



(12) 发明专利申请



(10) 申请公布号 CN 117377770 A

(43) 申请公布日 2024.01.09

(21) 申请号 202280034399.8

(22) 申请日 2022.03.14

(30) 优先权数据

2103470.7 2021.03.12 GB

PCT/GB2021/050633 2021.03.12 GB

PCT/GB2021/051668 2021.06.30 GB

(85) PCT国际申请进入国家阶段日

2023.11.10

(86) PCT国际申请的申请数据

PCT/GB2022/050649 2022.03.14

(87) PCT国际申请的公布数据

W02022/189811 EN 2022.09.15

(71) 申请人 布里斯托尔大学

地址 英国布里斯托尔

申请人 辛科纳知识产权控股(3)有限公司

(72) 发明人 M·萨莱姆-乌丁 G·威尔什

V·库兹穆克 A·W·格里菲思

(74) 专利代理机构 北京市柳沈律师事务所

11105

专利代理师 张文辉

(51) Int.Cl.

C12N 15/86 (2006.01)

权利要求书4页 说明书70页 附图21页

(54) 发明名称

最小肾病蛋白启动子

(57) 摘要

启动子,其包含(i)与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列,并且其中所述启动子具有约1.1kb或更短的长度。

1. 启动子,其包含(i)与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列,并且其中所述启动子具有约1.1kb或更短的长度。

2. 启动子,其包含根据SEQ ID NO:1的核苷酸序列或由根据SEQ ID NO:1的核苷酸序列组成,但其中:

(i) SEQ ID NO:1的位置1至位置n1缺失,其中n1为100至430的整数;和/或

(ii) SEQ ID NO:1的位置n2至位置n3缺失,其中 $n3 \geq n2$ ,n2为508至1061的整数,并且n3为508至1061的整数;

或与其具有至少70%同一性的核苷酸序列。

3. 根据权利要求2所述的启动子,其中:

(iii) 对应于SEQ ID NO:1的位置493的位置是G,对应于SEQ ID NO:1的位置1080的位置是T,对应于SEQ ID NO:1的位置1169的位置是C,对应于SEQ ID NO:1的位置1249的位置是G。

4. 根据权利要求2或3所述的启动子,其中所述启动子包含(i)与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列,和/或其中所述启动子具有约1.1kb或更短的长度。

5. 根据任一前述权利要求所述的启动子,其中所述启动子是肾特异性启动子。

6. 根据任一前述权利要求所述的启动子,其中所述启动子进一步包含(ii)与SEQ ID NO:5或18具有至少70%同一性的核苷酸序列、与SEQ ID NO:6或19具有至少70%同一性的核苷酸序列,和/或与SEQ ID NO:7或20具有至少70%同一性的核苷酸序列。

7. 根据任一前述权利要求所述的启动子,其中所述启动子进一步包含(iii)与SEQ ID NO:8具有至少70%同一性的核苷酸序列或其一个或多个片段。

8. 根据任一前述权利要求所述的启动子,其中所述启动子从5'至3'包含:(i)与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列;(iii)任选地与SEQ ID NO:8具有至少70%同一性的核苷酸序列,或其一个或多个片段;和(ii)与SEQ ID NO:5或18具有至少70%同一性的核苷酸序列、与SEQ ID NO:6或19具有至少70%同一性的核苷酸序列,和/或与SEQ ID NO:7或20具有至少70%同一性的核苷酸序列。

9. 根据任一前述权利要求所述的启动子,其中所述启动子具有约1.0kb或更短、约0.9kb或更短、约0.8kb或更短、约0.7kb或更短、约0.6kb或更短、约0.5kb或更短、约0.4kb或更短或约0.3kb或更短的长度。

10. 根据任一前述权利要求所述的启动子,其中所述启动子具有0.265kb至1.0kb、0.265kb至0.9kb、0.265kb至0.8kb、0.265kb至0.7kb、0.265kb至0.6kb、0.265kb至0.5kb、0.265kb至0.4kb或0.265kb至0.3kb的长度。

11. 根据任一前述权利要求所述的启动子,其中所述启动子包含:

(a) 视黄酸受体结合位点;

(b) WT1结合位点;

(c) 增强子框;

(d) 转录因子结合区;和/或

(e) 转录起始位点。

12. 根据权利要求11所述的启动子,其中所述视黄酸受体结合位点包含以下或由以下组成:如SEQ ID NO:10所示的核苷酸序列,或与SEQ ID NO:10相比具有一个或两个取代、缺

失或插入的核苷酸序列。

13. 根据权利要求11或12所述的启动子,其中所述WT1结合位点包含以下或由以下组成:如SEQ ID NO:11所示的核苷酸序列,或与SEQ ID NO:11相比具有一个、两个或三个取代、缺失或插入的核苷酸序列。

14. 根据权利要求11至13中任一项所述的启动子,其中所述增强子框包含以下或由以下组成:如SEQ ID NO:12所示的核苷酸序列,或与SEQ ID NO:12相比具有一个或两个取代、缺失或插入的核苷酸序列。

15. 根据权利要求1或4至14中任一项所述的启动子,其中(i)所述与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列包含以下的一个或多个:(a) 视黄酸受体结合位点;(b) WT1结合位点;和(c) 增强子框。

16. 根据权利要求1或4至15中任一项所述的启动子,其中以下核苷酸序列中的一个或多个存在于(i)所述与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列中:

(a) 对应于SEQ ID NO:4或17的位置7至位置13的位置处的GGGGTCA;

(b) 对应于SEQ ID NO:4或17的位置14至位置30的位置处的CGGAGGCTGGGGAGGCA;以及

(c) 对应于SEQ ID NO:4或17的位置49至位置53的位置处的ATGTG。

17. 根据权利要求11至16中任一项所述的启动子,其中所述转录因子结合区包含以下或由以下组成:如SEQ ID NO:13或43所示的核苷酸序列,或与SEQ ID NO:13或43相比具有一个、两个、三个、四个或五个取代、缺失或插入的核苷酸序列。

18. 根据权利要求11至17中任一项所述的启动子,其中所述转录起始位点包含“AG”二核苷酸或由“AG”二核苷酸组成。

19. 根据权利要求11至18中任一项所述的启动子,其中所述转录因子结合位点可操作地连接至所述转录起始位点,任选地其中所述转录因子结合位点直接位于所述转录起始位点的上游。

20. 根据任一前述权利要求所述的启动子,其中所述启动子包含与SEQ ID NO:2或SEQ ID NO:3具有至少70%同一性的核苷酸序列或由与SEQ ID NO:2或SEQ ID NO:3具有至少70%同一性的核苷酸序列组成。

21. 根据任一前述权利要求所述的启动子,其中所述启动子包含与SEQ ID NO:47具有至少70%同一性的核苷酸序列或由与SEQ ID NO:47具有至少70%同一性的核苷酸序列组成。

22. 根据任一前述权利要求所述的启动子,其中所述启动子包含与SEQ ID NO:47具有至少90%同一性的核苷酸序列或由与SEQ ID NO:47具有至少70%同一性的核苷酸序列组成。

23. 根据任一前述权利要求所述的启动子,其中所述启动子包含与SEQ ID NO:47具有至少95%同一性的核苷酸序列或由与SEQ ID NO:47具有至少95%同一性的核苷酸序列组成。

24. 根据任一前述权利要求所述的启动子,其中所述启动子包含与SEQ ID NO:47具有至少98%同一性的核苷酸序列或由与SEQ ID NO:47具有至少98%同一性的核苷酸序列组成。

25. 根据任一前述权利要求所述的启动子,其中所述启动子包含与SEQ ID NO:47具有

至少99%同一性的核苷酸序列或由与SEQ ID NO:47具有至少99%同一性的核苷酸序列组成。

26. 启动子,其由与SEQ ID NO:2具有至少70%同一性的核苷酸序列组成。

27. 启动子,其由与SEQ ID NO:3具有至少70%同一性的核苷酸序列组成。

28. 启动子,其由与SEQ ID NO:15或59具有至少70%同一性的核苷酸序列组成。

29. 启动子,其由与SEQ ID NO:47具有至少70%同一性的核苷酸序列组成。

30. 根据权利要求29所述的启动子,其中所述启动子由与SEQ ID NO:47具有至少90%同一性的核苷酸序列组成。

31. 根据权利要求29或30所述的启动子,其中所述启动子由与SEQ ID NO:47具有至少95%同一性的核苷酸序列组成。

32. 根据权利要求29-31中任一项所述的启动子,其中所述启动子由与SEQ ID NO:47具有至少98%同一性的核苷酸序列组成。

33. 根据权利要求29-32中任一项所述的启动子,其中所述启动子由与SEQ ID NO:47具有至少99%同一性的核苷酸序列组成。

34. 根据权利要求29-32任一项所述的启动子,其中所述启动子由SEQ ID NO:47的核苷酸序列组成。

35. 多核苷酸,其包含根据权利要求1至34中任一项所述的启动子。

36. 根据权利要求35所述的多核苷酸,其中所述启动子可操作地连接至蛋白质编码序列。

37. 根据权利要求35或36所述的多核苷酸,其中所述启动子可操作地连接至编码以下的蛋白质编码序列:NPFS2或其片段和/或变体;COL4A3多肽、COL4A4多肽或COL4A5多肽,或其片段或衍生物;或者CFI、CFH或FHL-1,或其片段和/或变体。

38. 根据权利要求35-37中任一项所述的多核苷酸,其中所述启动子可操作地连接至编码NPFS2,或其片段和/或变体的蛋白质编码序列,优选地其中所述蛋白质编码序列编码与SEQ ID NO:55至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段。

39. 根据权利要求35-37中任一项所述的多核苷酸,其中所述启动子可操作地连接至编码COL4A3多肽、COL4A4多肽或COL4A5多肽,或其片段或变体或衍生物的蛋白质编码序列,优选地其中所述蛋白质编码序列编码与SEQ ID NO:21-23中任一项至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段。

40. 根据权利要求35-37中任一项所述的多核苷酸,其中所述启动子可操作地连接至编码CFI、CFH或FHL-1,或其片段和/或变体的蛋白质编码序列,优选地其中所述蛋白质编码序列编码与SEQ ID NO:49、51或53中任一项至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段。

41. 根据权利要求35-40中任一项所述的多核苷酸,其中所述蛋白质编码序列可操作地连接至一个或多个进一步的调控元件,诸如转录后调控元件和/或多腺苷酸化序列。

42. 载体,其包含根据权利要求35至41中任一项所述的多核苷酸。

43. 根据权利要求42所述的载体,其中所述载体能够转导足细胞,任选地其中所述载体能够特异性转导足细胞。

44. 根据权利要求42或43所述的载体,其中所述载体是病毒载体,诸如腺相关病毒(AAV)载体、慢病毒载体、逆转录病毒载体、腺病毒载体、腺相关病毒载体、单纯疱疹病毒载体、甲病毒载体、黄病毒载体、弹状病毒载体、麻疹病毒载体、鸡新城疫病毒载体、痘病毒载体和小核糖核酸病毒载体,优选地其中所述载体是AAV载体。

45. 根据权利要求44所述的载体,其中所述病毒载体是以病毒载体颗粒的形式,优选其中所述病毒载体是以AAV载体颗粒的形式。

46. 根据权利要求42至45中任一项所述的载体,其中所述载体是由AAV3B、LK03或AAV9衣壳蛋白衣壳化的AAV载体颗粒的形式。

47. 细胞,其包含根据权利要求35至41中任一项所述的多核苷酸或根据权利要求42至46中任一项所述的载体。

48. 药物组合物,其包含根据权利要求35至41中任一项所述的多核苷酸、根据权利要求42至46中任一项所述的载体或根据权利要求47所述的细胞。

49. 根据权利要求35至41中任一项所述的多核苷酸、根据权利要求42至46中任一项所述的载体或根据权利要求47所述的细胞,其用于药物中。

50. 根据权利要求1至34中任一项所述的启动子用于驱动编码序列的表达的用途。

51. 根据权利要求50所述的用途,其中所述表达是肾特异性的。

## 最小肾病蛋白启动子

### 技术领域

[0001] 本发明与能够在足细胞中驱动转基因表达的启动子相关。本发明还与包含所述启动子的多核苷酸和载体相关。

[0002] 发明背景

[0003] 有多种通过攻击肾小球来影响肾功能的疾病。肾小球每天过滤大约180升的血浆，健康的肾小球滤过屏障具有在我们的一生中保留约99.9%的大蛋白(包括白蛋白)而不会堵塞的惊人的能力。肾小球滤过屏障(GFB)包含3个主要层:肾小球内皮细胞、肾小球基底膜(GBM)和足细胞。

[0004] GBM由IV型胶原、蛋白聚糖和层粘连蛋白的高度交联的大分子网络形成。肾小球疾病的遗传形式可以是由这些分子结构的遗传缺陷引起的。例如,奥尔波特综合征由COL4A3、COL4A4和COL4A5基因中的致病性变体引起,所述致病性变体导致基底膜的IV型胶原蛋白 $\alpha$ 345网络异常。在欧洲大陆和美国,奥尔波特综合征影响所有个体的大约五千分之一到万分之一。这种情况通常出现在童年期间,并与一系列表型相关联,包括肾功能的逐步丧失,还可能包括听力丧失和眼睛异常。其他GBM相关联的疾病包括皮尔逊综合征和指甲髌骨综合征(Chiang,C.K.和Inagi,R.,2010.Nature Reviews Nephrology,6(9),p.539)。

[0005] 足细胞也被认为是肾小球疾病进展中的关键细胞。足细胞是高度特化并仅在肾小球中发现的中胚层衍生细胞。它们展示出独特的特征,诸如足突和裂孔隔膜(slit diaphragm),这对肾小球滤过至关重要。足细胞相关联的遗传性肾小球疾病包括肾病综合征、弗雷泽综合征和德尼-德拉什综合征、Schimke免疫-骨发育不良、Epstein和Fechtner综合征(Chiang,C.K.和Inagi,R.,2010.Nature Reviews Nephrology,6(9),p.539)。

[0006] 因此,肾小球细胞,诸如足细胞,代表了基因疗法方法的潜在靶标。

[0007] 为了最大化基因疗法潜力,需要能够在肾小球细胞(诸如足细胞)中驱动转基因表达的启动子序列。Wong等人(American Journal of Physiology Renal Physiology;2000;279(6);F1027-32)描述了来自人肾病蛋白启动子和5'-侧翼区的1.25-kb DNA片段,其能够指导足细胞特异性表达。该启动子已用于使用AAV9载体来实现GFP的肾特异性表达(Picconi等人;2014;Molecular Therapy-Methods&Clinical Development;1,14014)。

[0008] 目前,大量的载体货物容量(例如AAV货物容量)被非编码元件,诸如肾病蛋白或足细胞素启动子、WPRE元件和多腺苷酸化序列占据。因此,在本领域中有对最小启动子的需要,其能够在肾小球细胞诸如足细胞中驱动转基因表达,以便为肾小球疾病的基因疗法方法提供最大的灵活性。

### 发明内容

[0009] 本发明基于发明人出人意料地提供了能够驱动肾细胞、特别是肾小球细胞诸如足细胞中的转基因表达的最小启动子。

[0010] 在一方面,本发明提供了启动子,其包含(i)与SEQ ID NO:4或17具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列,并且其中所述启动子具有约

1.1kb或更短的长度。

[0011] 在一些实施方案中,本发明提供了启动子,其包含根据SEQ ID NO:1的核苷酸序列或由根据SEQ ID NO:1的核苷酸序列组成,但其中:

[0012] (i) 缺失了SEQ ID NO:1的位置1至位置n1,其中n1为1至430的整数;和/或

[0013] (ii) 缺失了SEQ ID NO:1的位置n2至位置n3,其中 $n3 \geq n2$ , n2为508至1061的整数, n3为508至1061的整数;

[0014] 或与其具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列。

[0015] 在一些实施方案中:

[0016] (iii) 对应于SEQ ID NO:1的位置493的位置是G,对应于SEQ ID NO:1的位置1080的位置是T,对应于SEQ ID NO:1的位置1169的位置是C,对应于SEQ ID NO:1的位置1249的位置是G。

[0017] 在一方面,本发明提供了启动子,其包含与根据SEQ ID NO:1的核苷酸序列具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列或由与根据SEQ ID NO:1的核苷酸序列具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列组成,但其中:

[0018] (i) 缺失了SEQ ID NO:1的位置1至位置n1,其中n1为100至430的整数;和/或

[0019] (ii) 缺失了SEQ ID NO:1的位置n2至位置n3,其中 $n3 \geq n2$ , n2为508至1061的整数, n3为508至1061的整数。

[0020] 在一些实施方案中:

[0021] (iii) 用G取代SEQ ID NO:1的位置493,用T取代SEQ ID NO:1的位置1080,用C取代SEQ ID NO:1的位置1169,并且用G取代SEQ ID NO:1的位置1249。

[0022] 该启动子可以包含(i)与SEQ ID NO:4或17具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列,和/或该启动子可以具有约1.1kb或更短的长度。

[0023] 启动子可以是在肾细胞中可操作的。启动子可以能够驱动肾中的转基因表达。启动子可以是肾特异性启动子。启动子可以是在足细胞中可操作的。启动子可以能够驱动足细胞中的转基因表达。启动子可以是足细胞特异性启动子。

[0024] 合适地,启动子进一步包含(ii)与SEQ ID NO:5或18具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列、与SEQ ID NO:6或19具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列,和/或与SEQ ID NO:7或20具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列。在一些实施方案中,启动子进一步包含(ii)与SEQ ID NO:6或19具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列。

[0025] 合适地,启动子进一步包含(iii)与SEQ ID NO:8具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列或与SEQ ID NO:8具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列的一个或多个片段。

[0026] 合适地,根据任一前述权利要求的启动子,其中启动子从5'至3'包含:(i)与SEQ ID NO:4或17具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列;(iii)任选地与SEQ ID NO:8具有至少70%、至少80%、至少90%、至少95%或至少99%

同一性的核苷酸序列或其一个或多个片段;以及(ii)与SEQ ID NO:5或18具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列、与SEQ ID NO:6或19具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列,和/或与SEQ ID NO:7或20具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列。

[0027] 在一些实施方案中,启动子从5'至3'包含:(i)与SEQ ID NO:4具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列;以及(ii)与SEQ ID NO:7具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列。在一些实施方案中,启动子从5'至3'由以下组成:(i)与SEQ ID NO:4具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列;以及(ii)与SEQ ID NO:7具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列,并且其中该启动子具有约1.1kb或更短的长度。

[0028] 在一些实施方案中,启动子从5'至3'由以下组成:(i)与SEQ ID NO:17具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列;以及(ii)与SEQ ID NO:20具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列。在一些实施方案中,启动子从5'至3'由以下组成:(i)SEQ ID NO:17的核苷酸序列;以及(ii)SEQ ID NO:20的核苷酸序列。

[0029] 在一些实施方案中,启动子从5'至3'包含:(i)与SEQ ID NO:17具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列;以及(ii)与SEQ ID NO:20具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列,并且其中该启动子具有约1.1kb或更短的长度。在一些实施方案中,启动子从5'至3'包含:(i)SEQ ID NO:17的核苷酸序列;以及(ii)SEQ ID NO:20的核苷酸序列,并且其中该启动子具有约1.1kb或更短的长度。

[0030] 启动子可以具有约1.0kb或更短、约0.9kb或更短、约0.8kb或更短、约0.7kb或更短、约0.6kb或更短、约0.5kb或更短、约0.4kb或更短,或约0.3kb或更短的长度。启动子可以具有0.265kb至1.0kb、0.265kb至0.9kb、0.265kb至0.8kb、0.265kb至0.7kb、0.265kb至0.6kb、0.265kb至0.5kb、0.265kb至0.4kb、或0.265kb至0.3kb的长度。

[0031] 启动子可以包含:(a)视黄酸受体结合位点;(b)WT1结合位点;(c)增强子框;(d)转录因子结合区;和/或(e)转录起始位点。合适地,视黄酸受体结合位点包含以下或由以下组成:如SEQ ID NO:10所示的核苷酸序列,或与SEQ ID NO:10相比具有一个或两个取代、缺失或插入的核苷酸序列。合适地,WT1结合位点包含以下或由以下组成:如SEQ ID NO:11所示的核苷酸序列,或与SEQ ID NO:11相比具有一个、两个或三个取代、缺失或插入的核苷酸序列。合适地,增强子框包含以下或由以下组成:如SEQ ID NO:12所示的核苷酸序列,或与SEQ ID NO:12相比具有一个或两个取代、缺失或插入的核苷酸序列。合适地,转录因子结合区包含以下或由以下组成:如SEQ ID NO:13或43所示的核苷酸序列,或与SEQ ID NO:13或43相比具有一个、两个、三个、四个或五个取代、缺失或插入的核苷酸序列。合适地,转录起始位点包含“AG”二核苷酸或由“AG”二核苷酸组成。合适地,转录因子结合位点可操作地连接至所述转录起始位点,任选地其中所述转录因子结合位点直接位于所述转录起始位点的上

游。

[0032] 与SEQ ID NO:4或17具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列可以包含以下的一个或多个：(a) 视黄酸受体结合位点；(b) WT1结合位点；(c) 增强子框。一个或多个以下的核苷酸序列可以存在于与SEQ ID NO:4或17具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列中：(a) 对应于SEQ ID NO:4或17的位置7至位置13的位置处的GGGGTCA；(b) 对应于SEQ ID NO:4或17的位置14至位置30的位置处的CGGAGGCTGGGGAGGCA；以及(c) 对应于SEQ ID NO:4或17的位置49至位置53的位置处的ATGTG。

[0033] 在一些实施方案中，启动子包含以下或由以下组成：与SEQ ID NO:2或SEQ ID NO:3具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列。

[0034] 在一些实施方案中，启动子包含以下或由以下组成：与SEQ ID NO:15、SEQ ID NO:59或SEQ ID NO:47具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列。

[0035] 在优选的实施方案中，启动子包含以下或由以下组成：与SEQ ID NO:47具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列。在更优选的实施方案中，启动子包含SEQ ID NO:47的核苷酸序列或由SEQ ID NO:47的核苷酸序列组成。

[0036] 在一方面，本发明提供了启动子，其由与SEQ ID NO:2具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列组成。

[0037] 在一方面，本发明提供了启动子，其由与SEQ ID NO:3具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列组成。

[0038] 在一方面，本发明提供了启动子，其由与SEQ ID NO:15具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列组成。

[0039] 在一方面，本发明提供了启动子，其由与SEQ ID NO:59具有至少70%、至少80%、至少90%、至少95%或至少99%同一性的核苷酸序列组成。

[0040]

[0041] 在一方面，本发明提供了启动子，其由与SEQ ID NO:47具有至少70%、至少80%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列组成。

[0042] 在一方面，本发明提供了启动子，其由SEQ ID NO:47的核苷酸序列组成。

[0043] 在一方面，本发明提供了多核苷酸，其包含根据本发明的启动子。

[0044] 启动子可以可操作地连接至蛋白质编码序列。在一些实施方案中，蛋白质编码序列编码NPHS2或其片段和/或变体；COL4A3多肽、COL4A4多肽或COL4A5多肽、或其片段或衍生物；或CFI、CFH或FHL-1、或其片段和/或变体。

[0045] 在一些实施方案中，蛋白质编码序列编码NPHS2，或其片段和/或变体。在一些实施方案中，蛋白质编码序列编码与SEQ ID NO:55至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段。在一些实施方案中，蛋白质编码序列包含与SEQ ID NO:56至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段，或由与SEQ ID NO:56至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段组成。在一些实施方案中，蛋白质编码序列编

码与SEQ ID NO:57至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段,或由与SEQ ID NO:57至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段组成。

[0046] 在一些实施方案中,蛋白质编码序列编码COL4A3多肽、COL4A4多肽或COL4A5多肽、或其片段或衍生物。在一些实施方案中,蛋白质编码序列编码与SEQ ID NO:21-23之一至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段。在一些实施方案中,蛋白质编码序列包含与SEQ ID NO:24-26之一至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段,或由与SEQ ID NO:24-26之一至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段组成。

[0047] 在一些实施方案中,蛋白质编码序列编码CFI,或其片段和/或变体。在一些实施方案中,蛋白质编码序列编码与SEQ ID NO:49至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段。在一些实施方案中,蛋白质编码序列包含与SEQ ID NO:50至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段,或由与SEQ ID NO:50至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段组成。

[0048] 在一些实施方案中,蛋白质编码序列编码CFH,或其片段和/或变体。在一些实施方案中,蛋白质编码序列编码与SEQ ID NO:51至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段。在一些实施方案中,蛋白质编码序列包含与SEQ ID NO:52至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段,或由与SEQ ID NO:52至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段组成。

[0049] 在一些实施方案中,蛋白质编码序列编码FHL-1,或其片段和/或变体。在一些实施方案中,蛋白质编码序列编码与SEQ ID NO:53至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段。在一些实施方案中,蛋白质编码序列包含与SEQ ID NO:54至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段,或由与SEQ ID NO:54至少70%、至少80%、至少90%、至少95%、至少98%或至少99%相同的氨基酸序列或其片段组成。

[0050] 在一些实施方案中,蛋白质编码序列可操作地连接至一个或多个进一步的调控元件,诸如转录后调控元件和/或多腺苷酸化序列。

[0051] 在一方面,本发明提供了载体,其包含根据本发明的多核苷酸。合适地,载体能够转导足细胞,任选地其中所述载体能够特异性转导足细胞。

[0052] 合适地,载体是病毒载体,诸如腺相关病毒(AAV)载体、慢病毒载体、逆转录病毒载体、腺病毒载体、腺相关病毒载体、单纯疱疹病毒载体、甲病毒载体、黄病毒载体、弹状病毒载体、麻疹病毒载体、鸡新城疫病毒载体、痘病毒载体和小核糖核酸病毒载体,优选地其中所述载体是AAV载体。合适地,病毒载体是以病毒载体颗粒的形式,优选其中所述病毒载体是以AAV载体颗粒的形式。

[0053] 在一些实施方案中,载体是由AAV3B、LK03或AAV9衣壳蛋白衣壳化的AAV载体颗粒的形式。

[0054] 在一方面,本发明提供了细胞,其包含根据本发明的多核苷酸,或根据本发明的载体。

[0055] 在一方面,本发明提供了药物组合物,其包含根据本发明的多核苷酸、根据本发明的载体或根据本发明的细胞。

[0056] 在一方面,本发明提供了根据本发明的多核苷酸、根据本发明的载体或根据本发明的细胞,其用于药物中。

[0057] 在一方面,本发明提供了根据本发明的启动子用于驱动编码序列表达的用途。该表达可以是肾特异性的。合适地,该表达是足细胞特异性的。

## 附图说明

[0058] 图1-最小肾病蛋白启动子的示意图说明

[0059] (A) 示例性全长肾病蛋白启动子的长度可以是1249bp(不包括起始密码子),下文中称为“FL”肾病蛋白启动子。(B) 具有5' 区缺失的示例性最小肾病蛋白启动子的长度可以是819bp(不包括起始密码子),下文中称为“midi”肾病蛋白启动子。(C) 具有5' 区缺失和中心区缺失的示例性最小肾病蛋白启动子的长度可以是265bp(不包括起始密码子),下文中称为“mini”肾病蛋白启动子。(D) 表明了肾病蛋白启动子的以下区域:(i) 人小鼠同源区,(ii) 视黄酸受体(RAR)结合位点,(iii) WT1结合位点,(iv) 增强子框,(v) 转录因子结合区,以及(vi) 转录起始位点。

[0060] 图2-包含与midi肾病蛋白启动子可操作地偶联的GFP的慢病毒载体的示意图

[0061] (A) 将pACE\_hNPHS1启动子用作模板,以引入BamH1和Cla1限制性位点。(B) 包含与midi肾病蛋白启动子可操作地偶联的GFP的最终构建载体的示意图。

[0062] 图3-包含与mini肾病蛋白启动子可操作地偶联的GFP的慢病毒载体的示意图

[0063] (A) 将pACE\_hNPHS1启动子用作模板以进行PCR并对启动子的两部分进行凝胶提取。(B) 包含与mini肾病蛋白启动子可操作地偶联的GFP的最终构建载体的示意图。

[0064] 图4-用慢病毒载体转导后在CiPodocyte中表达GFP

[0065] 使用慢病毒方法生成稳定表达带有GFP标签的肾病蛋白启动子的人CiPodocyte。通过荧光显微镜观察GFP表达。(A) 未转导的CiPodocyte。(B) 稳定表达带有GFP标签的mini肾病蛋白启动子的CiPodocyte。(C) 稳定表达带有GFP标签的FL肾病蛋白启动子的CiPodocyte。

[0066] 图5-用慢病毒载体转导后在已分化的ciPodocyte中表达GFP

[0067] 将包含与肾病蛋白启动子可操作地连接的GFP的慢病毒载体转导到分化的条件永生足细胞(ciPodocyte)中。使用免疫沉淀(IP)来检测GFP表达。

[0068] 图6-用慢病毒-GFP.肾病蛋白启动子转导的人肾小球细胞(FL或265)

[0069] FACS分析使用Novocyte分析仪显示了条件永生足细胞(LY)和肾小球内皮细胞(GEnC)的所有活的单细胞的中值GFP荧光。将未转导的细胞(细胞对照)与用慢病毒构建体转导的细胞进行比较,该慢病毒构建体携带由全长人肾病蛋白启动子控制的GFP表达盒(hNPHS1.GFP)或由mini人肾病蛋白启动子控制的GFP表达盒(265.GFP)。所有细胞均分化10天,胰蛋白酶消化(100 $\mu$ L)并在PBS、2%FBS、1:1000DRAQ7(150 $\mu$ L)中稀释。数据和误差线代表3次技术重复(100 $\mu$ L,>2500个细胞) $\pm$ SEM。

[0070] 图7-包含与mini肾病蛋白启动子偶联的COL4A3、COL4A4和COL4A5的AAV转移质粒

[0071] (A) pAAV.265.Col4a3.3flag.sv40的示意图,其为包含与mini肾病蛋白启动子偶联的COL4A3的AAV质粒。显示了SmaI位点,并且在用SmaI限制后预计会出现以下片段:1.6238bp,2.2753bp,3.56bp,4.11bp,5.11bp。(B) pAAV.265.Col4a4.3flag.sv40的示意图,其为包含与mini肾病蛋白启动子偶联的COL4A4的AAV质粒。显示了SmaI位点,并且在用SmaI限制后预计会出现以下片段:1.4052bp,2.2753bp,3.2224bp,4.56bp,5.11Bp,6.11bp。(C) pAAV.265.Col4a5.3flag.sv40的示意图,其为包含与mini肾病蛋白启动子偶联的COL4A5的AAV质粒。显示了SmaI位点,并且在用SmaI限制后预计会出现以下片段:1.4272bp,2.2753bp,3.2032bp,4.56bp,5.11Bp,6.11bp。(D) 用SmaI进行限制性消化。MW=1kb DNA梯状条带,泳道1、2和3分别对应于(A)、(B)和(C)中所示的质粒的消化物。(E) 显示限制性消化的示意图。

[0072] 图8-用AAV.COL4.肾病蛋白265.Sv40病毒转导的足细胞

[0073] (A) 用抗FLAG抗体下拉的人已分化的CiPodocyte(条件永生化的)中的全长带有FLAG标签的Col4a3(LK03)或Col4a5(LK03)的免疫沉淀实验。抗FLAG抗体沉淀Col4a3和Col4a5两者。人FLAG IgG用作对照。(B) 蛋白质裂解物的Western印迹,其显示了人或小鼠分化的CiPodocyte中Col4a3(LK03衣壳血清型)、Col4a5(LK03)和Col4a5(2/9衣壳血清型)的表达水平。未感染的人和小鼠的Cipodocyte用作对照。(C) 共聚焦图像,其显示了人野生型ciPodocyte/Col4a5 3xFlag AAV ciPodocyte中转导的Col4a5与F-肌动蛋白的免疫荧光染色。与野生型对应物相比,在用Col4a5 3xFlagAAV病毒感染的人已分化的足细胞中,Col4a5以胞质水平存在。

[0074] 图9-HEK细胞中CFH、CFI和CFHL1的表达

[0075] 通过Western印迹分析293T人胚肾细胞中的蛋白质表达。用pAAV-265-CFH、pAAV-FL-CFI或pAAV-FL-CFHL1转染细胞。NT(未转染的293THEK细胞)。使用蛋白质特异性抗体或抗FLAG标签或抗MYC标签抗体,在细胞裂解物和/或培养基中证明每个转基因的表达。用在265bp最小肾病蛋白启动子控制下表达CFH转基因的质粒转染HEK293T细胞,表达并分泌人H因子。

[0076] 图10-用AAV2/9 265-CFH转导或用编码CFH的质粒转染H因子突变的足细胞(“人早期疾病(ED)足细胞”)

[0077] (A) 使用ELISA测定法分析人H因子浓度。用含有CFH转基因的AAV2/9病毒转导的足细胞在培养基中表现出比未转导的对照更高浓度的人H因子。(B) 来自(A)的平均人H因子浓度。(C) 使用ELISA测定法分析人H因子浓度。用在265bp最小肾病蛋白启动子的控制下表达CFH转基因的质粒转染的足细胞,表现出比未转染对照更高的人H因子浓度。

[0078] 图11-WT小鼠的肾中的CFH表达

[0079] 基因表达概况显示与用盐水对照注射的小鼠相比,用pAAV.NPHS1(265).hCFH.WPRE.bGH注射的野生型小鼠中的(A)病毒ITR,(B)人CFH cDNA,以及(C)病毒颗粒的增加。(D) 衍生自用盐水对照注射(m1)或用AAV基因疗法产品(pAAV.NPHS1(265).hCFH.WPRE.bGH,n=3)注射(m2/m3/m4)的野生型C57BL6小鼠的肾切片的免疫荧光成像。箭头表示小鼠肾小球中肾病蛋白和CFH的共定位)。缩放=20倍。

[0080] 图12-在肾细胞中使用265bp与818bp与FL最小肾病蛋白启动子(质粒转染)表达

## GFP

[0081] 用驱动eGFP的含有全长(PS0281)、818bp(Ps0301)和265bp(PS0282)最小肾病蛋白启动子的质粒转染的细胞中eGFP的表达。(A) HEK293T细胞中的转染结果。(B) 近端管上皮细胞(PTEC)中的转染结果。(C) 人足细胞中的转染结果。(D) 肾小球内皮细胞(GENC)中的转染结果。

[0082] 图13-在肾细胞中使用265bp与FL最小启动子(AAV转导)表达GFP

[0083] 用并入转基因盒的AAV颗粒转导的细胞中eGFP的表达,该转基因盒含有驱动eGFP的全长(PS0281)和265bp(PS0282)最小肾病蛋白启动子。(A) 人足细胞中的转导结果。(B) 近端管上皮细胞(PTEC)中的转导结果。(C) 肾小球内皮细胞(GENC)中的转导结果。(D) 系膜细胞中的转导结果。

[0084] 图14-使用265bp与FL最小肾病蛋白启动子表达人足细胞素

[0085] (A) 通过FACs分析Lenti-X 293T人胚肾细胞中的蛋白质表达。用3 $\mu$ g的以1:3的比率的质粒与PEI转染细胞。每个质粒表达由全长(FL.NPSH1-足细胞素-HA)或265bp(265.hNPSH1-足细胞素-HA)最小肾病蛋白启动子驱动的人足细胞素-HA。在转染后第2天,由抗HA抗体证明了人足细胞素的表达。 $N=3$ 数据表达为平均值 $\pm$ SD。(B) 通过FACs分析人足细胞中的蛋白质表达。用并入由全长(FL.NPSH1-足细胞素-HA)或265bp(265.hNPSH1-足细胞素-HA)最小肾病蛋白启动子驱动的人足细胞素-HA转基因盒的AAV颗粒转导细胞。在转导和分化后第9天,由抗HA抗体证明了人足细胞素的表达。

[0086] 图15-经由AAV转导使用265bp与FL最小肾病蛋白启动子体内表达GFP

[0087] 265肾病蛋白启动子和FL肾病蛋白启动子在WT小鼠(C57/B16)中进行了体内评价。生成在全长(FL-GFP)或者265bp(265-GFP)最小肾病蛋白启动子的控制下表达GFP的AAV2/9颗粒。每只小鼠尾静脉注射 $\sim 1 \times 10^{13}$ AAV2/9(每个治疗组2x小鼠,以及1x未治疗小鼠)。注射后4周进行肾组织的收获。进行qPCR分析,并将GFP拷贝数相对于18S(管家基因)归一化。

[0088] 发明详述

[0089] 现在将通过非限制性实例来描述本发明的各种优选特征和实施方案。

[0090] 必须注意的是,如本文和所附权利要求中所使用的,单数形式“一个(a,an)”和“该”包括复数指示物,除非上下文另外明确指出。

[0091] 本文使用的术语“包含(comprising,comprises和comprised of)”与“包括(including和includes)”或“含有(containing和contains)”同义,并且是包容性的或开放式的,并且不排除额外的、未陈述的成员、元素或方法步骤。术语“包含(comprising,comprises和comprised of)”还包括术语“由.....组成”。

[0092] 本文在讨论的出版物仅因为其在在本申请的提交日之前的公开而提供。本文中的任何内容均不应解释为承认此类出版物构成本文所附权利要求的现有技术。

[0093] 本公开不限于本文公开的示例性方法和材料,并且与本文描述的那些类似或等同的任何方法和材料可以用于本公开的实施方案的实践或测试中。数字范围包括定义该范围的数字。除非另有说明,否则任何核酸序列均以5'至3'方向从左到右书写;氨基酸序列分别以氨基到羧基的方向从左到右书写。

[0094] 除非另有说明,否则本发明的实践将采用化学、生物化学、分子生物学、微生物学和免疫学的常规技术,这些技术在本领域普通技术人员的能力之内。文献中对此类技术进

行了解释。参见,例如,Sambrook,J.,Fritsch,E.F.和Maniatis,T.(1989)Molecular Cloning:A Laboratory Manual,2nd Edition,Cold Spring Harbor Laboratory Press; Ausubel,F.M.等人(1995and periodic supplements)Current Protocols in MolecularBiology,Ch.9,13和16,JohnWiley&Sons;Roe,B.,Crabtree,J.和Kahn,A.(1996)DNA Isolation and Sequencing:Essential Techniques,JohnWiley&Sons;Polak,J.M.和 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;以及Lilley,D.M.和Dahlberg,J.E.(1992)Methods in Enzymology: DNA Structures Part A:Synthesis and Physical Analysis ofDNA,Academic Press。这些通用文本中的每个都通过引用并入本文。

[0095] 启动子

[0096] 本发明提供了用于肾小球基因疗法的启动子。

[0097] “启动子”在本文中根据其典型的含义使用,以指调控和/或起始从DNA序列转录RNA的DNA序列。

[0098] 合适地,启动子在哺乳动物细胞例如人细胞中是可操作的。启动子可以能够驱动哺乳动物细胞例如人细胞中的转基因表达。启动子可以是哺乳动物启动子,例如人启动子。

[0099] 启动子可以是在肾细胞中可操作的。本发明的启动子可以能够驱动肾中的转基因表达。肾细胞(启动子可以在其中是可操作的)的实例包括但不限于肾小球细胞。

[0100] 启动子可以是在肾小球细胞中可操作的。本发明的启动子可以能够驱动肾小球中的转基因表达。成熟的肾小球含有四种细胞类型:形成鲍曼囊的壁层上皮细胞(parietal epithelial cell)、覆盖肾小球滤过屏障最外层的足细胞、与血液直接接触的糖萼包被的有孔内皮细胞(fenestrated endothelial cell)以及位于毛细血管袢之间的系膜细胞(Vaughan,M.R.和Quaggin,S.E.,2008.Journal oftheAmerican Society ofNephrology, 19(1),pp.24-33)。

[0101] 启动子可以是在足细胞中可操作的。本发明的启动子可以能够驱动足细胞中的转基因表达。

[0102] 本发明的启动子可以是组织特异性启动子。如本文所用,“组织特异性启动子”是优先促进转基因在特定类型的细胞或组织中的表达的启动子。合适地,与其他细胞类型相比,组织特异性启动子可以促进转基因在一种细胞类型中更高的表达。例如,与其他细胞类型中的表达水平相比,组织特异性启动子可以是促进转基因在一种细胞类型中的表达水平至少10%更高、至少20%更高、至少30%更高、至少40%更高、至少50%更高、至少100%更高、至少200%更高、至少300%更高、至少400%更高、至少500%更高或至少1000%更高的启动子。

[0103] 合适地,启动子是肾特异性启动子。在一些实施方案中,启动子是肾小球特异性启动子。在一些实施方案中,启动子是足细胞特异性启动子。

[0104] 转基因表达可以通过本领域已知的任何合适的方法来测量。例如,通过测量可操作地连接至启动子的报告转基因例如绿色荧光蛋白(GFP)的表达,其中报告转基因的表达与启动子促进基因表达的能力相关。报告转基因例如GFP的表达可以通过任何合适的方法例如FACS来确定。例如,与其他细胞类型(例如CNS细胞、视网膜细胞、肺细胞、胰腺细胞、心

脏细胞或肌肉细胞)相比,在肾细胞中,肾特异性启动子可以促进报告转基因更高的表达。例如,与其他细胞类型相比,在条件永生化足细胞中,足细胞特异性启动子可以促进报告转基因更高的表达。合适的足细胞细胞系会是本领域技术人员众所周知的,例如CIHP-1。生成永生化足细胞的方法对于本领域技术人员会是众所周知的。Ni, L. 等人, 2012. *Nephrology*, 17(6), pp. 525-531中描述了合适的方法。

