



(12) 发明专利申请

(10) 申请公布号 CN 113853207 A

(43) 申请公布日 2021.12.28

(21) 申请号 202080032666.9

(74) 专利代理机构 中国专利代理(香港)有限公司

(22) 申请日 2020.04.28

72001

(30) 优先权数据

代理人 黄登高 初明明

62/840184 2019.04.29 US

(51) Int.Cl.

62/913314 2019.10.10 US

A61K 35/761 (2006.01)

62/924095 2019.10.21 US

(85) PCT国际申请进入国家阶段日

2021.10.29

(86) PCT国际申请的申请数据

PCT/US2020/030273 2020.04.28

(87) PCT国际申请的公布数据

W02020/223232 EN 2020.11.05

(71) 申请人 宾夕法尼亚州大学信托人

地址 美国宾夕法尼亚州

权利要求书3页 说明书36页

(72) 发明人 K·纳姆比亚 J·M·威尔逊

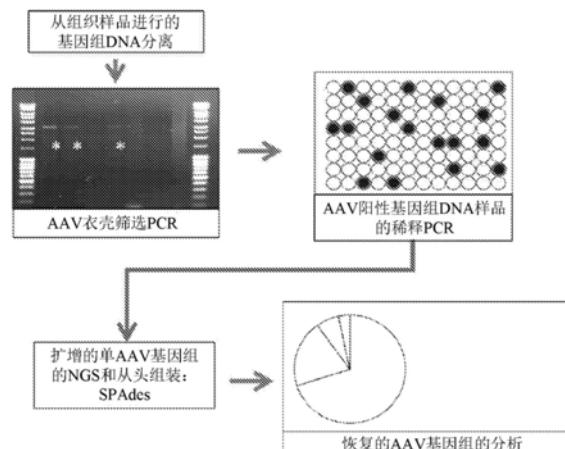
序列表13页 附图19页

(54) 发明名称

新型AAV衣壳和含有其的组合物

(57) 摘要

本文提供新型AAV衣壳和包含其的rAAV。在一个实施例中，相比于现有技术AAV，采用新型AAV衣壳的载体展示增加的所选靶组织的转导。



1. 一种重组腺病毒相关病毒(rAAV)，其具有包含衣壳蛋白的AAV衣壳，所述衣壳蛋白包含SEQ ID NO:2 (AAVrh.90) 的氨基酸序列，且在所述衣壳中包装有包含异源核酸序列的载体基因组。

2. 一种重组腺相关病毒(rAAV)，其具有包含衣壳蛋白的衣壳，所述衣壳蛋白通过SEQ ID NO:1的AAV衣壳序列，或与SEQ ID NO:1具有至少90%、至少95%、至少97%、至少98%、或至少99%一致性的序列的表达产生，且在所述衣壳中包装有包含异源核酸序列的载体基因组。

3. 根据权利要求1或2所述的rAAV，其中所述衣壳蛋白由SEQ ID NO:1编码。

4. 根据权利要求1至3中任何一项所述的rAAV，其进一步包含5' AAV反向末端重复序列(5' ITR)和3' AAV ITR以及与调节序列可操作地连接的异源核酸序列，所述调节序列引导由所述异源核酸序列编码的产物在靶细胞中的表达。

5. 根据权利要求4所述的rAAV，其中所述AAV ITR序列来自除AAVrh.90以外的AAV。

6. 根据权利要求5所述的rAAV，其中所述ITR序列来自AAV2。

7. 根据权利要求2至6中任一项所述的rAAV，其中所述AAV衣壳包含AAV衣壳蛋白，所述衣壳蛋白包含：

(1) 选自以下的AAVrh.90vp1蛋白的异质群体：通过由编码SEQ ID NO:2的1至738的预测氨基酸序列的核酸序列的表达产生的vp1蛋白、由SEQ ID NO:1产生的vp1蛋白、或由与SEQ ID NO:1具有至少70%一致性的核酸序列产生的vp1蛋白，所述核酸序列编码SEQ ID NO:2的1至738的预测氨基酸序列；

选自以下的AAVrh.90vp2蛋白的异质群体：通过由编码SEQ ID NO:2的至少约氨基酸138至738的预测氨基酸序列的核酸序列的表达产生的vp2蛋白、由包含SEQID NO:1的至少核苷酸412至2214的序列产生的vp2蛋白、或由与SEQ ID NO:1的至少核苷酸412至2214具有至少70%一致性的核酸序列产生的vp2蛋白，所述核酸序列编码SEQ ID NO:2的至少约氨基酸138至738的预测氨基酸序列；

选自以下的AAVrh.90vp3蛋白的异质群体：通过由编码SEQ ID NO:2的至少约氨基酸204至738的预测氨基酸序列的核酸序列的表达产生的vp3蛋白、由包含SEQID NO:1的至少核苷酸610至2214的序列产生的vp3蛋白、或由与SEQ ID NO:1的至少核苷酸610至2214具有至少70%一致性的核酸序列产生的vp3蛋白，所述核酸序列编码SEQ ID NO:2的至少约氨基酸204至738的预测氨基酸序列；和/或

(2) 作为编码SEQ ID NO:2的氨基酸序列的核酸序列的产物的vp1蛋白的异质群体、作为编码SEQ ID NO:2的至少约氨基酸138至738的氨基酸序列的核酸序列的产物的vp2蛋白的异质群体和作为编码SEQ ID NO:2的至少氨基酸204至738的核酸序列的产物的vp3蛋白的异质群体，其中：vp1、vp2和vp3蛋白含有具有氨基酸修饰的亚群，所述氨基酸修饰包含SEQ ID NO:2中的天冬酰胺-甘氨酸对中的至少两个高度脱酰胺化的天冬酰胺(N)，并且任选地进一步含有包含其它脱酰胺化的氨基酸的亚群，其中所述脱酰胺化引起氨基酸变化。

8. 根据权利要求7所述的rAAV，其中编码所述蛋白质的所述核酸序列为SEQ ID NO:1，或与SEQ ID NO:1具有至少80%到至少99%一致性的序列，所述序列编码SEQ ID NO:2的氨基酸序列。

9. 根据权利要求7或8所述的rAAV，其中所述核酸序列与SEQ ID NO:1具有至少80%至

97%一致性。

10. 根据权利要求7至9中任一权利要求所述的rAAV,其中所述AAV ITR序列为来自除AAVrh.90以外的AAV的5'ITR和3'ITR。

11. 根据权利要求10所述的rAAV,其中所述ITR序列来自AAV2。

12. 一种组合物,其至少包含根据权利要求1至11中任一项所述的rAAV和生理学上相容的载体、缓冲剂、佐剂和/或稀释剂。

13. 根据权利要求12所述的组合物,其中所述组合物被调配成用于鞘内递送,且所述载体基因组包含编码用于递送至中枢神经系统的基因产物的核酸序列。

14. 根据权利要求12所述的组合物,其中所述组合物被调配成用于静脉内递送。

15. 根据权利要求12所述的组合物,其中所述组合物被调配成用于鼻内或肌肉内递送。

16. 根据权利要求1至11中任一项所述的rAAV,其用于向有需要的受试者递送所需基因产物。

17. 一种根据权利要求1至11中任一项所述的rAAV或根据权利要求12至15中任一项所述的组合物的用途,其用于向有需要的受试者递送所需基因产物。

18. 一种适用于生产根据权利要求1至11中任一项所述的rAAV的生产系统,其中所述生产系统包含:

(a) 编码SEQ ID NO:2的氨基酸序列的核酸序列;

(b) 适合于包装到所述AAV衣壳中的核酸分子,所述核酸分子包含至少一个AAV反向末端重复序列(ITR)和编码基因产物的非AAV核酸序列,所述基因产物与引导所述产物在宿主细胞中的表达的序列可操作地连接;和

