



(12) 发明专利申请

(10) 申请公布号 CN 116782886 A

(43) 申请公布日 2023. 09. 19

(21) 申请号 202180072680.6

(22) 申请日 2021.08.30

(30) 优先权数据

63/073843 2020.09.02 US

(85) PCT国际申请进入国家阶段日

2023.04.24

(86) PCT国际申请的申请数据

PCT/US2021/048267 2021.08.30

(87) PCT国际申请的公布数据

W02022/051232 EN 2022.03.10

(71) 申请人 4D分子治疗有限公司

地址 美国加利福尼亚州

(72) 发明人 D·H·基恩 M·A·科特爾曼

D·沙菲尔 P·弗朗西斯

(74) 专利代理机构 中国专利代理(香港)有限公

司 72001

专利代理师 段菊兰 彭昶

(51) Int.Cl.

A61K 9/14 (2006.01)

权利要求书3页 说明书29页

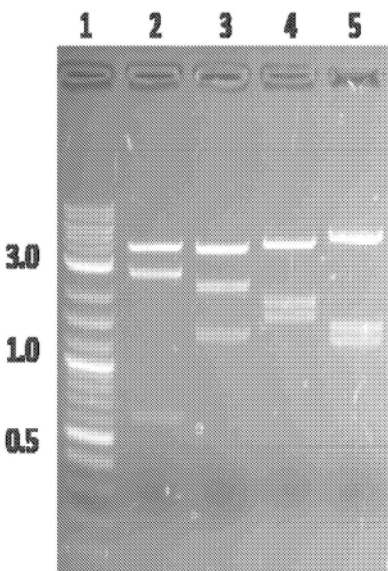
序列表15页 附图14页

(54) 发明名称

密码子优化的RPGRORF15基因及其用途

(57) 摘要

本公开提供了密码子优化的RPGRorf15序列、包含密码子优化的RPGRorf15序列的载体和宿主细胞,以及治疗视网膜病症,诸如XLRP的方法,所述方法包括向受试者施用密码子优化的RPGRorf15序列。



1. 编码SEQ ID NO:2的人视网膜色素变性GTP酶调节因子(RPGR)蛋白并经过密码子优化用于在人中表达的核酸,所述核酸包含如SEQ ID NO:1所示的核苷酸序列或包含与其具有至少95%同一性的核苷酸序列,其中所述核酸与SEQ ID NO:3的野生型RPGR核苷酸序列在其他方面相同的细胞中的表达水平相比以更高的水平表达。

2. 根据权利要求1所述的核酸,其中所述核苷酸序列具有至少0.89的密码子适应指数。

3. 根据权利要求1所述的核酸,其包含如SEQ ID NO:1所示的核苷酸序列。

4. 表达盒,其包含根据权利要求1至3任一项所述的核酸和可操作地连接至所述核酸序列并与所述核酸序列异源的表达控制序列。

5. 权利要求4所述的表达盒,其中所述表达控制序列是组成型启动子。

6. 权利要求4所述的表达盒,其中所述表达控制序列是指导所述核酸在视杆(rods)和视锥(cones)中的优先表达的启动子,优选人G蛋白偶联受体视紫红质激酶1(hGRK)启动子,其包含如SEQ ID NO:4所示的核苷酸序列或与其具有至少90%、至少95%、或至少98%同一性的序列。

7. 权利要求6所述的表达盒,其从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf15基因,(d) SV40多聚腺苷酸化序列,和(e) AAV2末端重复。

8. 权利要求7所述的表达盒,其中所述5'AAV2末端重复具有如SEQ ID NO:6所示的核苷酸序列和/或其中所述hGRK启动子具有如SEQ ID NO:4所示的核苷酸序列和/或其中所述SV40多聚腺苷酸化序列具有如SEQ ID NO:8所示的核苷酸序列和/或其中所述3'AAV2末端重复具有如SEQ ID NO:7所示的核苷酸序列。

9. 权利要求8所述的表达盒,其包含SEQ ID NO:5的核苷酸序列或与其具有至少90%、至少95%、至少98%同一性的序列,或由其组成。

10. 载体,其包含根据权利要求1至3任一项所述的核酸或根据权利要求4至9任一项所述的表达盒。

11. 权利要求10所述的载体,其中所述载体是重组腺相关(rAAV)载体。

12. 权利要求11所述的载体,其中所述rAAV载体包含血清型2、5或8的AAV衣壳或其变体。

13. 权利要求12所述的载体,其中所述rAAV载体包含AAV2衣壳或其变体。

14. 权利要求13所述的载体,其中所述rAAV载体包含含有衣壳蛋白的AAV2衣壳变体,所述衣壳蛋白包含SEQ ID NO:9的序列或由其组成。

15. 权利要求11-14任一项所述的载体,其中所述rAAV载体包含核酸,所述核酸从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf15基因,和(d) AAV2末端重复。

16. 权利要求15所述的载体,其中所述5'AAV2末端重复具有如SEQ ID NO:6所示的核苷酸序列和/或其中所述hGRK启动子具有如SEQ ID NO:4所示的核苷酸序列和/或其中所述SV40多聚腺苷酸化序列具有如SEQ ID NO:8所示的核苷酸序列和/或其中所述3'AAV2末端重复具有如SEQ ID NO:7所示的核苷酸序列。

17. 权利要求16所述的载体,其中所述rAAV载体包含核酸,所述核酸包含SEQ ID NO:5的核苷酸序列或与其具有至少90%、至少95%、或至少98%同一性的序列。

18. 权利要求17所述的载体,其中所述rAAV载体包含(i)含有衣壳蛋白的衣壳,所述衣壳蛋白包含SEQ ID NO:9的序列或由其组成,和(ii)包含SEQ ID NO:5的核苷酸序列或由其组成的核酸。

19. 宿主细胞,其包含根据权利要求1至3任一项所述的核酸或根据权利要求4至9任一项所述的表达盒。

20. 根据权利要求19所述的宿主细胞,其中所述宿主细胞是哺乳动物细胞。

21. 权利要求19或20所述的宿主细胞,其中所述宿主细胞是CHO细胞、HEK293细胞、HEK293T细胞、HeLa细胞、BHK21细胞或Vero细胞和/或其中所述宿主细胞在悬浮培养或细胞培养室培养中生长和/或其中所述宿主细胞是感光细胞、视网膜神经节细胞、胶质细胞、双极细胞、无长突细胞、水平细胞或视网膜色素上皮细胞。

22. 用于在有需要的受试者中治疗XLRP的方法,其包括向所述受试者施用治疗有效量的根据权利要求1-3任一项所述的核酸、根据权利要求4-9任一项所述的表达盒或根据权利要求10-18任一项所述的载体。

23. 用于在有需要的受试者中治疗XLRP的方法,其包括向所述受试者施用传染性rAAV,所述传染性rAAV包含(i) AAV衣壳和(ii) 核酸,所述核酸从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf15基因,(d) SV40多聚腺苷酸化序列,和(e) AAV2末端重复。

24. 根据权利要求23所述的方法,其中所述5' AAV2末端重复具有如SEQ ID NO:6所示的核苷酸序列和/或其中所述hGRK启动子具有如SEQ ID NO:4所示的核苷酸序列和/或其中所述SV40多聚腺苷酸化序列具有如SEQ ID NO:8所示的核苷酸序列和/或其中所述3' AAV2末端重复具有如SEQ ID NO:7所示的核苷酸序列。

25. 根据权利要求23或24所述的方法,其中所述rAAV包含(i)含有衣壳蛋白的衣壳,所述衣壳蛋白包含SEQ ID NO:9的序列或由其组成,和(ii)包含SEQ ID NO:5的核苷酸序列或由其组成的核酸。

26. 根据权利要求22-25任一项所述的方法,其中所述核酸或载体通过眼周、玻璃体内、脉络膜上或视网膜下注射施用于所述受试者和/或其中所述载体以大约 $10^{10}$ 个载体基因组(vg)/眼至大约 $10^{13}$ vg/眼,优选大约 $1 \times 10^{11}$ vg/眼至大约 $5 \times 10^{12}$ vg/眼的剂量,更优选以大约 $3 \times 10^{11}$ vg/眼的剂量或以大约 $1 \times 10^{12}$ vg/眼的剂量施用于所述受试者。

27. 用于治疗XLRP的根据权利要求1-3任一项所述的核酸、根据权利要求4-9任一项所述的表达盒、或根据权利要求10-18任一项所述的载体。

28. 用于制造用于治疗XLRP的药物的根据权利要求1-3任一项所述的核酸、根据权利要求4-9任一项所述的表达盒、或根据权利要求10-18任一项所述的载体。

29. 用于根据权利要求27或28所述的用途的核酸、表达盒或载体,其中所述核酸或载体通过眼周、玻璃体内、脉络膜上或视网膜下注射来施用和/或其中所述载体以大约 $10^{10}$ 个载体基因组(vg)/眼至大约 $10^{13}$ vg/眼,优选大约 $1 \times 10^{11}$ vg/眼至大约 $5 \times 10^{12}$ vg/眼的剂量施用,更优选以大约 $3 \times 10^{11}$ vg/眼的剂量或以大约 $1 \times 10^{12}$ vg/眼的剂量施用。

30. 用于治疗XLRP的传染性rAAV,其包含(i) AAV衣壳和(ii) 核酸,所述核酸从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf125基因,(d) SV40多聚腺苷酸化序列,和(e) AAV2末端重复。

31. 用于制造用于治疗XLRP的药物的传染性rAAV,其包含(i) AAV衣壳和(ii) 核酸,所述核酸从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf125基因,(d) SV40多聚腺苷酸化序列,和(e) AAV2末端重复。

32. 根据权利要求30或31所述的传染性rAAV,其中所述5' AAV2末端重复具有如SEQ ID NO:6所示的核苷酸序列和/或其中所述hGRK启动子具有如SEQ ID NO:4所示的核苷酸序列和/或其中所述SV40多聚腺苷酸化序列具有如SEQ ID NO:8所示的核苷酸序列和/或其中所述3' AAV2末端重复具有如SEQ ID NO:7所示的核苷酸序列。

33. 根据权利要求32所述的传染性rAAV,其中所述rAAV包含(i) 含有衣壳蛋白的衣壳,所述衣壳蛋白包含SEQ ID NO:9的序列或由其组成,和(ii) 包含SEQ ID NO:5的核苷酸序列或由其组成的核酸。

34. 用于根据权利要求30-33任一项所述的用途的传染性rAAV,其中所述rAAV通过玻璃体内注射来施用和/或其中所述载体以大约 $10^{10}$ 个载体基因组(vg)/眼至大约 $10^{13}$ vg/眼,优选大约 $1 \times 10^{11}$ vg/眼至大约 $5 \times 10^{12}$ vg/眼的剂量施用,更优选以大约 $3 \times 10^{11}$ vg/眼的剂量或以大约 $1 \times 10^{12}$ vg/眼的剂量施用。

35. 用于治疗哺乳动物中由降低的RPGRorf15水平介导的疾病或状况的方法,所述方法包括施用治疗有效量的根据权利要求1-3任一项所述的核酸、根据权利要求4-9任一项所述的表达盒、或根据权利要求10-18任一项所述的载体。

36. 用于提高哺乳动物中的RPGRorf15水平的方法,所述方法包括向所述哺乳动物施用根据权利要求1-3任一项所述的核酸、根据权利要求4-9任一项所述的表达盒、或根据权利要求10-18任一项所述的载体。

37. 药物组合物,其包含根据权利要求1-3任一项所述的核酸、根据权利要求4-9任一项所述的表达盒、或根据权利要求10-18任一项所述的载体,以及至少一种药学上可接受的赋形剂。

38. 药物组合物,其包含传染性rAAV,所述传染性rAAV包含(i) AAV衣壳和(ii) 核酸,所述核酸从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf基因,(d) SV40多聚腺苷酸化序列,和(e) AAV2末端重复。

39. 根据权利要求38所述的药物组合物,其中所述5' AAV2末端重复具有如SEQ ID NO:6所示的核苷酸序列和/或其中所述hGRK启动子具有如SEQ ID NO:4所示的核苷酸序列和/或其中所述SV40多聚腺苷酸化序列具有如SEQ ID NO:8所示的核苷酸序列和/或其中所述3' AAV2末端重复具有如SEQ ID NO:7所示的核苷酸序列。

40. 根据权利要求39所述的药物组合物,其中所述rAAV包含(i) 含有衣壳蛋白的衣壳,所述衣壳蛋白包含SEQ ID NO:9的序列或由其组成,和(ii) 包含SEQ ID NO:5的核苷酸序列或由其组成的核酸。

41. 根据权利要求38-40任一项所述的药物组合物,其中所述药物组合物包含 $10^9$ vg至 $10^{14}$ vg的rAAV,优选 $10^{10}$ vg至 $10^{13}$ vg的rAAV,更优选包含大约 $3 \times 10^{11}$ vg或大约 $1 \times 10^{12}$ vg的rAAV。

## 密码子优化的RPGRORF 15基因及其用途

### [0001] 相关申请的交叉引用

本申请要求2020年9月2日提交的美国临时专利申请序列号63/073,843的权益,其全部公开内容通过引用并入本文。

### [0002] 经由EFS-WEB提交序列表

在2021年8月11日或其前后创建的题为“090400-5012-W0-Sequence-Listing”的计算机可读文本文件(文件大小为大约37KB)含有本申请的序列表,并特此通过引用全文并入本文。

### [0003] 发明背景

X连锁视网膜色素变性(XLRP)是一种相对严重且遗传异质性的遗传性视网膜变性。大约70%的XLRP病例是由视网膜色素变性GTP酶调节因子(RPGR)基因中的突变引起的。RPGR基因编码广泛表达的几种不同的可变剪接转录本。编码的蛋白的功能尚不清楚,但研究表明其在称为纤毛的细胞结构中起重要作用。

[0004] 一种RPGR同工型含有称为ORF15的独特3'区域,其为567个氨基酸的富含Gly和Glu的羧基末端结构域。含有RPGR基因的外显子1-13和ORF15区域的这种版本的RPGR蛋白主要在视网膜中的光感受器中表达。RPGR的ORF15区域中的突变占有XLRP病例的大约60%。

[0005] 几项临床前研究支持使用RPGRorf15的野生型cDNA来挽救XLRP疾病表型。但是,野生型序列不佳的序列稳定性对在载体生产过程中维持序列完整性提出了挑战,并且野生型序列在人光感受器中的次优表达水平对治疗XLRP的基因疗法方法是一种挑战。

### [0006] 发明概述

公开了编码人视网膜色素变性GTP酶调节因子(RPGR)蛋白的密码子优化的核酸分子。在一方面,本公开提供了包含SEQ ID NO:1的核苷酸序列的核酸或包含与SEQ ID NO:1的核苷酸序列具有至少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的核酸,并且所述核酸编码具有SEQ ID NO:2的氨基酸序列的人RPGR多肽。在一些实施方案中,提供了包含SEQ ID NO:1的核苷酸序列或由其组成的核酸。在相关实施方案中,核酸以与野生型RPGR核酸序列(例如SEQ ID NO:3)在其他方面相同的细胞中的表达水平相比更高的水平表达。

[0007] 在一些方面,如本文中所述的密码子优化的核酸分子具有相对于野生型RPGR cDNA(GenBank检索号NM\_001034853;SEQ ID NO:3)的人密码子适应指数提高的人密码子适应指数。在一些实施方案中,密码子优化的核酸分子具有至少大约0.85、至少大约0.88、或至少大约0.89的人密码子适应指数。

[0008] 在某些实施方案中,核酸含有与SEQ ID NO:3中的G/C核苷酸百分比相比更高的G/C核苷酸百分比。在另一些实施方案中,核酸含有为至多大约59%、至多大约58%、或至多大约57%的G/C核苷酸百分比。在一些方面,核酸的平均G/C含量为大约55%至大约59%、大约56%至大约58%。在一些优选实施方案中,平均G/C含量为大约57%。

[0009] 在另一些实施方案中,核酸相对于SEQ ID NO:3包含一个或多个优化的参数,其选自负性顺式作用位点(包括但不限于TATA盒和剪接位点)的去除和最佳密码子频率的提高。

[0010] 在另一个实施方案中,核酸可操作地连接至至少一个转录控制序列,优选与核酸异源的转录控制序列。在一些方面,转录控制序列是导致核酸例如在感光细胞中的细胞特异性表达的细胞或组织特异性启动子,诸如人视杆光感受器特异性人G蛋白偶联受体视紫红质激酶1(hGRK)或人光感受器间类视黄醇结合蛋白(IRBP)启动子。在优选实施方案中,转录控制序列包含人视杆光感受器特异性人G蛋白偶联受体视紫红质激酶1(hGRK)启动子。在另一些方面,转录控制序列是组成型启动子,其在许多细胞类型中导致类似的核酸表达水平(例如CAG、CBA、CMV或PGK启动子)。在优选实施方案中,转录控制序列包含如Young等人, *Investigative Ophthalmology and Visual Science*, 44(9):4076-4085(2003)中所述的人G蛋白偶联受体激酶(hGRK,也称为视紫红质激酶)启动子。在一个特别优选的实施方案中,hGRK启动子包含SEQ ID NO:4的序列或包含与其具有至少95%、至少96%、至少97%、至少98%或至少99%同一性的序列:

```
GGGCCCCAGAAGCCTGGTGGTTGTTTGTCTTCTCAGGGGAAAAGTGAGGCGGCCCTT
TGGAGGAAGGGCGGGCAGAATGATCTAATCGGATTCCAAGCAGCTCAGGGGATTGTC
TTTTTCTAGCACCTTCTTGCCACTCCTAAGCGTCCTCCGTGACCCCGGCTGGGATTTAGCCTGGTGCT
GTGTCAGCCCCGGG(SEQ ID NO:4)。
```

[0011] 在相关实施方案中,本文中提供了包含可操作地连接至表达控制序列的核酸的表达盒,所述核酸包含SEQ ID NO:1的核苷酸序列或与其具有至少90%同一性的核苷酸序列。

[0012] 在相关实施方案中,本文中提供了包含核酸的载体,所述核酸包含SEQ ID NO:1的核苷酸序列或与其具有至少90%同一性的核苷酸序列。在优选实施方案中,载体是重组腺相关(rAAV)表达载体。在一些实施方案中,rAAV载体包含天然衣壳(例如AAV血清型2或AAV血清型5或AAV血清型8的衣壳)。在另一些实施方案中,rAAV载体包含相对于天然AAV衣壳经修饰(例如包含一个或多个肽插入和/或一个或多个氨基酸取代(例如酪氨酸至苯丙氨酸)和/或氨基酸插入或氨基酸缺失)的衣壳(例如相对于血清型2、5或8的AAV衣壳包含一种或多种修饰)。

[0013] 在另一个实施方案中,本文中提供了包含核酸的宿主细胞,所述核酸包含SEQ ID NO:1的核苷酸序列或与其具有至少90%同一性的核苷酸序列。在一些方面,宿主细胞是哺乳动物细胞,包括但不限于CHO细胞、HEK293细胞、HeLa细胞、BHK21细胞、Vero细胞或V27细胞。在相关方面,宿主细胞选自CHO细胞、HEK293细胞、HEK293T细胞、HeLa细胞、BHK21细胞和Vero细胞。在另一些方面,宿主细胞是感光细胞(例如视杆(rods);视锥(rods))、视网膜神经节细胞(RGC)、胶质细胞(例如穆勒胶质细胞、小胶质细胞)、双极细胞、无长突细胞、水平细胞或视网膜色素上皮(RPE)细胞。在相关实施方案中,本公开提供了提高SEQ ID NO:2的多肽的表达的方法,其包括在核酸分子表达SEQ ID NO:2的多肽的条件下培养宿主细胞,其中相对于包含含有SEQ ID NO:3的核苷酸序列(比较序列)的参考核酸的在相同条件下培养的宿主细胞,多肽的表达提高。

[0014] 在另一个实施方案中,本公开提供了提高人受试者中SEQ ID NO:2的多肽的表达的方法,其包括向受试者施用包含与SEQ ID NO:1的核苷酸序列具有至少85%、至少90%、至少95%、至少96%、至少97%、至少98%、或至少99%同一性并且编码具有SEQ ID NO:2的氨基酸序列的多肽的核苷酸序列的分离的核酸分子,或包含此类核苷酸序列的载体,其中相对于包含SEQ ID NO:3的核苷酸序列(比较序列)的参考核酸分子或包含参考核酸分子的

载体,多肽的表达提高。

[0015] 在一些实施方案中,本公开提供了在人受试者中治疗与RGRP ORF15活性不足相关的眼部病症的方法,其包括向受试者施用本文中公开的核酸分子或载体。在一些实施方案中,视网膜病症是X连锁视网膜色素变性。

## 附图说明

图1示出了pAAV-GRK-cohRPGRorf15-SV40的限制性消化物的凝胶电泳。用各种酶消化大量制备DNA (Maxiprep DNA) 并通过琼脂糖凝胶电泳进行分析:泳道1=2-log ladder;泳道2=BsrGI-H+BglIII;泳道3=Pml+Sph-HF;泳道4=HindIII-HF+Sph-HF;泳道5=Pst。所得限制性片段与所有消化物中的预测片段匹配(泳道2 3.9、2.5、0.6kb的片段;泳道3 3.7、2.1、1.3kb的片段;泳道4 3.9、1.7和1.5kb的片段;泳道5 4.6、1.4和1.2kb的片段)。以千碱基对计的突出的2-log ladder的尺寸显示在凝胶左侧。

[0017] 图2是来自用pAAV-GRK-cohRPGRorf15-SV40转染的HEK293T细胞的细胞裂解物的蛋白印迹。用指定的一抗(Sigma;CT-15;Polyglut GT335)评估人RPGRorf15蛋白在HEK293细胞中的表达。对于每种抗体,泳道1=未转染的对照;泳道2=pAAV-GRK-cohRPGRorf15-SV40;泳道3=pAAV-PGK-cohRPGRorf15-SV40。箭头指示hRPGRorf15蛋白。分子量标志物(以千道尔顿计)显示在左侧。

[0018] 图3用包含在hGRK1启动子的控制下的SEQ ID NO:1的密码子优化的RPGRorf15的重组AAV (rAAV) 病毒粒子的转导导致XLRP-iPSC来源的感光细胞中cohRPGRorf15 (SEQ ID NO:1) 转录本水平的强烈提高。在转导后三十天,在用包含pAAV-GRK-cohRPGRorf15-SV40和SEQ ID NO:9的衣壳的rAAV以50,000的MOI转导后,对从XLRP-iPSC来源的光感受器培养物中提取的RNA进行微滴式数字PCR。确定hRPGR1-19(内部对照) 和cohRPGRorf15转录本水平并量化为高于设定阈值的拷贝/mL,并按对数标尺绘图。在转导后,密码子优化的hRPGRorf15 (SEQ ID NO:1) 转录本水平在统计学上大于hRPGR1-19.NT=未转导,MOI=感染复数,hRPGR1-19=人视网膜色素变性GTP酶调节因子外显子1-19,组成型同工型,cohRPGRorf15=密码子优化的人视网膜色素变性GTP酶调节因子开放阅读框15,SEQ ID NO:1的视网膜特异性同工型。\*与MOI 50,000hRPGR1-19相比 $p \leq 0.05$ ,†与NT cohRPGRorf15相比 $p \leq 0.05$ 。误差条±标准偏差。每个患者n=3。Y轴以对数标尺表示。

[0019] 图4用包含在hGRK1启动子的控制下的SEQ ID NO:1的密码子优化的RPGRorf15的rAAV的转导提高了XLRP光感受器培养物中的hRPGRorf15蛋白水平。以50,000的MOI转导XLRP-iPSC来源的光感受器培养物,并在转导后30天收获蛋白裂解物。SDS-PAGE和蛋白印迹显示,对两个患者而言,与未转导细胞(NT) 相比,在127kDa处的hRPGRorf15(相对于加载的对照 $\alpha$ -微管蛋白归一化) 增加。量化条带强度并在患者之间取平均值。用rAAV转导产生hRPGRorf15蛋白的显著增加。\*与NT相比 $p \leq 0.05$ 。误差条±标准偏差。每个患者n=3。

[0020] 图5在XLRP光感受器培养物中用包含在hGRK启动子控制下的SEQ ID NO:1的密码子优化的RPGRorf15的rAAV转导后hRPGRorf15的谷氨酰化。以50,000的MOI转导XLRP-iPSC来源的光感受器培养物,并在转导后30天收获蛋白裂解物。SDS-PAGE和蛋白印迹分析显示,对两个患者而言,与未转导(NT) 对照相比,127kDa蛋白hRPGRorf15的谷氨酰化(相对于加载的对照 $\alpha$ -微管蛋白归一化) 增加,量化条带强度并在患者之间取平均值。用rAAV转导产生

hRPGRorf15蛋白的谷氨酰化的显著增加。GT335=抗谷氨酰化抗体,NT=未转导,MOI=感染复数,hRPGRorf15=人视网膜色素变性GTP酶调节因子开放阅读框15,视网膜特异性同工型。\*与NT相比 $p \leq 0.05$ 。误差条±标准偏差。每个患者 $n=3$ 。

[0021] 图6组成型启动子驱动XLRP光感受器培养物中hRPGRorf15蛋白和谷氨酰化的增加。用包含在PGK启动子的控制下的SEQ ID NO:1的密码子优化的RPGRorf15的rAAV以5,000、10,000和20,000的MOI转导XLRP-iPSC来源的光感受器培养物。在转导后30天收获蛋白裂解物。SDS-PAGE和蛋白印迹显示,对患者78而言,与未转导(NT)对照相比,hRPGRorf15和127kDa处的谷氨酰化(相对于加载的对照 $\alpha$ -微管蛋白归一化)增加。量化条带强度。转导产生hRPGRorf15蛋白的显著增加。NT=未转导,MOI=感染复数,hRPGRorf15=人视网膜色素变性GTP酶调节因子开放阅读框15,视网膜特异性同工型,GT335=抗谷氨酰化抗体。\*与NT相比 $p \leq 0.05$ 。误差条±标准偏差。 $n=3$ 。

[0022] 图7是SEQ ID NO:1的密码子优化的序列和编码的氨基酸序列。

[0023] 图8是下文实施例中描述的rAAV中所含转基因盒的示意图。转基因盒包含5' AAV2 ITR、人视紫红质激酶(亦称hGRK)启动子、SEQ ID NO:1的密码子优化的人RPGRorf15cDNA、晚期SV40多聚腺苷酸化信号和3' AAV2 ITR,并具有SEQ ID NO:5的核苷酸序列。

[0024] 图9示出了如通过房水闪辉、房水细胞和玻璃体细胞所评估的那样,通过定量眼部炎症的4D-125(包含图8中显示的转基因盒和SEQ ID NO:9的衣壳蛋白)的安全性。在高剂量下观察到短暂性轻度眼部炎症的眼底征。这些变化响应于全身类固醇治疗的增加。没有被认为与4D-125相关的不利发现。在不同的检查间隔下,所有动物的IOP值均在正常限度内。ERG值和包括黄斑形态的OCT图像也在正常限度内。

[0025] 图10示出了在玻璃体内施用4D-125的NHP中在3个尸检时间点处通过qPCR测得的所选视网膜、眼和非眼组织中的载体基因组生物分布。LOD=检测下限;为了可视化目的,所有样品“BL0D”以LOD值绘制。

[0026] 图11示出了在玻璃体内施用4D-125的NHP中在3个尸检时间点处通过RT-qPCR测得的所选视网膜、眼和非眼组织中的RPGR转基因mRNA表达。LOD=检测下限;为了可视化目的,所有样品“BL0D”以LOD值绘制。

[0027] 发明详述

定义

本文中所用的“密码子适应指数”是指密码子使用偏性的量度。密码子适应指数(CAI)度量给定蛋白编码基因序列相对于参考基因集的偏差(Sharp P M和Li W H, Nucleic Acids Res.15(3):1281-95(1987))。通过确定与基因序列长度上的每个密码子相关的权重的几何平均值(以密码子测量)来计算CAI:

$$CAI = \exp \left( 1/L \sum_{l=1}^L \ln(w_l(l)) \right), \quad (I)$$

[0028] 对每个氨基酸,将其密码子的每一个的权重(以CAI计)计算为该氨基酸的所观察到的密码子频率( $f_i$ )和同义密码子频率( $f_j$ )之间的比率:

$$w_i = \frac{f_i}{\max(f_j)} \quad ij \in [\text{synonymous codons for amino acid}] \quad (\text{II})$$

[0029] 术语“分离的”表示已经从其原始环境(其天然存在的环境)中取出的生物材料(细胞、核酸或蛋白)。例如,以天然状态存在于植物或动物中的多核苷酸不是分离的,但是与其天然存在的相邻核酸分离的相同多核苷酸被认为是“分离的”。

[0030] 术语“4D-125”是指重组AAV粒子,其包含(i)含有SEQ ID NO:9的氨基酸序列的衣壳蛋白和含有SEQ ID NO:5的核苷酸序列的异源核酸。

[0031] 术语“R100”是指包含SEQ ID NO:9的氨基酸序列的变体AAV衣壳蛋白。

[0032] 如本文中所用,“编码区”或“编码序列”是由可翻译成氨基酸的密码子组成的多核苷酸的一部分。尽管“终止密码子”(TAG、TGA或TAA)通常不翻译成氨基酸,但其可以被认为是编码区的一部分,但任何侧翼序列,例如启动子、核糖体结合位点、转录终止子、内含子等,不是编码区的一部分。编码区的边界通常由编码所得多肽的氨基末端的5'末端处的起始密码子和编码所得多肽的羧基末端的3'末端处的翻译终止密码子来确定。两个或更多个编码区可以存在于单个多核苷酸构建体中,例如在单个载体上,或存在于分开的多核苷酸构建体中,例如在分开的(不同的)载体上。因此,单个载体可以仅含有单个编码区,或包含两个或更多个编码区。