[0105] 合适地,启动子是最小肾特异性启动子。在一些实施方案中,启动子是最小肾小球特异性启动子。在一些实施方案中,启动子是最小足细胞特异性启动子。

[0106] 启动子可以具有约1.2kb或更短的长度。合适地,启动子具有约1.18kb或更短、约1.17kb或更短、约1.16kb或更短、约1.15kb或更短、约1.14kb或更短、约1.13kb或更短、约1.12kb或更短、约1.11kb或更短,或约1.10kb或更短的长度。合适地,启动子具有约1.15kb或更短的长度。

[0107] 启动子可以具有约1.1kb或更短的长度。在一些实施方案中,启动子具有约1.1kb或更短、约1.0kb或更短、约0.9kb或更短、约0.8kb或更短、约0.7kb或更短、约0.6kb或更短、约0.5kb或更短、约0.4kb或更短,或约0.3kb或更短的长度。

[0108] 在一些实施方案中,启动子具有约0.8kb或更短、约0.7kb或更短、约0.6kb或更短、约0.5kb或更短、约0.4kb或更短,或约0.3kb或更短的长度。在一些实施方案中,启动子具有818bp或更短的长度。在一些实施方案中,启动子具有800bp或更短的长度。

[0109] 在一些实施方案中,启动子具有约0.5kb或更短、约0.4kb或更短、或约0.3kb或更短的长度。在一些实施方案中,启动子具有约0.3kb或更短的长度。

[0110] 启动子可以具有约250bp或更长的长度。在一些实施方案中,启动子具有约250-1100bp、250-1000bp、250-900bp、250-800bp、250-700bp、250-600bp、250-500bp、250-400bp、250-300bp的长度。

[0111] 启动子可以具有约265bp或更长的长度。在一些实施方案中,启动子具有约265-1100bp、265-1000bp、265-900bp、265-800bp、265-700bp、265-600bp、265-500bp、265-400bp、265-300bp的长度。

[0112] 在一个实施方案中,启动子具有250-300bp、250-280bp、255-275bp、260-270bp或约265bp的长度。在一个实施方案中,启动子具有800-850bp、800-840bp、810-830bp、815-825bp、约819bp或约818bp的长度。

[0113] 肾病蛋白启动子

[0114] 启动子可以是合成的启动子。如本文所用,“合成的启动子”是不天然存在的启动子。通常,合成的启动子将通过现存的启动子的遗传修饰来产生。因此,合成的启动子可以衍生自现存的启动子,即除了包含一个或多个遗传修饰(例如取代、缺失或插入)外,与现存的启动子相同。

[0115] 合适地,启动子可以衍生自肾病蛋白(NPHS1)启动子。合适地,启动子可以是最小肾病蛋白启动子。NPHS1基因编码肾病蛋白,其在足细胞中选择性表达。

[0116] Moeller等人2002 *J Am Soc Nephrol*, 13(6):1561-7和Wong MA等人2000 *Am J Physiol Renal Physiol*, 279(6):F1027-32中已经描述了人NPHS1启动子。这个NPHS1启动子是一个1.2kb片段,并且似乎是足细胞特异性的。该1.2kb启动子区缺少TATA框,但具有其他转录因子的识别基序(例如PAX-2结合原件、E框和TATA共有序列)。

[0117] 示例性肾病蛋白启动子如SEQ ID NO:1、SEQ ID NO:14、SEQ ID NO:44,和SEQ ID NO:45所示。

[0118] 示例性肾病蛋白启动子-1,249bp (SEQ ID NO:1)

[0119] CCTGCAGGGCCACTAGTCTGTAATCCCAGCATTTTGGGAGGCTGAGGCAGATGGATCACCTGAGGTCA  
GGAGTTTCGAGACCAGCCTGGCCAACATGATGAAACCCCGTCTCTAGTAAAAATACAAAAATTAGCCAGGCATGGTGC  
TATATACCTGTAGTACCAGCTACTTGGGAGACAGAGGTGGGAGAATTACTTGAACCTGGGAGGTTCAAGCCATGGGA  
GGTGAAGTTGCAGTGAGCCGAGATGCCACTGCACTCCAGCCTGAGCAACAGAGCAAGACTATCTCAAGAAAAAAAA  
GAAAGAAAGAAAGGGACTTGCCAAGGTCATGTATCAGGGCAAGGAAGAGCTGGGGGCCAGCTGGCTGCTCCCCTGC  
TGAGCTGGGAGACCACCTTGATCTGACTTCTCCCATCTTCCCAGCCTAAGCCAGGCCCTGGGGTACGGAGGCTGGG  
GAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGAATTTTATGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGC  
TGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAGGTTGCACTGTGAGAATGAGCTCAAGCTGGGTC  
AGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGCTCTCCTAGGCTCCTGGCCTGCTGGACTCTGGG  
CTGCAGGTCCTTCTTGAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGTGGCAGGAGAGAAGATAGAGA  
TTGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAGACAAAAGGAGAGAGAACGGCTTAGACAAGGAG  
AGAAAGATGGAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCTGTAATCCCAACACTTGGGGAGGCCAAGGTGGG  
AGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCCCGTCTCTAAAAAAAAAAAAAGAAA  
AAAAAAAAAGAAAAAAGAAAAAAGTTTTTTTAAAGAGACAGAGAAAGAGACTCAGAGATTGAGACTGAGAGCAAGAC  
AGAGAGAGACACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGGAGGAGTAACGGAAAGAGATAAAAAAGA  
AAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGT  
GGGGTACAGTAGGGGGACCTGTG

[0120] 示例性变体肾病蛋白启动子-1,192bp (SEQ ID NO:14)

[0121] CACCTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGATGAAACCCCGTCTCTAGTAAAAATACAA  
AAATTAGCCAGGCATGGTGTATATACCTGTAGCACCAGCTACTTGGGAGACAGAGGTGGGAGAATTACTTGAACCT  
GGGAGGTTCAAGCCATGGGAGGTGGAAGTTGCAGTGAGCCGAGATGCCACTGCACTCCAGCCTGAGCAACAGAGCAA  
GACTATCTCAAGAAAAAGAAAGAAAGAAAGAGACTTGCCAAGGTCATGTATCAGGGCAAGGAAGAGCTGGGGGC  
CCAGCTGGCTGCTCCCCTGCTGAGCTGGGAGACCACCTTGATCTGACTTCTCCCATCTTCCCAGCCTAAGCCAGGCC  
CTGGGGTACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTATGCTCCAGCTGG  
GCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAGGTTGCACTGTGA  
GAATGAGCTCAAGCTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGCTCTCCTAGGCTC  
CTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGT  
GGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAGACAAAAGGAG  
AGAGAACGGCTTAGACAAGGAGAGAAAGATGAAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCTGTAATCCCAA  
CACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCC  
GTCTCTAAAAAAAAAAGAAAAAAGAAAAAAGAAAAAAGTTTTTTTAAAGAGACAGAGAAAGAGACTCAGAG  
ATTGAGACTGAGAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGGAGGAGAGTA  
ACGAAAGAGATAAAAAAGAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGG  
TGGCTGGCAGCGGGCGCTGTGGGGTACAGTAGGGGGACCTGTG

[0122] 示例性变体肾病蛋白启动子-1,192bp (SEQ ID NO:44)

[0123] CACCTGAGGTCAGGAGTTCGAGACCAGCGTGGCCAACATGATGAAACCCCGTCTCTAGTAAAAATACAA

AAATTAGCCAGGCATGGTGCTATATACCTGTAGCACCAGCTACTTGGGAGACAGAGGTGGGAGAATTACTTGAACCT  
GGGAGGTTCAAGCCATGGGAGGTGGAAGTTGCAGTGAGCCGAGATGCCACTGCACTCCAGCCTGAGCAACAGAGCAA  
GACTATCTCAAGAAAAGAAAAGAAAAGAAAAGAGACTTGCCAAGGTCATGTATCAGGGCAAGGAAGAGCTGGGGGC  
CCAGCTGGCTGCTCCCCTGCTGAGCTGGGAGACCACCTTGATCTGACTTCTCCCATCTTCCCAGCCTAAGCCAGGCC  
CTGGGGTCACGGAGGCTGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGAATTTTATGCTCCAGCTGG  
GCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAGGTTGCACTGTGA  
GAATGAGCTCAAGCTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGCTCTCCTAGGCTC  
CTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGT  
GGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAGACAAAAGGAG  
AGAGAACGGCTTAGACAAGGAGAGAAAAGATGGAAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCTGTAATCCCAA  
CACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCCC  
GTCTCTAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAAAGAGACTCAGAG  
ATTGAGACTGAGAGCAAGACAGAGAGAGACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGGAGGAGAGTA  
ACGAAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGG  
TGGCTGGCAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGTC

[0124] 示例性变体肾病蛋白启动子-1, 192bp (SEQ ID NO: 45)

[0125] CACCTGAGGTCAGGAGTTTCGAGACCAGCGTGGCCAACATGATGAAACCCCGTCTCTAGTAAAAATACAA  
AAATTAGCCAGGCATGGTGCTATATACCTGTAGCACCAGCTACTTGGGAGACAGAGGTGGGAGAATTACTTGAACCT  
GGGAGGTTCAAGCCATGGGAGGTGGAAGTTGCAGTGAGCCGAGATGCCACTGCACTCCAGCCTGAGCAACAGAGCAA  
GACTATCTCAAGAAAAGAAAAGAAAAGAAAAGAGACTTGCCAAGGTCATGTATCAGGGCAAGGAAGAGCTGGGGGC  
CCAGCTGGCTGCTCCCCTGCTGAGCTGGGAGACCACCTTGATCTGACTTCTCCCATCTTCCCAGCCTAAGCCAGGCC  
CTGGGGTCACGGAGGCTGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTATGCTCCAGCTGG  
GCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAGGTTGCACTGTGA  
GAATGAGCTCAAGCTGGGTCAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGCTCTCCTAGGCTC  
CTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCAGGAGTGAGGGGT  
GGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAGACAAAAGGAG  
AGAGAACGGCTTAGACAAGGAGAGAAAAGATGGAAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCTGTAATCCCAA  
CACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATAGTGAGACCCC  
GTCTCTAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAAAGAGACTCAGAG  
ATTGAGACTGAGAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAAGGGAGGAGAGTA  
ACGAAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGG  
TGGCTGGCAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGTC

[0126] 合适地, 本发明的启动子衍生自SEQ ID NO: 1或与SEQ ID NO: 1具有至少70%同一性的变体。合适地, 本发明的启动子与以下相比具有一个或多个缺失: SEQ ID NO: 1或与SEQ ID NO: 1具有至少70%同一性的变体。合适地, 变体与SEQ ID NO: 1可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。示例性变体是SEQ ID NO: 14、SEQ ID NO: 44或SEQ ID NO: 45。

[0127] 合适地, SEQ ID NO: 1的变体包含一个或多个以下取代: A493G、C1080T、G1169C和C1249G。合适地, SEQ ID NO: 1的变体包含取代A493G、C1080T、G1169C和C1249G的每个(参见

SEQ ID NO:14)。

[0128] 合适地,本发明的启动子衍生自SEQ ID NO:14、SEQ ID NO:44或SEQ ID NO:45,或与SEQ ID NO:14、SEQ ID NO:44或SEQ ID NO:45具有至少70%同一性的变体。合适地,本发明的启动子与以下相比,具有一个或多个缺失:SEQ ID NO:14、SEQ ID NO:44或SEQ ID NO:45,或与SEQ ID NO:14、SEQ ID NO:44或SEQ ID NO:45具有至少70%同一性的变体。合适地,变体与SEQ ID NO:14、SEQ ID NO:44或SEQ ID NO:45可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0129] 合适地,本发明的启动子衍生自SEQ ID NO:14或与SEQ ID NO:14具有至少70%同一性的变体。合适地,本发明的启动子与以下相比具有一个或多个缺失:SEQ ID NO:14或与SEQ ID NO:14具有至少70%同一性的变体。合适地,变体与SEQ ID NO:14可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0130] 合适地,本发明的启动子包含以下或由以下组成:根据SEQ ID NO:1并且具有一个或多个缺失,例如一个或两个缺失的核苷酸序列,或与其具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列。合适地,本发明的启动子包含以下或由以下组成:根据SEQ ID NO:1并且具有两个缺失的核苷酸序列,或与其具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列。缺失可以是任何大小。合适地,缺失各自的大小为至少50bp、至少100bp、至少150bp、至少200bp、至少250bp、至少300bp、至少350bp或至少400bp。合适地,缺失各自的大小为50至500bp、100至500bp、150至500bp、200至500bp、250至500bp、300至500bp、350至500bp、或400至500bp。

[0131] 在一些实施方案中,本发明的启动子包含根据SEQ ID NO:1的核苷酸序列或由根据SEQ ID NO:1的核苷酸序列组成,但其中:

[0132] (i) 缺失了SEQ ID NO:1的位置1至位置n1,其中n1为1至430的整数;和/或

[0133] (ii) 缺失了SEQ ID NO:1的位置n2至位置n3,其中 $n3 \geq n2$ , n2为508至1061的整数, n3为508至1061的整数;

[0134] 或与其具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、或至少99%的同一性的核苷酸序列。

[0135] 在一些实施方案中:

[0136] (iii) 对应于SEQ ID NO:1的位置493的位置是G,对应于SEQ ID NO:1的位置1080的位置是T,对应于SEQ ID NO:1的位置1169的位置是C,对应于SEQ ID NO:1的位置1249的位置是G。

[0137] 在优选的实施方案中:

[0138] (iii) 对应于SEQ ID NO:1的位置493的位置是G,对应于SEQ ID NO:1的位置1080的位置是T,对应于SEQ ID NO:1的位置1169的位置是C,对应于SEQ ID NO:1的位置1249的位置是G。

[0139] 尽管以上核苷酸位置是通过指SEQ ID NO:1中的位置来提及的,但是技术人员将能够容易地识别出其变体中的相应核苷酸位置,例如通过比对SEQ ID NO:1和变体序列。例如,SEQ ID NO:1的位置493对应于SEQ ID NO:14的位置436;SEQ ID NO:1的位置1080对应

于SEQ ID NO:14的位置1023;SEQ ID NO:1的位置1169对应于SEQ ID NO:14的位置1112;和SEQ ID NO:1的位置1249对应于SEQ ID NO:14的位置1192。

[0140] 例如,本发明的启动子可以包含以下或由以下组成:与根据SEQ ID NO:1的核苷酸序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸,但其中:

[0141] (i) 缺失了SEQ ID NO:1的位置1至位置n1,其中n1为1至430的整数;和/或

[0142] (ii) 缺失了SEQ ID NO:1的位置n2至位置n3,其中 $n3 \geq n2$ ,n2为508至1061的整数,n3为508至1061的整数。

[0143] 在一些实施方案中:

[0144] (iii) 用G取代SEQ ID NO:1的位置493,用T取代SEQ ID NO:1的位置1080,用C取代SEQ ID NO:1的位置1169,并且用G取代SEQ ID NO:1的位置1249。

[0145] 在优选的实施方案中:

[0146] (iii) 用G取代SEQ ID NO:1的位置493,用T取代SEQ ID NO:1的位置1080,用C取代SEQ ID NO:1的位置1169,并且用G取代SEQ ID NO:1的位置1249。

[0147] 合适地,n1是50至430、100至430、150至430、200至430、250至430、300至430、350至430或400至430的整数。在一些实施方案中,n1是100至430的整数。在一些实施方案中,n1=430,即在本发明的启动子中,缺失了SEQ ID NO:1的位置1至位置430。

[0148] n3和n2之间的差异指定了缺失的大小。合适地, $n3 \geq n2+49$ 、 $n3 \geq n2+99$ 、 $n3 \geq n2+149$ 、 $n3 \geq n2+199$ 、 $n3 \geq n2+249$ 、 $n3 \geq n2+299$ 、 $n3 \geq n2+349$ 、 $n3 \geq n2+399$ 、 $n3 \geq n2+449$ 、 $n3 \geq n2+499$ 或 $n3 \geq n2+549$ 。在一些实施方案中, $n3 \geq n2+49$ 。

[0149] n2和n3的值确定了缺失位于何处。合适地,n2和n3各自为550至1050的整数,n2和n3各自为600至1000的整数,n2和n3各自为650至950的整数,n2和n3各自为700至900的整数,n2和n3各自为750到850之间的整数。

[0150] 在一些实施方案中,n2=508并且n3=1061,即在本发明的启动子中,缺失了SEQ ID NO:1的位置508至1061。

[0151] 启动子区

[0152] 本发明人已经确定了驱动转基因表达的肾病蛋白启动子的区域。

[0153] 启动子通常包含“核心”和“近端”区。“核心启动子区”可以包含转录起始位点、RNA聚合酶结合位点和通用转录因子结合位点。“近端启动子区”可以包含例如促进有效且可控的转录所需的基本的(primary)调控元件和特异性转录因子结合位点。核心启动子区和近端启动子区两者的大小和组分通常以基因特异性方式变化。启动子还可以包含核心启动子区下游和起始密码子上游的5'非翻译区(5' UTR)(也称为前导序列)。(参见例如图1)。

[0154] 启动子可以是杂合启动子。如本文所用,“杂合启动子”包含衍生自不同启动子的元件的组合。例如,杂合启动子可以包含衍生自一个现存的启动子的近端启动子区和来自另一个现存的启动子的核心启动子,以实现期望的转基因表达。Piekarowicz,K.等人(2019).Methods&clinical development,15,157-169中描述了肌肉杂合启动子。

[0155] 在一些实施方案中,本发明的启动子包含(i)如SEQ ID NO:4所示的核苷酸序列或与SEQ ID NO:4是70%相同的变体。不希望受理论束缚,认为与SEQ ID NO:4具有至少约70%同一性的核苷酸序列可以提供近端启动子区。

[0156] 示例性近端启动子区 (SEQ ID NO:4)

[0157] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGAATTTA  
TGCTCCAG

[0158] 合适地,变体与SEQ ID NO:4可以是至少75%、至少80%、至少85%、至少90%、至少95%至少96%、至少97%、至少98%或至少99%相同的。启动子可以包含如SEQ ID NO:17所示的SEQ ID NO:4的变体。合适地,SEQ ID NO:4的变体包含取代A63G (参见SEQ ID NO:17)。

[0159] 在一些实施方案中,本发明的启动子包含 (i) 如SEQ ID NO:17所示的核苷酸序列或与SEQ ID NO:17至少70%相同的变体。合适地,变体与SEQ ID NO:17可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0160] 示例性变体近端启动子区 (SEQ ID NO:17)

[0161] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCC TGGCATGTGCTGACAGGGGATTT  
TATGCTCCAG

[0162] 在一些实施方案中,本发明的启动子包含 (ii) 如SEQ ID NO:5或18所示的核苷酸序列,或与SEQ ID NO:5或18至少70%相同的变体;如SEQ ID NO:6或19所示的核苷酸序列,或与SEQ ID NO:6或19至少70%相同的变体;和/或如SEQ ID NO:7或20所示的核苷酸序列,或与SEQ ID NO:7或20至少70%相同的变体。

[0163] 在一些实施方案中,本发明的启动子包含 (ii) 如SEQ ID NO:5所示的核苷酸序列或与SEQ ID NO:5至少70%相同的变体。不希望受理论束缚,认为与SEQ ID NO:5具有至少约70%同一性的核苷酸序列可以提供核心启动子区。

[0164] 示例性核心启动子区 (SEQ ID NO:5)

[0165] GAGCAAGACAGAGAGAGACACTCACAGGGAAG

[0166] 合适地,变体与SEQ ID NO:5可以是至少75%、至少80%、至少85%、至少90%、至少95%至少96%、至少97%、至少98%或至少99%相同的。启动子可以包含如SEQ ID NO:18所示的SEQ ID NO:5的变体。合适地,SEQ ID NO:5的变体包含取代C19T (参见SEQ ID NO:18)。

[0167] 在一些实施方案中,本发明的启动子包含 (ii) 如SEQ ID NO:18所示的核苷酸序列或与SEQ ID NO:18至少70%相同的变体。合适地,变体与SEQ ID NO:18可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0168] 示例性变体核心启动子区 (SEQ ID NO:18)

[0169] GAGCAAGACAGAGAGAGATACTCACAGGGAAG

[0170] 在一些实施方案中,本发明的启动子包含 (ii) 如SEQ ID NO:6所示的核苷酸序列或与SEQ ID NO:6至少70%相同的变体。不希望受理论束缚,认为与SEQ ID NO:6具有至少约70%同一性的核苷酸序列可以提供5' UTR。

[0171] 示例性5' UTR (SEQ ID NO:6)

[0172] AGGGGAAGAGGAAAACGAGAAAGGGAGGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGA  
GACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGTTCACAGTAGG  
GGGACCTGTC

[0173] 合适地,变体与SEQ ID NO:6可以是至少75%、至少80%、至少85%、至少90%、至

少95%、至少96%、至少97%、至少98%或至少99%相同的。启动子可以包含如SEQ ID NO:19所示的SEQ ID NO:6的变体。合适地,SEQ ID NO:6的变体包含取代G76C和/或C156G(参见SEQ ID NO:19)。

[0174] 在一些实施方案中,本发明的启动子包含(ii)如SEQ ID NO:19所示的核苷酸序列或与SEQ ID NO:19至少70%相同的变体。合适地,变体与SEQ ID NO:19可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0175] 示例性变体5' UTR(SEQ ID NO:19)

[0176] AGGGGAAGAGGAAAACGAGAAAGGGAGGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTACAGTAGGGGACCTGTG

[0177] 在一些实施方案中,本发明的启动子包含(ii)如SEQ ID NO:7所示的核苷酸序列或与SEQ ID NO:7至少70%相同的变体。不希望受理论束缚,认为与SEQ ID NO:7具有至少约70%同一性的核苷酸序列可以提供核心启动子区和5' UTR。

[0178] 示例性核心启动子区和5' UTR(SEQ ID NO:7)

[0179] GAGCAAGACAGAGAGAGACTCACAGGGAAGAGGGGAAGAGAAAACGAGAAAGGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTACAGTAGGGGGACCTGTC

[0180] 合适地,变体与SEQ ID NO:7可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。启动子可以包含如SEQ ID NO:20所示的SEQ ID NO:7的变体。合适地,SEQ ID NO:7的变体包含取代C19T、G108C和/或C188G(参见SEQ ID NO:20)。合适地,SEQ ID NO:7的变体包含取代C19T、G108C和C188G(参见SEQ ID NO:20)。

[0181] 在一些实施方案中,本发明的启动子包含(ii)如SEQ ID NO:20所示的核苷酸序列或与SEQ ID NO:20至少70%相同的变体。合适地,变体与SEQ ID NO:20可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0182] 示例性变体核心启动子区和5' UTR(SEQ ID NO:20)

[0183] GAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGAAAACGAGAAAGGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTACAGTAGGGGGACCTGTG

[0184] 在一些实施方案中,本发明的启动子包含(iii)与SEQ ID NO:8具有至少70%同一性的核苷酸序列或其一个或多个片段。合适地,在紧邻近端启动子区的下游和/或紧邻核心启动子区的上游,启动子包含与SEQ ID NO:8具有至少约70%同一性的核苷酸序列或其一个或多个片段。

[0185] 启动子可以包含与SEQ ID NO:8具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列或其一个或多个片段。启动子可以包含SEQ ID NO:8的核苷酸序列或其一个或多个片段。

[0186] 示例性任选的启动子区(SEQ ID NO:8)

[0187] CTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAGTTGCACTGTGAGAATGAGCTCAAGCTGGGTTCAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGC

TCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCA  
GGAGTGAGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAG  
ACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAAGATGGAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCT  
GTAATCCCAACACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATA  
GTGAGACCCCGTCTCTAAAAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAA  
AGAGACTCAGAGATTGAGACTGA

[0188] 合适地,一个或多个片段是(a)5'末端片段;和/或(b)3'末端片段。合适地,5'末端片段可以紧邻近端启动子区的下游。合适地,3'末端片段可以紧邻近端启动子区的上游。例如,本发明的启动子可以包含:

[0189] (a)与SEQ ID NO:8的位置1至x具有至少70%的核苷酸序列;和/或

[0190] (b)与SEQ ID NO:8的位置y至554具有至少70%同一性的核苷酸序列;

[0191] 其中x和y是整数,且 $y > x$ 。

[0192] SEQ ID NO:8的片段可以是任何长度。合适地,片段可以具有约500bp或更短、450bp或更短、400bp或更短、350bp或更短、300bp或更短、250bp或更短、200bp或更短、150bp或更短、100bp或更短、50bp或更短、40bp或更短、30bp或更短、20bp或更短、或10bp或更短的长度。

[0193] 在一些实施方案中,启动子不包含SEQ ID NO:8。

[0194] 在一些实施方案中,本发明的启动子包含与SEQ ID NO:9具有至少70%同一性的核苷酸序列或其片段。合适地,在紧邻近端启动子区的上游,启动子包含与SEQ ID NO:9具有至少约70%同一性的核苷酸序列或其片段。

[0195] 启动子可以包含与SEQ ID NO:9具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的核苷酸序列或其片段。启动子可以包含SEQ ID NO:9的核苷酸序列或其一个或多个片段。

[0196] 示例性任意的上游启动子区(SEQ ID NO:9)

[0197] CCTGCAGGGCCCACTAGTCTGTAATCCAGCATTTTGGGAGGCTGAGGCAGATGGATCACCTGAGGTCA  
GGAGTTTCAGACCAGCCTGGCCAACATGATGAAACCCCGTCTCTAGTAAAAATACAAAATTAGCCAGGCATGGTGC  
TATATACCTGTAGTACCAGCTACTTGGGAGACAGAGGTGGGAGAATTACTTGAACCTGGGAGGTTCAAGCCATGGGA  
GGTGAAGTTGCAGTGAGCCGAGATGCCACTGCACTCCAGCCTGAGCAACAGAGCAAGACTATCTCAAGAAAAAAAA  
GAAAGAAAGAAAGGGACTTGCCAAGGTCATGTATCAGGGCAAGGAAGAGCTGGGGGCCAGCTGGCTGCTCCCCTGC  
TGAGCTGGGAGACCACCTTGATCTGACTTCTCCCATCTTCCAGCCTAAGCCA

[0198] 合适地,该片段是3'末端片段。例如,本发明的启动子包含与SEQ ID NO:9的位置z至430具有至少70%同一性的核苷酸序列,其中z是整数。

[0199] SEQ ID NO:9的片段可以是任何长度。合适地,片段可以具有约400bp或更短、350bp或更短、300bp或更短、250bp或更短、200bp或更短、150bp或更短、100bp或更短、50bp或更短、40bp或更短、30bp或更短、20bp或更短、或10bp或更短的长度。

[0200] 在一些实施方案中,启动子不包含SEQ ID NO:9。

[0201] 在一些实施方案中,本发明的启动子从5'至3'包含以下或由以下组成:

[0202] (i)与SEQ ID NO:4或17具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列;

[0203] (iii) 任选地与SEQ ID NO:8具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列或其一个或多个片段;以及

[0204] (ii) 与SEQ ID NO:5或18具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列,与SEQ ID NO:6或19具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列,和/或与SEQ ID NO:7或20具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0205] 在一些实施方案中,本发明的启动子从5'至3'包含以下或由以下组成:

[0206] (i) 与SEQ ID NO:4或17具有至少90%同一性的核苷酸序列;

[0207] (iii) 任选地与SEQ ID NO:8具有至少90%同一性的核苷酸序列或其一个或多个片段;以及

[0208] (ii) 与SEQ ID NO:5或18具有至少90%同一性的核苷酸序列、与SEQ ID NO:6或19具有至少90%同一性的核苷酸序列、和/或与SEQ ID NO:7或20具有至少90%同一性的核苷酸序列。

[0209] 在一些实施方案中,本发明的启动子从5'至3'包含以下或由以下组成:

[0210] (i) 与SEQ ID NO:4或17具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列;

[0211] (iii) 任选地 (a) 与SEQ ID NO:8的5'末端片段具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列;和/或 (b) 与SEQ ID NO:8的3'末端片段具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、或至少99%或100%同一性的核苷酸序列;以及

[0212] (ii) 与SEQ ID NO:5或18具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列,与SEQ ID NO:6或19具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列,和/或与SEQ ID NO:7或20具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0213] 在一些实施方案中,本发明的启动子从5'至3'包含以下或由以下组成:

[0214] (i) 与SEQ ID NO:4或17具有至少90%同一性的核苷酸序列;

[0215] (iii) 任选地 (a) 与SEQ ID NO:8的5'末端片段具有至少90%同一性的核苷酸序列;和/或 (b) 与SEQ ID NO:8的3'末端片段具有至少90%同一性的核苷酸序列;以及

[0216] (ii) 与SEQ ID NO:5或18具有至少90%同一性的核苷酸序列、与SEQ ID NO:6或19具有至少90%同一性的核苷酸序列、和/或与SEQ ID NO:7或20具有至少90%同一性的核苷酸序列。

[0217] 启动子元件

[0218] 本发明人已经确定了驱动转基因表达的肾病蛋白启动子的功能性元件。

[0219] 本发明的启动子可以包含以下元件的一个或多个：(a) 视黄酸受体结合位点；(b) WT1结合位点；(c) 增强子框；(d) 转录因子结合区；和(e) 转录起始位点。

[0220] 合适地，本发明的启动子包含所有以下元件：(a) 视黄酸受体结合位点；(b) WT1结合位点；(c) 增强子框；(d) 转录因子结合区；和(e) 转录起始位点。

[0221] 视黄酸受体 (RAR) 结合位点指能够结合RAR $\alpha$ 、RAR $\beta$ 和/或RAR  $\gamma$  的多核苷酸序列。RAR结合位点可以包含以下或由以下组成：如SEQ ID NO:10所示的核苷酸序列，或与SEQ ID NO:10相比具有一个或两个取代、缺失或插入的核苷酸序列。取代、缺失或插入可以是单个核苷酸的任何取代、缺失或插入，使得该RAR结合位点保留至少一种其内源性功能。

[0222] 示例性RAR结合位点 (SEQ ID NO:10)

[0223] GGGGTCA

[0224] WT1结合位点指能够结合由肾母细胞瘤抑制基因WT1编码的锌指多肽的多核苷酸序列。WT1结合位点可以包含以下或由以下组成：如SEQ ID NO:11所示的核苷酸序列，或与SEQ ID NO:11相比具有一个、两个或三个取代、缺失或插入的核苷酸序列。取代、缺失或插入可以是单个核苷酸的任何取代、缺失或插入，使得该WT1结合区保留至少一种其内源性功能。

[0225] 示例性WT1结合位点 (SEQ ID NO:11)

[0226] CGGAGGCTGGGAGGCA

[0227] 增强子框指在一些真核生物中发现的DNA应答元件，其充当蛋白质结合位点。增强子框可以包含以下或由以下组成：如SEQ ID NO:12所示的核苷酸序列，或与SEQ ID NO:12相比具有一个或两个取代、缺失或插入的核苷酸序列。取代、缺失或插入可以是单个核苷酸的任何取代、缺失或插入，使得该增强子框保留至少一种其内源性功能。

[0228] 示例性增强子框 (SEQ ID NO:12)

[0229] ATGTG

[0230] (a) 视黄酸受体结合位点；(b) WT1结合位点；以及(c) 增强子框中的一个或多个可以存在于近端启动子区中。合适地，(a) 视黄酸受体结合位点；(b) WT1结合位点；以及(c) 增强子框各自存在于近端启动子区中。

[0231] 在一些实施方案中，以下元件的一个或多个存在于(i) 与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列中：(a) 大概对应于SEQ ID NO:4或17的位置7至位置13的位置处的RAR结合位点；(b) 大概对应于SEQ ID NO:4或17的位置14至位置30的位置处的WT1结合位点；以及(c) 大概对应于SEQ ID NO:4或17的位置49至位置53的位置处的增强子框。在一些实施方案中，元件各自存在于(i) 与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列中。

[0232] 在一些实施方案中，以下核苷酸序列的一个或多个存在于(i) 与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列中：(a) 对应于SEQ ID NO:4或17的位置7至位置13的位置处的GGGGTCA；(b) 对应于SEQ ID NO:4或17的位置14至位置30的位置处的CGGAGGCTGGGAGGCA；以及(c) 对应于SEQ ID NO:4或17的位置49至位置53的位置处的ATGTG。在一些实施方案中，该核苷酸序列各自存在于(i) 与SEQ ID NO:4或17具有至少70%同一性的核苷酸序列中。

[0233] 合适地，启动子可以包括包含以下或由以下组成的转录因子结合区：如SEQ ID

NO:13所示的核苷酸序列,或与SEQ ID NO:13相比具有一个、两个、三个、四个或五个取代、缺失或插入的核苷酸序列。取代、缺失或插入可以是单个核苷酸的任何取代、缺失或插入,使得该转录因子结合区保留至少一种其内源性功能。示例性变体如SEQ ID NO:43所示。合适地,SEQ ID NO:13的变体包含取代C19T(参见SEQ ID NO:43)。

[0234] 示例性转录因子结合区(SEQ ID NO:13)

[0235] GAGCAAGACAGAGAGAGACTCACAGGGA

[0236] 合适地,启动子可以包括包含以下或由以下组成的转录因子结合区:如SEQ ID NO:43所示的核苷酸序列,或与SEQ ID NO:43相比具有一个、两个、三个、四个或五个取代、缺失或插入的核苷酸序列。取代、缺失或插入可以是单个核苷酸的任何取代、缺失或插入,使得该转录因子结合区保留至少一种其内源性功能。

[0237] 示例性转录因子结合区(SEQ ID NO:43)

[0238] GAGCAAGACAGAGAGAGATACTCACAGGGA

[0239] 其他合适的转录因子结合区将是本领域技术人员众所周知的。例如,其他合适的转录因子结合区包括TACGAT(SEQ ID NO:37)、TATAAT(SEQ ID NO:38)、GATACT(SEQ ID NO:39)、TATGAT(SEQ ID NO:40)和TATGTT(SEQ ID NO:41)。

[0240] 合适地,启动子可以包括包含“AG”二核苷酸或由“AG”二核苷酸组成的转录起始位点。

[0241] 合适地,转录因子结合位点可操作地连接至转录起始位点。合适地,转录因子结合位点可以直接位于转录起始位点的上游。不希望受理论束缚,认为转录因子结合位点和转录起始位点可以提供核心启动子区。

[0242] 合适地,启动子可以包含5'非翻译区。5'非翻译区可以包含与SEQ ID NO:6或19具有至少约70%、80%、90%、95%或99%序列同一性的核苷酸序列或由与SEQ ID NO:6或19具有至少约70%、80%、90%、95%或99%序列同一性的核苷酸序列组成。5'非翻译区可以包含SEQ ID NO:6或19或由SEQ ID NO:6或19组成。

[0243] 合适地,5'非翻译区可操作地连接至转录起始位点。合适地,5'非翻译区可以直接位于转录起始位点的下游。

[0244] 在一些实施方案中,本发明的启动子从5'至3'包含以下或由以下组成:

[0245] (i) 与SEQ ID NO:4或17具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列,其中存在以下元件的每个:(a)大概对应于SEQ ID NO:4或17的位置7至位置13的位置处的RAR结合位点;(b)大概对应于SEQ ID NO:4或17的位置14至位置30的位置处的WT1结合位点;以及(c)大概对应于SEQ ID NO:4或17的位置49至位置53的位置处的增强子框;

[0246] (iii) 任选地与SEQ ID NO:8具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列或其一个或多个片段;以及

[0247] (ii) 与SEQ ID NO:5或18具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列,与SEQ ID NO:6或19具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列,或与SEQ ID NO:7或20

具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0248] 在一些实施方案中,本发明的启动子从5'至3'包含以下或由以下组成:

[0249] (i)与SEQ ID NO:4或17具有至少90%同一性的核苷酸序列,其中存在以下元件的每个:(a)对应于SEQ ID NO:4或17的位置7至位置13的位置处的RAR结合位点;(b)对应于SEQ ID NO:4或17的位置14至位置30的位置处的WT1结合位点;以及(c)对应于SEQ ID NO:4或17的位置49至位置53的位置处的增强子框;

[0250] (iii)任选地与SEQ ID NO:8具有至少90%同一性的核苷酸序列或其一个或多个片段;以及

[0251] (ii)与SEQ ID NO:5或18具有至少90%同一性的核苷酸序列、与SEQ ID NO:6或19具有至少90%同一性的核苷酸序列、或与SEQ ID NO:7或20具有至少90%同一性的核苷酸序列。

[0252] 示例性启动子

[0253] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:2所示的核苷酸序列或与SEQ ID NO:2至少70%相同的变体。

[0254] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:2所示的核苷酸序列或与SEQ ID NO:2至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0255] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:2所示的核苷酸序列或与SEQ ID NO:2至少70%相同的变体。

[0256] 示例性最小肾病蛋白启动子-819bp(SEQ ID NO:2)

[0257] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGAATTTTA  
TGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGGAG  
GTTGCACTGTGAGAATGAGCTCAAGCTGGGTGAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGC  
TCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCA  
GGAGTGAGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAG  
ACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAAGATGGAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCT  
GTAATCCCAACACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATA  
GTGAGACCCCGTCTCTAAAAAAAAAAAAAAAAAGAAAAAAAAAAGAAAAAAAAAGTTTTTTTTAAAGAGACAGAGAA  
AGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGACTCACAGGGAAGAGGGGAAGAGGAAAACGAGAAA  
GGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAAGCCA  
GACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGTC

[0258] 合适地,变体与SEQ ID NO:2可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。启动子可以包含如SEQ ID NO:15或SEQ ID NO:46所示的SEQ ID NO:2的变体或由如SEQ ID NO:15或SEQ ID NO:46所示的SEQ ID NO:2的变体组成。

[0259] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:15所示的核苷酸序列或与SEQ ID NO:15至少70%相同的变体。

[0260] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:15所示的核苷酸序列或与SEQ ID NO:15至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0261] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:15所示的核苷酸序列或与SEQ ID NO:15至少70%相同的变体。

[0262] 合适地,变体与SEQ ID NO:15可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0263] 示例性最小肾病蛋白启动子变体-819bp (SEQ ID NO:15)

[0264] GGCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTA  
TGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAG  
GTTGCACTGTGAGAATGAGCTCAAGCTGGGTGAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGC  
TCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGTCCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCA  
GGAGTGAGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAG  
AGACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAAGATGGAAGATAAAGAGACTGGGCGCAGTGGCTCACGC  
CTGTAATCCCAACACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACA  
TAGTGAGACCCCGTCTCTAAAAAAAAAAAAAGAAAAAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAA  
AGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGATACTCACAGGAAGAGGGGAAGAGGAAAACGAGAAA  
GGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAAGCCA  
GACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGTG

[0265] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:46所示的核苷酸序列或与SEQ ID NO:46至少70%相同的变体。

[0266] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:46所示的核苷酸序列或与SEQ ID NO:46至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0267] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:46所示的核苷酸序列或与SEQ ID NO:46至少70%相同的变体。

[0268] 合适地,变体与SEQ ID NO:46可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0269] 示例性最小肾病蛋白启动子变体-819bp (SEQ ID NO:46)

[0270] GGCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTA  
TGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAG  
GTTGCACTGTGAGAATGAGCTCAAGCTGGGTGAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGC  
TCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGTCCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCA  
GGAGTGAGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAG  
ACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAAGATGGAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCT  
GTAATCCCAACACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATA  
GTGAGACCCCGTCTCTAAAAAAAAAAAAAGAAAAAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAA  
AGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGATACTCACAGGAAGAGGGGAAGAGGAAAACGAGAAA  
GGGAGGAGAGTAACGGAAAGAGATAAAAAAGAAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCA  
GACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGTG

[0271] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:58所示的核苷酸序列或与SEQ ID NO:58至少70%相同的变体。

[0272] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:58所示的核苷酸序列或与

SEQ ID NO:58至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0273] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:58所示的核苷酸序列或与SEQ ID NO:58至少70%相同的变体。

[0274] 示例性最小肾病蛋白启动子-818bp (SEQ ID NO:58)

[0275] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGAATTTTA  
TGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAG  
GTTGCACTGTGAGAATGAGCTCAAGCTGGGTGAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGC  
TCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCA  
GGAGTGAGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAG  
ACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAAGATGAAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCT  
GTAATCCCAACACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACATA  
GTGAGACCCCGTCTCTAAAAAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAA  
AGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGACTCACAGGAAGAGGGGAAGAGGAAAACGAGAAA  
GGGAGGAGAGTAACGAAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAAGCCA  
GACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTACAGTAGGGGGACCTGT

[0276] 合适地,变体与SEQ ID NO:58可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。启动子可以包含如SEQ ID NO:59或SEQ ID NO:60所示的SEQ ID NO:58的变体或由如SEQ ID NO:59或SEQ ID NO:60所示的SEQ ID NO:58的变体组成。

[0277] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:59所示的核苷酸序列或与SEQ ID NO:59至少70%相同的变体。

[0278] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:59所示的核苷酸序列或与SEQ ID NO:59至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0279] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:59所示的核苷酸序列或与SEQ ID NO:59至少70%相同的变体。

[0280] 合适地,变体与SEQ ID NO:59可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0281] 示例性最小肾病蛋白启动子变体-818bp (SEQ ID NO:59)

