스 프로모터(예를 들어, CMV 프로모터)와 같은 이종성 프로모터일 수 있다.
- [0559] 벡터는 바이러스 인핸서 및 프로모터 서열이 결실된 자가-불활성화(SIN) 벡터일 수 있다. SIN 벡터는 야생형 벡터와 유사한 효능으로 생체내에서 생성되고 비-분열 세포를 형질도입할 수 있다. SIN 프로바이러스에서 긴 말단 반복부(LTR)의 전사 불활성화는 복제-적격 바이러스에 의한 동원을 방지해야한다. 이는 또한 LTR의 임의의 시스-작용 효과를 제거함으로써 내부 프로모터로부터 유전자의 조절된 발현을 가능하게 해야한다.
- [0560] 벡터는 통합-결합될 수 있다. 통합 결합 렌티바이러스 벡터(IDLV)는 예를 들어, 벡터를 촉매적으로 불활성인 인테그라제(예를 들어, 촉매 부위에 D64V 돌연변이를 갖는 HIV 인테그라제)로 패키징하거나 벡터 LTR로부터 필수 att 서열을 변형 또는 결실시킴으로써 또는 상기의 조합에 의해 생성될 수 있다.
- [0561] *아데노바이러스 벡터*
- [0562] 본 발명의 벡터는 아데노바이러스 벡터일 수 있다. 본 발명의 벡터는 아데노바이러스 벡터 입자일 수 있다.
- [0563] 아데노바이러스는 RNA 중간체를 거치지 않는 이중-가닥의 선형 DNA 바이러스이다. 50개 이상의 상이한 인간 혈청형의 아데노바이러스가 유전자 서열 상동성에 기초하여 6개의 서브그룹으로 나누어진다. 아데노바이러스의 자연 표적은 호흡기 및 위장 상피이며, 일반적으로 가벼운 증상만을 일으킨다. 혈청형 2 및 5(95% 서열 상동성)는 아데노바이러스 벡터 시스템에서 가장 일반적으로 사용되며, 일반적으로 청소년의 상기도 감염과 관련이 있다.
- [0564] 아데노바이러스는 유전자 요법 및 이종성 유전자의 발현을 위한 벡터로서 사용되어 왔다. 큰(36 kb) 게놈은 최대 8 kb의 외래 삽입 DNA를 수용할 수 있고, 보완 세포주에서 효율적으로 복제하여 최대  $10^{12}$ 의 매우 높은 역가를 생성할 수 있다. 따라서, 아데노바이러스는 일차 비-복제성 세포에서 유전자의 발현을 연구하기 위한 최상의 시스템 중 하나이다.
- [0565] 아데노바이러스 게놈으로부터의 바이러스 또는 외래 유전자의 발현은 복제 세포를 필요로 하지 않는다. 아데노 바이러스 벡터는 수용체 매개된 세포내 이입에 의해 세포에 진입한다. 일단 세포 내부로 들어가면, 아데노 바이러스 벡터는 숙주 염색체에 거의 통합되지 않는다. 대신, 이들은 숙주 핵에서 선형 게놈으로서 에피솜적으로(숙주 게놈과 독립적으로) 기능한다. 따라서, 재조합 아데노바이러스의 사용은 숙주 게놈으로의 무작위 통합과 관련된 문제를 완화시킨다.
- [0566] *단순 포진 바이러스 벡터*
- [0567] 본 발명의 벡터는 단순 포진 바이러스 벡터일 수 있다. 본 발명의 벡터는 단순 포진 바이러스 벡터 입자일 수 있다.
- [0568] 단순 포진 바이러스(HSV)는 유전자 전달 벡터로서 유리한 특성을 갖는 항신경성 DNA 바이러스이다. HSV는 감염성이 높으므로, HSV 벡터는 외인성 유전 물질을 세포로 전달하기 위한 효율적인 비히클이다. 바이러스 복제는 시험관내에서 트랜스로 보완되어 비-병원성 벡터의 고역가 순수한 제조물의 직접적인 생산을 가능하게 할 수 있

는 즉시 초기 유전자에서 널 돌연변이에 의해 용이하게 파괴된다. 게놈은 크고(152 Kb), 많은 바이러스 유전자는 시험관내 복제에 필수적이어서, 크거나 다수의 전이유전자로의 대체를 가능하게 한다. 야생형 바이러스에 의한 잠복 감염은 숙주 수명 기간 동안 감각 뉴런 핵에서 에피솜 바이러스 지속성을 초래한다. 백터는 비-병원성이며, 재활성화될 수 없고 장기간 지속될 수 없다. 잠복 활성 프로모터 복합체는 신경계에서 장기간 안정한 전이유전자 발현을 달성하기 위해 백터 설계에서 이용될 수 있다. HSV 백터는 바이러스에 의해 인식되는 세포 수용체의 광범위한 발현 패턴으로 인해 광범위한 조직을 형질도입한다. 세포 진입과 관련된 과정에 대한 이해가 증가함에 따라 HSV 백터의 지향성을 표적화할 수 있었다.

- [0569] *다른 바이러스 백터*
- [0570] 다른 적합한 바이러스 백터는 문헌 [Lundstrom, K., 2018. Diseases, 6(2), p.42]에 기재된 것들을 포함한다.
- [0571] 본 발명의 백터는 알파바이러스 백터일 수 있다. 본 발명의 백터는 알파바이러스 백터 입자일 수 있다. 본 발명의 백터는 플라비바이러스 백터일 수 있다. 본 발명의 백터는 플라비바이러스 백터 입자일 수 있다.
- [0572] 자가-증폭 ssRNA 바이러스는 양성 극성의 게놈을 갖는 알파바이러스(예를 들어, 썬리키 삼립 바이러스, 신드비스 바이러스, 베네수엘라 말 뇌염 바이러스 및 M1) 및 플라비바이러스(예를 들어, 쿤진 바이러스, 웨스트 나일 바이러스 및 뎅기 바이러스)를 포함한다. 알파바이러스는 주로 암 치료를 위한 전임상 유전자 요법 연구에 적용되었다. 알파바이러스 백터는 네이키드 RNA, 증상 플라스미드 DNA 백터 및 제조합 복제-결핍 또는-능숙 입자의 형태로 전달될 수 있다.
- [0573] 본 발명의 백터는 램도바이러스 백터일 수 있다. 본 발명의 백터는 램도바이러스 백터 입자일 수 있다. 본 발명의 백터는 홍역바이러스 백터일 수 있다. 본 발명의 백터는 홍역바이러스 백터 입자일 수 있다.
- [0574] 램도바이러스(예를 들어, 광견병 및 수포성 구내염 바이러스) 및 홍역 바이러스는 음성 가닥 게놈을 수반한다. 램도바이러스 중에서, 제조합 수포성 구내염 바이러스(VSV)가 전임상 유전자 요법 연구에 적용되었다. 홍역 바이러스(예를 들어, MV-Edm)는 다수의 유전자 요법 적용을 발견하였다.
- [0575] 본 발명의 백터는 뉴캐슬병 바이러스 백터일 수 있다. 본 발명의 백터는 뉴캐슬병 바이러스 백터 입자일 수 있다.
- [0576] ssRNA 파라믹소바이러스 뉴캐슬병 바이러스(NDV)는 종양 세포에서 특이적으로 복제하므로 암 유전자 요법에 자주 적용되었다.
- [0577] 본 발명의 백터는 폭스바이러스 백터일 수 있다. 본 발명의 백터는 폭스바이러스 백터 입자일 수 있다.
- [0578] 폭스바이러스의 특징은 이들의 dsDNA 게놈으로서, 30kb 초과외래 DNA를 관대하게 수용할 수 있다. 폭스바이러스는 유전자 요법 백터로서 여러 적용을 발견하였다. 예를 들어, 백시니아 바이러스 백터는 암의 치료 가능성을 입증하였다. 백시니아 바이러스는 대략 190 kb 선형, 이중-가닥 DNA 게놈을 갖는 큰 피막 폭스바이러스이다. 백시니아 바이러스는 최대 약 25kb의 외래 DNA를 수용할 수 있으며, 이는 또한 큰 유전자의 전달에 유용하다. 유전자 요법 적용에 적합한 다수의 약독화된 백시니아 바이러스 균주, 예를 들어, MVA 및 NYVAC 균주가 당 분야에 공지되어 있다.
- [0579] 본 발명의 백터는 피코르나바이러스 백터일 수 있다. 본 발명의 백터는 피코르나바이러스 백터 입자일 수 있다.
- [0580] 피코르노바이러스는 비-피막 ssRNA 바이러스이다. 피코르나바이러스 과에 속하는 콕사키에바이러스는 종양용해 백터로서 적용되었다.
- [0581] **변이체, 유도체, 유사체, 상동체 및 단편**
- [0582] 본원에 언급된 특정 단백질 및 뉴클레오티드 이외에, 본 발명은 또한 이의 변이체, 유도체, 상동체 및 단편을 포함한다.
- [0583] 본 발명의 맥락에서, 임의의 주어진 서열의 "변이체"는 잔기의 특정 서열이 (아미노산 또는 핵산 잔기이든) 해당 폴리펩티드 또는 폴리뉴클레오티드가 이의 내인성 기능 중 적어도 하나를 보유하는 방식으로 변형된 서열이다. 변이체 서열은 천연 발생 폴리펩티드 또는 폴리뉴클레오티드에 존재하는 적어도 하나의 잔기의 첨가, 결실, 치환, 변형, 대체 및/또는 변이에 의해 획득될 수 있다. 예를 들어, 변이체 프로모터 서열은 이것이 획득되는 프로모터 서열의 적어도 일부 수준의 활성 및 특이성을 보유한다.
- [0584] 본 발명의 단백질 또는 폴리펩티드와 관련하여 본원에서 사용되는 용어 "유도체"는 생성된 단백질 또는 폴리펩

티드가 이의 내인성 기능 중 적어도 하나를 유지하는 한, 서열로부터의 또는 서열로의 하나(또는 그 초과)의 아미노산 잔기의 임의의 치환, 변이, 변형, 대체, 결실 및/또는 부가를 포함한다.

[0585] 전형적으로, 아미노산 치환은 예를 들어, 1, 2 또는 3, 내지 10 또는 20개의 치환이 이루어질 수 있으며, 단, 변형된 서열은 요구되는 활성 또는 능력을 보유한다. 아미노산 치환은 비-자연 발생 유사체의 사용을 포함할 수 있다.

[0586] 본 발명에서 사용되는 단백질은 또한 침묵의 변화를 생성하고 기능적으로 동등한 단백질을 생성하는 아미노산 잔기의 결실, 삽입 또는 치환을 가질 수 있다. 고의적인 아미노산 치환은 내인성 기능이 유지되는 한 잔기의 극성, 전하, 용해도, 소수성, 친수성 및/또는 양친매성 성질의 유사성에 기초하여 이루어질 수 있다. 예를 들어, 음으로 하전된 아미노산은 아스파르트산 및 글루탐산을 포함하고; 양으로 하전된 아미노산은 리신 및 아르기닌을 포함하고; 유사한 친수성 값을 갖는 하전되지 않은 극성 헤드 기를 갖는 아미노산은 아스파라긴, 글루타민, 세린, 트레오닌 및 티로신을 포함한다.

[0587] 보존적 치환은 예를 들어, 하기 표에 따라 이루어질 수 있다. 제2 컬럼의 동일한 블록 및 바람직하게는 제3 컬럼의 동일한 라인에 있는 아미노산은 서로 치환될 수 있다:

표 1

지방족	비극성	G A P
		I L V
	극성 - 비하전됨	C S T M
		N Q
	극성-하전됨	D E
		K R H
방향족		F W Y

[0589] 본원에서 사용되는 용어 "상동체"는 야생형 아미노산 서열 또는 야생형 뉴클레오티드 서열과 특정 상동성을 갖는 변이체를 의미한다. 용어 "상동성"은 "동일성"과 동일시될 수 있다.

[0590] 본 문맥에서, 상동성 서열은 대상체 서열과 적어도 50%, 55%, 65%, 75%, 85% 또는 90% 동일할 수 있는, 바람직하게는 적어도 95%, 96% 또는 97% 또는 98% 또는 99% 동일할 수 있는 아미노산 서열을 포함하는 것으로 간주된다. 전형적으로, 상동체는 대상체 아미노산 서열과 동일한 활성 부위 등을 포함할 것이다. 상동성은 또한 유사성(즉, 유사한 화학적 특성/기능을 갖는 아미노산 잔기)의 관점에서 고려될 수 있지만, 본 발명의 맥락에서, 서열 동일성의 관점에서 상동성으로 표현하는 것이 바람직하다.

[0591] 본 문맥에서, 상동성 서열은 대상체 서열과 적어도 50%, 55%, 65%, 75%, 85% 또는 90% 동일할 수 있는, 바람직하게는 적어도 95%, 96% 또는 97% 또는 98% 또는 99% 동일할 수 있는 뉴클레오티드 서열을 포함하는 것으로 간주된다. 상동성은 또한 유사성 측면에서 고려될 수 있지만, 본 발명의 맥락에서, 서열 동일성 측면에서 상동성을 표현하는 것이 바람직하다.

[0592] 바람직하게는, 본원에 상세히 기재된 SEQ ID NO 중 어느 하나와 동일성 퍼센트를 갖는 서열에 대한 언급은 언급된 SEQ ID NO의 전체 길이에 걸쳐 언급된 동일성 퍼센트를 갖는 서열을 지칭한다.

[0593] 상동성 비교는 육안으로 또는 보다 일반적으로 용이하게 이용 가능한 서열 비교 프로그램의 도움으로 수행될 수 있다. 이러한 상업적으로 이용 가능한 컴퓨터 프로그램은 2개 이상의 서열 사이의 상동성 또는 동일성 퍼센트를 계산할 수 있다.

[0594] 상동성 퍼센트는 인접한 서열에 대해 계산될 수 있고, 즉, 하나의 서열은 다른 서열과 정렬되고, 한 서열의 각 아미노산 또는 뉴클레오티드는 한 번에 하나의 잔기로 다른 서열의 상응하는 아미노산 또는 뉴클레오티드와 직접 비교된다. 이것을 "갭핑되지 않은" 정렬이라한다. 전형적으로, 이러한 갭핑되지 않은 정렬은 비교적 적은 수의 잔기에 대해서만 수행된다.

[0595] 이는 매우 간단하고 일관된 방법이지만, 예를 들어, 달리 동일한 서열 쌍에서, 아미노산 또는 뉴클레오티드 서열에서 하나의 삽입 또는 결실이 다음 잔기 또는 코돈이 정렬을 이탈되게 하여 전체 정렬이 수행될 때 잠재적으로 상동성 퍼센트의 큰 감소를 초래할 수 있음은 고려되지 않았다. 결과적으로, 대부분의 서열 비교 방법은 전체 상동성 스코어에 부당한 페널티를 주지 않으면서 가능한 삽입 및 결실을 고려하는 최적의 정렬을 생성하도록

설계된다. 이는 국소 상동성을 최대화하기 위해 서열 정렬에 "갭"을 삽입함으로써 달성된다.

- [0596] 그러나, 이러한 보다 복잡한 방법은 정렬에서 발생하는 각 갭에 "갭 페널티"를 할당하여, 동일한 수의 동일한 아미노산 또는 뉴클레오티드에 있어서, 비교된 두 서열 사이의 더 높은 관련성을 반영하여 가능한 한 적은 갭을 갖는 서열 정렬이 많은 갭을 갖는 것보다 더 높은 스코어를 달성할 것이다. "아핀 갭 코스트"는 전형적으로 갭의 존재에 대해 비교적 높은 코스트를 부과하고 갭의 각각의 후속 잔기에 대해 더 적은 페널티를 부과하는 데 사용된다. 이것은 가장 일반적으로 사용되는 갭 스코어링 시스템이다. 높은 갭 페널티는 물론 더 적은 갭을 갖는 최적화된 정렬을 생성할 것이다. 대부분의 정렬 프로그램은 갭 페널티가 수정될 수 있도록 한다. 그러나, 서열 비교를 위해 이러한 소프트웨어를 사용할 때 디폴트 값을 사용하는 것이 바람직하다. 예를 들어, GCG Wisconsin Bestfit 패키지를 사용할 때, 아미노산 서열에 대한 디폴트 갭 페널티는 갭의 경우 -12이고 각 연장부의 경우 -4이다.
- [0597] 따라서, 최대 상동성 퍼센트의 계산은 갭 페널티를 고려하여 먼저 최적의 정렬의 생성을 필요로 한다. 이러한 정렬을 수행하기에 적합한 컴퓨터 프로그램은 GCG Wisconsin Bestfit 패키지이다(University of Wisconsin, USA; Devereux et al. (1984) Nucleic Acids Research 12: 387). 서열 비교를 수행할 수 있는 다른 소프트웨어의 예는 비제한적으로, BLAST 패키지(문헌 [Ausubel et al. (1999) *ibid* - Ch. 18] 참조), FASTA(Atschul et al. (1990) J. Mol. Biol. 403-410), EMBOSS Needle (Madeira, F., et al., 2019. Nucleic acids research, 47(W1), pp.W636-W641) 및 비교 툴의 GENWORKS 수트를 포함한다. BLAST 및 FASTA 둘 모두는 오프라인 및 온라인 검색에 이용 가능하다(문헌 [Ausubel et al. (1999) *ibid*, pages 7-58 to 7-60] 참조). 그러나, 일부 적용의 경우, GCG Bestfit 프로그램을 사용하는 것이 바람직하다. 또 다른 도구인 BLAST 2 Sequences는 또한 단백질 및 뉴클레오티드 서열을 비교하는데 이용 가능하다(FEMS Microbiol. Lett. (1999) 174(2):247-50; FEMS Microbiol. Lett. (1999) 177(1):187-8).
- [0598] 최종 상동성 퍼센트는 동일성의 관점에서 측정될 수 있지만, 정렬 과정 자체는 전형적으로 모두-또는-전무(all-or-nothing) 쌍 비교를 기반으로 하지 않는다. 대신, 화학적 유사성 또는 진화적 거리에 기반하여 각각의 쌍별 비교에 스코어를 할당하는 스케일링된 유사성 스코어 매트릭스가 일반적으로 사용된다. 일반적으로 사용되는 이러한 매트릭스의 예는 BLOSUM62 매트릭스(BLAST 슈트 프로그램에 대한 디폴트 매트릭스)이다. GCG 위스콘신 프로그램은 일반적으로 공개 디폴트 값 또는 제공되는 경우 사용자 정의 기호 비교 표를 사용한다(자세한 내용은 사용자 설명서 참조). 일부 적용의 경우, GCG 패키지에 있어서 공개 디폴트 값을 사용하거나, 다른 소프트웨어의 경우, BLOSUM62와 같은 디폴트 매트릭스를 사용하는 것이 바람직하다.
- [0599] 소프트웨어가 최적의 정렬을 생성하면, 상동성 퍼센트, 바람직하게는 서열 동일성 퍼센트를 계산하는 것이 가능하다. 소프트웨어는 전형적으로 서열 비교의 일부로서 이를 수행하고 수치 결과를 생성한다. 퍼센트 서열 동일성은 언급된 SEQ ID NO에서 총 잔기의 백분율로서 동일한 잔기의 수로서 계산될 수 있다.
- [0600] "단편"은 또한 변이체이고, 용어는 전형적으로 기능적으로 또는, 예를 들어, 검정에서 관심 있는 폴리펩티드 또는 폴리뉴클레오티드의 선택된 영역을 지칭한다. 따라서, "단편"은 전장 폴리펩티드 또는 폴리뉴클레오티드의 일부인 아미노산 또는 핵산 서열을 지칭한다.
- [0601] 이러한 변이체, 유도체, 상동체 및 단편은 부위-지정 돌연변이유발과 같은 표준 재조합 DNA 기술을 사용하여 제조될 수 있다. 삽입이 이루어질 경우, 삽입 부위의 어느 한 쪽에 자연-발생 서열에 상응하는 5' 및 3' 측접 영역과 함께 삽입부를 인코딩하는 합성 DNA가 만들어질 수 있다. 측접 영역은 자연 발생 서열의 부위에 상응하는 편리한 제한 부위를 함유하여 서열이 적절한 효소(들)로 절단되고 합성 DNA가 절단부에 리게이션될 수 있다. 이후, DNA는 본 발명에 따라 발현되어 인코딩된 단백질을 제조한다. 이러한 방법은 DNA 서열의 조작을 위해 당 분야에 공지된 수많은 표준 기술의 예시일 뿐이며, 다른 공지된 기술이 또한 사용될 수 있다.
- [0602] **세포**
- [0603] 일 양태에서, 본 발명은 본 발명의 폴리뉴클레오티드 또는 벡터를 포함하는 세포를 제공한다. 세포는 분리된 세포일 수 있다. 세포는 인간 세포, 적합하게는 분리된 인간 세포일 수 있다.
- [0604] 본 발명에서 사용되는 폴리뉴클레오티드를 포함하는 벡터는 형질감염, 형질도입 및 형질전환과 같은 당 분야에 공지된 다양한 기술을 사용하여 세포 내로 도입될 수 있다. 적합하게는, 본 발명의 벡터는 형질감염 또는 형질도입에 의해 세포 내로 도입된다.
- [0605] 세포는 종래 기술에 공지된 임의의 세포 유형일 수 있다.

- [0606] 적합하게는, 세포는 생산자 세포일 수 있다. 용어 "생산자 세포"는 바이러스 입자를 생산하는데 필요한 모든 요소의 일시적인 형질감염, 안정한 형질감염 또는 벡터 형질도입 후, 바이러스 입자를 생산하는 세포 또는 바이러스 입자를 생산하는 데 필요한 요소를 안정적으로 포함하도록 조작된 임의의 세포를 포함한다. 적합한 생산자 세포는 당업자에게 잘 알려져 있을 것이다. 적합한 생산자 세포주는 HEK 293(예를 들어, HEK 293T), HeLa 및 A549 세포주를 포함한다.
- [0607] 적합하게는, 세포는 패키징 세포일 수 있다. 용어 "패키징 세포"는 감염성 재조합 바이러스를 패키징하는데 필요한 요소의 일부 또는 전부를 함유하는 세포를 포함한다. 패키징 세포는 재조합 바이러스 벡터 게놈이 부족할 수 있다. 전형적으로, 이러한 패키징 세포는 바이러스 구조 단백질을 발현할 수 있는 하나 이상의 벡터를 함유한다. 피막 바이러스 입자의 생산에 필요한 요소의 단지 일부만을 포함하는 세포는 각각의 추가의 필요한 요소의 일시적인 형질감염, 형질도입 또는 안정적인 통합의 후속 단계를 통해 바이러스 입자 생산자 세포주의 생성에서 중간 시약으로서 유용하다. 이러한 중간 시약은 용어 "패키징 셀"에 포함된다. 적합한 패키징 세포는 당업자에게 잘 알려져 있을 것이다.
- [0608] 적합하게는, 세포는 신장 세포 또는 사구체 세포, 예를 들어, 발세포일 수 있다. 적합하게는, 세포는 불멸화 신장 세포 또는 사구체 세포, 예를 들어, 불멸화 발세포일 수 있다. 적합한 발세포 세포주, 예를 들어, CIHP-1은 당업자에게 잘 알려져 있을 것이다. 불멸화 발세포를 생성하는 방법은 당업자에게 잘 알려져 있을 것이다. 적합한 방법은 문헌[Ni, L., et al., 2012. Nephrology, 17(6), pp.525-531]에 기재되어 있다.
- [0609] **약학적 조성물**
- [0610] 일 양태에서, 본 발명은 본 발명의 폴리뉴클레오티드 또는 벡터 또는 본 발명의 세포를 포함하는 약학적 조성물을 제공한다.
- [0611] 약학적 조성물은 치료학적 유효량의 약학적 활성제, 즉, 벡터를 포함하거나 이로 구성된 조성물이다. 이는 바람직하게는 약학적으로 허용되는 담체, 희석제 또는 부형제(이의 조합 포함)를 포함한다.
- [0612] "약학적으로 허용되는"은 제형이 멸균되고 발열원이 없는 것을 포함한다. 담체, 희석제 및/또는 부형제는 벡터와 양립가능하고 이의 수용자에게 유해하지 않다는 의미에서 "허용"되어야 한다. 전형적으로, 담체, 희석제 및 부형제는 멸균되고 발열원이 없는 식염수 또는 주입 매질일 것이지만, 다른 허용되는 담체, 희석제 및 부형제가 사용될 수 있다.
- [0613] 치료용으로 허용되는 담체, 희석제 및 부형제는 약학 분야에 잘 알려져 있다. 약학적 담체, 부형제 또는 희석제의 선택은 의도된 투여 경로 및 표준 약학적 관행과 관련하여 선택될 수 있다. 약학적 조성물은 담체, 부형제 또는 희석제로서 - 또는 이에 추가하여, 임의의 적합한 결합제(들), 운할제(들), 현탁제(들), 코팅제(들) 또는 가용화제(들)를 포함할 수 있다.
- [0614] 약학적으로 허용되는 담체의 예는 예를 들어, 물, 염 용액, 알콜, 실리콘, 왁스, 바셀린, 식물성 오일, 폴리에틸렌 글리콜, 프로필렌 글리콜, 리포솜, 당, 젤라틴, 락토스, 아밀로스, 마그네슘 스테아레이트, 탈크, 계면활성제, 규산, 점성 과라판, 향유, 지방산 모노글리세리드 및 디글리세리드, 페트로에트랄 지방산 에스테르, 하이드록시메틸-셀룰로스, 폴리비닐피롤리돈 등을 포함한다.
- [0615] 본 발명에 따른 벡터, 세포 또는 약학적 조성물은 본원에 기재된 질환을 치료 및/또는 예방하기에 적합한 방식으로 투여될 수 있다. 투여의 양 및 빈도는 대상체의 상태, 및 대상체의 질병의 유형 및 중증도와 같은 인자에 의해 결정될 것이지만, 적절한 투여량은 임상 시험에 의해 결정될 수 있다. 약학적 조성물은 이에 따라 제형화될 수 있다.
- [0616] 본 발명에 따른 벡터, 세포 또는 약학적 조성물은 비경구적으로, 예를 들어, 정맥내로 또는 주입 기술에 의해 투여될 수 있다. 벡터, 세포 또는 약학적 조성물은 다른 물질, 예를 들어, 용액을 혈액과 등장성으로 만들기 위해 충분한 염 또는 글루코스를 함유할 수 있는 멸균 수용액의 형태로 투여될 수 있다. 수용액은 적합하게 완충될 수 있다(바람직하게는 pH 3 내지 9). 약학적 조성물은 이에 따라 제형화될 수 있다. 멸균 조건 하에 적합한 비경구 제형의 제조는 당업자에게 널리 공지된 표준 약학적 기술에 의해 용이하게 달성된다.
- [0617] 본 발명에 따른 벡터, 세포 또는 약학적 조성물은 전신으로, 예를 들어, 정맥내 주사에 의해 투여될 수 있다.
- [0618] 본 발명에 따른 벡터, 세포 또는 약학적 조성물은 국소적으로, 예를 들어, 신장으로의 표적 투여에 의해 투여될 수 있다. 적합하게는, 벡터, 세포 또는 약학적 조성물은 신장 동맥으로의 주사 또는 요관 또는 피막하 주사에

의해 투여될 수 있다.