[0033] 如本文中所用,术语“调控区”是指位于编码区的上游(5'非编码序列)、内部或下游(3'非编码序列)的核苷酸序列,并且其影响相关编码区的转录、RNA加工、稳定性或翻译。调控区可包括启动子、翻译前导序列、内含子、多聚腺苷酸化识别序列、RNA加工位点、效应子结合位点和茎-环结构。如果编码区意在在真核细胞中表达,则多聚腺苷酸化信号和转录终止序列将通常位于编码序列的3'。

[0034] 本文中使用的术语“核酸”可与“多核苷酸”或“核酸分子”互换,并且意在指核苷酸的聚合物。

[0035] 编码基因产物(例如多肽)的多核苷酸可包括与一个或多个编码区可操作地关联的启动子和/或其他转录或翻译控制元件。在可操作关联中,基因产物(例如多肽)的编码区与一个或多个调控区以将基因产物的表达置于一个或多个调控区的影响或控制下的方式关联。例如,如果启动子功能的诱导导致编码由编码区编码的基因产物的mRNA的转录,并且如果启动子和编码区之间的连接的性质不干扰启动子指导基因产物表达的能力或不干扰DNA模板转录的能力,则编码区和启动子“可操作地关联”。除启动子外,其他转录控制元件(例如增强子、操纵子、阻遏子(repressor)和转录终止信号)也可以与编码区可操作地关联以指导基因产物表达。

[0036] “转录控制序列”是指诸如启动子、增强子、终止子等DNA调控序列,其提供编码序列在宿主细胞中的表达。多种转录控制区是本领域技术人员已知的。这些包括但不限于在脊椎动物细胞中起作用的转录控制区,诸如但不限于来自巨细胞病毒(与内含子A结合的即刻早期启动子)、猿猴病毒40(早期启动子)和逆转录病毒(诸如劳斯肉瘤病毒)的启动子和增强子区段。其他转录控制区包括来源于脊椎动物基因(诸如肌动蛋白、热休克蛋白、牛生长激素和兔β-珠蛋白)的那些,以及能够控制真核细胞中基因表达的其他序列。另外的合适转录控制区包括组织特异性启动子和增强子,以及淋巴因子诱导型启动子(例如可由干扰素或白介素诱导的启动子)。

[0037] 类似地,多种翻译控制元件是本领域普通技术人员已知的。这些包括但不限于核糖体结合位点、翻译起始和终止密码子以及来源于小核糖核酸病毒的元件(特别是内部核糖体进入位点或IRES,也称为CITE序列)。

[0038] 本文中使用的术语“表达”是指多核苷酸产生基因产物(例如RNA或多肽)的过程。其包括但不限于将多核苷酸转录成信使RNA(mRNA)、转运RNA(tRNA)、小发夹RNA(shRNA)、小干扰RNA(siRNA)或任何其他RNA产物,以及将mRNA翻译成多肽。表达产生“基因产物”。如本文中所示,基因产物可以是核酸(例如通过基因转录产生的信使RNA),或从转录本翻译的多肽。本文中描述的基因产物进一步包括具有转录后修饰(例如多聚腺苷酸化或剪接)的核酸,或具有翻译后修饰(例如甲基化、糖基化、添加脂质、与其他蛋白亚基缔合、或溶蛋白性裂解(proteolytic cleavage))的多肽。

[0039] “载体”是指用于将核酸克隆和/或转移到宿主细胞中的任何媒介。载体可以是复制子,另一核酸区段可以与其连接以引起所连接的区段的复制。术语“载体”包括用于体外、离体或体内将核酸引入到细胞中的病毒和非病毒媒介。大量载体是本领域已知和使用的,包括例如质粒、修饰的真核病毒或修饰的细菌病毒。将多核苷酸插入合适的载体可以通过将合适的多核苷酸片段连接到具有互补粘性末端的所选载体中来实现。

[0040] 载体可以工程化以编码选择性标志物或报告子,所述标志物或报告子提供已掺入载体的细胞的选择或鉴定。选择性标志物或报告子的表达使得能够鉴定和/或选择掺入并表达载体上所含其他编码区的宿主细胞。本领域中已知和使用的选择性标志物基因的实例包括:提供对氨苄青霉素、链霉素、庆大霉素、卡那霉素、潮霉素、双丙氨膦除草剂、磺胺药物等的抗性的基因;以及用作表型标志物的基因,即花青素调控基因、异戊烯基转移酶(isopentenyl transferase)基因等。本领域中已知和使用的报告子的实例包括:荧光素酶(Luc)、绿色荧光蛋白(GFP)、氯霉素乙酰基转移酶(CAT)、-半乳糖苷酶(LacZ)、-葡萄糖醛酸糖苷酶(Gus)等。选择性标志物也可以被认为是报告子。

[0041] 可以使用的真核病毒载体包括但不限于腺病毒载体、逆转录病毒载体、腺相关病毒载体、痘病毒(例如痘苗病毒载体)、杆状病毒载体或疱疹病毒载体。非病毒载体包括质粒、脂质体、带电脂质(electrically charged lipid)(细胞转染素)、DNA-蛋白复合物和生物聚合物。

[0042] “启动子”和“启动子序列”可互换使用,并且是指能够控制编码序列或功能性RNA的表达的DNA序列。通常,编码序列位于启动子序列的3'。启动子可以全部来源于天然基因,或由来源于自然界中发现的不同启动子的不同元件构成,或甚至包含合成DNA区段。本领域技术人员应当理解的是,不同的启动子可以指导基因在不同组织或细胞类型中、或在不同发育阶段、或响应于不同的环境或生理条件的表达。引起基因在大多数时间在大多数细胞类型中表达的启动子通常被称为“组成型启动子”。引起基因在特定细胞类型中表达的启动子通常被称为“细胞特异性启动子”或“组织特异性启动子”。引起基因在特定的发育阶段或细胞分化阶段表达的启动子通常被称为“发育特异性启动子”或“细胞分化特异性启动子”。在用诱导启动子的试剂、生物分子、化学品、配体、光等对细胞进行暴露或处理后被诱导并引起基因表达的启动子通常被称为“诱导型启动子”或“可调节启动子”。进一步认识到,由于在大多数情况下调控序列的确切边界尚未完全确定,因此不同长度的DNA片段可具有相同的启动子活性。

[0043] 术语“质粒”是指染色体外元件,其通常携带并非细胞中心代谢的一部分的基因,并通常为环状双链DNA分子的形式。此类元件可以是来源于任何来源的单链或双链DNA或RNA的线性、环状或超螺旋的自主复制序列、基因组整合序列、噬菌体或核苷酸序列,其中许多核苷酸序列已经连接或重组成独特的构建体,该构建体能够将启动子片段和所选基因产物的DNA序列与适当的3'非翻译序列一起引入到细胞中。

[0044] 多核苷酸或多肽与另一多核苷酸或多肽具有一定百分比的“序列同一性”,意味着在比对时,当比较两个序列时,该百分比的碱基或氨基酸是相同的。序列相似性可以以许多不同的方式确定。为了确定序列同一性,可以使用包括可通过万维网在ncbi.nlm.nih.gov/BLAST/获得的BLAST的方法和计算机程序来比对序列。另一比对算法是FASTA,其可以在来自Madison,Wis.,USA的Genetics Computing Group(GCG)包中获得。用于比对的其他技术描述在Methods in Enzymology,第266卷:Computer Methods for Macromolecular Sequence Analysis(1996),ed.Doolittle,Academic Press,Inc中。特别引人关注的是允许序列中的空位的比对程序。Smith-Waterman是允许序列比对中的空位的一种类型的算法。参见Meth.Mol.Biol.70:173-187(1997)。此外,使用Needleman和Wunsch比对方法的GAP程序可用于比对序列。参见J.Mol.Biol.48:443-453(1970)。

[0045] 在一个实施方案中,本发明提供了包含编码SEQ ID NO:2的多肽(人RGPR ORF15)的核苷酸序列的修饰核酸分子,其中核酸序列已经经过密码子优化。在另一个实施方案中,编码SEQ ID NO:2的多肽并进行密码子优化的起始核酸序列具有SEQ ID NO:3所示的核苷酸序列。在优选实施方案中,编码SEQ ID NO:2的多肽的序列针对人表达进行密码子优化。SEQ ID NO:1是针对人表达进行优化的SEQ ID NO:3的密码子优化版本:ATGAGAGAACCCGAG GAACTGATGCCCGACTCTGGCGCCGTGTTTACCTTCGGCAAGAGCAAGTTCGCCGAGAACAACCCCGGCAAGTTCTG GTTCAAGAACGACGTGCCAGTGCACCTGAGCTGCGGAGATGAACACTCTGCCGTGGTCACCGGCAACAACAAGCTGT ACATGTTTCGGCAGCAACAACCTGGGGCCAGCTCGGCCTGGGATCTAAGTCTGCCATCAGCAAGCCTACCTGCGTGAAG GCCCTGAAGCCTGAGAAAGTGAACTGGCCGCTGCGGCAGAAATCACACCCTGGTTTCTACCGAAGGCGGCAATGT GTATGCCACCGGCGGAAACAATGAGGGACAGCTTGGACTGGGCGACACCGAGGAAAGAAACACCTTCCACGTGATCA GCTTTTTTACCAGCGAGCACAAGATCAAGCAGCTGAGCGCCGGCTCTAATACCTCTGCCGCTCTGACAGAGGACGGC AGACTGTTTATGTGGGGCGACAATTCTGAGGGCCAGATCGGACTGAAGAACGTGTCCAATGTGTGCGTGCCCCAGCA AGTGACAATCGGCAAGCCTGTGTCTTGGATCAGCTGCGGCTACTACCACAGCGCCTTTGTGACAACCGATGGCGAGC TGTATGTGTTTCGGCGAGCCAGAGAATGGCAAGCTGGGACTGCCTAACAGCTGCTGGGCAATCACAGAACCCCTCAG CTGGTGTCTGAGATCCCCGAAAAAGTGATCCAGGTGGCCTGTGGCGGAGAGCACACAGTGGTGTGACAGAGAATGC CGTGTACACCTTTGGCCTGGGCCAGTTTGGACAACCTCGGACTGGGAACCTTCCTGTTTCGAGACAAGCGAGCCCAAAG TGATCGAGAACATCCGGGACCAGACCATCAGCTACATCAGCTGTGGCGAGAACCACACAGCCCTGATCACAGACATC GGCCTGATGTACACATTCGGCGACGGAAGGCATGGAAAGCTCGGACTTGGCCTGGAAAACCTTCACCAACCACTTCAT CCCTACGCTGTGCAGCAACTTCCTGCGGTTTCATTGTGAAGCTGGTGGCCTGCGGAGGATGCCACATGGTGGTTTTTG CTGCCCCCTCACAGAGGCGTGGCCAAAGAGATTGAGTTCGACGAGATCAACGATACCTGCCTGAGCGTGGCCACCTTC CTGCCTTACAGCAGCCTGACATCTGGCAACGTGCTGCAGAGGACACTGAGCGCCAGAAT

GCGCAGACGGGAAAGAGAGAGAAGCCCCGACAGCTTCAGCATGAGAAGAACCCTGCC

TCCAATCGAGGGCACACTGGGCCTGTCTGCCTGCTTTCTGCCTAACAGCGTGTTCCTCCA

GATGCAGCGAGAGAAACCTGCAAGAGAGCGTGTGAGCGAGCAGGATCTGATGCAGCC

TGAGGAACCCGACTACCTGCTGGACGAGATGACCAAAGAGGCCGAGATCGACAACAGC  
AGCACAGTGGAAAGCCTGGGCGAGACAACCGACATCCTGAACATGACCCACATCATGA  
GCCTGAACAGCAACGAGAAGTCTCTGAAGCTGAGCCCCGTGCAGAAGCAGAAGAAGC  
AGCAGACCATCGGCGAGCTGACACAGGATACTGCCCTGACCGAGAACGACGACAGCGA  
CGAGTACGAAGAGATGAGCGAGATGAAGGAAGGCAAGGCCTGCAAGCAGCACGTGTC  
CCAGGGCATCTTTATGACCCAGCCTGCCACCACCATCGAGGCCTTTTCCGACGAGGAAG  
TGGAATCCCCGAGGAAAAAGAGGGCGCCGAGGACAGCAAAGGCAACGGCATTGAGG  
AACAAGAGGTGGAAGCCAACGAAGAGAACGTGAAGGTGCACGGCGGACGGAAGAA  
AAGACCGAGATCCTGAGCGACGACCTGACCGATAAGGCCGAGGTTTCCGAGGGCAAG  
CCAAGTCTGTGGGAGAAGCCGAGGATGGACCTGAAGGCCGCGGAGATGGAACCTGTGA  
AGAAGGATCTAGCGGAGCCGAGCACTGGCAGGATGAGGAACGCGAGAAGGGCGAGAA  
AGACAAAGGCAGAGGCGAGATGGAAAGACCCGGCGAGGGCGAAAAAGAGCTGGCCG  
AGAAAGAGGAATGGAAGAAACGCGACGGCGAAGAACAAGAGCAGAAAGAAAGAGAG  
CAGGGCCACCAGAAAGAACGGAATCAAGAGATGGAAGAAGGCGGCGAGGAAGAACAC  
GGCGAAGGGGAAGAAGAGGAAGGCGACCGAGAGGAAGAAGAAGAGAAAGAAGGCG  
AAGGCAAAGAAGAAGGCGAGGGCGAAGAGGTGGAAGGCGAGCGTGAAAAAGAAGAG  
GGCGAACGCAAGAAAGAAGAACGCGCCGAAAAGAGGAAAAAGGCGAGGAAGAGGG  
CGACCAAGGCGAAGGCGAGGAAGAAGAACTGAAGGCAGAGGGGAAGAGAAAGAGG  
AAGGCGGCGAAGTCAAGGCGGAGAGGTTGAAGAAGGCAAAGGCGAGCGAGAAGAG  
GAAGAAGAAGAAGGCGAAGGCGAGGAAGAGGAAGGCGAAGGCGAAGAGGAAGAAG  
GCGAAGGGGAAGAAGAAGAAGGCGAAGGCAAGGGCGAAGAGGAGGGCGAAGAAGGC  
GAGGGCGAAGAGGAGGGCGAAGAAGGCGAAGGCGAGGGCGAAGAAGAAGAAGGCGA  
AGGCGAAGGCGAGGAAGAAGGCGAAGGCGAAGGGGAAGAAGAGGAAGGCGAAGGCG  
AAGGCGAAGAAGAAGGCGAAGGCGAGGGCGAAGAGGAAGAAGGCGAAGGCAAAGGG  
GAAGAAGAAGGCGAGGAAGGCGAAGGCGAAGGCGAGGAAGAAGAAGGCGAAGGCGA  
GGGCGAAGATGGCGAAGGCGAAGGCGAAGAGGAAGAGGGCGAGTGGGAGGGCGAAG  
AAGAGGAAGGCGAAGGCGAGGGCGAAGAGGAAGGCGAAGGCGAGGGCGAAGAAGGC  
GAAGGCGAAGGCGAGGAAGAGGAAGGCGAAGGCGAAGGGGAAGAAGAAGAGGGCG  
AAGAAGAAGGCGAAGAGGAAGGCGAAGGGGAAGAAGAAGGCGAAGGCGAAGGCGA  
AGAAGAGGAAGAGGGCGAAGTTGAAGGCGAGGTTGAGGGCGAAGAAGGCGAAGGCG  
AAGGGGAAGAAGAAGAAGGCGAGGAAGAAGGGGAAGAGAGAGAAAAAGAAGGCGA  
GGGCGAAGAAAACCGCCGAACCGCGAAGAGGAAGAGGAAGAAGAGGGCAAGTACC  
AAGAGACTGGCGAGGAAGAGAACGAGCGGCAGGATGGCGAAGAGTACAAGAAGGTGT  
CCAAGATCAAGGGCAGCGTGAAGTACGGCAAGCACAAGACCTACCAGAAGAAGTCCGT  
CACCAACACGCAAGGCAATGGAAAAGAACAGCGGAGCAAGATGCCCCGTGCAGTCCAA  
GAGGCTGCTGAAGAATGGCCCTAGCGGCAGCAAGAAATTCTGGAACAATGTGCTGCCCCACTACCTCG

AGCTGAAGTGA (SEQ ID NO:1)。

[0046] 在一些实施方案中,提供了编码人RPGR ORF15的密码子优化的序列,其缺少SEQ ID NO:1的TGA终止密码子(即由SEQ ID NO:1的核苷酸1-3456组成)。

[0047] 在一方面,本公开提供了包含SEQ ID NO:1的核苷酸序列的多核苷酸,或包含与SEQ ID NO:1的核苷酸序列具有至少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的多核苷酸,并且所述多核苷酸编码具有SEQ ID NO:2的氨基酸序列的人RPGR多肽:

MREPEELMPDSGAVFTFGKSKFAENNPCKFWFKNDVPVHLSCGDEHSAVVTGNNKLYMFG  
SNNWQQLGLGSKSAISKPTCVKALKPEKVKLAACGRNHTLVSTEGGNVYATGGNNEGQLG  
LGDTEERNTFHVISFFTSEHKIKQLSAGSNTSAALTEDGRLFMWGDNSEGQIGLKNVSNVC  
VPQQVTIGKPVSWISCGYYHSAFVTTDGELYVFGEPENGKLGLPNQLLGNHRTPLVSEIPE  
KVIQVACGGEHTTVLTENAVYTFGLGQFGQLGLGTFLFETSEPKVIENIRDQTISYISCGENH  
TALITDIGLMTFGDGRHGKLGLGLENFTNHFIPTLCSNFLRFIVKLVACGGCHMVVFAAPH  
RGVAKEIEFDEINDTCLSVATFLPYSSLTSGNVLQRTLSARMRRRERERSPDSFSMRRTLPPIE  
GTLGLSACFLPNSVFPKCSERNLQESVLSEQDLMQPEEPDYLLDEMTKEAEIDNSSTVESLG  
ETTDILNMTHIMSLNSNEKSLKLSPVQKQKKQQTIGELTQDTALTENDDSDEYEEMSEMKE  
GKACKQHVSQGIFMTQPATTIEAFSDEEVEIPEEKEGAEDSKGNGIEEQEVEANEENVKVHG  
GRKEKTEILSDDLTDKAEVSEGKAKSVGEAEDGPEGRGDGTCEEGSSGAEHWQDEEREKG  
EKDKGRGEMERPGEKEKELAEKEEWKKRDGEEQEQKEREQGHQKERNQEMEEGGEEEH  
GEGEEEEGDREEEEEKEGEGKEEGEGEEVEGEREKEEGERKKKEERAGKEEKGEEEGDQGE  
GEEEETEGRGEEKEEGGEGVEGGEVEEGKGEREEEEEGEGEEEEEGEGEEEEEGEGEEEEEG  
KGGEEEGEGEGEEEEEGEGEGEEEEEGEGEGEEEEEGEGEGEEEEEGEGEGEEEEEG  
EGKGEEEGEGEGEEEEEGEGEGEDGEGEGEEEEEGEWEGEEEEEGEGEGEEEEEGEGEGEEG  
EGEGEEEEEGEGEGEEEEEGEGEGEEEEEGEGEGEEEEEGEGEGEEEEEGEGEGEEEE  
GEEEGEEREKEGEGEENRRNREEEEEEGKYQETGEEENERQDGEEYKKVSKIKGSVKYGHKTYQKK

SVTNTQNGKEQRSKMPVQSKRLLKNGPSGSKKFWNNVLPHYLELK (SEQ ID NO:2)。

[0048] 术语“密码子优化的”在其涉及用于转化各种宿主的核酸分子的基因或编码区时,是指改变核酸分子的基因或编码区中的密码子以反映宿主生物体的典型密码子使用而不改变由DNA编码的多肽。此类优化包括用在该生物体的基因中更频繁使用的一个或多个密码子替换至少一个、或超过一个、或大量的密码子。

[0049] 包含编码任何多肽链的氨基酸的密码子的核苷酸序列中的偏差允许编码基因的序列中的变化。由于每个密码子由三个核苷酸组成,并且构成DNA的核苷酸限于四个特定碱基,因此存在64种可能的核苷酸组合,其中61种编码氨基酸(其余三个密码子编码终止翻译的信号)。显示哪些密码子编码哪些氨基酸的“遗传密码”作为表1转载于本文中。因此,许多氨基酸由超过一个密码子指定。例如,氨基酸丙氨酸和脯氨酸由四个三联体编码,丝氨酸和精氨酸由六个,而色氨酸和蛋氨酸仅由一个三联体编码。这种简并性允许DNA碱基组成在宽范围内变化,而不改变由DNA编码的蛋白的氨基酸序列。

[0050] 表-US-00001表1标准遗传密码

|   |   |   |   |   |     |         |     |         |     |         |     |
|---|---|---|---|---|-----|---------|-----|---------|-----|---------|-----|
| T | C | A | G | T | TTT | Phe (F) | TCT | Ser (S) | TAT | Tyr (Y) | TGT |
| C | T | T | C | C | TTC | Phe (F) | TCC | Ser (S) | TAC | Tyr (Y) | TGC |
| T | T | A | T | A | TTA | Leu (L) | TCA | Ser (S) | TAA | Stop    | TGA |
| T | T | G | T | G | TTG | Leu (L) | TCG | Ser (S) | TAG | Stop    | TGG |
| T | C | T | T | C | CTT | Leu (L) | CCT | Pro (P) | CAT | His (H) | CGT |
| C | T | C | A | C | CTC | Leu (L) | CCC | Pro (P) | CAC | His (H) | CGC |
| C | T | A | C | A | CTA | Leu (L) | CCA | Pro (P) | CAA | Gln (Q) | CGA |
| C | T | G | C | G | CTG | Leu (L) | CCG | Pro (P) | CAG | Gln (Q) | CGG |
| A | T | T | A | T | ATT | Ile (I) | ACT | Thr (T) | AAT |         |     |

Asn(N) AGT Ser(S) ATC Ile(I) ACC Thr(T) AAC Asn(N) AGC Ser(S) ATA Ile(I) ACA Thr(T) AAA Lys(K) AGA Arg(R) ATG Met(M) ACG Thr(T) AAG Lys(K) AGG Arg(R) G GTT Val(V) GCT Ala(A) GAT Asp(D) GGT Gly(G) GTC Val(V) GCC Ala(A) GAC Asp(D) GGC Gly(G) GTA Val(V) GCA Ala(A) GAA Glu(E) GGA Gly(G) GTG Val(V) GCG Ala(A) GAG Glu(E) GGG Gly(G)。

[0051] 许多生物体表现出使用特定密码子来编码特定氨基酸在生长的肽链中的插入的偏性。密码子偏爱性或密码子偏性(生物体之间密码子使用的差异)由遗传密码的简并性提供,并在许多生物体中得到充分证明。密码子偏性通常与信使RNA(mRNA)的翻译效率相关,而信使RNA(mRNA)的翻译效率又被认为尤其取决于被翻译的密码子的性质和特定转运RNA(tRNA)分子的可用性。细胞中所选tRNA的优势通常反映了肽合成中最频繁使用的密码子。因此,可以基于密码子优化来定制基因以便在给定生物体中获得最佳基因表达。

[0052] 鉴于可用于多种动物、植物和微生物物种的大量基因序列,已经计算了密码子使用的相对频率。密码子使用表可以在例如“密码子使用数据库(Codon Usage Database)”处获得,所述“密码子使用数据库”可以在[www.kazusa.or.jp/codon/](http://www.kazusa.or.jp/codon/)(2012年6月18日访问)处获得。参见Nakamura,Y.等人,Nucl.Acids Res.28:292(2000)。

[0053] 以优化的频率随机分配密码子以编码给定的多肽序列可以通过计算每个氨基酸的密码子频率,并随后将密码子随机分配给多肽序列来手动完成。此外,可以使用各种算法和计算机软件程序来计算最佳序列。

#### [0054] 非病毒载体

在一些实施方案中,提供了包含本文中描述的修饰核酸的非病毒载体(例如表达质粒)。优选地,非病毒载体是包含SEQ ID NO:1的核酸序列或与其具有至少90%同一性的序列的质粒。

#### [0055] 病毒载体

在优选实施方案中,提供了包含本文中描述的修饰(密码子优化)核酸的病毒载体。优选地,病毒载体包含可操作地连接至表达控制序列的SEQ ID NO:1的核酸序列,或与其具有至少90%同一性的序列。合适的病毒载体的实例包括但不限于腺病毒、逆转录病毒、慢病毒、疱疹病毒和腺相关病毒(AAV)载体。

[0056] 在一个优选实施方案中,病毒载体包括细小病毒基因组的一部分,诸如rep和cap基因缺失和/或被修饰的RPGRorf15基因序列及其相关表达控制序列替换的AAV基因组。修饰的人RPGRorf15基因序列通常邻近一个或两个(即侧接)对病毒复制足够的AAV TR或TR元件插入(Xiao等人,1997,J.Virol.71(2):941-948),替代编码病毒rep和cap蛋白的核酸。还可以包括适合用于促进修饰的RPGRorf15基因序列在靶细胞中的组织特异性表达的其他调控序列。

[0057] 在一些优选实施方案中,AAV病毒载体包含核酸,所述核酸从5'至3'包含:(a)AAV2末端重复,(b)hGRK启动子,(c)如本文中所述的密码子优化的RPGRorf15基因,(d)多聚腺苷酸化序列,和(e)AAV2末端重复。在一个特别优选的实施方案中,AAV病毒载体包含核酸(转基因盒),所述核酸(转基因盒)包含SEQ ID NO:5的序列或与其具有至少90%、至少95%、至少98%或至少99%同一性的序列:

```
TTGGCCACTC CCTCTCTGCG CGCTCGCTCG CTCACTGAGG CCGGGCGACC AAAGGTCGCC 60
CGACGCCCCG GCTTTGCCCG GGCGGCCTCA GTGAGCGAGC GAGCGCGCAG AGAGGGAGTG 120
GCCAACTCCA TCACTAGGGG TTCCTATCGA TTGAATTCCC CGGGGATCCA CTAGTTATTA 180
ATAGTAATCA ATTACGGGGT CATTAGTTCA TAGCCCATAT ATGGAGTTCC GCGTTACATA 240
ACTTACGGTA AATGGCCCGC CTGGCTGACC GCCCAACGAC CCCCGCCCAT TGACGTCAAT 300
AATGACGTAT GTTCCCATAG TAACGCCAAT AGGGACTTTC CATTGACGTC AATGGGTGGA 360
GTATTTACGG TAAACTGCCC ACTTGGCAGT ACATCAAGTG TATCATATGC CAAGTACGCC 420
CCCTATTGAC GTCAATGACG GTAAATGGCC CGCCTGGCAT TATGCCCAGT ACATGACCTT 480
ATGGGACTTT CCTACTTGGC AGTACATCTA CGTATTAGTC ATCGCTATTA CCATGGTCGA 540
GGTGAGCCCC ACGTTCTGCT TCACTCTCCC CATCTCCCCC CCCTCCCCAC CCCCATTATT 600
GTATTTATTT ATTTTTTAAT TATTTTGTGC AGCGATGGGG GCGGGGGGGG GGGGGGGGCG 660
CGCGCCAGGC GGGGCGGGGC GGGGCGAGGG GCGGGGCGGG GCGAGGCGGA GAGGTGCGGC 720
GGCAGCCAAT CAGAGCGGCG CGCTCCGAAA GTTTCCTTTT ATGGCGAGGC GGCGGCGGCG 780
GCGGCCCTAT AAAAAGCGAA GCGCGCGGCG GCGGGGAGT CGCTGCGACG CTGCCTTCGC 840
```