[0282] GGCCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTA  
TGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGGCAGAGGCCAGAGCTGGGGGCTCTGGAAGGTACCTGGGGGAG  
GTTGCACTGTGAGAATGAGCTCAAGCTGGGTGAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGC  
TCTCCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGGTCCTTCTTGAAAGGCTGTGAGTAGTGAGACAAGGAGCA  
GGAGTGAGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAG  
AGACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAAGATGAAAAGATAAAGAGACTGGGCGCAGTGGCTCACGC  
CTGTAATCCCAACACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGGAAAGAGTCTGAGATCAACCTGGCCAACA  
TAGTGAGACCCCGTCTCTAAAAAAAAAAAAAAAAAGAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAA  
AGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGAGATACTCACAGGAAGAGGGGAAGAGGAAAACGAGAAA  
GGGAGGAGAGTAACGAAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAAAGCCA  
GACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTACAGTAGGGGGACCTGT

[0283] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:60所示的核苷酸序列或与SEQ ID NO:60至少70%相同的变体。

[0284] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:60所示的核苷酸序列或与SEQ ID NO:60至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0285] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:60所示的核苷酸序列或与SEQ ID NO:60至少70%相同的变体。

[0286] 合适地,变体与SEQ ID NO:60可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0287] 示例性最小肾病蛋白启动子变体-818bp (SEQ ID NO:60)

[0288] GGCCTGGGGTCACGGAGGCTGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGATTTTA  
TGCTCCAGCTGGGCCAGCTGGGAGGAGCCTGCTGGCAGAGGCCAGAGCTGGGGCTCTGGAAGGTACCTGGGGGAG  
GTTGCACTGTGAGAATGAGCTCAAGCTGGGTGAGAGAGCAGGGCTGACTCTGCCAGTGCCTGCATCAGCCTCATCGC  
TCTCTAGGCTCCTGGCCTGCTGGACTCTGGGCTGCAGTCTCTTCTGAAAGGCTGTGAGTAGTGAGACAAGGAGCA  
GGAGTGAGGGGTGGCAGGAGAGAAGATAGAGATTGAGAGAGAGAGAGAGAGACAGAGAGAGAGGAAGAGACAGAG  
ACAAAAGGAGAGAGAACGGCTTAGACAAGGAGAGAAAAGATGAAAAGATAAAGAGACTGGGCGCAGTGGCTCACGCCT  
GTAATCCCAACACTTGGGGAGGCCAAGGTGGGAGGATGGCTTGAAGAAAAGAGTCTGAGATCAACCTGGCCAACATA  
GTGAGACCCCGTCTCTAAAAAAAAAAAAAAAAAGAAAAAAAAAAAAAAAAAGAAAAAAAAAGTTTTTTTAAAGAGACAGAGAA  
AGAGACTCAGAGATTGAGACTGAGAGCAAGACAGAGAGATACTCACAGGAAGAGGGGAAGAGGAAAACGAGAAA  
GGGAGGAGAGTAACGAAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAAGCCA  
GACAGACGCAGGTGGCTGGCAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGT

[0289] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:3所示的核苷酸序列或与SEQ ID NO:3至少70%相同的变体。

[0290] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:3所示的核苷酸序列或与SEQ ID NO:3至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0291] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:3所示的核苷酸序列或与SEQ ID NO:3至少70%相同的变体。

[0292] 示例性最小肾病蛋白启动子-265bp (SEQ ID NO:3)

[0293] GGCCTGGGGTCACGGAGGCTGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGAATTTTA  
TGCTCCAGGAGCAAGACAGAGAGAGACACTCACAGGAAGAGGGGAAGAGGAAAACGAGAAAGGGAGGAGAGTAACG  
GAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAAGCCAGACAGACGCAGGTGG  
CTGGCAGCGGGCGCTGTGGGGGTCACAGTAGGGGGACCTGTC

[0294] 合适地,变体与SEQ ID NO:3可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。合适地,SEQ ID NO:3的变体包含取代A63G、C96T、G185C和C265G的一个或多个。合适地,SEQ ID NO:3的变体包含取代A63G、C96T、G185C和C265G(参见SEQ ID NO:17-20)。启动子可以包含如SEQ ID NO:16或SEQ ID NO:47所示的SEQ ID NO:3的变体或由如SEQ ID NO:16或SEQ ID NO:47所示的SEQ ID NO:3的变体组成。

[0295] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:16所示的核苷酸序列或与SEQ ID NO:16至少70%相同的变体。

[0296] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:16所示的核苷酸序列或与SEQ ID NO:16至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0297] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:16所示的核苷酸序列或与SEQ ID NO:16至少70%相同的变体。

[0298] 合适地,变体与SEQ ID NO:16可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0299] 示例性最小肾病蛋白启动子变体-265bp (SEQ ID NO:16)

[0300] GGCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTA  
TGCTCCAGGAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGAAAACGAGAAAGGGAGGAGAGTAACG  
GAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACAGAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGG  
CTGGCAGCGGGCGCTGTGGGGTTCACAGTAGGGGGACCTGTC

[0301] 在一些实施方案中,本发明的启动子包含以下或由以下组成:如SEQ ID NO:47所示的核苷酸序列或与SEQ ID NO:47至少70%相同的变体。

[0302] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:47所示的核苷酸序列或与SEQ ID NO:47至少70%相同的变体并且其中启动子具有约1.1kb或更短的长度。

[0303] 在一些实施方案中,本发明的启动子包含如SEQ ID NO:47所示的核苷酸序列或与SEQ ID NO:47至少70%相同的变体。

[0304] 合适地,变体与SEQ ID NO:47可以是至少75%、至少80%、至少85%、至少90%、至少95%至少96%、至少97%、至少98%或至少99%相同的。

[0305] 示例性最小肾病蛋白启动子变体-265bp (SEQ ID NO:47)

[0306] GGCCTGGGGTCACGGAGGCTGGGGAGGCACCGAGGAACGCGCCTGGCATGTGCTGACAGGGGATTTTA  
TGCTCCAGGAGCAAGACAGAGAGAGATACTCACAGGGAAGAGGGGAAGAGAAAACGAGAAAGGGAGGAGAGTAACG  
GAAAGAGATAAAAAAGAAAAGCAGGTGGCAGAGACACACAGAGAGGGACCCAGAGAAAGCCAGACAGACGCAGGTGG  
CTGGCAGCGGGCGCTGTGGGGTTCACAGTAGGGGGACCTGTG

[0307] 多核苷酸

[0308] 本发明提供了多核苷酸,其包含本发明的启动子。

[0309] 本发明的多核苷酸可以包含DNA或RNA,优选DNA。它们可以是单链的或双链的。多核苷酸诸如DNA多核苷酸可以重组产生、合成产生或通过本领域技术人员可获得的任何方式产生。它们还可以通过标准技术克隆。多核苷酸可以是分离的多核苷酸。

[0310] 更长的多核苷酸通常将使用重组方法产生,例如使用聚合酶链式反应 (PCR) 克隆技术。这将涉及制备在期望克隆的靶序列侧翼的一对引物(例如约15至30个核苷酸)、将引物送至与从动物细胞或人类细胞获得的mRNA或cDNA接触、在导致期望的区扩增的条件下进行聚合酶链式反应、分离扩增的片段(例如通过用琼脂糖凝胶纯化反应混合物)以及回收扩增的DNA。引物可以设计为含有合适的限制性内切酶识别位点,使得扩增的DNA可以克隆到合适的载体中。

[0311] 多核苷酸可以通过本领域可获得的任何方法进行修饰。可以进行此类修饰以增强本发明的多核苷酸的体内活性或寿命。

[0312] 蛋白质编码序列

[0313] 启动子可以可操作地连接至一个或多个蛋白质编码序列。

[0314] 蛋白质编码序列可以编码任何感兴趣的多肽。例如,蛋白质编码序列可以编码与肾小球疾病相关的任何多肽。蛋白质编码序列可以编码涉及与GBM相关联的遗传性肾小球疾病的多肽,例如奥尔波特综合征。蛋白质编码序列可以编码涉及足细胞相关联的遗传性肾小球疾病的多肽。

[0315] 合适地,蛋白质编码序列可以编码COL4A3、COL4A4、COL4A5、NPHS2、CFH、CFL、CFHL1、C1INH、C4BP、MASP2、C3、C5aR1、C5、C5a、CD55、CD35、CD46、CD59、玻连蛋白、簇集素、ADCK4、ALG1、ARHGAP24、ARGHD1A、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或VEGFC多肽。

[0316] 合适地,蛋白质编码序列编码具有1450个氨基酸或更长、1500个氨基酸或更长、1550个氨基酸或更长、1600个氨基酸或更长、或1650个氨基酸或更长的长度的多肽。

[0317] 对于蛋白质编码多核苷酸,技术人员应当理解,由于遗传密码的简并性,许多不同的多核苷酸可以编码相同的多肽。另外,应当理解,技术人员可以使用常规技术进行不影响本发明的多核苷酸编码的多肽序列的核苷酸取代,以反映任何特定宿主生物体(本发明的多肽在其中待表达)的密码子选用。

[0318] 蛋白质编码序列可以是经密码子优化的。不同的细胞在其特定密码子的选用中不同。这种密码子偏倚对应于细胞类型中特定tRNA的相对丰度中的偏倚。通过改变序列中的密码子,使其与相应tRNA的相对丰度匹配,可以增加表达。出于同样的原因,可以通过故意选择相应tRNA在特定细胞类型中已知少见的密码子来减少表达。因此,可获得额外程度的翻译控制。对于哺乳动物细胞(例如人)以及对于多种其他生物体来说,密码子选用表是本领域已知的。

[0319] 本文公开的蛋白质编码核苷酸序列可以在其3'末端包含或缺乏终止密码子。因此,本公开涵盖本文公开的存在或不存在终止密码子的SEQ ID NO。

[0320] COL4A3多肽、COL4A4多肽和COL4A5多肽

[0321] 蛋白质编码序列可编码COL4A3多肽、COL4A4多肽或COL4A5多肽、或其片段或衍生物。

[0322] COL4A3蛋白、COL4A4蛋白或COL4A5蛋白是大约170-185kDa的同源多肽,其包含经常由非胶原序列中断并形成三螺旋重复的胶原Gly-XY重复序列。每个多肽还在羧基末端处含有大的球状非胶原结构域。

[0323] 奥尔波特综合征(AS)由COL4A3基因、COL4A4基因和COL4A5基因的致病性变异引起,致病性变异导致基底膜的IV型胶原蛋白 $\alpha$ 345网络异常。

[0324] COL4A3多肽、COL4A4多肽或COL4A5多肽,或其片段或衍生物可以能够形成IV型胶原蛋白 $\alpha$ 345网络。

[0325] 可以从每个COL4A3多肽、COL4A4多肽和COL4A5多肽中除去大约200-300个氨基酸以产生适合于小基因方法的截短的转基因。可以从三螺旋重复中除去氨基酸。优选地,不从非胶原区中去除氨基酸。

[0326] 在一些实施方案中,COL4A3多肽、COL4A4多肽和COL4A5多肽是全长多肽。

[0327] 优选地, COL4A3多肽、COL4A4多肽或COL4A5多肽是人的。人COL4A3的一个实例是具有UniProtKB登录号Q01955的COL4A3。人COL4A4的一个实例是具有UniProtKB登录号P53420的COL4A3。人COL4A5的一个实例是具有UniProtKB登录号P29400的COL4A5。

[0328] 合适地, COL4A3肽可以包含以下或由以下组成: 如SEQ ID NO:21所示的多肽序列, 或与SEQ ID NO:21至少70%相同的变体。合适地, 变体与SEQ ID NO:21可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0329] 合适地, COL4A4肽可以包含以下或由以下组成: 如SEQ ID NO:22所示的多肽序列, 或与SEQ ID NO:22至少70%相同的变体。合适地, 变体与SEQ ID NO:22可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0330] 合适地, COL4A5肽可以包含以下或由以下组成: 如SEQ ID NO:23所示的多肽序列, 或与SEQ ID NO:23至少70%相同的变体。合适地, 变体与SEQ ID NO:23可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0331] 示例性COL4A3氨基酸序列-Uniprot索引Q01955 (SEQ ID NO:21)

[0332] MSARTAPRPQVLLPLLLVLLAAAPAASKGCVCKDKGQCFCDGAKGEKGEKGFPGPPGSPGQKGFPGPE  
GLPGPQGPQKGFPLPGLTGSKGVRGISGLPGFSGSPGLPGTPGNTGPYGLVGVPGCSGSKGEQGFPLPGLTGYPGI  
PGAAGLKGQKGAPAKEEDIELDAKGDPLPGAPGPQGLPGPPGFPGVPGPPGPPGFFGFGAMGPRGPKGHMGERVI  
GHKGERGVKGLTGPPGPPGTIVIVTLTGPDNRDLKGEKGDKGAMGEPGPPGSPGLPGESYGSEKAPGDPGLQKPKG  
KDGVPGFPGSEGVKGNRGFPLMGEDGKQKQKGDIGPPGFRGPTTEYYDITYQEKDDEGTPGPPGPRGARGPQGPSGPP  
GVPGSPGSSRPGLRGAPGWPLKKGSKGERGRPGKDAMGTPGSPGCAGSPGLPGSPGPPGPPGDIVFRKGPPGDHGLP  
GYLGSPGIPGVDGPKGEPGLLCTQCPYIPGPPGLPGLPGLHGVKIPGRQGAAGLKGSPGSPGNTGLPGFPFGPAQ  
GDPGLKGEKGETLQPEGQVGVPGDPGLRGQPRKGLDIPGTPGVKGLPGPKGELALSSEKGDQPPGDPGSPGSPG  
PAGPAGPPGYGPQGEPLQGTQGVPGAPPPGEAGPRGELSVSTPVPGPPGPPGHPGPQPPGIPGSLGKCGDP  
GLPGPDGEPGIPGIGFPGPPGPKGDQGFPGTKGSLGCPGKMGEPGLPGKPLGAKGEPVAVMPGGPGTPGFPGERG  
NSGEHGEI GLPGLPGLPGTPGNEGLDGRGDPGQPGPPGEQPPGRCIEGPRGAQGLPGLNGLKQQRGRRGKTGPKG  
DPGIPGLDRSGFPGETGSPGIPGHQGEMLGQGRGYPGNPILGPPGEDGVI GMMGFPGAIGPPGPPGNPPTGQRRG  
SPGIPGVKQQRGTPGAKGEQDKGNPSEISHVIGDKGEPGLKGFAGNPGEKGNRVPGMPLKGLKGLPGPAGPP  
GPRGDLGSTGNPGEPLRGIPGSMGMMPGSKGKRGTLGFPGRAGRPGLPGIHGLQDKGEPGYSEGTRPGLPPT  
GDPGLPGDMGKKGEMQPGPPHGLGAPGEGAPGSPGSPGLPGKPGPHGDLGFKGKIGLLGPPGIRGPPGLPGFPGS  
PGPMGIRGDQGRDIPGPAGEKGETGLLRAPPGRGNPGAQAKGDRGAPGFPLPGRKGAMGDAGPRGPTGIEGFP  
GPPGLPGAII PGQTGNRGPSPGSRGSPGAPGPPGPPGSHVIGIKGDKGSMGHPGPKGPPGTAGDMGPPGRLGAPGTPG  
LPGPRGDPGFQGFPGVKGEKGNPGLGSI GPPGPIGPKGPPGVRGDPGTLKII SLPGSPGPPGTPGEPGMQGEPPG  
GPPGNLPGCPGRGKPKGDKGKGTGPGPAGEKGNKSGKGEPPGAGSDGLPGLKGRGDSGSPATWTTTRGFVTRHSQTT  
AIPSCPEGTVPLYSGFSFLFVQGNQRAHGQDLGTLGSLQRFITMPFLFCNVNDVCNFASRNDYSYWLSTPALMPMN  
MAPITGRALEPYISRCTVCEGPAIAIAVHSQTTDIPCPHGWSLWKGFSEIFMFTSAGSEGTQALASPGSCLEEFR  
ASPFLCHGRGTCNYYSNSYSFWLASLNPERMFRKPIPSTVKAGELEKII SRCQVCMKKRH

[0333] 示例性COL4A4氨基酸序列-Uniprot索引P53420 (SEQ ID NO:22)

[0334] MWSLHIVLMRCSFRLTKSLATGPWSLILILFSVQYVYVYSGKKYIGPCGGRDCSVCHCVPEKGSRGPPGP  
PGPQGPPIGLGAPGPIGLSGEKGMGRGDRPPGAAGDKGDKGPTGVPGFPLDIPGHPGPPGPRGKPGMSGHNSRG  
DPGFPGGRGALGPGGPLGHPGEKGEKGNVSVILGAVKGIQDRGDPGLPGLPGSWGAGGAGPTGYPGEPGLVGGP

QPGRPGLKGNPGVGVKGMGDPGEVQQGSPGPTLLVEPPDFCLYKGEKGIKIPGMVGLPGPPGRKGESGIGAKGE  
 KGIPGFPGRGDPGSYSGSPGFPGLKGEGLVGDPLGFLIGPKGDGPNRHPGPPGVLVTPPLPLKGPDPGFPGR  
 YGETGDVGGPPGLGRPEACAGMIGPPGPQGFPLPLGPEAGIPGRPDSAPGKPKGSPGLPGAPLQGLPG  
 SSVIYCSVGNPGPQGIKGVKVPGRGPKGEKNEGLCACEPGMGPMPGPPGLPGRQGSKGDGLPGWLGTKGDPGP  
 PGAEGPPGLPGKHGASPPGNKGAKGMVVSrvKHGKGERGPDGPPGFPQGSXHRDGHAGEKGDPPGDHEDAT  
 PGGKGFPGPLGPPGKAGVGPGLGFPGPPGERGHPGVPHGVRGPDGLKQKGDITISCNVTPGRHGGPPGFDGPP  
 GPKGFPGPQAGPLSGSDGHKGRPGTPTAEIPGPPGFRGDMGDPGFGGEGKSSPVGPPGPPGSPGVNGQKGIKGPDP  
 AFGHLGPPGKRGLSGVPGIKGPRGDPGCPGAEGPAGIPGFLGLKGPKGREGHAGFPVPGPPGHSCERGAPGIPGQP  
 GLPGYPGSPGAPGGKQPGDVGPPGAGMKGLPLGRPGAHPGLPGIPGPFDDGLGPPGPKGRPLPGFPGF  
 PGERGKPGAEGCPGAKGEPGEKMSGLPGDRGLRGAIGAIGPPGDEGEMAIISQKGTPEGPPGDDGFPGERGDKG  
 TPGMQRRGEPGRYGPFGFHRGEPGEKQPGPPGPPGPGSTGLRGI GFPLPGDQEGPSPGPPGFSGIDGARGP  
 KGNKGDPAHFPPGPKGEPGSPGCPGHFGASGEQGLPGIQGPRGSPGRGPPGSSGPPGCPGDHGMPLRGQPGE  
 GDPGRPLQGDPIGPPGIKGPSGSPGLNGLHGLKQKGTGASGLHDVGGPPGVPVIGPLKGERGDPGSPGISPPG  
 PRGKKGPPGPPGSSGPPGAGATGRAPKIDPDPGPPGDQPPGPDGPRGAPGPPGLPGSVDLLRGEPCDGLGPPG  
 PPGPPGPPGYKGFPGCDGKDGQKGPVGFPGPQGFHGFPPGPEKGLPMPGRKGTGLPGRGEPGPPADVDDCPRI  
 PGLPGAPGMRGPEGAMGLPGMRGSPGCKGEPGLDGRRGVDGVPSPGPPGRKGTGEDGYGGGPPGPIGDGPP  
 KGFPGYLGGFLLVLHSQTDQEPTCPLGMPRLWTGYSLLYLEGQEKAHNQDLGLAGSCLPVFSTLPFAYCNIHQVCH  
 YAQRNDRSYWLASAAPLPMPLSEEAIRPYVSRCAVCEAPAQAVAVHSQDQSI PPCPQWRSWIGYSFLMHTGAGD  
 QGGGQALMSPGSCLEDFRAAPFLECQRQGTCHFFANKYSFWLTTVKADLQFSSAPADTLKESQAQRQKISRCQVC  
 VKYS

[0335] 示例性COL4A5氨基酸序列-Uniprot索引P29400 (SEQ ID NO:23)

[0336] MKLRGVSLAAGLFLALSLWGQPAEAAACYGCSPGSKDCSGIKGEKGERGFPLEGHPGLPGFPGPEG  
 PPGPRGQKGGDDGIPGPPGPKGIRGPPGLPGFPGTGLPGMPGHDGAPGQGIKGCNGTKGERGFPGSPGFPGLQGP  
 PPGIPGMKGEPSIIMSSLPGKGNPGYPGPPGIQGLPGPTGIPGIPGPPGPPGLMPPGPPGLPGPKGNMGLNFQ  
 GPKGEKGEQGLQGGPPGPPGQISEQKRPIDVEFQKGDQGLPGDRGPPGPPGIRGPPGPPGGEKGEKGEQGEKGRKGP  
 GKDGENGQPIPLPGDPGYGEPGRDGEKQKGDITGPPGPPGLVIPRPGTGITIGEKNIPLPLGPEKGERGFPG  
 IQGPPGLPGPPGAAMGPPGPPGFPGERGQKGDGPPGISIPGPPGLDQPGAPGLPGPPGAGPHIPPSDEICEPG  
 PPGPPGSPGDKGLQGEQGVKGDKGDTCFNCIGTGISGPPGQPLPLGPPGSLGFPGQKGEKQAGATGPKGLPGI  
 PGAPGAPGFPGSKGEPGDILTFPGMKGDKGELGSPGAPGLPLGTPGQDGLPLPGPKGEPGGITFKGERGPPGNP  
 GLPLPGNIGMPGPPGFPVGEKGIQGVAGNPGQPIGPKGDPGQTITQPKPLPGNPRGDVGLPGDPGL  
 PGQPGLPGIPGSKGEPGIPGIGLPGPPGPKGFPGIPGPPGAPGTGRIGLEGPPGPPGFPGPKGEPGFALGPPGPP  
 GLPGFKGALGPKGDRGFPPGPPGRTGLDGLPGPKGDVGNQPGMPGPPGLPGIGVQGGPPGPPGIPGIPGQPLH  
 GIPGEKGDGPPGLDVPGPPGERGSPGIPGAPGIPGPPGSPGLPGKAGASGFPGTKEMGMMGPPGPPGLGIPGRS  
 GVPGLKGGDGLQGPGLPGPTGEKGSKGEPLGPPGMPDNLGSKGEKGEPLGIPGVSGPKGYQGLPGDPGQP  
 GLSQPGLGPPGPKGNPLPGQPGLIGPPGLKGTIGDMGFPGPQVVEGPPGSPGVGQPGSPGLPGQKGDKDPGI  
 SSIPLPLGPKGEPGLPGYPGNPIKGSVGDPLPLGTPGAKGQPLPGFPGTTPGPPGPKGISGPPGNPLGPE  
 PGPVGGGGHPGQPPGGEKGPQDGI PGPAQKGEPPGQPGFNPMPGLPLSGQKGDGGLPGIPGNPLGPKGE  
 PGFHGFPVQGGPPGSPGPALEGPKGNPGQPPGRPLPGPEGPPGLPGNGGKGEKGNPGQPLPLPLGKGD  
 QGPPGLQGNPRPGLNGMKGDPGLPGVPGFPGMKGPSVPGSAGPEGEPGLIGPPGPPGLPGSPGQSI I IKGDAGPP

GIPGQPGLKGLPGPQGPQGLPGPTGPPGDPGRNGLPGFDGAGGRKGDPLPGQPGRGLDGGPPGDGLQGPPGPPGT  
SSVAHGFLITRHSQTTDAPQCPQGLTLQVYEGFSLLYVQGNKRAHQDLGTAGSCLRRFSTMPFMFCNINNVCFASR  
NDYSYWLSTPEPMPMSMQPLKQSIQPFISRCVCEAPAVVIAVHSQTIQIPHCPQGWDSLWIGYSFMMHTSAGAEG  
SGQALASPGSCLEEFRSAPFIECHGRGTCNYYANSYSFWLATVDVSDMFSKPQSETLKAGDLRTRISRCQVCMKRT

[0337] 蛋白质编码序列可以包含COL4A3转基因、COL4A4转基因或COL4A5转基因或由COL4A3转基因、COL4A4转基因或COL4A5转基因组成。

[0338] NM\_000091.5中提供了编码COL4A3的实例转基因。NM\_000092.5中提供了编码COL4A4的实例转基因。NM\_000495.5中提供了编码COL4A5的实例转基因。

[0339] 合适地，COL4A3转基因可以包含以下或由以下组成：如SEQ ID NO:24所示的多核苷酸序列，或与SEQ ID NO:24至少70%相同的变体。合适地，变体与SEQ ID NO:24可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0340] 合适地，COL4A4转基因可以包含以下或由以下组成：如SEQ ID NO:25所示的多核苷酸序列，或与SEQ ID NO:25至少70%相同的变体。合适地，变体与SEQ ID NO:25可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0341] 合适地，COL4A5转基因可以包含以下或由以下组成：如SEQ ID NO:26所示的多核苷酸序列，或与SEQ ID NO:26至少70%相同的变体。合适地，变体与SEQ ID NO:26可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0342] 示例性COL4A3转基因序列 (SEQ ID NO:24)

[0343] atgagcgcggaccgccccaggccgaggtgctcctgctgccgctcctgctggtgctcctggcggc  
ggcggccgcagccagcaagggttgtgtctgtaaagacaaaggccagtgcttctgtgacggggccaaaggggagaag  
ggggagaagggtttcctggacccccggttctcctggccagaaaggattcacaggtcctgaaggcttgccctggac  
cgcagggacccaagggtttccaggacttcaggactcacgggttccaaagggtgaaggggaataagtggattgcc  
aggattttctggttctcctggacttcaggcaccaggcaataaccggccttacggacttgtcggtgtaccagga  
tgcagtggttctaagggtgagcagggtttccaggactccagggacactgggctaccagggatcccggtgctg  
ctggtttgaaaggacaaaagggtgctcctgctaaagaagaagatatagaacttgatgcaaaaggcgacccccgggtt  
gccaggggctccaggacccccagggtttgccaggccctcagggtttcctgggctgttgggccacctggtcctccg  
ggattccttggtttccaggagcatgggacctagaggacctaaagggtcacatgggtgaaagagtgataggacata  
aaggagagcggggtgtgaaagggttaacaggacccccgggaccaccaggaacagttattgtgaccctaactggccc  
agataacagaacggacctcaagggggaaaaggagacaaggagcaatgggcgagcctggacctcctggacctca  
ggactgcctggagaatcatatggatctgaaaagggtgctcctggagacctggcctgcagggaaaaccggaaaag  
atggtgttctggttccctggaagtgaggagtcaggggcaacagggtttccctgggttaatgggtgaagatgg  
cattaagggacagaaaggggacattggccctccaggatttctggttccaacagaatattatgacacataccaggaa  
aaggagatgaaggcactccaggcccaggggccagaggagctcgtggccacaaggtcccagtggtccccccg  
gagttcctggaagtcctggatcatcaaggcctggcctcagaggagccccctggatggccaggcctgaaaggaagtaa  
aggggaacgaggccgcccaggaaaggatgcatggggactcctgggtccccagggtgtgctggttcaccaggtctt  
ccaggatcaccgggacctccaggaccgaggtgacatcgtttttcgaagggtccacctggagatcacggactgc

caggctatctagggctctccaggaatcccaggagttgatgggcccaggagaaccaggcctcctgtgtacacagtg  
cccttatatcccagggcctcccggctctcccaggattgccagggttacatgggtgtaaaaggaatcccaggaagacaa  
ggcgcagctggcttgaaaggaagcccagggtcccaggaatacaggctctccaggatttccaggtttcccagggtg  
cccagggtgaccaggacttaagggagaaaaaggtgaaacacttcagcctgaggggcaagtgggtgtcccagggtga  
cccggggctcagaggccaacctgggagaaagggttgatggaattcctggaactccgggagtgaaggattacca  
ggacctaaaggcgaactggctctgagtggtgagaaaggggaccaaggtcctccaggggatcctggctcccctgggt  
ccccaggacctgcaggaccagctggaccacctggctacggaccccaaggagaacctggctcctcaggggcacgcaagg  
agttcctggagccccggaccaccggagaagccggccctaggggagagctcagtgtttcaacaccagttccaggc  
ccaccaggacctccaggccccctggccatcctggcccccaagggtccacctggatcctcctggatcctggggaaat  
gtggagatcctggcttccagggtgatgggtgaaccaggaattccaggaattggatttctgggctcctggacc  
taagggagaccaaggttttccagggtacaaaaggatcactgggttgctcctggaaaaatgggagagcctgggttacct  
ggaaagccaggcctcccaggagccaagggagaaccagcagtagccatgcctggaggaccaggaacaccagggttttc  
caggagaaagaggcaattctggggaacatggagaaattggactcctggacttccaggtctcctggaactccagg  
aaatgaagggttgatggaccacgaggagatccagggcagcctggaccacctggagaacaaggacccccaggaagg  
tgcatagagggtcccaggggagcccaaggacttccaggcttaaatggattgaaagggaacaaggcagaagaggtg  
aaacggggccaaagggagaccaggaattccaggcttgatagatcaggatttctggagaaactggatcaccagg  
aattccagggtcatcaaggtgaaatgggaccactgggtcaaagaggatatccaggaaatccgggaattttaggggca  
ccagggtgaagatggagtgatgggatgatgggcttctcctggagccattggccctccaggggccccctgggaaccag  
gcacaccagggcagagggggagccctggaattccaggagtaaaaggccagagaggaaccccaggagccaaggggga  
acaaggagataaaaggaaatcccgggccttcagagatatcccacgtaataggggacaaaggagaaccagggtctcaa  
ggattcgcaggaaatccagggtgagaaaggaacagaggcgttccagggatgccaggtttaaaggcctcaaaggac  
taccggaccagcaggaccaccaggccccagaggagatttgggcagcactgggaatcctggagaaccaggactgcg  
tggtataccaggaagcatggggaacatgggcatgccaggttctaaaggaaaaagggaactttgggatccagggt  
cgagcaggaagaccaggcctcccaggtattcatggtctccaggagataagggagagccaggttattcagaaggtg  
caaggccaggaccaccgggaccaacgggggatccaggactgccgggtgatatgggaaagaaaggagaaatggggca  
acctggcccacctggacatttggggcctgctggacctgagggagccccctggaagtctggaagtctggcctccca  
ggaaagccagggtcctcatggtgatttgggttttaaggaatcaaaggcctcctgggccccctcaggaaatcagaggcc  
ctccagggtcttccaggatttccaggatctcctggaccaatgggtataagaggtagccaaggacgtgatggaattcc  
tggtccagccggagaaaaagggagaaacgggtttattgagggccccctccaggcccccaagggaacctgggtgctcaa  
ggagccaaaggagacaggggagccccaggttttctggcctcccgggcagaaaaggggccatgggagatgctggac  
ctcaggaccacagcatagaaggattcccaggggccaccagggtctgcccggtgcaattatcctggccagacagg  
aaatcgtgggtccaccaggctcaagaggaagcccagggtgcgctgggtccccctggacctccaggagatcatgtaata  
ggcataaaaggagacaaaagggtctatgggccaccctggcccaaaagggtccacctggaactgcaggagacatgggac  
caccaggctgcttgggagcaccaggtactccagggtcttccaggaccagaggtgatcctggattccagggggtttcc  
aggcgtgaaaggagaaaaagggtaatcctggatttctaggatccattggacctccaggaccaattgggccccaaagga  
ccacctgggtgtacgtggagacctggcacacttaagattatctccttccaggaaagcccaggggccacctggcacac  
ctggagaaccagggtatgcagggagaacctgggcccaccaggggccacctggaaacctaggacctgtgggccaagagg  
taagccaggcaaggatggaaaaccaggaactcctggaccagctggagaaaaaggcaacaaaggttctaaaggagag  
ccaggaccagctggatcagatggattgccaggtttgaaaggaaaacgtggagacagtgatcacctgcaacctgga

caacgagaggctttgtcttcacccgacacagtcaaacacagcaattccttcatgtccagaggggacagtgccact  
ctacagtgggttttcttttctttttgtacaaggaaatcaacgagcccacggacaagaccttggaaactcttggcagc  
tgctgcagcgatttaccacaatgccattcttattctgcaatgtcaatgatgtatgtaattttgcatctcgaaatg  
attattcatactggctgtcaacaccagctctgatgccaatgaacatggctcccattactggcagagcccttgagcc  
ttatataagcagatgcaactgtttgtgaaggctctgcgatcgccatagccggttcacagccaaaccactgacattcct  
ccatgtcctcacggctggatttctctctggaaggattttcattcatcatgttcacaagtgcaggttctgagggca  
ccgggcaagcactggcctcccctggctcctgcctggaagaattccgagccagcccatttctagaatgtcatggaag  
aggaacgtgcaactactattcaaattcctacagtttctggctggcttcattaaaccagaaaagaatgttcagaaaag  
cctattccatcaactgtgaaagctggggaattagaaaaataataagtcgctgtcaggtgtcatgaagaaaagac  
actga

[0344] 示例性COL4A4转基因序列 (SEQ ID NO:25)

[0345] atgtggtctctgcacatagtaactatgaggtgctccttcagattgaccaagtccttggccacaggtcc  
ctggctcacttatactcattctcttttctgtacaatatgtatatgggagtggaagaaatacattggctccttgtgga  
ggaagagattgctctgtttgccactgtgttctgaaaaggggtctcgggggtccaccaggaccaccaggggccacagg  
gtccaattggaccctgggagccccaggaccattgggctttcaggagagaaaggaatgagaggggaccgcggccc  
tcttgagcagcaggggacaaaggagataagggtccaactgggtgttctggatttccaggttagatggcatacct  
gggccccagggcctcctggaccagaggcaaacctgggtatgagtgggccacaatggctcaagaggtgaccagggt  
ttccaggaggaagaggagctcttggcccaggaggccccctaggccatcctggggaaaaggagaaaaggaaattc  
agtgttcatttttaggtgccgttaaaggtattcaggagacagaggggaccaggactgcctggcttaccaggatct  
tggggtgcaggaggaccggcaggtcccacaggatatcctggagagccagggttagtgggacctccgggccaaccag  
ggcgtccaggtttgaagggaaatcccgggtgtgggagtaaagggggcaaatgggagaccgggtgaggttggtcagca  
aggttctcctggaccaccctgttggtagagccacctgacttttgtctctataaaggagaaaagggtataaaagga  
attcctggaatggttgactgccaggaccaccaggacgcaagggagaatctggtattggggcaaaaggagaaaaag  
gtattcctggatttccaggcctcgggggatcctggttcctatggatctccaggttttccaggattaaagggaga  
actaggactggttgagatcctgggctatttggattaattggccaaagggggatcctggaaatcagggcaccaca  
ggaccaccaggtgttttgggtgactccacctcttccactcaaaggcccaccaggggaccagggttccctggccgct  
atggagaaacaggggatgttgaccacctgggtccccaggtctcttgggcagaccagggggaagcctgtgcaggcat  
gataggaccctggccacaaggatttctgggtcttctgggcttccaggagaagctggtattcctgggagacct  
gattctgctccaggaaaaccagggaagccaggatcacctggcttgcctggagcaccaggcctgcagggcctcccag  
gatcaagtgtgatatactgtagtgttgggaacccggaccacaaggaataaaaggcaaatgttggtccccaggagg  
aagaggcccaaaaggagaaaaaggaaatgaaggactctgtgcctgtgagcctggaccatgggccccctggccct  
ccaggacttctgggagcaggggagtaaggagacttggggctccctggctggcttggaaacaaaagggtgaccag  
gacctcctgggtgctgaaggacctccagggtaccaggaaagcatggtgcctctggaccacctggcaacaaaggggc  
gaagggtgacatggttgatcaagagttaaagggcacaaaggagaaagaggtcctgatgggccccaggatttcca  
gggcagccaggatcacatggtcgggatggacatgctggagaaaaaggggatccaggacctccagggatcatgaag  
atgacacccaggtggttaaaggatttctggacctctgggccccccaggcaaaagcaggacctgtggggccccccagg  
actgggatttctgggtccaccaggagagcaggccaccaggagttccaggccaccagggtgtgaggggccctgat  
ggcttgaaggtcagaaaggtgacacaatttcttgaacgtaacctaccctgggagcatggcctccaggttttg  
atggacctccaggtccgaaggatttccaggtcccaaggtgcccctgggctgagtggttcagatgggcataaagg

cagacctggcacaccaggaacagcggaaataccaggtccacctggttttcgtggtgacatgggagatccgggtttt  
ggaggtgaaaaggggtcctcccctggtgggccccaggccctcccggctcaccaggagtgaatggtcagaaaggaa  
tcccgggagacctgcatttggtcacctgggacccccgggaaagaggggtctttcaggagtccaggataaaagg  
accagaggtgatccgggatgtccaggggtgaagggccagctggcattcctggattcctaggtctcaaaggtccc  
aaaggcagagagggacatgctgggtttccaggtgtccaggtccacctggccattcctgtgaaagaggtgctccag  
ggataccagggcaaccgggactccctgggtatccaggtagcccaggtgctccaggtgggaaaggacagccgggaga  
tgtggggcctcccggccagctggaatgaaaggcctcccggactcccaggacggcctggggcacatggtccccc  
ggcctcccaggaatcccaggtccctttggagatgatgggctacctggctccaggtccaaaggacccccgggggc  
tgcctgggtttccaggtttcccggagaaagaggaaagcctgggtgcagagggatgtcctggcgaaagggagaacc  
tggagagaagggcatgtctggccttctggagaccggggactgagaggggccaaggagccataggacctcccgga  
gatgaaggagaaatggctatcatttcacaaaagggaacacctggggaacctggacctcctggagatgatggattcc  
caggagaaagaggtgataaaggaactcccgggatgcaagggagaagaggagagccgggaagatacggaccacctgg  
atctcacagaggggaacctgggtgagaaaggtcagccagggcctcctggacccccaggccctcaggctcaactggt  
ctaagagggttcattggttttccaggacttccaggtgaccagggtagccaggttctccaggtccccctggatttt  
caggaattgatggagcaagaggacctaaaggaaacaaaggtgacctgccagtcactttggtccacctggtccaaa  
gggtgagccaggtagccctggatgtccagggcattttggagcatccggagagcagggettgcctggtattcaagg  
cccagaggatcacctggaaggccagggccacctggctcctctggaccaccagggtgcccaggtgatcacgggatgc  
ctgggctgaggggacagccaggagaaatgggagacctgggccaagaggcctccaggggatccaggataccagg  
tctccgggaataaaaggtccctccggatcacctggcctgaacggcttgcatggattgaaaggtcagaaaggaact  
aaaggtgcttcaggtttgcatgatgtggggccacctggtccagtgggaataacctgggctaaaaggggagagaggag  
acctgggagcccaggaatctctcctcaggtcctcgtggaagaaaggtccccaggacccccagggagttcagg  
accacctggtcctgcaggtgccacaggaagagctcctaaggacattcctgacccgggtccacctggagatcaggga  
cctcctggctcctgatggccaagaggagcacctgggctccaggcctccctgggagtgttgacctctgagagggg  
agccaggtgactgtggtctaccagggccaccaggtccccctggcccaccaggccctccaggatacaaaggctttcc  
aggatgtgatggaaaagatggccagaaaggaccagtgggatcccccgggaccgcagggaccacatggatttctggg  
ccacctggagagaagggtttacctggacctccaggagaaaaagggcccactggcttccgggtcccagaggtgaac  
cggggccacctgcagatgtggatgactgtccccgaatcccaggccttctggggcggcaggcatgagaggaccaga  
aggagccatggggctccctggaatgagaggccctcaggaccagggtgcaaaggagagcctgggctggatggcagg  
aggggtgtggatggcgtccctgggtctcctgggctcccggacgtaaaggtgacacaggagaagacggctaccctg  
gaggaccagggcctcctggtccattggggatcctgggccccaaagggtttggcctggatacctcgggtggttctc  
cctggttctccacagtgcagcaggaccaggagcccacctgccccctgggcatgccaggctctggactgggtatagt  
ctgttataacctggaagggaagagaaagctcacaatcaagaccttggcttgccagggcttctgcttcccgtattta  
gcacgctgccctttgcctactgcaacatccaccaggtgtgccactatgccagagaaacgacagatcctactggct  
ggccagcgtgcgccccccccatgatgccactctctgaagaggcgatccgccccatgtcagccgctgtgcggt  
tgcgaggccccggcccaggcgggtggcgggtgcacagccaggaccagtcacccccatgtccgcagacctggagga  
gcctctggatcgggtattcattcctgatgcacacaggagctggggaccaaggaggaggcagggccttatgtcacc  
tggcagctgctggaagatttcagagcagcaccattccttgaatgccagggccggcagggaaacttgccacttttcc  
gcaataagtatagcttctggctcacaacgggtgaaagcagacttgacgttttctctgctccagcaccagacacct  
taaaagaaagccaggcccaacgccagaaaaatcagccgggtgccaggtctgcgtgaagtatagctag

[0346] 示例性COL4A5转基因序列(SEQ ID NO:26)