(c) 允许将所述核酸分子包装到所述rAAV衣壳中的足够的AAV rep功能和辅助功能。

19. 根据权利要求18所述的系统,其中(a)的所述核酸序列至少包含SEQ ID NO:1,或与SEQ ID NO:1具有至少70%到至少99%一致性的序列,所述序列编码SEQ ID NO:2的氨基酸序列。

20. 根据权利要求18或19中任一项所述的系统,其中所述细胞培养物包含人类胚肾293细胞。

21. 根据权利要求18至20中任一项所述的系统,其中所述AAV rep来自除AAVrh.90以外的AAV。

22. 根据权利要求21所述的系统,其中所述AAV rep来自AAV2。

23. 一种产生rAAV的方法,其包含培养宿主细胞的步骤,所述宿主细胞含有:(a) 编码包含SEQ ID NO:2的氨基酸序列的AAV衣壳蛋白的核酸分子;(b) 功能rep基因;(c) 包含AAV 5'ITR、AAV 3'ITR和转基因的小基因;和(d) 允许将所述小基因包装到AAV衣壳中的足够辅助功能。

24. 一种用根据权利要求1至11中任一项所述的rAAV转导的宿主细胞。

25. 一种将转基因递送至细胞的方法,所述方法包含使所述细胞与根据权利要求1至11中任一项所述的rAAV接触的步骤,其中所述rAAV包含所述转基因。

26. 一种核酸分子,其包含编码AAV衣壳蛋白的核酸序列,所述核酸序列至少包含SEQ ID NO:1,或与SEQ ID NO:1具有至少70%到至少99%一致性的序列,所述序列编码SEQ ID NO:2的氨基酸序列。

27. 根据权利要求26所述的核酸分子，其中所述分子为质粒。
28. 一种用根据权利要求26或27所述的核酸分子转染的宿主细胞。

## 新型AAV衣壳和含有其的组合物

### 背景技术

[0001] 腺相关病毒 (AAV) 在人类基因疗法中具有广阔的前景,且由于其能够提供长期基因表达且缺乏致病性,因此在各种研究中已广泛用于靶向肝脏、肌肉、心脏、脑、眼睛、肾脏和其它组织。AAV属于细小病毒家族,且每个都含有单股DNA,两侧是两个反向末端重复序列。已经报道了数十种天然存在的AAV衣壳;其独特衣壳结构使其能够识别和转导不同的细胞类型和器官。

[0002] 自1981年开始第一次试验以来,基于AAV载体的基因治疗的临床试验中没有报告任何与载体相关的毒性。AAV载体在临床试验中不断积累的安全记录,结合已证明的功效,表明AAV是一个有前景的基因递送平台。另一个有吸引力的特征是AAV相对容易操纵,因为它是一种单股DNA病毒,具有小基因组 (~4.7kb) 和简单的遗传成分-反向末端重复序列 (ITR) 以及Rep和Cap基因。AAV载体中只需要ITR和AAV衣壳蛋白,其中ITR作为载体生产的复制和包装信号,且衣壳蛋白不仅形成衣壳以容纳载体基因组DNA,而且决定组织向性以将载体基因组递送到靶细胞和组织。

[0003] 由于AAV的低免疫原性和非致病性,AAV是基因疗法最有效的候选载体之一。然而,尽管允许有效的基因转移,但目前在临床中使用的AAV载体可能会受到对病毒的预先存在的免疫力和限制的组织向性的阻碍。需要新的和更有效的AAV载体。

### 发明内容

[0004] 在一个实施例中,本文提供一种重组腺相关病毒 (rAAV),其具有包含衣壳蛋白的AAV衣壳,所述衣壳蛋白包含SEQ ID NO:2 (AAVrh.90) 的氨基酸序列,且在衣壳中包装有包含异源核酸序列的载体基因组。在某些实施例中,rAAV具有包含衣壳蛋白的衣壳,所述衣壳蛋白通过SEQ ID NO:1的AAV衣壳序列,或与SEQ ID NO:1具有至少90%、至少95%、至少97%、至少98%、或至少99%一致性的序列的表达产生,且在衣壳中包装有包含异源核酸序列的载体基因组。

[0005] 在某些实施例中,本文提供rAAV,其中AAV衣壳包含AAV衣壳蛋白,所述AAV衣壳蛋白包含:(1)选自以下的AAVrh.90vp1蛋白的异质群体:通过由编码SEQ ID NO:2的1至738的预测氨基酸序列的核酸序列的表达产生的vp1蛋白、由SEQ ID NO:1产生的vp1蛋白、或由与SEQ ID NO:1具有至少70%一致性的核酸序列产生的vp1蛋白,所述核酸序列编码SEQ ID NO:2的1至738的预测氨基酸序列;选自以下的AAVrh.90vp2蛋白的异质群体:通过由编码SEQ ID NO:2的至少约氨基酸138至738的预测氨基酸序列的核酸序列的表达产生的vp2蛋白、由包含SEQ ID NO:1的至少核苷酸412至2214的序列产生的vp2蛋白、或由与SEQ ID NO:1的至少核苷酸412至2214具有至少70%一致性的核酸序列产生的vp2蛋白,所述核酸序列编码SEQ ID NO:2的至少约氨基酸138至738的预测氨基酸序列;选自以下的AAVrh.90vp3蛋白的异质群体:通过由编码SEQ ID NO:2的至少约氨基酸204至738的预测氨基酸序列的核酸序列的表达产生的vp3蛋白、由包含SEQ ID NO:1的至少核苷酸610至2214的序列产生的vp3蛋白、或由与SEQ ID NO:1的至少核苷酸610至2214具有至少70%一致性的核酸序列产

生的vp3蛋白,所述核酸序列编码SEQ ID NO:2的至少约氨基酸204至738的预测氨基酸序列;和/或(2)作为编码SEQ ID NO:2的氨基酸序列的核酸序列的产物的vp1蛋白的异质群体、作为编码SEQ ID NO:2的至少约氨基酸138至738的氨基酸序列的核酸序列的产物的vp2蛋白的异质群体和作为编码SEQ ID NO:2的至少氨基酸204至738的核酸序列的产物的vp3蛋白的异质群体,其中:vp1、vp2和vp3蛋白含有具有氨基酸修饰的亚群,所述氨基酸修饰包含SEQ ID NO:2中的天冬酰胺-甘氨酸对中的至少两个高度脱酰胺化的天冬酰胺(N);并且任选地进一步含有包含其它脱酰胺化的氨基酸的亚群,其中所述脱酰胺化引起氨基酸变化。

[0006] 在另一实施例中,本文提供一种组合物,其至少包含rAAV和生理学上相容的载体、缓冲剂、佐剂和/或稀释剂。在某些实施例中,组合物被调配成用于鞘内递送,且载体基因组包含编码用于递送至中枢神经系统的基因产物的核酸序列。在另一实施例中,组合物被调配成用于静脉内递送、鼻内和/或肌肉内递送。

[0007] 在某些实施例中,提供一种适用于产生rAAV的系统。所述系统包含:(a)编码SEQ ID NO:2的氨基酸序列的核酸序列;(b)适合于包装至AAV衣壳中的核酸分子,其中所述核酸分子包含至少一个AAV反向末端重复序列(ITR)和编码基因产物的非AAV核酸序列,所述核酸序列与引导所述产物在宿主细胞中的表达的序列可操作地连接;和(c)允许将核酸分子包装至rAAV衣壳中的足够AAV rep功能和辅助功能。