CCCGTGCCCC GCTCCGCCGC CGCCTCGCGC CGCCCCCCCC GGCTCTGACT GACCGCGTTA 900  
CTCCCAACAGG TGAGCGGGCG GGACGGCCCT TCTCCTCCGG GCTGTAATTA GCGCTTGGTT 960  
TAATGACGGC TTGTTTCTTT TCTGTGGCTG CGTGAAAGCC TTGAGGGGCT CCGGGAGGGC 1020  
CCTTTGTGCG GGGGGAGCGG CTCGGGGGGT GCGTGCGTGT GTGTGTGCGT GGGGAGCGCC 1080  
GCGTGCGGCT CCGCGCTGCC CGGCGGCTGT GAGCGCTGCG GGCGCGGCGC GGGGCTTTGT 1140  
GCGCTCCGCA GTGTGCGCGA GGGGAGCGCG GCCGGGGGCG GTGCCCCGCG GTGCGGGGGG 1200  
GGCTGCGAGG GGAACAAAGG CTGCGTGCGG GGTGTGTGCG TGGGGGGGTG AGCAGGGGGT 1260  
GTGGGCGCGT CGGTCGGGCT GCAACCCCCC CTGCACCCCC CTCCCCGAGT TGCTGAGCAC 1320  
GGCCCCGCTT CGGGTGCGGG GCTCCGTACG GGGCGTGGCG CGGGGCTCGC CGTGCCGGGC 1380  
GGGGGGTGGC GGCAGGTGGG GGTGCCGGGC GGGGCGGGGC CGCCTCGGGC CGGGGAGGGC 1440  
TCGGGGGAGG GGC GCGCGG CGCCCCGAGC GCCGGCGGCT GTCGAGGCGC GGCGAGCCGC 1500  
AGCCATTGCC TTTTATGGTA ATCGTGCGAG AGGGCGCAGG GACTTCCTTT GTCCCAAATC 1560  
TGTGCGGAGC CGAAATCTGG GAGGCGCCGC CGCACCCCTT CTAGCGGGCG CGGGGCGAAG 1620  
CGGTGCGGCG CCGGCAGGAA GGAAATGGGC GGGGAGGGCC TTCGTGCGTC GCCGCGCCGC 1680  
CGTCCCCTTC TCCCTCTCCA GCCTCGGGGC TGTCCGCGGG GGGACGGCTG CCTTCGGGGG 1740  
GGACGGGGCA GGGCGGGGTT CGGCTTCTGG CGTGTGACCG GCGGCTCTAG AGCCTCTGCT 1800  
AACCATGTTC ATGCCTTCTT CTTTTTCCTA CAGTCTAGAG TCGACCTGCA GAAGCTTCCA 1860  
CCATGGCTGA TACACTGCCT TCTGAGTTTG ATGTGATCGT GATTGGAAT GGACTGCCTG 1920  
AGAGTATTAT TGCTGCTGCT TGTAGTAGAA GCGGCCGGAG AGTGCTGCAC GTGGACAGCA 1980  
GATCCTACTA TGGCGGCAAC TGGGCCTCTT TCAGCTTTTC CGGCCTGCTG AGCTGGCTGA 2040  
AGGAGTACCA GGAGAACTCC GACATCGTGT CTGATAGCCC CGTGTGGCAG GACCAGATCC 2100  
TGGAGAATGA GGAGGCCATC GCCCTGTCCA GGAAGGATAA GACCATCCAG CACGTGGAGG 2160  
TGTTCTGCTA TGCCAGCCAG GACCTGCACG AGGATGTGGA GGAGGCAGGC GCCCTGCAGA 2220  
AGAACCACGC CCTGGTGACC TCCGCCAATT CTACAGAGGC CGCCGACTCC GCCTTTCTGC 2280  
CTACCGAGGA TGAGTCCCTG TCTACAATGT CTTGTGAGAT GCTGACCGAG CAGACACCTA 2340  
GCTCCGATCC AGAGAACGCC CTGGAGGTCA ATGGCGCCGA GGTGACCGGC GAGAAGGAGA 2400  
ACCACTGCGA CGATAAGACC TGCCTGCCAA GCACATCCGC CGAGGACATG TCCGAGAACG 2460  
TGCCTATCGC CGAGGATACC ACAGAGCAGC CAAAGAAGAA TCGCATACA TACAGCCAGA 2520  
TCATCAAGGA GGGCAGGCGC TTCAATATCG ACCTGGTGTC TAAGCTGCTG TACAGCCGGG 2580  
GCCTGCTGAT CGATCTGCTG ATCAAGAGCA ACGTGTCCCG CTATGCCGAG TTCAAGAATA 2640  
TCACCAGAAAT CCTGGCCTTT CGGGAGGGAA GAGTGGAGCA GGTGCCCTGC AGCAGAGCCG 2700  
ACGTGTTCAA CTCCAAGCAG CTGACAATGG TGGAGAAGAG GATGCTGATG AAGTTCCTGA 2760  
CATTTTGTAT GGAGTACGAG AAGTATCCAG ATGAGTACAA GGGCTATGAG GAGATCACCT 2820  
TTTACGAGTA TCTGAAGACC CAGAAGCTGA CACCCAATCT GCAGTACATC GTGATGCACT 2880  
CCATCGCCAT GACCTCTGAG ACAGCCTCTA GCACCATCGA CGGCCTGAAG GCCACAAAGA 2940  
ACTTCCTGCA CTGCCTGGGC CGGTACGGCA ATACACCCTT CCTGTTTCCT CTGTATGGCC 3000  
AGGGCGAGCT GCCCCAGTGC TTCTGTAGAA TGTGCGCCGT GTTTGGCGGC ATCTATTGCC 3060  
TGAGGCACTC TGTGCAGTGT CTGGTGGTGG ACAAGGAGAG CCGCAAGTGT AAGGCCATCA 3120  
TCGATCAGTT TGGCCAGCGG ATCATCTCTG AGCACTTCCT GGTGGAGGAC AGCTACTTTC 3180  
CTGAGAACAT GTGCTCCAGG GTGCAGTATC GCCAGATCAG CCGGGCCGTG CTGATCACCG 3240  
ATAGATCCGT GCTGAAGACA GACAGCGATC AGCAGATCAG CATCCTGACC GTGCCAGCAG 3300  
AGGAGCCAGG CACCTTCGCC GTGAGAGTGA TCGAGCTGTG CTCCTCTACC ATGACATGTA 3360  
TGAAGGGCAC CTACCTGGTG CACCTGACCT GCACAAGCTC CAAGACAGCC CGCGAGGACC 3420  
TGGAGAGCGT GGTGCAGAAG CTGTTCTGTG CCTACACCGA GATGGAGATC GAGAACGAGC 3480  
AGGTGGAGAA GCCTAGAATC CTGTGGGCCC TGTACTTCAA CATGAGAGAC TCTAGCGATA 3540  
TCTCTAGGAG CTGTTACAAC GATCTGCCCT CTAACGTGTA CGTGTGCAGC GGACCTGACT 3600  
GTGGCCTGGG AAACGATAAT GCCGTGAAGC AGGCCGAGAC ACTGTTCCAG GAGATTTGCC 3660  
CTAACGAGGA CTTTTGTCCC CCTCCACCCA ATCCAGAGGA TATCATCTG GACGGCGATT 3720  
CCCTGCAGCC AGAGGCCTCT GAGTCTCTCTG CCATCCCCGA GGCCAATAGC GAAACATTCA 3780  
AAGAAAGCAC AAATCTGGGA AACCTGGAAG AAAGTAGTGA GTAAGCCTCG AGCAGCGCTG 3840  
CTCGAGAGAT CTGCGGCCGC GAGCTCGGGG ATCCAGACAT GATAAGATAC ATTGATGAGT 3900  
TTGGACAAAC CACAAC TAGA ATGCAGTGAA AAAAATGCTT TATTTGTGAA ATTTGTGATG 3960  
CTATTGCTTT ATTTGTAACC ATTATAAGCT GCAATAAACA AGTTAACAAC AACAATTGCA 4020  
TTCATTTTAT GTTTTCAGGT CAGGGGGAGG TGTGGGAGGT TTTTAAAGC AAGTAAACC 4080  
TCTACAAATG TGGTATGGCT GATTATGATC AATGCATCCT AGCCGGAGGA ACCCCTAGTG 4140  
ATGGAGTTGG CCACTCCCTC TCTGCGCGCT CGCTCGCTCA CTGAGGCCGC CCGGGCAAAG 4200  
CCGGGCGTGC GGGCGACCTT TGGTCGCCG GCCTCAGTGA GCGAGCGAGC GCGCAGAGAG 4260  
GGAGTGGCCA A 4271

(SEQ ID NO:5)。

[0058] 在下表2中鉴定了SEQ ID NO:5的转基因盒的组件及其相应的位置：

表2

| 位置(bp)    | 组件             | 长度(bp) |
|-----------|----------------|--------|
| 1-145     | 5' ITR         | 145    |
| 170-368   | GRK启动子         | 199    |
| 398-3856  | RPGRorf15 cDNA | 3459   |
| 3899-4143 | SV40多聚A        | 245    |
| 4159-4304 | 3' ITR         | 145    |

5' ITR具有以下序列：

TTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCCACGCCC  
GGGCTTTGCCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACTCCATCACTAGGGGTTCTCT  
(SEQ ID NO:6)

3' ITR具有以下序列：

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGCCGG  
GCAAAGCCCCGGGCGTCGGGCGACCTTTGGTCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAA  
(SEQ ID NO:7)

SV40多聚腺苷酸化序列具有以下序列：

GGGGATCCAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACCTAGAATGCAGTGAAAAAA  
ATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTAACAAC  
AACAATTGCATTCATTTTATGTTTCAGGTTTCAGGGGGAGGTGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACA  
AATGTGGTATGGCTGATTATGATCA (SEQ ID NO:8)。

[0059] 本领域技术人员将理解的是，包含转基因且缺少病毒复制所需的病毒蛋白（例如cap和rep）的AAV载体不能复制，因为此类蛋白是病毒复制和包装所必需的。辅助病毒通常包括腺病毒或单纯疱疹病毒。备选地，如下所述，可以向包装细胞提供辅助功能（E1a、E1b、E2a、E4和VA RNA），包括通过用编码各种辅助元件的一种或多种核酸转染该细胞和/或该细胞可包含编码辅助蛋白的核酸。例如，HEK 293通过用腺病毒5DNA转化人细胞来产生，并且其现在表达大量的腺病毒基因，包括但不限于E1和E3（参见例如Graham等人，1977，J.Gen.Virol.36:59-72）。由此，那些辅助功能可以由HEK 293包装细胞来提供，而不需要通过例如编码它们的质粒来将它们供应给细胞。

[0060] 病毒载体可以是任何合适的核酸构建体，诸如DNA或RNA构建体，并且可以是单链、双链或双链体的（即如WO 2001/92551中所述自身互补的）。

[0061] 包装的病毒载体的病毒衣壳组件可以是细小病毒衣壳。AAV Cap和嵌合衣壳是优选的。例如，病毒衣壳可以是AAV衣壳（例如AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAV1.1、AAV2.5、AAV6.1、AAV6.3.1、AAV9.45、AAVrh10、AAVrh74、RHM4-1、AAV2-TT、AAV2-TT-S312N、AAV3B-S312N、AAV-LK03、蛇AAV、鸟类AAV、牛AAV、犬AAV、马AAV、绵羊AAV、山羊AAV、虾AAV和目前已知或以后发现的任何其他AAV）。参见例如Fields等人，VIROLOGY，第2卷，第69章（4.sup.th ed., Lippincott-Raven Publishers）。

[0062] 在一些实施方案中，包装的病毒载体的病毒衣壳组件是天然AAV衣壳的变体（即相

对于天然AAV衣壳包含一个或多个修饰)。在一些实施方案中,衣壳是AAV2、AAV5或AAV8衣壳的变体。在优选实施方案中,衣壳是AAV2衣壳的变体,诸如美国专利申请公开号2019/0255192A1中描述的那些(例如包含SEQ ID NO:42-59中任一种的氨基酸序列)。在一个特别优选的实施方案中,衣壳包含具有以下氨基酸序列的VP1衣壳蛋白:MAADGYLPDWLEDTLSEGI RQWWKLKPGPPPKAAERHKDDSRGLVLPGYKYLGPFGNLDKGEPVNEADAAALEHDKAYDRQLDSGDNPYLKYNH ADAEFQERLKEDTSFGGNLGRAVFQAKKRVLEPLGLVEEPVKTPAGKKRPVEHSPVEPDSSSGTGKAGQQPARKRL NFGQTGDADSVDPDQPLGQPAPPSGLGTNTMATGSGAPMADNNEGADGVGNSSGNWHCDSTWMGDRVITTSTRTW ALPTYNNHLYKQISSQSGASNDNHYFGYSTPWGYFDFNRHFCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVKEV TQNDGTTTTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPQYGYLTLNNGSQAVGRSSFYCLEYFP SQMLRTGNNFTFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYLSRTNTPSGTTTQSRLQFSQAGASDIRDQSR NWLPGPCYRQQRVSKTSADNNNSEYSWTGATKYHLNGRDSLVNPGPAMASHKDDDEEKFFPQSGVLIFGKQGSEKTN VDIEKVMITDEEEIRTTNPVATEQYGSVSTNLQRGNLAISDQTKHARQAATADVNTQGVLPGMVWQDRDVYLQGP I WAKIPHTDGHFHPSPMLMGGFGLKHPPPQILIKNTVPANPSTTFSAAKFASFITQYSTGQVSVEIEWELQKENS KR WNPEIQYTSNYNKSNNVDFTVDTNGVYSEPRPIGTRYLTRNL (SEQ ID NO:9)。

[0063] SEQ ID NO:9的变体AAV衣壳蛋白相对于天然AAV2衣壳含有下列修饰:(i)位于组装的衣壳(仅VP1蛋白)内部的氨基酸位置34处的脯氨酸(P)至丙氨酸(A)突变,和(ii)存在于VP1、VP2和VP3中的氨基酸位置588处的10个氨基酸(亮氨酸-丙氨酸-异亮氨酸-丝氨酸-天冬氨酸-谷氨酰胺-苏氨酸-赖氨酸-组氨酸-丙氨酸/LAISDQTKHA)的插入。

[0064] 完整的AAV Cap蛋白包括VP1、VP2和VP3。包含编码AAV VP衣壳蛋白的核苷酸序列的ORF可以包含少于完整的AAV Cap蛋白或可以提供完整的AAV Cap蛋白。

[0065] 在再一个实施方案中,本发明提供了祖先AAV载体用于治疗性体内基因疗法的用途。具体而言,从头合成了基于计算机来源的序列并表征了其生物活性。这一努力导致了九个功能推定的祖先AAV的产生和AAV血清型1、2、8和9的预测祖先Anc80的鉴定(Zinn等人, 2015, Cell Reports 12:1056-1068)。除了组装到病毒颗粒中之外,还可以通过使用WO 2015/054653中描述的方法来实现此类祖先序列的预测和合成,其内容通过引用并入本文。值得注意的是,与当代病毒或其部分相比,使用由祖先病毒序列组装的病毒粒子可以表现出对当前人群中预先存在的免疫的降低的敏感性。

[0066] 本发明包括被“宿主细胞”涵盖的包装细胞,其可以进行培养以产生本发明的包装病毒载体。本发明的包装细胞通常包括具有异源(1)病毒载体功能、(2)包装功能、和(3)辅助功能的细胞。在以下章节中讨论这些组件功能的每一个。

[0067] 初始,可以通过技术人员已知的几种方法来制造载体(参见例如WO 2013/063379)。一种优选的方法描述在Grieger等人,2015, Molecular Therapy 24(2):287-297中,其内容出于所有目的通过引用并入本文。简而言之,将HEK293细胞的有效转染用作起始点,其中使用来自合格临床主细胞库的贴壁HEK293细胞系在摇瓶和允许快速和规模可变的rAAV生产的WAVE生物反应器中的无动物组分悬浮条件下生长。使用三重转染方法(例如WO 96/40240),悬浮HEK293细胞系在转染后48小时收获时产生超过 $10^5$ 个含有载体基因组的粒子(vg)/细胞或超过 $10^{14}$ vg/L的细胞培养物。更具体而言,三重转染是指用三种质粒转染包装细胞的事实:一种质粒编码AAV rep和cap基因,另一种质粒编码各种辅助功能(例如腺病毒或HSV蛋白,诸如E1a、E1b、E2a、E4和VA RNA),并且另一种质粒编码转基因及其各种控制

元件(例如修饰的RPGRorf15基因和hGRK启动子)。

[0068] 为了达到所需产量,优化了许多变量,诸如支持生长和转染两者的兼容无血清悬浮培养基的选择、转染试剂、转染条件和细胞密度的选择。还开发了基于离子交换色谱法的通用纯化策略,其产生AAV血清型1-6、8、9和各种嵌合衣壳的高纯度载体制备。这个便于使用的过程可以在一周内完成,产生高实壳颗粒(full particle)与空壳颗粒(empty particle)比(>90%实壳颗粒),提供适合于临床应用的纯化后产量( $>1 \times 10^{13}$  vg/L)和纯度,并且对于所有血清型和嵌合粒子是通用的。已经利用这种规模可变的制造技术来制造用于视网膜新血管形成(AAV2)、血友病B(scAAV8)、巨轴突神经病(scAAV9)和视网膜色素变性(AAV2)的GMP I期临床AAV载体,其已施用到患者中。此外,通过实施需要在转染后的许多时间点从培养基中收获rAAV的灌流法,总载体产生最少增加5倍。

[0069] 包装细胞包括病毒载体功能,以及包装和载体功能。病毒载体功能通常包括细小病毒基因组的一部分,诸如AAV基因组,其中rep和cap缺失并被修饰的RPGRorf15序列及其相关表达控制序列替换。病毒载体功能包括足够的表达控制序列,以产生用于包装的病毒载体的复制。通常,病毒载体包括细小病毒基因组的一部分,诸如rep和cap缺失并被转基因及其相关表达控制序列替换的AAV基因组。转基因通常侧接两个AAV TR,其代替缺失的病毒rep和cap ORF。包括适当的表达控制序列,诸如组织特异性启动子和适合用于促进转基因在靶细胞中的组织特异性表达的其他调控序列。转基因通常是可以表达以产生治疗性多肽或标志物多肽的核酸序列。

[0070] 选择用于病毒载体的末端重复(TR)(可拆分(resolvable)和不可拆分的)优选是AAV序列,其中血清型1、2、3、4、5和6是优选的。可拆分的AAV TR无需具有野生型TR序列(例如可以通过插入、缺失、截短或错义突变来改变野生型序列),只要TR介导所需功能,例如病毒包装、整合和/或原病毒拯救等即可。TR可以是用作AAV反向末端重复的合成序列,诸如Samulski等人的美国专利号5,478,745中描述的“双重D序列”,其全部公开内容通过引用整体并入本文。通常但不一定,TR来自相同的细小病毒,例如,两个TR序列均来自AAV2。

[0071] 包装功能包括衣壳组件。衣壳组件优选来自细小病毒衣壳,诸如AAV衣壳或嵌合AAV衣壳功能。合适的细小病毒衣壳组件的实例是来自细小病毒科(Parvoviridae),诸如自主性细小病毒或依赖病毒属(Dependovirus)的衣壳组件。例如,衣壳组件可选自AAV衣壳,例如AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9、AAV10、AAV11、AAV12、AAVrh10、AAVrh74、RHM4-1、RHM15-1、RHM15-2、RHM15-3/RHM15-5、RHM15-4、RHM15-6、AAV Hu.26、AAV1.1、AAV2.5、AAV6.1、AAV6.3.1、AAV9.45、AAV2i8、AAV2G9、AAV2i8G9、AAV2-TT、AAV2-TT-S312N、AAV3B-S312N和AAV-LK03,以及尚未鉴定或来自非人灵长类动物来源的其他新衣壳。衣壳组件可包括来自两种或更多种AAV衣壳的组件。

[0072] 包装的病毒载体通常包括侧翼为TR元件的修饰的RPGRorf15基因序列和表达控制序列,在本文中称为“转基因”或“转基因表达盒”,其足以导致载体DNA的包装和修饰的RPGRorf15基因序列在转导的细胞中的后续表达。病毒载体功能可以例如作为质粒或扩增子的组件供应给细胞。病毒载体功能可以在细胞系内染色体外存在和/或可以整合到细胞的染色体DNA中。

[0073] 可以采用将携带病毒载体功能的核苷酸序列引入到细胞宿主中以进行复制和包装的任何方法,包括但不限于电穿孔、磷酸钙沉淀、显微注射、阳离子或阴离子脂质体、以及

与核定位信号组合的脂质体。在通过使用病毒载体的转染来提供病毒载体功能的实施方案中;可以使用产生病毒感染的标准方法。

[0074] 包装功能包括用于病毒载体复制和包装的基因。由此,例如,包装功能可以按需包括病毒基因表达、病毒载体复制、从整合状态拯救病毒载体、病毒基因表达、以及将病毒载体包装成病毒粒子所必需的功能。包装功能可以使用遗传构建体(诸如质粒或扩增子、杆状病毒(Baculovirus)、或HSV辅助构建体)一起或分开供应给包装细胞。包装功能可以在包装细胞内染色体外存在,但优选整合到细胞的染色体DNA中。实例包括编码AAV Rep和Cap蛋白的基因。

[0075] 辅助功能包括建立包装细胞的主动感染(其为起始病毒载体的包装所需)所需的辅助病毒元件。实例包括来源于腺病毒、杆状病毒和/或疱疹病毒的足以导致病毒载体包装的功能。例如,腺病毒辅助功能通常将包括腺病毒组件E1a、E1b、E2a、E4和VA RNA。包装功能可以通过用所需病毒感染包装细胞来供给。包装功能可以使用遗传构建体(诸如质粒或扩增子)一起或分开供应给包装细胞。参见例如Rabinowitz等人,2002,J.Virol.76:791中描述的pXR辅助质粒,以及Grimm等人,1998,Human Gene Therapy 9:2745-2760中描述的pDG质粒。包装功能可以在包装细胞内染色体外存在,但优选整合到细胞的染色体DNA中(例如HEK 293细胞中的E1或E3)。

[0076] 可以采用任何合适的辅助病毒功能。例如,当包装细胞是昆虫细胞时,杆状病毒可以充当辅助病毒。疱疹病毒也可用作AAV包装方法中的辅助病毒。编码一种或多种AAV Rep蛋白的杂合疱疹病毒可以有利地促进更为规模可变的AAV载体产生方案。

[0077] 可以采用将携带辅助功能的核苷酸序列引入到细胞宿主中以进行复制和包装的任何方法,包括但不限于电穿孔、磷酸钙沉淀、显微注射、阳离子或阴离子脂质体、以及与核定位信号组合的脂质体。在通过使用病毒载体的转染或使用辅助病毒的感染来提供辅助功能的实施方案中;可以使用产生病毒感染的标准方法。

[0078] 本领域中已知的任何合适的感受细胞或包装细胞可用于生产包装的病毒载体。哺乳动物细胞或昆虫细胞是优选的。在本发明的实践中可用于生产包装细胞的细胞的实例包括例如人细胞系,诸如VERO、WI38、MRC5、A549、HEK293细胞(其在组成型启动子的控制下表达功能性腺病毒E1)、B-50或任何其他HeLa细胞、HepG2、Saos-2、Huh7和HT1080细胞系。在一方面,包装细胞能够在悬浮培养中生长,更优选地,细胞能够在无血清培养中生长。在一个实施方案中,包装细胞是在无血清培养基中悬浮生长的HEK293。在另一个实施方案中,包装细胞是美国专利号9,441,206中描述并保藏为ATCC号PTA 13274的HEK293细胞。许多rAAV包装细胞系在本领域中是已知的,包括但不限于WO 2002/46359中公开的那些。在另一方面,包装细胞以细胞培养室(例如用HEK293细胞接种的10层细胞培养室)的形式培养。

[0079] 用作包装细胞的细胞系包括昆虫细胞系。可以根据本发明使用允许AAV复制并可以在培养中维持的任何昆虫细胞。实例包括草地贪夜蛾(*Spodoptera frugiperda*),诸如Sf9或Sf21细胞系、果蝇属(*Drosophila* spp.)细胞系或蚊细胞系,例如白纹伊蚊(*Aedes albopictus*)来源的细胞系。优选的细胞系是草地贪夜蛾Sf9细胞系。以下参考文献因其涉及昆虫细胞用于表达异源多肽的用途、将核酸引入到此类细胞中的方法和在培养中维持此类细胞的方法的教导而并入本文:Methods in Molecular Biology,ed.Richard,Humana Press,N J(1995);O'Reilly等人,Baculovirus Expression Vectors:A Laboratory

Manual, Oxford Univ. Press (1994); Samulski等人, 1989, J. Virol. 63:3822-3828; Kajigaya等人, 1991, Proc. Nat'l. Acad. Sci. USA 88:4646-4650; Ruffing等人, 1992, J. Virol. 66:6922-6930; Kimbauer等人, 1996, Virol. 219:37-44; Zhao等人, 2000, Virol. 272:382-393; 和Samulski等人, 美国专利号6,204,059。

[0080] 根据本发明的病毒衣壳可以使用本领域已知的任何方法产生, 例如通过从杆状病毒表达(Brown等人, (1994) Virology 198:477-488)。作为进一步的备选方案, 例如由Urabe等人, 2002, Human Gene Therapy 13:1935-1943所述使用杆状病毒载体来递送rep/cap基因和rAAV模板, 可以在昆虫细胞中产生本发明的病毒载体。

[0081] 在另一方面, 本发明提供了在昆虫细胞中生产rAAV的方法, 其中可以通过将AAV Rep和Cap工程化到杆状病毒载体的多角体蛋白编码区中并通过转染到宿主细胞中产生病毒重组体来构建携带这些基因编码区的杆状病毒包装系统或载体。值得注意的是, 当使用杆状病毒生产AAV时, 优选AAV DNA载体产物是不使用对AAV ITR的突变的自身互补的AAV样分子。这似乎是由于功能性Rep酶活性的缺少而导致昆虫细胞中无效率的AAV rep切割的副产物, 其产生自身互补的DNA分子。宿主细胞是杆状病毒感染的细胞, 或者其中已经引入了编码杆状病毒辅助功能的另外的核酸, 或其中包括这些杆状病毒辅助功能。这些杆状病毒可以表达AAV组件并随后促进衣壳的产生。

[0082] 在生产过程中, 包装细胞通常包括一种或多种足以导致病毒载体的复制和包装的病毒载体功能以及辅助功能和包装功能。这些不同的功能可以使用遗传构建体(诸如质粒或扩增子)一起或分开供应给包装细胞, 并且它们可以在细胞系内染色体外存在或整合到细胞的染色体中。

[0083] 可以供应具有任何一种或多种已经掺入的所述功能的细胞, 例如具有一种或多种染色体外掺入或整合到细胞的染色体DNA中的载体功能的细胞系、具有一种或多种染色体外掺入或整合到细胞的染色体DNA中的包装功能的细胞系、或具有染色体外掺入或整合到细胞的染色体DNA中的辅助功能的细胞系。

[0084] rAAV载体可以通过本领域中标准的方法(诸如通过柱色谱或氯化铯梯度)进行纯化。用于纯化rAAV载体的方法在本领域中是已知的, 并包括Clark等人, 1999, Human Gene Therapy 10(6):1031-1039; Schenpp和Clark, 2002, Methods Mol. Med. 69:427-443; 美国专利号6,566,118和WO 98/09657中描述的方法。

#### [0085] 治疗方法

在某些实施方案中, 提供了用于在需要此类治疗的受试者中治疗XLRP的方法, 所述方法通过向该受试者施用治疗有效量的具有与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列的核酸, 或包含此类核酸和至少一种药学上可接受的赋形剂的药物组合物来进行。

[0086] 在相关方面, 提供了用于治疗XLRP的包含与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列的核酸。

[0087] 在另一些相关方面, 提供了包含与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列的核酸用于制造药物的用途。

[0088] 在另一些相关方面, 提供了包含与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列的核酸用于制造用于治疗XLRP的药物

的用途。

[0089] 在一些方面,与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列可操作地连接至表达控制序列。在一些实施方案中,SEQ ID NO:1的核苷酸序列可操作地连接至人G蛋白偶联受体视紫红质激酶1(hGRK)启动子。在一些优选实施方案中,hGRK启动子具有SEQ ID NO:4的序列。

[0090] 在一些实施方案中,与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列构成表达盒的一部分。在一些方面,表达盒从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf15基因,(d) SV40多聚腺苷酸化序列,和(e) AAV2末端重复。在优选实施方案中,5' AAV2末端重复具有如SEQ ID NO:6所示的核苷酸序列和/或hGRK启动子具有如SEQ ID NO:4所示的核苷酸序列和/或SV40多聚腺苷酸化序列具有如SEQ ID NO:8所示的核苷酸序列和/或3' AAV2末端重复具有如SEQ ID NO:7所示的核苷酸序列。在一个特别优选的实施方案中,表达盒包含含有SEQ ID NO:5的核苷酸序列或与其具有至少80%、至少81%、至少82%、至少83%、至少84%、至少85%、至少86%、至少87%、至少88%、至少89%、至少90%、至少91%、至少92%、至少93%、至少94%、至少95%、至少96%、至少97%、至少98%或至少99%同一性的序列的核酸。

[0091] 在其他实施方案中,提供了用于在需要此类治疗的受试者中治疗XLRP的方法,所述方法通过向该受试者施用治疗有效量的重组AAV(rAAV)病毒粒子,或包含其的药物组合物来进行,rAAV病毒粒子包含(i)可操作地连接至表达控制序列的核酸,所述核酸具有与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列,和(ii) AAV衣壳。

[0092] 在相关实施方案中,提供了重组AAV(rAAV)病毒粒子用于治疗XLRP的用途,所述rAAV病毒粒子包含(i)可操作地连接至表达控制序列的核酸,所述核酸具有与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列,和(ii) AAV衣壳。

[0093] 在另一些相关实施方案中,提供了重组AAV(rAAV)病毒粒子用于制造用于治疗XLRP的药物的用途,所述rAAV病毒粒子包含(i)可操作地连接至表达控制序列的核酸,所述核酸具有与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列,和(ii) AAV衣壳。

[0094] 在一些实施方案中,rAAV病毒粒子包含天然AAV2、AAV4、AAV5或AAV8衣壳。在另一些实施方案中,rAAV病毒粒子包含变体AAV衣壳,其相对于AAV2、AAV4、AAV5或AAV8包含一个或多个修饰。在一个优选实施方案中,AAV衣壳包含含有SEQ ID NO:9的序列的衣壳蛋白。

[0095] 在一些实施方案中,rAAV病毒粒子包含(i)天然AAV2衣壳或其变体,和(ii)表达盒,所述表达盒从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf15基因,(d) SV40多聚腺苷酸化序列,和(e) AAV2末端重复。在优选实施方案中,rAAV包含(i)含有SEQ ID NO:9的衣壳蛋白的衣壳,和(ii)含有SEQ ID NO:6的5' AAV2末端重复、SEQ ID NO:4的hGRK启动子、SEQ ID NO:8的SV40多聚腺苷酸化序列和SEQ ID NO:7的3' AAV2末端重复的核酸。在一个特别优选的实施方案中,rAAV包含(i)含有SEQ ID NO:9的衣壳蛋白的衣壳,和(ii)含有SEQ ID NO:5的核苷酸序列的表达盒。

[0096] 在特别优选的实施方案中,提供了rAAV在治疗XLRP中或用于制造用于治疗XLRP的药物的用途,其中rAAV包含(i)含有SEQ ID NO:5的核苷酸序列的核酸,和(ii)含有具有SEQ ID NO:9的氨基酸序列的衣壳蛋白的衣壳。在一些方面,rAAV通过玻璃体内注射来施用。

[0097] 在另一些特别优选的实施方案中,提供了治疗XLRP的方法,其包括向受试者施用有效量的rAAV,所述rAAV包含(i)含有SEQ ID NO:5的核苷酸序列的核酸,和(ii)含有具有SEQ ID NO:9的氨基酸序列的衣壳蛋白的衣壳。在一些方面,rAAV通过玻璃体内注射施用于受试者。