[0347] atgaaactgcgtggagtcagcctggctgccggcttgcttcttactggccctgagcttttgggggcagcc  
tgcagaggctgcggcttgctatgggtgttctccaggatcaaagtgtgactgcagtgccataaaaggggaaaagga  
gagagagggtttccagggtttggaaggacaccaggattgcctggatttccagggtccagaaggcctccggggcctc  
ggggacaaaagggtgatgatggaattccagggccaccaggacaaaaggaatcagaggtcctcctggacttcttg  
atctccaggacaccagggtcttctggaatgccaggccacgatggggccccaggacctcaaggtattcccggatgc  
aatggaaccaagggagaacgtggatttccaggcagtcctcggttttctgggttacagggtcctccaggacccccctg  
ggatcccagggtatgaagggtgaaccaggtagtataattatgtcatcactgccaggaccaaagggtaatccaggata  
tccagggtcctcctggaatacaaggcctacctgggtccactgggtataaccagggccaatgggtccccaggaccacca  
ggtttgatggggcctcctgggtccaccaggacttccaggacctaaagggaatatgggcttaaatttccagggacca  
aagggtgaaaaagggtgagcaagggtcttccaggccccacctggggccacctgggcagatcagtgaaacagaaaagaccaat  
tgatgtagagtttcagaaaggagatcagggacttcttggtgaccgaggccctcctggacctccagggatacgtggt  
cctccagggtccccagggtggtgagaaagggtgagaagggtgagcaaggagagccaggcaaaagaggtaaaccaggca  
aagatggagaaaaatggccaaccaggaattcctgggtttgcttggtgatcctgggtaccctgggtgaaccgggaaggga  
tggtgaaaagggccaaaaagggtgacactggcccacctggacctcctggacttgaattcctagacctgggactggt  
ataactataggagaaaaaggaaacattgggttgcttggttgcttgagaaaaaggagagcaggatttcttgga  
tacagggtccacctggccttctggacctccaggggctgcagttatgggtcctcctggcctcctggatttcttg  
agaaagggtcagaaagggtgatgaaggaccacctggaatttccattcctggacctcctggacttgacggacagcct  
ggggctcctgggcttccagggtcctggcctgctggcctcacattcctcctagtgatgagatatgtgaaccag  
gccctccaggccccccaggatcctcagggtgataaaggactccaaggagaacaaggagtgaagggtgacaaagggtga  
cacttgcttcaactgcattggaactggtatttccaggcctccagggtcaacctgggttgccagggttcccagggtcct  
ccaggatccttgggttccctggacagaaaggggaaaaaggacaagctgggtgcaactgggtcccaaaggattaccag  
gcattccaggagctccagggtgctccagggttctcctggatcctaaagggtgaacctgggtgatacctcacttttccagg  
aatgaagggtgacaaaggagagttgggttcccctggagctccagggttctcctgggttacctggcactcctggacag  
gatggattgccagggttctcctggcccgaaggagagcctgggtggaattacttttaagggtgaaagaggtccccctg  
ggaaccagggttaccaggcctcccagggaatataggcctatgggtccccctgggttccggcctccaggcccagt  
agggtgaaaaaggcatacaagggtgtggcaggaaatccaggccagccaggaataccagggtcctaaaggggatccagggt  
cagactataaccagccggggaagcctggcttgcttggttaaccaggcagagatggtgatgtaggtcttccagggtg  
acctggacttccagggaaccagggttccagggtatcctggtagcaaaaggagaaccagggtatcctggaattgg  
gcttctcctggaccacctgggtcccaaaggcttctcctggaattcaggacctccaggagacctgggacacctggaaga  
attggtctagaaggcctcctggggcaccggcttccaggaccaaagggtgaaccaggatttgcattacctgggc  
cacctggggcaccaggacttccagggttcaaggagcacttgggtccaaaagggtgatcgtggttcccaggacctcc  
gggtcctccaggacgcactggcttagatgggtcctcctggaccaaaagggtgatgttgaccaaatggacaacctgga  
ccaatgggacctcctgggtgccaggaatagggttccagggaccaccaggaccaccagggttctcctgggccaatag  
gtcaacctgggttacatggaataccaggagagaagggggatccaggacctcctggacttgatgttccaggaccccc  
agggtgaaagaggcagtcagggtccccggagcactgggtcctataggacctccaggatccaccagggttccaggga  
aaagcagggtgctcctggatttccagggtaccaagggtgaaatgggtatgatgggacctccaggcccaccaggacctt  
tgggaattcctggcaggagtggtgtacctgggtcttaagggtgatgatggccttcagggtcagccaggacttcttg  
ccctacaggagaaaaaggtagtaaaggagagcctggccttccaggccccctggaccaatggatccaaatcttctg

ggctcaaaaggagagaagggggaacctggcttaccaggtataacctggagtttcagggccaaaaggttatcagggtt  
tgcttgagaccagggaacctggactgagtgacaacctggattaccaggaccaccagggtcccaaaggtaaccc  
tggtctccctggacagccaggcttataggacctcctggacttaaggaacctcggatgatgggttttcaggg  
cctcagggtgtggaaggcctcctggaccttctggagttcctggacaacctggctcccaggattacctggacaga  
aaggcgacaaggatgatcctggtatttcaagcattggtcttccagggtctcctggtccaaagggtgagcctggtct  
gcctggataccagggaacctggtatcaaaggttctgtgggagatcctggtttgccggattaccaggaacctcct  
ggagcaaaaggacaaccaggccttctggattcccaggaacctcaggccctcctggaccaaaaggatttagtggcc  
ctcctgggaaccccgccctccaggagaacctggtcctgttaggtggtggaggatcctctgggcaaccagggcctcc  
aggcgaaaaaggcaaacccggatcaagatggtattcctggaccagctggacagaagggtgaaccagggtcaaccaggc  
tttggaacccaggaccctggacttccaggactttctggccaaaagggtgatggaggattacctgggattccag  
gaaatcctggccttccagggtccaaaggcgcaaccaggctttcacgggttccctggtgtgcagggtccccaggccc  
tcttggttctccgggtccagctctggaaggacctaaaggcaacctgggcccccaagggtcctcctgggagaccagggt  
ctaccagggtccagaaggctcctccagggtctcctggaaatggaggattataaaggagagaagggaatccaggccaac  
ctgggctacctggcttgctggtttgaaaggagatcaaggaccaccaggactccagggtaatcctggccggccggg  
tctcaatggaatgaaaggagatcctggtctcctggtgttccaggattccaggcatgaaaggaccagtgaggta  
cctggatcagctggccctgagggggaaccgggacttattggtcctccagggtcctcctggattacctggtccttcag  
gacagagtatacataataaaggagatgctggtcctccagggaatcctggccagcctgggctaaagggtctaccagg  
acccaaggacctcaaggcttaccagggtccaactggccctccaggagatcctggacgcaatggactccttggttt  
gatggtgcaggagggcgcaaggagaccagggtctgccaggacagccagggtaccgtggtttggatggtccccctg  
gtccagatggattgcaagggtccccagggtccccctggaacctcctctgttgacatggatttcttattacagcca  
cagccagacaacggatgcaccacaatgcccacagggaacacttcagggtctatgaaggcttttctctcctgtatgta  
caaggaaataaaagagcccacggatcaagacttggggacggctggcagctgccttcgtcgttttagtacctgcctt  
tcatgttctgcaacatcaataatgtttgcaactttgcttcaagaaatgactattcttactggtctctaccaccaga  
gcccattgccaatgagcatgcaaccctaaaggccagagcatccagccattcattagtcgatgtgcagtatgtgaa  
gctccagctgtggtgatcgagttcacagtcagacgatccagattccccattgtcctcagggtggtattctctgt  
ggattggttattccttcatgatgcatacaagtgcaggggcagaaggctcagggtcaagccctagcctcccctggttc  
ctgcttggaaagagtttctgctcagctccttcatcgaatgtcatgggaggggtacctgtaactactatgccaactcc  
tacagcttttggtggcaactgtagatgtgtcagacatgttcagtaaacctcagtcagaaacgctgaaagcaggag  
acttgaggacacgaattagccgatgtcaagtgtgcatgaagaggacataa

[0348] COL4A3转基因、COL4A4转基因或COL4A5转基因可包含内含子或内含子序列,其可以用于改善基因表达。COL4A3转基因、COL4A4转基因或COL4A5转基因可以包含蛋白质标签,诸如血凝素(HA)标签。HA可以用作表位标签,并且已被证明不会干扰其添加至的蛋白质的生物活性或生物分布。蛋白质标签可以促进转基因的检测、分离和纯化。其他合适的蛋白质标签可以包括Myc标签、多组氨酸标签和flag标签。

[0349] 在一些实施方案中,包含以下或由以下组成的启动子可操作地连接至编码COL4A3多肽、COL4A4多肽或COL4A5多肽或其片段或衍生物的蛋白质编码序列:与SEQ ID NO:47具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0350] 在一些实施方案中,由SEQ ID NO:47的核苷酸序列组成的启动子可操作地连接至

编码COL4A3多肽、COL4A4多肽或COL4A5多肽或其片段或衍生物的蛋白质编码序列。

[0351] 肾病综合征(NS)相关联的转基因

[0352] 蛋白质编码序列可以包含NS相关联的转基因或由NS相关联的转基因组成。

[0353] 肾病综合征(NS)是一种以显著的蛋白尿、低白蛋白血症、水肿和高脂血症为特征的慢性肾脏疾病。NS相关联的转基因可以是与单基因形式的NS相关联并在足细胞中表达的基因。

[0354] 合适的NS相关联的转基因包括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、KANK4、MAGI2、MYO1E、OCRL、PTPRO、SMARCAL1、SYNPO、TBC1D8B、XPO5、TNS2和NLRP3。

[0355] 在一些实施方案中,包含以下或由以下组成的启动子可操作地连接至NS相关联的转基因:与SEQ ID NO:47具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0356] 在一些实施方案中,由SEQ ID NO:47的核苷酸序列组成的启动子可操作地连接至NS相关联的转基因。

[0357] NPHS2

[0358] 蛋白质编码序列可以编码NPHS2,或其片段和/或变体。

[0359] “NPHS2”是由NPHS2基因编码的多肽的简称,并且也称为足细胞素。NPHS2是42kDa发夹样膜相关联的足细胞特异性蛋白,其是裂孔隔膜(相邻足细胞足突之间的细胞与细胞连接)处的蛋白复合物的关键组分。其定位于脂筏并与其他重要的裂孔隔膜蛋白(如肾病蛋白、CD2AP和TRPC6)相互作用。其在维持裂孔隔膜以及肾小球滤过屏障的完整性中非常重要。

[0360] NPHS2的片段和/或变体可以保留NPHS2活性。例如,足细胞素的片段和/或变体可以调控肾小球通透性。合适地,NPHS2的片段和/或变体可以具有与NPHS2相同或相似的活性,例如可以具有至少40%、至少50%、至少60%、至少70%、至少80%、至少90%、至少95%或至少100%的NPHS2的活性。

[0361] 基于NPHS2的已知结构和功能特征(参见例如Tabassum, A.等人, 2014. *Interdisciplinary Sciences: Computational Life Sciences*, 6(1), pp.32-39), 和/或基于已知的变体(参见例如NCBI Gene ID:7827和NCBI HomoloGene:22826),本领域技术人员将能够使用保守的取代来产生片段和/或变体。合适地,NPHS2的片段和/或变体包含跨膜结构域,与N末端和C末端处的两个胞质结构域。

[0362] NPHS2基因在黑猩猩、恒河猴、狗、牛、小鼠和大鼠中保守。NPHS2可以是人NPHS2。合适地,NPHS2可以包含UniProtKB登录号Q9NP85的多肽序列或其片段和/或变体或由UniProtKB登录号Q9NP85的多肽序列或其片段和/或变体组成。

[0363] 在一些实施方案中,NPHS2包含与SEQ ID NO:55至少70%相同的氨基酸序列或其片段或由与SEQ ID NO:55至少70%相同的氨基酸序列或其片段组成。合适地,NPHS2包含与SEQ ID NO:55至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至

少97%、至少98%或至少99%相同的氨基酸序列或其片段,或由与SEQ ID NO:55至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的氨基酸序列或其片段组成。

[0364] 在一些实施方案中,HPHS2包含SEQ ID NO:55或其片段或由SEQ ID NO:55或其片段组成。

[0365] MERRARSSSRESRGRGGRTPHKENKRAKAERSGGGRGRQEAGPEPSGSGRAGTPGEPRAAATVVDVDE  
VRGSGEETEVVALLESERPEEGTKSSGLGACEWLLVLISLLFIIMTFPFSIWFCVKVQVEYERVIIFRLGHLLPGR  
AKGPGLFFFLPCLDTYHKVDLRLQTLIPFHEIVTKDMFIMEIDAICYRMENASLLSSLAHVSKAVQFLVQTTMK  
RLLAHRSLTEILLERKSIAQDAKVALDSVTCIWGIKVERIEIKDVRLPAGLQHS�AVEAEAQRQAKVRMIAAEAEKA  
ASESLRMAAEILSGTPAAVQLRYLHTLQSLSTKSTVVLPLPFDLLNCLSSPSNRTQGSLPFPSPSKPVEPLNPKK  
KDSPML

[0366] 示例性NPHS2氨基酸序列(SEQ ID NO:55)

[0367] 在一些实施方案中,HPHS2转基因包含与SEQ ID NO:56至少70%相同的核苷酸序列或其片段或由与SEQ ID NO:56至少70%相同的核苷酸序列或其片段组成。合适地,NPHS2转基因包含与SEQ ID NO:56至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的核苷酸序列或其片段,或由与SEQ ID NO:56至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的核苷酸序列或其片段组成。

[0368] 在一些实施方案中,HPHS2转基因包含SEQ ID NO:56的核苷酸或其片段或由SEQ ID NO:56的核苷酸或其片段组成。

[0369] ATGGAGAGGAGGGCGCGGAGCTCCTCCAGGGAGTCCCGCGGGCGAGGCGGCAGGACTCCGCACAAGGAG  
AACAAGAGGGCAAAGGCCGAGAGGAGCGGCGGGGGCCGCGGGCGCCAGGAGGCTGGGCCCGAGCCGTCGGGCTCCGG  
ACGGGCGGGACCCCGGGGAGCCCCGAGCGCCCGCCACGGTGGTGGACGTGGATGAGGTCCGAGGCTCCGGCG  
AGGAGGGCACCGAGGTGGTGGCGCTGTTGGAGAGCGAGCGGCCCGAGGAAGGTACCAAATCCTCCGGCTTAGGGGCC  
TGTGAGTGGCTTCTTGTCTCATTTCCTGCTCTTCATCATCATGACCTTCCCTTTTCCATCTGGTTCTGCGTAAA  
GGTTGTACAAGAGTATGAAAGAGTAATTATATTCGACTGGGACATCTGCTTCCCTGGAAGAGCCAAAGGCCCTGGTC  
TTTTCTTTTTTTTGCCTGCCTGGATACCTACCACAAGGTTGACCTTCGTCTCCAACTCTGGAGATACCTTTTCAT  
GAGATCGTGACCAAAGACATGTTTATAATGGAGATAGATGCCATTTGCTACTACCGAATGGAAAATGCCTCTCTTCT  
CCTAAGCAGTCTTGTCTATGTATCTAAAGCTGTGCAATTCCTTGTGCAAACCACTATGAAGCGTCTCCTAGCACATC  
GATCCCTCACTGAAATTCTTCTAGAGAGGAAGACATCGCCCAAGATGCAAAGGTTGCCTTGATTGAGTACCTGT  
ATTTGGGGAATCAAAGTGGAGAGAATAGAAATTAAGATGTGAGGTTGCCAGCTGGGCTTCAGCACTCACTGGCTGT  
GGAGGCTGAAGCGCAAAGACAAGCCAAAGTGCAGGATGATTGCTGCAGAAGCGGAAAAGGCTGCTTCTGAGTCCCTGA  
GGATGGCAGCTGAGATTCTGTCAGGCACCCCTGCTGCTGTTGAGCTTCGATACCTCCACACCCTTCAGTCTCTGTCC  
ACAGAGAAGCCTTCCACTGTGGTTTTACCTTTGCCATTTGACCTACTGAATTGCCTGTCTTCTCCAGCAACAGAAC  
TCAGGGAAGCCTCCCCTTCCCAAGTCTTCCAAACCTGTTGAGCCACTAAATCCTAAAAAGAAAGACTCTCCCATGT  
TA

[0370] 示例性NPHS2转基因序列(SEQ ID NO:56)

[0371] 在一些实施方案中,HPHS2转基因包含与SEQ ID NO:57至少70%相同的核苷酸序列或其片段或由与SEQ ID NO:57至少70%相同的核苷酸序列或其片段组成。合适地,NPHS2

转基因包含与SEQ ID NO:57至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的核苷酸序列或其片段,或由与SEQ ID NO:57至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的核苷酸序列或其片段组成。

[0372] 在一些实施方案中,HPHS2转基因包含SEQ ID NO:57的核苷酸或其片段或由SEQ ID NO:57的核苷酸或其片段组成。

[0373] ATGGAGAGGAGGGCGCGAGCTCCTCCAGGGAGTCCCGCGGGCGAGGCGGCAGGACTCCGCACAAGGAG AACAAAGAGGGCAAAGGCCGAGAGGAGCGGCGGAGGCCGCGGGCGCCAGGAGGCTGGGCCCGAGCCGTCGGGCTCCGG ACGGGCGGGACCCCGGGGAGCCCGAGCGCCCGCCACGGTGGTGGACGTGGATGAGGTCCGAGGCTCCGGCG AGGAGGGCACCGAGGTGGTGGCGCTGTTGGAGAGCGAGCGGCCGAGGAAGGTACCAAATCCTCCGGCTTAGGGGCC TGTGAGTGGCTTCTTGCTCATTTCCCTGCTCTTCATCATCATGACCTTCCCTTTTTCCATCTGGTTCTGCGTAA GGTGTACAAGAGTATGAAAGAGTAATTATATTCGACTGGGACATCTGCTTCCCTGGAAGAGCCAAAGGCCCTGGTC TTTTCTTTTTTTTTGCCCTGCCTGGATACCTACCACAAGGTTGACCTTCGTCTCCAACTCTGGAGATACCTTTTCAT GAGATCGTGACCAAAGACATGTTTATAATGGAGATAGATGCCATTTGCTACTACCGAATGGAAAATGCCTCTCTTCT CCTAAGCAGTCTTGCTCATGTATCTAAAGCTGTGCAATTCCTGTGCAAACCACTATGAAGCGTCTCCTAGCACATC GATCCCTCACTGAAATTCTTCTAGAGAGGAAGAGCATCGCCCAAGATGCAAAGGTTGCCTTGGATTCAGTGACCTGT ATTTGGGGAATCAAAGTGGAGAGAATAGAAATTAAGATGTGAGGTTGCCAGCTGGGCTTCAGCACTCACTGGCTGT GGAGGCTGAAGCGCAAAGACAAGCCAAAGTGCGGATGATTGCTGCAGAAGCGAAAAGGCTGCTTCTGAGTCCCTGA GGATGGCAGCTGAGATTCTGTGAGGACCCCTGCTGCCGTTTCAGCTTCGATACCTCCACACCCTTCAGTCTCTGTCC ACAGAGAAGCCTTCCACTGTGGTTTTACCTTTGCCATTTGACCTACTGAATTGCCTGTCTTCTCCAGCAACAGAAC TCAGGGAAGCCTCCCCTTCCCAAGTCCCTCCAAACCTGTTGAGCCACTAAATCCTAAAAAGAAAGACTCTCCCATGT TATAG

[0374] 示例性NPHS2转基因序列(SEQ ID NO:57)

[0375] 在一些实施方案中,包含以下或由以下组成的启动子可操作地连接至编码NPHS2或其片段和/或变体的蛋白质编码序列:与SEQ ID NO:47具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0376] 在一些实施方案中,由SEQ ID NO:47的核苷酸序列组成的启动子可操作地连接至编码NPHS2或其片段和/或变体的蛋白质编码序列。

[0377] 血管内皮生长因子(VEGF)C转基因

[0378] 蛋白质编码序列可以包含血管内皮生长因子(VEGF)C转基因或由其组成。

[0379] VEGFC是淋巴管生成生长因子(lymphangiogenic growth factor),其已知经由两种受体(VEGFR-3(F1t4)和VEGFR-2(F1k4))进行信号传导。VEGFC由细胞以前肽形式产生,其进行二聚化,然后被切割成四聚体。

[0380] VEGFC转基因可包含编码任何形式的VEGFC例如前肽形式、四聚体形式、中间形式或完全加工的成熟VEGFC的多核苷酸。

[0381] 如果期望,编码不同形式的VEGFC多肽的多核苷酸可以以任何组合使用。优选地,VEGFC转基因包含编码以下的多核苷酸:一种或多种具有VEGFC生物活性的多肽,即可以结合并激活VEGFR-2和/或VEGRF-3的肽。更优选地,VEGFC转基因包含编码以下的多核苷酸:包

含VEGFC同源结构域并具有VEGFC生物活性的多肽,即可以结合并激活VEGFR-2和/或VEGFR-3的多肽。合适的VEGFC多核苷酸和多肽的进一步细节包括WO 2015/022447和US2014/0087002中描述的那些。

[0382] VEGFC多核苷酸可以包含SEQ ID NO:36的VEGFC开放阅读框(ORF)序列。VEGFC多核苷酸可以包含与SEQ ID NO:36的VEGFC ORF序列具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%至少96%、至少97%、至少98%或至少99%序列同一性的核苷酸序列。变体序列可以编码保留了结合和激活VEGFR-2和VEGFR-3的能力的VEGFC多肽。

[0383] 示例性VEGFC多核苷酸(SEQ ID NO:36)

[0384] ATGCACTTGCTGGGCTTCTTCTCTGTGGCGTGTCTCTGCTCGCCGCTGCGCTGCTCCCGGGTCTCGC  
GAGGCGCCCGCCGCCGCCGCCCTTCGAGTCCGGACTCGACCTCTCGGACGCGGAGCCCGACGCGGGCGAGGCCAC  
GGCTTATGCAAGCAAAGATCTGGAGGAGCAGTTACGGTCTGTGTCCAGTGTAGATGAACTCATGACTGTACTCTACC  
CAGAATATTGGAAAATGTACAAGTGTACAGCTAAGGAAAGGAGGCTGGCAACATAACAGAGAACAGGCCAACCTCAAC  
TCAAGGACAGAAGAGACTATAAAATTTGCTGCAGCACATTATAATACAGAGATCTTAAAAGTATTGATAATGAGTG  
GAGAAAGACTCAATGCATGCCACGGGAGGTGTGTATAGATGTGGGGAAGGAGTTTGGAGTCGCGACAAACACCTTCT  
TTAAACCTCCATGTGTGTCCGTCTACAGATGTGGGGGTTGCTGCAATAGTGAGGGGCTGCAGTGCATGAACACCAGC  
ACGAGCTACCTCAGCAAGACGTTATTTGAAATTACAGTGCCTCTCTCTCAAGGCCCAAACCAGTAACAATCAGTTT  
TGCCAATCACACTTCTGCGGATGCATGTCTAAACTGGATGTTTACAGACAAGTTCATTCCATTATTAGACGTTCCC  
TGCCAGCAACACTACCACAGTGTACAGCAGCGAACAAGACCTGCCCCACCAATTACATGTGGAATAATCACATCTGC  
AGATGCCTGGCTCAGGAAGATTTTATGTTTTCTCGGATGCTGGAGATGACTCAACAGATGGATTCCATGACATCTG  
TGGACCAAACAAGGAGCTGGATGAAGAGACCTGTCAGTGTGTCTGCAGAGCGGGGCTTCGGCCTGCCAGCTGTGGAC  
CCCACAAAGAACTAGACAGAACTCATGCCAGTGTGTCTGTAAAAACAACTCTTCCCCAGCCAATGTGGGGCCAAC  
CGAGAATTTGATGAAAACACATGCCAGTGTGTATGTAAAAGAACCTGCCCCAGAAATCAACCCCTAAATCCTGGAAA  
ATGTGCCTGTGAATGTACAGAAAAGTCCACAGAAATGCTTGTAAAAGGAAAGAAGTCCACCACCAAACATGCAGCT  
GTTACAGACGGCCATGTACGAACCGCCAGAAGGCTTGTGAGCCAGGATTTTCATATAGTGAAGAAGTGTGTCGTTGT  
GTCCCTTCATATTGGAAAAGACCACAAATGAGC

[0385] 在一些实施方案中,包含以下或由以下组成的启动子可操作地连接至VEGF(C)转基因:与SEQ ID NO:47具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0386] 在一些实施方案中,由SEQ ID NO:47的核苷酸序列组成的启动子可操作地连接至VEGF(C)转基因。

[0387] 补体蛋白和补体抑制剂

[0388] 蛋白质编码序列可以编码补体蛋白,或其片段和/或变体。

[0389] 如本文所用,“补体蛋白”是补体系统的一部分的蛋白质。补体系统,也称为补体级联,是充当抵御外来的和改变的宿主细胞的第一道防线的先天免疫的中心部分。补体系统由主要由肝脏产生的血浆蛋白或细胞表面上表达的膜蛋白组成。补体在血浆、组织或细胞内发挥作用。补体蛋白作为级联协作,以调理病原体并诱导一系列炎症应答,这帮助免疫细胞抵抗感染并维持稳态(Merle,N.S.等人,2015.Frontiers in immunology,6,262)。

[0390] 有三种补体激活的途径:经典途径、旁路途径和凝集素途径。三种补体途径的靶标识别机制中不同,但都集中在中心组分C3的激活上。激活后,C5被切割,并且开始攻膜复合

物(MAC)的组装。C3和C5的酶切割导致过敏毒素(anaphylotoxin)C3a和C5a的产生和释放。

[0391] 合适地,补体蛋白选自由CFI、CFH、FHL-1、C1INH、C4BP、MASP2、C3、C5aR1、C5、C5a、CD55、CD35、CD46、CD59、玻连蛋白和簇集素或其片段和/或其变体组成的列表。

[0392] 蛋白质编码序列可以编码补体系统的抑制剂,或其片段和/或变体。

[0393] 如本文所用,“补体系统的抑制剂”或“补体抑制剂”是防止补体系统激活的蛋白质。补体受到这些抑制剂的严格控制,它们天然保护自身细胞和组织免受不需要的补体激活。补体抑制剂可以调控经典途径、凝集素途径和旁路途径不同阶段的补体激活。合适地,补体抑制剂是天然存在的补体抑制剂,或其片段和/或变体。优选地,补体系统的抑制剂是人中的补体系统抑制剂。

[0394] 补体抑制剂分成两类:可溶性抑制剂和膜结合抑制剂。优选地,补体系统的抑制剂是可溶性补体抑制剂。可溶性补体抑制剂包括C1抑制剂(C1INH)、补体因子I(CFI)、补体因子H(CFH)、补体因子H样蛋白1(FHL-1)、C4结合蛋白(C4BP)、簇集素和玻连蛋白。膜结合调节因子包括CD46、CD55、CD59、CD35以及CUB和Sushi多结构域1(CSMD1)。

[0395] 补体系统的抑制剂可以选自:CFI、CFH、FHL-1、C1INH、C4BP、CD46、CD55、CD59、CD35、玻连蛋白、簇集素和CSMD1,或其片段和/或变体。

[0396] 优选地,补体系统的抑制剂选自:CFI、CFH和FHL-1或其片段和/或变体。

[0397] CFI

[0398] 蛋白质编码序列可以编码CFI,或其片段和/或变体。

[0399] 补体因子I(CFI)是一种胰蛋白酶样丝氨酸蛋白酶,其通过切割C3b的 $\alpha$ 链中的三个肽键和C4b的 $\alpha$ 链中的两个键来抑制补体系统,从而使这些蛋白质失活。

[0400] CFI是糖蛋白异二聚体,其由二硫键连接的重链和轻链组成。该重链具有四个结构域:FI攻膜复合物(FIMAC)结构域、CD5结构域以及低密度脂蛋白受体1(LDLr1)结构域和低密度脂蛋白受体2(LDLr2)结构域。重链在维持酶失活中发挥抑制作用,直到其与底物(C3b或者C4b)和辅因子蛋白(因子H、C4b结合蛋白、补体受体1和膜辅因子蛋白)形成的复合物相遇。酶与底物:辅因子复合物结合后,重链:轻链界面被扰乱,并且该酶通过变构激活。该轻链仅含有丝氨酸蛋白酶结构域。该结构域含有负责C3b和C4b的特异性切割的催化三联体His-362、Asp-411和Ser-507。

[0401] CFI或其片段和/或变体可以能够将C3b切割成iC3b和/或可以能够将iC3b切割成C3d,g。

[0402] CFI的片段和/或变体可以保留至少50%、60%、70%、80%、90%、95%或100%的天然CFI的C3b失活和iC3b降解活性。CFI和天然CFI的片段和/或变体的C3b失活和iC3b降解活性,可以使用本领域技术人员已知的任何合适的方法来确定。例如,使用蛋白水解测定。

[0403] 优选地,CFI是人CFI。人CFI的一个实例是具有UniProtKB登录号P05156的CFI。

[0404] 合适地,CFI可以包含以下或由以下组成:如SEQ ID NO:49所示的多肽序列,或与SEQ ID NO:49至少70%相同的变体。

[0405] 说明性CFI多肽序列(SEQ ID NO:49)

[0406] MKLLHVLLFLCFHLRFCKVTYTSQEDLVEKKCLAKKYTHLSCDKVFCQPWQRCIEGTCVCKLPYQCPK  
NGTAVCATNRRSFPTYCQQKSLECLHPGTKFLNNGTCTAEGKFSVSLKHGNTDSEGIVEVKLVQDKTMFICKSSWS  
MREANVACLDLGFQQGADTQRRFKLSDSLINSTECLHVHCRGLETSLAECTFTKRRTMGYQDFADVVCYTQKADSPM

DDFFQCVNGKYISQMKACDGINDCGDQSDDELCKACQKGFHCKSGVCIPSQYQCNGEVDICITGEDEVGCAGFASVT  
 QEETEILTADMDAERRRIKSLPKLSCGVKNRMHIRRKRIVGGKRAQLGDLPWQVAIKDASGITCGGIYIGGCWILT  
 AAHCLRASKTHRYQIWTTVVVDWIHPDLKRIVIEYVDRIIFHENYNAGTYQNDIALIEMKKDGNKKDCELPRSIPACV  
 PWSPYLFQPNDTCTIVSGWGREKDNERVFSLQWGEVKLISNCSKFYGNRFYEKEMECAGTYDGSIDACKGDSGGPLVC  
 MDANNVTYVWGVVSWGENCGKPEFPVYTKVANYFDWISYHVGRPFISQYNV

[0407] 合适地,变体与SEQ ID NO:49可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0408] 编码CFI的实例核苷酸序列的是NM\_000204.5。合适地,编码CFI的蛋白编码序列可以包含以下或由以下组成:如SEQ ID NO:50所示的多核苷酸序列,或与SEQ ID NO:50至少70%相同的变体。

[0409] 说明性CFI多核苷酸序列(SEQ ID NO:50)

[0410] ATGAAGCTTCTTCATGTTTTCTGTATTCTGTGCTTCCACTTAAGTTTTGCAAGGTCCTTATACA  
 TCTCAAGAGGATCTGGTGGAGAAAAAGTGCTTAGCAAAAAATATACTCACCTCTCCTGCGATAAAGTCTTCTGCCA  
 GCCATGGCAGAGATGCATTGAGGGCACCTGTGTTGTAACTACCGTATCAGTGCCCAAAGAATGGCACTGCAGTGT  
 GTGCAACTAACAGGAGAAGCTTCCCAACATACTGTCAACAAAAGAGTTTGAATGTCTTCATCCAGGGACAAAGTTT  
 TAAATAACGGAACATGCACAGCCGAAGGAAAGTTTAGTGTTCCTTGAAGCATGGAATACAGATTAGAGGGAAT  
 AGTTGAAGTAAACTTGTGGACCAAGATAAGACAATGTTTCATATGAAAAGCAGCTGGAGCATGAGGGAAGCCAACG  
 TGGCCTGCCTTGACCTTGGGTTTCAACAAGGTGCTGATACTCAAAGAAGGTTAAGTTGTCTGATCTCTATAAAT  
 TCCACTGAATGTCTACATGTGCATTGCCGAGGATTAGAGACCAGTTGGCTGAATGTACTTTACTAAGAGAAGAAC  
 TATGGGTTACCAGGATTTGCTGATGTGGTTTGTATACACAGAAAGCAGATTCTCCAATGGATGACTTCTTTCAGT  
 GTGTGAATGGGAAATACATTTCTCAGATGAAAGCCTGTGATGGTATCAATGATTGTGGAGACCAAAGTATGAACTG  
 TGTTGTAAAGCATGCCAAGGCAAAGGCTTCCATTGCAAATCGGGTGTTCATTCGAAGCCAGTATCAATGCAATGG  
 TGAGGTGGACTGCATTACAGGGGAAGATGAAGTTGGCTGTGCAGGCTTTCATCTGTGACTCAAGAAGAAACAGAAA  
 TTTTGACTGCTGCATGGATGCAGAAAGAAGACGGATAAAAATCATTATTACCTAACTATCTTGTGGAGTTAAAAAC  
 AGAATGCACATTGCAAGGAAACGAATTGTGGGAGGAAAGCGAGCACAACCTGGGAGACCTCCCATGGCAGGTGGCAAT  
 TAAGGATGCCAGTGAATCACCTGTGGGGGAATTTATATTGGTGGCTGTTGGATTCTGACTGCTGCACATTGTCTCA  
 GAGCCAGTAAACTCATCGTTACCAATATGGACAACAGTAGTAGACTGGATACACCCCGACCTTAAACGTATAGTA  
 ATTGAATACGTGGATAGAATTATTTCCATGAAAACATAATGCAGGCACTTACCAAAATGACATCGCTTTGATTGA  
 AATGAAAAAAGACGGAAACAAAAAAGATTGTGAGCTGCCTCGTTCCATCCCTGCCTGTGTCCCCTGGTCTCCTTACC  
 TATTCCAACCTAATGATACATGCATCGTTTCTGGCTGGGGACGAGAAAAAGATAACGAAAGAGTCTTTTCACTTCAG  
 TGGGGTGAAGTTAAACTAATAAGCAACTGCTCTAAGTTTTACGAAAATCGTTTCTATGAAAAAGAAATGGAATGTGC  
 AGGTACATATGATGGTTCCATCGATGCCTGTAAAGGGGACTCTGGAGGCCCTTAGTCTGTATGGATGCCAACAATG  
 TGAATTATGTCTGGGGTGTGTGAGTTGGGGGAAAACTGTGAAAACCAGAGTCCCAGGTGTTTACACCAAAGTG  
 GCCAATTATTTGACTGGATTAGCTACCATGTAGGAAGGCCCTTTATTTCTCAGTACAATGTATAA

[0411] 合适地,变体与SEQ ID NO:50可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0412] 在一些实施方案中,包含以下或由以下组成的启动子可操作地连接至编码CFI或其片段和/或变体的蛋白质编码序列:与SEQ ID NO:47具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的

核苷酸序列。

[0413] 在一些实施方案中,由SEQ ID NO:47的核苷酸序列组成的启动子可操作地连接至编码CFI或其片段和/或变体的蛋白质编码序列。

[0414] CFH

[0415] 蛋白质编码序列可以编码CFH,或其片段和/或变体。

[0416] 补体因子H (CFH) 调控自身细胞和表面上的补体激活。CFH竞争补体因子B (CFB) 与C3b的结合,作为CFI催化的C3b蛋白水解切割的辅助因子,并加速C3bBb和C3b2Bb不可逆解离成它们各自的组分。因此,CFH不仅抑制转化酶的形成,而且缩短其形成的任何转化酶复合物的寿命。

[0417] CFH是一种大的 (155kDa) 可溶性糖蛋白。CFH总共由20个结构域组成,每个结构域含有大约60个氨基酸残基,并且称为补体控制蛋白模块 (CCP) 或由3-8个残基组成的短接头连接的短共有重复序列。CCP模块从1-20 (从蛋白质的N末端) 编号:CCP 1-4和CCP 19-20与C3b结合,而CCP 7和CCP 19-20与GAG和唾液酸结合。

[0418] CFH或其片段和/或变体可以能够结合C3b和/或C3d;和/或充当C3b的CFI催化的蛋白水解切割的辅助因子;和/或增加C3bBb和C3b2Bb不可逆解离成它们各自的组分。CFH的片段和/或变体可以保留至少50%、60%、70%、80%、90%、95%或100%天然CFH的活性。CFH和天然CFH的片段和/或变体的活性可以使用本领域技术人员已知的任何合适的方法来确定。

[0419] 优选地,CFH是人CFH。人CFH的一个实例是具有UniProtKB登录号P08603的CFH。

[0420] 合适地,CFH可以包含以下或由以下组成:如SEQ ID NO:51所示的多肽序列,或与SEQ ID NO:51至少70%相同的变体。

[0421] 说明性CFH多肽序列 (SEQ ID NO:51)

[0422] MRLLAKIICLMLWAICVAEDCNELPPRRNTEILTGSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKG  
 EWVALNPLRKCQKRPCGHPGDTFPGTFTLTGGNVFEYGVKAVYTCNEGYQLLGEINYRECDTDGWTNDIPICEVVKC  
 LPVTAPENGIKIVSSAMEPDREYHFGQAVRFVCNSGYKIEGDEEMHCSDDGFWSKEKPKCVEISCKSPDVINGSPISQ  
 KIIYKENERFQYKCNMGYEYSERGDVCTESGWRPLPSCEEKSCDNPIPNGDYSPLRIKHRTGDEITYQCRNGFYF  
 ATRGNTAKCTSTGWIPAPRCTLKPCDYPDIKHGGLYHENMRRPYFPVAVGKYYSYCCDEHFETPSGSYWDHIHCTQD  
 GWSPAVPCLRKCYFPYLENGYNQNYGRKFVQGKSIDVACHPGYALPKAQTTVTCMENGWSPTPRCIRVKTCSSSID  
 IENGFISESQYTYALKEKAKYQCKLGYVTADGETSGSITCGKDGWSAQPTCIKSCDIPVFMNARTKNDFTWFKLNDT  
 LDYEC HDGYESNTGSTTGSIVCGYNGWSDLPICYERECELPKIDVHLPDRKKDQYKVGVEVLKFSCPKPGFTIVGPN  
 VQCYHFGLSPDLPICKEQVQSCGPPPELLNGNVKEKTKEEYGHSEVVEYCNPRFLMKGPNKIQCVDGEWTTLPVCI  
 VEESTCGDIPLEHGWAQLSSPPYYGDSVEFNCSSEFTMIGHRSITCIHGVWTQLPQCVAIDKLLKCKSSNLIILE  
 EHLKNKKEFDHNSNIRYRCRGKEGWIHTVCINGRWDPEVNCSMAQIQLCPPPPQIPNSHNMTTTLNYRDGEKVSVLC  
 QENYLIQEGEEITCKDGRWQSIPLCVEKIPCSQPPQIEHGTINSSRSSQESYAHGTKLSYTCGGFRISENETTCY  
 MGKWSPPQCEGLPCKSPPEISHGVVAHMSDSYQYGEEVYKCFEGFGIDGPAIAKCLGEKWSHPPSICKTDCLSLP  
 SFENAIPMGEEKDVYKAGEQVYTYTCATYYKMDGASNVTCINSRWTGRPTCRDTSVCNPPVQVQAYIVSRQMSKYP  
 ERVRYQCRSPYEMFGDEEVMCLNGNWTEPPQCKDSTGKCGPPPPIDNGDITSFPLSVYAPASSVEYQCQNLYQLEGN  
 KRITCRNGQWSEPPKCLHPCVISREIMENYNIALRWTAQKLYSRTGESVEFVCKRGYRLSSRSHTLRITTCWDGKLE  
 YPTCAKR

[0423] 合适地,变体与SEQ ID NO:51可以是至少75%、至少80%、至少85%、至少90%、至少95%至少96%、至少97%、至少98%或至少99%相同的。

[0424] 编码CFH的实例核苷酸序列的是NM\_000186.4。合适地,编码CFH的蛋白编码序列可以包含以下或由以下组成:如SEQ ID NO:52所示的多核苷酸序列,或与SEQ ID NO:52至少70%相同的变体。

[0425] 说明性CFH多核苷酸序列(SEQ ID NO:52)

[0426] ATGAGACTTCTAGCAAAGATTATTTGCCTTATGTTATGGGCTATTTGTGTAGCAGAAGATTGCAATGAACTTCCTCCAAGAAGAAATACAGAAATCTGACAGGTTCTGGTCTGACCAAACATATCCAGAAGGCACCCAGGCTATCTATAAATGCCGCCCTGGATATAGATCTCTTGAAATGTAATAATGGTATGCAGGAAGGGAGAATGGGTTGCTCTTAATCCATTAAGGAAATGTCAGAAAAGGCCCTGTGGACATCCTGGAGATACTCCTTTTGGTACTTTTACCCTTACAGGAGGAAATGTGTTTTGAATATGGTGTAAAAGCTGTGTATACATGTAATGAGGGTATCAATTGCTAGGTGAGATTAATTAACCGTGAATGTGACACAGATGGATGGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTACCAGTGACAGCACAGAGAATGGAAAAATTGTCAGTAGTGCAATGGAACCAGATCGGGAATACCATTTTGGACAAGCAGTACGGTTTTGTATGTAACTCAGGCTACAAGATTGAAGGAGATGAAGAAATGCATTGTTTCAGACGATGGTTTTTGGAGTAAAGAGAAACC AAAGTGTGTGGAAATTTTCATGCAAATCCCGAGATGTTATAAATGGATCTCCTATATCTCAGAAGATTATTTATAAGGAGAATGAACGATTTCAATATAAATGTAACATGGGTTATGAATACAGTGAAAGAGGAGATGCTGTATGCACTGAATCTGGATGGCGTCCGTTGCCCTTCATGTGAAGAAAAATCATGTGATAATCCTTATATTCCAAATGGTGACTACTCACCTTTAAGGATTAACACAGAACTGGAGATGAAATCACGTACCAGTGTAGAAATGGTTTTTATCCTGCAACCCGGGGAAATACAGCAAATGCACAAGTACTGGCTGGATACCTGCTCCGAGATGTACCTGAAACCTGTGATTATCCAGACATTAACATGGAGGTCTATATCATGAGAATATGCGTAGACCATACTTCCAGTAGCTGTAGGAAAATATTACTCCTATTACTGTGATGAACATTTTGAG