- [0619] 약학적 조성물은 주입 매질, 예를 들어, 멸균 등장액에 본 발명의 벡터 또는 세포를 포함할 수 있다. 약학적 조성물은 앰플, 일회용 주사기 또는 유리 또는 플라스틱으로 제조된 다회 용량 바이알에 포함될 수 있다.
- [0620] 벡터, 세포 또는 약학적 조성물은 단일 또는 다중 용량으로 투여될 수 있다. 특히, 벡터, 세포 또는 약학적 조성물은 단일 또는 일회성 복용량으로 투여될 수 있다. 약학적 조성물은 이에 따라 제형화될 수 있다.
- [0621] 벡터, 세포 또는 약학적 조성물은 다양한 용량으로 투여될 수 있다(예를 들어, kg 당 벡터 계놈(vg)으로 측정됨). 임의의 경우에 의사는 임의의 개별 대상체에 가장 적합할 실제 투여량을 결정할 것이며, 이는 특정 대상체의 연령, 체중 및 반응에 따라 달라질 것이다. 그러나, 전형적으로, 본 발명의 AAV 벡터의 경우,  $10^{10}$  내지  $10^{14}$  vg/kg, 또는  $10^{11}$  내지  $10^{13}$  vg/kg의 용량이 투여될 수 있다.
- [0622] 약학적 조성물은 하나 이상의 다른 치료제를 추가로 포함할 수 있다.
- [0623] 본 발명은 본 발명의 벡터, 세포 및/또는 약학적 조성물을 포함하는 키트의 용도를 추가로 포함한다. 바람직하게는, 상기 키트는 본원에 기재된 방법 및 용도, 예를 들어, 본원에 기재된 바와 같은 치료 방법에 사용하기 위한 것이다. 바람직하게는 상기 키트는 키트 성분의 사용 설명서를 포함한다.
- [0624] **질병의 치료 및/또는 예방 방법**
- [0625] 일 양태에서, 본 발명은 의약으로서 사용하기 위한 본 발명에 따른 벡터, 세포 또는 약학적 조성물을 제공한다.
- [0626] 일 양태에서, 본 발명은 의약의 제조에서의 본 발명에 따른 벡터, 세포 또는 약학적 조성물의 용도를 제공한다.
- [0627] 일 양태에서, 본 발명은 본 발명에 따른 벡터, 세포 또는 약학적 조성물을 이를 필요로 하는 대상체에게 투여하는 방법을 제공한다.
- [0628] 사구체 질환
- [0629] 본 발명에 따른 벡터, 세포 또는 약학적 조성물은 대상체에서 사구체 질환을 치료하는데 사용될 수 있다. 적합하게는, 대상체는 인간 대상체이다.
- [0630] 일 양태에서, 본 발명은 사구체 질환의 예방 또는 치료에 사용하기 위한 본 발명에 따른 벡터, 세포 또는 약학적 조성물을 제공한다.
- [0631] 관련 양태에서, 본 발명은 사구체 질환을 예방 또는 치료하기 위한 의약의 제조를 위한 본 발명에 따른 벡터, 세포 또는 약학적 조성물의 용도를 제공한다.
- [0632] 관련 양태에서, 본 발명은 본 발명에 따른 벡터, 세포 또는 약학적 조성물을 이를 필요로 하는 대상체에게 투여하는 것을 포함하는 사구체 질환의 예방 또는 치료 방법을 제공한다.
- [0633] 사구체 질환은 신증 또는 신염으로 분류될 수 있다. 콩팥 증후군은 종종 발세포-발세포 또는 발세포-GBM 상호작용의 완전성에 영향을 미치는 인자를 포함한다. 대조적으로, 콩팥 증후군의 병인에 관여하는 인자는 다양할 수 있지만, 순환 혈소판 및 백혈구, GBM, 상주 사구체 내피 세포, 및 혈관사이 세포를 포함할 수 있다. (Chiang, C.K. and Inagi, R., 2010. Nature Reviews Nephrology, 6(9), p.539).
- [0634] 적합하게는, 사구체 질환은 유전적 사구체 질환, 즉, 유전되는 사구체 질환이다. 유전자 사구체 질환은 콩팥 증후군과 같은 발세포-관련 유전적 사구체 질환, 및 알포트 증후군과 같은 GBM-관련 사구체 질환을 포함한다.
- [0635] 적합하게는, 사구체 질환은 발세포-관련 유전적 사구체 질환이다. 발세포-관련 유전 사구체 질환은 핀란드 유형의 선천성 콩팥 증후군, 선천성 콩팥 증후군 유형 2, 가족성 콩팥 증후군 유형 3, 프레이저 증후군 및 데니스-드라시 증후군, Schimke 면역-골이형성증, CD2AP의 돌연변이에 의해 유발된 콩팥 증후군, 액틴-4의 돌연변이에 의해 유발된 콩팥 증후군, TRPC6의 돌연변이에 의해 유발된 콩팥 증후군 및 엡스타인 및 페흐트너 증후군을 포함한다. 적합하게는, 사구체 질환은 콩팥 증후군이다.
- [0636] 적합하게는, 사구체 질환은 GBM-관련 유전적 사구체 질환이다. 발세포-관련 유전적 사구체 질환은 X-연결 알포트 증후군, 상염색체 열성 알포트 증후군, 상염색체 우성 알포트 증후군, 얇은 기저막 질환, 피어슨 증후군 및 손발톱-슬개골 증후군을 포함한다. 적합하게는, 사구체 질환은 알포트 증후군(AS)이다. AS는 또한 가족성 신장염, 유전성 신장염, 얇은 기저막 질환 및 얇은 기저막 신병증으로 공지되어 있다.

- [0637] **실시에**
- [0638] 본 발명은 이제 실시예에 의해 추가로 기술될 것이며, 이는 당업자가 본 발명을 수행하는 데 도움을 주기 위한 것이며 어떠한 방식으로든 본 발명의 범위를 제한하려는 것이 아니다.
- [0639] **실시예 1 - 최소 네프린 프로모터의 설계, 작제 및 시험**
- [0640] 최소 네프린 프로모터의 설계
- [0641] 인간 NPHS1 프로모터는 문헌[Moeller et al. 2002 J Am Soc Nephrol, 13(6):1561-7 and Wong MA et al. 2000 Am J Physiol Renal Physiol, 279(6):F1027-32]에 기재되어 있다. 이 NPHS1 프로모터는 1.2kb 단편이고, 발세포-특이적인 것으로 보인다. 이는 이하에서 "FL" 네프린 프로모터로 지칭되며, 도 1a에 도시되어 있다.
- [0642] FL 네프린 프로모터를 N-말단 서열을 결실시킴으로써 초기에 822 bp(시작 코돈을 제외한 819 bp)로 절단하였다. 이는 이하에서 "미디" 네프린 프로모터로 지칭되며, 도 1b에 도시되어 있다.
- [0643] 미디 네프린 프로모터를 중앙 영역으로부터 추정되는 일반 전사 도메인을 제거함으로써 268 bp(시작 코돈을 제외한 265 bp)로 추가로 절단되었다. 이는 이하에서 "미니" 네프린 프로모터로 지칭되며, 도 1c에 도시되어 있다.
- [0644] 벡터 작제물의 작제
- [0645] *미디 네프린 프로모터*
- [0646] pACE\_hNPHS1 프로모터를 도 2a에 도시된 바와 같이 BamHI 및 ClaI 제한 부위를 도입하기 위한 주형으로 사용하였다. 이어서, 단편을 겔 추출하고, pLenti GFP Blast 벡터에 결찰시키기 전에 37°C에서 1h 동안 ClaI 및 BamHI로 분해하였다. 결찰을 안정한 적격 E.Coli 세포로 추가로 형질전환시키고, DNA를 추출하고 시퀀싱하였다 (미디 프로모터). 최종 렌티바이러스 벡터는 도 2B에 도시되어 있다.
- [0647] *미니 네프린 프로모터*
- [0648] pACE\_hNPHS1 프로모터를 사용하여 도 3a에 도시된 바와 같이 오버행(OH)을 PCR하였다. OH를 함유하는 프로모터의 2개의 섹션을 NEBuilder HiFi 어셈블리 반응을 위해 pLenti GFP Blast 벡터로 겔 추출하였다. 이어서, 안정한 적격 E.Coli 세포로의 형질전환 전에, 결찰 반응을 DNA 클리핑 키트를 사용하여 세척하였다. DNA를 추출하고 시퀀싱하였다. 최종 렌티바이러스 벡터는 도 3b에 도시되어 있다.
- [0649] 벡터 작제물 시험
- [0650] 최소 네프린 프로모터를 사용하여 시험관내 세포 모델에서 GFP를 발현시켜 효능 및 발세포-특이성을 확인하였다.
- [0651] pLenti GFP 블라스트 네프린 프로모터 작제물(전장, 미디 및 미니)을 사용하여 48h 동안 HEK293T 세포를 형질감염시켜 바이러스를 제조하였고, 이를 추가로 사용하여 GFP-태깅된 FL NPHS1, 미디 또는 미니 프로모터를 안정적으로 발현하는 인간 조건부 불멸화 발세포를 생성하였다.
- [0652] 조건부 불멸화된 인간 발세포(ci 발세포)를 렌티바이러스 벡터로 형질감염시켜 최소 프로모터가 GFP 발현을 유도할 수 있는지의 여부를 결정하였다. 미디 및 미니 네프린 프로모터 둘 모두는 GFP 발현을 유도하는 것으로 나타났다. 도 4는 미니 네프린 프로모터로부터의 GFP 발현을 나타내는 대표적인 형광 현미경 이미지를 보여준다. 도 5는 미니 네프린 프로모터로부터의 GFP 발현을 나타내는 대표적인 웨스턴 블롯을 나타낸다. 이러한 결과는 미니 네프린 프로모터가 발세포에서 전이유전자 발현을 유도할 수 있음을 보여준다.
- [0653] 렌티바이러스 벡터는 또한 인간 사구체 세포를 형질도입하는데 사용되었다. ci 발세포 및 사구체 내피 세포는 미니 네프린 리포터에 커플링된 GFP를 포함하는 렌티바이러스로 형질도입하였다. 도 6a-c는 Novocyte Analyser를 사용하여 조건부 불멸화된 인간 발세포(LY) 및 사구체 내피 세포(GEnC)의 모든 살아있는 싱글렛의 중앙값 GFP 형광(AFU)을 나타내는 FACS 분석을 보여준다. 형질도입되지 않은 세포(세포 대조군)를 전장 인간 네프린 프로모터(hNPHS1.GFP) 또는 미니 인간 네프린 프로모터(265.GFP)에 의해 제어되는 GFP 발현 카세트를 보유하는 렌티바이러스 작제물로 형질도입된 세포와 비교하였다. 모든 세포를 10일 동안 분화시키고, 트립신 처리하고(100 uL), PBS, 2% FBS, 1:1000 DRAQ7(150uL)에 희석하였다. 데이터 및 오차 막대는 3개의 기술적 반복(100 uL, >2500개 세포) ±SEM을 나타낸다. 이러한 결과는 사구체 내피 세포와 비교할 때 최소 네프린 프로모터에 대한 발세포 특이성을 나타낸다.

- [0654] 실시예 2 - COL4A3, COL4A4 및 COL4A5에 커플링된 최소 네프린 프로모터의 설계, 작제 및 시험
- [0655] AAV 작제물의 설계 및 작제
- [0656] 미니 네프린 프로모터("265")에 커플링된 COL4A3, COL4A4 및 COL4A5를 포함하는 하기 AAV 전달 플라스미드를 설계하고 작제하였다:
- [0657] · pAAV.265.Col14a3.3flag.sv40, 미니 네프린 프로모터에 커플링된 COL4A3을 포함하는 AAV 플라스미드 (도 7a 참조).
- [0658] · pAAV.265.Col14a4.3flag.sv40, 미니 네프린 프로모터에 커플링된 COL4A4을 포함하는 AAV 플라스미드 (도 7b 참조).
- [0659] · pAAV.265.Col14a5.3flag.sv40, 미니 네프린 프로모터에 커플링된 COL4A5을 포함하는 AAV 플라스미드 (도 7c 참조).
- [0660] 플라스미드의 동일성을 확인하기 위해 SmaI 분해를 수행하였다(도 7d-e 참조). 이는 265bp 미니 네프린 프로모터 및 SV40 폴리A 꼬리를 갖는 AAV로의 COL4 a3, a4 및 a5의 성공적인 클로닝을 보여준다.
- [0661] AAV 작제물의 시험
- [0662] 하기 AAV 바이러스 벡터를 표준 방법을 사용하여 제조하였다:
- [0663] · LK03 혈청형을 갖는 AAV.COL4A3.nephr in265.Sv40
- [0664] · LK03 혈청형을 갖는 AAV.COL4A5.nephr in265.Sv40
- [0665] · 2/9 혈청형을 갖는 AAV.COL4A5.nephr in265.Sv40
- [0666] 도 8a는 항-FLAG 항체로 풀다운된 인간 분화된 ci발세포에서 전장 FLAG-태깅된 Co14a3(LK03) 또는 Co14a5(LK03)의 면역침전 실험을 보여준다. 항-FLAG 항체는 Co14a3 및 Co14a5 둘 모두를 침전시켰다. 인간-FLAG IgG를 대조군으로 사용하였다.
- [0667] 도 8b는 인간 또는 마우스 분화된 ci발세포에서 Co14a3(LK03 캡시드 혈청형), Co14a5(LK03) 및 Co14a5(2/9 캡시드 혈청형)의 발현 수준을 나타내는 단백질 용해물의 웨스턴 블롯을 보여준다. 감염되지 않은 인간 및 마우스 Ci발세포를 대조군으로 사용하였다.
- [0668] 도 8c는 F-액틴을 갖는 인간 야생형 Ci발세포/Co14a5 3xFlag AAV Ci발세포에서 형질도입된 Co14a5의 면역형광 염색을 나타내는 공초점 이미지를 보여준다. Co14a5는 야생형 대응물과 비교하여 Co14a5 3xFlag AAV 바이러스로 감염된 인간 분화된 발세포에서 세포질 수준으로 존재한다.
- [0669] 이러한 결과는 본 발명자들이 미니 네프린 프로모터에 커플링된 COL4a3 및 COL4a5 전장을 갖는 인간 발세포를 성공적으로 형질도입하였고, 미니 네프린 프로모터가 인간 발세포에서 전장 COL4a3 또는 COL4a5의 발현을 유도한다는 것을 보여준다.
- [0670] **실시예 3 - HEK 세포에서 CFH, CFI 및 CFHL1의 발현**
- [0671] 재료 및 방법
- [0672] 10% FBS가 보충된 DMEM에서 성장된 60-80% 컨플루언트 293T 인간 배아 신장 세포를 pHelper(HGTI1), 2개의 pAAV Rep-Cap(LK03 또는 AAV9) 중 하나 및 1) SEQ ID NO: 47의 265bp 최소 네프린 프로모터 하에 있는 CFH(pAAV-265-CFH), 2) SEQ ID NO: 14의 전장 최소 네프린 프로모터 하에 있는 CFI(pAAV-FL-CFI) 또는 3) SEQ ID NO: 14의 전장 최소 네프린 프로모터하의 CFHL1(pAAV-FL-CFHL1)를 함유하는 3개의 ITR-발현 플라스미드 중 하나로 삼중 형질감염시켰다. 모든 작제물은 MYC 및 FLAG로 태그하였다.
- [0673] 형질감염은 폴리에틸렌아민(PEI)의 존재 하에 무혈청 배지에서 150mm 배양 접시에서 수행되었다. 형질감염 다음 날 배지를 FBS를 갖는 DMEM으로 교체하였다. 형질감염 후 4일에, 배지 및 세포를 수집하고 별도로 처리하였다. 배지를 동결시키고 -80°C에서 저장하였다. 세포를 프로테이나제 억제제가 보충된 RIPA 완충제로 용해시키고 -80°C에서 저장하였다.
- [0674] 세포 용해물에서 단백질 농도를 Pierce BCA 단백질 검정을 사용하여 측정하고, 각 샘플로부터의 10 ug의 총 단백질을 4-15% 폴리아크릴아미드 Tris-Glycine 겔에 로딩하였다. 각 샘플로부터 총 2.6ul의 배지를 겔에 로딩하

였다. IBlot2 건조 블롯팅 시스템을 사용하여 단백질을 니트로셀룰로스 막으로 옮겼다. 하기 일차 항체를 단백질 검출에 사용하였다: 항-인자 H(Abcam, cat. ab124769), 항-인자 I(Abcam, ab278524), 항-MYC-태그(CST, cat.2276S), 항-FLAG-태그 (CST, cat. 14793S) 및 항-GAPDH (Millipore, cat. MAB374).

[0675] 결과

[0676] 인자 H는 세포 용해물 및 CFH 발현 플라스미드로 형질감염된 293T HEK 세포로부터의 배지 둘 모두에서 발현되었다(도 9, 라인 2, 3, 8, 9). 인자 H-특이적 항체를 사용하여 발현을 검출하였다. 인자 H는 형질감염되지 않은 세포 또는 CFI 또는 CFHL1 발현 플라스미드로 형질감염된 세포의 세포 용해물 또는 배지에서 검출되지 않았다(도 9, 라인 1, 4-7, 10-13). 이러한 결과는 265bp 최소 네프린 프로모터가 표적 세포에서 전이유전자의 발현을 허용한다는 것을 입증한다.

[0677] **실시예 4 - AAV2/9 265-CFH로의 인자 H 돌연변이된 발세포의 형질도입 또는 CFH를 인코딩하는 플라스미드로의 형질감염**

[0678] AAV2/9 265-CFH로의 인자 H 돌연변이된 발세포의 형질도입

[0679] 10% FBS 및 1% ITS가 보충된 RPMI에서 성장한 내인성 인자 H(Muehlig et al, 2020)에서 돌연변이를 갖는 조건부 불멸화 인간 발세포를 6-웰 배양 플레이트에 시딩하고 70-80% 컨플루언시까지 33°C에서 성장시켰다. 세포를 265bp 최소 네프린 프로모터(SEQ ID NO: 47)의 제어 하에 CFH 전이유전자를 함유하는 AAV2/9와 함께 인큐베이션하였다. 바이러스 없이 인큐베이션된 세포를 비-형질도입(NT) 대조군으로 사용하였다. 형질도입 후 동일 날에 세포를 37°C의 비-허용 온도로 옮겨, 세포 분화 및 유전자전이가 가능하게 하였다. 형질도입 후 다음 날에 배지를 2회 교체하였다. 형질도입 후 10일에, 세포 배지를 수집하고, 항-인자 H 항체(Abcam, cat. ab252359)를 사용하여 ELISA에 의해 인간 인자 H의 농도를 측정하였다.

[0680] 인자 H에 돌연변이를 갖는 인간 발세포를 265 bp 최소 네프린 프로모터의 제어 하에 CFH 전이유전자를 함유하는 AAV2/9 바이러스로 형질도입하고, 이는 형질도입되지 않은 세포보다 배양 배지에서 더 높은 농도의 인간 인자 H를 입증하였다(도 10a 및 10b). 이는 AAV에 의해 265bp 최소 네프린 프로모터의 제어 하에 전달된 인자 H 전이유전자가 인간 발세포에서 발현될 수 있음을 입증한다.

[0681] CFH를 인코딩하는 플라스미드로 인자 H 돌연변이된 발세포의 형질감염

[0682] 10% FBS 및 1% ITS가 보충된 RPMI에서 성장한 내인성 인자 H(Muehlig et al, 2020)에서 돌연변이를 갖는 조건부 불멸화 인간 발세포를 6-웰 배양 플레이트에 시딩하고 70-80% 컨플루언시까지 33°C에서 성장시켰다. 플라스미드 형질감염을 위해, 세포를 폴리에틸렌민 존재 하에 무혈청 배지에서 1.5ug의 발현 플라스미드와 함께 인큐베이션하였다. 플라스미드가 첨가되지 않은 세포를 비-형질감염(NT) 대조군으로 사용하였다. 배지를 다음날 FBS를 함유하는 배지로 교체하였다. 형질감염 후 3일에, 배지를 수집하고 ELISA(Abcam, cat. ab252359)에 의해 분석하였다.

[0683] 265bp 최소 네프린 프로모터의 제어 하에 CFH 전이유전자를 발현하는 플라스미드로 형질감염된 발세포는 형질감염되지 않은 대조군보다 더 높은 농도의 인간 인자 H를 나타내었다(도 10c).

[0684] **실시예 5 - WT 마우스의 신장에서 CFH의 발현**

[0685] 재료 및 방법

[0686] 100 µl의 AAV2/9 유전자 요법 생성물(pAAV.NPHS1(265).hCFH.WPRE.bGH) 또는 식염수를 IV 꼬리 정맥 주사에 의해 야생형 C57BL6 마우스에 투여하였다. AAV 발현된 태깅된 야생형 인간 CFH 전이유전자는 265bp 최소 네프린 프로모터(SEQ ID NO: 47)의 제어 하에 있었다. AAV를 수확하고 3일 후에 초원심분리에 의해 정제하고 PBS에서 적정하였다(약 1.5x10e13/ml). 모든 동물은 21일에 연구를 완료하고 도태시켰다. 신장을 액체 질소에서 급속 동결시키고, 제조사의 프로토콜에 따라 RNeasy Micro Kit(Qiagen Cat. No./ID: 74004)에 의한 RNA 추출에 사용하였다. 이어서, qPCR 분석 전에 고용량 RNA-to-cDNA™ 키트(4387406)를 사용하여 RNA를 cDNA로 전환시켰다.

[0687] pAAV\_CFH 주사된 마우스의 신장으로부터의 DNA 샘플에 대해 정량적 qPCR을 수행하였다. SYBR 녹색 시약 및 수동 ROX 방법을 사용한 표준 곡선 qPCR을 사용하여 신장 DNA 샘플에서 ITR 존재를 검출하였다. DNA의 pg 당 바이러스 계놈은 알려진 양의 ITR 앰플리콘의 표준 곡선을 사용하여 계산되었다. 마지막으로, 이배체 마우스 세포가 6pg의 DNA를 갖는다는 가정에 기초하여 세포 당 바이러스 계놈을 계산하였다.

[0688] 항-네프린 항체(PROGEN) 및 항-CFH 항체(ab124767)를 사용하여 5 마이크론 두께의 동결된 신장 섹션에 대해 면역

형광 염색을 수행하였다.

- [0689] 조직을 OCT 화합물(VWR, 카탈로그 번호 361603E)에 포매시키고 액체 질소에서 급속 동결시켰다. 저온유지장치 (Thermo, Cryostar NX270)를 사용하여 10uM 섹션을 절단하였다. 조직 섹션을 실온에서 20분 동안 4% 파라포름알데하이드로 고정하고, 실온에서 15분 동안 0.3% Triton-X로 투과시키고, 실온에서 30분 동안 5% BSA로 차단하였다.
- [0690] 일차 항체(인자 H에 대한 래빗 모노클로날 [EPR6226], Abcam 카탈로그 번호 ab124769 및 Nephtrin (NPHS1) (1243-1256) Guinea Pig Polyclonal Antibody, Origene 카탈로그 번호 BP5030)를 5% BSA에서 희석하였다. (항-인자 H의 경우 1:100 및 항-네프린의 경우 1:300). 인큐베이션을 실온에서 1시간 동안 수행하였다.
- [0691] 이차 항체(염소 항-토끼 IgG(H+L) 이차 항체, Alexa Fluor Plus 488, Invitrogen, 카탈로그 번호 A32731 및 염소 항-기니피그 IgG(H+L) Alexa Fluor 568, Invitrogen, 카탈로그 번호 A-11075)를 PBS에서 1:500으로 희석하고 인큐베이션을 실온에서 30분 동안 수행하였다.
- [0692] 슬라이드를 Fluoromount-G, 슬라이드 장착 매질(Southern Biotech, 카탈로그 번호 0100-01)에 장착하고 Leica DM750 형광 현미경에서 20X 배율로 영상화하였다.
- [0693] 결과
- [0694] 265bp 최소 네프린 프로모터 하에 인간 CFH를 함유하는 AAV로의 야생형 마우스의 주사는 식염수 대조군이 주사된 야생형 마우스와 비교하여 AAV에 의한 신장의 감염 및 신장에서 바이러스 및 인간 CFH의 더 큰 발현을 초래한다(도 11a-c). 면역형광 염색은 마우스 사구체에서 네프린 및 CFH의 공동-국소화를 입증한다(도 11d).
- [0695] **실시예 7 - 신장 세포에서 265bp v 818bp v FL 최소 네프린 프로모터(플라스미드 형질감염)를 사용한 GFP의 발현**
- [0696] (i) SEQ ID NO: 14의 전장(FL) 최소 네프린 프로모터(PS0281); (ii) SEQ ID NO: 59의 818bp 최소 네프린 프로모터(PS0301); 또는 (iii) SEQ ID NO: 47의 265bp 최소 네프린 프로모터(P20282)의 제어 하에 eGFP를 포함하는 플라스미드를 작제하였다.
- [0697] 발세포, 사구체 내피 세포(GENC) 및 근위 튜브 상피 세포(PTEC)를 6웰 플레이트의 웰 당 1.5e5 세포의 밀도로 시딩하고 밤새 부착되게 하였다. HEK293T 세포의 경우, 웰 당 5e5 세포를 사용하였다. 다음날 150 uL의 DMEM 중 3 ug의 플라스미드 DNA를 9uL PEI(또한 150uL DMEM 중)와 혼합하고, 실온에서 15분 동안 인큐베이션하고, 48시간 후에 eGFP 발현에 대해 유세포 측정기에 의해 분석하였다. 형질감염되지 않은 세포를 음성 및 양성 집단의 게이팅을 위한 음성 대조군으로 사용하였다.
- [0698] 도 12의 결과는 3개 모두의 프로모터가 시험된 모든 세포 유형에서 eGFP 발현을 유도할 수 있음을 보여준다. 이러한 결과는 전장 네프린 프로모터를 818bp 또는 265bp로 최소화하는 것이 네프린 프로모터 서열의 큰 영역을 제거함에도 불구하고 표적 세포에서 전이유전자의 발현을 여전히 허용한다는 것을 입증한다.
- [0699] **실시예 8 - 신장 세포에서 265bp v FL 최소 네프린 프로모터(AAV 형질도입)를 사용한 GFP의 발현**
- [0700] 재료 및 방법
- [0701] AAV 생성: 10% FBS가 보충된 DMEM에서 성장한 60-80% 컨플루언트 293T 인간 배아 신장 세포를 pHelper(HGTI1, PS0150), Rep-Cap 플라스미드(LK03, PS0240) 및 eGFP 발현을 유도하는 전장(FL) 최소 네프린 프로모터 또는 265bp 최소 네프린 프로모터(각각 PS0281 및 PS0282)를 함유하는 2개의 ITR-발현 플라스미드 중 하나로 삼중 형질감염시켰다. 형질감염은 폴리에틸렌이민(PEI)의 존재 하에 150mm 배양 접시에서 수행되었다. 형질감염 다음날 배지를 FBS를 갖는 DMEM으로 교체하였다. 형질감염 후 4일에, 배지 및 세포를 수집하고 별도로 처리하였다. PEG를 여과된 상청액에 첨가하여 바이러스 입자를 침전시키고, 이를 후속하여 1500g에서 20분 동안 원심분리하였다. 세포 펠렛을 TD 완충제에 재현탁시키고 -80도에서 동결시킨 다음, 37도에서 해동시키고 불텍싱하였다. 동결-해동 사이클을 5회 반복하였다. 이어서, PEG 처리된 상청액으로부터의 펠렛을 용해된 세포에 첨가하고, 벤조나제 및 소듐 데옥시콜레이트를 추가로 첨가하고 37도에서 30분 동안 인큐베이션하였다. 이어서, 샘플을 원심분리하고, 상청액을 여과한 후, 아이오딕사놀 구매 초원심분리를 통해 정제하였다.
- [0702] AAV QC: AAV 샘플을 qPCR을 통해 적정하여 바이러스 상청액의 mL 당 바이러스 게놈의 수를 결정하였다. eGFP 서열을 표적화하도록 설계된 프라이머를 사용하였다. 순도를 결정하기 위해 샘플에 대해 SDS-PAGE 분석을 수행하였다. 알칼리성 겔 전기영동을 사용하여 완전한 전이유전자 카세트의 바이러스 입자로의 혼입을 입증하였다.