[0008] 在某些实施例中,提供一种产生包含AAV衣壳的rAAV的方法。所述方法包含培养宿主细胞的步骤,所述宿主细胞含有:(a)编码包含SEQ ID NO:2的氨基酸序列的AAV衣壳蛋白的核酸分子;(b)功能rep基因;(c)包含AAV 5'ITR、AAV 3'ITR和转基因的小基因;和(d)允许将所述小基因包装到AAV衣壳中的足够辅助功能。

[0009] 在又一实施例中,提供含有本文所述的rAAV、表达盒或核酸分子的宿主细胞。

[0010] 在某些实施例中,提供将转基因递送到细胞的方法。所述方法包括使细胞与如本文所述的rAAV接触的步骤,其中rAAV包含转基因。

[0011] 这些组合物和方法的其它方面和优点在以下具体实施方式中进一步描述。

## 附图说明

[0012] 图1示出用于AAV-SGA工作流程的图。从恒河猴组织样本分离基因组DNA并且针对AAV衣壳基因的存在进行筛选。对AAV阳性DNA进行终点稀释且进行另一轮PCR。根据泊松分布,在不超过30%的孔中产生PCR产物的DNA稀释液在80%的时间内每个阳性PCR含有一个可扩增的DNA模板。使用Illumina MiSeq 2x150或2x250配对末端测序平台对阳性扩增子进行测序,并使用SPAdes组装器从头组装所得读段。

[0013] 图2是示出新型AAV天然分离物和代表性进化枝对照的DNA基因组序列的相邻连接系统发生的图。

[0014] 图3A-图3C示出AAVrh.90 (SEQ ID NO:1) 和AAV8 (SEQ ID NO:3) 衣壳的核酸序列的比对。

[0015] 图4示出AAVrh.90 (SEQ ID NO:2) 和AAV8 (SEQ ID NO:4) 衣壳的氨基酸序列的比对。

[0016] 图5A-图5D示出注射后14天小鼠组织中的eGFP转基因生物分布。(图5A和图5B) 以

每只小鼠 $1e12$  GC的剂量向C57BL/6小鼠IV注射含有CB7.CI.eGFP.WPRE.RBG的AAV衣壳( $n=5$ )。(图5C和图5D)以每只小鼠 $1e11$  GC的剂量向C57BL/6小鼠脑室内ICV注射含有CB7.CI.eGFP.WPRE.RBG的各种AAV衣壳(以 $6.9e10$  GC/小鼠给药的进化枝A载体)( $n=5$ )。值表示为平均值±SD;\* $p<0.01$ ,\*\* $p<0.001$ 。

[0017] 图6A和图6B示出IM递送AAV载体后肌肉中的LacZ表达的分析。向小鼠施用 $3e9$  GC的具有各种衣壳并在CMV启动子下表达LacZ的载体。在第20天,收集肌肉组织,并通过X-gal染色(较深染色)评估转基因表达。

[0018] 图7示出IM递送各种AAV载体后血清中mAb的含量。向B6小鼠施用 $1e11$  GC的载体,所述载体在tMCK启动子下表达3D6抗体。

[0019] 图8示出表达3D6或LacZ转基因的载体的产率(相对于AAV8)。

[0020] 图9示出NHP中汇集的带条形码的载体研究的实验设计(图10A-图10C中示出的数据)。五个新型衣壳和五个对照(AAVrh.90、AAVrh9.1、AAVrh.92、AAVrh.93、AAVrh.91.93、AAV8、AAV6.2、AAVrh32.33、AAV7和AAV9)用具有独特6bp条形码的改良的ATG耗尽GFP转基因包装。将载体以等量汇集并在食蟹猴体内IV或ICM注射(总剂量: $2e13$  GC/kg IV和 $3e13$  GC ICM)。IV注射的动物在基线时对AAV6、AAV8和AAVrh32.33呈血清阴性,并且针对AAV7和AAV9的中和抗体滴度分别为1:5和1:10。

[0021] 图10A-图10C是示出在IV递送(图10A和图10B)和ICM递送(图10C)之后带条形码的衣壳的RNA表达分析的图。IV施用 $-2e13$  GC/kg总剂量,第30天尸体剖检(此动物在基线时具有低含量的AAV7和AAV9 Nabs)。ICM施用 $-3e13$  GC/动物,第30天尸体剖检。每个组织RNA样品中的条形码频率被归一化为注射输入材料中的频率,使得每个条码在混合物中具有等效的表示(10%)。十个载体的输入量范围为8.5-12%。值表示为平均值±SEM,\*\* $p<0.001$ 。

[0022] 图11是示出NHP条形码研究中使用的各种AAV衣壳的小规模制剂滴度的条形图。每个点代表一个单独的小规模制剂。

[0023] 图12A和图12B示出AAVrh.90载体制剂的质谱分析结果。

## 具体实施方式

[0024] 通过使用AAV单基因组扩增来探索AAV在其天然哺乳动物宿主中的遗传变异,所述技术用于从病毒群体内准确分离单个AAV基因组(图1)。本文描述从恒河猴组织中分离出新的AAV序列,所述序列可以归类为各种进化枝。我们在静脉内(IV)和脑室内(ICV)递送后在小鼠中以及在IV和大池内(ICM)递送后在NHP中评估了天然分离物衍生的AAV载体的生物学特性。与原型进化枝成员对照相比时,结果确定了新AAV变体的进化枝特异性和可变转导模式。

[0025] 本文提供了重组AAVrh.90载体,其具有AAVrh.90衣壳和编码处于调节序列控制下的转基因的核酸,在递送至受试者后,所述调节序列引导所述转基因的表达。rAAVrh.90衣壳含有独立地具有SEQ ID NO:2的氨基酸序列的蛋白质。提供含有这些载体的组合物。本文所述的方法涉及使用rAAV靶向所关注的组织以治疗各种病状。

[0026] 在某些实施例中,本文提供包含AAVrh.90衣壳的载体,其非常适合将转基因递送至中枢神经系统的细胞。在某些实施例中,需要鞘内递送,包括例如通过ICM递送而递送至脑或脊髓。在某些实施例中,包含AAVrh.90衣壳的载体非常适合将转基因递送至外周器官,

包括肝脏、心脏、骨骼(横纹)肌、肾脏和胰腺的细胞。AAVrh.90载体可以全身递送或通过适合靶向这些组织的施用途径靶向递送。

[0027] 除非另外定义,否则本文所使用的技术和科学术语具有与本发明所属领域的普通技术人员通常所理解和参照公开的文本相同的含义,这些术语为本领域的技术人员提供了本申请中使用的许多术语的通用指南。以下定义仅出于清楚起见而提供且并不打算限制所要求的发明。如本文所用,术语“一个(a/an)”是指一个或多个,例如,“一个宿主细胞”应理解为表示一个或多个宿主细胞。如此,术语“一个(a/an)”、“一个或多个”和“至少一个”在本文可互换地使用。如本文所用,除非另有说明,否则术语“约”意指相对于给定参考的10%的变化性。虽然说明书中的各个实施例是使用“包含”语言来呈现的,但是在其它情况下,也意图使用“由……组成”或“基本上由……组成”的语言来解释和描述相关的实施例。

[0028] 关于以下描述,希望在另一实施例中,本文描述的每种组合物适用于本发明的方法。另外,还希望在另一实施例中,被描述为适用于所述方法的每种组合物本身是本发明的实施例。

[0029] “重组AAV”或“rAAV”是含有两个元件的DNA酶抗性病毒颗粒,所述两个元件即AAV衣壳和至少含有包装在AAV衣壳内的非AAV编码序列的载体基因组。除非另有说明,否则此术语可以与短语“rAAV载体”互换使用。rAAV是“复制缺陷型病毒”或“病毒载体”,因为其缺少任何功能性AAV rep基因或功能性AAV cap基因并且不能产生子代。在某些实施例中,仅AAV序列是AAV反向末端重复序列(ITR),通常定位在载体基因组的5'和3'最末端处,以允许定位在ITR之间的基因和调节序列包装在AAV衣壳内。