[0427] ACTCCGTCAGGAAGTTACTGGGATCACATTCATTGCACACAAGATGGAT

[0428] GGTCGCCAGCAGTACCATGCCTCAGAAAATGTTATTTTCCTTATTTGGAA

[0429] AATGGATATAATCAAAATCATGGAAGAAAGTTTGTACAGGGTAAATCTAT

[0430] AGACGTTGCCTGCCATCCTGGCTACGCTCTTCCAAAAGCGCAGACCACA

[0431] GTTACATGTATGGAGAATGGCTGGTCTCCTACTCCAGATGCATCCGTGT

[0432] CAAAACATGTTCCAAATCAAGTATAGATATTGAGAATGGGTTTATTTCTG

[0433] AATCTCAGTATACATATGCCTTAAAAGAAAAAGCGAAATATCAATGCAA

[0434] ACTAGGATATGTAACAGCAGATGGTGAACATCAGGATCAATTACATGT

[0435] GGGAAAGATGGATGGTCAGCTCAACCCACGTGCATTAATCTTGTGATA

[0436] TCCCAGTATTTATGAATGCCAGAACTAAAAATGACTTCACATGGTTTAAG

[0437] CTGAATGACACATTGGACTATGAATGCCATGATGGTTATGAAAGCAATAC

[0438] TGGAAGCACCCTGGTTCCATAGTGTGTGGTTACAATGGTTGGTCTGAT

[0439] TTACCCATATGTTATGAAAGAGAATGCGAACTTCTAAAATAGATGTACA

[0440] CTTAGTTCCTGATCGCAAGAAAGACCAGTATAAAGTTGGAGAGGTGTTG

[0441] AAATTCTCCTGCAAACCAGGATTTACAATAGTTGGACCTAATTCGGTTCA

[0442] GTGCTACCCTTTGGATTGTCTCCTGACCTCCCAATATGTAAAGAGCAA

[0443] GTACAATCATGTGGTCCACCTCCTGAACTCCTCAATGGGAATGTTAAGG

[0444] AAAAAACGAAAGAAGAATATGGACACAGTGAAGTGGTGAATATTATT

- [0445] GCAATCCTAGATTTCTAATGAAGGGACCTAATAAAATTCAATGTGTTGAT  
[0446] GGAGAGTGGACAACCTTTACCAGTGTGTATTGTGGAGGAGAGTACCTGT  
[0447] GGAGATATACCTGAACTTGAACATGGCTGGGCCAGCTTTCTTCCCCTC  
[0448] CTTATTACTATGGAGATTCAAGTGAATTCAATTGCTCAGAATCATTTACAA  
[0449] TGATTGGACACAGATCAATTACGTGTATTCATGGAGTATGGACCAACTT  
[0450] CCCCAGTGTGTGGCAATAGATAAACTTAAGAAGTGCAAATCATCAAAT  
[0451] TAATTATACTTGAGGAACATTTAAAAACAAGAAGGAATTCGATCATAAT  
[0452] TCTAACATAAGGTACAGATGTAGAGGAAAAGAAGGATGGATACACACA  
[0453] GTCTGCATAAATGGAAGATGGGATCCAGAAGTGAAGTGTCAATGGCAC  
[0454] AAATACAATTATGCCCACCTCCACCTCAGATTCCCAATTCTCACAATATG  
[0455] ACAACCACACTGAATTATCGGGATGGAGAAAAAGTATCTGTTCTTTGCC  
[0456] AAGAAAATTATCTAATTCAGGAAGGAGAAGAAATTACATGCAAAGATGG  
[0457] AAGATGGCAGTCAATACCACTCTGTGTTGAAAAAATTCATGTTTACAA  
[0458] CCACCTCAGATAGAACACGGAACCATTAATTCATCCAGGTCTTCACAAG  
[0459] AAAGTTATGCACATGGGACTAAATTGAGTTATACTTGTGAGGGTGGTTTC  
[0460] AGGATATCTGAAGAAAATGAAACAACATGCTACATGGGAAAATGGAGTT  
[0461] CTCCACCTCAGTGTGAAGGCCCTTCCTTGTAATCTCCACCTGAGATTTCT  
[0462] CATGGTGTGTAGCTCACATGTCAGACAGTTATCAGTATGGAGAAGAAG  
[0463] TTACGTACAAATGTTTTGAAGTTTTTGAATTGATGGGCCTGCAATTGCA  
[0464] AAATGCTTAGGAGAAAAATGGTCTCACCTCCATCATGCATAAAAAACAG  
[0465] ATTGTTCTCAGTTTACCTAGCTTTGAAAATGCCATACCCATGGGAGAGAA  
[0466] GAAGGATGTGTATAAGGCGGGTGAAGCAAGTACTTACACTTGTGCAACA  
[0467] TATTACAAAATGGATGGAGCCAGTAATGTAACATGCATTAATAGCAGATG  
[0468] GACAGGAAGGCCAACATGCAGAGACACCTCCTGTGTGAATCCGCCAC  
[0469] AGTACAAAATGCTTATATAGTGTGCGAGACAGATGAGTAAATATCCATCTG  
[0470] GTGAGAGAGTACGTTATCAATGTAGGAGCCCTTATGAAATGTTGGGGA  
[0471] TGAAGAAGTGTGTGTTTAAATGGAACTGGACGGAACCACCTCAATG  
[0472] CAAAGATTCTACAGGAAAATGTGGGCCCTCCACCTATTGACAATGGG  
[0473] GACATTACTTCATTCCCCTTGTGAGTATATGCTCCAGCTTCATCAGTTGA  
[0474] GTACCAATGCCAGAACTTGTATCAACTTGAGGGTAACAAGCGAATAACA  
[0475] TGTAGAAATGGACAATGGTCAGAACCACAAAATGCTTACATCCGTGTG  
[0476] TAATATCCCAGAAAATTATGGAAAATTATAACATAGCATTAAAGGTGGACA  
[0477] GCCAAACAGAAGCTTTATTCGAGAACAGGTGAATCAGTTGAATTTGTGT  
[0478] GTAAACGGGGATATCGTCTTTTCATCACGTTCTCACACATTGCGAACAAC  
[0479] ATGTTGGGATGGGAACTGGAGTATCCAATTGTGCAAAAAGATAG  
[0480] 合适地,变体与SEQ ID NO:52可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。  
[0481] CFH片段可以是剪接变体。例如,补体因子H样蛋白1 (FHL-1) 是CFH基因剪接变体,其与CFH的N端7个结构域 (CCP 1-7) 几乎相同。

[0482] 在一些实施方案中,包含以下或由以下组成的启动子可操作地连接至编码CFH或其片段和/或变体的蛋白质编码序列:与SEQ ID NO:47具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0483] 在一些实施方案中,由SEQ ID NO:47的核苷酸序列组成的启动子可操作地连接至编码CFH或其片段和/或变体的蛋白质编码序列。

[0484] FHL-1

[0485] 蛋白质编码序列可以编码FHL-1,或其片段和/或变体。

[0486] FHL-1或其片段和/或变体可以能够结合C3b和/或C3d。FHL-1的片段和/或变体可以保留至少50%、60%、70%、80%、90%、95%或100%的天然FHL-1的活性。FHL-1和天然FHL-1的片段和/或变体的活性可以使用本领域技术人员已知的任何合适的方法来确定。

[0487] 优选地,FHL-1是人FHL-1。人FHL-1的一个示例是具有NCBI参考序列:NP\_001014975.1的FHL-1。

[0488] 合适地,FHL-1可以包含以下或由以下组成:如SEQ ID NO:53所示的多肽序列,或与SEQ ID NO:53至少70%相同的变体。

[0489] 说明性FHL-1多肽序列(SEQ ID NO:53)

[0490] MRLLAKI ICLMLWAICVAEDCNELPPRRNTEILTGSWSDQTYPEGTQAIYKCRPGYRSLGNVIMVCRKG  
EWVALNPLRKCQKRPCGHPGDTPFGTFTLTGGNVFEYGVKAVYTCNEGYQLLGEINYRECDTDGWTNDIPICEVVKC  
LPVTAPENKIVSSAMEPDREYHFGQAVRFVCNSGYKIEGDEEMHCSDDGFWSKEKPKCVEISCKSPDVINGSPISQ  
KIIYKENERFQYKCNMGYEYSERGDVCTESGWRPLPSCEEKSCDNPYIPNGDYSPLRIKHRTGDEITYQCRNGFYF  
ATRGNTAKCTSTGWIPAPRCTLKPCDYPDIKHGGLYHENMRRPYFPVAVGKYYSYCCDEHFETPSGSYWDHIHCTQD  
GWSPA VPCLRKCYPYLENGYNQNHGRKFVQGKSIDVACHPGYALPKAQT TVTCMENGWSPTPRCIRVSFTL

[0491] 合适地,变体与SEQ ID NO:53可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0492] 编码FHL-1的实例核苷酸序列的是NM\_001014975.2。合适地,编码FHL-1的蛋白编码序列可以包含以下或由以下组成:如SEQ ID NO:54所示的多核苷酸序列,或与SEQ ID NO:54至少70%相同的变体。

[0493] 说明性FHL-1多核苷酸序列(SEQ ID NO:54)

[0494] ATGAGACTTCTAGCAAAGATTATTTGCCTTATGTTATGGGCTATTTGTGTAGCAGAAGATTGCAATGAA  
CTTCCTCCAAGAAGAAATACAGAAATTTCTGACAGGTTCCCTGGTCTGACCAAACATATCCAGAAGGCACCCAGGCTAT  
CTATAAATGCCGCCCTGGATATAGATCTCTTGAAAATGTAATAATGGTATGCAGGAAGGGAGAATGGGTTGCTCTTA  
ATCCATTAAGGAAATGTCAGAAAAGGCCCTGTGGACATCCTGGAGATACTCCTTTTGGTACTTTTACCCTTACAGGA  
GGAAATGTGTTTGAATATGGTGTAAAAGCTGTGTATACATGTAATGAGGGTATCAATTGCTAGGTGAGATTAATTA  
CCGTGAATGTGACACAGATGGATGGACCAATGATATTCCTATATGTGAAGTTGTGAAGTGTACCAGTGACAGCAC  
CAGAGAATGGAAAAATTGTCAGTAGTGCAATGGAACCAGATCGGGAATACCATTTTGGACAAGCAGTACGGTTTGTA  
TGTA ACTCAGGCTACAAGATTGAAGGAGATGAAGAAAATGCATTGTTTCAGACGATGGTTTTTGGAGTAAAGAGAAACC  
AAAGTGTGTGGAAATTTTCATGCAAAATCCCAGATGTTATAAATGGATCTCCTATATCTCAGAAGATTATTTATAAGG  
AGAATGAACGATTTCAATATAAATGTAACATGGGTTATGAATACAGTGAAGAGGAGATGCTGTATGCACTGAATCT  
GGATGGCGTCCGTTGCCTTCATGTGAAGAAAAATCATGTGATAATCCTTATATTCCAAATGGTGACTACTCACCTTT

AAGGATTA AACACAGAACTGGAGATGAAATCACGTACCAGTGTAGAAATGGTTTTATCCTGCAACCCGGGGAAATA  
CAGCAAAATGCACAAGTACTGGCTGGATACCTGCTCCGAGATGTACCTGAAACCTGTGATTATCCAGACATTA  
CATGGAGGTCTATATCATGAGAATATGCGTAGACCATACTTCCAGTAGCTGTAGGAAAATATTACTCCTATTACTG  
TGATGAACATTTTGGACTCCGTCAGGAAGTTACTGGGATCACATTCATTGCACACAAGATGGATGGTCGCCAGCAG  
TACCATGCCTCAGAAAATGTTATTTTCCTTATTTGGAAAATGGATATAATCAAAATCATGGAAGAAAGTTTGTACAG  
GGTAAATCTATAGACGTTGCCATCCTGGCTACGCTCTCCAAAAGCGCAGACCACAGTTACATGTATGGAGAA  
TGGCTGGTCTCCTACTCCCAGATGCATCCGTGTCAGCTTTACCCTCTGA

[0495] 合适地,变体与SEQ ID NO:54可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0496] 在一些实施方案中,包含以下或由以下组成的启动子可操作地连接至编码FHL-1或其片段和/或变体的蛋白质编码序列:与SEQ ID NO:47具有至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、至少99%或100%同一性的核苷酸序列。

[0497] 在一些实施方案中,由SEQ ID NO:47的核苷酸序列组成的启动子可操作地连接至编码FHL-1或其片段和/或变体的蛋白质编码序列。

[0498] 其他调控元件

[0499] 除了本发明的启动子之外,多核苷酸还可以包含在转录前或转录后可以发挥作用的一个或多个进一步的调控序列。合适地,蛋白质编码序列可以可操作地连接至一个或多个进一步的调控序列。一个或多个进一步的调控序列可以促进蛋白质在肾小球细胞(例如足细胞)中的表达。

[0500] “调控序列”是促进多肽表达,例如起到增加转录物的表达或增强mRNA稳定性的作用的任何序列。合适的进一步的调控序列包括例如增强子元件、转录后调控元件和多腺苷酸化位点。

[0501] 合适地,本发明的多核苷酸不包含任何进一步的调控序列(除了本发明的启动子之外)。

[0502] 增强子

[0503] 本发明的多核苷酸可以包含增强子。合适地,增强子可以可操作地连接至蛋白质编码序列。增强子可以促进蛋白质在肾小球细胞(例如足细胞)中的表达。合适地,增强子是哺乳动物增强子,例如人增强子。

[0504] “增强子”是DNA的区,其可以被蛋白质(激活剂)结合,以增加特定的基因转录将发生的可能性。增强子是顺式作用的。它们位于距基因最远达1Mbp(1,000,000bp)处,位于起始位点的上游或下游。可以使用任何合适的增强子,技术人员可以容易地进行此选择。

[0505] 本发明的多核苷酸可以包含组织特异性增强子。如本文所用,“组织特异性增强子”是优先促进基因在特定细胞或组织中的表达的增强子。合适地,与其他细胞类型相比,组织特异性增强子可以促进基因在特定细胞类型中的更高表达。较高的表达可以例如通过测量可操作地连接至增强子的转基因(例如GFP)的表达来测量,其中该转基因的表达与增强子促进基因表达的能力相关。例如,与其他细胞类型中的表达水平相比,组织特异性增强子可以是促进基因在特定细胞类型中的表达水平至少10%更高、至少20%更高、至少30%更高、至少40%更高、至少50%更高、至少100%更高、至少200%更高、至少300%更高、至少

400%更高、至少500%更高或至少1000%更高的增强子。

[0506] 合适的组织特异性增强子将是本领域技术人员众所周知的。增强子可以是肾特异性增强子,优选肾小球特异性增强子,更优选足细胞特异性增强子。合适地,增强子可以可操作地连接至蛋白质编码序列。

[0507] 合适地,增强子可以是与在人肾、肾小球细胞和/或足细胞中选择性表达的基因相关联的增强子或可以衍生自与在人肾、肾小球细胞和/或足细胞中选择性表达的基因相关联的增强子。鉴定与基因相关的增强子区域的方法将是本领域技术人员众所周知的。

[0508] 合适地,足细胞特异性增强子是NPHS1增强子或NPHS2增强子,或其片段或衍生物。合适地,足细胞特异性增强子是NPHS1增强子,或其片段或衍生物。

[0509] NPHS1增强子已在Guo,G.等人,2004. *Journal of the American Society of Nephrology*, 15(11), pp. 2851-2856. 中描述。当将来自人NPHS1启动子的186-bp片段放置在转基因小鼠中的异源最小启动子之前时,该片段能够指导 $\beta$ -半乳糖苷酶转基因的足细胞特异性表达。

[0510] 合适地,NPHS1增强子可以包含以下或由以下组成:如SEQ ID NO:27所示的核苷酸序列,或与SEQ ID NO:27至少70%相同的变体。

[0511] 示例性NPHS1增强子(SEQ ID NO:27):

[0512] ctgctgagctgggagaccacttgatctgacttctcccatcttcccagcctaagccaggccctggggt  
cacggaggctggggaggcaccgaggaacgcgcctggcatgtgctgacaggggattttatgctccagctgggccagc  
tgggaggagcctgctgggcagaggccagagctgggggctctg

[0513] 合适地,变体与SEQ ID NO:27可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、或至少99%相同的。

[0514] 合适地,本发明的多核苷酸不包含增强子。

[0515] 科扎克序列

[0516] 本发明的多核苷酸可以包含科扎克序列。合适地,科扎克序列可以可操作地连接至蛋白质编码序列。科扎克序列可以插入到蛋白质的起始密码子之前以改善翻译的起始。

[0517] 合适的科扎克序列将是本领域技术人员众所周知的。

[0518] 合适地,科扎克序列子可以包含以下或由以下组成:如SEQ ID NO:28所示的核苷酸序列,或与SEQ ID NO:28至少65%相同的变体。

[0519] 示例性科扎克序列(SEQ ID NO:28):

[0520] GCCGCCACCAUGG

[0521] 合适地,变体与SEQ ID NO:28可以是至少75%、至少85%或至少90%相同的。

[0522] 合适地,本发明的多核苷酸不包含科扎克序列。

[0523] 转录后调控元件

[0524] 本发明的多核苷酸可以包含转录后调控元件。合适地,转录后调控元件可以可操作地连接至蛋白质编码序列。转录后调控元件可以改善基因表达。

[0525] 多核苷酸可以包含土拨鼠肝炎病毒转录后调控元件(WPRE)。合适地,WPRE可以可操作地连接至蛋白质编码序列。

[0526] 在X抗原启动子和/或X抗原的起始密码子之内,WPRE序列可以具有取代、缺失或插入。这可以阻止功能性X抗原的产生。

[0527] 合适地, WPRE可以包含以下或由以下组成:如SEQ ID NO:29所示的多核苷酸序列, 或与SEQ ID NO:29至少70%相同的变体。

[0528] 示例性WPRE (SEQ ID NO:29):

[0529] aatcaacctctggattacaaaatttgtgaaagattgactggatttcttaactatggtgctccttttac  
gctatgtggatacgtgctttaatgcctttgtatcatgctattgcttcccgtatggctttcattttctcctccttg  
tataaatcctgggtgctgtctctttatgaggagttgtggcccgttgcaggcaacgtggcgtggtgtgcactgtgt  
ttgctgacgcaacccccactgggtggggcattgccaccacctgtcagctcctttccgggactttcgctttcccct  
ccctattgccacggcggaactcatcgccgcctgccttggccgctgctggacaggggctcggtggttgggcactgac  
aattccgtggtgtgtcggggaaatcatcgctcctttccttggctgctcgcctgtgttgcacctggattctgcgcg  
ggacgtccttctgctacgtcccttcggccctcaatccagcggaccttccctcccggcctgctgccggctctgcg  
gcctcttccgctcttccgcttccgctcagacgagtcggatctccctttgggccgcctccccgc

[0530] 合适地, 变体与SEQ ID NO:29可以是至少75%、至少80%、至少85%、至少90%、至少95%至少96%、至少97%、至少98%或至少99%相同的。

[0531] 合适地, 本发明的多核苷酸不包含转录后调控元件。

[0532] 多腺苷酸化信号

[0533] 本发明的多核苷酸可以包含多腺苷酸化信号。合适地, 多腺苷酸化信号可以可操作地连接至蛋白质编码序列。多腺苷酸化信号可以改善基因表达。

[0534] 合适的多腺苷酸化信号包括早期SV40多腺苷酸化信号 (SV40pA)、鸡β珠蛋白多腺苷酸化信号、牛生长激素多腺苷酸化信号 (bGH) 或可溶性神经纤毛蛋白1 (neuropilin-1) 多腺苷酸化信号。在一些实施方案中, 多腺苷酸化信号是早期SV40多腺苷酸化信号 (SV40pA) 或鸡β珠蛋白多腺苷酸化信号。优选地, 多腺苷酸化信号是早期SV40多腺苷酸化信号 (SV40pA)。

[0535] 合适地, 多腺苷酸化信号可以包含以下或由以下组成:如SEQ ID NO:30所示的多核苷酸序列, 或与SEQ ID NO:30至少70%相同的变体。合适地, 变体与SEQ ID NO:30可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0536] 示例性SV40pA信号序列 (SEQ ID NO:30):

[0537] aacttgttttattgcagcttataatggttacaataaagcaatagcatcacaatttcacaataaagc  
atTTTTTcac tgcattctagttgtggtttgtccaaactcatcaatgtatcttatcatgtctggatc

[0538] 合适地, 多腺苷酸化信号可以包含以下或由以下组成:如SEQ ID NO:31所示的多核苷酸序列, 或与SEQ ID NO:31至少70%相同的变体。合适地, 变体与SEQ ID NO:31可以是至少75%、至少80%、至少85%、至少90%、至少95%至少96%、至少97%、至少98%或至少99%相同的。

[0539] 示例性bGHpoly (A) 信号序列 (SEQ ID NO:31):

[0540] ctgtgccttctagttgccagccatctgttgtttgcccctccccgtgccttcttgaccctggaaggt  
gccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctgg  
ggggtgggggtggggcaggacagcaaggggaggattgggaagacaatagcaggcatgctgggatgcggtgggctc  
tatgg

[0541] 合适地, 多腺苷酸化信号可以包含以下或由以下组成:如SEQ ID NO:32所示的多

核苷酸序列,或与SEQ ID NO:32至少70%相同的变体。合适地,变体与SEQ ID NO:32可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0542] 示例性可溶性神经纤毛蛋白1多腺苷酸化信号 (SEQ ID NO:32):

[0543] aaataaaatacgaatg

[0544] 合适地,多腺苷酸化信号可以包含以下或由以下组成:如SEQ ID NO:42所示的多核苷酸序列,或与SEQ ID NO:42至少70%相同的变体。合适地,变体与SEQ ID NO:42可以是至少75%、至少80%、至少85%、至少90%、至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0545] 示例性鸡 $\beta$ -珠蛋白多腺苷酸化信号 (SEQ ID NO:42)

[0546] caataaaagatctttattttcattagatctgtgtgttggttttttgtgtg

[0547] 载体

[0548] 本发明提供了包含本发明的多核苷酸的载体。

[0549] 载体是允许或促进从一个环境转移实体到另一个环境的工具。载体的四种主要类型是质粒、病毒载体、黏粒和人工染色体。

[0550] 优选地,本发明的载体是病毒载体。本发明的载体优选是腺相关病毒 (AAV) 载体,但考虑了可以使用的其他病毒载体。

[0551] 本发明的载体可以是以病毒载体颗粒的形式。优选地,本发明的病毒载体是以AAV载体颗粒的形式。

[0552] 制备和修饰病毒载体和病毒载体颗粒 (诸如衍生自AAV的那些) 的方法也是本领域众所周知的。Ayuso, E. 等人, 2010. *Current gene therapy*, 10 (6), pp. 423-436, Merten, O. W. 等人, 2016. *Molecular Therapy-Methods&Clinical Development*, 3, p. 16017; 以及 Nadeau, I. 和 Kamen, A., 2003. *Biotechnology advances*, 20 (7-8), pp. 475-489 中描述了合适的方法。

[0553] 本发明的载体能够转导肾细胞。在一些实施方案中,本发明的载体能够特异性转导肾细胞。

[0554] 本发明的载体优选能够转导肾小球细胞 (例如足细胞)。在一些实施方案中,本发明的载体能够特异性转导肾小球细胞 (例如足细胞)。本发明的载体优选能够转导足细胞。在一些实施方案中,本发明的载体能够特异性转导足细胞。

[0555] 腺相关病毒 (AAV) 载体

[0556] 本发明的载体可以是腺相关病毒 (AAV) 载体。本发明的载体可以是以AAV载体颗粒的形式。

[0557] AAV基因组

[0558] AAV载体或AAV载体颗粒可以包含AAV基因组或其片段或衍生物。

[0559] AAV基因组是多核苷酸序列,其可以编码产生AAV颗粒所需的功能。这些功能包括在宿主细胞中AAV的复制和包装周期中发挥作用的功能,包括将AAV基因组衣壳化成AAV颗粒。天然存在的AAV是复制缺陷的,并且依赖于反式辅助功能的提供来完成复制和包装循环。因此,本发明的AAV载体的AAV基因组通常是复制缺陷的。

[0560] AAV基因组可以是以单链形式 (ssAAV) (正义链形式或者反义链形式),或替代性地

以双链形式(dsAAV)。双链形式的使用允许绕过靶细胞中的DNA复制步骤,因此可以加速转基因表达。单链形式的最大包装容量大于双链形式。合适地,AAV基因组是以单链形式。

[0561] 自然界中存在的AAV可以根据不同的生物系统进行分类。AAV基因组可以来自任何天然衍生的AAV血清型、分离株或进化枝。

[0562] AAV可以以其血清型指代。血清型对应于AAV的变体亚种,由于其衣壳表面抗原的表达概况,其具有可用于将其与其他变体亚种间进行区分的独特的反应性。通常,具有特定AAV血清型的AAV载体颗粒不与对任何其他AAV血清型特异的中和抗体高效地交叉反应。AAV血清型包括AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10和AAV11。在一些实施方案中,本发明的AAV载体可以是AAV3B、LK03、AAV9或AAV8血清型。在一些实施方案中,本发明的AAV载体可以是AAV3B、LK03或AAV9血清型。

[0563] AAV可以指的是进化枝或克隆。这指天然衍生的AAV的系统发育关系,通常指可以追溯到共同祖先的AAV系统发育群,并且包括其所有后代。另外,AAV可以指的是特定的分离株,即自然界中发现的特定AAV的遗传分离株。术语遗传分离株描述了AAV群体,其与其他天然存在的AAV已进行了有限的基因混合,从而在基因水平上定义了可识别的有区别的群体。

[0564] 通常,天然衍生的血清型、分离株或进化枝的AAV基因组包含至少一个末端反向重复序列(ITR)。ITR序列以顺式发挥作用以提供功能性复制起点,并允许从细胞基因组整合和切除载体。ITR可以是治疗性基因旁需要的唯一顺式序列。

[0565] AAV基因组还可以包含包装基因,诸如编码AAV颗粒的包装功能的rep和/或cap基因。启动子可以可操作地连接至每个包装基因。此类启动子的具体实例包括p5、p19和p40启动子。例如,p5和p19启动子一般用于表达rep基因,而p40启动子一般用于表达cap基因。rep基因编码蛋白质Rep78、Rep68、Rep52和Rep40或其变体中的一种或多种。cap基因编码一种或多种衣壳蛋白,诸如VP1、VP2和VP3或其变体。这些蛋白质构成决定AAV血清型的AAV颗粒的衣壳。VP1、VP2和VP3可以通过替代性mRNA剪接产生(Trempe, J.P.和Carter, B.J., 1988. *Journal of virology*, 62 (9), pp. 3356-3363)。因此,VP1、VP2和VP3可以具有相同的序列,但其中VP2相对于VP1在N末端处截短,并且VP3相对于VP2在N末端处截短。

[0566] AAV基因组可以是天然存在的AAV的完整基因组。例如,包含完整AAV基因组的载体可以用于制备AAV载体或载体颗粒。

[0567] 优选地,为了施用于患者的目的,对AAV基因组进行衍生化。此类衍生化是本领域标准的,并且本发明涵盖AAV基因组的任何已知衍生物,以及可以通过应用本领域已知的技术生成的任何衍生物的使用。AAV基因组可以是任何天然存在的AAV的衍生物。合适地,AAV基因组是AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10或AAV11的衍生物。合适地,AAV基因组是AAV2的衍生物。

[0568] AAV基因组的衍生物包括允许从本发明的AAV载体体内表达转基因的AAV基因组的任何截短形式或修饰形式。通常,可能显著截短AAV基因组以包含最小病毒序列,但仍保留上述功能。出于安全原因,这是优选的,以降低载体与野生型病毒重组的风险,并且还避免由于靶细胞中存在病毒基因蛋白而引发细胞免疫应答。

[0569] 通常,衍生物将包含至少一个末端反向重复序列(ITR),优选多于一个ITR,诸如两个或更多个ITR。一个或多个ITR可以衍生自具有不同血清型的AAV基因组,或者可以是嵌合的或突变的ITR。优选的突变体ITR是缺失了trs(末端解离位点,terminal resolution

site)的突变体ITR。这种缺失允许基因组继续复制以生成包含编码序列和互补序列的单链基因组,即自互补AAV(scAAV)基因组。这允许绕过靶细胞中的DNA复制,从而能够加速转基因表达。然而,scAAV的最大包装容量是减少的。合适地,AAV基因组不是scAAV基因组。

[0570] AAV基因组可包含来自AAV或其变体的任何天然衍生的血清型、分离株或进化枝的一个或多个ITR序列。AAV基因组可以包含至少一种,诸如两种AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10或AAV11 ITR或其变体。合适地,AAV基因组可包含至少一种,诸如两种AAV2 ITR。

[0571] 优选包含一个或多个ITR以帮助AAV载体在宿主细胞核中形成多联体,例如在通过宿主细胞DNA聚合酶的作用将单链载体DNA转换成双链DNA之后。这种附加体多联体的形成在宿主细胞的生命周期中保护AAV载体,从而允许转基因在体内延长表达。

[0572] 合适地,在衍生物中,ITR元件将是天然AAV基因组中保留的唯一序列。衍生物优选不包括天然基因组的rep和/或cap基因和天然基因组的任何其他序列。由于上述原因,以及减少载体整合到宿主细胞基因组中的可能性,这是优选的。另外,减小AAV基因组的大小允许提高除了转基因之外,在载体之内并入其他序列元件(诸如调控元件)的灵活性。

[0573] 因此,在本发明的衍生物中可以去除以下部分:一个末端反向重复(ITR)序列、复制(rep)基因和衣壳(cap)基因。然而,另外,衍生物可以包含AAV基因组的rep和/或cap基因或其他病毒序列中的一个或多个。天然存在的AAV在人19号染色体上的特定位点以高频率整合,并且显示出可忽略不计的随机整合的频率,使得在治疗环境中可以容忍AAV载体中的整合能力的保留。

[0574] 另外,本发明涵盖提供以与天然AAV基因组的顺序和构型不同的顺序和构型的AAV基因组的序列。本发明还涵盖用来自另一种病毒的序列或由来自多于一种病毒的序列组成的嵌合基因置换一个或多个AAV序列或基因。此类嵌合基因可以由来自不同病毒种的两个或更多个相关的病毒蛋白的序列组成。

[0575] AAV血清型和衣壳蛋白

[0576] AAV载体颗粒可以由衣壳蛋白衣壳化。血清型可以促进肾小球细胞(例如足细胞)的转导,例如肾小球细胞(例如足细胞)的特异性转导。

[0577] AAV载体颗粒可以是肾特异性载体颗粒。优选地,AAV载体颗粒是肾小球特异性(例如足细胞特异性)载体颗粒。AAV载体颗粒可以由肾小球特异性(例如足细胞特异性)衣壳衣壳化。AAV载体颗粒可以包含肾小球特异性(例如足细胞特异性)衣壳蛋白。

[0578] 合适地,AAV载体颗粒可以是转衣壳形式,其中具有一种血清型的ITR的AAV基因组或衍生物包装在不同血清型的衣壳中。AAV载体颗粒还包括嵌合体形式,其中来自两种或更多种不同血清型的未修饰的衣壳蛋白的混合物构成病毒衣壳。AAV载体颗粒还包括带有吸附至衣壳表面的配体的化学修饰形式。例如,此类配体可包括用于靶向特定细胞表面受体的抗体。

[0579] 当衍生物包含衣壳蛋白,即VP1、VP2和/或VP3时,该衍生物可以是一种或多种天然存在的AAV的嵌合的、改组的(shuffled)或衣壳修饰的衍生物。具体地,本发明涵盖在同一载体(即假型载体)之内提供来自AAV的不同血清型、进化枝、克隆或分离株的衣壳蛋白序列。AAV载体可以是以假型AAV载体颗粒的形式。

[0580] 通常,选择嵌合的、改组的或衣壳修饰的衍生物来为AAV载体提供一种或多种期望

的功能。因此,与包含天然存在的AAV基因组的AAV载体相比,这些衍生物可以表现出增加的基因递送效率、降低的免疫原性(体液的或细胞的)、改变的嗜性(tropism)范围和/或改善的足细胞靶向性。基因的递送效率的提高可以通过改善细胞表面的受体或共受体结合、改善内化、改善细胞之内的运输和进入细胞核的运输、改善病毒颗粒的脱壳以及改善单链基因组向双链形式的转换来实现。提高效率也可以与改变的嗜性范围或足细胞的靶向有关,使得载体剂量不会因施用于不需要的组织而被稀释。

[0581] 嵌合衣壳蛋白包括通过在天然存在的AAV血清型的两个或更多个衣壳编码序列之间重组产生的那些。这可以例如通过标记获救方法来进行,其中一种血清型的非感染性衣壳序列与不同血清型的衣壳序列共转染,并且使用定向选择来选择具有期望特性的衣壳序列。不同血清型的衣壳序列可以通过细胞之内的同源重组来改变,以产生新的嵌合衣壳蛋白。

[0582] 嵌合衣壳蛋白还包括通过工程化衣壳蛋白序列以在两个或更多个衣壳蛋白之间,例如在不同血清型的两个或更多个衣壳蛋白之间转移特定衣壳蛋白结构域、表面环或特定氨基酸残基而生成的那些。

[0583] 改组衣壳蛋白或嵌合衣壳蛋白还可以通过DNA改组或易错PCR生成。杂交AAV衣壳基因可以通过随机片段化相关AAV基因(例如编码多种不同血清型衣壳蛋白的基因)的序列,和然后随后在自引发的聚合酶反应中重新组装片段来创建,这也可以导致序列同源性区中的交叉。以这种方式,可以筛选通过改组几种血清型的衣壳基因创建的杂交AAV基因文库,以鉴定具有期望功能的病毒克隆。类似地,易错PCR可用于随机突变AAV衣壳基因,以创建不同的变体文库,然后可以针对期望的特性对其进行选择。

[0584] 还可以对衣壳基因的序列进行遗传修饰以引入相对于天然野生型序列的特异性缺失、取代或插入。具体地,衣壳基因可以通过在衣壳编码序列的开放阅读框之内或在衣壳编码序列的N末端和/或C末端处插入不相关的蛋白质或肽的序列来修饰。不相关的蛋白质或肽可以有利地是充当特定细胞类型的配体的蛋白质或肽,从而赋予与靶细胞的改善的结合或改善载体的靶向至特定细胞群的特异性。不相关的蛋白质还可以是帮助作为产生过程的一部分的病毒颗粒的纯化的蛋白质,即表位或亲和标签。通常选择插入位点以便不干扰病毒颗粒的其他功能,例如病毒颗粒的内化、运输。

[0585] 衣壳蛋白可以是人工衣壳蛋白或突变衣壳蛋白。如本文所用的术语“人工衣壳”意指衣壳颗粒包含自然界中不存在的氨基酸序列或包含已从天然存在的衣壳氨基酸序列工程化(例如修饰)的氨基酸序列。换句话说,在人工衣壳氨基酸序列与亲本衣壳氨基酸序列比对时,与衍生人工衣壳蛋白的亲本衣壳的序列相比,该人工衣壳蛋白在氨基酸序列中包含突变或变异。

[0586] 衣壳蛋白可以包含相对于野生型衣壳蛋白的突变或修饰,其相对于未修饰的病毒颗粒或野生型病毒颗粒提高了转导足细胞的能力。改善的转导足细胞的能力可以例如通过测量AAV载体颗粒携带的转基因(例如GFP)的表达来测量,其中足细胞中的转基因的表达与AAV载体颗粒转导足细胞的能力相关。

[0587] AAV载体颗粒可以是AAV3B、LK03、AAV9或AAV8载体颗粒。本发明人已经证明具有AAV3B、LK03、AAV9和AAV8血清型的AAV载体颗粒可以转导足细胞。优选地,AAV载体颗粒是AAV3B载体颗粒或LK03载体颗粒。更优选地,AAV载体颗粒是AAV3B载体颗粒。

[0588] AAV载体颗粒可以包含AAV3B、LK03、AAV9或AAV8衣壳蛋白。优选地,AAV载体颗粒包含AAV3B衣壳蛋白或LK03衣壳蛋白。更优选地,AAV载体颗粒包含AAV3B衣壳蛋白。

[0589] AAV载体颗粒可以包含AAV3B、LK03、AAV9或AAV8衣壳蛋白VP1、VP2和VP3。优选地,AAV载体颗粒包含AAV3B或LK03衣壳蛋白VP1、VP2和VP3。更优选地,AAV载体颗粒包含AAV3B衣壳蛋白VP1、VP2和VP3。

[0590] AAV载体颗粒可以包含一个或多个AAV2 ITR序列和AAV3B衣壳蛋白、LK03衣壳蛋白、AAV9衣壳蛋白或AAV8衣壳蛋白。优选地,AAV载体颗粒包含一个或多个AAV2 ITR序列和AAV3B或LK03衣壳蛋白。更优选地,AAV载体颗粒包含一个或多个AAV2 ITR序列和AAV3B衣壳蛋白。

[0591] AAV载体颗粒可以具有AAV2基因组和AAV3B衣壳蛋白(AAV2/3B)、AAV2基因组和LK03衣壳蛋白、AAV2基因组和AAV9衣壳蛋白(AAV2/9)、或AAV2基因组和AAV8衣壳蛋白(AAV2/8)。优选地,AAV载体颗粒包含AAV2基因组和AAV3B或LK03衣壳蛋白。更优选地,AAV载体颗粒包含AAV2基因组和AAV3B衣壳蛋白。

[0592] 命名法AAVX/Y可以表示假型AAV,例如其中ITR序列来自AAVX并且位于携带被衣壳化成血清型AAVY(即具有AAVY衣壳蛋白)的有效载荷的盒的翼侧。

[0593] AAV3B血清型

[0594] AAV载体颗粒可以包含AAV3B衣壳蛋白。合适地,AAV载体颗粒可以由AAV3B衣壳蛋白衣壳化。

[0595] 已克隆出两个不同的AAV3分离株(AAV3A和AAV3B)。与基于其他AAV血清型的载体相比,认为AAV3载体不能有效地转导大多数细胞类型。然而,AAV3B可以有效地转导足细胞。AA3B增强子已在Rutledge, E. A. 等人, 1998. 1998. Journal of virology, 72(1), pp. 309-319 中描述。

[0596] AAV载体颗粒可以包含AAV3B VP1衣壳蛋白、AAV3B VP2衣壳蛋白和/或AAV3B VP3衣壳蛋白。合适地,AAV载体颗粒可以由AAV3B VP1衣壳蛋白、AAV3B VP2衣壳蛋白和/或AAV3B VP3衣壳蛋白衣壳化。合适地,AAV载体颗粒可以由AAV3B VP1、VP2和VP3衣壳蛋白衣壳化。

[0597] 合适地,AAV3B VP1衣壳蛋白可以包含以下或由以下组成:如SEQ ID NO:33所示的氨基酸序列,或与SEQ ID NO:33至少90%相同的变体。

[0598] 示例性AAV3B VP1衣壳蛋白(SEQ ID NO:33):

[0599] MAADGYLPDWLEDNLSEGIREWWALKPGVPQPKANQQHQDNRRGLVLPGYKYLPGNGLDKGEVNEAD  
AAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLQEDTSFGGNLGRAVFQAKKRILEPLGLVEEAAKTAPGKKRPV  
DQSPQEPDSSSGVGKSGKQPARKRLNFGQTGDSESVDPDQPLGEPPAAPTSLGSNTMASGGGAPMADNNEGADGVGN  
SSGNWHCDSQWLGDREVITTTSTRTWALPTYNNHLYKQISSQSGASNDNHYFGYSTPWGYFDNRFHCHFSPRDWQRLI  
NNNWGFRPKKLSFKLFNIQVKEVTQNDGTTTIANLNTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPQYGY  
LTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYLNRQTGTTSG  
TTNQSRLIFSQAGPQMSLQARNWLPGPCYRQRLSKTANDNNSNFPWTAASKYHLNGRDSLNVNPGPAMASHKDD  
EKFFPMHGNIIFGKEGTTASNAELDNVMI TDEEE IRTTNPVATEQYGTVANLQSSNTAPTTRTVNDQGALPGMVWQ  
DRDVYLQGPWAKIPHTDGHFHPSPMLGGFGLKHPQQIMIKNTPVPANPPTTFSPAKFASFITQYSTGQVSVEIEW  
ELQKENSKRWNPEIQYTSNYNKSXNVDFVDTNGVYSEPRPIGTRYLTRN

[0600] 合适地,变体与SEQ ID NO:33可以是至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0601] 合适地,AAV3B VP2和VP3衣壳蛋白可以是SEQ ID NO:33的N末端截短物、或与SEQ ID NO:33至少90%相同、至少95%、至少96%、至少97%、至少98%或至少99%相同的变体的N端截短物。

[0602] LK03血清型

[0603] AAV载体颗粒可以包含LK03衣壳蛋白。合适地,AAV载体颗粒可以由LK03衣壳蛋白衣壳化。

[0604] AAV-LK03 cap序列由来自七种不同野生型血清型(AAV1、2、3B、4、6、8、9)的片段组成,并在Lisowski,L.,等人,2014.Nature,506(7488),pp.382-386中描述。本发明人已经证明AAV-LK03载体可以在人足细胞中实现接近100%的体外高转导。