- [0703] AAV 형질도입: 세포를 1.5e5 세포/웰의 밀도로 6-웰 플레이트의 웰에 시딩하고 밤새 부착되게 하였다. 다음날 바이러스 입자를 세포 당 5e5 입자의 감염 다중도(MOI)로 첨가하였다. 세포를 37도에 놓아 분화를 개시하고 10 일 동안 배양하였다. 10일 후, 세포를 유세포 측정에 의해 eGFP 발현에 대해 분석하였다.
- [0704] 결과
- [0705] 도 13의 결과는 두 프로모터 모두가 시험된 모든 세포 유형에서 eGFP 발현을 유도할 수 있음을 보여준다. 265bp 최소 네프린 프로모터는 시험된 각각의 세포에서 더 높은 eGFP 발현 수준을 나타내었다.
- [0706] 이러한 데이터는 265bp 최소 네프린 프로모터가 표적 세포에서 eGFP 발현을 유도할 수 있음을 다시 입증한다. 또한, 이 프로모터는 유전자 요법에 사용하기 위해 AAV 벡터에 혼입될 수 있다. 전장 최소 네프린 프로모터 대신에 265bp 최소 네프린 프로모터의 사용은 AAV 목적을 위해 전이유전자 카세트에 더 많은 유전자 물질이 혼입되게 할 수 있다.
- [0707] **실시예 9 - HEK에서 265bp v FL 최소 네프린 프로모터(플라스미드 형질감염)를 사용한 인간 포도신의 발현**
- [0708] (i) SEQ ID NO: 14의 전장(FL) 최소 네프린 프로모터(FL.NPSH1-Podocin-HA); 또는 (ii) SEQ ID NO: 47의 265bp 최소 네프린 프로모터(265.hNPSH1-Podocin-HA)의 제어 하에 HA-태깅된 포도신을 포함하는 플라스미드를 작제하였다.
- [0709] 10% FBS, 1% 페니실린/스트렙토마이신 및 1% 소듐 피루베이트가 보충된 DMEM에서 성장한 293T 인간 배아 신장 세포를 6-웰 배양 플레이트에 5e5의 밀도로 시딩하였다. 세포가 70-80% 컨플루언시에 도달한 경우 다음 날 형질 감염을 수행하였다. 무혈청 배지에서 3ug의 발현 플라스미드를 폴리에틸렌이민(PEI)과 1:3의 비율로 복합화시킴으로써 형질감염 시약 혼합물을 제조하였다. 형질감염되지 않은 대조군은 임의의 발현 플라스미드 없이 동일한 비율로 PEI를 함유하는 형질감염 믹스를 수용하였다. 배지가 완전 배지로 교체된 경우 다음날까지 세포를 37°C, 5% CO2에서 형질감염 믹스와 함께 인큐베이션하였다. 형질감염 후 2일에 세포를 수확하고, 항-HA 항체로 HA-태그에 대해 염색함으로써 포도신의 발현에 대해 FAC에 의해 분석하였다.
- [0710] 도 14a의 결과는 두 프로모터 모두가 HEK 세포에서 인간 포도신-HA 발현을 유도할 수 있음을 보여준다. 이는 전장 최소 네프린 프로모터를 265bp로 최소화하는 것이 프로모터 서열의 대부분을 제거함에도 불구하고 여전히 표적 세포에서 전이유전자의 발현을 허용한다는 것을 입증한다.
- [0711] **실시예 10 - 인간 발세포에서 265bp v FL 최소 네프린 프로모터(AAV 형질도입)를 사용한 인간 포도신의 발현**
- [0712] 재료 및 방법
- [0713] AAV 생성: 10% FBS, 1% 페니실린/스트렙토마이신 및 1% 소듐 피루베이트가 보충된 DMEM에서 성장한 60-80% 컨플루언트 293T 인간 배아 신장 세포를 pHelper(HGT11, PS0150), pAAV Rep-Cap (LK03, PS0240), 및 포도신-HA 발현을 유도하는 전장(FL) 최소 네프린 프로모터 또는 265bp 최소 네프린 프로모터를 함유하는 2개의 ITR-발현 플라스미드(각각 FL.NPSH1-Podocin-HA 및 265.hNPSH1-Podocin-HA) 중 하나로 삼중 형질감염시켰다. 형질감염은 1:2 비율의 폴리에틸렌이민(PEI)의 존재 하에 150mm 배양 접시에서 수행되었다. 형질감염 후 다음 날 배지를 FBS를 갖는 DMEM으로 교체하였다. 형질감염 후 4일에, 배지 및 세포를 수집하고 별도로 처리하였다. PEG를 여과된 상청액에 첨가하여 바이러스 입자를 침전시키고, 이를 후속하여 1500g에서 20분 동안 원심분리하였다. 세포 펠렛을 TD 완충제에 재현탁시키고 -80도에서 동결시킨 다음, 37도에서 해동시키고 볼텍싱하였다. 동결-해동 사이클을 5회 반복하였다. 이어서, PEG 처리된 상청액으로부터의 펠렛을 용해된 세포에 첨가하였다. 벤조나제 및 소듐 테옥시콜레이트를 추가로 첨가하고 37도에서 30분 동안 인큐베이션하였다. 이어서, 샘플을 원심분리하고, 상청액을 여과한 후, 아이오덕사놀 구배 초원심분리를 통해 정제하였다.
- [0714] AAV QC: AAV 샘플을 qPCR을 통해 적정하여 바이러스 상청액의 ml 당 바이러스 게놈의 수를 결정하였다. 포도신 서열을 표적화하도록 설계된 프라이머를 사용하였다.
- [0715] AAV 형질도입: 세포를 0.5e5 세포/웰의 밀도로 12-웰 플레이트의 웰에 시딩하고 밤새 부착되게 하였다. 다음날 바이러스 입자를 세포 당 5e5 입자의 감염 다중도(MOI)로 첨가하였다. 세포를 37도에 놓아 분화를 개시하고 9일 동안 배양하였다. 9일 후 세포를 수확하고 포도신 발현의 척도로서 유세포 측정에 의한 HA 발현에 대해 항-HA 항체로 염색하였다.
- [0716] 결과
- [0717] HEK 세포에서 플라스미드 형질감염 데이터로 입증된 바와 같이(도 11a), 도 14b에 제시된 결과는 두 프로모터

모두가 인간 발세포에서 전이유전자의 발현을 유도할 수 있음을 추가로 확인시켜 준다. 265bp 최소 네프린 프로모터는 인간 발세포에서 더 높은 포도신 발현 수준을 나타내었다.

[0718] **실시예 11 - AAV 형질도입을 통한 265bp v FL 프로모터를 사용한 GFP의 생체내 발현**

[0719] 265- 및 FL- 네프린 프로모터를 WT 마우스의 생체내에서 평가하였다(C57/BI6). AAV9(PS0241)를 LK03 대신 사용한 것을 제외하고는 실시예 8에 기재된 바와 같은 AAV 입자를 마우스당  $\sim 1 \times 10^{13}$  AAV2/9의 꼬리 정맥 주사를 통해 투여하였다(처리 그룹당 2x 마우스 및 1x 미처리 마우스). 신장의 조직 수확은 주사 후 4주에 발생하였다. qPCR 분석을 신장에서 수행하고, GFP 복사체 수를 18S(하우스키핑 유전자)로 표준화하였다. 18S rRNA는 실시간 PCR을 위한 신뢰할 수 있는 표준화 유전자이다.

[0720] 도 15의 결과는 265bp 및 전장 최소 네프린 프로모터 둘 모두를 사용하여 신장에서 검출가능한 GFP 발현이 있음을 보여준다. 신장에서의 발현은 265bp 최소 네프린 프로모터를 사용하여 더 높았다.

[0721] 상기 명세서에 언급된 모든 간행물은 본원에 참조로 포함된다. 본 발명의 개시된 방법, 세포, 조성물 및 용도의 다양한 변경 및 변형은 본 발명의 범위 및 사상을 벗어나지 않으면서 당업자에게 명백할 것이다. 본 발명이 특정 바람직한 구체예와 관련하여 개시되었지만, 청구된 바와 같은 본 발명은 이러한 특정 구체예로 과도하게 제한되어서는 안된다는 것이 이해되어야 한다. 실제로, 당업자에게 명백한 본 발명을 수행하기 위한 개시된 모드의 다양한 변형은 하기 청구범위의 범위 내에 있는 것으로 의도된다.

[0722] **구체예**

[0723] 본 발명의 다양한 바람직한 특징 및 구체예는 이제 하기 번호가 매겨진 단락을 참조하여 설명될 것이다.

[0724] 1. (i) SEQ ID NO: 4에 대해 적어도 70%의 동일성을 갖는 뉴클레오티드 서열을 포함하는 프로모터로서, 상기 프로모터는 약 1.1 kb 이하의 길이를 갖는, 프로모터.

[0725] 2. SEQ ID NO: 1에 따른 뉴클레오티드 서열에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성된 프로모터로서, 여기서:

[0726] (i) SEQ ID NO: 1의 위치 1 내지 위치 n1 (여기서 n1은 100 내지 430의 정수임)이 결실되고/거나;

[0727] (ii) SEQ ID NO: 1의 위치 n2 내지 위치 n3 (여기서  $n3 \geq n2$ 이고, n2는 508 내지 1061의 정수이고, n3은 508 내지 1061의 정수임)이 결실되는, 프로모터.

[0728] 3. 단락 2에 있어서, 프로모터가 (i) SEQ ID NO: 4에 대해 적어도 70%의 동일성을 갖는 뉴클레오티드 서열을 포함하고/거나 프로모터는 약 1.1 kb 이하의 길이를 갖는, 프로모터.

[0729] 4. 단락 1 내지 3 중 어느 한 단락에 있어서, 프로모터가 발세포-특이적 프로모터인, 프로모터.

[0730] 5. 단락 1 내지 4 중 어느 한 단락에 있어서, 프로모터가 (ii) SEQ ID NO: 5에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열 및/또는 SEQ ID NO: 7에 대해 적어도 70%의 동일성을 갖는 뉴클레오티드 서열을 추가로 포함하는, 프로모터.

[0731] 6. 단락 1 내지 5 중 어느 한 단락에 있어서, 프로모터가 (iii) SEQ ID NO: 8에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열 또는 이의 하나 이상의 단편을 추가로 포함하는, 프로모터.

[0732] 7. 단락 1 내지 6 중 어느 한 단락에 있어서, 프로모터가 5'에서 3'으로 (i) SEQ ID NO: 4에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열; (iii) 임의적으로 SEQ ID NO: 8에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열 또는 이의 하나 이상의 단편; 및 (ii) SEQ ID NO: 5에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열, SEQ ID NO: 6에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열 및/또는 SEQ ID NO: 7에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하는, 프로모터.

[0733] 8. 단락 1 내지 7 중 어느 한 단락에 있어서, 프로모터가 약 1.0 kb 이하, 약 0.9 kb 이하, 약 0.8 kb 이하, 약 0.7 kb 이하, 약 0.6 kb 이하, 약 0.5 kb 이하, 약 0.4 kb 이하 또는 약 0.3 kb 이하의 길이를 갖는, 프로모터.

[0734] 9. 단락 1 내지 8 중 어느 한 단락에 있어서, 프로모터가 0.265 kb 내지 1.0 kb, 0.265 kb 내지 0.9 kb, 0.265 kb 내지 0.8 kb, 0.265 kb 내지 0.7 kb, 0.265 kb 내지 0.6 kb, 0.265 kb 내지 0.5 kb, 0.265 kb 내지 0.4 kb 또는 0.265 kb 내지 0.3 kb의 길이를 갖는, 프로모터.

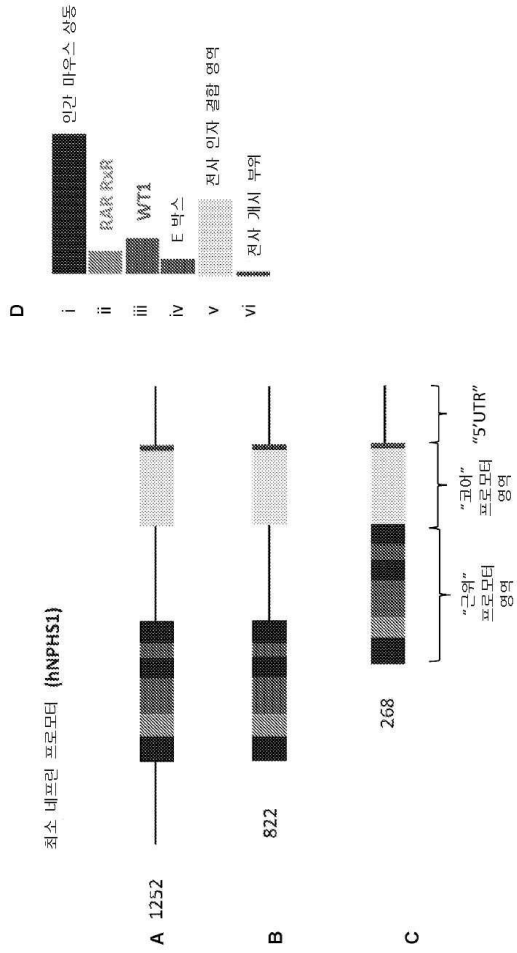
- [0735] 10. 단락 1 내지 9 중 어느 한 단락에 있어서, 프로모터가
- [0736] (a) 레티노산 수용체 결합 부위;
- [0737] (b) WT1 결합 부위;
- [0738] (c) 인헨서 박스;
- [0739] (d) 전사 인자 결합 영역; 및/또는
- [0740] (e) 전사 개시 부위를 포함하는, 프로모터.
- [0741] 11. 단락 10에 있어서, 레티노산 수용체 결합 부위가 SEQ ID NO: 10으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 10과 비교하여 1 또는 2개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.
- [0742] 12. 단락 10 또는 11에 있어서, WT1 결합 부위가 SEQ ID NO: 11로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 11과 비교하여 1, 2 또는 3개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.
- [0743] 13. 단락 10 내지 12 중 어느 한 단락에 있어서, 인헨서 박스가 SEQ ID NO: 12로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 12와 비교하여 1 또는 2개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.
- [0744] 14. 단락 1 또는 3 내지 13 중 어느 한 단락에 있어서, (i) SEQ ID NO: 4와 적어도 70% 동일성을 갖는 뉴클레오티드 서열이 (a) 레티노산 수용체 결합 부위; (b) WT1 결합 부위; 및 (c) 인헨서 박스 중 하나 이상을 포함하는, 프로모터.
- [0745] 15. 단락 1 또는 단락 3 내지 14 중 어느 한 단락에 있어서, 하기 뉴클레오티드 서열 중 하나 이상이 (i) SEQ ID NO: 4에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열에 존재하는, 프로모터:
- [0746] (a) SEQ ID NO: 4의 위치 7 내지 위치 13에 상응하는 위치에서의 GGGGTCA;
- [0747] (b) SEQ ID NO: 4의 위치 14 내지 위치 30에 상응하는 위치에서의 CGGAGGCTGGGGAGGCA; 및
- [0748] (c) SEQ ID NO: 4의 위치 49 내지 위치 53에 상응하는 위치에서의 ATGTG.
- [0749] 16. 단락 10 내지 15 중 어느 한 단락에 있어서, 전사 인자 결합 영역이 SEQ ID NO: 13으로 제시된 뉴클레오티드 서열 또는 SEQ ID NO: 13과 비교하여 1, 2, 3, 4 또는 5개의 치환, 결실 또는 삽입을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.
- [0750] 17. 단락 10 내지 16 중 어느 한 단락에 있어서, 전사 개시 부위가 "AG" 디뉴클레오티드를 포함하거나 이로 구성되는, 프로모터.
- [0751] 18. 단락 10 내지 17 중 어느 한 단락에 있어서, 전사 인자 결합 부위가 전사 개시 부위에 작동가능하게 연결되며, 임의적으로 전사 인자 결합 부위는 전사 개시 부위의 바로 업스트림에 있는, 프로모터.
- [0752] 19. 단락 1 내지 18 중 어느 한 단락에 있어서, 프로모터가 SEQ ID NO: 2 또는 SEQ ID NO: 3에 대해 적어도 70% 동일성을 갖는 뉴클레오티드 서열을 포함하거나 이로 구성되는, 프로모터.
- [0753] 20. SEQ ID NO: 2에 대해 적어도 70%의 동일성을 갖는 뉴클레오티드 서열로 구성된 프로모터.
- [0754] 21. SEQ ID NO: 3에 대해 적어도 70%의 동일성을 갖는 뉴클레오티드 서열로 구성된 프로모터.
- [0755] 22. 단락 1 내지 21 중 어느 한 단락에 따른 프로모터를 포함하는 폴리뉴클레오티드.
- [0756] 23. 단락 22에 있어서, 프로모터가 단백질 코딩 서열에 작동가능하게 연결되는, 폴리뉴클레오티드.
- [0757] 24. 단락 23에 있어서, 단백질 코딩 서열이 전사-후 조절 요소 및/또는 폴리아데닐화 서열과 같은 하나 이상의 추가 조절 요소에 작동가능하게 연결되는, 폴리뉴클레오티드.
- [0758] 25. 단락 22 내지 24 중 어느 한 단락에 따른 폴리뉴클레오티드를 포함하는 벡터.
- [0759] 26. 단락 25에 있어서, 벡터가 발세포를 형질도입할 수 있고, 임의적으로 벡터는 발세포를 특이적으로 형질도입

할 수 있는, 벡터.

- [0760] 27. 단락 25 또는 26에 있어서, 벡터가 바이러스 벡터, 예컨대, 아데노-관련 바이러스(AAV) 벡터, 렌티바이러스 벡터, 레트로바이러스 벡터, 아데노바이러스 벡터, 아데노-관련 바이러스 벡터, 단순 포진 바이러스 벡터, 알파 바이러스 벡터, 플라비바이러스 벡터, 랩도바이러스 벡터, 홍역 바이러스 벡터, 뉴캐슬병 바이러스 벡터, 폭스 바이러스 벡터 및 피코르나바이러스 벡터이고, 바람직하게는 벡터는 AAV 벡터인, 벡터.
- [0761] 28. 단락 27에 있어서, 바이러스 벡터가 바이러스 벡터 입자의 형태이고, 바람직하게는 바이러스 벡터는 AAV 벡터 입자의 형태인 벡터.
- [0762] 29. 단락 25 내지 28 중 어느 한 단락에 있어서, 벡터가 AAV3B, LK03, 또는 AAV9 캡시드 단백질에 의해 캡시드화된 AAV 벡터 입자의 형태인, 벡터.
- [0763] 30. 단락 22 내지 24 중 어느 한 단락에 따른 폴리뉴클레오티드 또는 단락 25 내지 29 중 어느 한 단락에 따른 벡터를 포함하는 세포.
- [0764] 31. 단락 22 내지 24 중 어느 한 단락에 따른 폴리뉴클레오티드, 단락 25 내지 29 중 어느 한 단락에 따른 벡터 또는 단락 30에 따른 세포를 포함하는 약학적 조성물.
- [0765] 32. 의약에 사용하기 위한 단락 22 내지 24 중 어느 한 단락에 따른 폴리뉴클레오티드, 단락 25 내지 29 중 어느 한 단락에 따른 벡터, 또는 단락 30에 따른 세포.
- [0766] 33. 코딩 서열의 발현을 유도하기 위한 단락 1 내지 21 중 어느 한 단락에 따른 프로모터의 용도.
- [0767] 34. 단락 32에 있어서, 발현이 발세포-특이적인, 용도.