[0098] 在另一些方面,提供了药物组合物,其包含任选可操作地连接至表达控制序列的核酸,和至少一种药学上可接受的赋形剂,所述核酸具有与SEQ ID NO:1的核苷酸序列具有至少90%、至少95%、至少98%同一性或100%同一性的核苷酸序列。

[0099] 在一些实施方案中,药物组合物包含可操作地连接至组成型启动子(优选具有与SEQ ID NO:4的核苷酸序列具有至少90%、至少95%至少98%同一性或100%同一性的序列的hGRK启动子)的核酸,所述核酸包含SEQ ID NO:1的核苷酸序列。

[0100] 在另一些方面,提供了药物组合物,其包含至少一种药学上可接受的赋形剂和传染性rAAV,所述传染性rAAV包含(i) AAV衣壳和(ii) 核酸,所述核酸从5'至3'包含:(a) AAV2末端重复,(b) hGRK启动子,(c) SEQ ID NO:1的密码子优化的RPGRorf15基因,(d) SV40多聚腺苷酸化序列,和(e) AAV2末端重复。在相关实施方案中,药物组合物包含 $10^9$ 至 $10^{14}$ vg,优选 $10^{10}$ 至 $10^{13}$ vg的rAAV,更优选包含 $3 \times 10^{11}$ vg或 $1 \times 10^{12}$ vg的rAAV。

[0101] 在优选实施方案中,药物组合物包含rAAV,所述rAAV包含(i) 含有衣壳蛋白的衣壳,所述衣壳蛋白包含SEQ ID NO:9的序列或由其组成,和(ii) 包含SEQ ID NO:6的5' AAV2末端重复和/或SEQ ID NO:4的hGRK启动子和/或SEQ ID NO:8的SV40多聚腺苷酸化序列和/或SEQ ID NO:7的AAV2末端重复的核酸。在相关实施方案中,药物组合物包含 $10^9$ vg至 $10^{14}$ vg,优选 $10^{10}$ vg至 $10^{13}$ vg的rAAV,更优选包含大约 $3 \times 10^{11}$ vg或大约 $1 \times 10^{12}$ vg的rAAV。

[0102] 在一些实施方案中,提供了用于在人受试者的一种或多种感光细胞中表达RPGR的方法,其包括向人受试者施用有效量的如本文中描述的传染性rAAV,其中RPGR在一种或多种感光细胞中表达。在一些优选实施方案中,传染性rAAV的有效量是 $10^9$ 至 $10^{14}$ vg/眼和/或向人受试者玻璃体内施用(双侧或单侧)单剂量的rAAV和/或rAAV包含SEQ ID NO:9的衣壳和/或rAAV包含含有SEQ ID NO:5的核苷酸序列的异源核酸。

[0103] 在一个特别优选的实施方案中,提供了药物组合物,其包含至少一种药学上可接受的赋形剂和传染性rAAV,所述传染性rAAV包含(i) 含有衣壳蛋白的衣壳,所述衣壳蛋白包含SEQ ID NO:9的序列或由其组成,和(ii) 含有SEQ ID NO:5的核苷酸序列或由其组成的核酸。在相关实施方案中,药物组合物包含 $10^{10}$ vg至 $10^{13}$ vg的rAAV,优选包含大约 $3 \times 10^{11}$ vg或大约 $1 \times 10^{12}$ vg的rAAV。

[0104] 在一些实施方案中,通过眼周或眼内(玻璃体内、脉络膜上或视网膜下)注射向患有XLRP的人施用本文中描述的核酸或传染性rAAV,由此在受试者中治疗XLRP。在另一些实施方案中,将本文中描述的核酸或传染性rAAV视网膜下或玻璃体内施用于患有XLRP的人,由此在受试者中治疗XLRP。在优选实施方案中,向患有XLRP的人受试者施用本文中描述的rAAV的单次玻璃体内注射(双侧或单侧)。

[0105] 在相关方面,在受治疗的受试者中治疗XLRP包括在治疗后例如6个月、12个月或24

个月(i)相对于对照(例如相对于治疗前受治疗的患者的基线测量值,如果核酸或rAAV是单侧施用的话,则相对于未治疗的眼,或相对于未治疗的XLRP患者的同期或历史对照组)在视觉功能或功能性视觉方面得到改善(即增益)和/或(ii)与对照(例如同一患者或未治疗的对照组的未治疗的眼)相比,受治疗的眼的视觉功能丧失和/或视网膜变性减少。这些改善可以通过适当的眼科测试来评估,所述眼科测试包括但不限于视觉敏锐度测试、微视野检查和其他视野测试、解剖测试,诸如光学相干断层扫描和眼底自发荧光成像、视网膜电生理学 and/或生活质量(QoL)评估。

[0106] 在一些方面,本文中描述的核酸或rAAV(或包含其的药物组合物)的有效量是在人患者中有效治疗XLRP的量。在相关方面,本文中描述的rAAV的有效量是 $10^9$ 至 $10^{14}$ 个rAAV粒子(或载体基因组(vg))/眼,优选 $10^{10}$ 至 $10^{13}$ vg/眼、或 $1 \times 10^{11}$ vg/眼至 $5 \times 10^{12}$ vg/眼,更优选为大约 $3 \times 10^{11}$ vg/眼或大约 $1 \times 10^{12}$ vg/眼。在一些优选实施方案中,将大约 $3 \times 10^{11}$ vg/眼或大约 $1 \times 10^{12}$ vg/眼的单一剂量玻璃体内施用于患有XLRP的人患者,由此治疗XLRP。

## 实施例

[0108] 以下实施例示出了本发明的优选实施方案,并且并非意在以任何方式限制本发明的范围。虽然本发明已就其优选实施方案进行了描述,但通过阅读本申请,其各种修改对本领域技术人员将是显而易见的。

[0109] 实施例1——具有改善的稳定性的RPGRorf15 cDNA序列的密码子优化人视网膜色素变性GTP酶调节因子开放阅读框15(hRPGRorf15)序列含有高度重复的、富含嘌呤的区域,这导致转基因盒克隆和质粒扩增过程中的序列不稳定性。hRPGRorf15 cDNA序列(NCBI参考序列NM\_001034853.1)经过密码子优化以生成具有提高的在人细胞中的表达以及改善的序列稳定性的RPGRorf15 cDNA序列。

[0110] 密码子优化的核苷酸序列如下所示:

```
ATGAGAGAGCCTGAAGAGCTGATGCCTGATAGCGGAGCAGTGTTTACCTTTGGGAAGA
GCAAGTTCGCAGAGAATAACCCTGGGAAATTCTGGTTTAAGAACGACGTGCCCCGTGCAC
CTGAGCTGTGGCGATGAGCACTCCGCCGTGGTGACAGGCAACAATAAGCTGTACATGTT
CGGCTCTAACAATTGGGGACAGCTGGGCCTGGGAAGCAAGTCCGCCATCAGCAAGCCA
ACCTGCGTGAAGGCCCTGAAGCCCGAGAAGGTGAAGCTGGCCGCTGTGGCAGAAAC
CACACACTGGTGAGCACCGAGGGAGGAAACGTGTACGCAACAGGAGGCAACAATGAA
GGCCAGCTGGGCCTGGGCGACACAGAGGAGAGGAATACCTTTCACGTGATCAGCTTCTT
TACCTCCGAGCACAAGATCAAGCAGCTGTCCGCCGGCTCTAACACAAGCGCCGCCCTG
ACCGAGGACGGCCGCCTGTTTCATGTGGGGCGATAATAGCGAGGGCCAGATCGGCCTGA
AGAACGTGTCCAACGTGTGCGTGCCTCAGCAGGTGACCATCGGCAAGCCAGTGTCTCTG
GATCTCTTGTGGCTACTATCACAGCGCCTTCGTGACCACAGATGGCGAGCTGTACGTGTT
TGGAGAGCCAGAGAACGGCAAGCTGGGCCTGCCTAACCAGCTGCTGGGCAATCACCGG
ACACCCCAGCTGGTGTCCGAGATCCCTGAGAAAGTGATCCAGGTGGCATGCGGAGGAG
AGCACACAGTGGTGCTGACCGAGAATGCCGTGTATACCTTCGGCCTGGGACAGTTTGGA
CAGCTGGGCCTGGGCACATTCTGTTTGAGACAAGCGAGCCAAAAGTGATCGAGAACA
TCCGCGACCAGACAATCAGCTACATCTCTGCGGCGAGAATCACACAGCCCTGATCACC
```

GACATCGGCCTGATGTATACCTTTGGCGATGGCCGGCACGGCAAGCTGGGCCTGGGCCT  
GGAGAACTTCACAAATCACTTTATCCCCACCCTGTGCTCTAACTTCCTGCGGTTTCATCGT  
GAAGCTGGTGGCCTGCGGCGGCTGTCACATGGTGGTGTTCGCAGCACCTCACAGGGGA  
GTGGCCAAGGAGATCGAGTTTGACGAGATCAACGATACATGCCTGTCCGTGGCCACCTT  
CCTGCCATACAGCTCCCTGACATCCGGCAATGTGCTGCAGCGCACCTGTCTGCCAGGA  
TGCGGAGAAGGGAGAGGGAGCGGTCCCCTGACTCTTTCAGCATGAGGCGGACACTGCC  
ACCTATCGAGGGCACCTGGGCCTGTCTGCCTGCTTCCTGCCTAACAGCGTGTTCCCAA  
GATGTAGCGAGAGGAATCTGCAGGAGTCTGTGCTGAGCGAGCAGGATCTGATGCAGCC  
AGAGGAGCCCGACTACCTGCTGGATGAGATGACAAAGGAGGCCGAGATCGACAACTCT  
AGCACCGTGGAGAGCCTGGGCGAGACAACAGATATCCTGAATATGACACACATCATGTC  
CCTGAACTCTAATGAGAAGTCTCTGAAGCTGAGCCAGTGCAGAAGCAGAAGAAGCAG  
CAGACCATCGGCGAGCTGACCCAGGACACAGCCCTGACCGAGAACGACGATTCTGATG  
AGTATGAGGAGATGAGCGAGATGAAGGAGGGCAAGGCCTGTAAGCAGCACGTGTCCCA  
GGGCATCTTCATGACCCAGCCAGCCACCACAATCGAGGCCTTTTCTGACGAAGAGGTGG  
AGATCCCCGAGGAGAAGGAGGGCGCCGAGGATAGCAAGGGCAATGGCATCGAGGAGC  
AGGAGGTGGAGGCCAACGAGGAGAATGTGAAGGTGCACGGCGGCAGAAAGGAGAAG  
ACAGAGATCCTGTCCGACGATCTGACCGACAAGGCCGAGGTGTCCGAGGGCAAGGCCA  
AGTCTGTGGGAGAGGCAGAGGACGGACCAGAGGGACGCGGCGATGGAACCTGCGAGG  
AGGGATCCTCTGGAGCAGAGCACTGGCAGGACGAAGAAAGAGAGAAGGGCGAGAAGG  
ATAAGGGCAGAGGAGAGATGGAGAGGCCTGGAGAGGGAGAGAAGGAGCTGGCAGAGA  
AGGAGGAGTGGAAGAAGAGGGACGGCGAGGAGCAGGAGCAGAAGGAGAGAGAGCAG  
GGCCACCAGAAGGAGAGGAACCAGGAGATGGAGGAGGGAGGAGAGGAGGAGCACGG  
CGAGGGAGAGGAGGAGGAGGGCGATAGAGAGGAAGAAGAGGAGAAGGAGGGAGAGG  
GCAAGGAGGAAGGCGAGGGAGAGGAGGTGGAGGGAGAAAGGGAGAAGGAGGAGGG  
AGAGCGCAAGAAGGAAGAAAGAGCAGGCAAGGAAGAGAAGGGAGAGGAGGAGGGC  
GATCAGGGCGAAGGAGAGGAGGAGGAGACAGAGGAAGGGGAGAGGAGAAGGAGGA  
GGGAGGAGAGGTCTGAAGGAGGAGAAGTGGAGGAGGGCAAGGGCGAAAGAGAAGAGG  
AGGAGGAGGAAGGCGAGGGCGAAGAAGAGGAGGGCGAGGGCGAGGAAGAAGAGGG  
CGAGGGCGAAGAGGAAGAAGGCGAGGGCAAGGGCGAGGAGGAGGGCGAAGAAGGCG  
AAGGGGAGGAGGAGGGCGAAGAGGGAGAGGGCGAGGGCGAGGAGGAAGAAGGCGA  
AGGCGAAGGCGAAGAAGAAGGAGAAGGAGAGGGCGAAGAGGAGGAAGGCGAAGGA  
GAAGGAGAGGAGGAAGGAGAAGGGGAGGGCGAAGAGGAGGAGGGAGAAGGCAAGG  
GAGAAGAAGAAGGCGAAGAAGGCGAGGGAGAAGGCGAGGAAGAAGAAGGCGAGGG  
AGAGGGAGAGGACGGCAAGGCGAGGGCGAGGAAGAGGAAGGAGAGTGGGAGGGCG  
AGGAAGAGGAGGGAGAAGGAGAAGGCGAAGAAGAAGGGGAAGGAGAGGGCGAGGA  
AGGAGAAGGCGAAGGCGAAGAGGAGGAGGGGGAAGGGGAGGGCGAGGAGGAAGAG  
GGAGAAGAGGAAGGCGAAGAAGAGGGAGAAGGCGAAGAGGAAGGAGAAGGCGAGG  
GAGAAGAAGAGGAGGAGGGCGAGGTCTGAAGGCGAGGTGGAGGGCGAAGAGGGGGAA  
GGCGAAGGCGAGGAGGAGGAAGGGGAAGAAGAAGGCGAGGAGAGAGAGAAAGAAG

GCGAGGGCGAGGAGAACAGAAGGAATCGCGAAGAAGAAGAGGAAGAAGAGGGCAAG  
TACCAGGAGACAGGCGAGGAGGAGAACGAGCGGCAGGATGGCGAGGAGTATAAGAAG  
GTGTCCAAGATCAAGGGCTCTGTGAAGTACGGCAAGCACAAGACCTATCAGAAGAAGA  
GCGTGACCAACACACAGGGCAATGGCAAGGAGCAGCGCAGCAAGATGCCTGTGCAGTC  
CAAGCGGCTGCTGAAGAATGGCCCCTCTGGGAGCAAGAAGTTTTGGAATAATGTCCTGCCACACTACC  
TGGAGCTGAAATGA (SEQ ID NO:10)。

[0111] 通过GenScript构建含有在人G蛋白偶联受体激酶1启动子(也称为人视紫红质激酶启动子(hGRK))或普遍存在的3-磷酸甘油酸激酶(PGK)启动子的控制下的密码子优化的hRPGRorf15基因(SEQ ID NO:10)的AAV质粒。

[0112] 20ng的AAV质粒DNA用于转化感受态大肠杆菌(E.coli)(Cat.#C3040H, New England BioLabs, Ipswich, MA), 并将细胞涂布在卡那霉素50μg/ml板(#L1025, Teknova, Hollister, CA)上。从所得菌落生长微量制备培养物, 用GeneJET质粒微量制备试剂盒(Cat.#0503, ThermoFisher, Waltham, MA)制备DNA并进行限制性消化以鉴定阳性克隆。

[0113] 尽管经过密码子优化, 在限制性消化后仍检测到质粒生产过程中密码子优化的hRPGRorf15(SEQ ID NO:10)的序列不稳定性。

[0114] 使用不同的优化算法开发第二密码子优化的hRPGRorf15序列, 所述优化算法包括但不限于密码子使用偏性、GC含量、富含AT或富含GC的区域、mRNA二级结构、RNA不稳定性基序、隐蔽剪接位点、内部chi位点和核糖体结合位点、以及重复序列的参数。通过将密码子适应指数(CAI)提升至0.89来改变人中的密码子使用偏性。平均GC含量从天然序列中的59.16优化为优化序列中的57以延长mRNA的半衰期。所得密码子优化的核苷酸序列(如本文中SEQ ID NO:1所示)含有改善的密码子使用、改变的GC含量、更好的mRNA稳定性和负性顺式作用元件的修饰。

[0115] 构建包含SEQ ID NO:5(SEQ ID NO:5包含(i)5'AAV2 ITR(SEQ ID NO:6); (ii)在hGRK启动子(SEQ ID NO:4)的控制下的密码子优化的hRPGRorf15 cDNA(SEQ ID NO:1); (iii)SV40晚期多聚A元件(SEQ ID NO:8)和(iv)3'AAV2 ITR(SEQ ID NO:7))的核苷酸序列的AAV质粒(pAAV-GRK启动子-cohRPGRorf15-SV40)。

[0116] 如下制备pAAV-GRK启动子-cohRPGRorf15-SV40 DNA。来自GenScript的质粒DNA(20ng)用于转化感受态大肠杆菌(Cat.#C3040H, New England BioLabs, Ipswich, MA), 并将细胞涂布在卡那霉素50μg/ml板(#L1025, Teknova, Hollister, CA)上。从所得菌落生长微量制备培养物, 用GeneJET质粒微量制备试剂盒(Cat.#0503, ThermoFisher, Waltham, MA)制备DNA并进行限制性消化以鉴定阳性克隆。从一个阳性克隆生长50mL Terrific Broth中的培养物, 并用Qiagen EndoFree Plasmid Maxi试剂盒(Cat.#12362, Qiagen, Hilden, Germany)制备DNA。用多种限制酶消化pAAV-GRK-cohRPGRorf15-SV40的大量制备物以验证质粒的身份。限制性消化物和预期片段的凝胶电泳显示在图1中。所有实际片段与预期片段匹配。通过Sanger DNA测序验证表达盒的序列。

[0117] 结论: 通过限制性消化正确映射了pAAV-GRK-cohRPGRorf15-SV40的大量制备物, 并通过Sanger DNA测序验证了其完整性。由此, 如SEQ ID NO:1所示的密码子优化的hRPGRorf15序列相对于SEQ ID NO:3的天然序列和SEQ ID NO:10的密码子优化的序列表现出优异的稳定性。

[0118] 实施例2——由SEQ ID NO:1的密码子优化的hRPGRorf15表达的人RPGRorf15蛋白的表达和活性

在转染的HEK293T细胞中评估由pAAV-GRK-cohRPGRorf15-SV40表达的人RPGRorf15蛋白的表达和活性。

[0119] 简而言之,将HEK293T细胞以 $2.0 \times 10^5$ 个细胞/孔接种在12孔板中的1.0mL DMEM/10% FBS培养基中。由于HEK293T细胞的高可转染性和蛋白表达而使用其。第二天,将与3.0  $\mu$ l FuGene6 (Cat.#E2691, Promega, Madison, WI) 复合的1.0 $\mu$ g AAV质粒DNA添加到在一式两份的孔中的细胞中。转染后两天,用PBS洗涤细胞,并在0.25mL含有1 $\times$  Halt蛋白酶抑制剂 (ThermoFisher) 的1 $\times$  被动裂解缓冲液 (Promega) 中裂解,在室温下摇动15分钟。通过在微量离心机中在4 $^{\circ}$ C下以12,000g离心10分钟来沉淀 (pellet) 细胞碎片。收集上清液并储存在-20 $^{\circ}$ C下。转染中分别包括无质粒和pAAV-PGK启动子-cohRPGRorf15-SV40样品作为阴性和阳性对照。pAAV-PGK启动子-cohRPGRorf15-SV40与上述AAV载体相同,除了密码子优化的hRPGRorf15可操作地连接至普遍存在的启动子3-磷酸甘油酸激酶 (PGK) 启动子而不是hGRK启动子。

[0120] 将细胞裂解物 (20 $\mu$ l) 与10 $\mu$ l 4 $\times$  LDS、4 $\mu$ l 10 $\times$  还原剂和6 $\mu$ l水混合 (最终体积=40 $\mu$ l) 并在70 $^{\circ}$ C下变性10分钟。将样品加载到12孔Bolt 4-12% Bis-Tris Plus聚丙烯酰胺凝胶 (Invitrogen, NW04122BOX) 上,并在1 $\times$  MOPS缓冲液中在200V下运行32分钟。用iBlot 2装置 (ThermoFisher) 将分离的蛋白转移至硝酸纤维素滤膜10分钟并使用iBind Flex装置 (ThermoFisher) 用第一抗RPGR (Sigma HPA001593 1:2000和GenScript CT-15U1729DC260\_16 1:500) 和抗多聚谷氨酰化GT335 (AG-20B-0020 1:500, Adipogen, San Diego, CA) 抗体进行探测。二抗是HRP缀合的山羊抗兔 (ThermoFisher 31460) (用于抗RPGR一抗) 和HRP缀合的山羊抗小鼠 (ThermoFisher 31430) (用于抗多聚谷氨酰化一抗)。用SuperSignal West Dura化学发光底物 (ThermoFisher 34076) 使蛋白可视化,并在ChemiDoc MP (BioRad, Hercules, CA) 上成像。使用的所有抗体列举在下表3中。

[0121] 表3:蛋白印迹抗体

| 抗体              | 宿主物种 | 销售商       | 目录#           | 稀释度     |
|-----------------|------|-----------|---------------|---------|
| 抗RPGR多克隆        | 兔    | Sigma     | HPA001593     | 1:2,000 |
| 抗CT-15          | 兔    | GenScript | U1729DC260_16 | 1:500   |
| 抗多聚谷氨酰化GT335    | 小鼠   | Adipogen  | AG-20B-0020   | 1:500   |
| HRP抗兔IgG (H+L)  | 山羊   | Thermo    | 31460         | 1:5,000 |
| HRP抗小鼠IgG (H+L) | 山羊   | Thermo    | 31430         | 1:5,000 |

图2显示了来自转染的HEK293T细胞的裂解物的代表性蛋白印迹的图像。CT-15和Sigma抗体检测到似乎是RPGRorf15的相同135-140kD物类 (species), 因为其存在于RPGR转染的裂解物中, 但不存在于未转染的裂解物中, 具有正确的尺寸并且被多聚谷氨酰化检测抗体GT335识别。当由普遍存在的PGK启动子 (其并不优先在感光细胞中活跃) 驱动时, 表达更高。

[0122] 结论——来自转染的HEK293T细胞的裂解物的蛋白印迹分析表明了由SEQ ID NO:1的密码子优化的hRPGRorf15表达的正确尺寸的hRPGRorf15蛋白的表达和多聚谷氨酰化。

[0123] 实施例3——hRPGRorf15在人XLRP体外模型中的功能性表达生成人体外模型系

统,以评估用具有SEQ ID NO:1的核苷酸序列的密码子优化的人RPGRorf15核酸对X连锁视网膜色素变性(XLRP)疾病表型的校正。为此,构建AAV载体,其包含由人G蛋白偶联受体视紫红质激酶1(hGRK)启动子驱动的SEQ ID NO:1的核苷酸序列(即实施例1和2中描述的AAV载体骨架,具有SEQ ID NO:5的序列)和具有SEQ ID NO:9的氨基酸序列的变体衣壳蛋白。选择hGRK启动子以将RPGRorf15的表达限制于光感受器。

[0124] 从患有XLRP的个体抽取的全血中分离外周血单核细胞(PBMC),并使用CytoTune iPS 2.0Sendai重编程试剂盒(Thermo Fisher Scientific,Waltham,MA)重编程为诱导的多能干细胞(iPSC)。通过免疫细胞化学检查iPSC标志物(包括Sox2、Oct4和Nanog)来确认多能干细胞的多能性。随后通过Gonzalez-Cordero等人,Stem Cell Report,9,820:837(2017);Gonzalez Cordero等人,Human Gene Therapy,29(1)(2018);和Meyer等人,Stem Cells,29(8):1206-1218(2011)中描述的方法将诱导的多能干细胞分化为光感受器。通过免疫细胞化学检查特异性标志物(恢复蛋白和视紫红质)来确认光感受器分化。证实了光感受器缺少hRPGRorf15蛋白表达和已知赋予功能性的hRPGorf15蛋白的谷氨酰化。

[0125] 免疫细胞化学如下:将细胞用4%多聚甲醛(PFA)(Santa Cruz Biotechnologies,Dallas,TX)在4℃下固定15分钟。在含有0.2% Triton-X100(Sigma-Aldrich)、2%牛血清白蛋白(Millipore Sigma,Burlington,MA)和5%山羊血清(Thermo Fisher Scientific)的PBS封闭溶液中进行所有抗体染色。一抗在4℃下进行温育过夜。随后将细胞与二抗在室温下温育一小时,并随后在室温下用DAPI(Sigma Aldrich)在PBS中复染五分钟。使用Zeiss Axio Observer.D1荧光显微镜对细胞进行成像。使用Zeiss Zen 2软件(Carl Zeiss Microscopy LLC,White Plains,NY)进行图像处理。一抗和二抗的列表在下表4中提供:

表4

| 抗体                   | 宿主 | 公司目录号                        | 稀释度    |
|----------------------|----|------------------------------|--------|
| 一抗                   |    |                              |        |
| OCT4                 | 小鼠 | Millipore- MAB4401           | 1:50   |
| Nanog                | 兔  | Abcam- ab21624               | 1:50   |
| SOX2                 | 兔  | Abcam- ab92494               | 1:50   |
| $\beta$ -微管蛋白 III    | 小鼠 | Sigma- T8578                 | 1:200  |
| HNF4- $\alpha$       | 兔  | Santa Cruz-SC-8987           | 1:100  |
| A-SMA                | 小鼠 | Sigma Aldrich- A2547         | 1:500  |
| 恢复蛋白                 | 兔  | EMD Millipore- AB5585        | 1:100  |
| 视紫红质                 | 小鼠 | Abcam- AB98887               | 1:100  |
| RPGR                 | 兔  | Sigma- HPA001593             | 1:2000 |
| GT335                | 小鼠 | Fisher Adipogen- 50-463-394# | 1:4000 |
| $\alpha$ 微管蛋白        | 兔  | Abcam- ab52866               | 1:4000 |
| 二抗                   |    |                              |        |
| Alexa Fluor488 抗兔    | 山羊 | Invitrogen-A11078            | 1: 500 |
| Alexa Fluor555 抗兔    | 山羊 | Invitrogen-A21428            | 1:500  |
| Alexa Fluor680 抗兔    | 山羊 | Invitrogen-A21109            | 1:500  |
| Alexa Fluor488 抗小鼠   | 山羊 | Invitrogen-A11029            | 1:500  |
| Alexa Fluor555 抗小鼠   | 山羊 | Invitrogen-A21422            | 1:500  |
| Alexa Fluor680 抗小鼠   | 山羊 | Invitrogen-35518             | 1:500  |
| 辣根过氧化物酶抗兔 IgG (H+L)  | 山羊 | Thermo- 31460                | 1:5000 |
| 辣根过氧化物酶抗小鼠 IgG (H+L) | 山羊 | Thermo- 31430                | 1:5000 |

为了评估在转导到XLRP-iPSC来源的患病的光感受器中后密码子优化的RPGRorf15转基因的转录本水平,用上述AAV载体以50,000的感染复数(MOI,每个细胞的病毒基因组)转导XLRP光感受器(PR),以确保水平高于测定的检测限。转导后30天分离RNA并合成cDNA。对制备的样品运行微滴式数字PCR,并分析作为拷贝/mL值的每个微滴的转录本水平。检查含有引物/探针组的转录本的微滴数量(高于设定阈值)的定量。创建两个引物/探针组以特异性区分密码子优化的人RPGRorf15转基因与内源性人RPGR1-19组成型同工型(hRPGR1-19)。

[0126] 如预期的那样,未转导的XLRP患病细胞表达低的、背景水平的cohRPGRorf15转录本。在用AAV载体转导后,与hRPGR1-19相比,细胞显示出cohRPGRorf15转录本水平超过400倍的提高。与未转导的细胞cohRPGRorf15水平相比,转导的细胞在cohRPGRorf15转录本方面表现出超过1000倍的提高。未转导的细胞具有比cohRPGRorf15更高的hRPGR1-19水平。参见图3。一式三份进行分析,并将水平取平均值。用包含SEQ ID NO:1的密码子优化的hRPGRorf15的AAV载体的转导显著提高了光感受器培养物中cohRPGRorf15的转录本水平。

[0127] 为了评估通过用AAV载体转导XLRP-iPSC来源的感光细胞产生的密码子优化的人RPGRorf15转基因的蛋白水平,以50,000vg/细胞的MOI转导XLRP-iPSC来源的患病光感受

器。转导后30天收集细胞裂解物,并进行SDS-PAGE和蛋白印迹分析以评估hRPGRorf15蛋白水平。将条带强度量化并在图4中描绘为直方图。与未转导的细胞相比,用AAV载体的转导引起人RPGRorf15蛋白表达的显著增加。

[0128] 为了确定外源引入到光感受器中的cohRPGRorf15蛋白是否是功能性的,检测了谷氨酰化(功能的代表)。根据公开的著作,hRPGRorf15的谷氨酰化和蛋白功能强烈相关。(Fischer等人,2017;Rao等人,2016;Sun等人,2016)。以50,000vg/细胞的MOI转导XLRP-iPSC来源的患病PR。转导后30天收集细胞裂解物,并进行SDS-PAGE和蛋白印迹分析以评估表达的hRPGRorf15蛋白的谷氨酰化。通过用谷氨酰化特异性抗体GT335探测膜并在hRPGRorf15的尺寸127kDa处检查阳性条带模式来确定谷氨酰化。将条带强度量化并在图5中描绘为直方图。在两种XLRP患者来源的患病光感受器中,与未转导的细胞相比,用包含密码子优化的hRPGRorf15核苷酸序列的AAV载体转导PR细胞导致人RPGRorf15蛋白的谷氨酰化的显著增加。