[0605] AAV载体颗粒可以包含LK03 VP1衣壳蛋白、LK03 VP2衣壳蛋白和/或LK03 VP3衣壳蛋白。合适地,AAV载体颗粒可以由LK03 VP1衣壳蛋白、LK03 VP2衣壳蛋白和/或LK03 VP3衣壳蛋白衣壳化。合适地,AAV载体颗粒可以由LK03 VP1、VP2和VP3衣壳蛋白衣壳化。

[0606] 合适地,LK03 VP1衣壳蛋白可以包含以下或由以下组成:如SEQ ID NO:34所示的氨基酸序列,或与SEQ ID NO:34至少90%相同的变体。

[0607] 示例性LK03 VP1衣壳蛋白(SEQ ID NO:34):

[0608] MAADGYLPDWLEDNLSEGIREWALQPGAPKPKANQQHQDNARGLVLPGYKYLPGNGLDKGEPVNAAD  
AAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKEDTSFGGNLGRAVFQAKKRILLEPLGLVEEAAKTAPGKKRPV  
DQSPQEPDSSSGVGKSGKQPARKRLNFGQTGDSSEVPDPQPLGEPPAAPTSLGSNTMASGGGAPMADNNEGADGVGN  
SSGNWHCDSQWLGRVITTTSTRTWALPTYNNHLYKQISSQSGASNDNHYFGYSTPWGYFDFNRFHCHFSPRDWQRLI  
NNNWGFRPKKLSFKLFNIQVKEVTQNDGTTTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPQYGY  
LTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYLNRQTGTTSG  
TTNQSRLFLFSQAGPQSMQLQARNWLPGPCYRQQRLSKTANDNNNSNFPWTAASKYHLNGRDLSLVNPGPAMASHKDD  
EKFFPMHGNLIFGKEGTTASNAELDNVMITDEEEIRTTNPVATEQYGTVANLQSSNTAPTTRTVNDQGALPGMVWQ  
DRDVYLQGPWIWAKIPHTDGHFHPSPMLGGFGLKHPPPQIMIKNTPVPANPPTTFSPAKFASFITQYSTGQVSVEIEW  
ELQKENSKRWNPEIQYTSNYSVNVDFTVDTNGVYSEPRPIGTRYLTRPL

[0609] 合适地,变体与SEQ ID NO:34可以是至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0610] 合适地,LK03 VP2和VP3衣壳蛋白可以是SEQ ID NO:34的N末端截短物、或与SEQ ID NO:34至少90%相同、至少95%、至少96%、至少97%、至少98%或至少99%相同的变体的N端截短物。

[0611] AAV9血清型

[0612] AAV载体颗粒可以包含AAV9衣壳蛋白。合适地,AAV载体颗粒可以由AAV9衣壳蛋白衣壳化。

[0613] 本发明人已经证明AAV9载体可以在人足细胞中实现体外高转导。

[0614] AAV载体颗粒可以包含AAV9 VP1衣壳蛋白、AAV9 VP2衣壳蛋白和/或AAV9 VP3衣壳蛋白。合适地,AAV载体颗粒可以由AAV9 VP1衣壳蛋白、AAV9 VP2衣壳蛋白和/或AAV9 VP3衣壳蛋白衣壳化。合适地,AAV载体颗粒可以由AAV9 VP1、VP2和VP3衣壳蛋白衣壳化。

[0615] 合适地, AAV9 VP1衣壳蛋白可以包含以下或由以下组成:如SEQ ID NO:35所示的氨基酸序列,或与SEQ ID NO:35至少90%相同的变体。

[0616] 示例性AAV9 VP1衣壳蛋白(SEQ ID NO:35):

[0617] MAADGYLPDWLEDNLSEGIREWALKPGAPQPKANQQHQDNARGLVLPGYKYLPGNGLDKGEPVNAAD  
AAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKEDTSFGGNGRAVVFQAKKRLLEPLGLVEEAAKTAPGKKRPV  
EQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDTEVPDPQPIGEPPAAPSGVGSGLTMASGGGAPVADNNEGADGVGS  
SSGNWHCDSQWLGDREVITTTSTRTWALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQR  
LINNNWGFPRKRLNFKLFNIQVKEVTDNNGVKTIANLNTSTVQVFTDSYQLPYVLGSAHEGCLPPFPADVFMIPQY  
GYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLDRLMNPLIDQYLYYLSKTINGS  
GQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQNNSEFAWPGASSWALNGRNSLMNPGPAMASHKEGE  
DRFFPLSGSLIFGKQGTGRDNDADKVMITNEEEIKTTNPVATESYQVATNHQSAQAQAQTGWVQNGGILPGMVWQ  
DRDVYLQGPWAKIPHTDGNFHPSPMLGGFGMKHPPPQILIKNTPVPADPPTAFNKDKLNSFITQYSTGQVSVEIEW  
ELQKENSKRWNPEIQYTSNYYKSNVFAVNTGQVYSEPRPIGTRYLTRNL

[0618] 合适地,变体与SEQ ID NO:35可以是至少95%、至少96%、至少97%、至少98%或至少99%相同的。

[0619] 合适地, AAV9 VP2和VP3衣壳蛋白可以是SEQ ID NO:35的N末端截短物、或与SEQ ID NO:35至少90%相同、至少95%、至少96%、至少97%、至少98%或至少99%相同的变体的N端截短物。

[0620] 其他病毒载体

[0621] 逆转录病毒载体和慢病毒载体

[0622] 本发明的载体可以是逆转录病毒载体或慢病毒载体。本发明的载体可以是逆转录病毒载体颗粒或慢病毒载体颗粒。

[0623] 逆转录病毒载体可以衍生自或可衍生(derivable)自任何合适的逆转录病毒。已鉴定出大量不同的逆转录病毒。实例包括鼠白血病病毒(MLV)、人T细胞白血病病毒(HTLV)、小鼠乳腺肿瘤病毒(MMTV)、Rous肉瘤病毒(RSV)、Fujinami肉瘤病毒(FuSV)、Moloney鼠白血病病毒(Mo-MLV)、FBR鼠骨肉瘤病毒(FBRMSV)、Moloney鼠肉瘤病毒(Mo-MSV)、Abelson鼠白血病病毒(A-MLV)、禽骨髓细胞瘤(myelocytomatosis)病毒29(MC29)和禽成红细胞增多症(erythroblastosis)病毒(AEV)。

[0624] 逆转录病毒可大致分成两类:“简单”和“复合”。逆转录病毒甚至可以进一步分成七组。这些组的五个组代表具有致癌潜力的逆转录病毒。其余的两组是慢病毒和泡沫病毒(spumaviruse)。

[0625] 逆转录病毒基因组和慢病毒基因组的基本结构共有许多共同特征,诸如5' LTR和3' LTR。使基因组能够被包装的包装信号、引物结合位点、使能够整合到宿主细胞基因组中的整合位点,以及编码包装组分的gag、pol和env基因位于它们之间或之内——这些是病毒颗粒的组装所需要的。慢病毒具有其他特征,诸如HIV中的rev和RRE序列,其能够将整合的原病毒的RNA转录物从细胞核有效导出到受感染的靶细胞的细胞质。

[0626] 在原病毒中,称为长末端重复序列(LTR)的区位于这些基因两端处的翼侧。LTR负责原病毒整合和转录。LTR还充当增强子-启动子序列,并且可以控制病毒基因的表达。

[0627] LTR本身是相同的序列,可分为三个元件:U3、R和U5。U3衍生自RNA的3'末端特有的

序列。R衍生自RNA的两端重复的序列。U5衍生自RNA的5'末端特有的序列。不同逆转录病毒中这三个元件的大小可以差异很大。

[0628] 在缺陷的逆转录病毒载体基因组中,gag、pol和env可能不存在或不是功能性的。

[0629] 在典型的逆转录病毒载体中,复制所必需的一个或多个蛋白质编码区的至少一部分可以从病毒中去除。这使得病毒载体是复制缺陷的。病毒基因组的部分也可以通过编码候选调制部分的文库替换,该候选调制部分可操作地连接至载体基因组中的调控控制区和报告部分,以生成包含能够转导靶宿主细胞和/或将其基因组整合到宿主基因组中的候选调制部分的载体。

[0630] 慢病毒载体是较大组的逆转录病毒载体的一部分。简而言之,慢病毒可以分为灵长类组和非灵长类组。灵长类慢病毒的实例包括但不限于人免疫缺陷病毒(HIV),人获得性免疫缺陷综合症(AIDS)的病原体(causative agent);和猿猴免疫缺陷病毒(SIV)。非灵长类慢病毒的实例包括原型“慢病毒”visna/maedi病毒(VMV),以及相关的山羊关节炎脑炎病毒(CAEV)、马感染性贫血病毒(EIAV)和最近描述的猫免疫缺陷病毒(FIV)和牛免疫缺陷病毒(BIV)。

[0631] 慢病毒家族与逆转录病毒在以下中不同:慢病毒具有感染分裂细胞和非分裂细胞两者的能力。相比之下,其他逆转录病毒,诸如MLV,无法感染非分裂或缓慢分裂的细胞,诸如构成例如肌肉、脑、肺和肝组织的那些。

[0632] 如本文所用,慢病毒载体是包含可衍生自慢病毒的至少一个组分部分的载体。优选地,该组分部分涉及该载体感染细胞、表达基因或复制依据的生物机制。

[0633] 慢病毒载体可以是“灵长类”载体。慢病毒载体可以是“非灵长类”载体(即衍生自不主要感染灵长类、尤其是人类的病毒)。非灵长类慢病毒的实例可以是不天然感染灵长类的慢病毒科的任何成员。

[0634] 作为基于慢病毒的载体的实例,基于HIV-1的载体和基于HIV-2的载体如下所述。

[0635] HIV-1载体含有简单逆转录病毒中也发现的顺式作用元件。已证明,延伸至gag开放阅读框的序列对HIV-1的包装非常重要。因此,HIV-1载体通常含有gag的相关部分,其中翻译起始密码子已突变。此外,大多数HIV-1载体还含有env基因的一部分,其包括RRE。Rev与RRE结合,这允许将全长或单剪接mRNA从细胞核转运至细胞质。在没有Rev和/或RRE的情况下,全长HIV-1RNA在细胞核中积累。或者,来自某些简单逆转录病毒诸如Mason-Pfizer猴病毒的组成型转运元件可用于减轻对Rev和RRE的需要。从HIV-1LTR启动子的有效转录需要病毒蛋白Tat。

[0636] 大多数基于HIV-2的载体在结构上与HIV-1载体非常类似。与基于HIV-1的载体类似,HIV-2载体也需要RRE来有效运输全长或单剪接病毒RNA。

[0637] 优选地,本发明中使用的病毒载体具有最小病毒基因组。

[0638] 所谓“最小病毒基因组”应当理解已对病毒载体进行操作从而除去非必需元件并保留必需元件,以提供将感兴趣的核苷酸序列感染、转导和递送至靶宿主细胞所需的功能。该策略的进一步细节可以在W0 1998/017815中找到。

[0639] 优选地,用于在宿主细胞/包装细胞之内产生病毒基因组的质粒载体将具有足够的慢病毒遗传信息,以允许在包装组分存在的情况下,将RNA基因组包装成能够感染靶细胞的病毒颗粒,但不能独立复制以在最终靶细胞之内产生感染性病毒颗粒。优选地,载体缺乏

功能性gag-pol和/或env基因和/或复制必需的其他基因。

[0640] 然而,用于在宿主细胞/包装细胞之内产生病毒基因组的质粒载体还将包含与慢病毒基因组可操作地连接的转录调控控制序列,以指导宿主细胞/包装细胞中的基因组的转录。这些调控序列可以是与转录的病毒序列相关联的天然序列(即5' U3区),或者它们可以是异源启动子,诸如另一种病毒启动子(例如CMV启动子)。

[0641] 载体可以是自失活(SIN)载体,其中病毒增强子和启动子序列已缺失。SIN载体可以在体内生成并以与野生型载体相似的功效转导非分裂细胞。SIN原病毒中的长末端重复序列(LTR)的转录失活应该阻止通过复制型(replication-competent)病毒的动员。这还应该能够通过消除LTR的任何顺式作用效果来调控从内部启动子的基因表达。

[0642] 载体可以是整合缺陷的。例如,可以通过将载体与催化失活的整合酶(诸如催化位点上带有D64V突变的HIV整合酶)进行包装,或者通过从载体LTR中修改或缺失必需的att序列,或者通过以上的组合来产生整合缺陷型慢病毒载体(IDLV)。

[0643] 腺病毒载体

[0644] 本发明的载体可以是腺病毒载体。本发明的载体可以是腺病毒载体颗粒。

[0645] 腺病毒是不经过RNA中间体的双链线性DNA病毒。有超过50种不同人血清型的腺病毒,基于基因序列同源性分为6个亚组。腺病毒的天然靶标是呼吸上皮细胞和胃肠上皮细胞,通常仅引起轻微症状。血清型2和5(具有95%序列同源性)最常用于腺病毒载体系统,并且通常与年轻人中的上呼吸道感染相关联。

[0646] 腺病毒已用作基因治疗和异源基因表达的载体。大基因组(36kb)可以容纳最多达8kb的外源插入DNA,并且能够在互补细胞系中高效复制,以产生最高达 $10^{12}$ 的非常高的滴度。因此,腺病毒是研究原代非复制细胞中基因表达的最佳系统之一。

[0647] 来自腺病毒基因组的病毒基因或外源基因的表达不需要正在复制的细胞。腺病毒载体通过受体介导的内吞作用进入细胞。一旦进入细胞,腺病毒载体很少整合到宿主染色体中。相反,它们在宿主细胞核中作为线性基因组以附加体(独立于宿主基因组)发挥功能。因此,重组腺病毒的使用减轻了与随机整合到宿主基因组中相关联的问题。

[0648] 单纯疱疹病毒载体

[0649] 本发明的载体可以是单纯疱疹病毒载体。本发明的载体可以是单纯疱疹病毒载体颗粒。

[0650] 单纯疱疹病毒(HSV)是具有作为基因传递载体的良好特性的嗜神经DNA病毒。HSV具有高感染性,因此HSV载体是将外源遗传物质递送至细胞的高效媒介物。病毒复制很容易被立即早期基因中的无效突变所破坏,这些突变可以在体外进行反式互补,这能够直接生产非致病性载体的高滴度纯制剂。基因组很大(152kb),并且许多病毒基因对于体外复制来说是不必要的,这允许用大的转基因或多个转基因来置换它们。用野生型病毒的潜伏感染导致附加体病毒在宿主一生的持续时间内存在于感觉神经元核中。这些载体是非致病性的,不能再激活并长期存在。潜伏活性启动子复合物可以用于载体设计,以在神经系统中实现长期稳定的转基因表达。因为由病毒识别的细胞受体的广泛的表达模式,所以HSV载体转导广谱的组织。对涉及细胞进入过程的渐增的理解已经允许靶向HSV载体的嗜性。

[0651] 其他病毒载体

[0652] 其他合适的病毒载体包括Lundstrom, K., 2018. *Diseases*, 6 (2), p. 42中描述的那

些。

[0653] 本发明的载体可以是甲病毒载体。本发明的载体可以是甲病毒载体颗粒。本发明的载体可以是黄病毒载体。本发明的载体可以是黄病毒载体颗粒。

[0654] 自扩增ssRNA病毒包括具有正极性基因组的甲病毒(例如Semliki Forest病毒、辛德比斯病毒、委内瑞拉马脑炎病毒和M1)和黄病毒(例如Kunjin病毒、西尼罗病毒和登革热病毒)。甲病毒主要应用于癌症治疗的临床前基因疗法研究。甲病毒载体可以以裸RNA、分层质粒DNA载体和重组复制缺陷或复制良好颗粒的形式递送。

[0655] 本发明的载体可以是弹状病毒载体。本发明的载体可以是弹状病毒载体颗粒。本发明的载体可以是麻疹病毒载体。本发明的载体可以是麻疹病毒载体颗粒。

[0656] 弹状病毒(例如狂犬病病毒和水疱性口炎病毒)和麻疹病毒携带负链基因组。在弹状病毒中,重组水泡性口炎病毒(VSV)已应用于临床前基因疗法研究。已发现麻疹病毒(例如MV-Edm)许多基因疗法应用。

[0657] 本发明的载体可以是鸡新城疫病毒载体。本发明的载体可以是鸡新城疫病毒载体颗粒。

[0658] ssRNA副黏病毒鸡新城疫病毒(NDV)在肿瘤细胞中特异性复制,并且因此已经经常应用于癌症基因疗法。

[0659] 本发明的载体可以是痘病毒载体。本发明的载体可以是痘病毒载体颗粒。

[0660] 痘病毒的特征是其dsDNA基因组,其可以容纳超过30kb的外源DNA。已发现痘病毒作为基因疗法载体的多种应用。例如,已证明痘苗病毒载体具有治疗癌症的潜力。痘苗病毒是一种大囊膜痘病毒,其具有大约190kb的线性双链DNA基因组。痘苗病毒可以容纳最高达大约25kb的外源DNA,这也使其用于递送大基因。许多减毒痘苗病毒株是本领域已知的,其适合于基因疗法应用,例如MVA株和NYVAC株。

[0661] 本发明的载体可以是小核糖核酸病毒载体。本发明的载体可以是小核糖核酸病毒载体颗粒。

[0662] 小核糖核酸病毒是非囊膜ssRNA病毒。属于小核糖核酸病毒科的柯萨奇病毒已应用为溶瘤载体。

[0663] 变体、衍生物、类似物、同源物和片段

[0664] 除了本文提及的特定蛋白质和核苷酸之外,本发明还涵盖其变体、衍生物、同源物和片段。

[0665] 在本发明的上下文中,任何给定序列的“变体”是其中已经以此类方式修饰残基的特定序列(无论是氨基酸残基还是核酸残基),使得所讨论的多肽或多核苷酸保留至少一种其内源性功能的序列。变体序列可以通过天然存在的多肽或多核苷酸中存在的至少一个残基的添加、缺失、取代、修饰、置换和/或变异来获得。例如,变体启动子序列保留其所获得自的启动子序列的至少一定水平的活性和特异性。

[0666] 本文所用的与本发明的蛋白质或多肽相关的术语“衍生物”包括来自序列的和/或对序列的一个(或多个)氨基酸残基的任何取代、变异、修饰、置换、缺失和/或添加,条件是所得蛋白质或多肽保留至少一种其内源功能。

[0667] 通常,可以进行氨基酸取代,例如1个、2个或3个至10个或20个取代,条件是修饰的序列保留所需的活性或能力。氨基酸取代可以包括使用非天然存在的类似物。

[0668] 本发明所用蛋白质还可以具有氨基酸残基的缺失、插入或取代,其产生沉默变化并且导致功能上等同的蛋白质。只要保留内源性功能,可以基于残基的极性、电荷、溶解度、疏水性、亲水性和/或两亲性质中的相似性进行有意的氨基酸取代。例如,带负电荷的氨基酸包括天冬氨酸和谷氨酸;带正电荷的氨基酸包括赖氨酸和精氨酸;以及有具有相似亲水性值的不带电荷的极性头基团的氨基酸包括天冬酰胺、谷氨酰胺、丝氨酸、苏氨酸和酪氨酸。

[0669] 例如可以根据下表进行保守取代。第二列中同一组中且优选第三列中同一行中的氨基酸可以彼此取代:

[0670]	脂肪族	非极性	GAP
			ILV
		极性-不带电荷	CSTM
			NQ
		极性-带电荷	DE
			KRH
芳香族		FWY	

[0671] 本文所用的术语“同源物”意指与野生型氨基酸序列或野生型核苷酸序列具有一定同源性的变体。术语“同源性”可以等同于“同一性”。

[0672] 在本文中,认为同源序列包括与主题序列可以是至少50%、55%、65%、75%、85%或90%相同的、优选至少95%、96%或97%或98%或99%相同的氨基酸序列。通常,同源物将包含与主题氨基酸序列相同的活性位点等。尽管同源性也可以就相似性(即,具有相似化学特性/功能的氨基酸残基)而言来考虑,但在本发明的上下文中,优选就序列同一性而言来表达同源性。

[0673] 在本文中,认为同源序列包括与主题序列可以是至少50%、55%、65%、75%、85%或90%相同的、优选至少95%、96%或97%或98%或99%相同的核苷酸序列。尽管同源性也可以就相似性而言来考虑,但在本发明的上下文中,优选就序列同一性而言来表达同源性。

[0674] 优选地,指与本文详述的任一SEQ ID NO具有百分比同一性的序列指与所指的SEQ ID NO的全长上具有所述百分比同一性的序列。

[0675] 同源性比较可以通过肉眼进行,或更常见的是在容易获得的序列比较程序的帮助下进行。这些商业上可获得的计算机程序可以计算两个或更多个序列之间的百分比同源性或同一性。

[0676] 可以计算相对于连续序列的同源性百分比,即将一个序列与另一个序列进行比对,并且将一个序列中的每个氨基酸或核苷酸与另一个序列中的相应氨基酸或核苷酸直接比较,一次一个残基。这称为“无空位”比对。通常,这种无空位比对仅在相对于相对少数量的残基进行。

[0677] 尽管这是一种非常简单且一致的方法,但它没有考虑到,例如,在其他方面相同的一对序列中,氨基酸或核苷酸序列中的一个插入或缺失可能会导致随后的残基或密码子置

于比对之外,因此在进行全局比对时可能会导致百分比同源性大幅降低。因此,设计了大多数序列比对方法以产生考虑可能的插入和缺失,而不会过度惩罚总体同源性得分的最佳比对。这是通过在序列比对中插入“空位”以尝试最大化局部同源性来实现的。

[0678] 然而,这些更复杂的方法对比对中出现的每个空位分配“空位罚分”,使得对于相同数量的相同氨基酸或核苷酸,序列比对具有尽可能少的空位,这反映了两个比较序列之间更高的相关性,将比具有许多空位的序列达到更高的得分。通常使用“仿射空位成本”,其对空位的存在收取相对较高的成本,并对空位中的每个随后的残基收取较小的罚分。这是最常用的空位评分系统。高空位罚分当然会产生具有更少空位的优化的比对。大多数比对程序允许修改空位罚分。然而,当使用此类软件进行序列比较时,优选使用默认值。例如,当使用GCG Wisconsin Bestfit包时,氨基酸序列的对于空位的默认空位罚分为-12,并且对于每个延伸的默认空位罚分为-4。

[0679] 因此,计算最大百分比同源性首先需要产生最佳比对,这考虑到空位罚分。用于执行此类比对的合适计算机程序是GCG Wisconsin Bestfit包(University of Wisconsin, USA; Devereux等人(1984) *Nucleic Acids Research* 12:387)。可以进行序列比较的其他软件的实例包括但不限于BLAST包(参见Ausubel等人(1999)同上-Ch.18),FASTA(Atschul等人(1990) *J. Mol. Biol.* 403-410),EMBOSS Needle(Madeira,F.等人,2019. *Nucleic acids research*, 47 (W1), pp. W636-W641)和GENEWORKS比较工具套件。BLAST和FASTA两者都可用于离线和在线搜索(参见Ausubel等人(1999)同上,7-58至7-60页)。然而,对于某些应用程序,优选使用GCG Bestfit程序。另一种工具BLAST 2Sequences也可用于比较蛋白质序列和核苷酸序列(*FEMS Microbiol. Lett.* (1999) 174 (2):247-50; *FEMS Microbiol. Lett.* (1999) 177 (1):187-8)。

[0680] 尽管最终的百分比同源性可以就同一性而言来测量,但比对过程本身通常不是基于全有或全无对比较(all-or-nothing pair comparison)。相反,通常使用缩放的相似性得分矩阵,根据化学相似性或进化距离给每个成对比较分配评分。常用的此类矩阵的实例是BLOSUM62矩阵(BLAST程序套件的默认矩阵)。GCG Wisconsin程序通常使用公共默认值或者自定义符号比较表(如果提供)(有关进一步细节参见用户手册)。对于一些应用程序,优选使用GCG包的公共默认值,或在其他软件的情况下,使用默认矩阵,诸如BLOSUM62。

[0681] 一旦软件已产生了最佳比对,就可以计算百分比同源性,优选是百分比序列同一性。软件通常将此作为序列比较的一部分并生成数值结果。百分比序列同一性可以计算为相同的残基的数目占所指的SEQ ID NO中总残基的百分比。

[0682] “片段”也是变体,并且该术语通常指在功能上或例如在测定中感兴趣的多肽或多核苷酸的选定区域。因此,“片段”指是全长多肽或多核苷酸的一部分的氨基酸序列或核酸序列。

[0683] 此类变体、衍生物、同源物和片段可以使用标准重组DNA技术诸如定点诱变来制备。当要进行插入时,可以制备编码该插入以及对应于插入位点两侧的天然存在序列的5'侧翼区和3'侧翼区的合成DNA。侧翼区将含有对应于天然存在的序列中的位点的方便的限制性位点,使得可以用适当的酶切割该序列并将合成DNA连接到切口中。然后根据本发明表达DNA以制备编码的蛋白质。这些方法仅是本领域已知的用于操作DNA序列的多种标准技术的说明性的,并且也可以使用其他已知的技术。

**[0684] 细胞**

[0685] 在一方面,本发明提供了细胞,其包含本发明的多核苷酸或载体。细胞可以是分离的细胞。细胞可以是人细胞,合适地是分离的人细胞。

[0686] 包含本发明中使用的多核苷酸的载体可以使用本领域已知的多种技术,诸如转染、转导和转化引入到细胞中。合适地,本发明的载体通过转染或转导引入到细胞中。

[0687] 细胞可以是现有技术中已知的任何细胞类型。

[0688] 合适地,细胞可以是生产细胞。术语“生产细胞”包括在瞬时转染、稳定转染或载体转导产生病毒颗粒所需的所有元件之后产生病毒颗粒的细胞,或工程化以稳定地包含产生病毒颗粒所需的元件的任何细胞。合适的生产细胞是本领域技术人员众所周知的。合适的生产细胞系包括HEK293(例如HEK 293T)细胞系、HeLa细胞系和A549细胞系。

[0689] 合适地,细胞可以是包装细胞。术语“包装细胞”包括含有包装感染性重组病毒所需的一些或全部元件的细胞。包装细胞可以缺乏重组病毒载体基因组。通常,此类包装细胞含有能够表达病毒结构蛋白的一种或多种载体。通过随后的瞬时转染、转导或稳定整合每个额外所需元件的步骤,仅包含产生囊膜病毒颗粒所需的一些元件的细胞,可用作生成病毒颗粒生产细胞系中的中间试剂。通过术语“包装细胞”涵盖这些中间试剂。合适的包装细胞是本领域技术人员众所周知的。

[0690] 合适地,细胞可以是肾细胞或肾小球细胞,例如足细胞。合适地,细胞可以是永生化肾细胞或永生化肾小球细胞,例如永生化足细胞。合适的足细胞细胞系将是本领域技术人员众所周知的,例如CIHP-1。生成永生化足细胞的方法对于本领域技术人员会是众所周知的。Ni, L. 等人, 2012. *Nephrology*, 17(6), pp. 525-531中描述了合适的方法。

**[0691] 药物组合物**

[0692] 在一方面,本发明提供了药物组合物,其包含本发明的多核苷酸或载体或本发明的细胞。

[0693] 药物组合物是包含治疗有效量的药学上的活性剂(即载体)或由治疗有效量的药学上的活性剂(即载体)组成的组合物。优选地,其包括药学上可接受的载剂、稀释剂或赋形剂(包括其组合)。

[0694] “药学上可接受的”包括配制剂是无菌的且无热原的。载剂、稀释剂和/或赋形剂在与载体相容并且对其接受者无害的意义上必须是“可接受的”。通常,载剂、稀释剂和赋形剂将是无菌且无热原的盐水或输注介质,然而,可以使用其他可接受的载剂、稀释剂和赋形剂。

[0695] 用于治疗用途的可接受的载剂、稀释剂和赋形剂是药学领域众所周知的。药物载剂、赋形剂或稀释剂的选择可以就预期的施用途径和标准药物实践而言来选择。药物组合物可包含作为载剂、赋形剂或稀释剂的或除了载剂、赋形剂或稀释剂之外的任何合适的黏合剂、润滑剂、悬浮剂、包衣剂或溶解剂。

[0696] 药学上可接受的载剂的实例包括例如水、盐溶液、乙醇、硅树脂、蜡、凡士林、植物油、聚乙二醇、丙二醇、脂质体、糖、明胶、乳糖、直链淀粉、硬脂酸镁、滑石、表面活性剂、硅酸、粘性石蜡、芳香油、脂肪酸单甘油酯和脂肪酸二甘油酯、石油醚脂肪酸酯、羟甲基纤维素、聚乙烯吡咯烷酮等。

[0697] 根据本发明的载体、细胞或药物组合物可以以适合于治疗和/或预防本文所述的

疾病的方式施用。施用的量和频率将由如受试者的病况以及受试者的疾病的类型和严重性的此类因素来确定,但是适当的剂量可以通过临床试验来确定。可以相应地配制药剂组合物。

[0698] 根据本发明的载体、细胞或药物组合物可以肠胃外施用,例如静脉内施用,或通过输注技术施用。载体、细胞或药物组合物可以以无菌水溶液的形式施用,该无菌水溶液可以含有其他物质,例如足够的盐或葡萄糖以使溶液与血液等渗。水溶液可以是适当缓冲的(优选3至9的pH)。可以相应地配制药剂组合物。通过本领域技术人员熟知的标准制药技术可以容易地在无菌条件下实现制备合适的肠胃外制剂。

[0699] 根据本发明的载体、细胞或药物组合物可以全身施用,例如通过静脉内注射施用。

[0700] 根据本发明的载体、细胞或药物组合物可以局部施用,例如通过靶向施用至肾。合适地,载体、细胞或药物组合物可以通过注射到肾动脉中或通过输尿管或被膜下注射来施用。

[0701] 药物组合物可以包含在输注介质例如无菌等渗溶液中的本发明的载体或细胞。药物组合物可以封装在由玻璃或塑料制成的安瓿、一次性注射器或多剂量小瓶中。

[0702] 载体、细胞或药物组合物可以以单剂量或多剂量施用。特别地,载体、细胞或药物组合物可以以单次、一次性剂量施用。可以相应地配制药剂组合物。

[0703] 载体、细胞或药物组合物可以以不同剂量施用(例如以载体基因组(vg)每kg测量)。无论如何,医生将确定最适合任何个体受试者的实际剂量,并且剂量将随着特定受试者的年龄、体重和反应而变化。然而,通常,对于本发明的AAV载体,可以施用 $10^{10}$ vg/kg至 $10^{14}$ vg/kg,或 $10^{11}$ vg/kg至 $10^{13}$ vg/kg的剂量。

[0704] 药物组合物可以进一步包含一种或多种其他治疗剂。

[0705] 本发明还包括包含本发明的载体、细胞和/或药物组合物的试剂盒的用途。优选地,试剂盒用于本文所述的方法中并如本文所述使用,例如本文所述的治疗方法。优选地,试剂盒包含试剂盒组分的使用说明书。

[0706] 治疗和/或预防疾病的方法

[0707] 在一方面,本发明提供了用作药物的根据本发明的载体、细胞或药物组合物。

[0708] 在一方面,本发明提供了根据本发明的载体、细胞或药物组合物在制备药物中的用途。

[0709] 在一方面,本发明提供了向有此需要的受试者施用根据本发明的载体、细胞或药物组合物的方法。

[0710] 肾小球疾病

[0711] 根据本发明的载体、细胞或药物组合物可以用于治疗受试者中的肾小球疾病。合适地,受试者是人受试者。

[0712] 在一方面,本发明提供了根据本发明的载体、细胞或药物组合物,其用于预防或治疗肾小球疾病。

[0713] 在相关的方面,本发明提供了根据本发明的载体、细胞或药物组合物在制备用于预防或治疗肾小球疾病的药物中的用途。

[0714] 在相关的方面,本发明提供了预防或治疗肾小球疾病的方法,其包括向有此需要的受试者施用根据本发明的载体、细胞或药物组合物。

[0715] 肾小球疾病可分为肾病性或者肾炎性。肾病综合征通常涉及影响足细胞-足细胞相互作用或足细胞-GBM相互作用的完整性的因素。相比之下,涉及肾炎综合征病因的因素可以有所不同,但可以包括循环血小板和白细胞、GBM、驻留肾小球内皮细胞和系膜细胞。(Chiang,C.K.和Inagi,R.,2010.Nature Reviews Nephrology,6(9),p.539)。

[0716] 合适地,肾小球疾病是遗传性肾小球疾病,即遗传的肾小球疾病。遗传性肾小球疾病包括足细胞相关联的遗传性肾小球疾病,诸如肾病综合征以及GBM相关联的肾小球疾病,诸如奥尔波特综合征。

[0717] 合适地,肾小球疾病是足细胞相关联的遗传性肾小球疾病。足细胞相关联的遗传性肾小球疾病包括芬兰型的先天性肾病综合征、先天性肾病综合征2型、家族性肾病综合征3型、弗雷泽综合征和德尼-德拉什综合征、Schimke免疫-骨发育不良、由CD2AP中的突变引起的肾病综合征、由肌动蛋白4中的突变引起的肾病综合征、由TRPC6中的突变引起的肾病综合征,以及Epstein和Fechtner综合征。合适地,肾小球疾病是肾病综合征。

[0718] 合适地,肾小球疾病是GBM相关联的遗传性肾小球疾病。足细胞相关联的遗传性肾小球疾病包括X连锁奥尔波特综合征、常染色体隐性奥尔波特综合征、常染色体显性奥尔波特综合征、薄基底膜疾病、Pierson综合征和指甲髌骨综合征。合适地,肾小球疾病是奥尔波特综合征(AS)。AS又称家族性肾炎、遗传性肾炎、薄基底膜病、薄基底膜肾病。

## 实施例

[0719] 现在将通过实施例进一步描述本发明,这些实施例意指在实施本发明中帮助本领域普通技术人员,并且不旨在以任何方式限制本发明的范围。

[0720] 实施例1-最小肾病蛋白启动子的设计、构建和测试

[0721] 最小肾病蛋白启动子的设计

[0722] Moeller等人2002J Am Soc Nephrol,13(6):1561-7和Wong MA等人2000Am J Physiol Renal Physiol,279(6):F1027-32中已经描述了人NPHS1启动子。这个NPHS1启动子是一个1.2kb片段,并且似乎是足细胞特异性的。这在下文中称为“FL”肾病蛋白启动子,并显示在图1A中。

[0723] 通过缺失N末端序列,FL肾病蛋白启动子最初被切割为822bp(除起始密码子外819bp)。这在下文中称为“midi”肾病蛋白启动子,并显示在图1B中。

[0724] 通过从中心区去除推定的通用转录结构域,将midi肾病蛋白启动子进一步切割为268bp(除起始密码子外265bp)。这在下文中称为“mini”肾病蛋白启动子,并显示在图1C中。

[0725] 载体构建体的构建

[0726] Midi肾病蛋白启动子

[0727] 如图2A中所示,将pACE\_hNPHS1启动子用作模板以引入BamHI和ClaI限制性位点。然后对片段进行凝胶提取并用ClaI和BamHI在37℃消化1小时,然后连接到pLenti GFPBlast载体中。连接物进一步转化至稳定的感受态大肠杆菌细胞中,提取DNA并测序DNA(Midi启动子)。最终的慢病毒载体如图2B中所示。

[0728] Mini肾病蛋白启动子

[0729] 如图3A中所示,pACE\_hNPHS1启动子用于PCR单链突出端(OH)。对含有OH的启动子的两部分进行凝胶提取,用于到pLenti GFP Blast载体中的NEBuilder HiFi组装反应。然

后使用DNA清理试剂盒清理连接反应,然后将其转化到稳定的感受态大肠杆菌细胞中。提取DNA并测序DNA。最终的慢病毒载体如图3B中所示。

#### [0730] 测试载体构建体

[0731] 最小肾病蛋白启动子用于在体外细胞模型中表达GFP以检验功效和足细胞特异性。

[0732] pLenti GFP Blast肾病蛋白启动子构建体(全长、Midi和Mini)用于转染HEK 293T细胞48小时以制备病毒,该病毒进一步用于创建稳定表达GFP标记的FLNPHS1、midi启动子或mini启动子的人条件永生化足细胞。

[0733] 用慢病毒载体转染条件永生化人足细胞(ciPodocyte),以确定最小启动子是否能够驱动GFP表达。显示了midi和mini肾病蛋白启动子驱动GFP表达。图4显示了代表性荧光显微镜图像,其显示了mini肾病蛋白启动子的GFP表达。图5显示了代表性Western印迹,其显示了mini肾病蛋白启动子的GFP表达。这些结果显示了,mini肾病蛋白启动子能够驱动足细胞中的转基因表达。

[0734] 慢病毒载体也用于转导人肾小球细胞。ciPodocyte和肾小球内皮细胞用包含与mini肾病蛋白报告基因偶联的GFP的慢病毒转导。图6A-C显示了FACS分析,该分析展示了使用Novocyte分析仪对条件永生化人足细胞(LY)和肾小球内皮细胞(GEnC)的所有活的单细胞的中值GFP荧光。将未转导的细胞(细胞对照)与用慢病毒构建体转导的细胞进行比较,该慢病毒构建体携带由全长人肾病蛋白启动子控制的GFP表达盒(hNPHS1.GFP)或mini人肾病蛋白启动子控制的GFP表达盒(265.GFP)。所有细胞均分化10天,胰蛋白酶消化(100 $\mu$ L)并在PBS、2%FBS、1:1000DRAQ7(150 $\mu$ L)中稀释。数据和误差线代表3次技术重复(100 $\mu$ L,>2500个细胞) $\pm$ SEM。这些结果显示,当与肾小球内皮细胞相比时,足细胞对最小肾病蛋白启动子具有特异性。

[0735] 实施例2-与COL4A3、COL4A4和COL4A5偶联的最小肾病蛋白启动子的设计、构建和测试

#### [0736] AAV构建体的设计和构建

[0737] 设计并构建了包含与mini肾病蛋白启动子(“265”)偶联的COL4A3、COL4A4和COL4A5的以下AAV转移质粒:

[0738] • pAAV.265.Col4a3.3flag.sv40, AAV质粒,其包含与mini肾病蛋白启动子偶联的COL4A3(参见图7A)。

[0739] • pAAV.265.Col4a4.3flag.sv40, AAV质粒,其包含与mini肾病蛋白启动子偶联的COL4A4(参见图7A)

[0740] • pAAV.265.Col4a5.3flag.sv40, AAV质粒,其包含与mini肾病蛋白启动子偶联的COL4A5(参见图7C)

[0741] 进行SmaI消化以确认质粒的身份(参见图7D-E)。这表明带有265bp mini肾病蛋白启动子和SV40聚腺苷酸尾的COL4 a3、a4和a5成功克隆到AAV中。

#### [0742] AAV构建体的测试

[0743] 使用标准方法制备以下AAV病毒载体:

[0744] • 具有LK03血清型的AAV.COL4A3.nephrin265.Sv40

[0745] • 具有LK03血清型的AAV.COL4A5.nephrin265.Sv40

[0746] • 具有2/9血清型的AAV.COL4A5.nephrin265.Sv40

[0747] 图8A显示了用抗FLAG抗体下拉的人已分化的ciPodocyte中全长带有FLAG标签的Col14a3 (LK03) 或Col14a5 (LK03) 的免疫沉淀实验。抗FLAG抗体沉淀Col14a3和Col14a5两者。人FLAG IgG用作对照。

[0748] 图8B显示了蛋白质裂解物的蛋白质印迹,其显示了人或小鼠分化的ciPodocyte中Col14a3 (LK03衣壳血清型)、Col14a5 (LK03) 和Col14a5 (2/9衣壳血清型) 的表达水平。未感染的人和小鼠的cipodocyte用作对照。

[0749] 图8C显示了共聚焦图像,其显示了人野生型ciPodocyte/Col14a5 3xFlag AAV ciPodocyte中转导的Col14a5与F-肌动蛋白的免疫荧光染色。与野生型对应物相比,在用Col14a5 3xFlag AAV病毒感染的人已分化的足细胞中,Col14a5以胞质水平存在。

[0750] 这些结果表明,我们已成功用与mini肾病蛋白启动子偶联的全长COL4a3和COL4a5转导人足细胞,并且mini肾病蛋白启动子驱动全长COL4a3或COL4a5在人足细胞中的表达。

[0751] 实施例3-在HEK细胞中表达CFH、CFI和CFHL1

[0752] 材料和方法

[0753] 在补充有10%FBS的DMEM中生长的60%-80%汇合的293T人胚肾细胞,用pHelper (HGT11)、两个pAAV Rep-Cap (LK03或AAV9) 之一和三个含有以下的ITR表达质粒进行三重转染:1) SEQ ID NO:47的265bp最小肾病蛋白启动子下的CFH (pAAV-265-CFH), 2) SEQ ID NO:14的全长最小肾病蛋白启动子下的CFI (pAAV-FL-CFI) 或3) SEQ ID NO:14的全长最小肾病蛋白启动子下的CFHL1 (pAAV-FL-CFHL1)。所有构建体均标记有MYC和FLAG。

[0754] 在聚乙烯亚胺 (PEI) 存在的条件下,在无血清培养基中,在150mm培养皿上进行转染。转染后第二天将培养基更换为含FBS的DMEM。在转染后第4天,分别收集和处理培养基和细胞。将培养基冷冻并储存于-80℃。用补充有蛋白酶抑制剂的RIPA缓冲液裂解细胞并储存于-80℃。