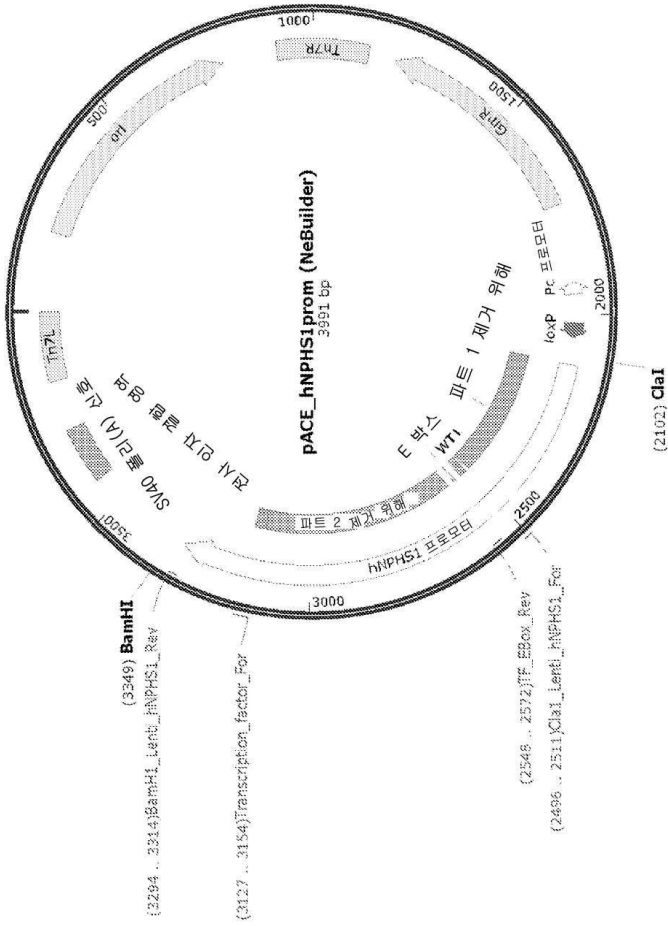
도면

도면1



도면2a

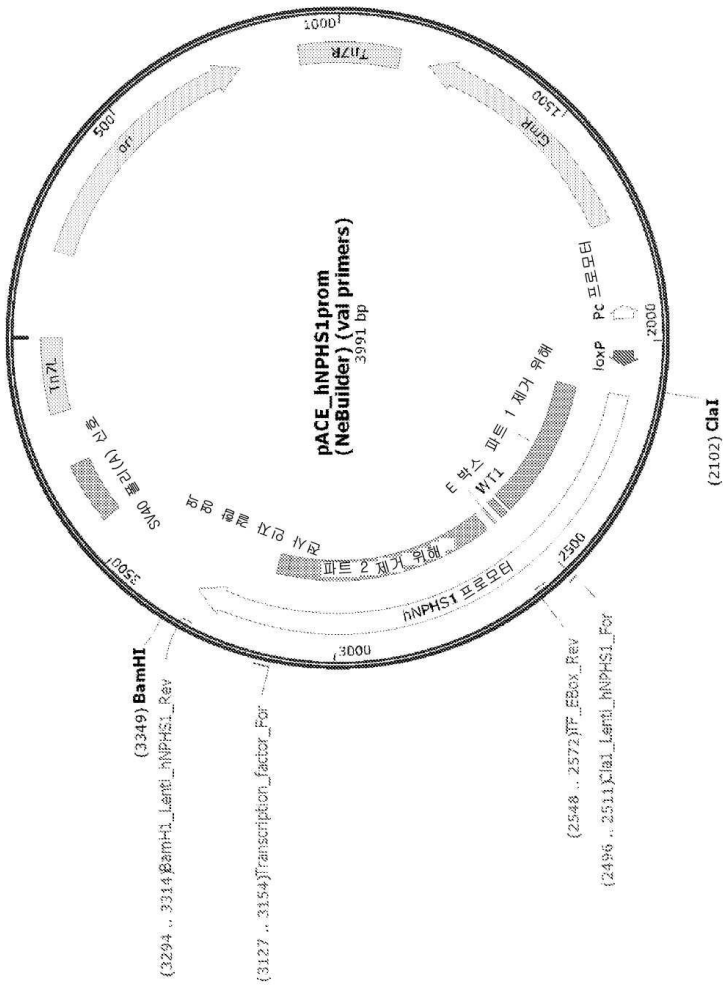
A





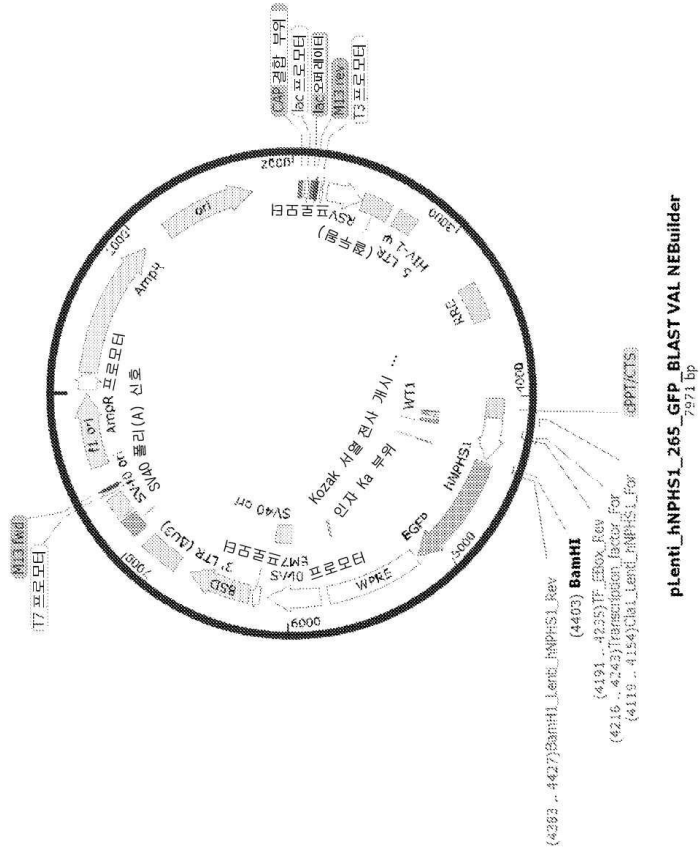
도면3a

A

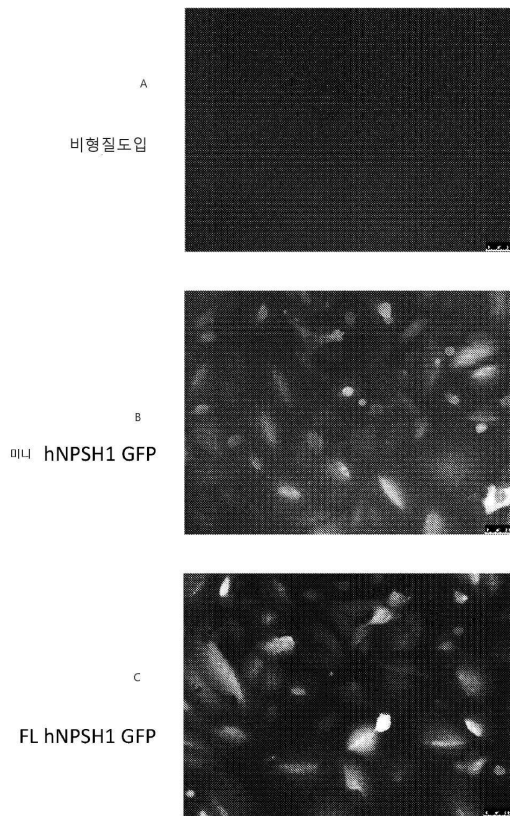


도면3b

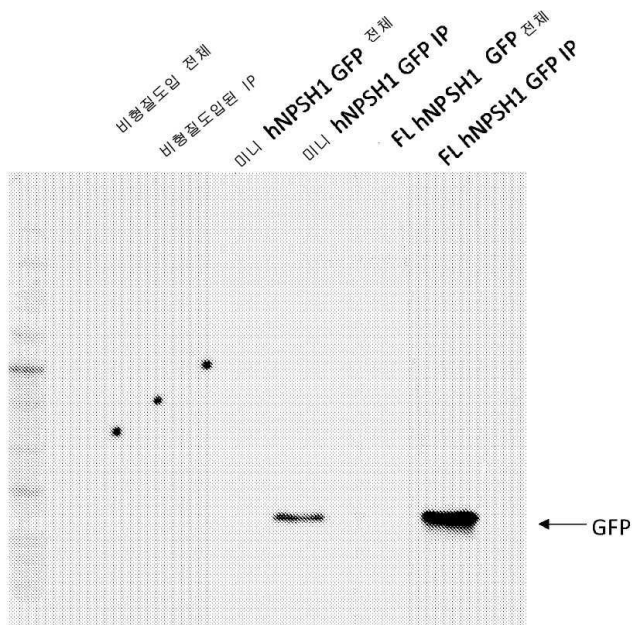
B



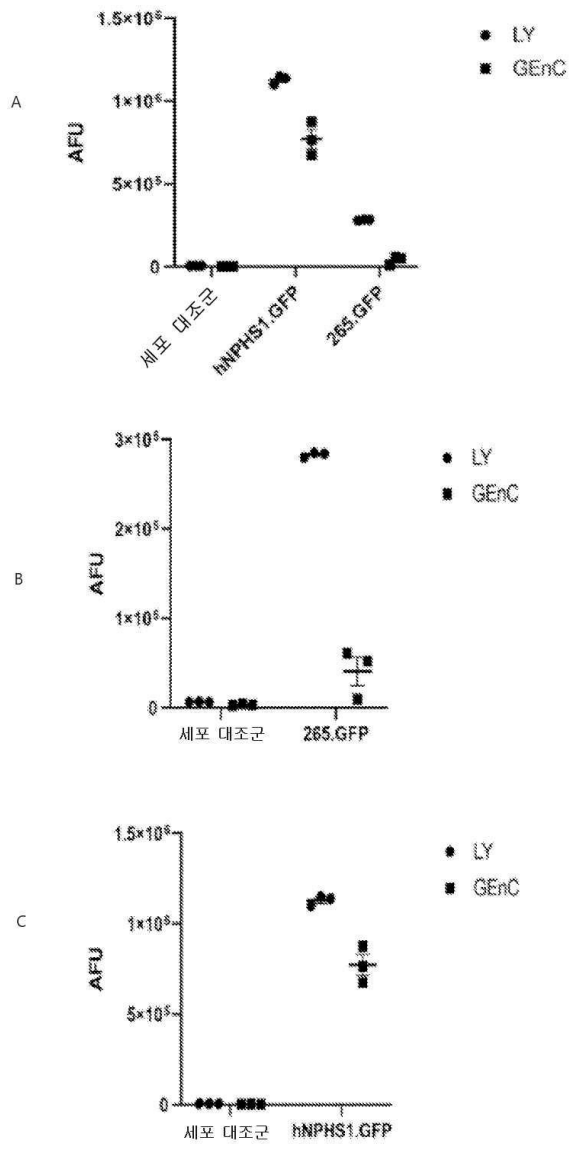
도면4



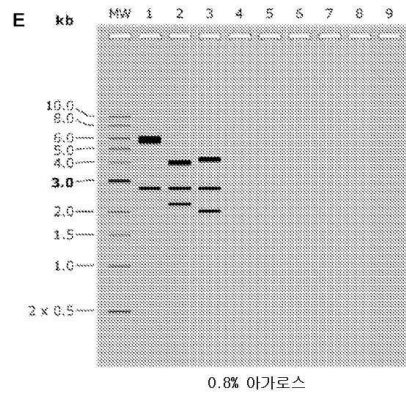
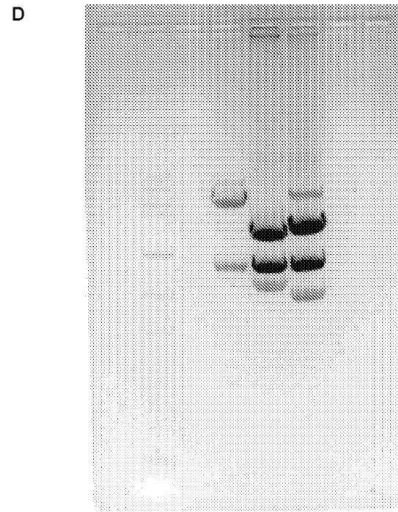
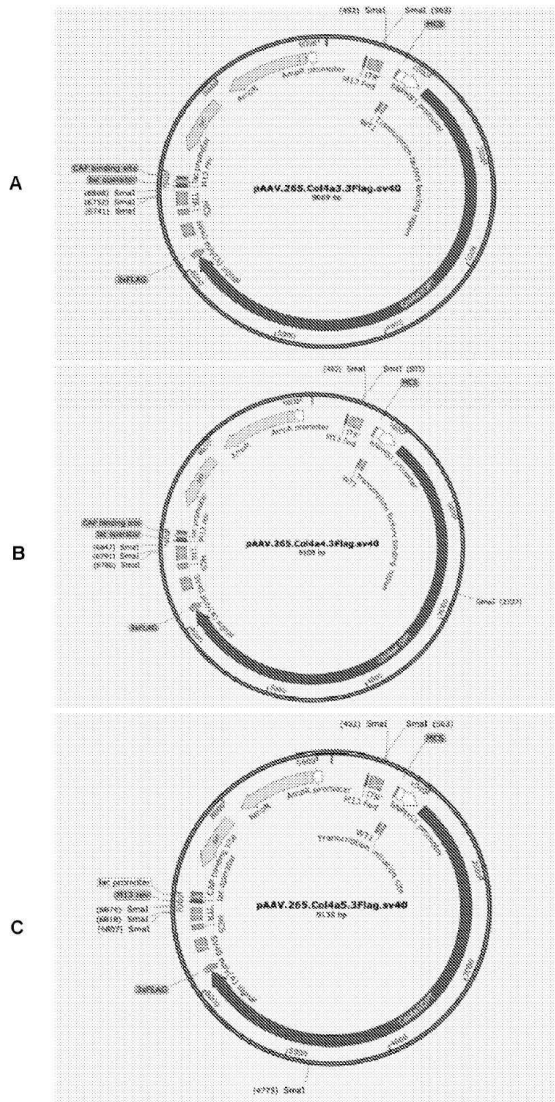
도면5



도면6

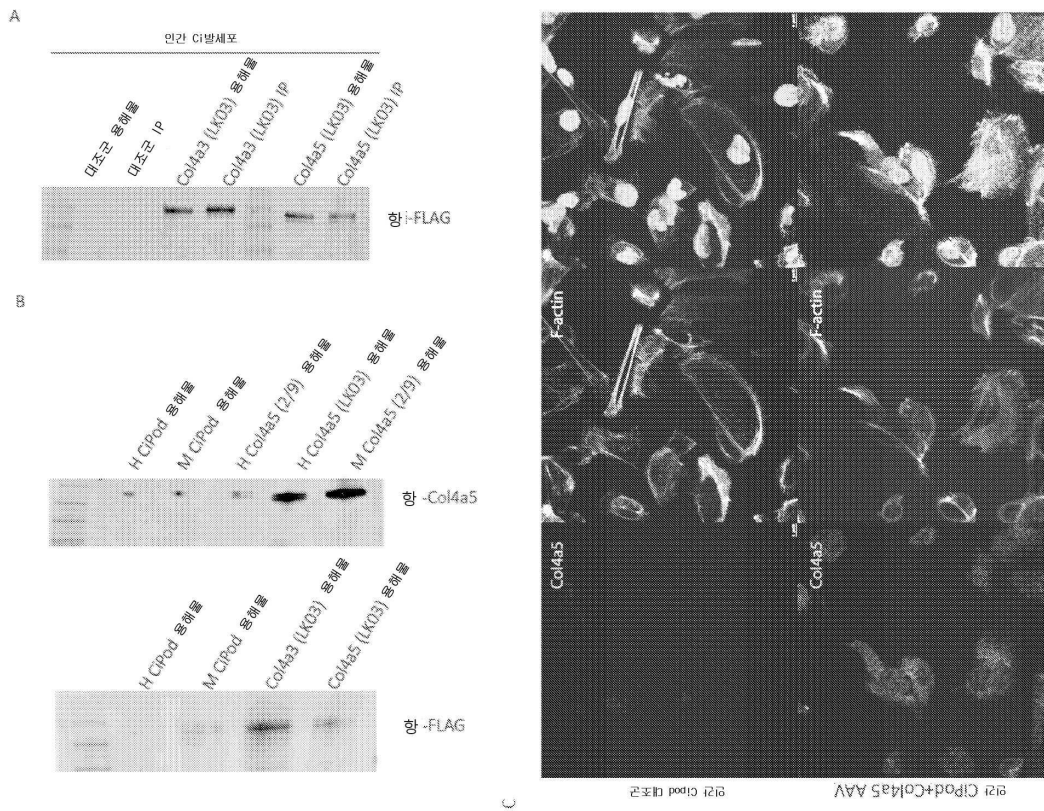


도면7

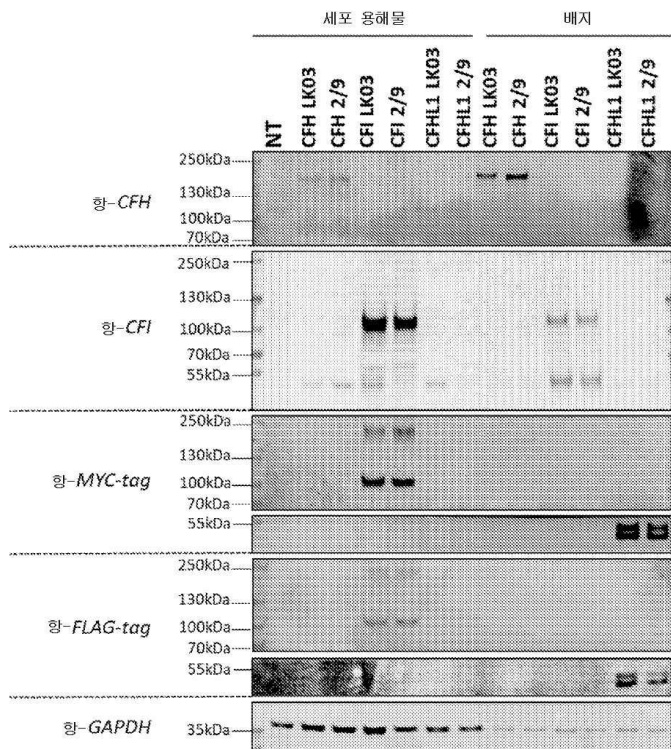


0.8% 아가로스

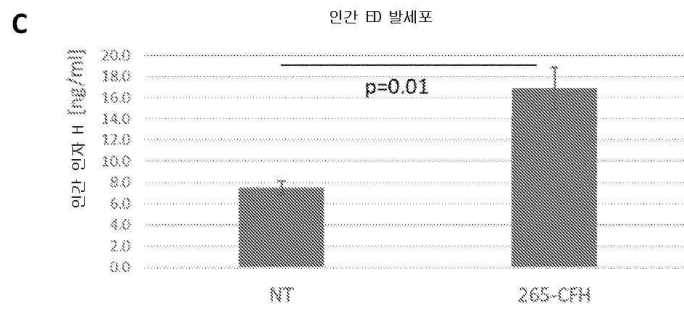
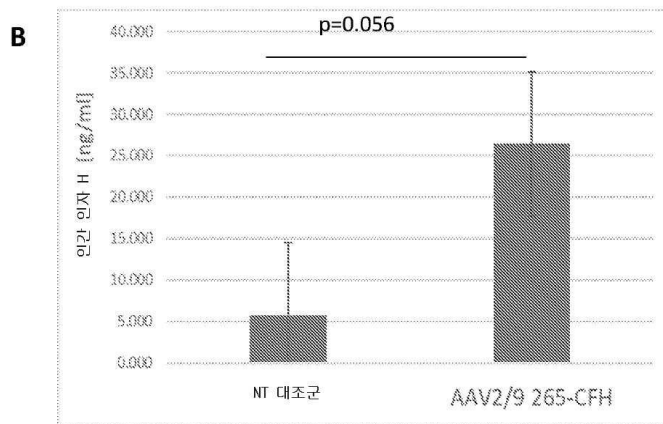
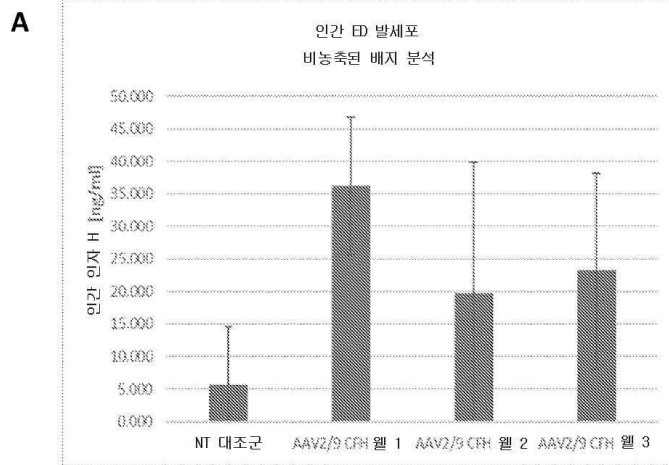
도면8



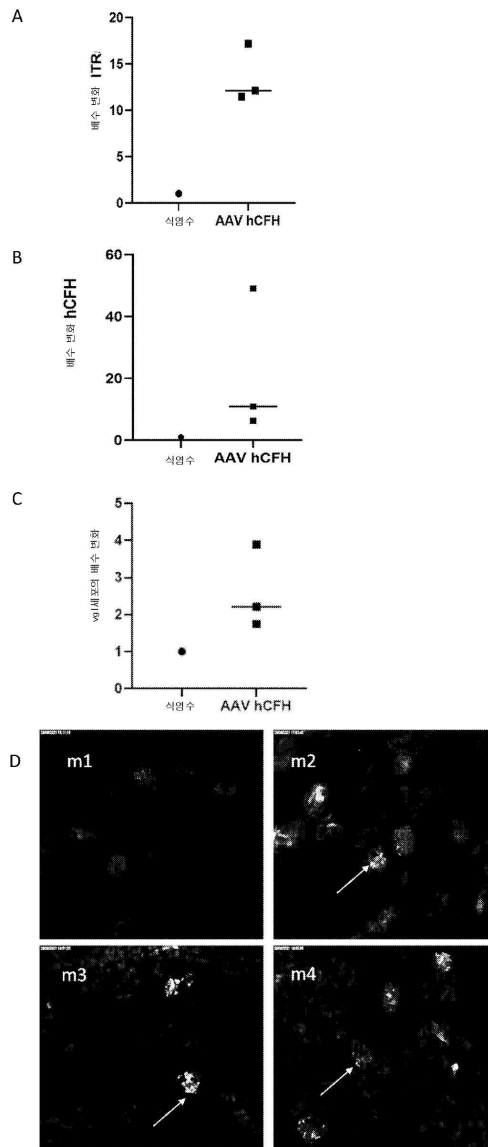
도면9



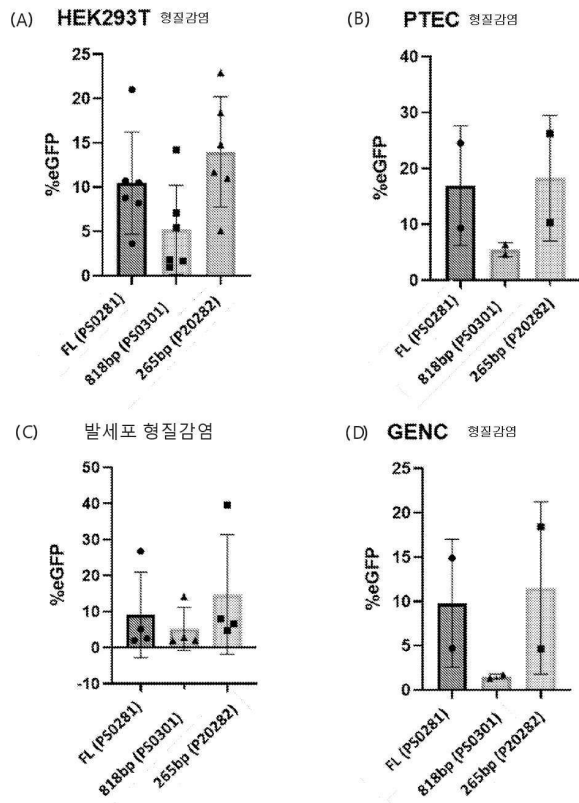
도면10



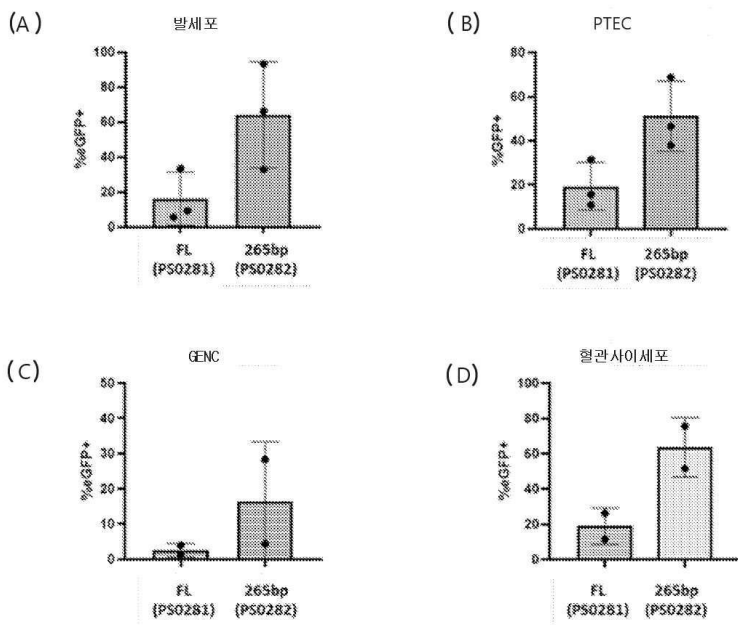
도면11



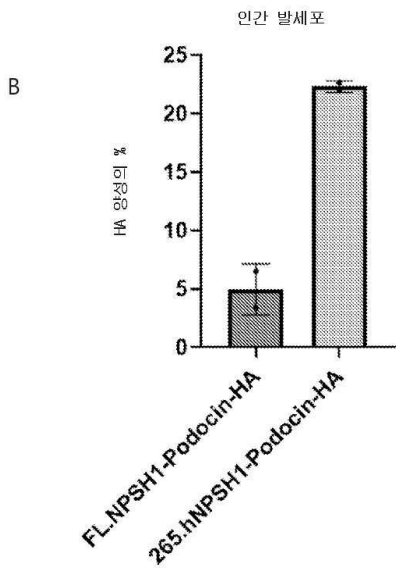
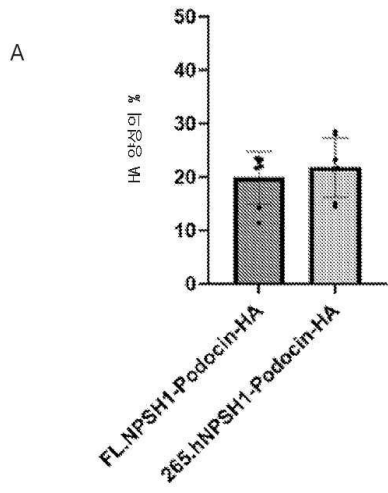
도면12



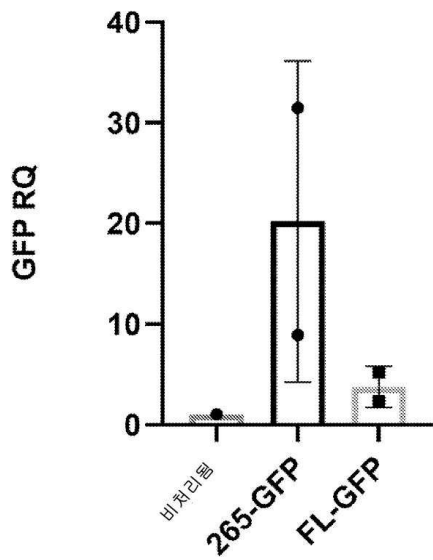
도면13



도면14



도면15



서열 목록

SEQUENCE LISTING

- <110> The University of Bristol  
Syncona Investment Management Limited  
Purespring Therapeutics Limited
- <120> PROMOTER
- <130> P121490PCT
- <140> PCT/GB2022/050649
- <141> 2022-03-14
- <150> GB 2103470.7
- <151> 2021-03-12
- <150> PCT/GB2021/050633
- <151> 2021-03-12
- <150> PCT/GB2021/051668
- <151> 2021-06-30
- <160> 63
- <170> PatentIn version 3.5
- <210> 1
- <211> 1249
- <212> DNA
- <213> Unknown

<220><223> Exemplary nephrin promoter

<400

> 1

```

cctgcagggc ccactagtct gtaatcccag cattttggga ggctgaggca gatggatcac      60
ctgaggtcag gaggttcgaga ccagcctggc caacatgatg aaaccccgctc tctagtaaaa    120
atacaaaaat tagccaggca tgggtctata tacctgtagt accagctact tgggagacag      180
aggtgggaga attacttgaa cctgggaggt tcaagccatg ggaggtggaa gttgcagtga    240
gccgagatgc cactgcactc cagcctgagc aacagagcaa gactatctca agaaaaaaaa    300
gaaagaaaga aagggacttg ccaaggtcat gtatcagggc aaggaagagc tgggggcca      360
gctggctgct ccctgctga gctgggagac caccttgatc tgacttctcc catcttcca      420

gcctaagcca ggccttgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat      480
gtgctgacag ggaattttat gctccagctg ggccagctgg gaggagcctg ctgggcagag     540
gccagagctg ggggctctgg aaggtacctg ggggaggttg cactgtgaga atgagctcaa     600
gctgggtcag agagcagggc tgactctgcc agtgcctgca tcagcctcat cgctctccta     660
ggctcctggc ctgctggact ctgggctgca ggtccttctt gaaaggctgt gagtagtgag     720
acaaggagca ggagtgaggg gtggcaggag agaagataga gattgagaga gagagagaga     780
gagacagaga gagaggaaga gacagagaca aaaggagaga gaacggctta gacaaggaga     840

gaaagatgga aagataaaga gactgggcgc agtggctcac gcctgtaatc ccaacacttg     900
gggaggccaa ggtgggagga tggcttgaag gaaagagtct gagatcaacc tggccaacat     960
agtgagaccg cgctctaaa aaaaaaaaaa aaaaaaaaaa gaaaaaagaa aaaaaagttt    1020
ttttaagag acagagaaag agactcagag attgagactg agagcaagac agagagagac    1080
actcacaggg aagaggggaa gaggaaaacg agaaaggag gagagtaacg gaaagagata    1140
aaaaagaaaa gcaggtggca gagacacaga gagagggacc cagagaaagc cagacagacg    1200
caggtggctg gcagcggcgc ctgtgggggt cacagtaggg ggacctgtc                  1249

```

<210> 2

<211> 819

<212> DNA

<213> Unknown

<220><223> Exemplary minimal nephrin promoter

<400> 2

```

ggccctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag      60
ggaattttat gctccagctg ggccagctgg gaggagcctg ctgggcagag gccagagctg    120

```

ggggctctgg aaggtacctg ggggagggtg cactgtgaga atgagctcaa gctgggtcag 180  
 agagcagggc tgactctgcc agtgctgca tcagcctcat cgctctccta ggctcctggc 240  
 ctgctggact ctgggctgca ggtccttctt gaaaggctgt gagtagtgag acaaggagca 300  
 ggagtgaggg gtggcaggag agaagataga gattgagaga gagagagaga gagacagaga 360

gagaggaaga gacagagaca aaaggagaga gaacggctta gacaaggaga gaaagatgga 420  
 aagataaaga gactgggcgc agtggtcac gcctgtaatc ccaacacttg gggaggccaa 480  
 ggtgggagga tggcttgaag gaaagagtct gagatcaacc tggccaacat agtgagaccc 540  
 cgtctctaaa aaaaaaaaaa aaaaaaaaaa gaaaaaagaa aaaaaagttt ttttaaagag 600  
 acagagaaag agactcagag attgagactg agagcaagac agagagagac actcacaggg 660  
 aagaggggaa gaggaaaacg agaaaggag gagagtaacg gaaagagata aaaaagaaaa 720  
 gcaggtggca gagacacaga gagagggacc cagagaaagc cagacagacg caggtggctg 780

gcagcgggcg ctgtgggggt cacagtaggg ggacctgtc 819

<210> 3

<211> 265

<212> DNA

<213> Unknown

<220><223> Exemplary minimal nephrin promoter

<400> 3

ggccctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag 60  
 ggaattttat gctccaggag caagacagag agagacactc acaggaaga ggggaagagg 120  
 aaaacgagaa agggaggaga gtaacgaaa gagataaaaa agaaaagcag gtggcagaga 180  
 cacagagaga gggaccaga gaaagccaga cagacgcagg tggctggcag cgggcgctgt 240  
 ggggtcaca gtagggggac ctgtc 265

<210> 4

<211> 77

<212> DNA

<213> Unknown

<220><223> Exemplary proximal promoter region

<400> 4

ggccctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag 60  
 ggaattttat gctccag 77

<210> 5

<211> 32  
 <212> DNA  
 <213> Unknown  
 <220><223> Exemplary core promoter region  
 <400> 5  
 gagcaagaca gagagagaca ctcacaggga ag 32  
 <210> 6  
 <211> 156  
 <212> DNA  
 <213> Unknown  
 <220><  
 223> Exemplary 5'UTR  
 <400> 6  
 aggggaagag gaaaacgaga aaggaggag agtaacggaa agagataaaa aagaaaagca 60  
 ggtggcagag acacagagag agggaccag agaagccag acagacgcag gtggctggca 120  
 gcgggcgctg tgggggtcac agtagggga cctgtc 156  
 <210> 7  
 <211> 188  
 <212> DNA  
 <213> Unknown  
 <220><223> Exemplary core promoter region and 5'UTR  
 <400> 7  
 gagcaagaca gagagagaca ctcacaggga agagggaag aggaaaacga gaaaggagg 60  
 agagtaacgg aaagagataa aaaagaaaag caggtggcag agacacagag agagggaccc 120  
 agagaaagcc agacagacgc aggtggctgg cagcggcgc tgtgggggtc acagtagggg 180  
 gacctgtc 188  
 <210> 8  
 <211> 554  
 <212> DNA  
 <213> Unknown  
 <220><223> Exemplary optional promoter region  
 <400> 8  
 ctgggccagc tgggaggagc ctgctgggca gaggccagag ctgggggctc tggaaggtac 60  
 ctgggggagg ttgcactgtg agaatgagct caagctgggt cagagagcag ggctgactct 120

gccagtgcct gcatcagcct catcgctctc ctaggctcct ggccctgctgg actctgggct 180  
gcaggtcctt ctgaaaggc tgtgagtagt gagacaagga gcaggagtga ggggtggcag 240

gagagaagat agagattgag agagagagag agagagacag agagagagga agagacagag 300  
acaaaaggag agagaacggc ttagacaagg agagaaagat ggaaagataa agagactggg 360  
cgcagtggct caccgctgta atcccaacac ttggggaggc caaggtggga ggatggcttg 420  
aaggaaagag tctgagatca acctggccaa catagtgaga ccccgctctt aaaaaaaaaa 480  
aagaaaaaaaa aaagaaaaaaaa gaaaaaaaaag tttttttaa gagacagaga aagagactca 540  
gagattgaga ctga 554

<210> 9  
<211> 430  
<212> DNA  
<213> Unknown

<220><223> Exemplary optional upstream promoter region

<400> 9

cctgcagggc ccactagtct gtaatcccag cattttggga ggctgaggca gatggatcac 60  
ctgaggtcag gatttcgaga ccagcctggc caacatgatg aaaccccgtc tctagtaaaa 120  
atacaaaaat tagccaggca tgggtctata tacctgtagt accagctact tgggagacag 180  
aggtgggaga attacttgaa cctgggaggt tcaagccatg ggaggtggaa gttgcagtga 240  
gccgagatgc cactgcactc cagcctgagc aacagagcaa gactatctca agaaaaaaaa 300  
gaaagaaaga aagggacttg ccaaggtcat gtatcagggc aaggaagagc tgggggcccc 360

gctggctgct cccctgctga gctgggagac caccttgatc tgacttctcc catcttccca 420  
gcctaagcca 430