[0129] 由于在使用高MOI的情况下在蛋白印迹中检测到低hRPGRorf15蛋白水平,因此验证了hRPGRorf15密码子优化的转基因(cohRPGRorf15)的剂量反应。为此,构建AAV载体,其包含可操作地连接至普遍存在的启动子3-磷酸甘油酸激酶(PGK)的SEQ ID NO:1的密码子优化的RPGRorf15序列和SEQ ID NO:9的衣壳(除启动子外,该AAV载体与上述AAV载体相同)。以三种MOI(5,000、10,000和20,000)转导患病的光感受器。转导后30天收集细胞裂解物,并进行SDS-PAGE和蛋白印迹分析以评估hRPGRorf15蛋白水平和谷氨酰化(GT335=抗谷氨酰化抗体)。将条带强度量化并描绘为直方图(图6)。尽管由于培养物的异质性,因此存在高变异性,但使用组成型启动子来驱动cohRPGRorf15表达,在较低的MOI下观察到hRPGRorf15蛋白和hRPGRorf15的谷氨酰化。

[0130] 结论——用iPSC来源的光感受器进行的体外研究已经表明,AAV介导的SEQ ID NO:1的密码子优化的hRPGRorf15的递送恢复了人XLRP患病光感受器中的人RPGRorf15转录本和转基因表达。此外,在4D-125转导后表达的RPGRorf15蛋白被翻译后谷氨酰化。基于公开的文献,谷氨酰化赋予了RPGRorf15的功能性。

[0131] 实施例4——在非人灵长类动物中经由玻璃体内施用由R100递送的密码子优化的RPGRorf15 cDNA序列的安全性和生物分布的评估

#### 材料和方法

#### GLP毒理学和生物分布研究

2-14岁的雄性食蟹猴(*cynomolgus macaque*) (食蟹猴(*macaca fascicularis*))经由两次50μL玻璃体内注射通过巩膜向每只眼中给药,总剂量体积为100μL/眼。评估 $1 \times 10^{11}$ vg/眼和 $1 \times 10^{12}$ vg/眼的剂量。用氯胺酮IM麻醉动物并给予局部眼用溶液以消除疼痛。注射后每周通过IM注射施用20-80mg的甲基强的松龙。在施用后第3周、第13周和第26周由受过训练的兽医人员进行安乐死。

[0132] 使用经验证的符合GLP的qPCR测定评估所有主要眼部腔室(视网膜、视神经、睫状体、虹膜、小梁网)和主要全身器官(包括睾丸)中的4D-125(包含SEQ ID NO:9的衣壳蛋白和含有SEQ ID NO:5的核苷酸序列的异源核酸的rAAV)基因组生物分布。在检测到基因组的组织中,通过合格的符合GLP的RT-qPCR测定评估转基因表达。

[0133] 在研究中进行的系列毒理学评估是:临床眼部评估(完整的眼科检查,包括SD-OCT

成像和ERG)、全身评估、临床病理学、大体病理学和显微病理学。对测定进行验证以确定抗衣壳和抗转基因抗体反应。对ELISpot测定进行验证以检测对R100 (包含SEQ ID NO:9的变体衣壳蛋白)衣壳和表达的蛋白的细胞反应。

[0134] 中和抗体测定

在感染前24小时,以 $3 \times 10^4$ 个细胞/孔的密度将2v6.11细胞铺板。在感染前,将编码由CAG启动子驱动的萤火虫荧光素酶的rAAV载体与个体血清样品一起在37℃下温育1小时,并随后以1,000的基因组MOI感染细胞。感染后48小时,使用Luc-Screen Extended-Glow 荧光素酶报告基因测定系统 (Invitrogen) 或ONE-Glo 荧光素酶测定系统 (Promega) 评估荧光素酶活性,并使用BioTek Cytation 3细胞成像多功能酶标仪和Gen5软件来量化。

[0135] 在参与研究之前,筛选非人灵长类动物(NHP)血清中针对R100的中和抗体的存在。当样品在1:10血清稀释度下导致AAV转导的中和小于50%时,将NHP纳入研究。

[0136] AAV制造

通过在HEK293细胞中瞬时转染产生重组R100病毒载体。细胞在补充有FBS的DMEM中培养,并维持在37℃、5% CO<sub>2</sub>环境中。使用聚乙烯亚胺(PEI)将细胞三次转染(有效负载、衣壳和辅助质粒)。转染后48-96小时,从细胞和/或上清液中收获病毒粒子,并经由微射流裂解细胞。酶促处理细胞裂解物和/或上清液以降解质粒和宿主细胞DNA,随后通过切向流过滤(TFF)澄清和浓缩。随后将TFF渗余物加载到亲和树脂柱上以便纯化。在pH梯度洗脱后,将亲和后材料进行缓冲液交换,随后通过阴离子交换色谱进一步纯化(如果需要的话)。随后将纯化的rAAV配制到含有0.001%聚山梨醇酯-20的DPBS中,无菌过滤,并灌装以产生rAAV药物产品。

[0137] 结果

4D-125递送是安全的,并导致NHP中治疗性转基因的表达

4D-125 (R100.GRK-cohRPGRorf15) 已进入1-2期临床试验。该产品的研究性新药(IND)申报数据包括在6个月良好实验室规范(GLP)毒理学和生物分布研究中的评估(表5)。通过玻璃体内注射单眼施用来注射30个NHP的总计30只眼。

[0138] 表5:良好实验室规范(GLP)毒理学和生物分布研究

| 4DMT<br>研究编号 | 批号            | 编号 | 性别 | 眼  | 剂量         | 存活期  |
|--------------|---------------|----|----|----|------------|------|
| 4D18-08      | N/A           | 1  | 雄性 | OD | 溶媒         | 3 周  |
|              | 4DEP000008.01 | 5  | 雄性 | OD | 1E+11 vg/眼 |      |
|              |               | 5  | 雄性 | OD | 1E+12 vg/眼 |      |
|              | N/A           | 1  | 雄性 | OD | 溶媒         | 13 周 |
|              | 4DEP000008.01 | 5  | 雄性 | OD | 1E+11 vg/眼 |      |
|              |               | 5  | 雄性 | OD | 1E+12 vg/眼 |      |
|              | N/A           | 1  | 雄性 | OD | 溶媒         | 26 周 |
|              | 4DEP000008.01 | 5  | 雄性 | OD | 1E+11 vg/眼 |      |
|              |               | 5  | 雄性 | OD | 1E+12 vg/眼 |      |

如通过临床观察、组织病理学、OCT或ERG所确定的那样,在任一剂量水平下用4D-

110未观察到显著毒性。将4D-110施用到单只眼中导致仅最小至轻度的前葡萄膜炎,其限于施用后即刻时期并在第3周消退(图9);在一些情况下,全身类固醇剂量瞬时提高。

[0139] 在所有时间点(第3周,左图;第13周,中图;第26周,右图),在经处理的眼的视网膜中存在非常高水平的载体基因组,这表明载体在眼部组织中的持久性(图10)。除视网膜外,在所有时间点在经处理的眼中来自房水、玻璃体液、虹膜/睫状体和视神经的样品内均检测到载体基因组。非眼部组织通常没有可检测的载体基因组,除了肝、脾和淋巴结中的低水平(图10)。在来自低剂量组和高剂量组的经处理的视网膜和虹膜/睫状体中检测到R100载体来源的转基因表达(图11)。基因表达是剂量依赖性的,并且从第3周至第13周增加,并在第26周保持稳定(图11,分别为左图、中图和右图)。在第26周末检测到非眼部载体表达(图11)。

[0140] 使用ELISpot测定来评估细胞免疫反应,没有动物对R100衣壳肽或转基因肽产生显著反应(数据未显示)。用4D-125给药的大多数动物在施用后产生抗衣壳抗体反应(数据未显示)。

[0141] 总结

4D-125 (R100.GRK-cohRPGRorf15) 近来已转化为遗传性视网膜疾病X连锁视网膜色素变性的临床试验(NCT04517149)。已经在GLP毒理学和生物分布研究中评估了该治疗产品(表5)。通过单眼施用注射总共30个NHP;总计注射30只NHP眼。没有报告显著的测试品相关不良事件或T细胞反应。观察到轻度至中度的一过性皮质类固醇反应性前葡萄膜炎。转基因表达定位于视网膜,并且在评估的全身器官的任一个中均未检测到表达。人类临床试验正在进行中,以便确定通过玻璃体内注射的该产品的安全性、药效学和功效(包括通过连续视野测试和光学相干断层扫描)。

[0142] 实施例5——在人X连锁视网膜色素变性患者中经由玻璃体内施用由R100递送的密码子优化的RPGRorf15 cDNA序列的安全性的评估

初始1期剂量递增安全性和耐受性数据汇总

临床试验设计和招募

临床试验采用标准的“3+3”剂量递增,其设计用于评估在两个剂量水平(3E11或1E12 vg/眼)下单次玻璃体内注射4D-125的安全性、耐受性和生物活性。在剂量递增群组中招募总共6个患者,每种剂量水平三个。患者接受逐渐减量的标准免疫抑制方案;调整由研究者确定。所述结果基于施用后1-9个月之间的数据截止值。

[0143] 初始耐受性和不良事件概况

如在治疗中出现的不良事件(AE)汇总表(表6)中概述的那样,4D-125在整个评估期内耐受性良好:

表6.不良事件汇总

|              |                |
|--------------|----------------|
| 招募的患者#       | 6              |
| 剂量           | 3E11或1E12 vg/眼 |
| 数据截止时的随访(月)  | 4-9个月          |
| 剂量限制性毒性(DLT) | 0(0%)          |
| 严重AE         | 0(0%)          |
| 任何≥3级的CTCAE  | 0(0%)          |

|                   |           |
|-------------------|-----------|
| 视网膜AE (任何等级)      | 0 (0%)    |
| 葡萄膜炎CTCAE 2级 (中度) | 1/6 (17%) |
| 葡萄膜炎CTCAE 1级 (轻度) | 2/6 (33%) |

临床评估

使用微视野检查 (MP) 测量视网膜敏感性并使用SD-OCT测量椭圆体带面积 (EZA) , 来评估初步生物活性。七名受试者 (中位年龄42.5岁; 范围27-56岁) 接受了4D-125 ( $3 \times 10^{11}$ vg/眼 (n=3) 和 $1 \times 10^{12}$ vg/眼 (n=4)) , 进行4.2-12.5个月的随访。眼内炎症 (4/7受试者) 为轻度或中度、一过性 (持续时间0.9-1.6个月) 和类固醇反应性的。大多数受试者患有晚期疾病, 仅2名在两只眼中在基线 (BL) 和至少4个月的随访时具有可测量的EZA和平均MP视网膜敏感性 (mMPRS) 。两名受试者在受治疗的眼对未治疗的眼中在mMPRS方面 (在9个月时+1.65dB对+0.25dB和在4个月时+0.50dB对+0.10dB; BL值1.5-3.2dB) 和获得 $\geq 7$ dB灵敏性的基因座数量方面 (在9个月时6对1和在4个月时3对0) 与BL相比均有更大的提高。对于两名受试者, EZA在受治疗的眼对未治疗的眼中与BL相比的相对降低较少 (在9个月时-12.4%对-16.2%和在6个月时-20.2%对-28.7%) 。

[0144] 在1/2期研究过程中, 密切监测患者的眼部和全身状态, 包括详细的眼科评估和视网膜成像以及必要时的血液测试和全身检查。进行各种视觉功能和解剖学评估以检测任何初步功效信号。这些评估包括但不限于椭圆体带 (EZ) 面积的测量、眼底自发荧光、微视野检查、静态自动视野检查和最佳矫正视力 (BCVA) 。

[0145] 结论

玻璃体内施用4D-125耐受性良好, 具有轻度或中度、一过性和类固醇反应性的眼内炎症。基于微视野检查和SD-OCT, 在2名可评估的剂量递增受试者中观察到生物活性的初步迹象。这些发现支持在正在进行的1/2期研究中在患有病情较轻的晚期疾病 (less advanced disease) 的XLRP受试者中以 $1 \times 10^{12}$ vg/眼的剂量进行剂量扩展。

[0146] 虽然已经就优选实施方案描述了本发明的材料和方法, 对于本领域技术人员显而易见的是, 在不脱离本发明的概念、精神和范围的情况下, 可以对本文中描述的方法施以改变。对于本领域技术人员显而易见的所有此类类似的替代和修改均被认为在本发明的精神、范围和概念内。

- <110> 4D Molecular Therapeutics, Inc.  
 <120> 密码子优化的RPGRORF15基因及其用途  
 <130> 090400-5012 W0  
 <150> US 63/073,843  
 <151> 2020-09-02  
 <160> 10  
 <170> PatentIn version 3.5  
 <210> 1  
 <211> 3459  
 <212> DNA  
 <213> 人工序列  
 <220>  
 <223> 密码子优化的RPGRorf15  
 <400> 1

```

atgagagaac ccgaggaact gatgccccgac tctggcgccg tgtttacctt cggcaagagc      60
aagttcgccg agaacaaccc cggcaagttc tggttcaaga acgacgtgcc agtgcacctg      120
agctgcggag atgaacactc tgccgtggtc accggcaaca acaagctgta catgttcggc      180
agcaacaact ggggccagct cggcctggga tctaagtctg ccatcagcaa gcctacctgc      240
gtgaaggccc tgaagcctga gaaagtgaat ctggccgcct gcggcagaaa tcacaccctg      300
gtttctaccg aaggcggcaa tgtgtatgcc accggcggaa acaatgaggg acagcttgga      360
ctgggcgaca ccgaggaaag aaacaccttc cacgtgatca gctttttcac cagcgagcac      420
aagatcaagc agctgagcgc cggctctaata acctctgccg ctctgacaga ggacggcaga      480
ctgtttatgt ggggcgacaa ttctgagggc cagatcggac tgaagaacgt gtccaatgtg      540
tgcgtgcccc agcaagtgc aatcggcaag cctgtgtctt ggatcagctg cggctactac      600
cacagcgcc tttgtgacaac cgatggcgag ctgtatgtgt tcggcgagcc agagaatggc      660
aagctgggac tgcctaacca gctgctgggc aatcacagaa cccctcagct ggtgtctgag      720
atccccgaaa aagtgatcca ggtggcctgt ggcgagagc acacagtggg gctgacagag      780
aatgccgtgt acacctttgg cctgggccag tttggacaac tcggactggg aaccttcctg      840
ttcgagacaa gcgagcccaa agtgatcgag aacatccggg accagaccat cagctacatc      900
agctgtggcg agaaccacac agccctgata acagacatcg gcctgatgta cacattcggc      960
gacggaaggc atggaaaagct cggacttggc ctggaaaact tcaccaacca cttcatccct     1020
acgctgtgca gcaacttcct gcggttcatt gtgaagctgg tggcctgcgg aggatgccac     1080
atggtggttt ttgctgcccc tcacagaggc gtggccaaag agattgagtt cgacgagatc     1140
aacgatacct gcctgagcgt ggccaccttc ctgccttaca gcagcctgac atctggcaac     1200
gtgctgcaga ggacactgag cgccagaatg cgcagacggg aaagagagag aagccccgac     1260
agcttcagca tgagaagaac cctgcctcca atcgagggca cactgggcct gtctgcctgc     1320
tttctgccta acagcgtgtt ccccgatgac agcgagagaa acctgcaaga gagcgtgctg     1380
agcgagcagg atctgatgca gcctgaggaa cccgactacc tgctggacga gatgacccaa     1440
  
```

|             |            |             |            |            |            |      |
|-------------|------------|-------------|------------|------------|------------|------|
| gaggccgaga  | tcgacaacag | cagcacagt   | gaaagcctgg | gcgagacaac | cgacatcctg | 1500 |
| aacatgaccc  | acatcatgag | cctgaacagc  | aacgagaagt | ctctgaagct | gagccccgtg | 1560 |
| cagaagcaga  | agaagcagca | gaccatcggc  | gagctgacac | aggatactgc | cctgaccgag | 1620 |
| aacgacgaca  | gcgacgagta | cgaagagatg  | agcgagatga | aggaaggcaa | ggcctgcaag | 1680 |
| cagcacgtgt  | cccagggcat | ctttatgacc  | cagcctgcc  | ccaccatcga | ggccttttcc | 1740 |
| gacgaggaag  | tggaaatccc | cgaggaaaaa  | gagggcgccg | aggacagcaa | aggcaacggc | 1800 |
| attgaggaac  | aagaggtgga | agccaacgaa  | gagaacgtga | aggtgcacgg | cggacggaaa | 1860 |
| gaaaagaccg  | agatcctgag | cgacgacctg  | accgataagg | ccgaggtttc | cgagggcaaa | 1920 |
| gccaaagtctg | tgggagaagc | cgaggatgga  | cctgaaggcc | gcggagatgg | aacctgtgaa | 1980 |
| gaaggatcta  | gcggagccga | gcaactggcag | gatgaggaac | gcgagaaggg | cgagaaagac | 2040 |
| aaaggcagag  | gcgagatgga | aagacccggc  | gagggcgaaa | aagagctggc | cgagaaagag | 2100 |
| gaatggaaga  | aacgcgacgg | cgaagaacaa  | gagcagaaag | aaagagagca | gggccaccag | 2160 |
| aaagaacgga  | atcaagagat | ggaagaaggc  | ggcgaggaag | aacacggcga | aggggaagaa | 2220 |
| gaggaaggcg  | accgagagga | agaagaagag  | aaagaaggcg | aaggcaaaga | agaaggcgag | 2280 |
| ggcgaagagg  | tggaaggcga | gcgtgaaaaa  | gaagagggcg | aacgcaagaa | agaagaacgc | 2340 |
| gccggaaaag  | aggaaaaagg | cgaggaagag  | ggcgaccaag | gcgaaggcga | ggaagaagaa | 2400 |
| actgaaggca  | gaggggaaga | gaaagaggaa  | ggcggcgaa  | tcgaaggcgg | agaggttgaa | 2460 |
| gaaggcaaag  | gcgagcgaga | agaggaagaa  | gaagaaggcg | aaggcgagga | agaggaaggc | 2520 |
| gaaggcgaag  | aggaagaagg | cgaaggggaa  | gaagaagaag | gcgaaggcaa | gggcgaagag | 2580 |
| gagggcgaag  | aaggcgaggg | cgaagaggag  | ggcgaagaag | gcgaaggcga | gggcgaagaa | 2640 |
| gaagaaggcg  | aaggcgaagg | cgaggaagaa  | ggcgaaggcg | aagggaaga  | agaggaaggc | 2700 |
| gaaggcgaag  | gcgaagaaga | aggcgaaggc  | gagggcgaag | aggaagaagg | cgaaggcaaa | 2760 |
| ggggaagaag  | aaggcgagga | aggcgaaggc  | gaaggcgagg | aagaagaagg | cgaaggcgag | 2820 |
| ggcgaagatg  | gcgaaggcga | aggcgaagag  | gaagagggcg | agtgggaggg | cgaagaagag | 2880 |
| gaaggcgaag  | gcgagggcga | agaggaaggc  | gaaggcgagg | gcgaagaagg | cgaaggcgaa | 2940 |
| ggcgaggaag  | aggaaggcga | aggcgaaggg  | gaagaagaag | agggcgaaga | agaaggcgaa | 3000 |
| gaggaaggcg  | aagggaaga  | agaaggcgaa  | ggcgaaggcg | aagaagagga | agaggcgcaa | 3060 |
| gttgaaggcg  | aggttgaggg | cgaagaaggc  | gaaggcgaag | gggaagaaga | agaaggcgag | 3120 |
| gaagaagggg  | aagagagaga | aaaagaaggc  | gagggcgaag | aaaaccgccg | gaaccgcgaa | 3180 |
| gaggaagagg  | aagaagaggg | caagtaccaa  | gagactggcg | aggaagagaa | cgagcggcag | 3240 |
| gatggcgaag  | agtacaagaa | ggtgtccaag  | atcaagggca | gcgtgaagta | cggcaagcac | 3300 |
| aagacctacc  | agaagaagtc | cgtcaccaac  | acgcaaggca | atggaaaaga | acagcggagc | 3360 |
| aagatgcccc  | tgcatgctca | gaggtgtctg  | aagaatggcc | ctagcggcag | caagaaattc | 3420 |
| tggaacaatg  | tgctgcccc  | ctacctcgag  | ctgaagtga  |            |            | 3459 |
| <210>       | 2          |             |            |            |            |      |
| <211>       | 1152       |             |            |            |            |      |
| <212>       | PRT        |             |            |            |            |      |
| <213>       | 智人         |             |            |            |            |      |
| <400>       | 2          |             |            |            |            |      |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Met | Arg | Glu | Pro | Glu | Glu | Leu | Met | Pro | Asp | Ser | Gly | Ala | Val | Phe | Thr |
| 1   |     |     |     | 5   |     |     |     |     | 10  |     |     |     |     | 15  |     |
| Phe | Gly | Lys | Ser | Lys | Phe | Ala | Glu | Asn | Asn | Pro | Gly | Lys | Phe | Trp | Phe |
|     |     |     | 20  |     |     |     |     | 25  |     |     |     |     | 30  |     |     |
| Lys | Asn | Asp | Val | Pro | Val | His | Leu | Ser | Cys | Gly | Asp | Glu | His | Ser | Ala |
|     |     | 35  |     |     |     |     | 40  |     |     |     |     | 45  |     |     |     |
| Val | Val | Thr | Gly | Asn | Asn | Lys | Leu | Tyr | Met | Phe | Gly | Ser | Asn | Asn | Trp |
|     | 50  |     |     |     |     | 55  |     |     |     |     | 60  |     |     |     |     |
| Gly | Gln | Leu | Gly | Leu | Gly | Ser | Lys | Ser | Ala | Ile | Ser | Lys | Pro | Thr | Cys |
| 65  |     |     |     |     | 70  |     |     |     | 75  |     |     |     |     |     | 80  |
| Val | Lys | Ala | Leu | Lys | Pro | Glu | Lys | Val | Lys | Leu | Ala | Ala | Cys | Gly | Arg |
|     |     |     |     | 85  |     |     |     | 90  |     |     |     |     | 95  |     |     |
| Asn | His | Thr | Leu | Val | Ser | Thr | Glu | Gly | Gly | Asn | Val | Tyr | Ala | Thr | Gly |
|     |     |     | 100 |     |     |     |     | 105 |     |     |     |     | 110 |     |     |
| Gly | Asn | Asn | Glu | Gly | Gln | Leu | Gly | Leu | Gly | Asp | Thr | Glu | Glu | Arg | Asn |
|     |     | 115 |     |     |     |     | 120 |     |     |     |     | 125 |     |     |     |
| Thr | Phe | His | Val | Ile | Ser | Phe | Phe | Thr | Ser | Glu | His | Lys | Ile | Lys | Gln |
|     | 130 |     |     |     |     | 135 |     |     |     |     | 140 |     |     |     |     |
| Leu | Ser | Ala | Gly | Ser | Asn | Thr | Ser | Ala | Ala | Leu | Thr | Glu | Asp | Gly | Arg |
| 145 |     |     |     |     | 150 |     |     |     | 155 |     |     |     |     |     | 160 |
| Leu | Phe | Met | Trp | Gly | Asp | Asn | Ser | Glu | Gly | Gln | Ile | Gly | Leu | Lys | Asn |
|     |     |     |     | 165 |     |     |     | 170 |     |     |     |     | 175 |     |     |
| Val | Ser | Asn | Val | Cys | Val | Pro | Gln | Gln | Val | Thr | Ile | Gly | Lys | Pro | Val |
|     |     |     | 180 |     |     |     |     | 185 |     |     |     |     | 190 |     |     |
| Ser | Trp | Ile | Ser | Cys | Gly | Tyr | Tyr | His | Ser | Ala | Phe | Val | Thr | Thr | Asp |
|     | 195 |     |     |     |     |     | 200 |     |     |     |     | 205 |     |     |     |
| Gly | Glu | Leu | Tyr | Val | Phe | Gly | Glu | Pro | Glu | Asn | Gly | Lys | Leu | Gly | Leu |
|     | 210 |     |     |     |     | 215 |     |     |     |     | 220 |     |     |     |     |
| Pro | Asn | Gln | Leu | Leu | Gly | Asn | His | Arg | Thr | Pro | Gln | Leu | Val | Ser | Glu |
| 225 |     |     |     |     | 230 |     |     |     | 235 |     |     |     |     |     | 240 |
| Ile | Pro | Glu | Lys | Val | Ile | Gln | Val | Ala | Cys | Gly | Gly | Glu | His | Thr | Val |
|     |     |     |     | 245 |     |     |     | 250 |     |     |     |     | 255 |     |     |
| Val | Leu | Thr | Glu | Asn | Ala | Val | Tyr | Thr | Phe | Gly | Leu | Gly | Gln | Phe | Gly |
|     |     |     | 260 |     |     |     |     | 265 |     |     |     |     | 270 |     |     |
| Gln | Leu | Gly | Leu | Gly | Thr | Phe | Leu | Phe | Glu | Thr | Ser | Glu | Pro | Lys | Val |
|     |     | 275 |     |     |     |     | 280 |     |     |     |     | 285 |     |     |     |
| Ile | Glu | Asn | Ile | Arg | Asp | Gln | Thr | Ile | Ser | Tyr | Ile | Ser | Cys | Gly | Glu |
|     | 290 |     |     |     |     | 295 |     |     |     |     | 300 |     |     |     |     |
| Asn | His | Thr | Ala | Leu | Ile | Thr | Asp | Ile | Gly | Leu | Met | Tyr | Thr | Phe | Gly |

|                                                                 |     |     |     |
|-----------------------------------------------------------------|-----|-----|-----|
| 305                                                             | 310 | 315 | 320 |
| Asp Gly Arg His Gly Lys Leu Gly Leu Gly Leu Glu Asn Phe Thr Asn |     |     |     |
|                                                                 | 325 | 330 | 335 |
| His Phe Ile Pro Thr Leu Cys Ser Asn Phe Leu Arg Phe Ile Val Lys |     |     |     |
|                                                                 | 340 | 345 | 350 |
| Leu Val Ala Cys Gly Gly Cys His Met Val Val Phe Ala Ala Pro His |     |     |     |
|                                                                 | 355 | 360 | 365 |
| Arg Gly Val Ala Lys Glu Ile Glu Phe Asp Glu Ile Asn Asp Thr Cys |     |     |     |
|                                                                 | 370 | 375 | 380 |
| Leu Ser Val Ala Thr Phe Leu Pro Tyr Ser Ser Leu Thr Ser Gly Asn |     |     |     |
| 385                                                             | 390 | 395 | 400 |
| Val Leu Gln Arg Thr Leu Ser Ala Arg Met Arg Arg Arg Glu Arg Glu |     |     |     |
|                                                                 | 405 | 410 | 415 |
| Arg Ser Pro Asp Ser Phe Ser Met Arg Arg Thr Leu Pro Pro Ile Glu |     |     |     |
|                                                                 | 420 | 425 | 430 |
| Gly Thr Leu Gly Leu Ser Ala Cys Phe Leu Pro Asn Ser Val Phe Pro |     |     |     |
|                                                                 | 435 | 440 | 445 |
| Arg Cys Ser Glu Arg Asn Leu Gln Glu Ser Val Leu Ser Glu Gln Asp |     |     |     |
|                                                                 | 450 | 455 | 460 |
| Leu Met Gln Pro Glu Glu Pro Asp Tyr Leu Leu Asp Glu Met Thr Lys |     |     |     |
| 465                                                             | 470 | 475 | 480 |
| Glu Ala Glu Ile Asp Asn Ser Ser Thr Val Glu Ser Leu Gly Glu Thr |     |     |     |
|                                                                 | 485 | 490 | 495 |
| Thr Asp Ile Leu Asn Met Thr His Ile Met Ser Leu Asn Ser Asn Glu |     |     |     |
|                                                                 | 500 | 505 | 510 |
| Lys Ser Leu Lys Leu Ser Pro Val Gln Lys Gln Lys Lys Gln Gln Thr |     |     |     |
|                                                                 | 515 | 520 | 525 |
| Ile Gly Glu Leu Thr Gln Asp Thr Ala Leu Thr Glu Asn Asp Asp Ser |     |     |     |
|                                                                 | 530 | 535 | 540 |
| Asp Glu Tyr Glu Glu Met Ser Glu Met Lys Glu Gly Lys Ala Cys Lys |     |     |     |
| 545                                                             | 550 | 555 | 560 |
| Gln His Val Ser Gln Gly Ile Phe Met Thr Gln Pro Ala Thr Thr Ile |     |     |     |
|                                                                 | 565 | 570 | 575 |
| Glu Ala Phe Ser Asp Glu Glu Val Glu Ile Pro Glu Glu Lys Glu Gly |     |     |     |
|                                                                 | 580 | 585 | 590 |
| Ala Glu Asp Ser Lys Gly Asn Gly Ile Glu Glu Gln Glu Val Glu Ala |     |     |     |
|                                                                 | 595 | 600 | 605 |
| Asn Glu Glu Asn Val Lys Val His Gly Gly Arg Lys Glu Lys Thr Glu |     |     |     |
| 610                                                             | 615 | 620 |     |