[0755] 使用Pierce BCA蛋白质测定测量细胞裂解物中的蛋白质浓度,并将来自每个样品的10μg总蛋白质加载到4%-15%聚丙烯酰胺Tris-甘氨酸凝胶上。将来自每个样品的总共2.6μL培养基加载到凝胶上。使用iBlot2干印迹系统将蛋白质转移到硝酸纤维素膜上。以下一抗用于蛋白质检测:抗H因子 (Abcam, 目录ab124769)、抗I因子 (Abcam, ab278524)、抗MYC标签 (CST, 目录2276S)、抗FLAG标签 (CST, 目录14793S) 和抗GAPDH (Millipore, 目录MAB374)。

[0756] 结果

[0757] H因子在来自用CFH表达质粒转染的293T HEK细胞的细胞裂解物和培养基中表达 (图9,泳道2、3、8、9)。使用H因子特异性抗体检测表达。在未转染细胞或用CFI表达质粒或CFHL1表达质粒转染的细胞的细胞裂解物或培养基中未检测到H因子 (图9,泳道1、4-7、10-13)。这些结果证明,265bp最小肾病蛋白启动子允许转基因在靶细胞中表达。

[0758] 实施例4-用AAV2/9 265-CFH转导或用编码CFH的质粒转染H因子突变的足细胞

[0759] 用AAV2/9 265-CFH转导H因子突变的足细胞

[0760] 将生长在补充有10%FBS和1%ITS的RPMI中的具有内源性H因子中的突变的条件永生化人足细胞 (Muehlig等人, 2020) 接种到6孔培养板中并在33℃下生长直至70%-80%汇合。将细胞与含有在265bp最小肾病蛋白启动子 (SEQ ID NO:47) 的控制下的CFH转基因的

AAV2/9一起孵育。不与病毒一起孵育的细胞用作非转导 (NT) 对照。在病毒转导后的同一天, 将细胞转移至37°C的非允许 (non-permissive) 温度, 以允许细胞分化和转基因表达。转导之后的随后几天更换培养基两次。在转导后第10天, 收集细胞培养基并使用抗H因子抗体 (Abcam, 目录ab252359) 通过ELISA测量人H因子的浓度。

[0761] 用含有265bp最小肾病蛋白启动子控制下的CFH转基因的AAV2/9病毒转导含有H因子中的突变的人足细胞, 表现出与未转导的细胞相比, 培养基中的人H因子浓度更高 (图10A和10B)。这证明由AAV递送并在265bp最小肾病蛋白启动子控制下的H因子转基因可以在人足细胞中表达。

[0762] 用编码CFH的质粒转染H因子突变的足细胞

[0763] 将生长在补充有10% FBS和1% ITS的RPMI中的具有内源性H因子中的突变的条件永生人足细胞 (Muehlig等人, 2020) 接种到6孔培养板中并在33°C下生长直至70%-80%汇合。对于质粒转染, 在聚乙烯亚胺存在的条件下, 将细胞与1.5 $\mu$ g的表达质粒的无血清培养基中培养。未添加质粒的细胞用作未转染 (NT) 对照。第二天将培养基改为含有FBS的培养基。在转染后第3天, 收集培养基并通过ELISA (Abcam, 目录ab252359) 进行分析。

[0764] 用表达265bp最小肾病蛋白启动子的控制下的CFH转基因的质粒转染的足细胞, 表现出比未转染对照更高的人因子H浓度 (图10C)。

[0765] 实施例5-在WT小鼠的肾中表达CFH

[0766] 材料和方法

[0767] 通过静脉注射尾静脉注射将100 $\mu$ l的AAV2/9基因疗法产品 (pAAV.NPHS1 (265).hCFH.WPRE.bGH) 或盐水施用至野生型C57BL6小鼠。AAV表达带有标签的在265bp最小肾病蛋白启动子 (SEQ ID NO:47) 的控制下的野生型人CFH转基因。3天后通过超速离心收获和纯化AAV, 并在PBS中滴定 (~1.5x10<sup>13</sup>/ml)。所有动物均在第21天完成研究并杀死。将肾在液氮中快速冷冻, 并按照制造商的方案通过RNeasy Micro试剂盒 (Qiagen目录号:174004) 用于RNA提取。然后使用High-Capacity RNA-to-cDNA™试剂盒 (4387406) 将RNA转化成cDNA, 然后进行qPCR分析。

[0768] 对来自pAAV\_CFH注射小鼠的肾的DNA样品进行定量qPCR。用SYBR green试剂和被动ROX方法的标准曲线qPCR用于检测肾DNA样品中的ITR存在。使用已知数量的ITR扩增子的标准曲线计算每pg DNA的病毒基因组。最后, 基于二倍体小鼠细胞具有6pg的DNA的假设计算每个细胞的病毒基因组。

[0769] 使用抗肾病蛋白抗体 (PROGEN) 和抗CFH抗体 (ab124767) 对5微米厚的冰冻肾切片进行免疫荧光染色。

[0770] 将组织包埋在OCT化合物 (VWR, 目录号361603E) 中并在液氮中快速冷冻。使用低温恒温器 (Thermo, CryostarNX270) 切割10 $\mu$ m切片。组织切片用4%多聚甲醛在室温下固定20分钟, 用0.3% Triton-X在室温下透化15分钟, 并用5% BSA在室温下封闭30分钟。

[0771] 将一抗 (H因子的兔单克隆 [EPR6226], Abcam目录号ab124769和肾病蛋白 (NPHS1) (1243-1256) 豚鼠多克隆抗体, Origene目录号BP5030) 用5% BSA稀释。(抗H因子为1:100, 抗肾病蛋白为1:300)。在室温下孵育1小时。

[0772] 二抗 (山羊抗兔IgG (H+L) 二抗, Alexa Fluor Plus 488, Invitrogen, 目录号A32731 和山羊抗豚鼠IgG (H+L) Alexa Fluor 568, Invitrogen, 目录号A-11075) 在PBS中1:500稀

释,并在室温下进行30分钟孵育。

[0773] 将载玻片安置在Fluoromount-G、载玻片安置介质(SouthernBiotech,目录号0100-01)中,并在Leica DM750荧光显微镜上以20X放大倍数成像。

[0774] 结果

[0775] 与用盐水对照注射的野生型小鼠相比,用含有265bp最小肾病蛋白启动子下的人CFH的AAV注射野生型小鼠,导致肾被AAV感染以及肾中病毒和人CFH的更高的表达(图11A-C)。免疫荧光染色证明小鼠肾小球中肾病蛋白和CFH的共定位(图11D)。

[0776] 实施例7-在肾细胞中使用265bp与818bp与FL最小肾病蛋白启动子(质粒转染)表达GFP

[0777] 构建的质粒包含在以下控制下的eGFP:(i) SEQ ID NO:14(PS0281)的全长(FL)最小肾病蛋白启动子;(ii) SEQ ID NO:59(PS0301)的818bp最小肾病蛋白启动子;(iii) SEQ ID NO:47(P20282)的265bp最小肾病蛋白启动子。

[0778] 将足细胞、肾小球内皮细胞(GENC)和近端管上皮细胞(PTEC)以每个六孔板孔 $1.5 \times 10^5$ 个细胞的密度接种,并允许其贴附过夜。对于HEK293T细胞,每孔使用 $5 \times 10^5$ 个细胞。第二天,将150 $\mu$ L DMEM中的3 $\mu$ g质粒DNA与9 $\mu$ L PEI(也在150 $\mu$ L DMEM中)混合,在室温下孵育15分钟,并在48小时后通过流式细胞术对eGFP表达进行分析。未转染的细胞用作阴性群和阳性群的门控的阴性对照。

[0779] 图12中的结果显示了,所有三种启动子都能够所有测试的细胞类型中驱动eGFP表达。这些结果证明,尽管去除了肾病蛋白启动子序列的较大区域,将全长肾病蛋白启动子最小化至818bp或者265bp仍然允许转基因在靶细胞中表达。

[0780] 实施例8-在肾细胞中使用265bp与FL最小肾病蛋白启动子(AAV转导)表达GFP

[0781] 材料和方法

[0782] AAV产生:用pHelper(HGTI1,PS0150)、Rep-Cap质粒(LK03,PS0240)和含有驱动eGFP表达的全长(FL)最小肾病蛋白启动子或265bp最小肾病蛋白启动子(分别为PS0281和PS0282)的两种ITR表达质粒之一,三重转染在补充有10%FBS的DMEM中生长的60%-80%汇合293T人胚肾细胞。在聚乙烯亚胺(PEI)存在的条件下,在150mm培养皿上进行转染。转染后第二天将培养基更换为含FBS的DMEM。在转染后第4天,分别收集和处培养基和细胞。将PEG添加到过滤的上清液中以沉淀病毒颗粒,随后将其在1500g下离心20分钟。将细胞沉淀重悬于TD缓冲液中并在-80度下冷冻,然后在37度下解冻并涡旋。冻融循环重复5次。然后将来自PEG处理过的上清液的沉淀添加到裂解的细胞中,并进一步添加Benzonase和脱氧胆酸钠并在37度下孵育30分钟。然后将样品离心,并过滤上清液,并随后经由碘克沙醇梯度超速离心纯化。

[0783] AAV质量控制:经由qPCR滴定AAV样品,以确定每mL病毒上清液中病毒基因组的数量。使用设计用于靶向eGFP序列的引物。对样品进行SDS-PAGE分析以确定纯度。使用碱性凝胶电泳以证明完整转基因盒并入到病毒颗粒中。

[0784] AAV转导:将细胞以 $1.5 \times 10^5$ 个细胞/孔的密度接种到6孔板的孔中,并允许其贴壁过夜。第二天,添加病毒颗粒,感染复数(MOI)为每个细胞 $5 \times 10^5$ 个颗粒。将细胞置于37度以起始分化并培养10天。10天后,通过流式细胞术分析细胞的eGFP表达。

[0785] 结果

[0786] 图13中的结果显示了,两种启动子都能够所有测试的细胞类型中驱动eGFP表达。265bp最小肾病蛋白启动子在每个测试的细胞中显示出较高的eGFP表达水平。

[0787] 这些数据再次证明265bp最小肾病蛋白启动子能够在靶细胞中驱动eGFP表达。另外,该启动子可以并入AAV载体中以用于基因治疗。使用265bp最小肾病蛋白启动子代替全长最小肾病蛋白启动子,可以允许将更多遗传物质并入到用于AAV目的的转基因盒中。

[0788] 实施例9-在HEK中使用265bp与FL最小肾病蛋白启动子(质粒转染)表达人足细胞素

[0789] 构建包含在以下控制下的带有HA标签的足细胞素的质粒:(i) SEQ ID NO:14 (FL.NPSH1-Podocin-HA)的全长(FL)最小肾病蛋白启动子;或(ii) SEQ ID NO:47 (265.hNPSH1-Podocin-HA)的265bp最小肾病蛋白启动子。

[0790] 将在补充有10%FBS、1%青霉素/链霉素和1%丙酮酸钠的DMEM中生长的293T人胚肾细胞以 $5 \times 10^5$ 的密度接种在6孔培养板中。第二天当细胞达到70%-80%汇合时进行转染。通过将3 $\mu$ g的表达质粒与聚乙烯亚胺(PEI)以1:3的比率在无血清培养基中复合来制备转染试剂混合物。未转染对照接受含有相同比率的PEI但不含任何表达质粒的转染混合物。将细胞与转染混合物在37 $^{\circ}$ C、5%CO<sub>2</sub>下孵育,直至将培养基更换为完全培养基的第二天。转染后第2天收获细胞,并通过用抗HA抗体对HA标签进行染色,通过FAC分析足细胞素的表达。

[0791] 图14A中的结果显示了,两种启动子都能够HEK细胞中驱动人足细胞素-HA表达。这证明,尽管去除了大多数的启动子序列,将全长最小肾病蛋白启动子最小化至265bp仍然允许转基因在靶细胞中表达。

[0792] 实施例10-在人足细胞中使用265bp与FL最小肾病蛋白启动子(AAV转导)表达人足细胞素

[0793] 材料与方法

[0794] AAV产生:用pHelper(HGT11,PS0150)、pAAV Rep-Cap(LK03,PS0240)和含有驱动足细胞素-HA表达的全长(FL)最小肾病蛋白启动子或265bp最小肾病蛋白启动子(分别为FL.NPSH1-足细胞素-HA和265.hNPSH1-足细胞素-HA)的两种ITR表达质粒之一,三重转染在补充有10%FBS、1%青霉素/链霉素和1%丙酮酸钠的DMEM中生长的60%-80%汇合293T人胚肾细胞。在1:2的比率的聚乙烯亚胺(PEI)存在的条件下,在150mm培养皿上进行转染。转染后第二天将培养基更换为含FBS的DMEM。在转染后第4天,分别收集和处理培养基和细胞。将PEG添加到过滤的上清液中以沉淀病毒颗粒,随后将其在1500g下离心20分钟。将细胞沉淀重悬于TD缓冲液中并在-80度下冷冻,然后在37度下解冻并涡旋。冻融循环重复5次。然后将来自PEG处理后的上清液的沉淀添加到裂解的细胞。进一步加入Benzonase和脱氧胆酸钠并在37度下孵育30分钟。然后将样品离心,并过滤上清液,并随后经由碘克沙醇梯度超速离心纯化。

[0795] AAV质量控制:经由qPCR滴定AAV样品,以确定每毫升病毒上清液中病毒基因组的数量。使用设计用于靶向足细胞素序列的引物。

[0796] AAV转导:将细胞以 $0.5 \times 10^5$ 个细胞/孔的密度接种到12孔板的孔中,并允许其贴壁过夜。第二天,添加病毒颗粒,感染复数(MOI)为每个细胞 $5 \times 10^5$ 个颗粒。将细胞置于37度以起始分化并培养9天。9天后,通过将流式细胞术作为足细胞素表达的量度,收获细胞并用抗HA抗体对HA表达进行染色。

[0797] 结果

[0798] 正如HEK细胞中的质粒转染数据所证明的(图11A),图14B中的结果进一步证实,两种启动子都能够在人足细胞中驱动转基因的表达。265bp最小肾病蛋白启动子在人足细胞中显示出更高的足细胞素表达水平。

[0799] 实施例11-经由AAV转导使用265bp与FL启动子体内表达GFP

[0800] 265肾病蛋白启动子和FL肾病蛋白启动子在WT小鼠(C57/B16)中进行了体内评价。将如实施例8所述的AAV颗粒(除了使用AAV9(PS0241)代替LK03外)经由约 $1 \times 10^{13}$  AAV2/9每只小鼠的尾静脉注射进行施用(每个治疗组2x小鼠,以及1x未治疗小鼠)。注射后4周进行肾组织的收获。对肾进行qPCR分析,并将GFP拷贝数相对于18S(管家基因)归一化。18S rRNA是实时PCR的可靠的归一化基因。

[0801] 图15中的结果显示了,在使用265bp最小肾病蛋白启动子和全长最小肾病蛋白启动子的肾中,有可检测的GFP表达。在使用265bp最小肾病蛋白启动子的肾中的表达更高。

[0802] 上述说明书中提及的所有出版物均在此通过引用并入。在不脱离本发明的范围和精力的情况下,本发明所公开的方法、细胞、组合物和用途的各种修改和变化对于技术人员来说将是显而易见的。尽管已经结合特定的优选实施方案公开了本发明,但是应当理解,所要求保护的发明不应不合适地限于这些特定的实施方案。实际上,对于本领域技术人员来说显而易见的用于实施本发明的本公开模式的各种修改旨在落入所附权利要求的范围内。

[0803] 实施方案

[0804] 现在将参考以下的编号的段落来描述本发明的各种优选特征和实施方案。

[0805] 1.启动子,其包含(i)与SEQ ID NO:4具有至少70%同一性的核苷酸序列,并且其中所述启动子具有约1.1kb或更短的长度。

[0806] 2.启动子,其包含与根据SEQ ID NO:1的核苷酸序列具有至少70%同一性的核苷酸序列,或由与根据SEQ ID NO:1的核苷酸序列具有至少70%同一性的核苷酸序列组成,但其中:

[0807] (i)缺失了SEQ ID NO:1的位置1至位置n1,其中n1为100至430的整数;和/或

[0808] (ii)缺失了SEQ ID NO:1的位置n2至位置n3,其中 $n_3 \geq n_2$ ,n2为508至1061的整数,n3为508至1061的整数。

[0809] 3.根据段落2的启动子,其中所述启动子包含(i)与SEQ ID NO:4具有至少70%同一性的核苷酸序列,和/或其中所述启动子具有约1.1kb或更短的长度。

[0810] 4.根据任一前述段落的启动子,其中所述启动子是足细胞特异性启动子。

[0811] 5.根据任一前述段落的启动子,其中所述启动子进一步包含(ii)与SEQ ID NO:5具有至少70%同一性的核苷酸序列、与SEQ ID NO:6具有至少70%同一性的核苷酸序列、和/或与SEQ ID NO:7具有至少70%同一性的核苷酸序列。

[0812] 6.根据任一前述段落的启动子,其中所述启动子进一步包含(iii)与SEQ ID NO:8具有至少70%同一性的核苷酸序列或其一个或多个片段。

[0813] 7.根据任一前述段落的启动子,其中所述启动子从5'至3'包含:(i)与SEQ ID NO:4具有至少70%同一性的核苷酸序列;(iii)任选地与SEQ ID NO:8具有至少70%同一性的核苷酸序列,或其一个或多个片段;和(ii)与SEQ ID NO:5具有至少70%同一性的核苷酸序

列、与SEQ ID NO:6具有至少70%同一性的核苷酸序列、和/或与SEQ ID NO:7具有至少70%同一性的核苷酸序列。

[0814] 8. 根据任一前述段落的启动子,其中所述启动子具有约1.0kb或更短、约0.9kb或更短、约0.8kb或更短、约0.7kb或更短、约0.6kb或更短、约0.5kb或更短、约0.4kb或更短,或约0.3kb或更短的长度。

[0815] 9. 根据任一前述段落的启动子,其中所述启动子具有0.265kb至1.0kb、0.265kb至0.9kb、0.265kb至0.8kb、0.265kb至0.7kb、0.265kb至0.6kb、0.265kb至0.5kb、0.265kb至0.4kb、或0.265kb至0.3kb的长度。

[0816] 10. 根据任一前述段落的启动子,其中所述启动子包含:

[0817] (a) 视黄酸受体结合位点;

[0818] (b) WT1结合位点;

[0819] (c) 增强子框;

[0820] (d) 转录因子结合区;和/或

[0821] (e) 转录起始位点。

[0822] 11. 根据段落10的启动子,其中所述视黄酸受体结合位点包含以下或由以下组成:如SEQ ID NO:10所示的核苷酸序列,或与SEQ ID NO:10相比具有一个或两个取代、缺失或插入的核苷酸序列。

[0823] 12. 根据段落10或11的启动子,其中所述WT1结合位点包含以下或由以下组成:如SEQ ID NO:11所示的核苷酸序列,或与SEQ ID NO:11相比具有一个、两个或三个取代、缺失或插入的核苷酸序列。

[0824] 13. 根据段落10至12中任一段的启动子,其中所述增强子框包含以下或由以下组成:如SEQ ID NO:12所示的核苷酸序列,或与SEQ ID NO:12相比具有一个或两个取代、缺失或插入的核苷酸序列。

[0825] 14. 根据段落1或3至13中任一段的启动子,其中(i)所述与SEQ ID NO:4具有至少70%同一性的核苷酸序列包含以下的一个或多个:(a) 视黄酸受体结合位点;(b) WT1结合位点;和(c) 增强子框。

[0826] 15. 根据段落1或3至14中任一段的启动子,其中以下核苷酸序列中的一个或多个存在于(i)所述与SEQ ID NO:4具有至少70%同一性的核苷酸序列中:

[0827] (a) 对应于SEQ ID NO:4的位置7至位置13的位置处的GGGGTCA;

[0828] (b) 对应于SEQ ID NO:4的位置14至位置30的位置处的CGGAGGCTGGGGAGGCA;以及

[0829] (c) 对应于SEQ ID NO:4的位置49至位置53的位置处的ATGTG。

[0830] 16. 根据段落10至15中任一段的启动子,其中所述转录因子结合区包含以下或由以下组成:如SEQ ID NO:13所示的核苷酸序列,或与SEQ ID NO:13相比具有一个、两个、三个、四个或五个取代、缺失或插入的核苷酸序列。

[0831] 17. 根据段落10至16中任一段的启动子,其中所述转录起始位点包含“AG”二核苷酸或由“AG”二核苷酸组成。

[0832] 18. 根据段落10至17中任一段的启动子,其中所述转录因子结合位点可操作地连接至所述转录起始位点,任选地其中所述转录因子结合位点直接位于所述转录起始位点的上游。

- [0833] 19. 根据任一前述段落的启动子,其中所述启动子包含与SEQ ID NO:2或SEQ ID NO:3具有至少70%同一性的核苷酸序列或由与SEQ ID NO:2或SEQ ID NO:3具有至少70%同一性的核苷酸序列组成。
- [0834] 20. 启动子,其由与SEQ ID NO:2具有至少70%同一性的核苷酸序列组成。
- [0835] 21. 启动子,其由与SEQ ID NO:3具有至少70%同一性的核苷酸序列组成。
- [0836] 22. 多核苷酸,其包含段落1至21中任一段的启动子。
- [0837] 23. 根据段落22的多核苷酸,其中所述启动子可操作地连接至蛋白质编码序列。
- [0838] 24. 根据段落23的多核苷酸,其中所述蛋白质编码序列可操作地连接至一个或多个进一步的调控元件,诸如转录后调控元件和/或多腺苷酸化序列。
- [0839] 25. 载体,其包含根据段落22至24中任一段的多核苷酸。
- [0840] 26. 根据段落25的载体,其中所述载体能够转导足细胞,任选地其中所述载体能够特异性转导足细胞。
- [0841] 27. 根据段落25或26的载体,其中所述载体是病毒载体,诸如腺相关病毒(AAV)载体、慢病毒载体、逆转录病毒载体、腺病毒载体、腺相关病毒载体、单纯疱疹病毒载体、甲病毒载体、黄病毒载体、弹状病毒载体、麻疹病毒载体、鸡新城疫病毒载体、痘病毒载体和小核糖核酸病毒载体,优选地其中所述载体是AAV载体。
- [0842] 28. 根据段落27的载体,其中所述病毒载体是以病毒载体颗粒的形式,优选地其中所述病毒载体是以AAV载体颗粒的形式。
- [0843] 29. 根据段落25至28中任一段的载体,其中所述载体是由AAV3B、LK03或AAV9衣壳蛋白衣壳化的AAV载体颗粒的形式。
- [0844] 30. 细胞,其包含根据段落22至24中任一段的多核苷酸、或根据段落25至29中任一段的载体。
- [0845] 31. 药物组合物,其包含根据段落22至24中任一段的多核苷酸、根据段落25至29中任一段的载体、或根据段落30的细胞。
- [0846] 32. 根据段落22至24中任一段的多核苷酸、根据段落25至29中任一段的载体或根据段落30的细胞,其用于药物中。
- [0847] 33. 根据段落1至21中任一段的启动子用于驱动编码序列的表达的用途。
- [0848] 34. 根据段落32的用途,其中所述表达是足细胞特异性的。

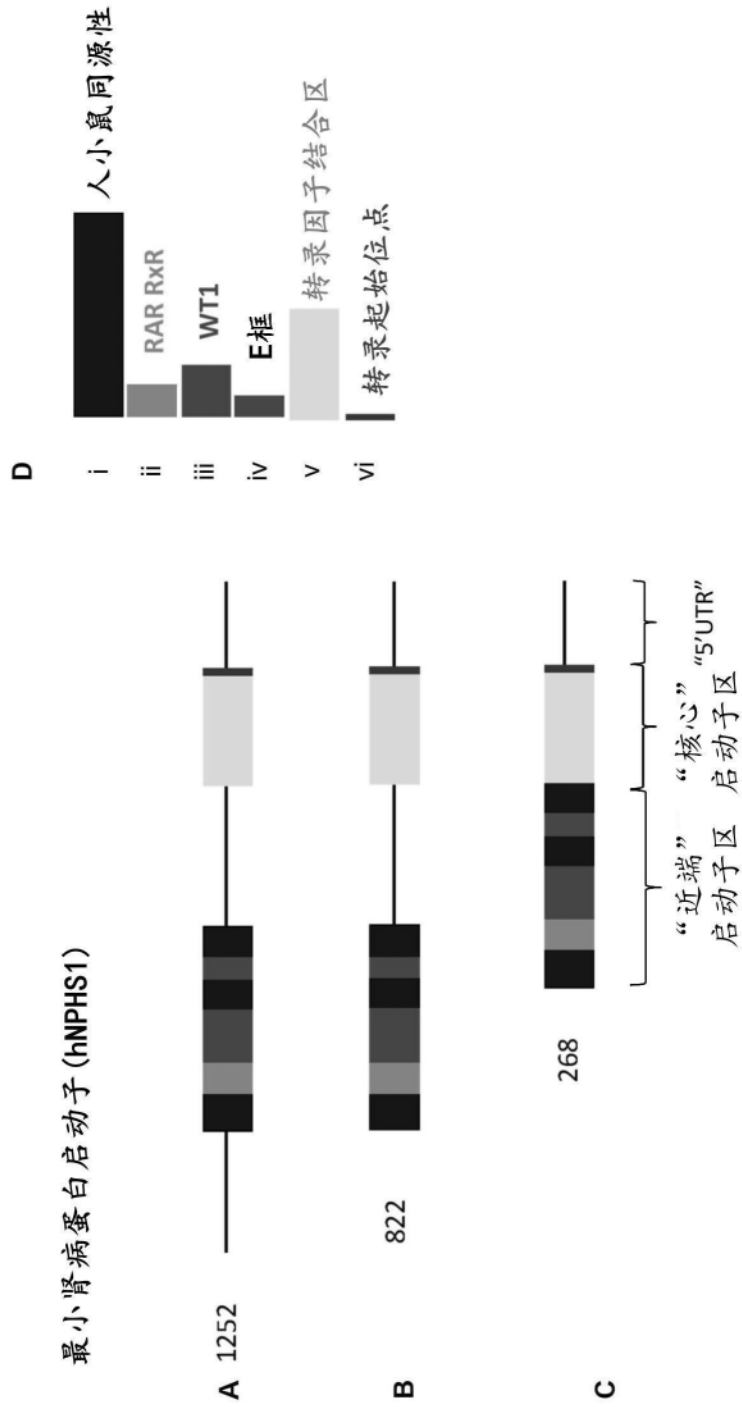


图1

A

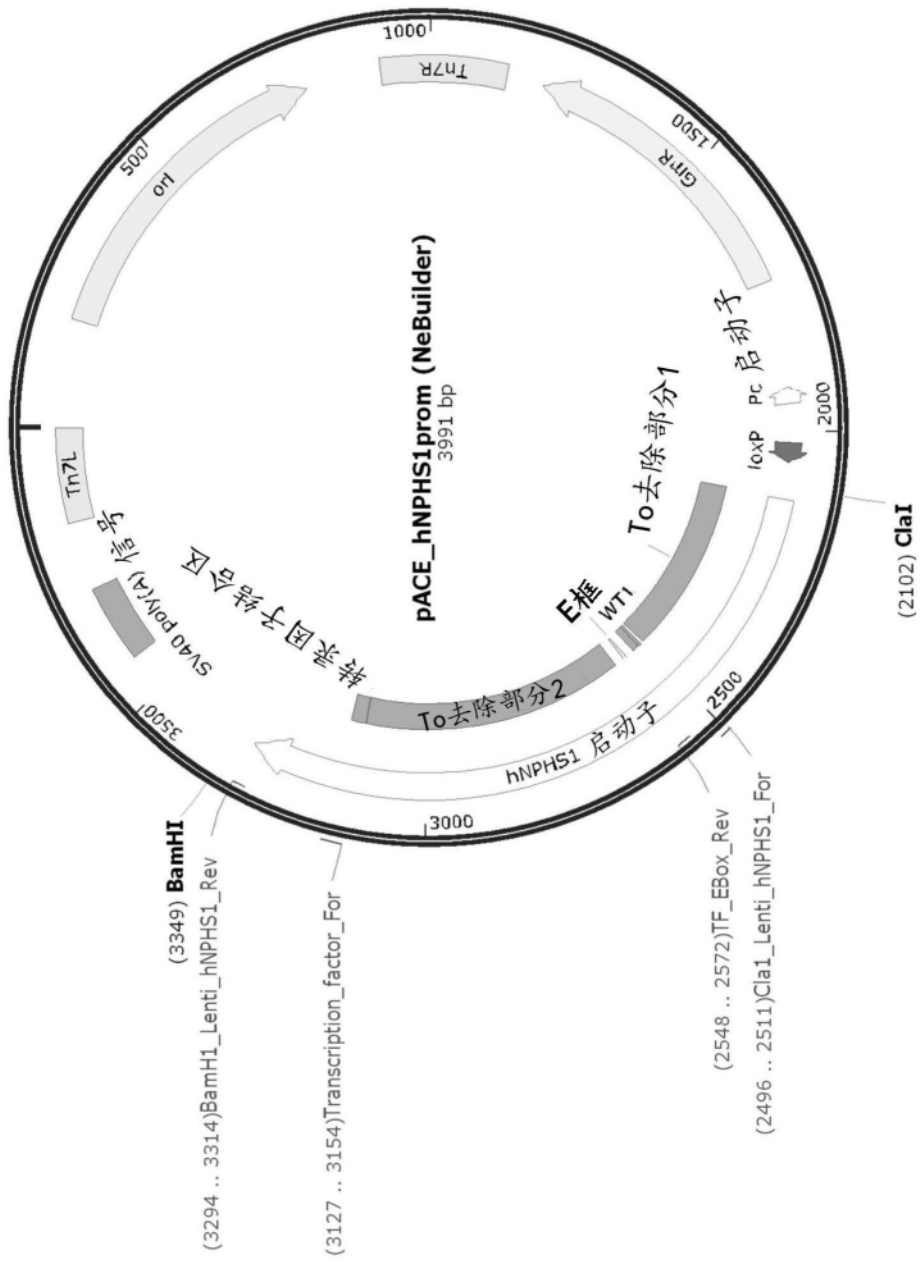


图2

B

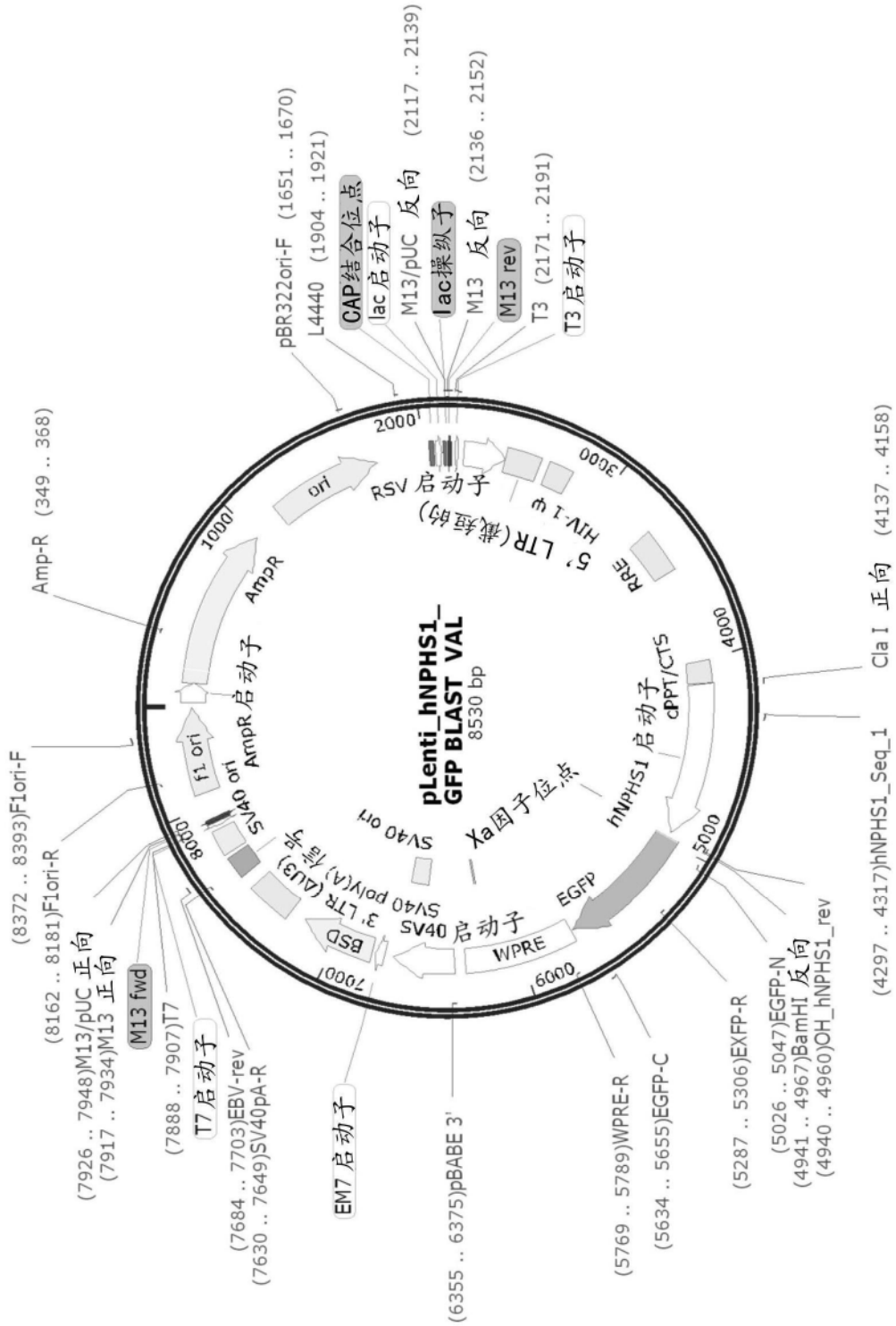


图2 (续)