[0030] 如本文所用,“载体基因组”是指包装在形成病毒颗粒的rAAV衣壳内部的核酸序列。此类核酸序列含有AAV反向末端重复序列(ITR)。在本文的实例中,载体基因组至少含有从5'到3'的AAV 5' ITR、编码序列和AAV 3' ITR。在某些实施例中,ITR来自AAV2,可以选择与衣壳不同的源AAV,或另一全长ITR。在某些实施例中,ITR来自与在产生或反式补充AAV期间提供rep功能的AAV源相同的AAV。另外,可以使用其它ITR。另外,载体基因组含有引导基因产物表达的调节序列。在本文中更详细地论述载体基因组的合适成分。载体基因组在本文中有时被称作“小基因”。

[0031] 术语“表达盒”是指包含转基因序列和其调节序列(例如启动子、增强子、polyA)的核酸分子,所述表达盒可以被包装到病毒载体(例如病毒颗粒)的衣壳中。通常,这种用于产生病毒载体的表达盒含有转基因序列,所述转基因序列的两侧是病毒基因组的包装信号和其它表达控制序列,例如本文所述的那些。例如,对于AAV病毒载体,包装信号是5'反向末端重复序列(ITR)和3' ITR。在某些实施例中,术语“转基因”可以与“表达盒”互换使用。在其它实施例中,术语“转基因”仅仅是指所选基因的编码序列。

[0032] rAAV由AAV衣壳和载体基因组构成。AAV衣壳是vp1的异质群体、vp2的异质群体和vp3蛋白的异质群体的组装。如本文所用,当用于指vp衣壳蛋白时,术语“异源”或其任何语法变型是指由不相同的元件组成的群体,例如具有含有不同的经过修饰的氨基酸序列的vp1、vp2或vp3单体(蛋白)。

[0033] 如本文所用,与vp1、vp2和vp3蛋白(可替代地被称为同种型)结合使用的术语“异质群体”是指衣壳内的vp1、vp2和vp3蛋白的氨基酸序列中的差异。AAV衣壳含有具有来自预测的氨基酸残基的修饰的vp1蛋白内、vp2蛋白内和vp3蛋白内的亚群。这些亚群至少包括某

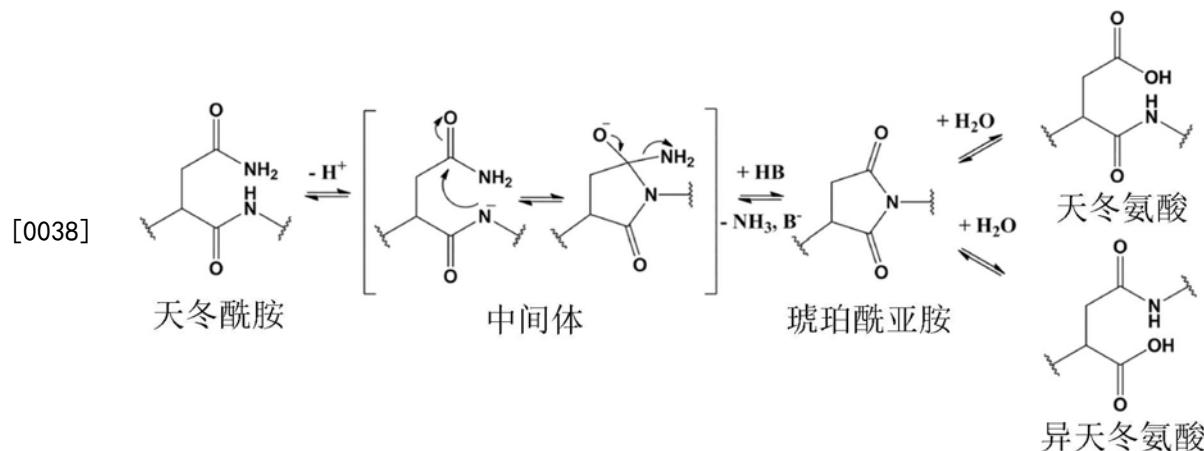
些脱酰胺的天冬酰胺(N或Asn)残基。例如，某些亚群包含天冬酰胺-甘氨酸对中的至少一个、两个、三个或四个高度脱酰胺的天冬酰胺(N)位置，并且任选地进一步包含其它脱酰胺的氨基酸，其中脱酰胺引起氨基酸变化和其它任选的修饰。参见2019年2月27日提交的PCT/US19/019804和2019年2月27日提交的PCT/US19/019861，其各自以引用的方式并入本文中。

[0034] 如本文所用,除非另有说明,否则vp蛋白的“亚群”是指一组vp蛋白,所述一组vp蛋白具有至少一个限定的共同特性,并且由参考组的至少一个组成员到少于所有成员组成。例如,除非另有说明,否则vp1蛋白的“亚群”可以是组装的AAV衣壳中的至少一种(1) vp1蛋白,并且少于所有vp1蛋白。除非另有说明,否则vp3蛋白的“亚群”可以是少于组装的AAV衣壳中的所有vp3蛋白的一种(1) vp3蛋白。例如,vp1蛋白可以是vp蛋白的亚群;vp2蛋白可以是vp蛋白的单独的亚群,并且vp3是组装的AAV衣壳中的vp蛋白的又另外的亚群。在另一实例中,vp1、vp2和vp3蛋白可含有具有不同修饰的亚群,例如,至少一种、两种、三种或四种高度脱酰胺化的天冬酰胺,例如在天冬酰胺-甘氨酸对处。

[0035] 除非另有说明,否则高度脱酰胺化是指与在参考氨基酸位置处的预测的氨基酸序列相比,在参考的氨基酸位置处被至少45%脱酰胺化、至少50%脱酰胺化、至少60%脱酰胺化、至少65%脱酰胺化、至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少99%或至多约100%脱酰胺化。此类百分比可以使用2D凝胶、质谱技术或其它合适的技术来确定。

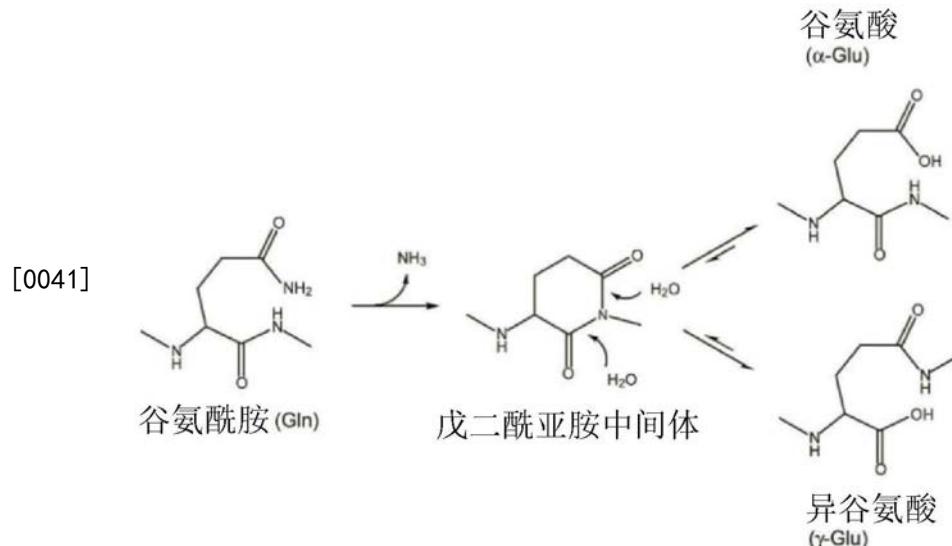
[0036] 不希望受理论束缚,AAV衣壳中的vp蛋白中的至少高度脱酰胺的残基的脱酰胺化被认为本质上主要是非酶促的,这是由衣壳蛋白内的官能团引起的,所述官能团使所选天冬酰胺脱酰胺化,并在较小程度上使谷氨酰胺残基脱酰胺化。大多数脱酰胺化vp1蛋白的有效衣壳组装表明这些事件在衣壳组装之后发生,或者单独单体(vp1、vp2或vp3)中的脱酰胺化在结构上具有良好的耐受性,并且在很大程度上不会影响组装动力学。VP1独特(VP1-u)区(约aa 1-137)中的广泛的脱酰胺化通常被认为在细胞进入之前定位在内部,这表明VP脱酰胺化可以在衣壳组装之前发生。