|     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |     |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Ile | Leu | Ser | Asp | Asp | Leu | Thr | Asp | Lys | Ala | Glu | Val | Ser | Glu | Gly | Lys |
| 625 |     |     |     |     | 630 |     |     |     |     | 635 |     |     |     |     | 640 |
| Ala | Lys | Ser | Val | Gly | Glu | Ala | Glu | Asp | Gly | Pro | Glu | Gly | Arg | Gly | Asp |
|     |     |     |     | 645 |     |     |     |     | 650 |     |     |     |     | 655 |     |
| Gly | Thr | Cys | Glu | Glu | Gly | Ser | Ser | Gly | Ala | Glu | His | Trp | Gln | Asp | Glu |
|     |     |     | 660 |     |     |     |     | 665 |     |     |     |     | 670 |     |     |
| Glu | Arg | Glu | Lys | Gly | Glu | Lys | Asp | Lys | Gly | Arg | Gly | Glu | Met | Glu | Arg |
|     | 675 |     |     |     |     |     | 680 |     |     |     |     | 685 |     |     |     |
| Pro | Gly | Glu | Gly | Glu | Lys | Glu | Leu | Ala | Glu | Lys | Glu | Glu | Trp | Lys | Lys |
|     | 690 |     |     |     |     | 695 |     |     |     |     | 700 |     |     |     |     |
| Arg | Asp | Gly | Glu | Glu | Gln | Glu | Gln | Lys | Glu | Arg | Glu | Gln | Gly | His | Gln |
| 705 |     |     |     | 710 |     |     |     |     | 715 |     |     |     |     |     | 720 |
| Lys | Glu | Arg | Asn | Gln | Glu | Met | Glu | Glu | Gly | Gly | Glu | Glu | Glu | His | Gly |
|     |     |     | 725 |     |     |     |     |     | 730 |     |     |     |     | 735 |     |
| Glu | Gly | Glu | Glu | Glu | Glu | Gly | Asp | Arg | Glu | Glu | Glu | Glu | Glu | Lys | Glu |
|     |     |     | 740 |     |     |     |     | 745 |     |     |     |     | 750 |     |     |
| Gly | Glu | Gly | Lys | Glu | Glu | Gly | Glu | Gly | Glu | Glu | Val | Glu | Gly | Glu | Arg |
|     | 755 |     |     |     |     |     | 760 |     |     |     |     | 765 |     |     |     |
| Glu | Lys | Glu | Glu | Gly | Glu | Arg | Lys | Lys | Glu | Glu | Arg | Ala | Gly | Lys | Glu |
|     | 770 |     |     |     |     | 775 |     |     |     |     | 780 |     |     |     |     |
| Glu | Lys | Gly | Glu | Glu | Glu | Gly | Asp | Gln | Gly | Glu | Gly | Glu | Glu | Glu | Glu |
| 785 |     |     |     | 790 |     |     |     |     | 795 |     |     |     |     |     | 800 |
| Thr | Glu | Gly | Arg | Gly | Glu | Glu | Lys | Glu | Glu | Gly | Gly | Glu | Val | Glu | Gly |
|     |     |     | 805 |     |     |     |     |     | 810 |     |     |     |     | 815 |     |
| Gly | Glu | Val | Glu | Glu | Gly | Lys | Gly | Glu | Arg | Glu | Glu | Glu | Glu | Glu | Glu |
|     |     |     | 820 |     |     |     |     | 825 |     |     |     |     | 830 |     |     |
| Gly | Glu | Gly | Glu | Glu | Glu | Glu | Gly | Glu | Gly | Glu | Glu | Glu | Glu | Gly | Glu |
|     | 835 |     |     |     |     |     | 840 |     |     |     |     | 845 |     |     |     |
| Gly | Glu | Glu | Glu | Glu | Gly | Glu | Gly | Lys | Gly | Glu | Glu | Glu | Gly | Glu | Glu |
|     | 850 |     |     |     |     | 855 |     |     |     | 860 |     |     |     |     |     |
| Gly | Glu | Gly | Glu | Glu | Glu | Gly | Glu | Glu | Gly | Glu | Gly | Glu | Gly | Glu | Glu |
| 865 |     |     |     | 870 |     |     |     |     | 875 |     |     |     |     | 880 |     |
| Glu | Glu | Gly | Glu | Gly | Glu | Gly | Glu | Glu | Glu | Gly | Glu | Gly | Glu | Gly | Glu |
|     |     |     | 885 |     |     |     |     | 890 |     |     |     |     | 895 |     |     |
| Glu | Glu | Glu | Gly | Glu | Gly | Glu | Gly | Glu | Glu | Glu | Gly | Glu | Gly | Glu | Gly |
|     |     |     | 900 |     |     |     |     | 905 |     |     |     | 910 |     |     |     |
| Glu | Glu | Glu | Glu | Gly | Glu | Gly | Lys | Gly | Glu | Glu | Glu | Gly | Glu | Glu | Gly |
|     | 915 |     |     |     |     |     | 920 |     |     |     |     | 925 |     |     |     |
| Glu | Gly | Glu | Gly | Glu | Glu | Glu | Glu | Gly | Glu | Gly | Glu | Gly | Glu | Asp | Gly |

|                                                                   |      |      |
|-------------------------------------------------------------------|------|------|
| 930                                                               | 935  | 940  |
| Glu Gly Glu Gly Glu Glu Glu Glu Gly Glu Trp Glu Gly Glu Glu Glu   |      |      |
| 945                                                               | 950  | 955  |
| Glu Gly Glu Gly Glu Gly Glu Glu Glu Gly Glu Gly Glu Gly Glu Glu   |      |      |
| 965                                                               | 970  | 975  |
| Gly Glu Gly Glu Gly Glu Glu Glu Glu Gly Glu Gly Glu Gly Glu Glu   |      |      |
| 980                                                               | 985  | 990  |
| Glu Glu Gly Glu Glu Glu Gly Glu Glu Glu Gly Glu Gly Glu Glu Glu   |      |      |
| 995                                                               | 1000 | 1005 |
| Gly Glu Gly Glu Gly Glu Glu Glu Glu Glu Gly Glu Val Glu Gly       |      |      |
| 1010                                                              | 1015 | 1020 |
| Glu Val Glu Gly Glu Glu Gly Glu Gly Glu Gly Glu Glu Glu Glu       |      |      |
| 1025                                                              | 1030 | 1035 |
| Gly Glu Glu Glu Gly Glu Glu Arg Glu Lys Glu Gly Glu Gly Glu       |      |      |
| 1040                                                              | 1045 | 1050 |
| Glu Asn Arg Arg Asn Arg Glu Glu Glu Glu Glu Glu Glu Gly Lys       |      |      |
| 1055                                                              | 1060 | 1065 |
| Tyr Gln Glu Thr Gly Glu Glu Glu Asn Glu Arg Gln Asp Gly Glu       |      |      |
| 1070                                                              | 1075 | 1080 |
| Glu Tyr Lys Lys Val Ser Lys Ile Lys Gly Ser Val Lys Tyr Gly       |      |      |
| 1085                                                              | 1090 | 1095 |
| Lys His Lys Thr Tyr Gln Lys Lys Ser Val Thr Asn Thr Gln Gly       |      |      |
| 1100                                                              | 1105 | 1110 |
| Asn Gly Lys Glu Gln Arg Ser Lys Met Pro Val Gln Ser Lys Arg       |      |      |
| 1115                                                              | 1120 | 1125 |
| Leu Leu Lys Asn Gly Pro Ser Gly Ser Lys Lys Phe Trp Asn Asn       |      |      |
| 1130                                                              | 1135 | 1140 |
| Val Leu Pro His Tyr Leu Glu Leu Lys                               |      |      |
| 1145                                                              | 1150 |      |
| <210> 3                                                           |      |      |
| <211> 3459                                                        |      |      |
| <212> DNA                                                         |      |      |
| <213> 智人                                                          |      |      |
| <400> 3                                                           |      |      |
| atgagggagc cggaagagct gatgcccgat tcgggtgctg tgtttacatt tgggaaaagt |      | 60   |
| aaatttgctg aaaataatcc cggtaaattc tggtttaaaa atgatgtccc tgtacatctt |      | 120  |
| tcatgtggag atgaacattc tgctgttggt accggaaata ataaacttta catgtttggc |      | 180  |
| agtaacaact ggggtcagtt aggattagga tcaaagtcag ccatcagcaa gccaacatgt |      | 240  |
| gtcaaagctc taaaacctga aaaagtga ttagctgcct gtggaaggaa ccacaccctg   |      | 300  |

|             |             |             |            |            |            |      |
|-------------|-------------|-------------|------------|------------|------------|------|
| gtgtcaacag  | aaggaggcaa  | tgtatatgca  | actggtggaa | ataatgaagg | acagttgggg | 360  |
| cttggtgaca  | ccgaagaaag  | aaacactttt  | catgtaatta | gcttttttac | atccgagcat | 420  |
| aagattaagc  | agctgtctgc  | tggatctaata | acttcagctg | ccctaactga | ggatggaaga | 480  |
| ctttttatgt  | ggggtgacaa  | ttccgaaggg  | caaattgggt | taaaaaatgt | aagtaatgtc | 540  |
| tgtgtccctc  | agcaagtgc   | cattgggaaa  | cctgtctcct | ggatctcttg | tggatattac | 600  |
| cattcagctt  | ttgtaacaac  | agatggtgag  | ctatatgtgt | ttggagaacc | tgagaatggg | 660  |
| aagttaggtc  | ttcccaatca  | gctcctgggc  | aatcacagaa | caccccagct | ggtgtctgaa | 720  |
| attccggaga  | aggtgatcca  | agtagcctgt  | ggtggagagc | atactgtggt | tctcacggag | 780  |
| aatgctgtgt  | ataccttttg  | gctgggacaa  | tttggtcagc | tgggtcttgg | cacttttctt | 840  |
| tttgaaaactt | cagaacccaa  | agtcattgag  | aatattaggg | atcaaacaat | aagttatatt | 900  |
| tcttgtggag  | aaaaatcacac | agctttgata  | acagatatcg | gccttatgta | tacttttggg | 960  |
| gatggtcgcc  | acggaaaatt  | aggacttggg  | ctggagaatt | ttaccaatca | cttcattcct | 1020 |
| actttgtgct  | ctaatttttt  | gaggtttata  | gttaaattgg | ttgcttgtgg | tggatgtcac | 1080 |
| atggtagttt  | ttgctgctcc  | tcatcgtggt  | gtggcaaaag | aaattgaatt | cgatgaaata | 1140 |
| aatgatactt  | gcttatctgt  | ggcgactttt  | ctgccgtata | gcagtttaac | ctcaggaaat | 1200 |
| gtactgcaga  | ggactctatc  | agcacgtatg  | cggcgaagag | agaggagag  | gtctccagat | 1260 |
| tctttttcaa  | tgaggagaac  | actacctcca  | atagaaggga | ctcttggcct | ttctgcttgt | 1320 |
| tttctcccca  | attcagtctt  | tccacgatgt  | tctgagagaa | acctccaaga | gagtgtctta | 1380 |
| tctgaacagg  | acctcatgca  | gccagaggaa  | ccagattatt | tgctagatga | aatgacccaa | 1440 |
| gaagcagaga  | tagataattc  | ttcaactgta  | gaaagccttg | gagaaactac | tgatatctta | 1500 |
| aacatgacac  | acatcatgag  | cctgaattcc  | aatgaaaagt | cattaaaatt | atcaccagtt | 1560 |
| cagaaacaaa  | agaaacaaca  | aacaattggg  | gaactgacgc | aggatacagc | tcttactgaa | 1620 |
| aacgatgata  | gtgatgaata  | tgaagaaatg  | tcagaaatga | aagaagggaa | agcatgtaaa | 1680 |
| caacatgtgt  | cacaagggat  | tttcatgacg  | cagccagcta | cgactatcga | agcattttca | 1740 |
| gatgaggaag  | tagagatccc  | agaggagaag  | gaaggagcag | aggattcaaa | aggaaatgga | 1800 |
| atagaggagc  | aagaggtaga  | agcaaatgag  | gaaaatgtga | aggtgcatgg | aggaagaaag | 1860 |
| gagaaaacag  | agatcctatc  | agatgacctt  | acagacaaag | cagaggtgag | tgaaggcaag | 1920 |
| gcaaaatcag  | tgggagaagc  | agaggatggg  | cctgaaggta | gaggggatgg | aacctgtgag | 1980 |
| gaaggtagtt  | caggagcaga  | acactggcaa  | gatgaggaga | gggagaaggg | ggagaaagac | 2040 |
| aagggtagag  | gagaaaatgga | gaggccagga  | gaggagagag | aggaactagc | agagaaggaa | 2100 |
| gaatggaaga  | agagggatgg  | ggaagagcag  | gagcaaaagg | agaggagca  | gggccatcag | 2160 |
| aaggaaaaga  | accaagagat  | ggaggaggga  | ggggaggagg | agcatggaga | aggagaagaa | 2220 |
| gaggaggagg  | acagagaaga  | ggaagaagag  | aaggaggagg | aagggaaga  | ggaaggagaa | 2280 |
| ggggaagaag  | tggagggaga  | acgtgaaaag  | gaggaaggag | agaggaaaaa | ggaggaaaga | 2340 |
| gcgggggaagg | aggagaaagg  | agaggaagaa  | ggagaccaag | gagaggggga | agaggaggaa | 2400 |
| acagaggggga | gaggggagga  | aaaagaggag  | ggaggggaag | tagagggagg | ggaagtagag | 2460 |
| gaggggaaag  | gagagaggga  | agaggaagag  | gaggagggtg | agggggaaga | ggaggaaggg | 2520 |
| gagggggaag  | aggaggaagg  | ggagggggaa  | gaggaggaag | gagaagggaa | aggggaggaa | 2580 |
| gaaggggaag  | aaggagaagg  | ggaggaagaa  | ggggaggaag | gagaagggga | gggggaagag | 2640 |

|                                                                    |      |
|--------------------------------------------------------------------|------|
| gaggaaggag aaggggaggg agaagaggaa ggagaagggg agggagaaga ggaggaagga  | 2700 |
| gaaggggagg gagaagagga aggagaaggg gagggagaag aggaggaagg agaagggaaa  | 2760 |
| ggggaggagg aaggagagga aggagaaggg gagggggaag aggaggaagg agaaggggaa  | 2820 |
| ggggaggatg gagaagggga gggggaagag gaggaaggag aatgggaggg ggaagaggag  | 2880 |
| gaaggagaag gggaggggga agaggaagga gaaggggaag gggaggaagg agaaggggag  | 2940 |
| ggggaagagg aggaaggaga aggggagggg gaagaggagg aaggggaaga agaaggggag  | 3000 |
| gaagaaggag agggagagga agaaggggag ggagaagggg aggaagaaga ggaaggggaa  | 3060 |
| gtggaagggg aggtggaagg ggaggaagga gagggggaag gagaggaaga ggaaggagag  | 3120 |
| gaggaaggag aagaaaaggga aaaggagggg gaaggagaag aaaacaggag gaacagagaa | 3180 |
| gaggaggagg aagaagaggg gaagtatcag gagacaggcg aagaagagaa tgaaaggcag  | 3240 |
| gatggagagg agtacaaaaa agtgagcaaa ataaaaggat ctgtgaaata tggcaaacat  | 3300 |
| aaaacatatc aaaaaaagtc agttactaac acacagggaa atgggaaaga gcagaggtcc  | 3360 |
| aaaatgccag tccagtcaaa acgactttta aaaaacgggc catcaggttc caaaaagttc  | 3420 |
| tggaataatg tattaccaca ttacttgaa ttgaagtaa                          | 3459 |

&lt;210&gt; 4

&lt;211&gt; 199

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; hGRK启动子

&lt;400&gt; 4

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| gggccccaga agcctggtgg ttgtttgtcc ttctcagggg aaaagtgagg cggcccccttg | 60  |
| gaggaagggg ccgggcagaa tgatctaatac ggattccaag cagctcaggg gattgtcttt | 120 |
| ttctagcacc ttcttgccac tcctaagcgt cctccgtgac cccggctggg atttagcctg  | 180 |
| gtgctgtgtc agccccggg                                               | 199 |

&lt;210&gt; 5

&lt;211&gt; 4303

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; cohRPGR表达盒

&lt;400&gt; 5

|                                                                    |     |
|--------------------------------------------------------------------|-----|
| ttggccactc cctctctgcg cgctcgtcgc ctactgagg ccgggcgacc aaaggtcgcc   | 60  |
| cgacgcccgg gctttgcccc ggcggcctca gtgagcgagc gagcgcgag agagggagtg   | 120 |
| gccaactcca tcactagggg ttctatcga ttgaattccc cggggatccg ggccccagaa   | 180 |
| gcctgggtgtg tgtttgtcct tctcagggga aaagtgaggc ggcccccttg aggaaggggc | 240 |
| cgggcagaat gatctaatac gattccaagc agctcagggg attgtctttt tctagcacct  | 300 |
| tcttgccact cctaagcgtc ctccgtgacc ccggctggga tttagcctgg tgctgtgtca  | 360 |
| gccccgggtc tagagtcgac ctgcagaagc ttccaccatg agagaacccg aggaactgat  | 420 |

|            |            |            |             |            |            |      |
|------------|------------|------------|-------------|------------|------------|------|
| gccccactct | ggcgccgtgt | ttaccttcgg | caagagcaag  | ttcgccgaga | acaacccccg | 480  |
| caagtctctg | ttcaagaacg | acgtgccagt | gcacctgagc  | tgcggagatg | aacactctgc | 540  |
| cgtggtcacc | ggcaacaaca | agctgtacat | gttcggcagc  | aacaactggg | gccagctcgg | 600  |
| cctgggatct | aagtctgcc  | tcagcaagcc | tacctgcgtg  | aaggccctga | agcctgagaa | 660  |
| agtgaaactg | gccgcctgcg | gcagaaatca | caccctgggt  | tctaccgaag | gcggcaatgt | 720  |
| gtatgccacc | ggcggaacaa | atgagggaca | gcttggactg  | ggcgacaccg | aggaaagaaa | 780  |
| caccttccac | gtgatcagct | ttttcaccag | cgagcacaag  | atcaagcagc | tgagcgccgg | 840  |
| ctctaatacc | tctgccgctc | tgacagagga | cggcagactg  | tttatgtggg | gcgacaattc | 900  |
| tgagggccag | atcggactga | agaacgtgtc | caatgtgtgc  | gtgccccagc | aagtgacaat | 960  |
| cggcaagcct | gtgtcttga  | tcagctgcgg | ctactaccac  | agcgcccttg | tgacaaccga | 1020 |
| tgcgagctg  | tatgtgttcg | gcgagccaga | gaatggcaag  | ctgggactgc | ctaaccagct | 1080 |
| gctgggcaat | cacagaaccc | ctcagctggg | gtctgagatc  | cccgaaaaag | tgatccaggt | 1140 |
| ggcctgtggc | ggagagcaca | cagtgggtgt | gacagagaat  | gccgtgtaca | cctttggcct | 1200 |
| gggccagttt | ggacaactcg | gactgggaac | cttcctgttc  | gagacaagcg | agcccaaagt | 1260 |
| gatcgagaac | atccgggacc | agaccatcag | ctacatcagc  | tgtggcgaga | accacacagc | 1320 |
| cctgatcaca | gacatcggcc | tgatgtacac | attcggcgac  | ggaaggcatg | gaaagctcgg | 1380 |
| acttggcctg | gaaaacttca | ccaaccactt | catccctacg  | ctgtgcagca | acttcctgcg | 1440 |
| gttcattgtg | aagctggtgg | cctgcggagg | atgccacatg  | gtggtttttg | ctgcccctca | 1500 |
| cagaggcgtg | gccaagaga  | ttgagttcga | cgagatcaac  | gatacctgcc | tgagcgtggc | 1560 |
| caccttcctg | ccttacagca | gcctgacatc | tggcaacgtg  | ctgcagagga | cactgagcgc | 1620 |
| cagaatgcgc | agacgggaaa | gagagagaag | ccccgacagc  | ttcagcatga | gaagaacctt | 1680 |
| gcctccaatc | gagggcacac | tgggcctgtc | tgctgtcttt  | ctgcctaaca | gcgtgttccc | 1740 |
| cagatgcagc | gagagaaacc | tgcaagagag | cgtgtctgagc | gagcaggatc | tgatgcagcc | 1800 |
| tgaggaaccc | gactacctgc | tggacgagat | gaccaaagag  | gccgagatcg | acaacagcag | 1860 |
| cacagtggaa | agcctgggcg | agacaaccga | catcctgaac  | atgaccacaa | tcatgagcct | 1920 |
| gaacagcaac | gagaagtctc | tgaagctgag | ccccgtgcag  | aagcagaaga | agcagcagac | 1980 |
| catcggcgag | ctgacacagg | atactgccct | gaccgagaac  | gacgacagcg | acgagtacga | 2040 |
| agagatgagc | gagatgaagg | aaggcaaggc | ctgcaagcag  | cacgtgtccc | agggcattct | 2100 |
| tatgaccacg | cctgccacca | ccatcgaggc | cttttccgac  | gaggaagtgg | aaatccccga | 2160 |
| ggaaaaagag | ggcgccgagg | acagcaaagg | caacggcatt  | gaggaacaag | aggtggaagc | 2220 |
| caacgaagag | aacgtgaagg | tgcacggcgg | acggaaagaa  | aagaccgaga | tcctgagcga | 2280 |
| cgacctgacc | gataaggccc | aggtttccga | gggcaaagcc  | aagtctgtgg | gagaagccga | 2340 |
| ggatggacct | gaaggccgcg | gagatggaac | ctgtgaagaa  | ggatctagcg | gagccgagca | 2400 |
| ctggcaggat | gaggaacgcg | agaagggcga | gaaagacaaa  | ggcagaggcg | agatggaaag | 2460 |
| acccggcgag | ggcgaaaaag | agctggccga | gaaagaggaa  | tggaagaaac | gcgacggcga | 2520 |
| agaacaagag | cagaaagaaa | gagagcaggg | ccaccagaaa  | gaacggaatc | aagagatgga | 2580 |
| agaaggcggc | gaggaagaac | acggcgaagg | ggaagaagag  | gaaggcgacc | gagaggaaga | 2640 |
| agaagagaaa | gaaggcgaag | gcaaagaaga | aggcgagggc  | gaagaggtgg | aaggcgagcg | 2700 |
| tgaaaaagaa | gagggcgaac | gcaagaaaga | agaacgcgcc  | ggaaaagagg | aaaaaggcga | 2760 |

|                                                                        |      |
|------------------------------------------------------------------------|------|
| ggaagagggc gaccaaggcg aaggcgagga agaagaaact gaaggcagag gggaagagaa      | 2820 |
| agaggaaggc ggcgaagtcg aaggcggaga ggttgaagaa ggcaaaggcg agcgagaaga      | 2880 |
| ggaagaagaa gaaggcgaag gcgaggaaga ggaaggcgaa ggcgaagagg aagaaggcga      | 2940 |
| aggggaagaa gaagaaggcg aaggcaaggg cgaagaggag ggcgaagaag gcgagggcga      | 3000 |
| agaggagggc gaagaaggcg aaggcgaggg cgaagaagaa gaaggcgaag gcgaaggcga      | 3060 |
| ggaagaaggc gaaggcgaag gggaagaaga ggaaggcgaa ggcgaaggcg aagaagaagg      | 3120 |
| cgaaggcgag ggcgaagagg aagaaggcga agggcaaagg gaagaagaag gcgaggaagg      | 3180 |
| cgaaggcgaa ggcgaggaag aagaaggcga aggcgagggc gaagatggcg aaggcgaagg      | 3240 |
| cgaagaggaa gagggcgagt gggagggcga agaagaggaa ggcgaaggcg agggcgaaga      | 3300 |
| ggaaggcgaa ggcgagggcg aagaaggcga aggcgaaggc gaggaagagg aaggcgaagg      | 3360 |
| cgaaggggaa gaagaagagg gcgaagaaga aggcgaagag gaaggcgaag gggaagaaga      | 3420 |
| aggcgaaggc gaaggcgaag aagaggaaga gggcgaagtt gaaggcgagg ttgagggcga      | 3480 |
| agaaggcgaa ggcgaagggg aagaagaaga aggcgaggaa gaaggggaag agagagaaaa      | 3540 |
| agaaggcgag ggcgaagaaa accgccggaa ccgcgaagag gaagaggaag aagagggcaa      | 3600 |
| gtaccaagag actggcgagg aagagaacga gcggcaggat ggcgaagagt acaagaaggt      | 3660 |
| gtccaagatc aagggcagcg tgaagtacgg caagcacaag acctaccaga agaagtccgt      | 3720 |
| caccaacacg caaggcaatg gaaaagaaca gcggagcaag atgcccgtgc agtccaagag      | 3780 |
| gctgctgaag aatggcccta gcggcagcaa gaaattcttg aacaatgtgc tgccccacta      | 3840 |
| cctcgagctg aagtgagcct cgagcagcgc tgctcgagag atctgcggcc gcgagctcgg      | 3900 |
| ggatccagac atgataagat acattgatga gtttggacaa accacaacta gaatgcagtg      | 3960 |
| aaaaaaaaatgc tttattttgtg aaattttgtga tgctattgct ttattttgtaa ccattataag | 4020 |
| ctgcaataaa caagttaaca acaacaattg cattcatttt atgtttcagg ttcaggggga      | 4080 |
| ggtgtgggag gtttttttaa gcaagtaaaa cctctacaaa tgttggtatgg ctgattatga     | 4140 |
| tcaatgcatc ctagccggag gaacccttag tgatggagtt ggccactccc tctctgcgcg      | 4200 |
| ctcgcctcgt cactgaggcc gcccgggcaa agcccgggcg tcgggcgacc tttggtcgcc      | 4260 |
| cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa                        | 4303 |
| <210> 6                                                                |      |
| <211> 145                                                              |      |
| <212> DNA                                                              |      |
| <213> 腺相关病毒2                                                           |      |
| <400> 6                                                                |      |
| ttggccactc cctctctgcg cgctcgtcgc ctactgagg ccgggcgacc aaaggtcgcc       | 60   |
| cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg       | 120  |
| gccaactcca tcactagggg ttcct                                            | 145  |
| <210> 7                                                                |      |
| <211> 145                                                              |      |
| <212> DNA                                                              |      |
| <213> 腺相关病毒2                                                           |      |
| <400> 7                                                                |      |

|                                                                     |     |
|---------------------------------------------------------------------|-----|
| aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg    | 60  |
| ccgccccgggc aaagccccggg cgtcgggcga cctttggtcg cccggcctca gtgagcgagc | 120 |
| gagcgcgcag agagggagtg gccaa                                         | 145 |
| <210> 8                                                             |     |
| <211> 245                                                           |     |
| <212> DNA                                                           |     |
| <213> 人工序列                                                          |     |
| <220>                                                               |     |
| <223> SV40多聚腺苷酸化序列                                                  |     |
| <400> 8                                                             |     |
| ggggatccag acatgataag atacattgat gagtttgac aaaccacaac tagaatgcag    | 60  |
| tgaaaaaaat gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata   | 120 |
| agctgcaata aacaagttaa caacaacaat tgcattcatt ttatgtttca ggttcagggg   | 180 |
| gaggtgtggg aggtttttta aagcaagtaa aacctctaca aatgtggtat ggctgattat   | 240 |
| gatca                                                               | 245 |
| <210> 9                                                             |     |
| <211> 745                                                           |     |
| <212> PRT                                                           |     |
| <213> 人工序列                                                          |     |
| <220>                                                               |     |
| <223> rAAV变体衣壳蛋白                                                    |     |
| <400> 9                                                             |     |
| Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Thr Leu Ser     |     |
| 1 5 10 15                                                           |     |
| Glu Gly Ile Arg Gln Trp Trp Lys Leu Lys Pro Gly Pro Pro Pro Pro     |     |
| 20 25 30                                                            |     |
| Lys Ala Ala Glu Arg His Lys Asp Asp Ser Arg Gly Leu Val Leu Pro     |     |
| 35 40 45                                                            |     |
| Gly Tyr Lys Tyr Leu Gly Pro Phe Asn Gly Leu Asp Lys Gly Glu Pro     |     |
| 50 55 60                                                            |     |
| Val Asn Glu Ala Asp Ala Ala Leu Glu His Asp Lys Ala Tyr Asp         |     |
| 65 70 75 80                                                         |     |
| Arg Gln Leu Asp Ser Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala     |     |
| 85 90 95                                                            |     |
| Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly     |     |
| 100 105 110                                                         |     |
| Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Val Leu Glu Pro     |     |
| 115 120 125                                                         |     |
| Leu Gly Leu Val Glu Glu Pro Val Lys Thr Ala Pro Gly Lys Lys Arg     |     |

|                     |                                 |                     |
|---------------------|---------------------------------|---------------------|
| 130                 | 135                             | 140                 |
| Pro Val Glu His Ser | Pro Val Glu Pro Asp Ser         | Ser Ser Gly Thr Gly |
| 145                 | 150                             | 155                 |
| Lys Ala Gly Gln Gln | Pro Ala Arg Lys Arg Leu Asn Phe | Gly Gln Thr         |
| 165                 | 170                             | 175                 |
| Gly Asp Ala Asp Ser | Val Pro Asp Pro Gln Pro Leu Gly | Gln Pro Pro         |
| 180                 | 185                             | 190                 |
| Ala Ala Pro Ser Gly | Leu Gly Thr Asn Thr Met Ala Thr | Gly Ser Gly         |
| 195                 | 200                             | 205                 |
| Ala Pro Met Ala Asp | Asn Asn Glu Gly Ala Asp Gly     | Val Gly Asn Ser     |
| 210                 | 215                             | 220                 |
| Ser Gly Asn Trp His | Cys Asp Ser Thr Trp Met Gly     | Asp Arg Val Ile     |
| 225                 | 230                             | 235                 |
| Thr Thr Ser Thr Arg | Thr Trp Ala Leu Pro Thr Tyr     | Asn Asn His Leu     |
| 245                 | 250                             | 255                 |
| Tyr Lys Gln Ile Ser | Ser Gln Ser Gly Ala Ser Asn Asp | Asn His Tyr         |
| 260                 | 265                             | 270                 |
| Phe Gly Tyr Ser Thr | Pro Trp Gly Tyr Phe Asp Phe     | Asn Arg Phe His     |
| 275                 | 280                             | 285                 |
| Cys His Phe Ser Pro | Arg Asp Trp Gln Arg Leu Ile     | Asn Asn Asn Trp     |
| 290                 | 295                             | 300                 |
| Gly Phe Arg Pro Lys | Arg Leu Asn Phe Lys Leu Phe     | Asn Ile Gln Val     |
| 305                 | 310                             | 315                 |
| Lys Glu Val Thr Gln | Asn Asp Gly Thr Thr Thr         | Ile Ala Asn Asn Leu |
| 325                 | 330                             | 335                 |
| Thr Ser Thr Val Gln | Val Phe Thr Asp Ser Glu Tyr     | Gln Leu Pro Tyr     |
| 340                 | 345                             | 350                 |
| Val Leu Gly Ser Ala | His Gln Gly Cys Leu Pro Pro     | Phe Pro Ala Asp     |
| 355                 | 360                             | 365                 |
| Val Phe Met Val Pro | Gln Tyr Gly Tyr Leu Thr         | Leu Asn Asn Gly Ser |
| 370                 | 375                             | 380                 |
| Gln Ala Val Gly Arg | Ser Ser Phe Tyr Cys Leu Glu     | Tyr Phe Pro Ser     |
| 385                 | 390                             | 395                 |
| Gln Met Leu Arg Thr | Gly Asn Asn Phe Thr Phe         | Ser Tyr Thr Phe Glu |
| 405                 | 410                             | 415                 |
| Asp Val Pro Phe His | Ser Ser Tyr Ala His Ser         | Gln Ser Leu Asp Arg |
| 420                 | 425                             | 430                 |
| Leu Met Asn Pro Leu | Ile Asp Gln Tyr Leu Tyr Tyr     | Leu Ser Arg Thr     |
| 435                 | 440                             | 445                 |

|                                                                 |         |         |
|-----------------------------------------------------------------|---------|---------|
| Asn Thr Pro Ser Gly Thr Thr Thr Gln Ser Arg Leu Gln Phe Ser Gln |         |         |
| 450                                                             | 455     | 460     |
| Ala Gly Ala Ser Asp Ile Arg Asp Gln Ser Arg Asn Trp Leu Pro Gly |         |         |
| 465                                                             | 470     | 475 480 |
| Pro Cys Tyr Arg Gln Gln Arg Val Ser Lys Thr Ser Ala Asp Asn Asn |         |         |
|                                                                 | 485 490 | 495     |
| Asn Ser Glu Tyr Ser Trp Thr Gly Ala Thr Lys Tyr His Leu Asn Gly |         |         |
|                                                                 | 500 505 | 510     |
| Arg Asp Ser Leu Val Asn Pro Gly Pro Ala Met Ala Ser His Lys Asp |         |         |
|                                                                 | 515 520 | 525     |
| Asp Glu Glu Lys Phe Phe Pro Gln Ser Gly Val Leu Ile Phe Gly Lys |         |         |
|                                                                 | 530 535 | 540     |
| Gln Gly Ser Glu Lys Thr Asn Val Asp Ile Glu Lys Val Met Ile Thr |         |         |
| 545                                                             | 550     | 555 560 |
| Asp Glu Glu Glu Ile Arg Thr Thr Asn Pro Val Ala Thr Glu Gln Tyr |         |         |
|                                                                 | 565 570 | 575     |
| Gly Ser Val Ser Thr Asn Leu Gln Arg Gly Asn Leu Ala Ile Ser Asp |         |         |
|                                                                 | 580 585 | 590     |
| Gln Thr Lys His Ala Arg Gln Ala Ala Thr Ala Asp Val Asn Thr Gln |         |         |
|                                                                 | 595 600 | 605     |
| Gly Val Leu Pro Gly Met Val Trp Gln Asp Arg Asp Val Tyr Leu Gln |         |         |
| 610                                                             | 615     | 620     |
| Gly Pro Ile Trp Ala Lys Ile Pro His Thr Asp Gly His Phe His Pro |         |         |
| 625                                                             | 630     | 635 640 |
| Ser Pro Leu Met Gly Gly Phe Gly Leu Lys His Pro Pro Pro Gln Ile |         |         |
|                                                                 | 645 650 | 655     |
| Leu Ile Lys Asn Thr Pro Val Pro Ala Asn Pro Ser Thr Thr Phe Ser |         |         |
|                                                                 | 660 665 | 670     |
| Ala Ala Lys Phe Ala Ser Phe Ile Thr Gln Tyr Ser Thr Gly Gln Val |         |         |
|                                                                 | 675 680 | 685     |
| Ser Val Glu Ile Glu Trp Glu Leu Gln Lys Glu Asn Ser Lys Arg Trp |         |         |
| 690                                                             | 695     | 700     |
| Asn Pro Glu Ile Gln Tyr Thr Ser Asn Tyr Asn Lys Ser Val Asn Val |         |         |
| 705                                                             | 710     | 715 720 |
| Asp Phe Thr Val Asp Thr Asn Gly Val Tyr Ser Glu Pro Arg Pro Ile |         |         |
|                                                                 | 725 730 | 735     |
| Gly Thr Arg Tyr Leu Thr Arg Asn Leu                             |         |         |
|                                                                 | 740 745 |         |