A

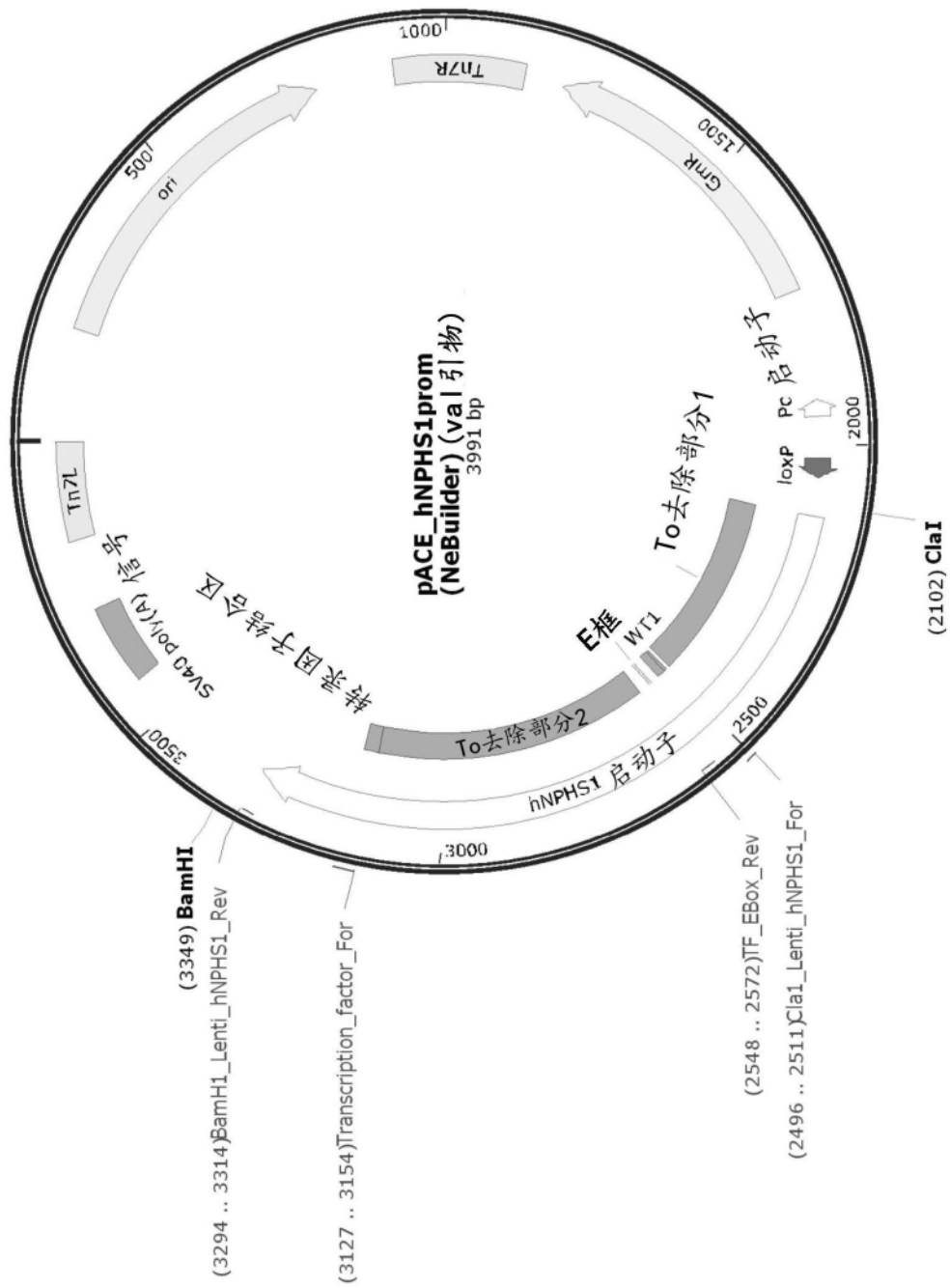


图3



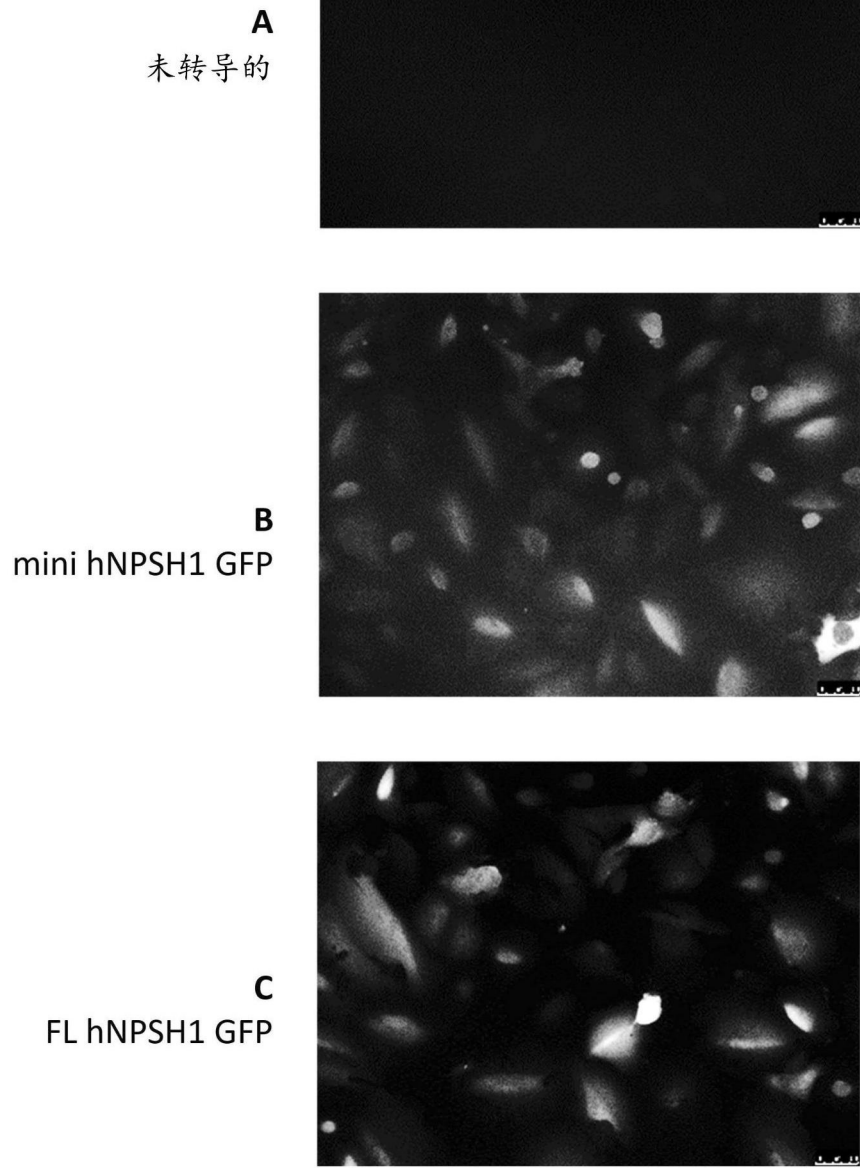


图4

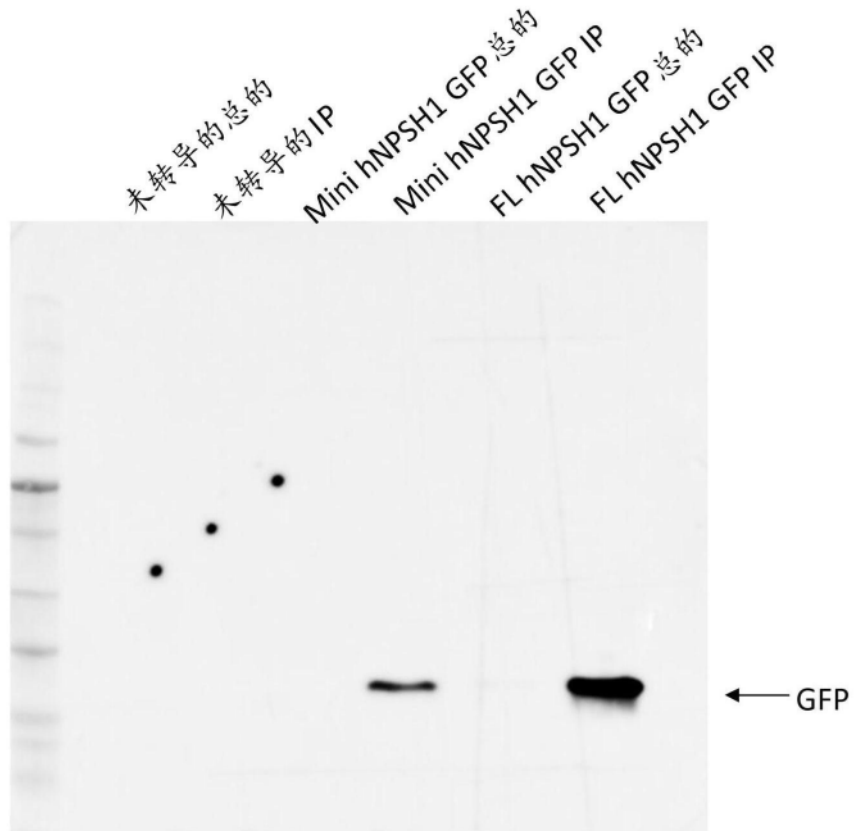


图5

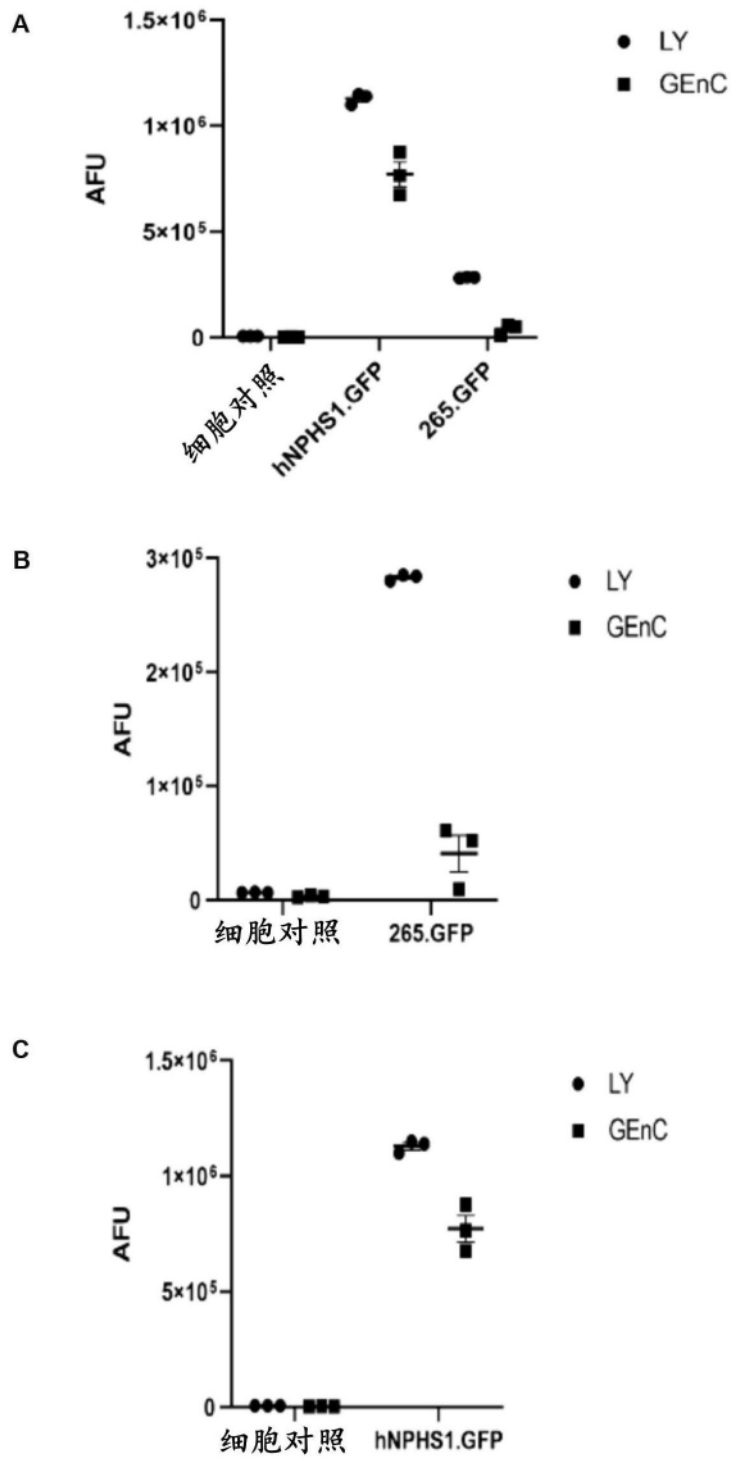


图6

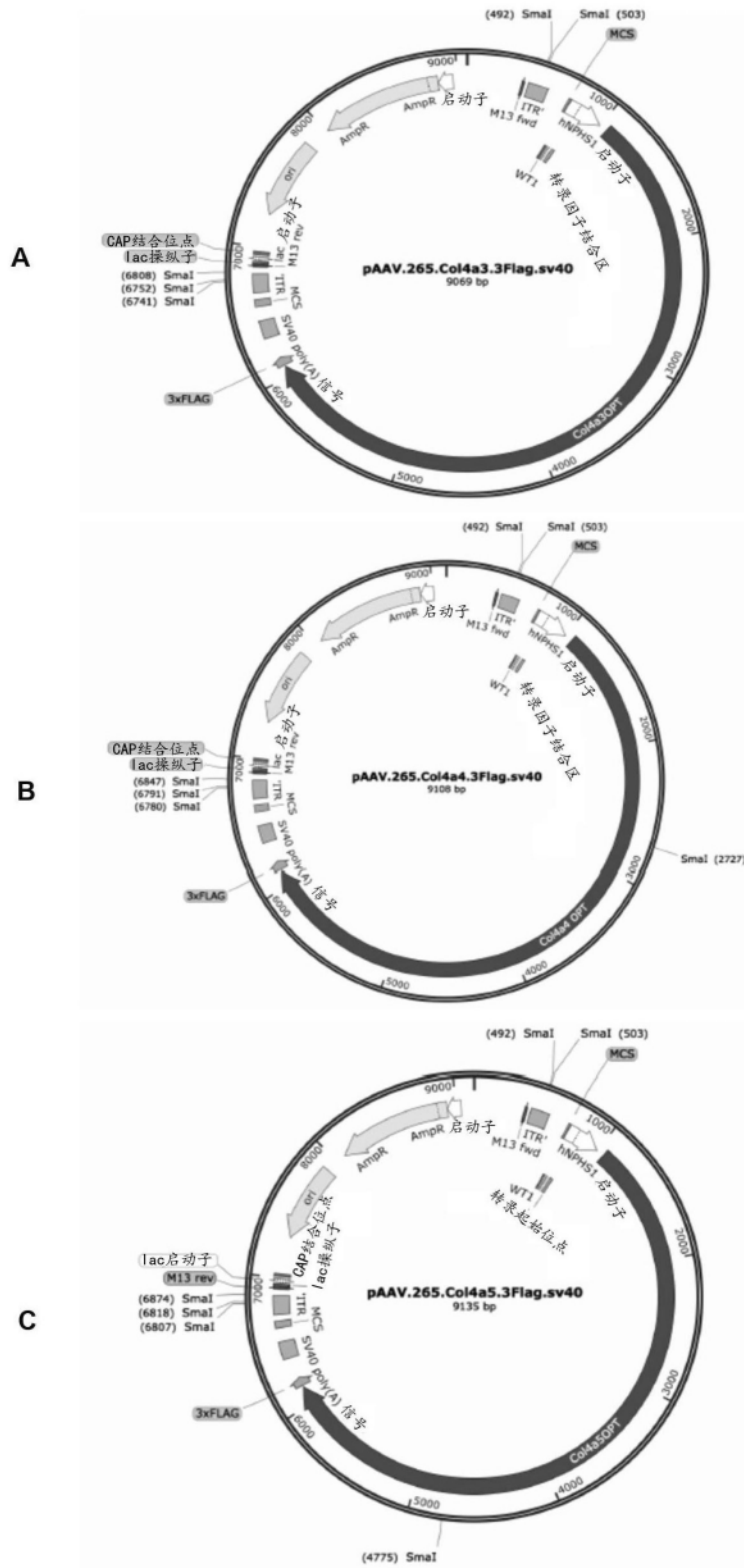
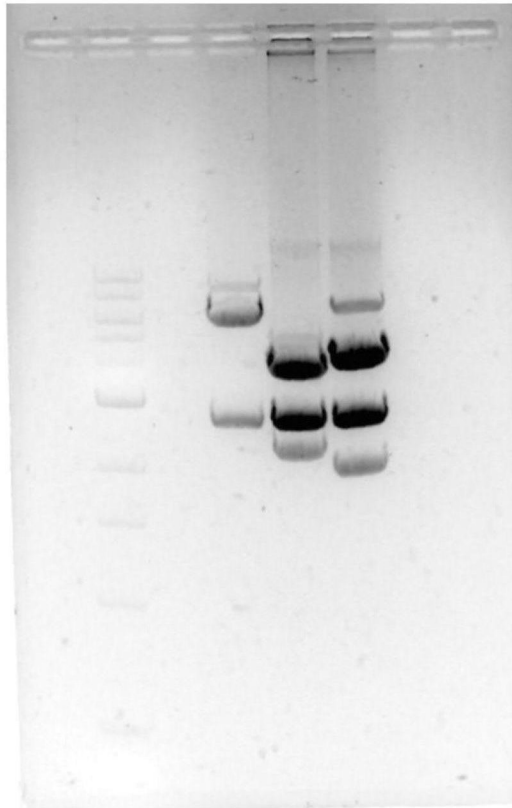
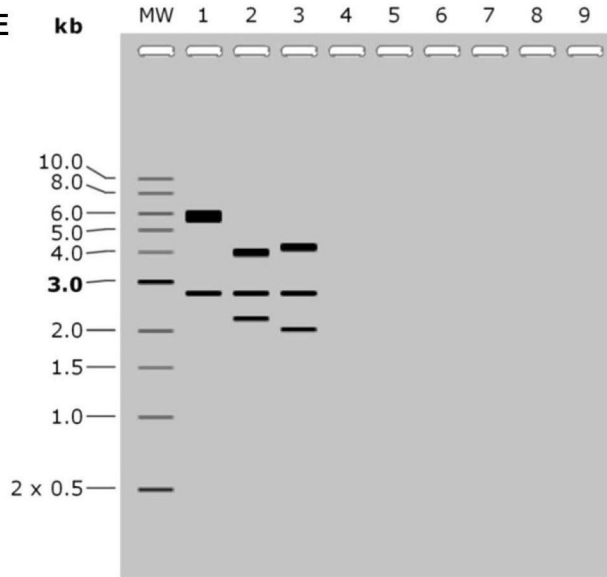


图7

D



E kb



0.8%琼脂糖

图7(续)

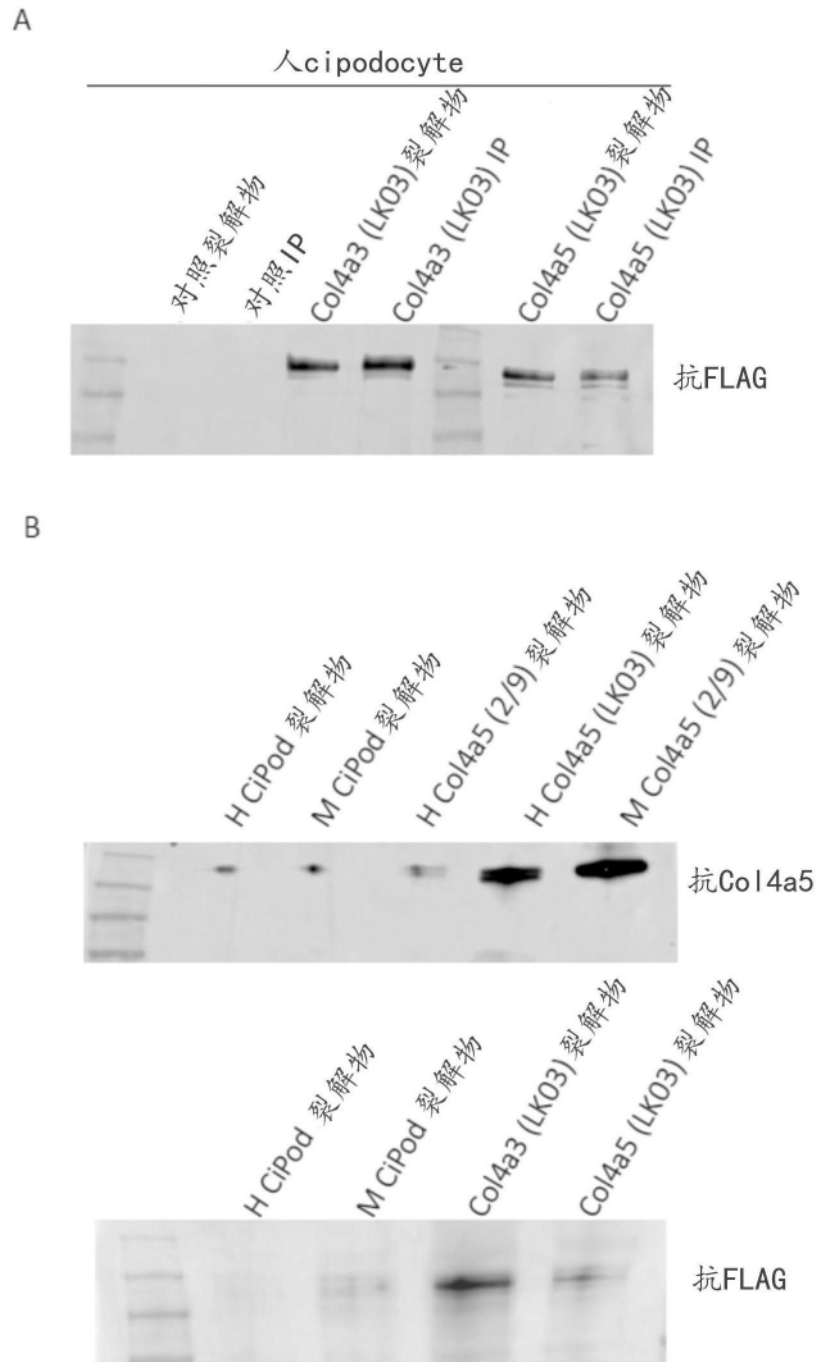


图8

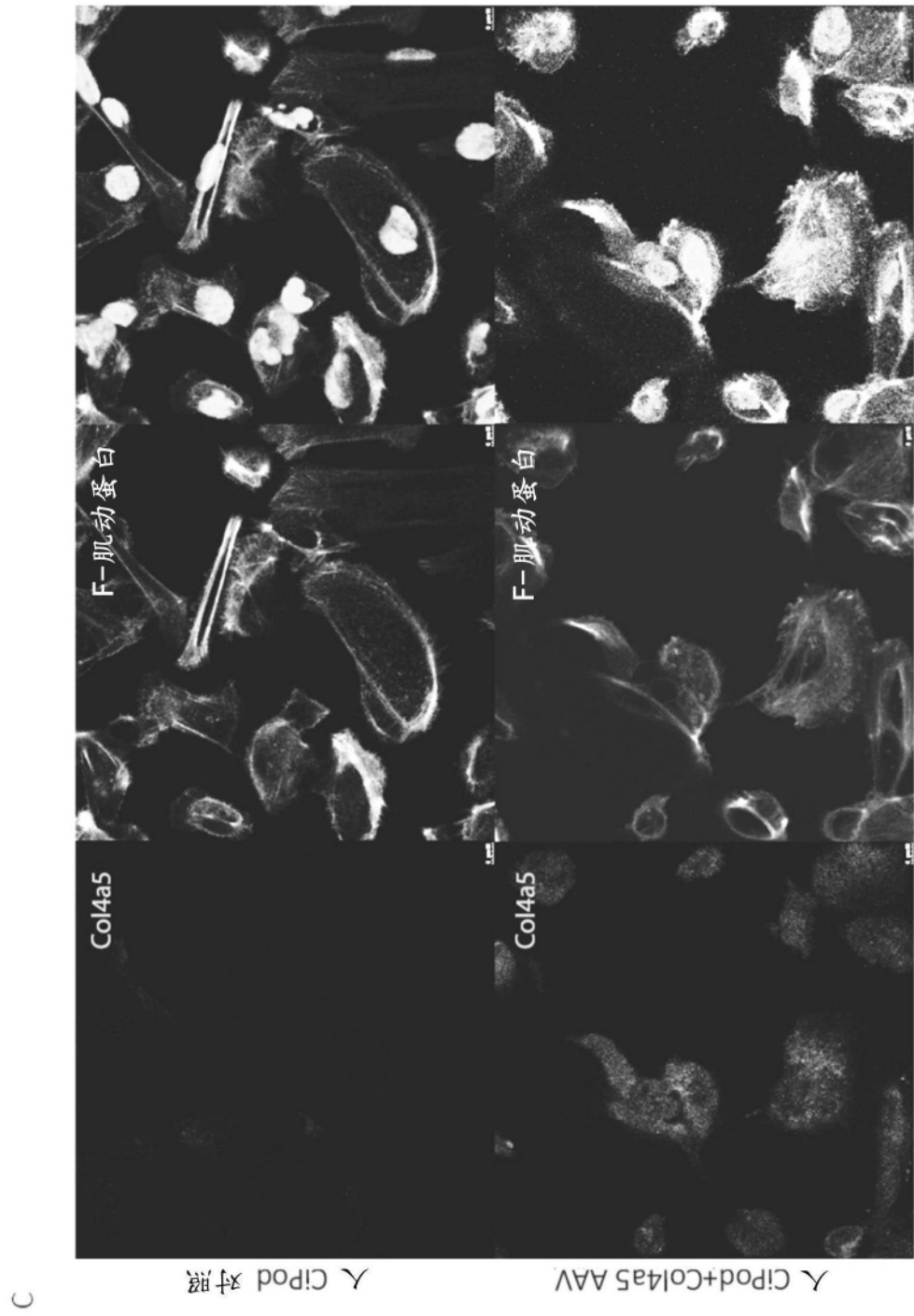


图8(续)

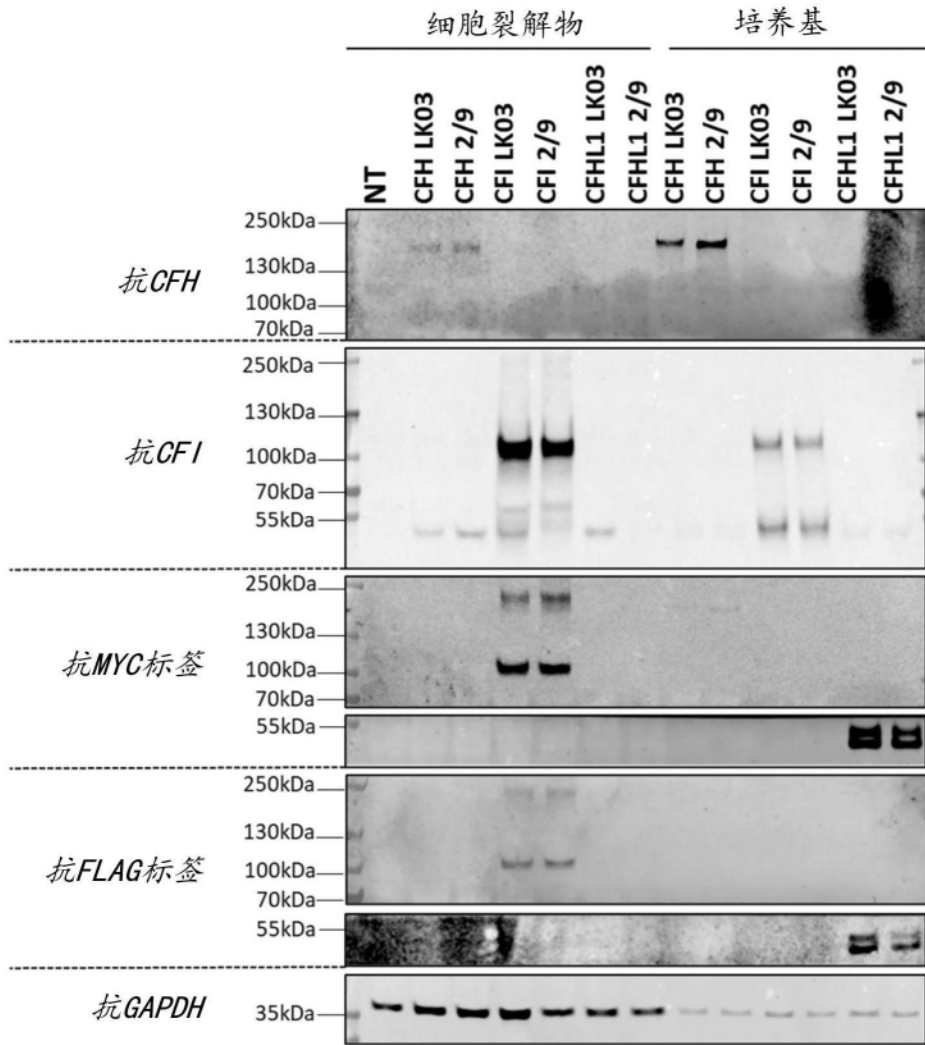
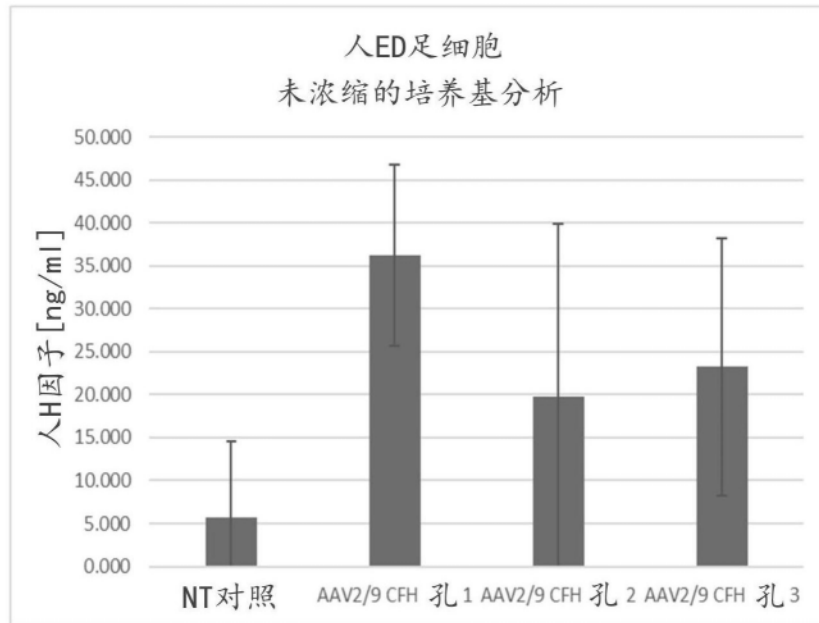


图9

**A**



**B**

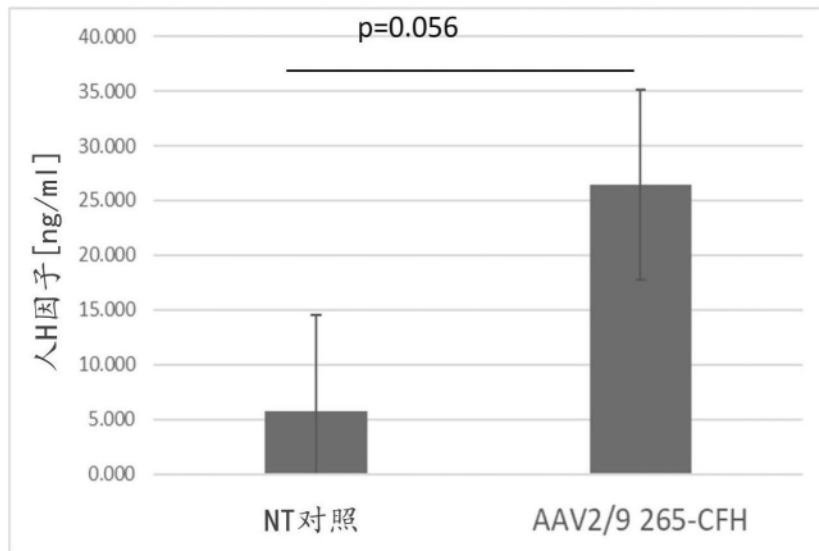


图10

C

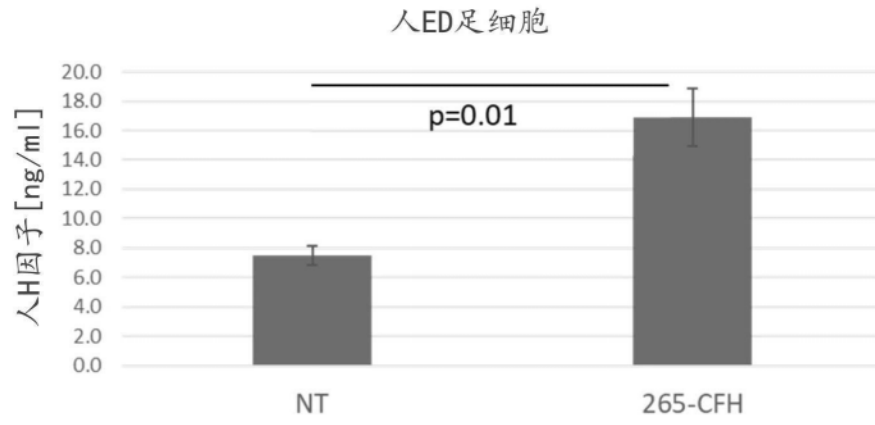


图10(续)

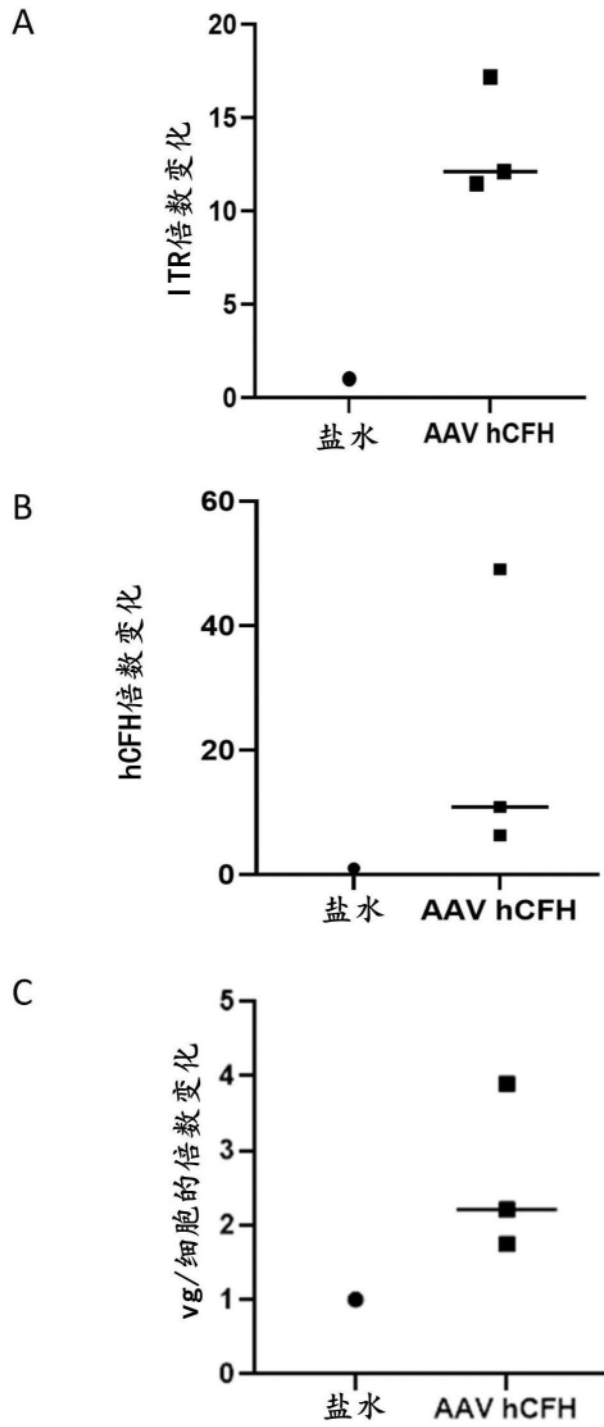


图11

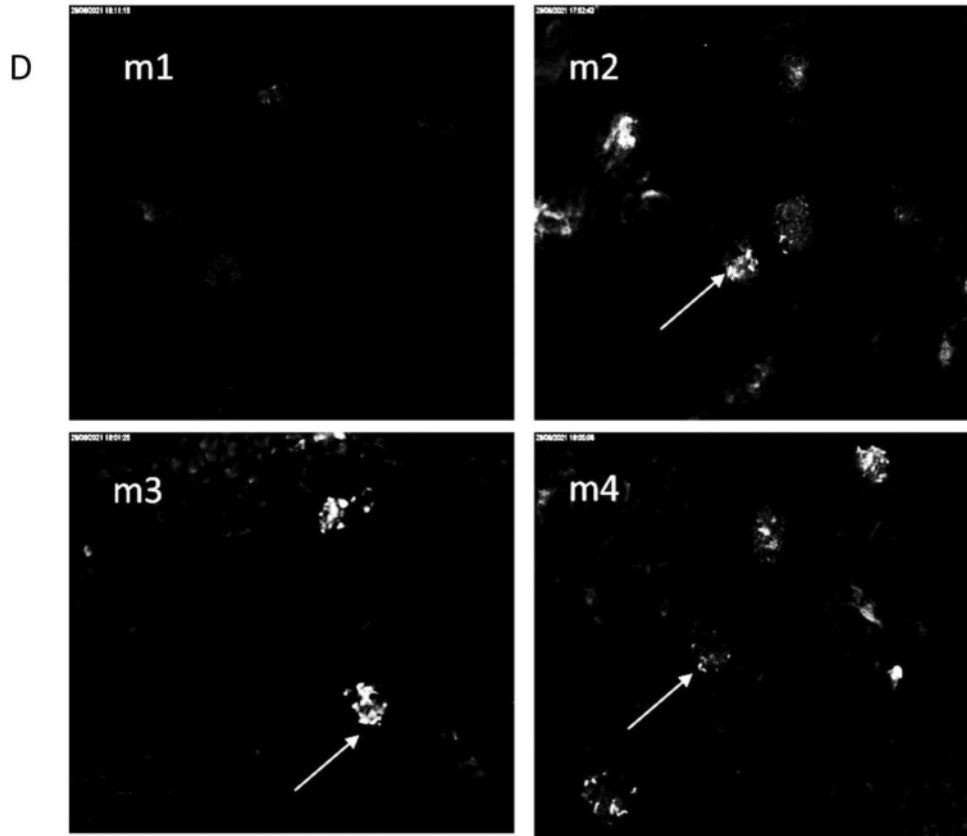


图11(续)

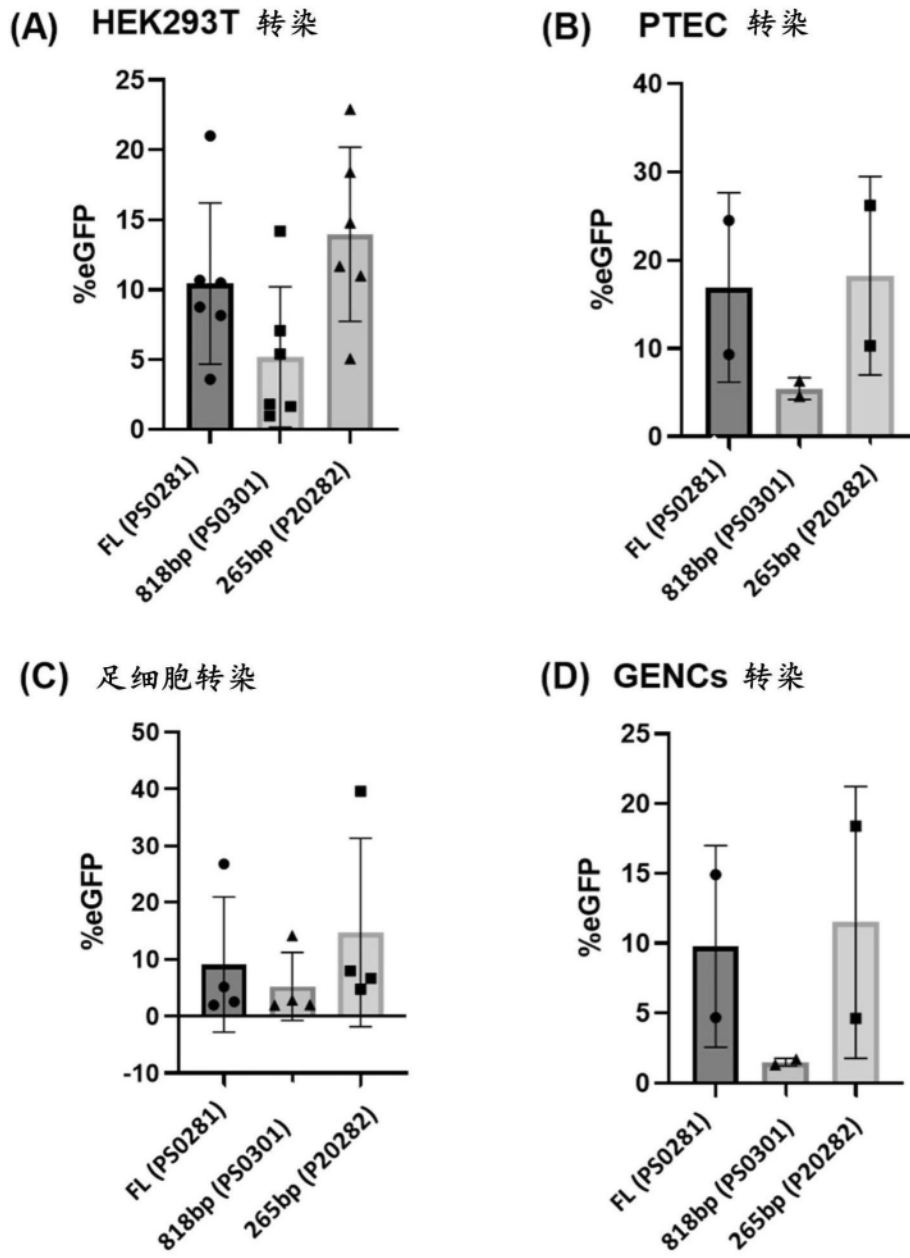


图12

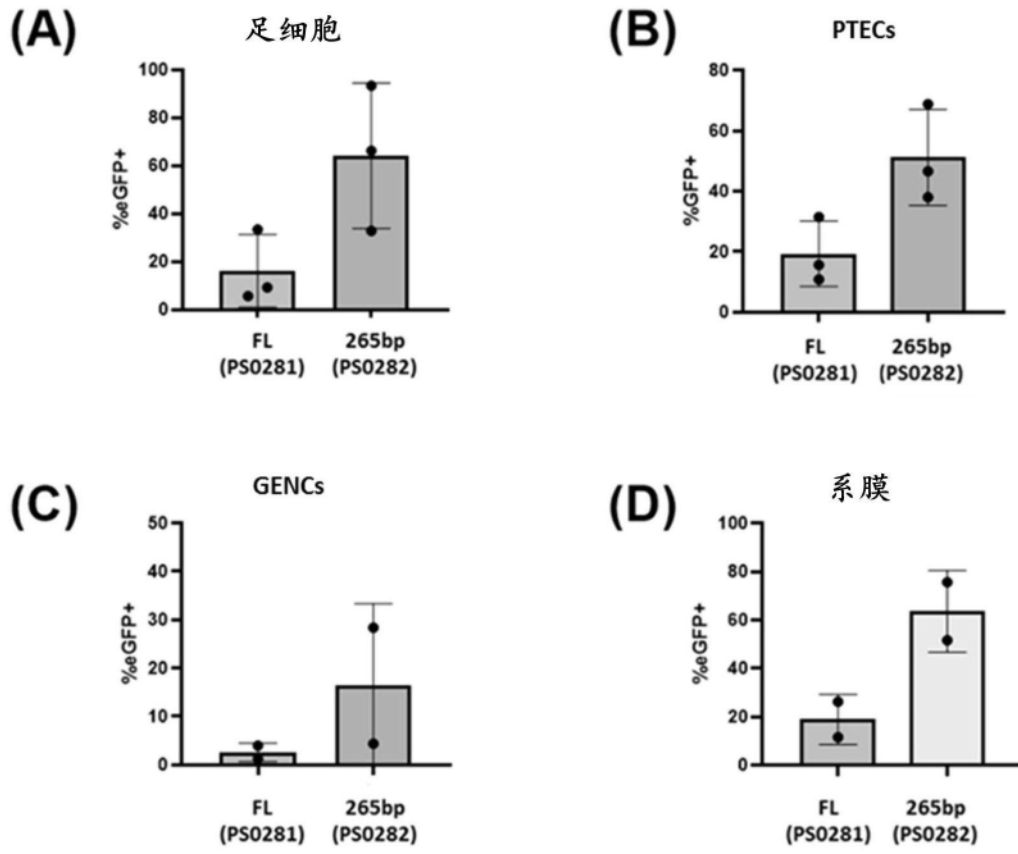


图13

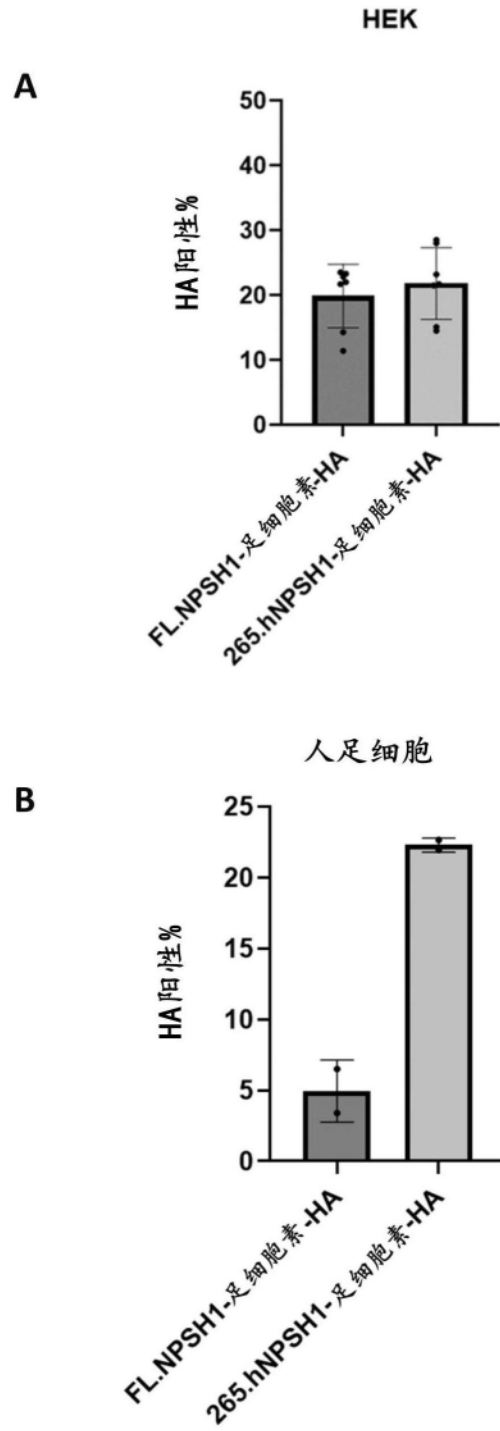


图14

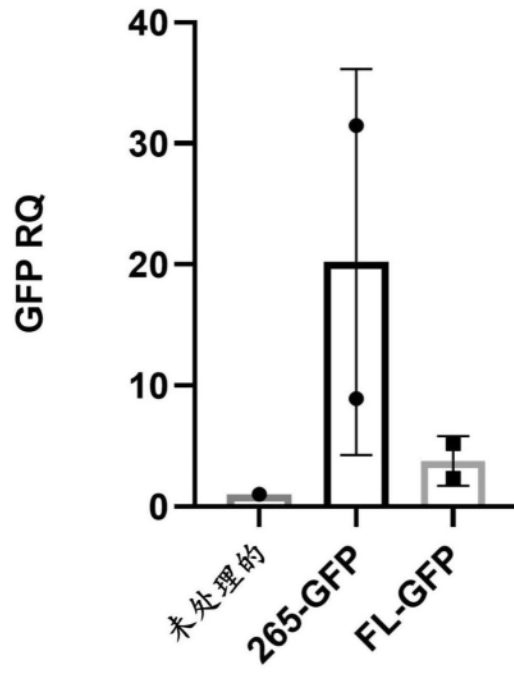


图15