<210> 10  
<211> 7  
<212> DNA  
<213> Unknown

<220><223> Exemplary RAR binding site

<400> 10

ggggtca 7

<210> 11  
<211> 17  
<212> DNA

<213> Unknown	
<220><223> Exemplary WT1 binding site	
<400> 11	
cggaggctgg ggaggca	17
<210> 12	
<211> 5	
<212> DNA	
<213> Unknown	
<220><223> Exemplary enhancer box	
<400> 12	
atgtg	5
<210> 13	
<211> 30	
<212> DNA	
<213> Unknown	
<220><223> Exemplary transcription factor binding region	
<400> 13	
gagcaagaca gagagagaca ctcacaggga	30
<210> 14	
<211> 1192	
<212> DNA	
<213> Unknown	
<220><223> Exemplary variant nephrin promoter	
<400> 14	
cacctgaggt caggagtgcg agaccagcgt ggccaacatg atgaaacccc gtctctagta	60
aaaatacaaa aattagccag gcatggtgct atatacctgt agcaccagct acttgggaga	120
cagaggtggg agaattactt gaacctggga ggttcaagcc atgggaggtg gaagttgcag	180
tgagccgaga tgccactgca ctccagcctg agcaacagag caagactatc tcaagaaaag	240
aaagaaagaa agaaagagac ttgccaaggt catgtatcag ggcaaggaag agctgggggc	300
ccagctggct gctccccctg tgagctggga gaccacctg atctgacttc tccatcttc	360
ccagcetaag ccaggcctg gggtcacgga ggtggggag gcaccgagga acgcgcctgg	420
catgtgctga caggggattt tatgtccag ctgggccagc tgggaggagc ctgctgggca	480

gaggccagag ctgggggctc tggaaggtag ctgggggagg ttgcactgtg agaatgagct 540  
 caagctgggt cagagagcag ggctgactct gccagtgcct gcatcagcct catcgctctc 600  
 ctagctcctt ggctgctgg actctgggct gcaggtcctt cttgaaagc tgtgagtgt 660  
 gagacaagga gcaggagtga ggggtggcag gagagaagat agagattgag agagagagag 720  
 agagagagac agagagagag gaagagacag agacaaaagg agagagaacg gcttagacaa 780  
 ggagagaaag atggaaagat aaagagactg ggcgcagtgg ctcacgcctg taatccaac 840  
 acttggggag gccaaaggtgg gaggatggct tgaaggaaag agtctgagat caacctggcc 900  
  
 aacatagtga gacccctct ctaaaaaaaaa aagaaaaaaaa aaagaaaaaaaa gaaaaaaaaag 960  
 tttttttaa gagacagaga aagagactca gagattgaga ctgagagcaa gacagagaga 1020  
 gatactcaca ggaagaggg gaagaggaaa acgagaaagg gaggagagta acggaagag 1080  
 ataaaaaga aaagcaggtg gcagagacac acagagaggg acccagagaa agccagacag 1140  
 acgcaggtgg ctggcagcgg gcgctgtggg ggtcacagta gggggacctg tg 1192  
  
 <210> 15  
 <211> 819  
 <212> DNA  
 <213> Unknown  
 <220><223> Exemplary minimal nephrin promoter variant  
 <400> 15  
 gccctgggg tcacggaggc tggggaggca ccgaggaacg cgctggcat gtgctgacag 60  
  
 gggattttat gctccagctg ggccagctgg gaggagcctg ctgggcagag gccagagctg 120  
 ggggctctgg aaggtacctg ggggaggttg cactgtgaga atgagctcaa gctgggtcag 180  
 agagcagggc tgactctgcc agtgctgca tcagcctcat cgctctccta ggctcctggc 240  
 ctgctggact ctgggctgca ggtccttctt gaaaggctgt gagtgtgag acaaggagca 300  
 ggagtgaggg gtggcaggag agaagataga gattgagaga gagagagaga gagagacaga 360  
 gagagaggaa gagacagaga caaaaggaga gagaacggct tagacaagga gagaaagatg 420  
 gaaagataaa gagactgggc gcagtggctc acgctgtaa tccaacact tggggaggcc 480  
  
 aaggtgggag gatggcttga aggaaagagt ctgagatcaa cctggccaac atagtgagac 540  
 cccgtctcta aaaaaaaaaa aaaaaaaaaa gaaaaaagaa aaaaaagttt ttttaagag 600  
 acagagaaag agactcagag attgagactg agagcaagac agagagagat actcacaggg 660  
 aagaggggaa gaggaaaacg agaaagggag gagagtacg gaaagagata aaaaagaaaa 720  
 gcaggtggca gagacacaca gagagggacc cagagaaagc cagacagacg caggtggctg 780  
 gcagcgggcg ctgtgggggt cacagtaggg ggacctgtg 819

<210> 16  
 <211> 265  
 <212> DNA  
 <213> Unknown

<220><223> Exemplary minimal nephrin promoter variant  
 <400> 16  
 ggcctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag 60  
 gggattttat gctccaggag caagacagag agagatactc acaggaaga ggggaagagg 120  
 aaaacagaa agggaggaga gtaacggaaa gagataaaaa agaaaagcag gtggcagaga 180  
 cacagagaga gggaccaga gaaagccaga cagacgcagg tggctggcag cgggcgctgt 240  
 gggggtcaca gtagggggac ctgtc 265

<210> 17  
 <211> 77  
 <212> DNA  
 <213> Unknown

<220>  
 <223> Exemplary variant proximal promoter region  
 <400> 17  
 ggcctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag 60  
 gggattttat gctccag 77

<210> 18  
 <211> 32  
 <212> DNA  
 <213> Unknown

<220><223> Exemplary variant core promoter region  
 <400> 18  
 gagcaagaca gagagagata ctcacaggga ag 32

<210> 19  
 <211> 156  
 <212> DNA  
 <213> Unknown

<220><223> Exemplary variant 5'UTR  
 <400> 19

aggggaagag gaaaacgaga aaggaggag agtaacggaa agagataaaa aagaaaagca 60  
 ggtggcagag acacacagag agggaccag agaaagccag acagacgcag gtggctggca 120  
 gcgggcgctg tgggggtcac agtagggga cctgtg 156

<210> 20

<211> 188

<212> DNA

<213> Unknown

<220><223> Exemplary variant core promoter region and 5'UTR

<400> 20

gagcaagaca gagagagata ctcacagga agagggaag aggaaaacga gaaaggagg 60  
 agagtaacgg aaagagataa aaaagaaaag caggtggcag agacacacag agagggacc 120

agagaaagcc agacagacgc aggtggctgg cagcgggccc tgtgggggtc acagtagggg 180  
 gacctgtg 188

<210> 21

<211> 1670

<212> PRT

<213> Unknown

<220><223> Exemplary COL4A3 amino acid sequence

<400> 21

Met Ser Ala Arg Thr Ala Pro Arg Pro Gln Val Leu Leu Leu Pro Leu

1                    5                    10                    15

Leu Leu Val Leu Leu Ala Ala Ala Pro Ala Ala Ser Lys Gly Cys Val

20                    25                    30

Cys Lys Asp Lys Gly Gln Cys Phe Cys Asp Gly Ala Lys Gly Glu Lys

35                    40                    45

Gly Glu Lys Gly Phe Pro Gly Pro Pro Gly Ser Pro Gly Gln Lys Gly

50                    55                    60

Phe Thr Gly Pro Glu Gly Leu Pro Gly Pro Gln Gly Pro Lys Gly Phe

65                    70                    75                    80

Pro Gly Leu Pro Gly Leu Thr Gly Ser Lys Gly Val Arg Gly Ile Ser

85                    90                    95

Gly Leu Pro Gly Phe Ser Gly Ser Pro Gly Leu Pro Gly Thr Pro Gly



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





1085	1090	1095
Pro Glu Gly Ala Pro Gly Ser	Pro Gly Ser Pro Gly	Leu Pro Gly
1100	1105	1110
Lys Pro Gly Pro His Gly Asp	Leu Gly Phe Lys Gly	Ile Lys Gly
1115	1120	1125
Leu Leu Gly Pro Pro Gly Ile	Arg Gly Pro Pro Gly	Leu Pro Gly
1130	1135	1140
Phe Pro Gly Ser Pro Gly Pro	Met Gly Ile Arg Gly	Asp Gln Gly
1145	1150	1155
Arg Asp Gly Ile Pro Gly Pro	Ala Gly Glu Lys Gly	Glu Thr Gly
1160	1165	1170
Leu Leu Arg Ala Pro Pro Gly	Pro Arg Gly Asn Pro	Gly Ala Gln
1175	1180	1185
Gly Ala Lys Gly Asp Arg Gly	Ala Pro Gly Phe Pro	Gly Leu Pro
1190	1195	1200
Gly Arg Lys Gly Ala Met Gly	Asp Ala Gly Pro Arg	Gly Pro Thr
1205	1210	1215
Gly Ile Glu Gly Phe Pro Gly	Pro Pro Gly Leu Pro	Gly Ala Ile
1220	1225	1230
Ile Pro Gly Gln Thr Gly Asn	Arg Gly Pro Pro Gly	Ser Arg Gly
1235	1240	1245
Ser Pro Gly Ala Pro Gly Pro	Pro Gly Pro Pro Gly	Ser His Val
1250	1255	1260
Ile Gly Ile Lys Gly Asp Lys	Gly Ser Met Gly His	Pro Gly Pro
1265	1270	1275
Lys Gly Pro Pro Gly Thr Ala	Gly Asp Met Gly Pro	Pro Gly Arg
1280	1285	1290
Leu Gly Ala Pro Gly Thr Pro	Gly Leu Pro Gly Pro	Arg Gly Asp
1295	1300	1305
Pro Gly Phe Gln Gly Phe Pro	Gly Val Lys Gly Glu	Lys Gly Asn
1310	1315	1320

Pro Gly Phe Leu Gly Ser Ile Gly Pro Pro Gly Pro Ile Gly Pro  
 1325 1330 1335

Lys Gly Pro Pro Gly Val Arg Gly Asp Pro Gly Thr Leu Lys Ile  
 1340 1345 1350

Ile Ser Leu Pro Gly Ser Pro Gly Pro Pro Gly Thr Pro Gly Glu  
 1355 1360 1365

Pro Gly Met Gln Gly Glu Pro Gly Pro Pro Gly Pro Pro Gly Asn  
 1370 1375 1380

Leu Gly Pro Cys Gly Pro Arg Gly Lys Pro Gly Lys Asp Gly Lys  
 1385 1390 1395

Pro Gly Thr Pro Gly Pro Ala Gly Glu Lys Gly Asn Lys Gly Ser  
 1400 1405 1410

Lys Gly Glu Pro Gly Pro Ala Gly Ser Asp Gly Leu Pro Gly Leu  
 1415 1420 1425

Lys Gly Lys Arg Gly Asp Ser Gly Ser Pro Ala Thr Trp Thr Thr  
 1430 1435 1440

Arg Gly Phe Val Phe Thr Arg His Ser Gln Thr Thr Ala Ile Pro  
 1445 1450 1455

Ser Cys Pro Glu Gly Thr Val Pro Leu Tyr Ser Gly Phe Ser Phe  
 1460 1465 1470

Leu Phe Val Gln Gly Asn Gln Arg Ala His Gly Gln Asp Leu Gly  
 1475 1480 1485

Thr Leu Gly Ser Cys Leu Gln Arg Phe Thr Thr Met Pro Phe Leu  
 1490 1495 1500

Phe Cys Asn Val Asn Asp Val Cys Asn Phe Ala Ser Arg Asn Asp  
 1505 1510 1515

Tyr Ser Tyr Trp Leu Ser Thr Pro Ala Leu Met Pro Met Asn Met  
 1520 1525 1530

Ala Pro Ile Thr Gly Arg Ala Leu Glu Pro Tyr Ile Ser Arg Cys  
 1535 1540 1545

Thr Val Cys Glu Gly Pro Ala Ile Ala Ile Ala Val His Ser Gln

1550                      1555                      1560  
 Thr Thr Asp Ile Pro Pro Cys Pro His Gly Trp Ile Ser Leu Trp

1565                      1570                      1575  
 Lys Gly Phe Ser Phe Ile Met Phe Thr Ser Ala Gly Ser Glu Gly  
 1580                      1585                      1590

Thr Gly Gln Ala Leu Ala Ser Pro Gly Ser Cys Leu Glu Glu Phe  
 1595                      1600                      1605

Arg Ala Ser Pro Phe Leu Glu Cys His Gly Arg Gly Thr Cys Asn  
 1610                      1615                      1620

Tyr Tyr Ser Asn Ser Tyr Ser Phe Trp Leu Ala Ser Leu Asn Pro  
 1625                      1630                      1635

Glu Arg Met Phe Arg Lys Pro Ile Pro Ser Thr Val Lys Ala Gly  
 1640                      1645                      1650

Glu Leu Glu Lys Ile Ile Ser Arg Cys Gln Val Cys Met Lys Lys  
 1655                      1660                      1665

Arg His  
 1670

<210> 22

<211> 1690

<212> PRT

<213> Unknown

<220><223> Exemplary COL4A4 amino acid sequence

<400> 22

Met Trp Ser Leu His Ile Val Leu Met Arg Cys Ser Phe Arg Leu Thr  
 1                      5                      10                      15

Lys Ser Leu Ala Thr Gly Pro Trp Ser Leu Ile Leu Ile Leu Phe Ser  
                     20                      25                      30

Val Gln Tyr Val Tyr Gly Ser Gly Lys Lys Tyr Ile Gly Pro Cys Gly  
                     35                      40                      45

Gly Arg Asp Cys Ser Val Cys His Cys Val Pro Glu Lys Gly Ser Arg  
                     50                      55                      60

Gly Pro Pro Gly Pro Pro Gly Pro Gln Gly Pro Ile Gly Pro Leu Gly



Phe Pro Gly Leu Lys Gly Glu Leu Gly Leu Val Gly Asp Pro Gly Leu  
 325 330 335  
 Phe Gly Leu Ile Gly Pro Lys Gly Asp Pro Gly Asn Arg Gly His Pro  
 340 345 350  
  
 Gly Pro Pro Gly Val Leu Val Thr Pro Pro Leu Pro Leu Lys Gly Pro  
 355 360 365  
 Pro Gly Asp Pro Gly Phe Pro Gly Arg Tyr Gly Glu Thr Gly Asp Val  
 370 375 380  
 Gly Pro Pro Gly Pro Pro Gly Leu Leu Gly Arg Pro Gly Glu Ala Cys  
 385 390 395 400  
 Ala Gly Met Ile Gly Pro Pro Gly Pro Gln Gly Phe Pro Gly Leu Pro  
 405 410 415  
  
 Gly Leu Pro Gly Glu Ala Gly Ile Pro Gly Arg Pro Asp Ser Ala Pro  
 420 425 430  
 Gly Lys Pro Gly Lys Pro Gly Ser Pro Gly Leu Pro Gly Ala Pro Gly  
 435 440 445  
 Leu Gln Gly Leu Pro Gly Ser Ser Val Ile Tyr Cys Ser Val Gly Asn  
 450 455 460  
 Pro Gly Pro Gln Gly Ile Lys Gly Lys Val Gly Pro Pro Gly Gly Arg  
 465 470 475 480  
  
 Gly Pro Lys Gly Glu Lys Gly Asn Glu Gly Leu Cys Ala Cys Glu Pro  
 485 490 495  
 Gly Pro Met Gly Pro Pro Gly Pro Pro Gly Leu Pro Gly Arg Gln Gly  
 500 505 510  
 Ser Lys Gly Asp Leu Gly Leu Pro Gly Trp Leu Gly Thr Lys Gly Asp  
 515 520 525  
 Pro Gly Pro Pro Gly Ala Glu Gly Pro Pro Gly Leu Pro Gly Lys His  
 530 535 540  
  
 Gly Ala Ser Gly Pro Pro Gly Asn Lys Gly Ala Lys Gly Asp Met Val  
 545 550 555 560  
 Val Ser Arg Val Lys Gly His Lys Gly Glu Arg Gly Pro Asp Gly Pro



Phe Pro Gly Val Pro Gly Pro Pro Gly His Ser Cys Glu Arg Gly Ala  
 820 825 830  
 Pro Gly Ile Pro Gly Gln Pro Gly Leu Pro Gly Tyr Pro Gly Ser Pro  
 835 840 845  
 Gly Ala Pro Gly Gly Lys Gly Gln Pro Gly Asp Val Gly Pro Pro Gly  
 850 855 860  
  
 Pro Ala Gly Met Lys Gly Leu Pro Gly Leu Pro Gly Arg Pro Gly Ala  
 865 870 875 880  
 His Gly Pro Pro Gly Leu Pro Gly Ile Pro Gly Pro Phe Gly Asp Asp  
 885 890 895  
 Gly Leu Pro Gly Pro Pro Gly Pro Lys Gly Pro Arg Gly Leu Pro Gly  
 900 905 910  
 Phe Pro Gly Phe Pro Gly Glu Arg Gly Lys Pro Gly Ala Glu Gly Cys  
 915 920 925  
  
 Pro Gly Ala Lys Gly Glu Pro Gly Glu Lys Gly Met Ser Gly Leu Pro  
 930 935 940  
 Gly Asp Arg Gly Leu Arg Gly Ala Lys Gly Ala Ile Gly Pro Pro Gly  
 945 950 955 960  
 Asp Glu Gly Glu Met Ala Ile Ile Ser Gln Lys Gly Thr Pro Gly Glu  
 965 970 975  
 Pro Gly Pro Pro Gly Asp Asp Gly Phe Pro Gly Glu Arg Gly Asp Lys  
 980 985 990  
  
 Gly Thr Pro Gly Met Gln Gly Arg Arg Gly Glu Pro Gly Arg Tyr Gly  
 995 1000 1005  
 Pro Pro Gly Phe His Arg Gly Glu Pro Gly Glu Lys Gly Gln Pro  
 1010 1015 1020  
 Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Ser Thr Gly Leu Arg  
 1025 1030 1035  
 Gly Phe Ile Gly Phe Pro Gly Leu Pro Gly Asp Gln Gly Glu Pro  
 1040 1045 1050  
  
 Gly Ser Pro Gly Pro Pro Gly Phe Ser Gly Ile Asp Gly Ala Arg

1055	1060	1065
Gly Pro Lys Gly Asn Lys Gly Asp Pro Ala Ser His Phe Gly Pro		
1070	1075	1080
Pro Gly Pro Lys Gly Glu Pro Gly Ser Pro Gly Cys Pro Gly His		
1085	1090	1095
Phe Gly Ala Ser Gly Glu Gln Gly Leu Pro Gly Ile Gln Gly Pro		
1100	1105	1110
Arg Gly Ser Pro Gly Arg Pro Gly Pro Pro Gly Ser Ser Gly Pro		
1115	1120	1125
Pro Gly Cys Pro Gly Asp His Gly Met Pro Gly Leu Arg Gly Gln		
1130	1135	1140
Pro Gly Glu Met Gly Asp Pro Gly Pro Arg Gly Leu Gln Gly Asp		
1145	1150	1155
Pro Gly Ile Pro Gly Pro Pro Gly Ile Lys Gly Pro Ser Gly Ser		
1160	1165	1170
Pro Gly Leu Asn Gly Leu His Gly Leu Lys Gly Gln Lys Gly Thr		
1175	1180	1185
Lys Gly Ala Ser Gly Leu His Asp Val Gly Pro Pro Gly Pro Val		
1190	1195	1200
Gly Ile Pro Gly Leu Lys Gly Glu Arg Gly Asp Pro Gly Ser Pro		
1205	1210	1215
Gly Ile Ser Pro Pro Gly Pro Arg Gly Lys Lys Gly Pro Pro Gly		
1220	1225	1230
Pro Pro Gly Ser Ser Gly Pro Pro Gly Pro Ala Gly Ala Thr Gly		
1235	1240	1245
Arg Ala Pro Lys Asp Ile Pro Asp Pro Gly Pro Pro Gly Asp Gln		
1250	1255	1260
Gly Pro Pro Gly Pro Asp Gly Pro Arg Gly Ala Pro Gly Pro Pro		
1265	1270	1275
Gly Leu Pro Gly Ser Val Asp Leu Leu Arg Gly Glu Pro Gly Asp		
1280	1285	1290

Cys Gly Leu Pro Gly Pro Pro Gly Pro Pro Gly Pro Pro Gly Pro  
 1295 1300 1305  
 Pro Gly Tyr Lys Gly Phe Pro Gly Cys Asp Gly Lys Asp Gly Gln  
 1310 1315 1320  
 Lys Gly Pro Val Gly Phe Pro Gly Pro Gln Gly Pro His Gly Phe  
 1325 1330 1335  
 Pro Gly Pro Pro Gly Glu Lys Gly Leu Pro Gly Pro Pro Gly Arg  
 1340 1345 1350  
  
 Lys Gly Pro Thr Gly Leu Pro Gly Pro Arg Gly Glu Pro Gly Pro  
 1355 1360 1365  
 Pro Ala Asp Val Asp Asp Cys Pro Arg Ile Pro Gly Leu Pro Gly  
 1370 1375 1380  
 Ala Pro Gly Met Arg Gly Pro Glu Gly Ala Met Gly Leu Pro Gly  
 1385 1390 1395  
 Met Arg Gly Pro Ser Gly Pro Gly Cys Lys Gly Glu Pro Gly Leu  
 1400 1405 1410  
  
 Asp Gly Arg Arg Gly Val Asp Gly Val Pro Gly Ser Pro Gly Pro  
 1415 1420 1425  
 Pro Gly Arg Lys Gly Asp Thr Gly Glu Asp Gly Tyr Pro Gly Gly  
 1430 1435 1440  
 Pro Gly Pro Pro Gly Pro Ile Gly Asp Pro Gly Pro Lys Gly Phe  
 1445 1450 1455  
 Gly Pro Gly Tyr Leu Gly Gly Phe Leu Leu Val Leu His Ser Gln  
 1460 1465 1470  
  
 Thr Asp Gln Glu Pro Thr Cys Pro Leu Gly Met Pro Arg Leu Trp  
 1475 1480 1485  
 Thr Gly Tyr Ser Leu Leu Tyr Leu Glu Gly Gln Glu Lys Ala His  
 1490 1495 1500  
 Asn Gln Asp Leu Gly Leu Ala Gly Ser Cys Leu Pro Val Phe Ser  
 1505 1510 1515  
 Thr Leu Pro Phe Ala Tyr Cys Asn Ile His Gln Val Cys His Tyr

1520	1525	1530
Ala Gln Arg Asn Asp Arg Ser Tyr Trp Leu Ala Ser Ala Ala Pro		
1535	1540	1545
Leu Pro Met Met Pro Leu Ser Glu Glu Ala Ile Arg Pro Tyr Val		
1550	1555	1560
Ser Arg Cys Ala Val Cys Glu Ala Pro Ala Gln Ala Val Ala Val		
1565	1570	1575
His Ser Gln Asp Gln Ser Ile Pro Pro Cys Pro Gln Thr Trp Arg		
1580	1585	1590
Ser Leu Trp Ile Gly Tyr Ser Phe Leu Met His Thr Gly Ala Gly		
1595	1600	1605
Asp Gln Gly Gly Gly Gln Ala Leu Met Ser Pro Gly Ser Cys Leu		
1610	1615	1620
Glu Asp Phe Arg Ala Ala Pro Phe Leu Glu Cys Gln Gly Arg Gln		
1625	1630	1635
Gly Thr Cys His Phe Phe Ala Asn Lys Tyr Ser Phe Trp Leu Thr		
1640	1645	1650
Thr Val Lys Ala Asp Leu Gln Phe Ser Ser Ala Pro Ala Pro Asp		
1655	1660	1665
Thr Leu Lys Glu Ser Gln Ala Gln Arg Gln Lys Ile Ser Arg Cys		
1670	1675	1680
Gln Val Cys Val Lys Tyr Ser		
1685	1690	
<210> 23		
<211> 1685		
<212> PRT		
<213> Unknown		
<220><223> Exemplary COL4A5 amino acid sequence		
<400> 23		
Met Lys Leu Arg Gly Val Ser Leu Ala Ala Gly Leu Phe Leu Leu Ala		
1	5	10
		15

Leu Ser Leu Trp Gly Gln Pro Ala Glu Ala Ala Ala Cys Tyr Gly Cys  
 20 25 30  
 Ser Pro Gly Ser Lys Cys Asp Cys Ser Gly Ile Lys Gly Glu Lys Gly  
 35 40 45  
 Glu Arg Gly Phe Pro Gly Leu Glu Gly His Pro Gly Leu Pro Gly Phe  
 50 55 60  
 Pro Gly Pro Glu Gly Pro Pro Gly Pro Arg Gly Gln Lys Gly Asp Asp  
 65 70 75 80  
  
 Gly Ile Pro Gly Pro Pro Gly Pro Lys Gly Ile Arg Gly Pro Pro Gly  
 85 90 95  
 Leu Pro Gly Phe Pro Gly Thr Pro Gly Leu Pro Gly Met Pro Gly His  
 100 105 110  
 Asp Gly Ala Pro Gly Pro Gln Gly Ile Pro Gly Cys Asn Gly Thr Lys  
 115 120 125  
 Gly Glu Arg Gly Phe Pro Gly Ser Pro Gly Phe Pro Gly Leu Gln Gly  
 130 135 140  
  
 Pro Pro Gly Pro Pro Gly Ile Pro Gly Met Lys Gly Glu Pro Gly Ser  
 145 150 155 160  
 Ile Ile Met Ser Ser Leu Pro Gly Pro Lys Gly Asn Pro Gly Tyr Pro  
 165 170 175  
 Gly Pro Pro Gly Ile Gln Gly Leu Pro Gly Pro Thr Gly Ile Pro Gly  
 180 185 190  
 Pro Ile Gly Pro Pro Gly Pro Pro Gly Leu Met Gly Pro Pro Gly Pro  
 195 200 205  
  
 Pro Gly Leu Pro Gly Pro Lys Gly Asn Met Gly Leu Asn Phe Gln Gly  
 210 215 220  
 Pro Lys Gly Glu Lys Gly Glu Gln Gly Leu Gln Gly Pro Pro Gly Pro  
 225 230 235 240  
 Pro Gly Gln Ile Ser Glu Gln Lys Arg Pro Ile Asp Val Glu Phe Gln  
 245 250 255  
 Lys Gly Asp Gln Gly Leu Pro Gly Asp Arg Gly Pro Pro Gly Pro Pro

260	265	270	
Gly Ile Arg Gly Pro Pro Gly Pro Pro Gly Gly Glu Lys Gly Glu Lys			
275	280	285	
Gly Glu Gln Gly Glu Pro Gly Lys Arg Gly Lys Pro Gly Lys Asp Gly			
290	295	300	
Glu Asn Gly Gln Pro Gly Ile Pro Gly Leu Pro Gly Asp Pro Gly Tyr			
305	310	315	320
Pro Gly Glu Pro Gly Arg Asp Gly Glu Lys Gly Gln Lys Gly Asp Thr			
325	330	335	
Gly Pro Pro Gly Pro Pro Gly Leu Val Ile Pro Arg Pro Gly Thr Gly			
340	345	350	
Ile Thr Ile Gly Glu Lys Gly Asn Ile Gly Leu Pro Gly Leu Pro Gly			
355	360	365	
Glu Lys Gly Glu Arg Gly Phe Pro Gly Ile Gln Gly Pro Pro Gly Leu			
370	375	380	
Pro Gly Pro Pro Gly Ala Ala Val Met Gly Pro Pro Gly Pro Pro Gly			
385	390	395	400
Phe Pro Gly Glu Arg Gly Gln Lys Gly Asp Glu Gly Pro Pro Gly Ile			
405	410	415	
Ser Ile Pro Gly Pro Pro Gly Leu Asp Gly Gln Pro Gly Ala Pro Gly			
420	425	430	
Leu Pro Gly Pro Pro Gly Pro Ala Gly Pro His Ile Pro Pro Ser Asp			
435	440	445	
Glu Ile Cys Glu Pro Gly Pro Pro Gly Pro Pro Gly Ser Pro Gly Asp			
450	455	460	
Lys Gly Leu Gln Gly Glu Gln Gly Val Lys Gly Asp Lys Gly Asp Thr			
465	470	475	480
Cys Phe Asn Cys Ile Gly Thr Gly Ile Ser Gly Pro Pro Gly Gln Pro			
485	490	495	
Gly Leu Pro Gly Leu Pro Gly Pro Pro Gly Ser Leu Gly Phe Pro Gly			
500	505	510	