&lt;210&gt; 10

&lt;211&gt; 3459

&lt;212&gt; DNA

&lt;213&gt; 人工序列

&lt;220&gt;

&lt;223&gt; 密码子优化的人RPGR

&lt;400&gt; 10

|             |            |             |            |            |            |      |
|-------------|------------|-------------|------------|------------|------------|------|
| atgagagagc  | ctgaagagct | gatgcctgat  | agcggagcag | tgtttacctt | tgggaagagc | 60   |
| aagttcgcag  | agaataaccc | tgggaaattc  | tggtttaaga | acgacgtgcc | cgtgcacctg | 120  |
| agctgtggcg  | atgagcactc | cgccgtgggtg | acaggcaaca | ataagctgta | catgttcggc | 180  |
| tctaacaatt  | ggggacagct | gggcctggga  | agcaagtccg | ccatcagcaa | gccaacctgc | 240  |
| gtgaaggccc  | tgaagcccga | gaaggtgaag  | ctggccgcct | gtggcagaaa | ccacacactg | 300  |
| gtgagcaccg  | agggaggaaa | cgtgtacgca  | acaggaggca | acaatgaagg | ccagctgggc | 360  |
| ctgggcgaca  | cagaggagag | gaataccttt  | cacgtgatca | gcttctttac | ctccgagcac | 420  |
| aagatcaagc  | agctgtccgc | cggctctaac  | acaagcgccg | ccctgaccga | ggacggccgc | 480  |
| ctgttcatgt  | ggggcgataa | tagcgagggc  | cagatcggcc | tgaagaacgt | gtccaacgtg | 540  |
| tgcgtgcctc  | agcaggtgac | catcggaag   | ccagtgtcct | ggatctcttg | tggctactat | 600  |
| cacagcgctt  | tcgtgaccac | agatggcgag  | ctgtacgtgt | ttggagagcc | agagaacggc | 660  |
| aagctgggcc  | tgccaaacca | gctgctgggc  | aatcaccgga | caccccagct | ggtgtccgag | 720  |
| atccctgaga  | aagtgatcca | ggtggcatgc  | ggaggagagc | acacagtggg | gctgaccgag | 780  |
| aatgccgtgt  | ataccttcgg | cctgggacag  | tttgacagc  | tgggcctggg | cacattcctg | 840  |
| tttgagacaa  | gcgagccaaa | agtgatcgag  | aacatccgcg | accagacaat | cagctacatc | 900  |
| tcctgcggcg  | agaatcacac | agccctgata  | accgacatcg | gcctgatgta | tacctttggc | 960  |
| gatggccggc  | acggcaagct | gggcctgggc  | ctggagaact | tcacaaatca | ctttatcccc | 1020 |
| accctgtgct  | ctaacttcct | gcggttcata  | gtgaagctgg | tggcctgcgg | cggctgtcac | 1080 |
| atgggtgggt  | tcgcagcacc | tcacagggga  | gtggccaagg | agatcgagtt | tgacgagatc | 1140 |
| aacgatacat  | gcctgtccgt | ggccaccttc  | ctgccataca | gctccctgac | atccggcaat | 1200 |
| gtgctgcagc  | gcaccctgtc | tgccaggatg  | cggagaaggg | agagggagcg | gtcccctgac | 1260 |
| tctttcagca  | tgaggcggac | actgccacct  | atcgagggca | ccctgggcct | gtctgcctgc | 1320 |
| ttcctgccta  | acagcgtgtt | cccaagatgt  | agcgagagga | atctgcagga | gtctgtgctg | 1380 |
| agcgagcagg  | atctgatgca | gccagaggag  | cccactacc  | tgctggatga | gatgacaaag | 1440 |
| gaggccgaga  | tcgacaactc | tagcaccgtg  | gagagcctgg | gcgagacaac | agatatcctg | 1500 |
| aatatgacac  | acatcatgtc | cctgaactct  | aatgagaagt | ctctgaagct | gagcccagtg | 1560 |
| cagaagcaga  | agaagcagca | gaccatcggc  | gagctgacct | aggacacagc | cctgaccgag | 1620 |
| aacgacgatt  | ctgatgagta | tgaggagatg  | agcgagatga | aggagggcaa | ggcctgtaag | 1680 |
| cagcacgtgt  | cccagggcct | cttcatgacc  | cagccagcca | ccacaatcga | ggccttttct | 1740 |
| gacgaagagg  | tggagatccc | cgaggagaag  | gagggcgccg | aggatagcaa | gggcaatggc | 1800 |
| atcgaggagc  | aggaggtgga | ggccaacgag  | gagaatgtga | aggtgcacgg | cggcagaaag | 1860 |
| gagaagacag  | agatcctgtc | cgacgatctg  | accgacaagg | ccgaggtgtc | cgagggcaag | 1920 |
| gccaaagtctg | tgggagaggc | agaggacgga  | ccagagggac | gcggcgatgg | aacctgcgag | 1980 |

|            |             |            |            |            |            |      |
|------------|-------------|------------|------------|------------|------------|------|
| gagggatcct | ctggagcaga  | gcactggcag | gacgaagaaa | gagagaaggg | cgagaaggat | 2040 |
| aagggcagag | gagagatgga  | gaggcctgga | gagggagaga | aggagctggc | agagaaggag | 2100 |
| gagtggaaga | agagggacgg  | cgaggagcag | gagcagaagg | agagagagca | gggccaccag | 2160 |
| aaggagagga | accaggagat  | ggaggaggga | ggagaggagg | agcacggcga | gggagaggag | 2220 |
| gaggagggcg | atagagagga  | agaagaggag | aaggagggag | agggcaagga | ggaaggcgag | 2280 |
| ggagaggagg | tggagggaga  | aagggagaag | gaggagggag | agcgcaagaa | ggaagaaaga | 2340 |
| gcaggcaagg | aagagaaggg  | agaggaggag | ggcgatcagg | gcgaaggaga | ggaggaggag | 2400 |
| acagagggaa | ggggagagga  | gaaggaggag | ggaggagagg | tcgaaggagg | agaagtggag | 2460 |
| gagggcaagg | gcgaaaagaga | agaggaggag | gaggaaggcg | agggcgaaga | agaggagggc | 2520 |
| gagggcgagg | aagaagaggg  | cgagggcgaa | gaggaagaag | gcgagggcaa | gggcgaggag | 2580 |
| gagggcgaag | aaggcgaagg  | ggaggaggag | ggcgaagagg | gagagggcga | gggcgaggag | 2640 |
| gaagaaggcg | aaggcgaagg  | cgaagaagaa | ggagaaggag | agggcgaaga | ggaggaaggc | 2700 |
| gaaggagaag | gagaggagga  | aggagaaggg | gagggcgaag | aggaggaggg | agaaggcaag | 2760 |
| ggagaagaag | aaggcgaaga  | aggcgaggga | gaaggcgagg | aagaagaagg | cgagggagag | 2820 |
| ggagaggacg | gcgaaggcga  | gggcgaggaa | gaggaaggag | agtgggaggg | cgaggaagag | 2880 |
| gagggagaag | gagaaggcga  | agaagaaggg | gaaggagagg | gcgaggaagg | agaaggcgaa | 2940 |
| ggcgaagagg | aggaggggga  | aggggagggc | gaggaggaag | agggagaaga | ggaaggcgaa | 3000 |
| gaagagggag | aaggcgaaga  | ggaaggagaa | ggcgagggag | aagaagagga | ggagggcgag | 3060 |
| gtcgaaggcg | aggtggaggg  | cgaagagggg | gaaggcgaag | gcgaggagga | ggaaggggaa | 3120 |
| gaagaaggcg | aggagagaga  | gaaagaaggc | gagggcgagg | agaacagaag | gaatcgcgaa | 3180 |
| gaagaagagg | aagaagaggg  | caagtaccag | gagacaggcg | aggaggagaa | cgagcggcag | 3240 |
| gatggcgagg | agtataagaa  | ggtgtccaag | atcaagggct | ctgtgaagta | cggcaagcac | 3300 |
| aagacctatc | agaagaagag  | cgtgaccaac | acacagggca | atggcaagga | gcagcgcagc | 3360 |
| aagatgcctg | tgcagtccaa  | gcggctgctg | aagaatggcc | cctctgggag | caagaagttt | 3420 |
| tggaataatg | tcctgccaca  | ctacctggag | ctgaaatga  |            |            | 3459 |

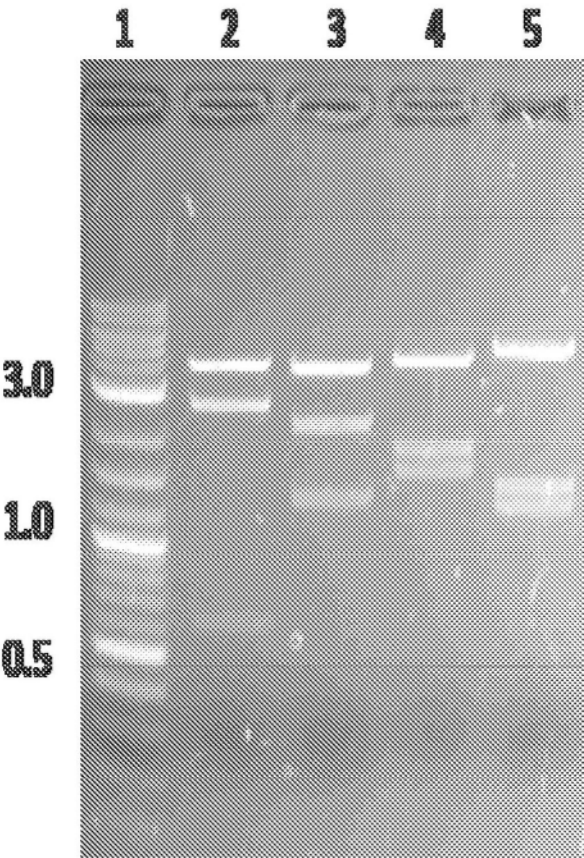


图1

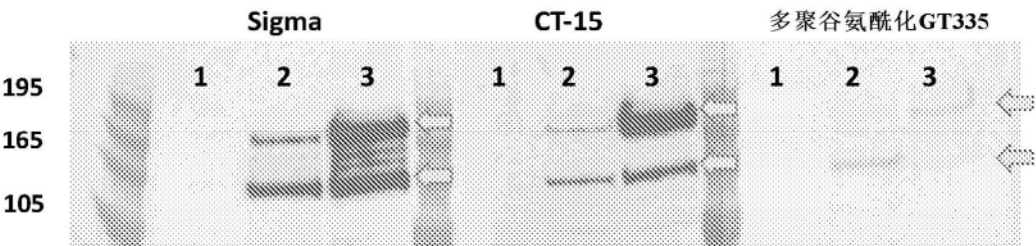


图2

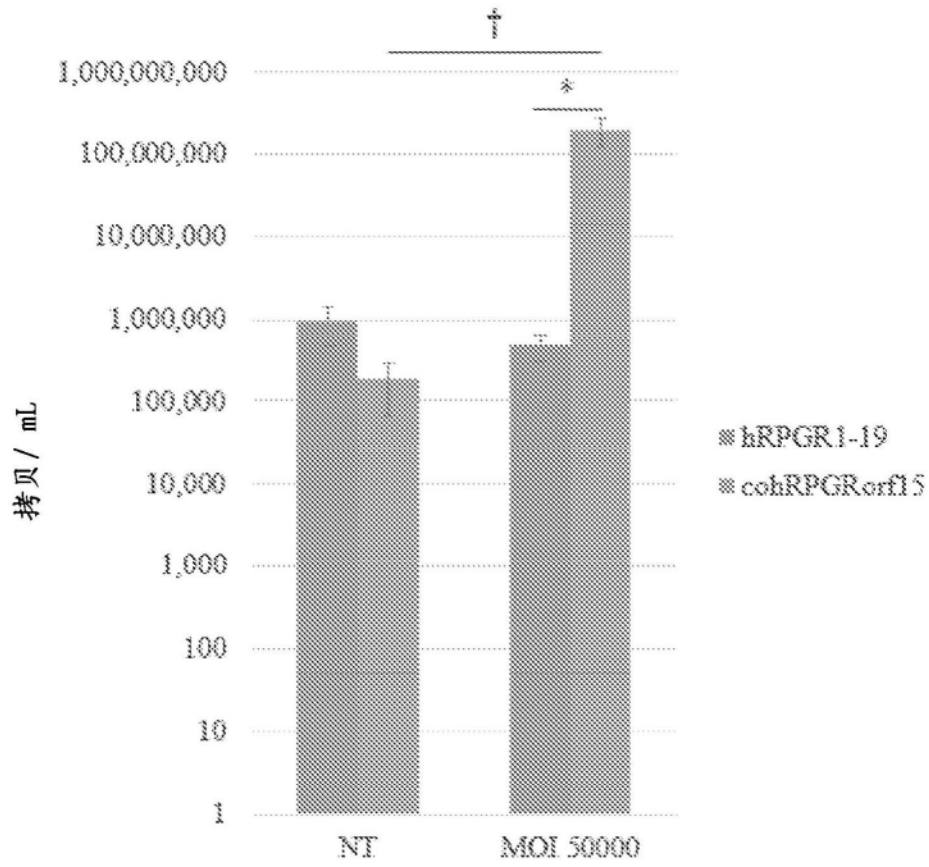


图3

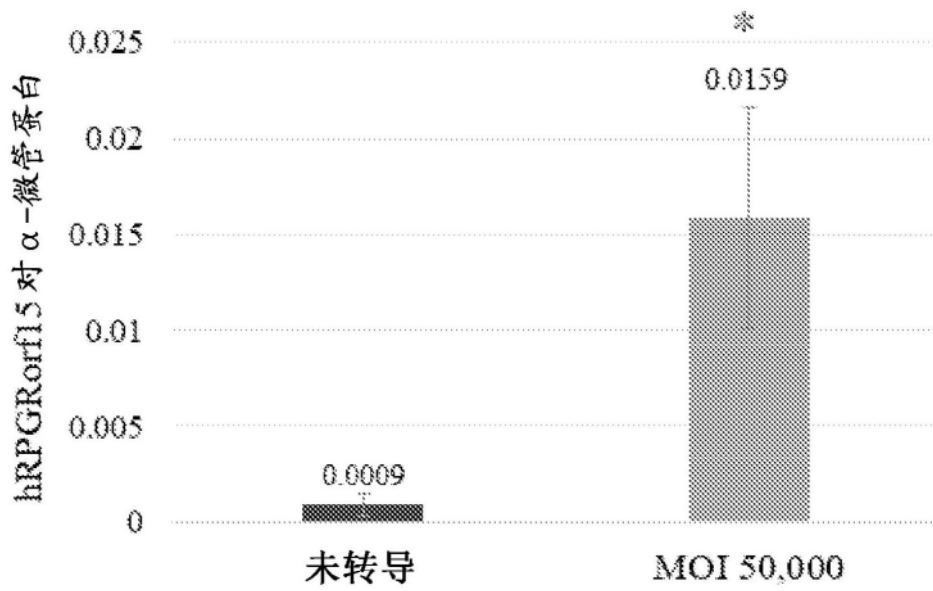


图4

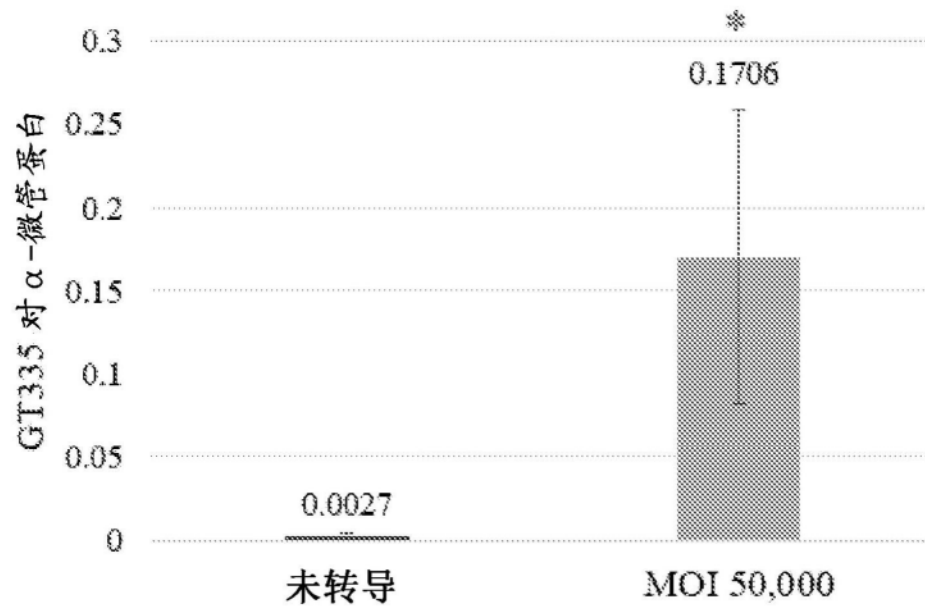


图5

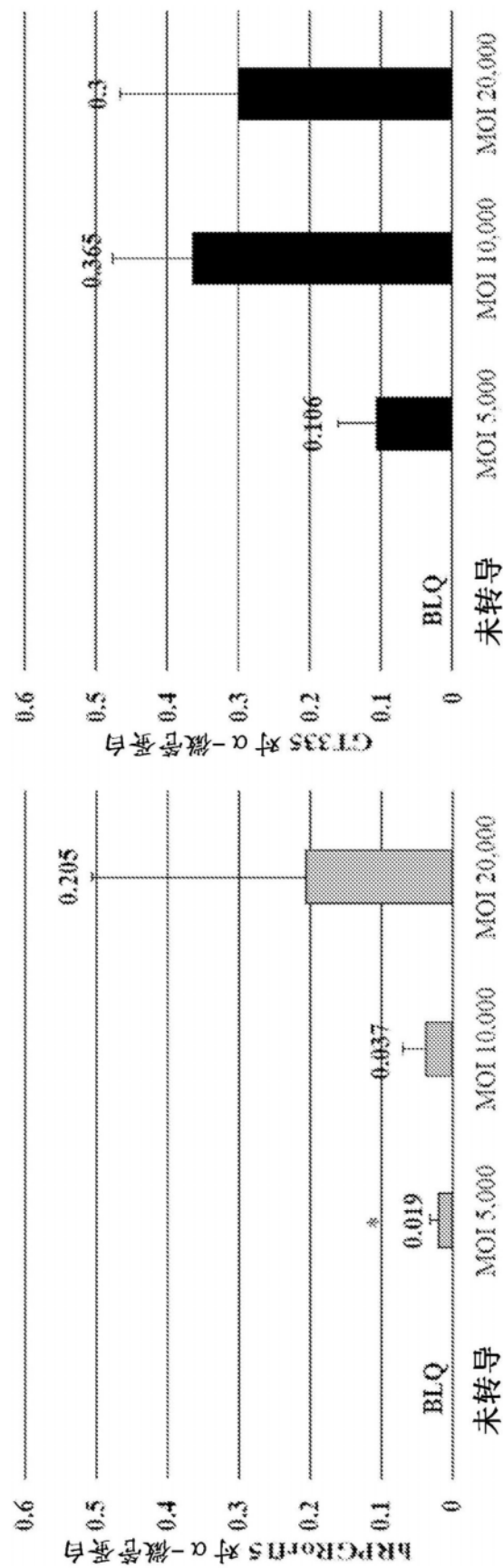


图6

*NarI*  
*KasI*

ATGAGAGAACCCGAGGAACTGATGCCCCACTCTGGCGCCGTGTTTACCTTCGGCAAGAGC  
 1 -----+-----+-----+-----+-----+-----+  
 TACTCTCTTGGGCTCCTTGACTACGGGCTGAGACCGCGGCACAAATGGAAGCCGTTCTCG  
 M R E P E E L M P D S G A V F T F G K S  
 1 3 5 7 9 11 13 15 17 19

AAGTTCGCCGAGAAACAACCCCGCAAGTTCTGGTTCAAGAACGACGTGCCAGTGCACCTG 61 --  
 -----+-----+-----+-----+-----+-----+  
 TTCAAGCGGCTCTTGTGGGGCCGTTCAAGACCAAGTTCTTGCTGCACGGTCACGTGGAC  
 K F A E N N P G K F W F K N D V P V H L  
 21 23 25 27 29 31 33 35 37 39

*BstEII*

AGCTGCGGAGATGAACACTCTGCCGTGGTCACCGGCAACAACAAGCTGTACATGTTCCGGC  
 121 -----+-----+-----+-----+-----+-----+  
 TCGACGCCTCTACTTGTGAGACGGCACCAGTGGCCGTTGTTGTTTCGACATGTACAAGCCG  
 S C G D E H S A V V T G N N K L Y M F G  
 41 43 45 47 49 51 53 55 57 59

*BspMI*

AGCAACAACCTGGGGCCAGCTCGGCCTGGGATCTAAGTCTGCCATCAGCAAGCCTACCTGC  
 181 -----+-----+-----+-----+-----+-----+  
 TCGTTGTTGACCCCGGTGAGCCGGACCCTAGATTGAGACGGTAGTCGTTCCGATGGACG  
 S N N W G Q L G L G S K S A I S K P T C  
 61 63 65 67 69 71 73 75 77 79

GTGAAGGCCCTGAAGCCTGAGAAAGTGAAACTGGCCGCTGCGGCAGAAATCACACCCTG 241 -  
 -----+-----+-----+-----+-----+-----+  
 CACTTCCGGGACTTCGGACTCTTTCACTTTGACCGCGGACGCCGCTTTTGTGTTGGGAC  
 V K A L K P E K V K L A A C G R N H T L  
 81 83 85 87 89 91 93 95 97 99

GTTTCTACCGAAGGCGGCAATGTGTATGCCACCGCGGAAACAATGAGGGACAGCTTGA 301 -  
 -----+-----+-----+-----+-----+-----+  
 CAAAGATGGCTTCCGCCGTTACACATACGGTGGCCGCTTTGTTACTCCCTGTGCAACCT  
 V S T E G G N V Y A T G G N N E G Q L G  
 101 103 105 107 109 111 113 115 117 119

*BclI*

CTGGGCGACACCGAGGAAAGAAACACCTTCCACGTGATCAGCTTTTTTACCAGCGAGCAC  
 361 -----+-----+-----+-----+-----+-----+  
 GACCCGCTGTGGCTCCTTTCTTTGTGGAAGGTGCACTAGTCGAAAAAGTGGTCGCTCGTG  
 L G D T E E R N T F H V I S F F T S E H  
 121 123 125 127 129 131 133 135 137 139

*PvuII*

AAGATCAAGCAGCTGAGCGCCGGCTCTAATACCTCTGCCGCTCTGACAGAGGACGGCAGA  
 421 -----+-----+-----+-----+-----+-----+  
 TTCTAGTTTCGTCGACTCGCGGCCGAGATTATGGAGACGGCGAGACTGTCTCCTGCCGTCT  
 K I K Q L S A G S N T S A A L T E D G R  
 141 143 145 147 149 151 153 155 157 159

CTGTTTATGTGGGGCGACAATTCTGAGGGCCAGATCGGACTGAAGAACGTGTCCAATGTG 481 -  
 -----+-----+-----+-----+-----+-----+  
 GACAAATACACCCGCTGTTAAGACTCCCGGTCTAGCCTGACTTCTGCACAGGTTACAC  
 L F M W G D N S E G Q I G L K N V S N V  
 161 163 165 167 169 171 173 175 177 179

图7

*PvuII*

TGCGTGCCCCAGCAAGTGACAATCGGCAAGCCTGTGTCTTGGATCAGCTGCGGCTACTAC  
 541 -----+-----+-----+-----+-----+-----+-----+  
 ACGCACGGGGTCGTTCACTGTTAGCCGTTCCGACACAGAACCTAGTCGACGCCGATGATG  
 C V P Q Q V T I G K P V S W I S C G Y Y  
 181 183 185 187 189 191 193 195 197 199

*PflMI*

CACAGCGCCTTTGTGACAACCGATGGCGAGCTGTATGTGTTCCGCGAGCCAGAGAATGGC  
 601 -----+-----+-----+-----+-----+-----+-----+  
 GTGTCGCGGAAACACTGTTGGCTACCGCTCGACATACACAAGCCGCTCGGTCTCTTACCG  
 H S A F V T T D G E L Y V F G E P E N G  
 201 203 205 207 209 211 213 215 217 219

*PvuII*  
*PflMI* *PvuII*

AAGCTGGGACTGCCTAACCAGCTGCTGGGCAATCACAGAACCCCTCAGCTGGTGTCTGAG  
 661 -----+-----+-----+-----+-----+-----+-----+  
 TTCGACCCTGACGGATTGGTCGACGACCCGTTAGTGTCTTGGGGAGTCGACCACAGACTC  
 K L G L P N Q L L G N H R T P Q L V S E  
 221 223 225 227 229 231 233 235 237 239

ATCCCCGAAAAAGTGATCCAGGTGGCCTGTGGCGAGAGCACACAGTGGTGTGACAGAG 721 -  
 -----+-----+-----+-----+-----+-----+-----+  
 TAGGGGCTTTTCACTAGGTCCACCGACACCGCCTCTCGTGTGTACACGACTGTCTC  
 I P E K V I Q V A C G G E H T V V L T E  
 241 243 245 247 249 251 253 255 257 259

AATGCCGTGTACACCTTTGGCCTGGGCCAGTTTGGACAACCTCGGACTGGGAACCTTCCTG 781 -  
 -----+-----+-----+-----+-----+-----+-----+  
 TTACGGCACATGTGGAACCGGACCCGGTCAAACCTGTTGAGCCTGACCCTTGAAGGAC  
 N A V Y T F G L G Q F G Q L G L G T F L  
 261 263 265 267 269 271 273 275 277 279