[0037] 不希望受理论束缚,N的脱酰胺化可以通过其C端残基的骨架氮原子对Asn的侧链酰胺基碳原子进行亲核攻击而发生。据信形成了中间闭环的琥珀酰亚胺残基。然后,琥珀酰亚胺残基进行快速水解以产生最终产物天冬氨酸(Asp)或异天冬氨酸(IsoAsp)。因此,在某些实施例中,天冬酰胺(N或Asn)的脱酰胺化产生Asp或IsoAsp,其可以通过琥珀酰亚胺中间体相互转化,例如如下文所示的。



[0039] 如本文所提供的,VP1、VP2或VP3中的每个脱酰胺的N可以独立地是天冬氨酸(Asp)、异天冬氨酸(isoAsp)、天冬氨酸和/或Asp和isoAsp的相互转化的共混物或其组合。可以存在任何合适比率的 $\alpha$ -和异天冬氨酸。例如,在某些实施例中,比率可以是10:1到1:10天冬氨酸:异天冬氨酸、约50:50天冬氨酸:异天冬氨酸或约1:3天冬氨酸:异天冬氨酸,或另一所选比率。

[0040] 在某些实施例中,一种或多种谷氨酰胺(Q)可以脱酰胺化为谷氨酸(Glu),即 $\alpha$ -谷氨酸、 $\gamma$ -谷氨酸(Glu)或 $\alpha$ -和 $\gamma$ -谷氨酸的共混物,其可以通过常见的戊二酰亚胺(glutarimide)中间体相互转化。可以存在任何合适比率的 $\alpha$ -和 $\gamma$ -谷氨酸。例如,在某些实施例中,比率可以是10:1到1:10 $\alpha$ : $\gamma$ 、约50:50 $\alpha$ : $\gamma$ 或约1:3 $\alpha$ : $\gamma$ 或另一所选比率。



[0042] 因此, rAAV包括vp1、vp2和/或vp3蛋白的rAAV衣壳内具有脱酰胺化的氨基酸的亚群,至少包括至少一个包含至少一种高度脱酰胺化的天冬酰胺的亚群。另外,其它修饰可以包括异构化,特别是在所选天冬氨酸(D或Asp)残基位置处。在仍其它实施例中,修饰可以包括在Asp位置处的酰胺化。

[0043] 在某些实施例中,AAV衣壳含有具有至少1个、至少2个、至少3个、至少4个、至少5个到至少约25个脱酰胺化的氨基酸残基位置的vp1、vp2和vp3的亚群,与vp蛋白的经过编码的氨基酸序列相比,所述vp1、vp2和vp3的至少1%到10%、至少10%到25%、至少25%到50%、至少50%到70%、至少70%到100%、至少75%到100%、至少80%到100%或至少90%到100%被脱酰胺化。这些中的大部分可以是N残基。然而,Q残基也可以被脱酰胺化。

[0044] 如本文所用,“经编码的氨基酸序列”是指基于被翻译成氨基酸的参考核酸序列的已知DNA密码子的翻译而预测的氨基酸。下表展示了DNA密码子和二十种常见氨基酸,分别示出了单字母代码(SLC)和三个字母代码(3LC)。

[0045]

氨基酸	SLC	3 LC	DNA 密码子
异亮氨酸	I	Ile	ATT、ATC、ATA
亮氨酸	L	Leu	CTT、CTC、CTA、CTG、TTA、TTG
缬氨酸	V	Val	GTT、GTC、GTA、GTG
苯丙氨酸	F	Phe	TTT、TTC
甲硫氨酸	M	Met	ATG
半胱氨酸	C	Cys	TGT、TGC
丙氨酸	A	Ala	GCT、GCC、GCA、GCG
甘氨酸	G	Gly	GGT、GGC、GGA、GGG
脯氨酸	P	Pro	CCT、CCC、CCA、CCG
苏氨酸	T	Thr	ACT、ACC、ACA、ACG
丝氨酸	S	Ser	TCT、TCC、TCA、TCG、AGT、AGC
酪氨酸	Y	Tyr	TAT、TAC
色氨酸	W	Trp	TGG
谷氨酰胺	Q	Gln	CAA、CAG
天冬酰胺	N	Asn	AAT、AAC
组氨酸	H	His	CAT、CAC
谷氨酸	E	Glu	GAA、GAG
天冬氨酸	D	Asp	GAT、GAC
赖氨酸	K	Lys	AAA、AAG
精氨酸	R	Arg	CGT、CGC、CGA、CGG、AGA、AGG
终止密码子	终止		TAAT、TAG、TGA

[0046] 在某些实施例中, rAAV具有含有vp1、vp2和vp3蛋白的AAV衣壳, 所述蛋白具有包含在本文提供的表中所阐述的位置处的两个、三个、四个、五个或更多个脱酰胺化的残基的组合的亚群, 并以引用的方式并入本文中。

[0047] rAAV中的脱酰胺化可以使用2D凝胶电泳和/或质谱和/或蛋白质建模技术来确定。在线色谱可以使用Acclaim PepMap柱和与Q Exactive HF和NanoFlex源(赛默飞世尔科技公司(Thermo Fisher Scientific))耦合的Thermo UltiMate 3000RSLC系统(赛默飞世尔科技公司)执行。MS数据是使用用于Q Exactive HF的数据依赖性前20种方法获取的, 所述方法从调查扫描(200-2000m/z)中动态选择最丰富的尚未测序的前体离子。测序通过高能碰撞解离片段进行, 其中通过预测性自动增益控制确定的目标值为1e5离子, 并且以4m/z的窗口进行前体分离。在m/z 200下以120,000的分辨率获取调查扫描。HCD光谱的分辨率可以在m/z 200下设置为30,000, 其中最大离子注入时间为50毫秒, 并且归一化碰撞能量为30。S-透镜RF水平可以设置为50, 以使消化肽所占据的m/z区达到最佳传输。可以从片段选择中排除具有单个、未分配或六个和更高电荷状态的前体离子。BioPharma Finder 1.0软件(赛默飞世尔科技公司)可以用于分析所获取的数据。对于肽作图, 使用单进入蛋白FASTA数据库进行搜索, 其中脲基甲基化设置为固定的修饰; 并将氧化、脱酰胺化和磷酸化设置为可变