Gln Lys Gly Glu Lys Gly Gln Ala Gly Ala Thr Gly Pro Lys Gly Leu  
 515 520 525

Pro Gly Ile Pro Gly Ala Pro Gly Ala Pro Gly Phe Pro Gly Ser Lys  
 530 535 540

Gly Glu Pro Gly Asp Ile Leu Thr Phe Pro Gly Met Lys Gly Asp Lys  
 545 550 555 560

Gly Glu Leu Gly Ser Pro Gly Ala Pro Gly Leu Pro Gly Leu Pro Gly  
 565 570 575

Thr Pro Gly Gln Asp Gly Leu Pro Gly Leu Pro Gly Pro Lys Gly Glu  
 580 585 590

Pro Gly Gly Ile Thr Phe Lys Gly Glu Arg Gly Pro Pro Gly Asn Pro  
 595 600 605

Gly Leu Pro Gly Leu Pro Gly Asn Ile Gly Pro Met Gly Pro Pro Gly  
 610 615 620

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

Gly Asn Pro Gly Gln Pro Gly Ile Pro Gly Pro Lys Gly Asp Pro Gly  
 645 650 655

Gln Thr Ile Thr Gln Pro Gly Lys Pro Gly Leu Pro Gly Asn Pro Gly  
 660 665 670

Arg Asp Gly Asp Val Gly Leu Pro Gly Asp Pro Gly Leu Pro Gly Gln  
 675 680 685

Pro Gly Leu Pro Gly Ile Pro Gly Ser Lys Gly Glu Pro Gly Ile Pro  
 690 695 700

Gly Ile Gly Leu Pro Gly Pro Pro Gly Pro Lys Gly Phe Pro Gly Ile  
 705 710 715 720

Pro Gly Pro Pro Gly Ala Pro Gly Thr Pro Gly Arg Ile Gly Leu Glu  
 725 730 735

Gly Pro Pro Gly Pro Pro Gly Phe Pro Gly Pro Lys Gly Glu Pro Gly  
 740 745 750

Phe Ala Leu Pro Gly Pro Pro Gly Pro Pro Gly Leu Pro Gly Phe Lys



Gly Asn Pro Gly Leu Pro Gly Gln Pro Gly Leu Ile Gly Pro Pro  
 1010 1015 1020

Gly Leu Lys Gly Thr Ile Gly Asp Met Gly Phe Pro Gly Pro Gln  
 1025 1030 1035

Gly Val Glu Gly Pro Pro Gly Pro Ser Gly Val Pro Gly Gln Pro  
 1040 1045 1050

Gly Ser Pro Gly Leu Pro Gly Gln Lys Gly Asp Lys Gly Asp Pro  
 1055 1060 1065

Gly Ile Ser Ser Ile Gly Leu Pro Gly Leu Pro Gly Pro Lys Gly  
 1070 1075 1080

Glu Pro Gly Leu Pro Gly Tyr Pro Gly Asn Pro Gly Ile Lys Gly  
 1085 1090 1095

Ser Val Gly Asp Pro Gly Leu Pro Gly Leu Pro Gly Thr Pro Gly  
 1100 1105 1110

Ala Lys Gly Gln Pro Gly Leu Pro Gly Phe Pro Gly Thr Pro Gly  
 1115 1120 1125

Pro Pro Gly Pro Lys Gly Ile Ser Gly Pro Pro Gly Asn Pro Gly  
 1130 1135 1140

Leu Pro Gly Glu Pro Gly Pro Val Gly Gly Gly Gly His Pro Gly  
 1145 1150 1155

Gln Pro Gly Pro Pro Gly Glu Lys Gly Lys Pro Gly Gln Asp Gly  
 1160 1165 1170

Ile Pro Gly Pro Ala Gly Gln Lys Gly Glu Pro Gly Gln Pro Gly  
 1175 1180 1185

Phe Gly Asn Pro Gly Pro Pro Gly Leu Pro Gly Leu Ser Gly Gln  
 1190 1195 1200

Lys Gly Asp Gly Gly Leu Pro Gly Ile Pro Gly Asn Pro Gly Leu  
 1205 1210 1215

Pro Gly Pro Lys Gly Glu Pro Gly Phe His Gly Phe Pro Gly Val  
 1220 1225 1230

Gln Gly Pro Pro Gly Pro Pro Gly Ser Pro Gly Pro Ala Leu Glu

1235	1240	1245
Gly Pro Lys Gly Asn Pro Gly	Pro Gln Gly Pro Pro	Gly Arg Pro
1250	1255	1260
Gly Leu Pro Gly Pro Glu Gly	Pro Pro Gly Leu Pro	Gly Asn Gly
1265	1270	1275
Gly Ile Lys Gly Glu Lys Gly	Asn Pro Gly Gln Pro	Gly Leu Pro
1280	1285	1290
Gly Leu Pro Gly Leu Lys Gly	Asp Gln Gly Pro Pro	Gly Leu Gln
1295	1300	1305
Gly Asn Pro Gly Arg Pro Gly	Leu Asn Gly Met Lys	Gly Asp Pro
1310	1315	1320
Gly Leu Pro Gly Val Pro Gly	Phe Pro Gly Met Lys	Gly Pro Ser
1325	1330	1335
Gly Val Pro Gly Ser Ala Gly	Pro Glu Gly Glu Pro	Gly Leu Ile
1340	1345	1350
Gly Pro Pro Gly Pro Pro Gly	Leu Pro Gly Pro Ser	Gly Gln Ser
1355	1360	1365
Ile Ile Ile Lys Gly Asp Ala	Gly Pro Pro Gly Ile	Pro Gly Gln
1370	1375	1380
Pro Gly Leu Lys Gly Leu Pro	Gly Pro Gln Gly Pro	Gln Gly Leu
1385	1390	1395
Pro Gly Pro Thr Gly Pro Pro	Gly Asp Pro Gly Arg	Asn Gly Leu
1400	1405	1410
Pro Gly Phe Asp Gly Ala Gly	Gly Arg Lys Gly Asp	Pro Gly Leu
1415	1420	1425
Pro Gly Gln Pro Gly Thr Arg	Gly Leu Asp Gly Pro	Pro Gly Pro
1430	1435	1440
Asp Gly Leu Gln Gly Pro Pro	Gly Pro Pro Gly Thr	Ser Ser Val
1445	1450	1455
Ala His Gly Phe Leu Ile Thr	Arg His Ser Gln Thr	Thr Asp Ala
1460	1465	1470

Pro Gln Cys Pro Gln Gly Thr Leu Gln Val Tyr Glu Gly Phe Ser  
 1475 1480 1485

Leu Leu Tyr Val Gln Gly Asn Lys Arg Ala His Gly Gln Asp Leu  
 1490 1495 1500

Gly Thr Ala Gly Ser Cys Leu Arg Arg Phe Ser Thr Met Pro Phe  
 1505 1510 1515

Met Phe Cys Asn Ile Asn Asn Val Cys Asn Phe Ala Ser Arg Asn  
 1520 1525 1530

Asp Tyr Ser Tyr Trp Leu Ser Thr Pro Glu Pro Met Pro Met Ser  
 1535 1540 1545

Met Gln Pro Leu Lys Gly Gln Ser Ile Gln Pro Phe Ile Ser Arg  
 1550 1555 1560

Cys Ala Val Cys Glu Ala Pro Ala Val Val Ile Ala Val His Ser  
 1565 1570 1575

Gln Thr Ile Gln Ile Pro His Cys Pro Gln Gly Trp Asp Ser Leu  
 1580 1585 1590

Trp Ile Gly Tyr Ser Phe Met Met His Thr Ser Ala Gly Ala Glu  
 1595 1600 1605

Gly Ser Gly Gln Ala Leu Ala Ser Pro Gly Ser Cys Leu Glu Glu  
 1610 1615 1620

Phe Arg Ser Ala Pro Phe Ile Glu Cys His Gly Arg Gly Thr Cys  
 1625 1630 1635

Asn Tyr Tyr Ala Asn Ser Tyr Ser Phe Trp Leu Ala Thr Val Asp  
 1640 1645 1650

Val Ser Asp Met Phe Ser Lys Pro Gln Ser Glu Thr Leu Lys Ala  
 1655 1660 1665

Gly Asp Leu Arg Thr Arg Ile Ser Arg Cys Gln Val Cys Met Lys  
 1670 1675 1680

Arg Thr  
 1685

<210> 24

<211> 5013

<212> DNA

<213> Unknown

<220><223> Exemplary COL4A3 transgene sequence

<400> 24

```

atgagcgcgc ggaccgcccc caggccgcag gtgctcctgc tgccgctcct gctggtgctc      60
ctggcggcgg cgcccgcagc cagcaagggt tgtgtctgta aagacaaagg ccagtgtctc      120
tgtgacgggg ccaaagggga gaagggggag aagggtcttc ctggaccccc cggttctcct      180
ggccagaaag gattcacagg tctgaaggc ttgctggac cgcagggacc caagggttt      240
ccaggacttc caggactcac gggttccaaa ggtgtaaggg gaataagtgg attgccagga      300
ttttctggtt ctctggact tccaggcacc ccaggcaata cgggcctta cggactgtc      360
ggtgtaccag gatgcagtgg ttctaagggt ggcagggggt ttccaggact cccagggaca      420

ctgggctacc cagggatccc ggggtctgct ggtttgaaag gacaaaaggg tgctcctgct      480
aaagaagaag atatagaact tgatgcaaaa ggcgaccccc ggttgccagg ggctccagga      540
ccccagggtt tgccaggccc tccaggtttt cctgggectg ttggcccacc tggtectcgg      600
ggattctttg gctttccagg agccatggga cctagaggac ctaagggtca catgggtgaa      660
agagtgatag gacataaagg agagcggggt gtgaaagggt taacaggacc cccgggacca      720
ccaggaaacag ttattgtgac cctaactggc ccagataaca gaacggacct caagggggaa      780
aaggagaca agggagcaat gggcgagcct ggacctctg gacctcagg actgcctgga      840

gaatcatatg gatctgaaaa ggggtctcct ggagacctg gcctgcaggg aaaaccggga      900
aaagatggtg ttcttgctt ccctggaagt gagggagtca agggcaacag gggtttcct      960
gggttaatgg gtgaagatgg cattaaggga cagaaagggg acattggccc tccaggattt      1020
cgtggtccaa cagaatatta tgacacatac caggaaaagg gagatgaagg cactccaggc      1080
ccaccagggc ccagaggagc tcgtggccca caaggtccca gtggtcccc cggagtctct      1140
ggaagtctg gatcatcaag gcttggcctc agaggagccc ctggatggcc aggcctgaaa      1200
ggaagtaaag gggaacgagg ccgcccagga aaggatgcca tggggactcc tgggtccca      1260

ggttgtgctg gttcaccagg tcttcagga tcaccgggac ctccaggacc gccagtgac      1320
atcgtttttc gcaagggtcc acctggagat cacggactgc caggctatct aggtctcca      1380
ggaatcccag gagtgtatgg gccc aaagga gaaccaggcc tcctgtgtac acagtgcct      1440
tatatcccag ggcctcccgg tctcccagga ttgccagggt tacaatggtgt aaaaggaatc      1500
ccaggaagac aaggcgcagc tggcttgaaa ggaagcccag ggtccccagg aaatacaggt      1560
cttccaggat ttccaggttt cccaggtgcc cagggtgacc caggacttaa aggagaaaaa      1620

```

ggtgaaacac ttcagcctga ggggcaagtg ggtgtcccag gtgacccggg gctcagaggc 1680  
  
 caacctggga gaaagggctt ggatggaatt cctggaactc cgggagtgaaggattacca 1740  
 ggacctaaag gcgaactggc tctgagtggg gagaaagggg accaaggtcc tccaggggat 1800  
 cctggctccc ctgggtcccc aggacctgca ggaccagctg gaccacctgg ctacggaccc 1860  
 caaggagaac ctggtctcca gggcacgcaa ggagtctctg gagcccccg accaccgga 1920  
 gaagccggcc ctaggggaga gctcagtgtt tcaacaccag tccaggccc accaggacct 1980  
 ccagggcccc ctggccatcc tggccccca ggtccacctg gtatccctgg atccctgggg 2040  
 aatgtggag atcctggtct tccagggcct gatggtgaac caggaattcc aggaattgga 2100  
  
 tttcctgggc ctctggacc taaggagac caaggtttc cagtlacaaa aggatcactg 2160  
 ggttgtcctg gaaaaatggg agagcctggg ttacctggaa agccaggcct cccaggagcc 2220  
 aagggagaac cagcagtagc catgcttggg ggaccaggaa caccaggttt tccaggagaa 2280  
 agaggcaatt ctggggaaca tggagaaatt ggactccctg gacttccagg tctccctgga 2340  
 actccaggaa atgaagggt tgatggacca cgaggagatc cagggcagcc tggaccacct 2400  
 ggagaacaag gacccccagg aaggtgcata gagggtccca ggggagccca aggacttcca 2460  
 ggcttaaatg gattgaaagg gcaacaaggc agaagaggta aaacggggcc aaagggagac 2520  
  
 ccaggaattc caggcttga tagatcagga tttcctggag aaactggatc accaggaatt 2580  
 ccaggctatc aaggtgaaat gggaccactg ggtcaaagag gatatccagg aaatccggga 2640  
 attttagggc caccaggtga agatggagtg attgggatga tgggctttcc tggagccatt 2700  
 ggccctccag ggccccctgg gaaccaggc acaccagggc agagggggag ccttggaaatt 2760  
 ccaggagtaa agggccagag aggaacccca ggagccaagg gggaacaagg agataaagga 2820  
 aatccgggc ctccagagat atcccagta ataggggaca aaggagaacc aggtctcaaa 2880  
 ggattcgag gaaatccagg tgagaaagga aacagaggcg tccagggat gccaggttta 2940  
  
 aagggcctca aaggactacc cggaccagca ggaccaccag gccccagagg agatttgggc 3000  
 agcactggga atcctggaga accaggactg cgtggtatac caggaagcat ggggaacatg 3060  
 ggcattgccag gtictaaagg aaaaagggga actttgggat tcccaggctg agcaggaaga 3120  
 ccaggcctcc caggtattca tggctccag ggagataagg gagagccagg ttattcagaa 3180  
 ggtacaaggc caggaccacc gggaccaacg ggggatccag gactgccggg tgatatggga 3240  
 aagaaaggag aaatgggca acctggccca cctggacatt tgggcctgc tggacctgag 3300  
 ggagcccctg gaagtctgg aagtctggc ctcccaggaa agccaggtcc tcatggtgat 3360

ttgggtttta aaggaatcaa aggcctcctg ggcctccag gaatcagagg ccctccaggt 3420  
 cttccaggat ttccaggatc tcttggacca atgggtataa gaggtgacca aggacgtgat 3480  
 ggaattcctg gtccagccgg agaaaaggga gaaacgggtt tattgagggc ccctccaggc 3540  
 ccaagaggga accctgggtc tcaaggagcc aaaggagaca ggggagcccc aggttttctt 3600  
 ggcctcccgg gcagaaaagg ggccatggga gatgctggac ctgaggacc cacaggcata 3660  
 gaaggattcc cagggccacc aggtctgccc ggtgcaatta tccctggcca gacaggaaat 3720  
 cgtggtccac caggctcaag aggaagccca ggtgcgctg gtccccctgg acctccaggg 3780  
  
 agtcatgtaa taggcataaa aggagacaaa ggtctatgg gccaccctgg cccaaaaggt 3840  
 ccacctggaa ctgcaggaga catgggacca ccaggctgctc tgggagcacc aggtactcca 3900  
 ggtcttccag gaccagagg tgatcctgga ttccaggggt ttccaggcgt gaaaggagaa 3960  
 aagggtaatc ctggatttct aggatccatt ggacctccag gaccaattgg gccaaaagga 4020  
 ccacctggtg tacgtggaga ccctggcaca cttaaagatta tctccttcc aggaagccca 4080  
 gggccacctg gcacacctgg agaaccaggg atgcaggag aacctgggcc accagggccca 4140  
 cctggaaac taggacctg tggccaaga ggtaaagccag gcaaggatgg aaaaccagga 4200  
  
 actcctggac cagctggaga aaaaggcaac aaaggttcta aaggagagcc aggaccagct 4260  
 ggatcagatg gattgccagg tttgaaagga aaactggag acagtggatc acctgcaacc 4320  
 tggacaacga gaggccttgt ctacccega cacagtcaa ccacagcaat tcttcatgt 4380  
 ccagagggga cagtgccact ctacagtggg ttttctttc ttttgtaca aggaaatcaa 4440  
 cgagcccacg gacaagacct tggaaactct ggcagctgcc tgcagcgatt taccacaatg 4500  
 ccattcttat tctgcaatgt caatgatgta tgtaattttg catctgaaa tgattattca 4560  
 tactgctgt caacaccagc tctgatgcca atgaacatgg ctcccattac tggcagagcc 4620  
  
 cttgagcctt atataagcag atgcactgtt tgtgaaggtc ctgcatcgc catagccgtt 4680  
 cacagccaaa cactgacat tcttccatgt cctcacggct ggatttctct ctggaaagga 4740  
 ttttattca tcatgttac aagtgcaggt tctgagggca ccgggcaagc actggcctcc 4800  
 cctggctcct gectggaaga attccgagcc agcccattc tagaatgtca tggagagga 4860  
 acgtgcaact actattcaaa ttctacagt ttctggctgg cttcattaaa ccagaaaga 4920  
 atgttcagaa agcctattcc atcaactgtg aaagctgggg aattagaaaa aataataagt 4980  
 cgctgtcagg tgtgcatgaa gaaaagacac tga 5013

<210> 25

<211> 5073

<212> DNA

<213> Unknown

<220><223> Exemplary COL4A4 transgene sequence

<400> 25

```

atgtggtctc tgcacatagt actaatgagg tgctccttca gattgaccaa gtccttggcc      60
acaggtccct ggtcacttat actcattctc ttttctgtac aataigtata tgggagtgga      120
aagaaataca ttggtccttg tggaggaaga gattgctctg tttgccactg tgttctgaa      180
aaggggtctc ggggtccacc aggaccacca gggccacagg gtccaattgg acccctggga      240
gccccaggac ccattgggct ttcaggagag aaaggaatga gaggggaccg cggccctcct      300
ggagcagcag gggacaaaagg agataagggt ccaactgggt ttctctggatt tccaggttta      360

gatggcatac ctgggcaccc agggcctcct ggaccagag gcaaacctgg tatgagtggc      420
cacaatggct caagaggtga cccagggttt ccaggaggaa gaggagctct tggcccagga      480
ggccccctag gccatcctgg ggaaaaggga gaaaaaggaa attcagtgtt cattttaggt      540
gccgttaaag gtattcaggg agacagaggg gaccaggac tgcttggctt accaggatct      600
tggggtgcag gaggaccggc aggtcccaca ggatatctg gagagccagg gttagtggga      660
cctccgggcc aaccaggcg tccaggtttg aagggaaatc ccggtgtggg agtaaagggg      720
caaatgggag acccgggtga ggttggctag caaggttctc ctggaccac cctgttggtg      780

gagccacctg acttttgtct ctataaagga gaaaagggtg taaaaggaat tcttggaatg      840
gttgactgc caggaccacc aggacgaag ggagaatctg gtattggggc aaaaggagaa      900
aaaggtattc ctggatttcc agggcctcgg ggggatcctg gttcctatgg atctccaggt      960
tttccaggat taaagggaga actaggactg gttggagatc ctgggctatt tggattaatt      1020
ggcccaaagg gggatcctgg aaatcgaggg caccaggac caccaggtgt tttggtgact      1080
ccacctctc cactcaaagg cccaccaggg gaccagggt tccttggcgg ctatggagaa      1140
acaggggatg ttggaccacc tggccccca ggtctcttgg gcagaccagg ggaagcctgt      1200

gcaggcatga taggaccccc tgggccacaa ggatttctg gtcttcttgg gcttccagga      1260
gaagctggta ttctgggag acctgattct gctccaggaa aaccaggga gccaggatca      1320
cctggcttgc ctggagcacc aggcctgcag gcctcccag gatcaagtgt gatatactgt      1380
agtgttggga accccggacc acaaggaata aaaggcaaag ttggtcccc aggaggaaga      1440
ggcccaaaag gagaaaaagg aaatgaagga ctctgtgcct gtgagcctgg acctatgggc      1500
ccccctggcc ctccaggact tcttgggagg caggggagta agggagactt ggggctccct      1560
ggctggcttg gaacaaaagg tgaccagga cctcctgggt ctgaaggacc tccagggcta      1620

ccaggaaagc atggtgcctc tggaccacct ggcaacaaag gggcgaagg tgacatggtt      1680

```

gatatcaagag ttaaaggcca caaaggagaa agaggctctg atgggcccc aggatttcca 1740  
 gggcagccag gatcacatgg tccggatgga catgctggag aaaaaggga tccaggacct 1800  
 ccaggggatc atgaagatgc gaccccaggt ggtaaaggat ttcttgacc tctgggcccc 1860  
 ccaggcaaag caggacctgt ggggccccca ggactgggat ttcttggtcc accaggagag 1920  
 cgaggccacc caggagtcc aggccacca ggtgtgagg gccctgatgg cttgaagggt 1980  
 cagaaagggt acacaatttc ttgcaacgta acctacctg ggaggcatgg cctccaggt 2040  
  
 ttgatggac ctccaggtcc gaagggattt ccaggtcccc aaggtgcccc tgggctgagt 2100  
 ggttcagat ggcataaagg cagacctggc acaccaggaa cagcggaat accaggtcca 2160  
 cctggttttc gtggtgacat gggagatccg ggttttgag gtgaaaagg gtcctccct 2220  
 gttggcccc caggccctcc cggctacca ggagtgaatg gtcagaaagg aatcccggga 2280  
 gacctgcat ttgtcacct gggaccccc ggaaaggagg gtccttcagg agtgccagg 2340  
 ataaaaggac ccagaggatga tccgggatgt ccagggctg aaggccagc tggcattcct 2400  
 ggattcctag gtctcaaagg tcccaaaggc agagaggac atgctgggtt tccaggtgtc 2460  
  
 ccaggtccac ctggccattc ctgtgaaaga ggtgctccag ggataccagg gcaaccggga 2520  
 ctccctgggt atccaggtag cccaggtgct ccaggtggga aaggacagcc gggagatgtg 2580  
 gggcctccc ggccagctgg aatgaaaggc ctccccggac tcccaggac gctggggca 2640  
 catggtcccc caggcctccc aggaatccca ggtcccttg gagatgatgg gctacctgt 2700  
 cctccaggtc caaaggacc ccgggggctg cctggtttcc caggtttcc cggagaaaga 2760  
 ggaaagcctg gtcagaggg atgtcctggc gcaaaggag aacctggaga gaaggcatg 2820  
 tctggccttc ctggagaccg gggactgaga gggccaaag gagccatagg acctccgga 2880  
  
 gatgaaggag aaatggctat catttcaca aagggaacac ctggggaacc tggacctcct 2940  
 ggagatgatg gattcccagg agaaagaggt gataaaggaa ctcccggat gcaagggaga 3000  
 agaggagagc cggaagata cggaccact ggatttcaca gaggggaacc tggtagaaa 3060  
 ggtcagccag ggcctcctgg acccccagc cctccaggct caactggtct aagagggttc 3120  
 attggttttc caggacttcc aggtgaccag ggtgagccag gttctccagg tcccctgga 3180  
 ttttcaggaa ttgatggagc aagaggacct aaaggaaca aaggtgacct tgccagtcac 3240  
 tttggtccac ctggtcaaaa ggtgagcca ggtagcctg gatgtccagg gcattttgga 3300  
  
 gcatccggag agcagggtt gcctggtatt caaggccca gaggatcacc tggaggcca 3360  
 gggccacctg gtcctctgg accaccagg tgcccaggat atcacggat gcctgggctg 3420  
 aggggacagc caggagaaat gggagacct gggccaagag gcctccagg ggatccagg 3480  
 ataccaggtc ctccgggat aaaaggctc tccgatcac ctggcctgaa cgcttgcat 3540

ggattgaaag gtcagaaagg aactaaaggt gcttcaggtt tgcatgatgt ggggccacct 3600  
 ggtccagtgg gaatacctgg gctaaaaggg gagagaggag accctgggag cccaggaatc 3660  
 tctcctccag gtctctgtgg aaagaaaggt cccccaggac ccccaggagag ttcaggacca 3720  
  
 cctggctctg caggtgccac aggaagagct cctaaggaca ttcctgacct ggtccacct 3780  
 ggagatcagg gacctctgg tctgatggc ccaagaggag cacctgggcc tccaggcctc 3840  
 cctgggagtg ttgaccttct gagaggggag ccaggtgact gtggtctacc agggccacca 3900  
 ggtccccctg gccaccagg ccctccagga tacaaggct ttccaggatg tgatggaaaa 3960  
 gatggccaga aaggaccagt gggattcccg ggaccgcagg gaccacatgg atttctctgg 4020  
 ccacctggag agaagggttt acctggacct ccaggagaa aaggccacc tggcttccg 4080  
 ggtccagag gtgaaccggg gccacctgca gatgtgatg actgtccccg aatcccaggc 4140  
  
 ctctctgggg cgccaggcat gagaggacca gaaggagcca tggggctccc tggaatgaga 4200  
 ggcccctcag gaccagggtg caaaggagag cctgggctgg atggcaggag ggtgtggat 4260  
 ggcgctcctg ggtctctgg gcctcccga cgtaaagggtg acacaggaga agacggctac 4320  
 cctggaggac cagggcctcc tggteccatt ggggatctg ggcccaaagg gtttggcct 4380  
 ggatacctcg gtggcttct cctggttctc cacagtcaga cggaccagga gccacctgc 4440  
 cccctgggca tgcccaggct ctggactggg tatagtctgt tatacctgga agggcaagag 4500  
 aaagtcaca atcaagacct tggctctggca gggtcttgcc tteccgtatt tagcacgctg 4560  
  
 ccctttgctt actgcaacat ccaccagggtg tgccactatg cccagagaaa cgacagatcc 4620  
 tactggctgg ccagcgtgc gccctcccc atgatgccac tctctgaaga ggcgatccgc 4680  
 ccctatgtca gccgctgtgc ggtatgcgag gccccggccc aggcggtggc ggtgcacagc 4740  
 caggaccagt ccatcccc atgtccgag acctggagga gcctctggat cgggtattca 4800  
 ttctgatgc acacaggagc tggggaccaa ggaggaggc aggccttat gtcacctggc 4860  
 agctgcctgg aagatttcag agcagacca ttcttgaat gccagggccg gcagggaact 4920  
 tgccactttt tcgcaaataa gtatagcttc tggtcaciaa cggtgaaagc agacttgcag 4980  
  
 tttctctg ctccagcacc agacacctta aaagaaagcc aggcccaacg ccagaaaatc 5040  
 agccggtgcc aggtctgctg gaagtatagc tag 5073  
  
 <210> 26  
 <211> 5058  
 <212> DNA  
 <213> Unknown  
 <220><223> Exemplary COL4A5 transgene sequence