*PvuII*

TTCGAGACAAGCGAGCCCCAAAGTGATCGAGAACATCCGGGACCAGACCATCAGCTACATC  
 841 -----+-----+-----+-----+-----+-----+-----+  
 AAGCTCTGTTTCGCTCGGGTTTCACTAGCTCTTGTAGGCCCTGGTCTGGTAGTCGATGTAG  
 F E T S E P K V I E N I R D Q T I S Y I  
 281 283 285 287 289 291 293 295 297 299

*BclI*

AGCTGTGGCGAGAACCACACAGCCCTGATCACAGACATCGGCCTGATGTACACATTCCGGC  
 901 -----+-----+-----+-----+-----+-----+-----+  
 TCGACACCGCTCTTGGTGTGTCGGGACTAGTGTCTGTAGCCGGACTACATGTGTAAGCCG  
 S C G E N H T A L I T D I G L M Y T F G  
 301 303 305 307 309 311 313 315 317 319

GACGGAAGGCATGGAAAGCTCGGACTTGGCCTGGAAAACCTCACCAACCACTTCATCCCT 961  
 -----+-----+-----+-----+-----+-----+-----+  
 CTGCCTTCCGTACCTTTTCGAGCCTGAACCGGACCTTTTGAAGTGGTTGGTGAAGTAGGGA  
 D G R H G K L G L G L E N F T N H F I P  
 321 323 325 327 329 331 333 335 337 339

*PflMI*

ACGCTGTGCAGCAACTTCCTGCGGTTTATTGTGAAGCTGGTGGCCTGCGGAGGATGCCAC  
 1021 -----+-----+-----+-----+-----+-----+-----+  
 TGCGACACGTCGTTGAAGGACGCCAAGTAACACTTCGACCACCGGACGCCTCCTACGGTG  
 T L C S N F L R F I V K L V A C G G C H  
 341 343 345 347 349 351 353 355 357 359

图7(续)

ATGGTGTTTGTGCTGCCCCCTCACAGAGGCGTGGCCAAAGAGATTGAGTTCGACGAGATC  
1081 -----+-----+-----+-----+-----+-----+  
TACCACCAAAAACGACGGGGAGTGTCTCCGCACCGGTTTCTTAAGCTCAAGCTGCTCTAG  
M V V F A A P H R G V A K E I E F D E I  
361 363 365 367 369 371 373 375 377 379

*BspMI*  
AACGATACCTGCCTGAGCGTGGCCACCTTCCTGCCTTACAGCAGCCTGACATCTGGCAAC  
1141 -----+-----+-----+-----+-----+-----+  
TTGCTATGGACCGACTCGCACCGGTGGAAGGACGGAATGTCGTCGGACTGTAGACCGTTG  
N D T C L S V A T F L P Y S S L T S G N  
381 383 385 387 389 391 393 395 397 399

*PstI*  
GTGCTGCAGAGGACACTGAGCGCCAGAATGCGCAGACGGGAAAGAGAGAGAAGCCCCGAC  
1201 -----+-----+-----+-----+-----+-----+  
CACGACGTCTCCTGTGACTCGCGGTCTTACGCGTCTGCCCTTTCTCTCTCTCGGGGCTG  
V L Q R T L S A R M R R R E R E R S P D  
401 403 405 407 409 411 413 415 417 419

AGCTTCAGCATGAGAAGAACCCTGCCTCCAATCGAGGGCACACTGGGCCTGTCTGCCTGC  
1261 -----+-----+-----+-----+-----+-----+  
TCGAAGTCGTACTCTTCTTGGGACGGAGGTAGCTCCCGTGTGACCCGGACAGACGGACG  
S F S M R R T L P P I E G T L G L S A C  
421 423 425 427 429 431 433 435 437 439

*BspMI*  
TTTCTGCCTAACAGCGTGTTCCCCAGATGCAGCGAGAGAAACCTGCAAGAGAGCGTGCTG  
1321 -----+-----+-----+-----+-----+-----+  
AAAGACGGATTGTGCGACAAGGGGTCTACGTGCGTCTCTTTGGACGTTCTCTCGCACGAC  
F L P N S V F P R C S E R N L Q E S V L  
441 443 445 447 449 451 453 455 457 459

*BspMI*  
AGCGAGCAGGATCTGATGCAGCCTGAGGAACCCGACTACCTGCTGGACGAGATGACCAAA  
1381 -----+-----+-----+-----+-----+-----+  
TCGCTCGTCCTAGACTACGTGCGACTCCTTGGGCTGATGGACGACCTGCTCTACTGGTTT  
S E Q D L M Q P E E P D Y L L D E M T K  
461 463 465 467 469 471 473 475 477 479

GAGGCCGAGATCGACAACAGCAGCACAGTGGAAAGCCTGGGCGAGACAACCGACATCCTG  
1441 -----+-----+-----+-----+-----+-----+  
CTCCGGCTCTAGCTGTTGTCGTCGTGTACCTTTCCGACCCGCTCTGTTGGCTGTAGGAC  
E A E I D N S S T V E S L G E T T D I L  
481 483 485 487 489 491 493 495 497 499

AACATGACCCACATCATGAGCCTGAACAGCAACGAGAAGTCTCTGAAGCTGAGCCCCGTG  
1501 -----+-----+-----+-----+-----+-----+  
TTGTACTGGGTGTAGTACTCGGACTTGTGCTTGCTCTTCAGAGACTTCGACTCGGGGCAC  
N M T H I M S L N S N E K S L K L S P V  
501 503 505 507 509 511 513 515 517 519

CAGAAGCAGAAGAAGCAGCAGACCATCGGCGAGCTGACACAGGATACTGCCCTGACCGAG  
1561 -----+-----+-----+-----+-----+-----+  
GTCTTCGTCTTCTTCGTGCTGCTGGTAGCCGCTCGACTGTGTCCTATGACGGGACTGGCTC  
Q K Q K K Q Q T I G E L T Q D T A L T E  
521 523 525 527 529 531 533 535 537 539

图7(续)

*StuI*

1621 -----+-----+-----+-----+-----+-----+-----+  
 AACGACGACAGCGACGAGTACGAAGAGATGAGCGAGATGAAGGAAGGCAAGGCCTGCAAG  
 TTGCTGCTGTCGCTGCTCATGCTTCTCTACTCGCTCTACTTCCTTCCGTTCCGGACGTTT  
 N D D S D E Y E E M S E M K E G K A C K  
 541 543 545 547 549 551 553 555 557 559

*StuI*

1681 -----+-----+-----+-----+-----+-----+-----+  
 CAGCACGTGTCCCAGGGCATCTTTATGACCCAGCCTGCCACCACCATCGAGGCCTTTTCC  
 GTCGTGCACAGGGTCCCGTAGAAATACTGGGTCCGACGGTGGTGGTAGCTCCGAAAAGG  
 Q H V S Q G I F M T Q P A T T I E A F S  
 561 563 565 567 569 571 573 575 577 579

*NarI*  
*KasI*

1741 -----+-----+-----+-----+-----+-----+-----+  
 GACGAGGAAGTGGAAATCCCCGAGGAAAAAGAGGGCGCCGAGGACAGCAAAGGCAACGGC  
 CTGCTCCTTCACCTTTAGGGGCTCCTTTTCTCCCGCGGCTCCTGTCTGTTCCGTTGCCG  
 D E E V E I P E E K E G A E D S K G N G  
 581 583 585 587 589 591 593 595 597 599

1801 -----+-----+-----+-----+-----+-----+-----+  
 ATTGAGGAACAAGAGGTGGAAGCCAACGAAGAGAACGTGAAGGTGCACGGCGGACGGAAA  
 TAACTCCTTGTCTCCACCTTCGGTTGCTTCTTGCACCTCCACGTGCCGCTGCCTTT  
 I E E Q E V E A N E E N V K V H G G R K  
 601 603 605 607 609 611 613 615 617 619

1861 -----+-----+-----+-----+-----+-----+-----+  
 GAAAAGACCGAGATCCTGAGCGACGACCTGACCGATAAGGCCGAGGTTTCCGAGGGCAAA  
 CTTTTCTGGCTCTAGGACTCGCTGCTGGACTGGCTATTCCGGCTCAAAGGCTCCCGTTT  
 E K T E I L S D D L T D K A E V S E G K  
 621 623 625 627 629 631 633 635 637 639

*SacII*

1921 -----+-----+-----+-----+-----+-----+-----+  
 GCCAAGTCTGTGGGAGAAGCCGAGGATGGACCTGAAGGCCGCGGAGATGGAACCTGTGAA  
 CGGTTTCAGACACCTCTTCGGCTCCTACCTGGACTTCCGGCGCCTCTACCTTGGACACTT  
 A K S V G E A E D G P E G R G D G T C E  
 641 643 645 647 649 651 653 655 657 659

1981 -----+-----+-----+-----+-----+-----+-----+  
 GAAGGATCTAGCGGAGCCGAGCACTGGCAGGATGAGGAACGCGAGAAGGGCGAGAAAGAC  
 CTTCTAGATCGCCTCGGCTCGTGACCGTCTACTCCTTGCCTCTTCCGCTCTTTCTG  
 E G S S G A E H W Q D E E R E K G E K D  
 661 663 665 667 669 671 673 675 677 679

2041 -----+-----+-----+-----+-----+-----+-----+  
 AAAGGCAGAGGCGAGATGGAAGACCCGGCGAGGGCGAAAAAGAGCTGGCCGAGAAAGAG  
 TTTCCGCTCCGCTCTACCTTTCTGGGCGGCTCCCGCTTTTCTCGACCGGCTCTTTCTC  
 K G R G E M E R P G E G E K E L A E K E  
 681 683 685 687 689 691 693 695 697 699

2101 -----+-----+-----+-----+-----+-----+-----+  
 GAATGGAAGAAACGCGACGGCGAAGAACAAGAGCAGAAAGAAAGAGAGCAGGGCCACCAG  
 CTTACCTTCTTTCGCTGCCGCTTCTTGTCTCGTCTTCTTCTCTCGTCCCGGTGGTC  
 E W K K R D G E E Q E Q K E R E Q G H Q  
 701 703 705 707 709 711 713 715 717 719

图7(续)

AAAGAACGGAATCAAGAGATGGAAGAAGGCGGCGAGGAAGAACACGGCGAAGGGGAAGAA  
2161 -----+-----+-----+-----+-----+-----+  
TTTCTTGCCTTAGTTCTCTACCTTCTTCCGCCGCTCCTTCTGTGCCGCTCCCTTCTT  
K E R N Q E M E E G G E E E H G E G E E  
721 723 725 727 729 731 733 735 737 739

GAGGAAGGCGACCGAGAGGAAGAAGAAGAGAAAGAAGGCGAAGGCAAAGAAGAAGGCGAG  
2221 -----+-----+-----+-----+-----+-----+  
CTCCTTCCGCTGGCTCTCCTTCTTCTTCTTCTTCTTCCGCTTCCGTTTCTTCTTCCGCTC  
E E G D R E E E E E K E G E G K E E G E  
741 743 745 747 749 751 753 755 757 759

GGCGAAGAGGTGGAAGGCGAGCGTGAAAAAGAAGAGGGCGAACGCAAGAAAGAAGAACGC  
2281 -----+-----+-----+-----+-----+-----+  
CCGCTTCTCCACCTTCCGCTCGCACTTTTCTTCTCCGCTTGCCTTCTTCTTCTTCTTGC  
G E E V E G E R E K E E G E R K K E E R  
761 763 765 767 769 771 773 775 777 779

GCCGAAAAGAGGAAAAGGCGAGGAAGAGGGCGACCAAGGCGAAGGCGAGGAAGAAGAA  
2341 -----+-----+-----+-----+-----+-----+  
CGGCCTTTTCTCCTTTTCCGCTCCTTCTCCGCTGGTTCGCTTCCGCTCCTTCTTCTT  
A G K E E K G E E E G D Q G E G E E E E  
781 783 785 787 789 791 793 795 797 799

ACTGAAGGCAGAGGGGAAGAGAAAGAGGAAGGCGGCGAAGTCGAAGGCGGAGAGGTTGAA  
2401 -----+-----+-----+-----+-----+-----+  
TGACTTCCGCTCTCCCTTCTCTTTCTCCTTCCGCCGCTTCCGCTTCCGCTCTCCAATT  
T E G R G E E K E E G G E V E G G E V E  
801 803 805 807 809 811 813 815 817 819

GAAGGCAAAGGCGAGCGAGAAGAGGAAGAAGAAGGCGAAGGCGAGGAAGAGGAAGGC  
2461 -----+-----+-----+-----+-----+-----+  
CTTCCGTTTCCGCTCGCTCTTCTCCTTCTTCTTCTTCCGCTTCCGCTCCTTCTCCTTCCG  
E G K G E R E E E E E E G E G E E E E G  
821 823 825 827 829 831 833 835 837 839

GAAGGCGAAGAGGAAGAAGGCGAAGGGGAAGAAGAAGAAGGCGAAGGCAAAGGGCGAAGAG  
2521 -----+-----+-----+-----+-----+-----+  
CTTCCGCTTCTCCTTCTTCCGCTTCCCTTCTTCTTCTTCCGCTTCCGTTCCGCTTCTC  
E G E E E E G E G E E E E G E G K G E E  
841 843 845 847 849 851 853 855 857 859

GAGGGCGAAGAAGGCGAGGGCGAAGAGGGCGAAGAAGGCGAAGGCGAGGGCGAAGAA  
2581 -----+-----+-----+-----+-----+-----+  
CTCCCGCTTCTTCCGCTCCGCTTCTCCTCCGCTTCTTCCGCTTCCGCTCCGCTTCTT  
E G E E G E G E E E G E E G E G E G E E  
861 863 865 867 869 871 873 875 877 879

GAAGAAGGCGAAGGCGAAGGCGAGGAAGAAGGCGAAGGCGAAGGGGAAGAAGGAAGGC  
2641 -----+-----+-----+-----+-----+-----+  
CTTCTTCCGCTTCCGCTTCCGCTCCTTCTTCCGCTTCCGCTTCCCTTCTTCTCCTTCCG  
E E G E G E G E E E G E G E G E E E E G  
881 883 885 887 889 891 893 895 897 899

GAAGGCGAAGGCGAAGAAGAAGGCGAAGGCGAGGGCGAAGAGGAAGAAGGCGAAGGCAAA  
2701 -----+-----+-----+-----+-----+-----+  
CTTCCGCTTCCGCTTCTTCTTCCGCTTCCGCTTCCGCTTCTCCTTCTTCCGCTTCCGTTT  
E G E G E E G E G E G E E E E E G K  
901 903 905 907 909 911 913 915 917 919

图7(续)

```
GGGGAAGAAGAAGGCGAGGAAGGCGAAGGCGAAGGCGAGGAAGAAGAAGGCGAAGGCGAG
2761 -----+-----+-----+-----+-----+-----+
      CCCCTTCTTCTCCGCTCCTTCCGCTTCCGCTTCCGCTCCTTCTTCTCCGCTTCCGCTC
      G E E E G E E G E G E G E E E E G E G E
      921  923  925  927  929  931  933  935  937  939

GGCGAAGATGGCGAAGGCGAAGGCGAAGAGAGGGCGAGTGGGAGGGCGAAGAAGAG
2821 -----+-----+-----+-----+-----+
      CCGCTTCTACCGCTTCCGCTTCCGCTTCTCCTTCTCCCGCTCACCTCCCGCTTCTTCTC
      G E D G E G E G E E E E G E W E G E E E
      941  943  945  947  949  951  953  955  957  959

GAAGGCGAAGGCGAGGGCGAAGAGGAAGGCGAAGGCGAGGGCGAAGAAGGCGAAGGCGAA
2881 -----+-----+-----+-----+-----+
      CTTCCGCTTCCGCTCCCGCTTCTCCTTCCGCTTCCGCTTCCGCTTCTTCCGCTTCCGCTT
      E G E G E G E E E G E G E G E E G E G E
      961  963  965  967  969  971  973  975  977  979

GGCGAGGAAGAGGAAGGCGAAGGCGAAGGGGAAGAAGAAGAGGGCGAAGAAGAAGGCGAA
2941 -----+-----+-----+-----+-----+
      CCGCTCCTTCTCCTTCCGCTTCCGCTTCCCTTCTTCTTCTCCGCTTCTTCTTCCGCTT
      G E E E E G E G E G E E E E G E E E G E
      981  983  985  987  989  991  993  995  997  999

GAGGAAGGCGAAGGGGAAGAAGAAGGCGAAGGCGAAGGCGAAGAAGAGGAAGAGGGCGAA
3001 -----+-----+-----+-----+-----+
      CTCCTTCCGCTTCCCCTTCTTCTTCCGCTTCCGCTTCCGCTTCTTCTCCTTCTCCGCTT
      E E G E G E E E G E G E G E E E E E G E
      1001 1003 1005 1007 1009 1011 1013 1015 1017 1019

GTTGAAGGCGAGGTTGAGGGCGAAGAAGGCGAAGGCGAAGGGGAAGAAGAAGAAGGCGAG
3061 -----+-----+-----+-----+-----+
      CAACTTCCGCTCCAACCTCCGCTTCTTCCGCTTCCGCTTCCCCTTCTTCTTCTTCCGCTC
      V E G E V E G E E G E G E G E E E E G E
      1021 1023 1025 1027 1029 1031 1033 1035 1037 1039

GAAGAAGGGGAAGAGAGAGAAAAAGAAGGCGAGGGCGAAGAAAACCGCCGGAACCGCGAA
3121 -----+-----+-----+-----+-----+
      CTTCTTCCCCTTCTCTCTCTTTTCTTCCGCTCCCGCTTCTTTTGGCGGCCTTGGCGCTT
      E E G E E R E K E G E G E E N R R N R E
      1041 1043 1045 1047 1049 1051 1053 1055 1057 1059

GAGGAAGAGGAAGAAGAGGGCAAGTACCAAGAGACTGGCGAGGAAGAGAACGAGCGGCAG
3181 -----+-----+-----+-----+-----+
      CTCCTTCTCCTTCTTCTCCGTTTCATGGTTCTCTGACCGCTCCTTCTTCTGCTCGCCGTC
      E E E E E E G K Y Q E T G E E E N E R Q
      1061 1063 1065 1067 1069 1071 1073 1075 1077 1079

GATGGCGAAGAGTACAAGAAGGTGTCCAAGATCAAGGGCAGCGTGAAGTACGGCAAGCAC
3241 -----+-----+-----+-----+-----+
      CTACCGCTTCTCATGTTCTTCCACAGTTCTAGTTCCCGTCGCACTTCATGCCGTTCTGTG
      D G E E Y K K V S K I K G S V K Y G K H
      1081 1083 1085 1087 1089 1091 1093 1095 1097 1099

AAGACCTACCAGAAGAAGTCCGTCACCAACACGCAAGGCAATGGAAAAGAACAGCGGAGC
3301 -----+-----+-----+-----+-----+
      TTCTGGATGGTCTTCTTCCAGGAGTGGTTGTGCGTTCCGTTACCTTTTCTTGTGCGCTCG
      K T Y Q K K S V T N T Q G N G K E Q R S
      1101 1103 1105 1107 1109 1111 1113 1115 1117 1119
```

图7(续)

```

      AAGATGCCCCGTGCAGTCCAAGAGGCTGCTGAAGAATGGCCCTAGCGGCAGCAAGAAATTC
3361 -----+-----+-----+-----+-----+-----+-----+
      TTCTACGGGCACGTCAGGTTCTCCGACGACTTCTTACCGGGATCGCCGTCGTTCTTTAAG
      K M P V Q S K R L L K N G P S G S K K F
      1121 1123 1125 1127 1129 1131 1133 1135 1137 1139

                        XhoI
      TGGAACAATGTGCTGCCCCACTACCTCGAGCTGAAGTGA
3421 -----+-----+-----+-----+-----+-----+
      ACCTTGTTACACGACGGGGTGATGGAGCTCGACTTCACT
      W N N V L P H Y L E L K *
      1141 1143 1145 1147 1149 1151 1153

```

图7(结束)

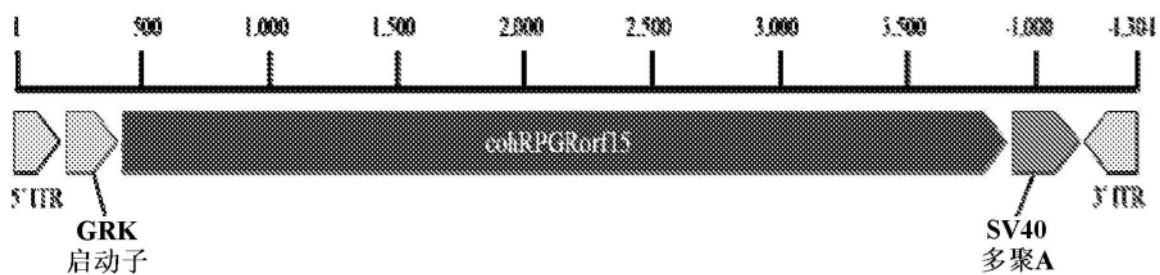


图8

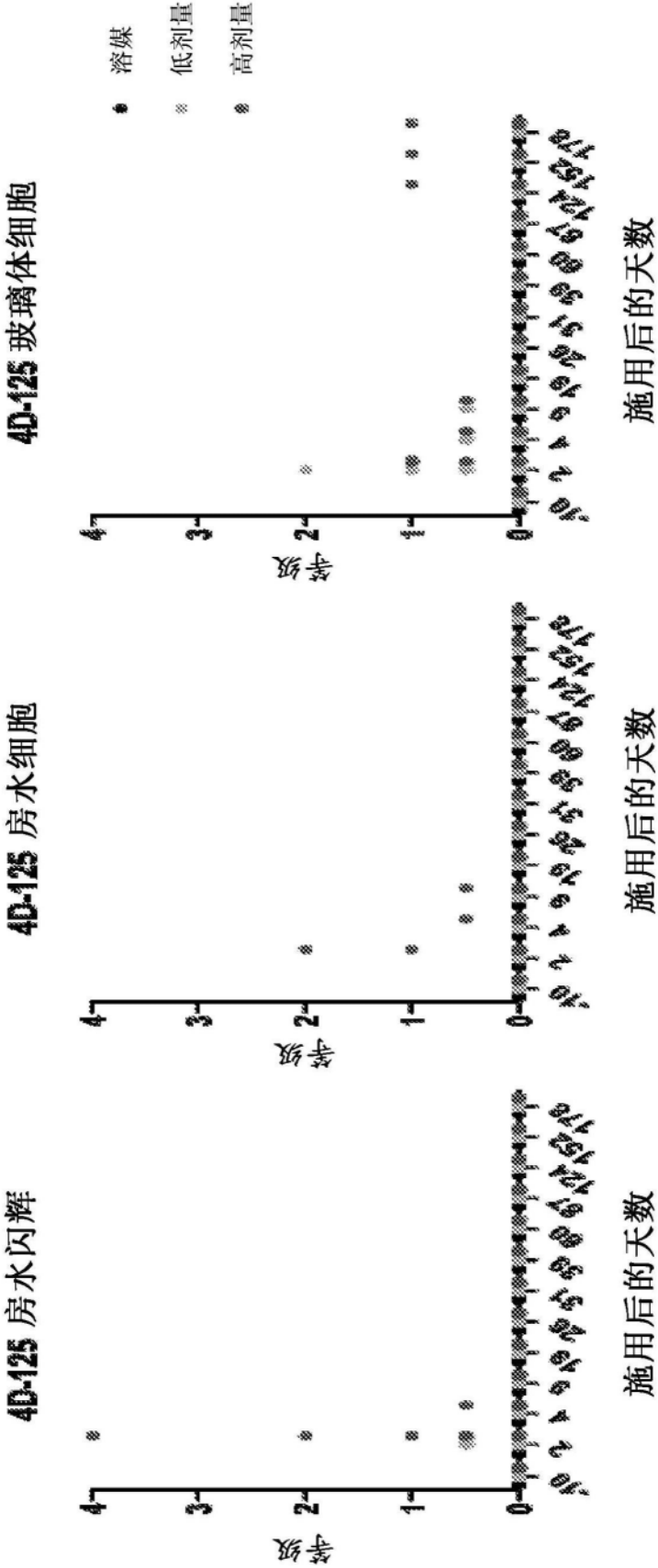


图9

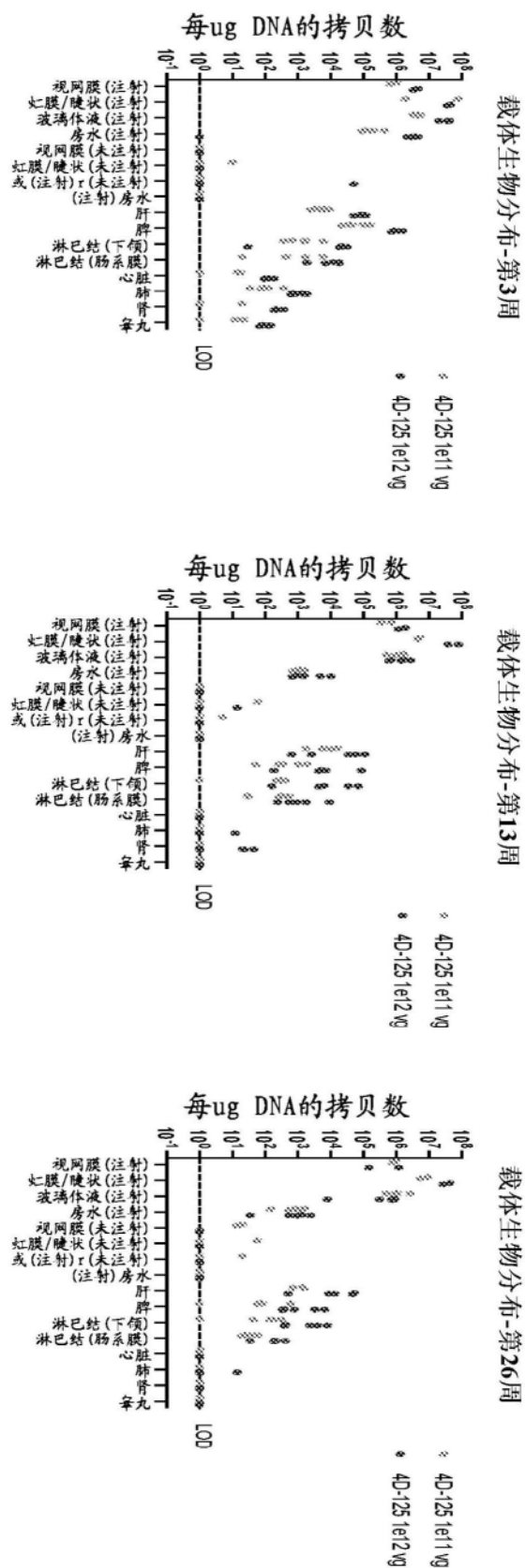


图10

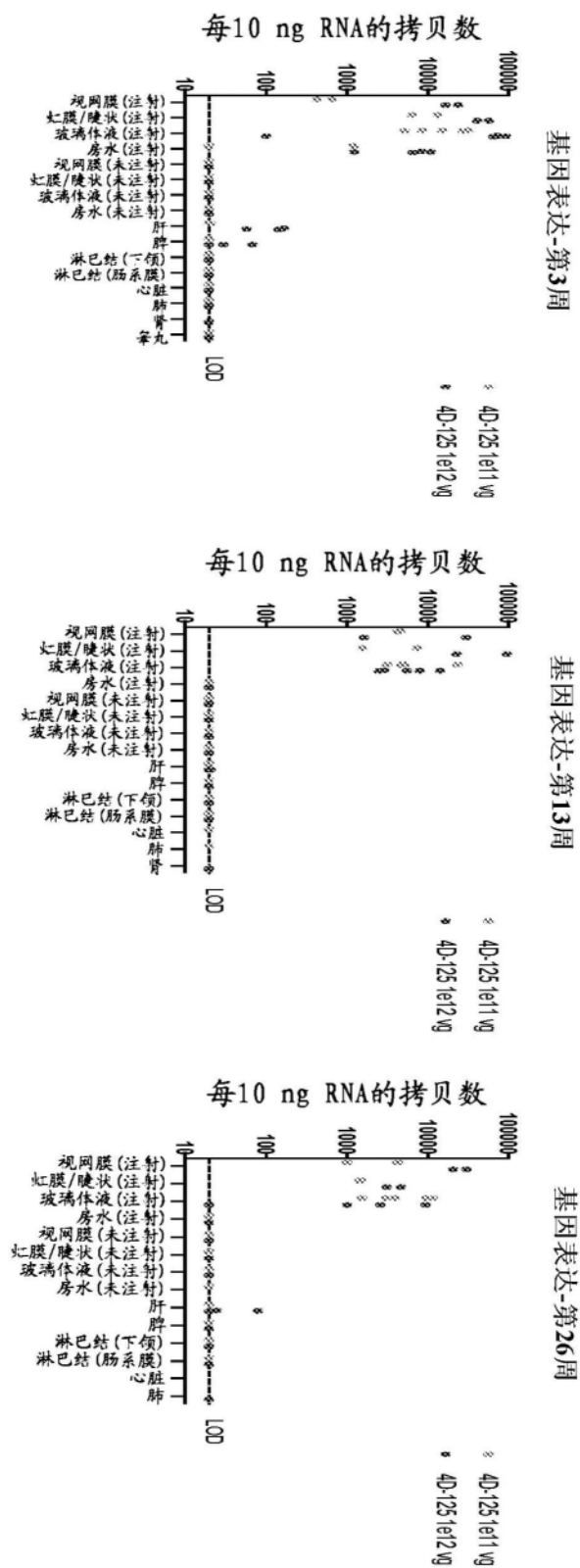


图11