修饰、10ppm质量准确度、高蛋白酶特异性和置信水平为0.8的MS/MS光谱。合适的蛋白酶的实例可包括例如胰蛋白酶或胰凝乳蛋白酶。脱酰胺化的肽的质谱鉴定相对简单,因为脱酰胺化向完整分子的质量添加了+0.984Da (-OH基团与-NH<sub>2</sub>基团之间的质量差)。特定肽的脱酰胺化百分比通过将脱酰胺化的肽的质量面积除以脱酰胺化的和天然的肽的面积之和来确定。考虑到可能的脱酰胺化位点的数量,在不同位点处被脱酰胺化的同量异位物种可以在单个峰中共迁移。因此,源自具有多个潜在的脱酰胺化位点的肽的片段离子可以用于定位或区分多个脱酰胺化位点。在这些情况下,观察到的同位素图案内的相对强度可以用于特异性确定不同的脱酰胺化的肽异构体的相对丰度。此方法假设所有异构物种的片段化效率是相同的,并且在脱酰胺化位点上是独立的。本领域的技术人员将理解的是,可以使用这些说明性方法的多种变型。例如,合适的质谱仪可包括例如四极杆飞行时间质谱仪(QTOF),如Waters Xevo或Agilent 6530,或Orbitrap仪器,如Orbitrap Fusion或Orbitrap Velos(赛默飞世尔科技公司)。合适的液相色谱系统包括例如来自沃特世(Waters)的Acquity UPLC系统或Agilent系统(1100或1200系列)。合适的数据分析软件可包括例如MassLynx(沃特世)、Pinpoint和Petfinder(赛默飞世尔科技公司)、Mascot(矩阵科学公司(Matrix Science))、Peaks DB(生物信息学解决方案公司(Bioinformatics Solutions))。可以在例如X.Jin等人于2017年6月16日在线公开的《人类基因疗法方法(Hu Gene Therapy Methods)》,第28卷,第5期,第255-267页中描述其它技术。

[0048] 除脱酰胺化之外,可能发生不会导致一个氨基酸转化为不同的氨基酸残基的其它修饰。这种修饰可包括乙酰化残基、异构化、磷酸化或氧化。

[0049] 脱酰胺化的调节:在某些实施例中,AAV被修饰成改变天冬酰胺-甘氨酸对中的甘氨酸,以降低脱酰胺化。在其它实施例中,将天冬酰胺改变为不同的氨基酸,例如以较慢速率进行脱酰胺化的谷氨酰胺;或改变为缺乏酰胺基的氨基酸(例如,含有酰胺基的谷氨酰胺和天冬酰胺);和/或改变为缺乏胺基的氨基酸(例如,含有胺基的赖氨酸、精氨酸和组氨酸)。如本文所用,缺乏酰胺或胺侧基的氨基酸是指例如甘氨酸、丙氨酸、缬氨酸、亮氨酸、异亮氨酸、丝氨酸、苏氨酸、胱氨酸、苯丙氨酸、酪氨酸或色氨酸和/或脯氨酸。如所描述的修饰可以在经过编码的AAV氨基酸序列中存在的天冬酰胺-甘氨酸对中的一个、两个或三个天冬酰胺-甘氨酸对中。在某些实施例中,在所有四个天冬酰胺-甘氨酸对中没有进行此类修饰。因此,一种用于降低具有较低脱酰胺化速率的AAV和/或工程化AAV变体的脱酰胺化的方法。另外地或可替代地,可以将一种或多种其它酰胺氨基酸改变为非酰胺氨基酸以降低AAV的脱酰胺化。在某些实施例中,本文所描述的突变AAV衣壳含有天冬酰胺-甘氨酸对中的突变,使得甘氨酸变为丙氨酸或丝氨酸。突变AAV衣壳可以含有一个、两个或三个突变体,其中参考AAV天然地含有四个NG对。在某些实施例中,AAV衣壳可以含有一个、两个、三个或四个此类突变体,其中参考AAV天然地含有五个NG对。在某些实施例中,突变AAV衣壳含有NG对中的仅单个突变。在某些实施例中,突变AAV衣壳含有两个不同的NG对中的突变。在某些实施例中,突变AAV衣壳含有定位在AAV衣壳中的结构上分开的位置中的两个不同的NG对中的突变。在某些实施例中,突变不在VP1独特区中。在某些实施例中,突变中的一个突变不在VP1独特区中。任选地,突变AAV衣壳不含有NG对中的修饰,但是含有突变以最小化或消除定位在NG对的外部的一个或多个天冬酰胺或谷氨酰胺中的脱酰胺化。

[0050] 在某些实施例中,提供了一种增加rAAV载体的效力的方法,所述方法包含使AAV衣

壳工程化,这消除了野生型AAV衣壳中的NG中的一个或多个NG。在某些实施例中,“NG”的“G”的编码序列被工程化为对另一种氨基酸进行编码。在下文的某些实例中,“S”或“A”被取代。然而,可以选择其它合适的氨基酸编码序列。

[0051] 这些氨基酸修饰可以通过常规的基因工程化技术进行。例如,可以产生含有经过修饰的AAV vp密码子的核酸序列,其中修饰天冬酰胺-甘氨酸对中对甘氨酸进行编码的密码子中的一到三个密码子以对除甘氨酸之外的氨基酸进行编码。在某些实施例中,含有经过修饰的天冬酰胺密码子的核酸序列可以在天冬酰胺-甘氨酸对中的一到三个天冬酰胺-甘氨酸对处被工程化,使得经过修饰的密码子对除天冬酰胺之外的氨基酸进行编码。每个经修饰的密码子可以对不同的氨基酸进行编码。或者,改变的密码子中的一个或多个密码子可以对相同的氨基酸进行编码。在某些实施例中,这些经过修饰的AAVrh.90核酸序列可以用于产生具有比天然AAVrh.90衣壳脱酰胺化程度更低的衣壳的突变rAAV。此类突变rAAV可以具有降低的免疫原性和/或增加储存时的稳定性,特别是以悬浮液形式储存时的稳定性。

[0052] 本文还提供了对具有降低的脱酰胺化的AAV衣壳进行编码的核酸序列。设计对此AAV衣壳进行编码的核酸序列,包括DNA(基因组或cDNA)或RNA(例如,mRNA)在本领域的技术范围内。此类核酸序列可以被密码子优化以在所选系统(即,细胞类型)中进行表达并且可以通过各种方法设计。可以使用在线可用的方法(例如,GeneArt)、公开的方法或提供密码子优化服务的公司(例如,DNA2.0)(加利福尼亚州门洛帕克市(Menlo Park,CA))来执行此优化。例如,在国际专利公开第WO 2015/012924号中描述了一种密码子优化方法,其以全文引用的方式并入本文中。还参见例如美国专利公开案第2014/0032186号和美国专利公开案第2006/0136184号。适合地,修饰产物的开放阅读框(ORF)的全长。然而,在一些实施例中,仅ORF的片段可以被改变。通过使用这些方法中的一种方法,可以将频率应用于任何给定的多肽序列,并产生对多肽进行编码的经密码子优化的编码区的核酸片段。许多选项可用于对密码子进行实际改变或者可用于合成如本文所述地设计的密码子优化的编码区。此类修饰或合成可以使用本领域普通技术人员已熟知的标准和常规分子生物学操作来进行。在一种方法中,通过标准方法合成各自长度为80-90个核苷酸且跨越希望的序列的长度的一系列互补寡核苷酸对。这些寡核苷酸对被合成为使得在退火时它们形成80-90个碱基对的双股片段,所述双股片段含有粘性末端,例如在所述对中的各寡核苷酸被合成为延伸超过与所述对中另一寡核苷酸互补的区域3个、4个、5个、6个、7个、8个、9个、10个或更多个碱基。每对寡核苷酸的单股末端被设计为用另一对寡核苷酸的单股末端粘接。允许这些寡核苷酸对粘接,并且然后允许约五至六个这些双链片段经由粘性单股末端一起粘接,并且随后它们连接在一起并克隆到标准细菌克隆载体中,例如可获自加利福尼亚州卡尔斯巴德英杰公司(Invitrogen Corporation,Carlsbad,Calif)的TOPO®载体。然后通过标准方法对构建体进行测序。制备由连接在一起的80到90个碱基对片段的5到6个片段(即约500个碱基对的片段)组成的这些构建体中的若干构建体,以使得整个希望的序列以一系列质粒构建体表示。然后用适当的限制性酶切割这些质粒的插入物并且将其连接在一起以形成最终构建体。然后将最终构建体克隆到标准细菌克隆载体中,并进行测序。另外的方法对于技术人员而言将立即是清楚的。另外,基因合成易于商购获得。