<400> 26

atgaaactgc gtggagtcag cctggctgcc ggcttgttct tactggcctc gagtctttgg 60  
 gggcagcctg cagaggctgc ggcttgctat ggggtgttctc caggatcaaa gtgtgactgc 120  
 agtggcataa aaggggaaaa gggagagaga gggtttccag gtttggaaagg acaccaggga 180  
 ttgcctggat ttccaggfcc agaagggcct cggggcctc ggggacaaaa gggatgatgat 240  
  
 ggaattccag ggcaccagg accaaaagga atcagaggtc ctctggact tectggattt 300  
 ccagggacac caggtcttcc tggaatgcca ggccacgatg gggccccagg acctcaaggt 360  
 attcccggat gcaatggaac caagggagaa cgtggatttc caggcagtcc cggttttctc 420  
 ggtttacagg gtctccagg acccctggg atcccaggta tgaagggtga accaggtagt 480  
 ataattatgt catcactgcc aggaccaaag ggtaatccag gatatccagg tctcctgga 540  
 atacaaggcc tacctggtcc cactggtata ccagggcaa ttggtcccc aggaccacca 600  
 ggtttgatgg gcctcctgg tccaccagga ctccaggac ctaaggggaa tatgggctta 660  
  
 aatttccagg gacccaaagg tgaaaaaggt gagcaaggtc ttcagggcc acctgggcca 720  
 cctgggcaga tcagtgaaca gaaaagacca attgatgtag agtttcagaa aggagatcag 780  
 ggacttctg gtgaccagg gcctcctgga cctccaggga tacgtggtcc tccaggtccc 840  
 ccaggtggtg agaaaggtga gaagggtgag caaggagagc caggcaaaag aggtaacca 900  
 ggcaaagatg gagaaaaagg ccaaccagga attcctggtt tgctggtga tctggttac 960  
 cctggtgaac ccggaagga ttggtaaaag ggccaaaaag gtgacactgg cccacctgga 1020  
 cctcctggac ttgtaattcc tagacctggg actggtataa ctataggaga aaaaggaaac 1080  
  
 attgggttgc ctgggttgc tggagaaaa ggagagcgag gatttctgg aatacagggt 1140  
 ccacctggcc ttctggacc tccaggggct gcagttatgg gtctcctgg ccctcctgga 1200  
 tttcctggag aaaggggtca gaaaggtgat gaaggaccac ctggaatttc cattcctgga 1260  
 cctcctggac ttgacggaca gcctggggct cctgggcttc caggcctcc tggccctgct 1320  
 ggccctcaca ttctcctag tgatgagata tgtgaaccag gccctccagg cccccagga 1380  
 tctccagggtg ataaaggact ccaaggagaa caaggagtga aaggtgacaa aggtgacact 1440  
 tgcttcaact gcattggaac tggatattca gggcctccag gtcaacctgg ttgcccagg 1500  
  
 ctcccaggtc ctccaggatc tcttggtttc cctggacaga aaggggaaaa aggacaagct 1560  
 ggtgcaactg gtccaaagg attaccaggc attccaggag ctccagggtc tccaggcttt 1620  
 cctggatcta aaggtgaacc tggatgatac ctacttttc caggaatgaa gggtgacaaa 1680  
 ggagagttag gtccccctgg agctccaggg ctctcctggt tacctggcac tectggacag 1740  
 gatggattgc cagggcttcc tggcccgaag ggagagcctg gtggaattac ttttaagggt 1800

gaaagaggtc cccctgggaa cccaggttta ccaggcctcc cagggaaatag agggcctatg 1860  
 ggtccccctg gtttcggccc tccaggccca gtaggtgaaa aaggcataca aggtgtggca 1920  
  
 ggaaatccag gccagccagg aataccaggt cctaaagggg atccaggtca gactataacc 1980  
 cagccgggga agcctggctt gcctggtaac ccaggcagag atggtgatgt aggtcttcca 2040  
 ggtgaccctg gacttccagg gcaaccaggc ttgccaggga tacctggtag caaaggagaa 2100  
 ccaggatate ctggaattgg gcttctctga ccacctggtc ccaaaggctt tectggaatt 2160  
 ccaggacctc caggagcacc tgggacacct ggaagaattg gtctagaagg cctcctggg 2220  
 ccaccggctt tccaggacc aaagggtgaa ccaggatttg cattacctgg gccacctggg 2280  
 ccaccaggac tccagggtt caaaggagca cttggtccaa aaggtgatcg tggtttccca 2340  
  
 ggacctccgg gtctccagg acgcaactggc ttagatgggc tccctggacc aaaaggtgat 2400  
 gttggaccaa atggacaacc tggaccaatg ggacctctg ggctgccagg aataggtgtt 2460  
 cagggaccac caggaccacc agggattcct gggccaatag gtcaacctgg ttacatgga 2520  
 ataccaggag agaaggggga tccaggacct cctggacttg atgttccagg acccccaggt 2580  
 gaaagagca gtccagggat ccccgagca cctggctcta taggacctcc aggatcacca 2640  
 gggcttccag gaaaagcagg tgcctctgga tttccaggta ccaaaggtga aatgggtatg 2700  
 atgggacctc caggcccacc aggacctttg ggaattctg gcaggagtgg tgtacctggt 2760  
  
 cttaaaggtg atgatggctt gcagggtcag ccaggacttc ctggcctac aggagaaaaa 2820  
 ggtagttaaag gagagcctgg ccttccaggc cctcctggac caatggatcc aaatcttctg 2880  
 ggctcaaaag gagagaagg ggaacctggc ttaccaggta tacctggagt ttcagggcca 2940  
 aaaggtatc agggtttggc tggagacca gggcaacctg gactgagtgg acaacctgga 3000  
 ttaccaggac caccaggtcc caaaggtaac cctggtctcc ctggacagcc aggtcttata 3060  
 ggacctctg gacttaaagg aacctcggg gatatgggtt ttcagggcc tcagggtgtg 3120  
 gaaggcctc ctggacctt tggagttcct ggacaacctg gctcccagg attacctgga 3180  
  
 cagaaaggcg acaaaggtga tcttggtatt tcaagcattg gtcttccagg tcttctggt 3240  
 ccaaagggtg agcctggtct gcctggatac ccagggaacc ctggtatcaa aggttctgtg 3300  
 ggagatcctg gtttggccgg attaccagga acccctggag caaaaggaca accaggcctt 3360  
 cctggattcc caggaacecc aggcctcct ggacaaaag gtattagtgg cctcctggg 3420  
 aacccggcc tccaggaga acctggtcct gtaggtggtg gaggtcatcc tgggcaacca 3480  
 gggcctccag gcgaaaaagg caaacccgt caagatggtg tccctggacc agctggacag 3540

aagggtgaac caggtaacc aggccttggga aaccaggac cccctggact tccaggactt 3600  
  
tctggccaaa aggtgatgg aggattacct gggattccag gaaatcctgg ccttccaggt 3660  
ccaaagggcg aaccaggctt tcacggtttc cctgggtgtgc agggcccccc aggccctcct 3720  
ggttctccgg gtccagctct ggaaggacct aaaggcaacc ctgggccccca aggtcctcct 3780  
gggagaccag gtctaccagg tccagaaggt cctccaggtc tccctggaaa tggaggtatt 3840  
aaaggagaga agggaaatcc aggccaacct gggctacctg gcttgcctgg ttgaaagga 3900  
gatcaaggac caccaggact ccagggtaat cctggccggc cgggtctcaa tggaatgaaa 3960  
ggagatcctg gtctccctgg tgttccagga ttcccaggca tgaaaggacc cagtggagta 4020  
  
cctggatcag ctggccctga gggggaaccg ggacttattg gtcctccagg tcctcctgga 4080  
ttacctggtc cttcaggaca gagtatcata attaaaggag atgctggctc tccaggaatc 4140  
cctggccagc ctgggctaaa gggcttacca ggacccaag gacctcaagg cttaccaggt 4200  
ccaactggcc ctccaggaga tcctggacgc aatggactcc ctggctttga tggatgcagga 4260  
gggcgcaaag gagaccagg tctgccagga cagccaggta cccgtggttt ggatggctcc 4320  
cctggctcag atggattgca aggtccccca ggtccccctg gaacctctc tgttgacat 4380  
ggatttctta ttacacgcca cagccagaca acggatgcac cacaatgcc acaggaaca 4440  
  
cttcaggctt atgaagctt ttctctctg tatgtacaag gaaataaaag agccccaggt 4500  
caagacttgg ggacggctgg cagctgcctt cgtcgcttta gtaccatgcc tttcatgttc 4560  
tgcaacatca ataagtgtg caactttgct tcaagaaatg actattctta ctggctctct 4620  
accccagagc ccatgccaat gagcatgcaa ccctaaagg gccagagcat ccagccattc 4680  
attagtcgat gtgcagtatg tgaagctcca gctgtggtga tcgcagttca cagtcagacg 4740  
atccagattc cccattgtcc tcagggatgg gattctctgt ggattggttt ttccttcag 4800  
atgcatacaa gtgcaggggc agaaggctca ggtcaagccc tagcctcccc tggttcctgc 4860  
  
ttggaagagt ttcgttcagc tccttcatc gaatgtcatg ggaggggtac ctgtaactac 4920  
tatgccaaact cctacagctt ttggctggca actgtagatg tgtcagacat gttcagtaaa 4980  
cctcagtcag aaacgctgaa agcaggagac ttgaggacac gaattagccg atgtcaagtg 5040  
tgcatagaaga ggacataa 5058  
<210> 27  
<211> 186  
<212> DNA  
<213> Unknown  
<220><223> Exemplary NPHS1 enhancer

<400> 27  
ctgctgagct gggagaccac cttgatctga cttctcccat cttccagcc taagccaggc 60  
cctgggtca cggaggctgg ggaggcaccg aggaacgcgc ctggcatgtg ctgacagggg 120

atztatgct ccagctgggc cagctgggag ggcctgctg ggcagaggcc agagctgggg 180  
gctctg 186

<210> 28  
<211> 13  
<212> DNA  
<213> Exemplary Kozak sequence

<400> 28  
gccgccacca ugg 13

<210> 29  
<211> 589  
<212> DNA  
<213> Unknown

<220><223> Exemplary WPRE

<400> 29  
aatcaacctc tggattacaa aatttgtgaa agattgactg gtattcttaa ctatgttgct 60  
ccttttacgc tatgtggata cgctgcttta atgcctttgt atcatgctat tgcttcccg 120

atggctttca ttttctctc cttgtataaa tcttggttgc tgtctcttta tgaggagt 180  
tggcccgttg tcaggcaacg tggcgtgggtg tgcactgtgt ttgctgacgc aacccccact 240  
ggttggggca ttgccaccac ctgtcagctc ctttccggga ctttcgcttt cccctcct 300  
attgccacgg cggaactcat cgccgcctgc cttgcccgct gctggacagg ggctcggctg 360  
ttgggcactg acaattcctg ggtgtgtcg gggaaatcat cgtcctttcc ttggtgctc 420  
gcctgtgttg ccacctggat tctgcgctgg acgtccttct gctacgtccc ttcggccctc 480  
aatccagcgg accttccttc ccgcggcctg ctgccggctc tgcggcctct tccgctctt 540

cgcttcgcc ctcagacgag tcggatctcc ctttgggccc cctccccgc 589

<210> 30  
<211> 135  
<212> DNA  
<213> Unknown

<220><223> Exemplary SV40pA signal sequence

<400> 30  
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 60  
aataaagcat ttttttcaact gcattctagt tgtggtttgt ccaaactcat caatgtatct 120  
tatcatgtct ggatc 135

<210> 31

<211> 225

<212> DNA

<213> Unknown

<220><223> Exemplary bGH poly(A) signal sequence

<400> 31  
ctgtgcttc tagttgccag ccactctgtt tttgccctc ccccgctcct tccttgacc 60  
tggaagggtc cactcccact gtcctttcct aataaaatga ggaaattgca tcgcattgtc 120  
tgagtaggtg tcattctatt ctgggggggtg ggggggggca ggacagcaag ggggaggatt 180  
gggaagacaa tagcaggcat gctggggatg cggggggctc tatgg 225

<210> 32

<211> 17

<212> DNA

<213> Unknown

<220><223> Exemplary soluble neuropilin-1 polyadenylation signal

<400> 32  
aaataaata cgaaatg 17

<210> 33

<211> 735

<212> PRT

<213> Unknown

<220><223> Exemplary AAV3B VP1 capsid protein

<400> 33

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 Val Pro Gln Pro  
                  20                    25                    30  
Lys Ala Asn Gln Gln His Gln Asp Asn Arg Arg Gly Leu Val Leu Pro



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



1                    5                    10                    15  
 Glu Gly Ile Arg Glu Trp Trp Ala Leu Gln Pro Gly Ala Pro Lys 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 Asp Gln Ser Pro Gln Glu Pro Asp Ser Ser Ser Gly Val Gly  
 145                    150                    155                    160  
 Lys Ser Gly Lys Gln Pro Ala Arg Lys Arg Leu Asn Phe Gly Gln Thr  
                          165                    170                    175  
 Gly Asp Ser Glu Ser Val Pro Asp Pro Gln Pro Leu Gly Glu Pro Pro  
                          180                    185                    190  
 Ala Ala Pro Thr Ser Leu Gly Ser Asn Thr Met Ala Ser Gly Gly Gly  
  
                          195                    200                    205  
 Ala Pro Met Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Asn 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 Ser Gln Ser Gly Ala Ser Asn Asp Asn His Tyr  
 260 265 270  
 Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg Phe His  
 275 280 285  
 Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn Asn Trp  
 290 295 300  
 Gly Phe Arg Pro Lys Lys Leu Ser Phe Lys Leu Phe Asn Ile Gln Val  
 305 310 315 320  
 Lys Glu Val Thr Gln Asn Asp Gly Thr Thr Thr Ile Ala Asn Asn Leu  
 325 330 335  
 Thr Ser Thr Val Gln Val Phe Thr Asp Ser Glu Tyr Gln Leu Pro Tyr  
 340 345 350  
 Val Leu Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe Pro Ala Asp  
 355 360 365  
 Val Phe Met Val Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asn Gly Ser  
 370 375 380  
 Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe Pro Ser  
 385 390 395 400  
 Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Thr Phe Glu  
 405 410 415  
 Asp Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu Asp Arg  
 420 425 430  
 Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Asn Arg Thr  
 435 440 445  
 Gln Gly Thr Thr Ser Gly Thr Thr Asn Gln Ser Arg Leu Leu Phe Ser  
 450 455 460  
 Gln Ala Gly Pro Gln Ser Met Ser Leu Gln Ala Arg Asn Trp Leu Pro  
 465 470 475 480  
 Gly Pro Cys Tyr Arg Gln Gln Arg Leu Ser Lys Thr Ala Asn Asp Asn  
 485 490 495  
 Asn Asn Ser Asn Phe Pro Trp Thr Ala Ala Ser Lys Tyr His Leu Asn

500 505 510  
 Gly Arg Asp Ser Leu Val Asn Pro Gly Pro Ala Met Ala Ser His Lys

515 520 525  
 Asp Asp Glu Glu Lys Phe Phe Pro Met His Gly Asn Leu Ile Phe Gly

530 535 540  
 Lys Glu Gly Thr Thr Ala Ser Asn Ala Glu Leu Asp Asn Val Met Ile

545 550 555 560  
 Thr Asp Glu Glu Glu Ile Arg Thr Thr Asn Pro Val Ala Thr Glu Gln

565 570 575  
 Tyr Gly Thr Val Ala Asn Asn Leu Gln Ser Ser Asn Thr Ala Pro Thr

580 585 590  
 Thr Arg Thr Val Asn Asp Gln Gly Ala 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 His Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Leu

625 630 635 640  
 Lys His Pro Pro Pro Gln Ile Met Ile Lys Asn Thr Pro Val Pro Ala

645 650 655  
 Asn Pro Pro Thr Thr Phe Ser Pro Ala Lys Phe Ala 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 Asn Lys Ser Val Asn Val Asp Phe Thr Val Asp Thr Asn Gly Val

705 710 715 720  
 Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Pro Leu

725 730 735

<210> 35

<211> 736

<212> PRT

<213> Unknown

<220><223> Exemplary AAV9 VP1 capsid protein

<400> 35

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



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



<211> 6  
 <212> DNA  
 <213> Unknown  
 <220><223> transcription factor binding region  
 <400> 37  
 tacgat 6  
 <210> 38  
 <211> 6  
 <212> DNA  
 <213> Unknown  
 <220><223> transcription factor binding region  
 <400> 38  
 tataat 6  
  
 <210> 39  
 <211> 6  
 <212> DNA  
 <213> Unknown  
 <220><223> transcription factor binding region  
 <400> 39  
 gatact 6  
 <210> 40  
 <211> 6  
 <212> DNA  
 <213> transcription factor binding region  
 <400> 40  
 tatgat 6  
 <210> 41  
 <211> 6  
 <212> DNA  
 <213> Unknown  
 <220><223> transcription factor binding region  
 <400> 41  
 tatggt 6

<210> 42  
 <211> 50  
 <212> DNA  
 <213> Unknown  
 <220><223> Exemplary chicken beta-globin polyadenylation signal  
 <400> 42  
 caataaaaga tcittatattt cattagatct gtgtgttggg tttttgtgtg 50  
 <210> 43  
 <211> 30  
 <212> DNA  
 <213> Unknown  
 <220><223> Exemplary transcription factor binding region  
 <400> 43  
 gagcaagaca gagagagata ctcacagga 30  
 <210> 44  
 <211> 1192  
 <212> DNA  
 <213> Unknown  
 <220><223> Exemplary variant nephrin promoter  
 <400> 44  
 cacctgaggt caggagtctg agaccagcgt ggccaacatg atgaaacccc gtctctagta 60  
 aaaatacaaa aattagccag gcatggtgct atatacctgt agcaccagct acttgggaga 120  
 cagaggtggg agaattactt gaacctggga ggttcaagcc atgggagggtg gaagttgcag 180  
 tgagccgaga tgccactgca ctccagcctg agcaacagag caagactatc tcaagaaaag 240  
 aaagaaagaa agaaagagac ttgccaaggt catgtatcag ggcaaggaag agctgggggc 300  
 ccagctggct gtccectcgt tgagctggga gaccaccttg atctgacttc tccatcttc 360  
 ccagcctaag ccaggccctg gggtcacgga ggctggggag gcaccgagga acgcgcctgg 420  
 catgtgctga caggaattt tatgetccag ctgggccagc tgggaggagc ctgctgggca 480  
 gaggccagag ctgggggctc tggaaggtac ctgggggagg ttgactgtg agaatgagct 540  
 caagctgggt cagagagcag ggctgactct gccagtgcct gcatcagcct catcgtctc 600  
 ctagctcct ggctgctgg actctgggct gcaggctctt cttgaaagc tgtgagtagt 660  
 gagacaagga gcaggagtga ggggtggcag gagagaagat agagattgag agagagagag 720  
 agagagagac agagagagag gaagagacag agacaaaagg agagagaacg gcttagacaa 780

ggagagaaag atggaaagat aaagagactg ggcgcagtgg ctcacgcctg taatccaac 840  
 acttggggag gccaaagtg gaggatggct tgaaggaag agtctgagat caacctggcc 900  
 aacatagtga gaccccgctct ctaaaaaaaaa aagaaaaaaaa aaagaaaaaaaa gaaaaaaaaag 960  
 tttttttaa gagacagaga aagagactca gagattgaga ctgagagcaa gacagagaga 1020  
 gacactcaca gggaagaggg gaagaggaaa acgagaaagg gaggagagta acggaagag 1080  
 ataaaaaga aaagcaggtg gcagagacac agagagaggg acccagagaa agccagacag 1140  
 acgcaggtgg ctggcagcgg gcgctgtggg ggtcacagta gggggacctg tc 1192  
 <210> 45  
 <211> 1192  
 <212> DNA  
 <213> Unknown

<220><223> Exemplary variant nephrin promoter

<400> 45  
 cacctgaggt caggagtctg agaccagcgt ggccaacatg atgaaacccc gtctctagta 60  
 aaaatacaaa aattagccag gcatggtgct atatacctgt agcaccagct acttgggaga 120  
 cagaggtggg agaattactt gaacctggga ggttcaagcc atgggaggtg gaagttgcag 180  
 tgagccgaga tgccactgca ctccagcctg agcaacagag caagactatc tcaagaaaag 240  
 aaagaaagaa agaaagagac ttgccaaggt catgtatcag ggcaaggaag agctgggggc 300  
 ccagctggct gctcccctgc tgagctggga gaccaccttg atctgacttc tcccatcttc 360  
  
 ccagcctaag ccaggccctg gggtcacgga ggctggggag gcaccgagga acgcgcctgg 420  
 catgtctgta caggggattt tatgtctcag ctgggccagc tgggaggagc ctgctgggca 480  
 gagccagag ctgggggctc tggaaggtac ctgggggagg ttgactgtg agaatgagct 540  
 caagctgggt cagagagcag ggctgactct gccagtgcct gcatcagcct catcgctctc 600  
 ctaggtctct ggctgtctgg actctgggct gcaggtcctt cttgaaagc tgtgagtgt 660  
 gagacaagga gcaggagtga ggggtggcag gagagaagat agagattgag agagagagag 720  
 agagagagac agagagagag gaagagacag agacaaaagg agagagaacg gcttagacaa 780  
  
 ggagagaaag atggaaagat aaagagactg ggcgcagtgg ctcacgcctg taatccaac 840  
 acttggggag gccaaagtg gaggatggct tgaaggaag agtctgagat caacctggcc 900  
 aacatagtga gaccccgctct ctaaaaaaaaa aagaaaaaaaa aaagaaaaaaaa gaaaaaaaaag 960  
 tttttttaa gagacagaga aagagactca gagattgaga ctgagagcaa gacagagaga 1020  
 gatactcaca gggaagaggg gaagaggaaa acgagaaagg gaggagagta acggaagag 1080

ataaaaaaga aaagcaggtg gcagagacac agagagaggg acccagagaa agccagacag 1140  
 acgcaggtgg ctggcagcgg gcgctgtggg ggtcacagta gggggacctg tc 1192

<210> 46

<211> 819

<212> DNA

<213> Unknown

<220><223> Exemplary minimal nephrin promoter variant

<400> 46

ggccctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag 60  
 gggattttat gctccagctg ggccagctgg gaggagcctg ctgggcagag gccagagctg 120  
 ggggctctgg aaggtacctg ggggaggttg cactgtgaga atgagctcaa gctgggtcag 180  
 agagcagggc tgactctgcc agtgctgca tcagcctcat cgctctccta ggctcctggc 240  
 ctgctggact ctgggctgca ggtccttctt gaaaggctgt gtagtagag acaaggagca 300  
 ggagtgaggg gtggcaggag agaagataga gattgagaga gagagagaga gagacagaga 360

gagaggaaga gacagagaca aaaggagaga gaacggctta gacaaggaga gaaagatgga 420  
 aagataaaga gactgggcgc agtggtcac gcctgtaatc ccaacacttg gggaggccaa 480  
 ggtgggagga tgcttgaag gaaagagtct gagatcaacc tggccaacat agtgagacct 540  
 cgtctctaaa aaaaaaaaaa aaaaaaaaaa gaaaaaagaa aaaaaagttt ttttaagag 600  
 acagagaaag agactcagag attgagactg agagcaagac agagagagat actcacaggg 660  
 aagaggggaa gaggaaaacg agaaaggag gagagtaacg gaaagagata aaaaagaaaa 720  
 gcaggtggca gagacacaga gagaggacc cagagaaagc cagacagacg caggtggctg 780

gcagcgggcg ctgtgggggt cacagtaggg ggacctgtc 819

<210> 47

<211> 265

<212> DNA

<213> Unknown

<220><223> Exemplary minimal nephrin promoter variant

<400> 47

ggccctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag 60  
 gggattttat gctccaggag caagacagag agagatactc acaggggaaga ggggaagagg 120  
 aaaacagaaa agggaggaga gtaacgaaa gagataaaaa agaaaagcag gtggcagaga 180  
 cacacagaga gggaccaga gaaagccaga cagacgcagg tggctggcag cgggcgctgt 240

gggggtcaca gtagggggac ctgtg

265

<210> 48

<211> 0

<212> DNA

<213> Unknown

<220><223> n/a

<400> 48

000

<210> 49

<211> 583

<212> PRT

<213> Unknown

<220><223> Illustrative CFI polypeptide sequence

<400> 49

Met Lys Leu Leu His Val Phe Leu Leu Phe Leu Cys Phe His Leu Arg

1                    5                    10                    15

Phe Cys Lys Val Thr Tyr Thr Ser Gln Glu Asp Leu Val Glu Lys Lys

                  20                    25                    30

Cys Leu Ala Lys Lys Tyr Thr His Leu Ser Cys Asp Lys Val Phe Cys

                  35                    40                    45

Gln Pro Trp Gln Arg Cys Ile Glu Gly Thr Cys Val Cys Lys Leu Pro

50                    55                    60

Tyr Gln Cys Pro Lys Asn Gly Thr Ala Val Cys Ala Thr Asn Arg Arg

65                    70                    75                    80

Ser Phe Pro Thr Tyr Cys Gln Gln Lys Ser Leu Glu Cys Leu His Pro

                  85                    90                    95

Gly Thr Lys Phe Leu Asn Asn Gly Thr Cys Thr Ala Glu Gly Lys Phe

                  100                    105                    110

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

115                    120                    125

Val Lys Leu Val Asp Gln Asp Lys Thr Met Phe Ile Cys Lys Ser Ser

130                    135                    140

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

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

<210> 50

<211> 1752

<212> DNA

<213> Unknown

<220><223> Illustrative CFI polynucleotide sequence

<400> 50

atgaagcttc ttcattgtttt cctgttattt ctgtgcttcc acttaagggtt ttgcaaggtc

60

acttatacat ctcaagagga tctgggtggag aaaaagtgct tagcaaaaaa atatactcac 120

ctctcctgcg ataaagtctt ctgccagcca tggcagagat gcattgaggg cacctgtgtt 180

tgtaaaactac cgtatcagtg cccaaagaat ggcaactgcag tgtgtgcaac taacaggaga 240

agcttcccaa catactgtca acaaaagagt ttggaatgtc ttcattccagg gacaaagttt 300

ttaaataacg gaacatgcac agccgaagga aagtttagtg tttccttgaa gcatggaaat 360

acagattcag agggaatagt tgaagtaaaa cttgtggacc aagataagac aatgttcata 420

tgcaaaagca gctggagcat gaggaagcc aacgtggcct gccttgacct tgggtttcaa 480

caaggtgctg atactcaag aaggtttaag ttgtctgac tctctataaa ttccactgaa 540

tgtctacatg tgcattgccg aggattagag accagtttgg ctgaatgtac ttttactaag 600

agaagaacta tgggttacca ggatttcgct gatgtggttt gttatacaca gaaagcagat 660

tctccaatgg atgacttctt tcagtggtg aatgggaaat acatttctca gatgaaagcc 720

tgtgatggta tcaatgattg tggagaccaa agtgatgaac tgtgttgtaa agcatgccaa 780

ggcaaaggct tccattgcaa atcgggtgtt tgcattccaa gccagtatca atgcaatggt 840

gaggtggact gcattacagg ggaagatgaa gttggctgtg caggctttgc atctgtgact 900

caagaagaaa cagaaatfff gactgctgac atggatgcag aaagaagacg gataaaatca 960

ttattaccta aactatcttg tggagttaa aacagaatgc acattcgaag gaaacgaatt 1020

gtgggaggaa agcgagcaca actgggagac ctcccatggc aggtggcaat taaggatgcc 1080

agtggaaatca cctgtggggg aatttatatt ggtggctgtt ggattctgac tgctgacat 1140

tgtctcagag ccagtaaaac tcatcgttac caaataatgga caacagtagt agactggata 1200

caccccgacc ttaaacgat agtaattgaa tacgtggata gaattatfff ccatgaaaac 1260

tacaatgcag gcaattacca aatgacatc gctttgattg aatgaaaaa agacggaaaac 1320

aaaaaagatt gtgagctgcc tcgttccatc cctgcctgtg tcccctggtc tecttaccta 1380

ttccaaccta atgatacatg catcgtttct ggctggggac gaaaaaaga taacgaaaga 1440

gtcttttcac ttcagtgggg tgaagttaa ctaataagca actgctctaa gttttacgga 1500

aatcgtttct atgaaaaaga aatggaatgt gcaggtacat atgatggttc catcgatgcc 1560

tgtaaaaggg actctggagg ccccttagtc tgtatggatg ccaacaatgt gacttatgtc 1620

tgggtgtgtg tgagttgggg ggaaaactgt ggaaaaccag agttcccagg tgtttacacc 1680

aaagtggcca attatfffga ctggattagc taccatgtag gaaggcfff tatttctcag 1740

tacaatgtat aa 1752

<210> 51

<211> 1231

<212> PRT

<213> Unknown

<220><223> Illustrative CFH polypeptide sequence

<400> 51

Met Arg Leu Leu Ala Lys Ile Ile Cys Leu Met Leu Trp Ala Ile Cys

1                    5                    10                    15

Val Ala Glu Asp Cys Asn Glu Leu Pro Pro Arg Arg Asn Thr Glu Ile

                  20                    25                    30

Leu Thr Gly Ser Trp Ser Asp Gln Thr Tyr Pro Glu Gly Thr Gln Ala

                  35                    40                    45

Ile Tyr Lys Cys Arg Pro Gly Tyr Arg Ser Leu Gly Asn Val Ile Met

50                    55                    60

Val Cys Arg Lys Gly Glu Trp Val Ala Leu Asn Pro Leu Arg Lys Cys

65                    70                    75                    80

Gln Lys Arg Pro Cys Gly His Pro Gly Asp Thr Pro Phe Gly Thr Phe

                  85                    90                    95

Thr Leu Thr Gly Gly Asn Val Phe Glu Tyr Gly Val Lys Ala Val Tyr

                  100                    105                    110

Thr Cys Asn Glu Gly Tyr Gln Leu Leu Gly Glu Ile Asn Tyr Arg Glu

115                    120                    125

Cys Asp Thr Asp Gly Trp Thr Asn Asp Ile Pro Ile Cys Glu Val Val

130                    135                    140

Lys Cys Leu Pro Val Thr Ala Pro Glu Asn Gly Lys Ile Val Ser Ser

145                    150                    155                    160

Ala Met Glu Pro Asp Arg Glu Tyr His Phe Gly Gln Ala Val Arg Phe

                  165                    170                    175

Val Cys Asn Ser Gly Tyr Lys Ile Glu Gly Asp Glu Glu Met His Cys

180                    185                    190

Ser Asp Asp Gly Phe Trp Ser Lys Glu Lys Pro Lys Cys Val Glu Ile

195                    200                    205

Ser Cys Lys Ser Pro Asp Val Ile Asn Gly Ser Pro Ile Ser Gln Lys  
 210 215 220  
 Ile Ile Tyr Lys Glu Asn Glu Arg Phe Gln Tyr Lys Cys Asn Met Gly  
 225 230 235 240  
  