[0053] 在某些实施例中,提供了AAV衣壳,所述AAV衣壳具有含有多个高度脱酰胺化的

“NG”位置的AAV衣壳同种型(即,VP1、VP2、VP3)的异质群体。在某些实施例中,高度脱酰胺化的位置在下文参考预测的全长VP1氨基酸序列鉴定的位置中。在其它实施例中,衣壳基因被修饰成使得参考的“NG”被消融,并且突变体“NG”被工程化到另一个位置中。

[0054] 如本文所用,术语“靶细胞”和“靶组织”可以指打算被受试者AAV载体转导的任何细胞或组织。所述术语可以指肌肉、肝脏、肺、气道上皮、中枢神经系统、神经元、眼睛(眼细胞)或心脏中的任何一种或多种。在一个实施例中,靶组织为肝脏。在另一实施例中,靶组织为心脏。在另一实施例中,靶组织为脑。在另一实施例中,靶组织为肌肉。

[0055] 如本文所用,术语“哺乳动物受试者”或“受试者”包括需要本文所述的治疗或预防方法的任何哺乳动物,尤其包括人类。需要此类治疗或预防的至少哺乳动物包括狗、猫或其它驯养动物、马、家畜、实验动物,包括非人类灵长类动物等。受试者可以是雄性或雌性。

[0056] 如本文所使用,术语“宿主细胞”可以指重组AAV由质粒产生的包装细胞系。在替代方案中,术语“宿主细胞”可以指需要转基因表达的靶细胞。

[0057] A.AAV衣壳

[0058] 本文提供一种新型AAV衣壳蛋白,其具有SEQ ID NO:2中阐述的vp1序列。AAV衣壳由三个重叠编码序列组成,由于使用替代起始密码子,所述编码序列的长度不同。这些可变蛋白被称为VP1、VP2和VP3,其中VP1是最长的且VP3是最短的。AAV粒子由比率为约1:1:10 (VP1:VP2:VP3)的所有三种衣壳蛋白组成。VP3包含在N端的VP1和VP2中,是构建粒子的主要结构成分。可使用若干不同编号系统来指代衣壳蛋白。为方便起见,如本文所用,使用VP1编号来指代AAV序列,所述编号从VP1的第一个残基的aa 1开始。但是,本文所述的衣壳蛋白包括VP1、VP2和VP3(在本文中可与vp1、vp2和vp3互换使用)。衣壳的可变蛋白的编号如下:

[0059] 核苷酸(nt)

[0060] AAVrh.90:vp1-nt 1至2214;vp2-nt 412至2214;vp3-SEQ ID NO:1的nt 610至2214

[0061] AAVrh.90衣壳的核酸序列与AAV8衣壳序列的比对在图3A-图3C中示出。

[0062] 氨基酸(aa)

[0063] AAVrh.90:aa vp1-1至738;vp2-aa 138至738;vp3-SEQ ID NO:2的aa 204至738

[0064] AAVrh.90衣壳的氨基酸序列衣壳与AAV8衣壳序列的比对在图4中示出。

[0065] 本文中包括包含AAVrh.90 (SEQ ID NO:2) 的vp1、vp2和vp3中的至少一个的rAAV。本文还提供包含由AAVrh.90 (SEQ ID NO:1) 的vp1、vp2和vp3中的至少一个编码的AAV衣壳的rAAV。

[0066] 在一个实施例中,提供一种组合物,其包括重组腺相关病毒(rAAV)的混合群体,所述rAAV中的每一个包含:(a)包含约60个由vp1蛋白、vp2蛋白和vp3蛋白构成的衣壳蛋白的AAV衣壳,其中vp1、vp2和vp3蛋白为:由编码所选AAV vp1氨基酸序列的核酸序列产生的vp1蛋白的异质群体、由编码所选AAV vp2氨基酸序列的核酸序列产生的vp2蛋白的异质群体、由编码所选AAV vp3氨基酸序列的核酸序列产生的vp3蛋白的异质群体,其中:vp1、vp2和vp3蛋白含有具有氨基酸修饰的亚群,所述氨基酸修饰包含AAV衣壳中的天冬酰胺-甘氨酸对中的至少两个高度脱酰胺化的天冬酰胺(N),且任选地进一步含有包含其它脱酰胺化的氨基酸的亚群,其中所述脱酰胺化引起氨基酸变化;以及(b)AAV衣壳中的载体基因组,所述载体基因组含有包含AAV反向末端重复序列和编码产物的非AAV核酸序列的核酸分子,所述

核酸序列与引导宿主细胞中产物的表达的序列可操作地连接。

[0067] 在某些实施例中,所述脱酰胺化的天冬酰胺被脱酰胺化为天冬氨酸、异天冬氨酸、相互转化的天冬氨酸/异天冬氨酸对或其组合。在某些实施例中,所述衣壳进一步包含脱酰胺化的谷氨酰胺,所述脱酰胺化的谷氨酰胺被脱酰胺化为( $\alpha$ -谷氨酸、 $\gamma$ -谷氨酸、相互转化的( $\alpha$ -谷氨酸/ $\gamma$ -谷氨酸对或其组合。

[0068] 在某些实施例中,提供了新型的分离的AAVrh.90衣壳。编码AAVrh.90衣壳的核酸序列提供于SEQ ID N0:1中,且编码的氨基酸序列提供于SEQ ID N0:2中。本文提供包含AAVrh.90 (SEQ ID N0:1) 的vp1、vp2和vp3中的至少一个的rAAV。本文还提供包含由AAVrh.90 (SEQ ID N0:1) 的vp1、vp2和vp3中的至少一个编码的AAV衣壳的rAAV。在某些实施例中, vp1、vp2和/或vp3是AAVrh.90 (SEQ ID N0:2) 的全长衣壳蛋白。在其它实施例中, vp1、vp2和/或vp3具有N端和/或C端截短(例如,约1至约10个氨基酸的截短)。

[0069] 在另一方面,提供了一种重组腺相关病毒(rAAV),所述rAAV包含:(A)AAVrh.90衣壳,所述衣壳包含以下中的一种或多种:(1)AAVrh.90衣壳蛋白,其包含:选自以下的AAVrh.90vp1蛋白的异质群体:通过由编码SEQ ID N0:2的1至738的预测氨基酸序列的核酸序列的表达产生的vp1蛋白、由SEQ ID N0:2产生的vp1蛋白、或由与SEQ ID N0:1具有至少70%一致性的核酸序列产生的vp1蛋白,所述核酸序列编码SEQ ID N0:2的1至738的预测氨基酸序列;选自以下的AAVrh.90vp2蛋白的异质群体:通过由编码SEQ ID N0:2的至少约氨基酸138至738的预测氨基酸序列的核酸序列的表达产生的vp2蛋白、由包含SEQ ID N0:2的至少核苷酸412至2214的序列产生的vp2蛋白、或由与SEQ ID N0:1的至少核苷酸412至2214具有至少70%一致性的核酸序列产生的vp2蛋白,所述核酸序列编码SEQ ID N0:2的至少约氨基酸138至738的预测氨基酸序列;选自以下的AAVrh.90vp3蛋白的异质群体:通过由编码SEQ ID N0:2的至少约氨基酸204至738的预测氨基酸序列的核酸序列的表达产生的vp3蛋白、由包含SEQ ID N0:2的至少核苷酸610至2214的序列产生的vp3蛋白、或由与SEQ ID N0:1的至少核苷酸610至2214具有至少70%一致性的核酸序列产生的vp3蛋白,所述核酸序列编码SEQ ID N0:2的至少约氨基酸204至738的预测氨基酸序列;和/或(2)作为编码SEQ ID N0:2的氨基酸序列的核酸序列的产物的vp1蛋白的异质群体、作为编码SEQ ID N0:2的至少约氨基酸138至738的氨基酸序列的核酸序列的产物的vp2蛋白的异质群体和作为编码SEQ ID N0:2的至少氨基酸204至738的核酸序列的产物的vp3蛋白的异质群体,其中:vp1、vp2和vp3蛋白含有具有氨基酸修饰的亚群,所述氨基酸修饰包含SEQ ID N0:2中的天冬酰胺-甘氨酸对中的至少两个高度脱酰胺化的天冬酰胺(N),并且任选地进一步含有包含其它脱酰胺化的氨基酸的亚群,其中所述脱酰胺化引起氨基酸变化;以及(B)AAVrh.90衣壳中的载体基因组,所述载体基因组含有包含AAV反向末端重复序列和编码产物的非AAV核酸序列的核酸分子,所述核酸序列与引导产物在宿主细胞中的表达的序列可操作地连接。