 Tyr Glu Tyr Ser Glu Arg Gly Asp Ala Val Cys Thr Glu Ser Gly Trp  
 245 250 255  
 Arg Pro Leu Pro Ser Cys Glu Glu Lys Ser Cys Asp Asn Pro Tyr Ile  
 260 265 270  
 Pro Asn Gly Asp Tyr Ser Pro Leu Arg Ile Lys His Arg Thr Gly Asp  
 275 280 285  
 Glu Ile Thr Tyr Gln Cys Arg Asn Gly Phe Tyr Pro Ala Thr Arg Gly  
 290 295 300  
  
 Asn Thr Ala Lys Cys Thr Ser Thr Gly Trp Ile Pro Ala Pro Arg Cys  
 305 310 315 320  
 Thr Leu Lys Pro Cys Asp Tyr Pro Asp Ile Lys His Gly Gly Leu Tyr  
 325 330 335  
 His Glu Asn Met Arg Arg Pro Tyr Phe Pro Val Ala Val Gly Lys Tyr  
 340 345 350  
 Tyr Ser Tyr Tyr Cys Asp Glu His Phe Glu Thr Pro Ser Gly Ser Tyr  
 355 360 365  
  
 Trp Asp His Ile His Cys Thr Gln Asp Gly Trp Ser Pro Ala Val Pro  
 370 375 380  
 Cys Leu Arg Lys Cys Tyr Phe Pro Tyr Leu Glu Asn Gly Tyr Asn Gln  
 385 390 395 400  
 Asn Tyr Gly Arg Lys Phe Val Gln Gly Lys Ser Ile Asp Val Ala Cys  
 405 410 415  
 His Pro Gly Tyr Ala Leu Pro Lys Ala Gln Thr Thr Val Thr Cys Met  
 420 425 430  
  
 Glu Asn Gly Trp Ser Pro Thr Pro Arg Cys Ile Arg Val Lys Thr Cys  
 435 440 445  
 Ser Lys Ser Ser Ile Asp Ile Glu Asn Gly Phe Ile Ser Glu Ser Gln







Leu Tyr Ser Arg Thr Gly Glu Ser Val Glu Phe Val Cys Lys Arg  
 1190 1195 1200  
 Gly Tyr Arg Leu Ser Ser Arg Ser His Thr Leu Arg Thr Thr Cys  
 1205 1210 1215  
 Trp Asp Gly Lys Leu Glu Tyr Pro Thr Cys Ala Lys Arg  
 1220 1225 1230

<210> 52

<211> 3696

<212> DNA

<213> Unknown

<220><223> Illustrative CFH polynucleotide sequence

<400> 52

atgagacttc tagcaaagat tatttgcctt atgttatggg ctatttgtgt agcagaagat 60

tgcaatgaac ttctccaag aagaaataca gaaattctga caggttctg gctgaccaa 120

acatatccag aaggcaccca ggctatctat aaatgccgcc ctggatatag atctcttgga 180

aatgtaataa tggatgcag gaagggagaa tgggttgctc ttaatccatt aaggaaatgt 240

cagaaaagc cctgtggaca tcttgagat actccttttg gtaactttac cttacagga 300

ggaaatgtgt ttgaatatgg tgtaaaagct gtgtatacat gtaatgaggg gtatcaattg 360

ctaggtgaga ttaattaccg tgaatgtgac acagatggat ggaccaatga tattcctata 420

tgtgaagtgg tgaagtgttt accagtgaca gcaccagaga atggaaaaat tgtcagtagt 480

gcaatggaac cagatcggga ataccatttt ggacaagcag tacggtttgt atgtaactca 540

ggctacaaga ttgaaggaga tgaagaaatg cattgttcag acgatggttt ttggagtaaa 600

gagaaaccaa agigtgtgga aatttcatgc aaatccccag atgttataaa tggatctcct 660

atatctcaga agattathta taaggagaat gaacgatttc aatataaatg taacatgggt 720

tatgaataca gtgaaagagg agatgctgta tgcactgaat ctggatggcg tccgttgctt 780

tcatgtgaag aaaaatcatg tgataatcct tatattccaa atggtgacta ctcaccttta 840

aggattaaac acagaactgg agatgaaatc acgtaccagt gtagaaatgg tttttatcct 900

gcaacccggg gaaatacagc aaaatgcaca agtactggct ggatacctgc tccgagatgt 960

accttgaac ctgtgatta tccagacatt aaacatggag gtctatatca tgagaatag 1020

cgtagaccat actttccagt agctgtagga aaatattact cctattactg tgatgaacat 1080

tttgagactc cgtcaggaag ttactgggat cacattcatt gcacacaaga tggatggtcg 1140

ccagcagtac catgcctcag aaaatgttat tttccttatt tggaaaatgg atataatcaa 1200

aatcatggaa gaaagtttgt acagggtaaa tctatagacg ttgctgcca tcttgctac 1260  
gctcttccaa aagcgcagac cacagttaca tgtatggaga atggctggtc tctactccc 1320  
  
agatgcatcc gtgtcaaaac atgttccaaa tcaagtatag atattgagaa tgggtttatt 1380  
tctgaatctc aglatacata tgccttaaaa gaaaaagcga aatatcaatg caaactagga 1440  
tatgtaacag cagatgggtga aacatcagga tcaattacat gtgggaaaga tggatggca 1500  
gctcaaccca cgtgcattaa atcttggat atcccagtat ttatgaatgc cagaactaaa 1560  
aatgacttca catggtttaa gctgaatgac acattggact atgaatgcca tgatggttat 1620  
gaaagcaata ctggaagcac cactggttcc atagtgtgtg gttacaatgg ttggtctgat 1680  
ttacccatat gttatgaaag agaatgcgaa cttcctaaaa tagatgtaca cttagttcct 1740  
  
gatcgcaaga aagaccagta taaagttgga gaggtgttga aattctcctg caaaccagga 1800  
tttacaatag ttggacctaa ttccgttcag tctaccact ttggattgtc tectgacctc 1860  
ccaatatgta aagagcaagt acaatcatgt ggtccacctc ctgaactcct caatgggaat 1920  
gttaaggaaa aaacgaaaga agaatatgga cacagtgaag tgggtgaata ttattgcaat 1980  
cctagatttc taatgaaggg acctaatata attcaatgtg ttgatggaga gtggacaact 2040  
ttaccagtgt gtattgtgga ggagagtacc tgtggagata tacctgaact tgaacatggc 2100  
tgggcccage tttcttcccc tcttattac tatggagatt cagtggaatt caattgctca 2160  
  
gaatcattta caatgattgg acacagatca attacgtgta ttcattggagt atggacccaa 2220  
cttccccagt gtgtggcaat agataaactt aagaagtca aatcatcaa ttaattata 2280  
cttgaggaaac atttaaaaa caagaaggaa ttcgatcata attctaacat aaggtacaga 2340  
tgtagaggaa aagaaggatg gatacacaca gtctgcataa atggaagatg ggatccagaa 2400  
gtgaactgct caatggcaca aatacaatta tgcccacctc cacctcagat tcccaattct 2460  
cacaatatga caaccacact gaattatcgg gatggagaaa aagtatctgt tctttgccaa 2520  
gaaaattatc taattcagga aggagaagaa attacatgca aagatggaag atggcagtca 2580  
  
ataccactct gtgttgaaaa aattccatgt tcacaaccac ctgagataga acacggaacc 2640  
attaattcat ccaggtcttc acaagaaagt tatgcacatg ggactaaatt gagttatact 2700  
tgtgagggtg gtttcaggat atctgaagaa aafgaaaca catgctacat gggaaaatgg 2760  
agttctccac ctgagtgtga aggcttctct tgtaaatctc cacctgagat ttctcatggt 2820  
gtttagctc acatgtcaga cagttatcag tatggagaag aagttacgta caaatgtttt 2880  
gaaggttttg gaattgatgg gcctgcaatt gcaaaatgct taggagaaaa atggtctcac 2940

cctccatcat gcataaaaac agattgtctc agtttaccta gctttgaaaa tgccataccc 3000

atgggagaga agaaggatgt gtataaggcg ggtgagcaag tgacttacac ttgtgcaaca 3060

tattacaaaa tggatggagc cagtaatgta acatgcatta atagcagatg gacaggaagg 3120

ccaacatgca gagacacctc ctgtgtgaat ccgccacag taaaaatgc ttatatagtg 3180

tcgagacaga tgagtaaata tccatctggt gagagagtac gttatcaatg taggagcctt 3240

tatgaaatgt ttggggatga agaagtgatg tgtttaaatg gaaactggac ggaaccacct 3300

caatgcaaag attctacagg aaaatgtggg cccctccac ctattgaaa tggggacatt 3360

acttcattcc cgttgtcagt atatgctcca gttcatcag ttgagtacca atgccagaac 3420

ttgtatcaac ttgagggtaa caagcgaata acatgtagaa atggacaatg gtcagaacca 3480

ccaaaatgct tacatccgtg tgtaatatcc cgagaaatta tggaaaatta taacatagca 3540

ttaagtgga cagccaaaca gaagctttat tcgagaacag gtgaatcagt tgaatttgtg 3600

tgtaaacggg gatatcgtct ttcacacgt tctcacacat tgcgaacaac atgttgggat 3660

gggaaactgg agtatccaac ttgtgcaaaa agatag 3696

<210> 53

<211> 449

<212> PRT

<213> Unknown

<220><223> Illustrative FHL-1 polypeptide sequence

<400> 53

Met Arg Leu Leu Ala Lys Ile Ile Cys Leu Met Leu Trp Ala Ile Cys

1                    5                    10                    15

Val Ala Glu Asp Cys Asn Glu Leu Pro Pro Arg Arg Asn Thr Glu Ile

                  20                    25                    30

Leu Thr Gly Ser Trp Ser Asp Gln Thr Tyr Pro Glu Gly Thr Gln Ala

                  35                    40                    45

Ile Tyr Lys Cys Arg Pro Gly Tyr Arg Ser Leu Gly Asn Val Ile Met

                  50                    55                    60

Val Cys Arg Lys Gly Glu Trp Val Ala Leu Asn Pro Leu Arg Lys Cys

65                    70                    75                    80

Gln Lys Arg Pro Cys Gly His Pro Gly Asp Thr Pro Phe Gly Thr Phe

                  85                    90                    95

Thr Leu Thr Gly Gly Asn Val Phe Glu Tyr Gly Val Lys Ala Val Tyr  
 100 105 110  
 Thr Cys Asn Glu Gly Tyr Gln Leu Leu Gly Glu Ile Asn Tyr Arg Glu  
 115 120 125  
 Cys Asp Thr Asp Gly Trp Thr Asn Asp Ile Pro Ile Cys Glu Val Val  
 130 135 140  
 Lys Cys Leu Pro Val Thr Ala Pro Glu Asn Gly Lys Ile Val Ser Ser  
 145 150 155 160  
 Ala Met Glu Pro Asp Arg Glu Tyr His Phe Gly Gln Ala Val Arg Phe  
 165 170 175  
 Val Cys Asn Ser Gly Tyr Lys Ile Glu Gly Asp Glu Glu Met His Cys  
 180 185 190  
 Ser Asp Asp Gly Phe Trp Ser Lys Glu Lys Pro Lys Cys Val Glu Ile  
 195 200 205  
 Ser Cys Lys Ser Pro Asp Val Ile Asn Gly Ser Pro Ile Ser Gln Lys  
 210 215 220  
 Ile Ile Tyr Lys Glu Asn Glu Arg Phe Gln Tyr Lys Cys Asn Met Gly  
 225 230 235 240  
 Tyr Glu Tyr Ser Glu Arg Gly Asp Ala Val Cys Thr Glu Ser Gly Trp  
 245 250 255  
 Arg Pro Leu Pro Ser Cys Glu Glu Lys Ser Cys Asp Asn Pro Tyr Ile  
 260 265 270  
 Pro Asn Gly Asp Tyr Ser Pro Leu Arg Ile Lys His Arg Thr Gly Asp  
 275 280 285  
 Glu Ile Thr Tyr Gln Cys Arg Asn Gly Phe Tyr Pro Ala Thr Arg Gly  
 290 295 300  
 Asn Thr Ala Lys Cys Thr Ser Thr Gly Trp Ile Pro Ala Pro Arg Cys  
 305 310 315 320  
 Thr Leu Lys Pro Cys Asp Tyr Pro Asp Ile Lys His Gly Gly Leu Tyr  
 325 330 335  
 His Glu Asn Met Arg Arg Pro Tyr Phe Pro Val Ala Val Gly Lys Tyr



atatctcaga agattattta taaggagaat gaacgatttc aatataaatg taacatgggt 720  
 tatgaataca gtgaaagagg agatgctgta tgcactgaat ctggatggcg tccgttgcct 780  
  
 tcatgtgaag aaaaatcatg tgataatect tatattccaa atggtgacta ctcaccttta 840  
 aggattaaac acagaactgg agatgaaatc acgtaccagt gtagaaatgg tttttatcct 900  
 gcaacccggg gaaatacagc aaaatgcaca agtactggct ggatacctgc tccgagatgt 960  
 accttgaaac cttgtgatta tccagacatt aaacatggag gtctatatca tgagaatag 1020  
 cgtagacat acittccagt agctgtagga aaatattact cctattactg tgatgaacat 1080  
 tttgagactc cgtcaggaag ttaactggat cacattcatt gcacacaaga tggatggctc 1140  
 ccagcagtac catgcctcag aaaatgttat tttccttatt tggaaaatgg atataatcaa 1200  
  
 aatcatggaa gaaagtttgt acagggtaaa tctatagacg ttgcctgcca tcttgctac 1260  
 gctcttccaa aagcgcagac cacagttaca tgtatggaga atggctggtc tectactccc 1320  
 agatgcatcc gtgtcagctt taccctctga 1350

<210> 55

<211> 383

<212> PRT

<213> Unknown

<220><223> Exemplary NPHS2 amino acid sequence

<400> 55

Met Glu Arg Arg Ala Arg Ser Ser Ser Arg Glu Ser Arg Gly Arg Gly  
 1                    5                    10                    15  
 Gly Arg Thr Pro His Lys Glu Asn Lys Arg Ala Lys Ala Glu Arg Ser  
  
                   20                    25                    30  
 Gly Gly Gly Arg Gly Arg Gln Glu Ala Gly Pro Glu Pro Ser Gly Ser  
                   35                    40                    45  
 Gly Arg Ala Gly Thr Pro Gly Glu Pro Arg Ala Pro Ala Ala Thr Val  
                   50                    55                    60  
 Val Asp Val Asp Glu Val Arg Gly Ser Gly Glu Glu Gly Thr Glu Val  
 65                    70                    75                    80  
 Val Ala Leu Leu Glu Ser Glu Arg Pro Glu Glu Gly Thr Lys Ser Ser  
  
                   85                    90                    95  
 Gly Leu Gly Ala Cys Glu Trp Leu Leu Val Leu Ile Ser Leu Leu Phe



Pro Ser Asn Arg Thr Gln Gly Ser Leu Pro Phe Pro Ser Pro Ser Lys  
 355 360 365

Pro Val Glu Pro Leu Asn Pro Lys Lys Lys Asp Ser Pro Met Leu  
 370 375 380

<210> 56

<211> 1149

<212> DNA

<213> Unknown

<220><223> Exemplary NPHS2 transgene sequence

<400> 56

atggagagga gggcgcggag ctctccagg ggtcccgcg ggcgagcgg caggactccg 60

cacaaggaga acaagagggc aaagccgag aggagcggcg ggggccgcgg gcgccaggag 120

gctgggccc agccgtcggg ctccggacgg gcggggacc cgggggagcc cggagcggcc 180

gcccaccgg tgggtggact ggatgaggtc cgaggctccg gcgaggaggg caccgaggtg 240

gtggcctgt tggagagcga gcggcccag gaaggtacca aatcctccg cttaggggcc 300

tgtgagtggc ttcttgcct catttccctg ctcttcatca tcatgacctt ccctttttcc 360

atctggttct gcgtaaaggt tgtacaagag tatgaaagag taattatatt ccgactggga 420

catctgcttc ctggaagagc caaaggcct ggtcttttct tttttttgcc ctgcctggat 480

acctaccaca aggttgacct tcgtctccaa actctggaga taccttttca tgagatcgtg 540

accaaagaca tgittataat ggagatagat gccatttgc actaccgaat ggaaaatgcc 600

tctcttctcc taagcagtct tgctcatgta tctaaagctg tgcaattcct tgtgcaaacc 660

actatgaagc gtctcctagc acatcgatcc ctactgaaa ttcttctaga gaggaagagc 720

atcgccaag atgcaaaggt tgccttggat tcagtacct gtatttgggg aatcaaagtg 780

gagagaatag aaattaaaga tgtgaggttg ccagctgggc ttcagcactc actggctgtg 840

gaggctgaag cgcaaagaca agccaaagtg cggatgattg ctgcagaagc ggaaaaggct 900

gcttctgagt ccctgaggat ggcagctgag attctgtcag gcaccctgc tgctgttcag 960

cttcgatacc tccacacct tcagtctctg tccacagaga agccttccac tgtggtttta 1020

cttttgccat ttgacctact gaattgcctg tcttctccca gcaacagaac tcaggaagc 1080

ctccccttc caagtcctc caaacctgt gagccactaa atcctaaaaa gaaagactct 1140

cccatgtta 1149

<210> 57

<211> 1152

<212> DNA

<213> Unknown

<220><223> Exemplary NPHS2 transgene sequence

<400> 57

```

atggagagga gggcgcggag ctctccagg gagtcccgcg ggcgaggcgg caggactccg      60

cacaaggaga acaagagggc aaaggccgag aggagcggcg gaggccgcgg gcgccaggag      120
gttgggcccg agccgtcggg ctccggacgg gcggggaccc cgggggagcc ccgagcgcgc      180
gccgccacgg tggtaggact ggatgaggtc cgaggctccg gcgaggaggg caccgaggtg      240
gtggcctgt tggagagcga gcggcccag gaaggtacca aatcctccgg cttaggggcc      300
tgtgagtggc ttcttgtcct catttccctg ctcttcatca tcatgacctt cctttttcc      360
atctggttct gcgtaaaggt tgtacaagag tatgaaagag taattatatt ccgactggga      420
catctgcttc ctggaagagc caaaggccct ggtcttttct tttttttgcc ctgcctggat      480

acctaccaca aggttgacct tcgtctcaa actctggaga taccttttca tgagatcgtg      540
accaaagaca tgttataat ggagatagat gccatttgcct actaccgaat ggaaaatgcc      600
tctcttctcc taagcagtct tgctcatgta tctaaagctg tgcaattcct tgtgcaaacc      660
actatgaage gtctcctagc acatcgatcc ctactgaaa ttcttctaga gaggaagagc      720
atcgccaag atgcaaaggt tgccttggat tcagtgcacct gtatttgggg aatcaaagtg      780
gagagaatag aaattaaaga tgtgaggttg ccagctgggc ttcagcactc actggctgtg      840
gaggctgaag cgcaaagaca agccaaagtg cggatgattg ctgcagaagc ggaaaaggct      900

gcttctgagt ccctgaggat ggcagctgag attctgtcag gcacccctgc tgccgttcag      960
cttcgatacc tccacacct tcagtctctg tccacagaga agccttccac tgtggtttta      1020
cctttgccat ttgacctact gaattgcctg tcttctceca gcaacagaac tcagggaagc      1080
ctccccttcc caagtccttc caaacctgtt gagccactaa atcctaaaaa gaaagactct      1140
cccatgttat ag

```

<210> 58

<211> 818

<212> DNA

<213> Unknown

<220><223> Exemplary minimal nephrin promoter

<400> 58

```

ggccctgggg tcacggaggg tggggaggca ccgaggaacg cgcctggcat gtgctgacag      60

```

ggaattttat gctccagctg ggccagctgg gaggagcctg ctgggcagag gccagagctg 120  
 ggggctctgg aaggtacctg ggggaggttg cactgtgaga atgagctcaa gctgggtcag 180  
 agagcagggc tgactctgcc agtgcctgca tcagcctcat cgctctccta ggctcctggc 240  
 ctgctggact ctgggctgca ggtccttctt gaaaggctgt gagttagtgag acaaggagca 300  
 ggagtgaggg gtggcaggag agaagataga gattgagaga gagagagaga gagacagaga 360  
 gagaggaaga gacagagaca aaaggagaga gaacggctta gacaaggaga gaaagatgga 420  
 aagataaaga gactgggcgc agtggctcac gcctgtaatc ccaacacttg gggaggccaa 480

ggtgggagga tggcttgaag gaaagagtct gagatcaacc tggccaacat agtgagacc 540  
 cgctctctaaa aaaaaaaaaag aaaaaaaaaa gaaaaaagaa aaaaaagttt ttttaagag 600  
 acagagaaag agactcagag attgagactg agagcaagac agagagagac actcacaggg 660  
 aagaggggaa gaggaaaacg agaaaggag gagagtaacg gaaagagata aaaaagaaaa 720  
 gcaggtggca gagacacaga gagagggacc cagagaaagc cagacagacg caggtggctg 780  
 gcagcgggcg ctgtgggggt cacagtaggg ggacctgt 818

<210> 59

<211> 818

<212> DNA

<213> Unknown

<220><223> Exemplary minimal nephrin promoter variant

<400> 59

ggccctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag 60  
 gggattttat gctccagctg ggccagctgg gaggagcctg ctgggcagag gccagagctg 120  
 ggggctctgg aaggtacctg ggggaggttg cactgtgaga atgagctcaa gctgggtcag 180  
 agagcagggc tgactctgcc agtgcctgca tcagcctcat cgctctccta ggctcctggc 240  
 ctgctggact ctgggctgca ggtccttctt gaaaggctgt gagttagtgag acaaggagca 300  
 ggagtgaggg gtggcaggag agaagataga gattgagaga gagagagaga gagacacaga 360  
 gagagaggaa gagacagaga caaaaggaga gagaacggct tagacaagga gagaaagatg 420  
 gaaagataaa gagactgggc gcagtggctc acgcctgtaa tccaacact tggggaggcc 480  
 aaggtgggag gatggcttga aggaaagagt ctgagatcaa cctggccaac atagtgagac 540  
 cccgtctcta aaaaaaaaaag aaaaaaaaaa gaaaaaagaa aaaaaagttt ttttaagag 600  
 acagagaaag agactcagag attgagactg agagcaagac agagagagat actcacaggg 660  
 aagaggggaa gaggaaaacg agaaaggag gagagtaacg gaaagagata aaaaagaaaa 720

gcaggtggca gagacacaca gagagggacc cagagaaagc cagacagacg caggtggctg 780

gcagcgggcg ctgtgggggt cacagtaggg ggacctgt 818

<210> 60

<211> 818

<212> DNA

<213> Unknown

<220><223> Exemplary minimal nephrin promoter variant

<400> 60

ggccctgggg tcacggaggc tggggaggca ccgaggaacg cgcctggcat gtgctgacag 60

gggattttat gctccagctg ggccagctgg gaggagcctg ctgggcagag gccagagctg 120

ggggctctgg aaggtacctg ggggaggttg cactgtgaga atgagctcaa gctgggtcag 180

agagcagggc tgactctgcc agtgacctgca tcagcctcat cgctctccta ggctcctggc 240

ctgctggact ctgggctgca ggtccttctt gaaaggctgt gtagtagtag acaaggagca 300

ggagtgaggg gtggcaggag agaagataga gattgagaga gagagagaga gagacagaga 360

gagaggaaga gagacagaca aaaggagaga gaacggctta gacaaggaga gaaagatgga 420

aagataaaga gactgggcbc agtggtcac gcctgtaatc ccaacacttg gggaggccaa 480

ggtgggagga tgcttgaag gaaagagtct gagatcaacc tggccaacat agtgagacct 540

ctctctaaa aaaaaaaaaa aaaaaaaaaa gaaaaaagaa aaaaaagttt ttttaagag 600

acagagaaag agactcagag attgagactg agagcaagac agagagagat actcacaggg 660

aagaggggaa gaggaaaacg agaaaggag gagagtaacg gaaagagata aaaaagaaaa 720

gcaggtggca gagacacaga gagagggacc cagagaaagc cagacagacg caggtggctg 780

gcagcgggcg ctgtgggggt cacagtaggg ggacctgt 818

<210> 61

<211> 7

<212> DNA

<213> Unknown

<220><223> sequence

<400> 61

ggggtca 7

<210> 62

<211> 17

<212> DNA

<213> Unknown

<220><223> sequence

<400> 62

cggaggctgg ggaggca

17

<210> 63

<211> 5

<212>

DNA

<213> Unknown

<220><223> Sequence

<400> 63

atgtg

5