[0070] 在某些实施例中,AAVrh.90vp1、vp2和vp3蛋白含有具有氨基酸修饰的亚群,所述氨基酸修饰包含SEQ ID N0:2中的天冬酰胺-甘氨酸对中的至少两个高度脱酰胺化的天冬酰胺(N),并且任选地进一步含有包含其它脱酰胺化的氨基酸的亚群,其中所述脱酰胺化引起氨基酸变化。相对于SEQ ID N0:2的编号,在N-G对N57、~N263、~N385和/或~N514处观察到高水平的脱酰胺化。如下表和图12B中所示,已在其它残基中观察到脱酰胺化。在某些实施例中,AAVrh.90可具有其它脱酰胺化的残基(例如~N305、~N499和/或~N599,通常为

小于20%)和/或可具有其它修饰,包括磷酸化(例如当存在时,在约2至约30%,或约2至约20%,或约2至约10%范围内)(例如在S149处)或氧化(例如在~W23、~M204、~M212、W248、W282、M405、M473、W480、W505、M526、~N544、M561和/或~M607中的一个或多个处)。任选地,W可以氧化成犬尿氨酸。

[0071] 表-AAVrh.90脱酰胺化

[0072]

基于VP1编号的AAVrh.90脱酰胺化	脱酰胺化%
N57+脱酰胺化	65-90、70-95、80-95、75-100、80-100或90-100
N94+脱酰胺化	2-15或2-5
~N263+脱酰胺化	65-90、70-95、80-95、75-100、80-100或90-100
~N305+脱酰胺化	5-30、5-20或10-20
~N385+脱酰胺化	65-90、70-95、80-95、75-100、80-100或90-100
~N499+脱酰胺化	2-15、2-10或5-10
~N514+脱酰胺化	65-90、70-95、80-95、75-100、80-100或90-100
~N599+脱酰胺化	2-15、2-10或5-10

[0073] 在某些实施例中,如使用胰蛋白酶质谱所测定,在提供的范围内,在上表中鉴定的一个或多个位置中修饰AAVrh.90衣壳。在某些实施例中,如本文所述地修饰一个或多个位置或N之后的甘氨酸。残基数基于本文所提供的AAVrh.90序列。参见SEQ ID NO:2。

[0074] 在某些实施例中,AAVrh.90衣壳包含:作为编码SEQ ID NO:2的氨基酸序列的核酸序列的产物的vp1蛋白的异质群体、作为编码SEQ ID NO:2的至少约氨基酸138到738的氨基酸序列的核酸序列的产物的vp2蛋白的异质群体、以及作为编码SEQ ID NO:2的至少氨基酸204到738的核酸序列的产物的vp3蛋白的异质群体。

[0075] 在某些实施例中,编码AAVrh.90vp1衣壳蛋白的核酸序列提供于SEQ ID NO:1中。在其它实施例中,可以选择与SEQ ID NO:1具有70%到99.9%一致性的核酸序列以表达AAVrh.90衣壳蛋白。在某些其它实施例中,核酸序列与SEQ ID NO:1至少约75%一致、至少80%一致、至少85%、至少90%、至少95%、至少97%一致或至少99%到99.9%一致。然而,可以选择编码SEQ ID NO:2的氨基酸序列的其它核酸序列用于产生rAAV衣壳。在某些实施例中,核酸序列具有SEQ ID NO:1的核酸序列,或与SEQ ID NO:1至少70%至99.9%一致、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%或至少99%一致的序列,所述序列编码SEQ ID NO:2。在某些实施例中,核酸序列具有SEQ ID NO:1的核酸序列,或与SEQ ID NO:1的约nt 412至约nt 2214至少70%至99.9%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少99%一致的序列,所述序列编码SEQ ID NO:2的vp2衣壳蛋白(约aa 138至738)。在某些实施例中,核酸序列具有SEQ ID NO:1的约nt 610至约nt 2214的核酸序列,或与SEQ ID NO:1的约nt 610至约nt 2214至少70%至99.9%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%或至少99%一致的序列,所述序列编码SEQ ID NO:2的vp3衣壳蛋白(约aa 204至738)。

[0076] 本发明还涵盖编码AAVrh.90衣壳序列(SEQ ID NO:2)或突变型AAVrh.90的核酸序列,其中一个或多个残基已被改变以减少脱酰胺化,或本文中鉴定的其它修饰。此类核酸序列可用于产生突变型AAVrh.90衣壳。

[0077] 在某些实施例中,本文提供一种核酸分子,其具有SEQ ID NO:1的序列,或与SEQ

ID NO:1至少70%、至少75%、至少80%、至少85%、至少90%、至少95%、至少97%、至少99%一致的序列，所述序列编码具有如本文所述的修饰(例如，脱酰胺化的氨基酸)的SEQ ID NO:2的vp1氨基酸序列。在某些实施例中，提供具有本文所述的核酸序列的质粒。

[0078] 当提及核酸或其片段时，术语“基本同源性”或“基本类似性”表示当与另一个核酸(或其互补股)的适当核苷酸插入或缺失进行最佳比对时，至少约95到99%的比对序列具有核苷酸序列一致性。优选地，同源性在全长序列、或其开放阅读框或长度为至少15个核苷酸的另一合适的片段上。本文描述了合适片段的实例。

[0079] 在核酸序列的情况下，术语“百分比(%)一致性”、“序列一致性”、“百分比序列一致性”或“百分比一致”是指当比对对应关系时相同的两个序列中的残基。期望序列一致性比较的长度可以超过基因组的全长、基因编码序列的全长或至少约500到5000个核苷酸的片段。然而，也可以期望较小片段之间的一致性，例如至少约九个核苷酸，通常至少约20到24个核苷酸、至少约28到32个核苷酸、至少约36个或更多个核苷酸。

[0080] 可以容易地确定蛋白质、多肽、约32个氨基酸、约330个氨基酸或其肽片段或对应核酸序列编码序列的全长上的氨基酸序列的百分比一致性。合适的氨基酸片段可以是至少约8个氨基酸长度，并且可以是至多约700个氨基酸。通常，当提及两个不同序列之间的“一致性”、“同源性”或“类似性”时，参考“比对”序列来确定“一致性”、“同源性”或“类似性”。“比对”序列或“比对”是指与参考序列相比，通常含有对缺失的或额外的碱基或氨基酸的校正的多个核酸序列或蛋白质(氨基酸)序列。

[0081] 一致性可以通过制备序列的比对和通过使用本领域中已知或可商购的多种算法和/或计算机程序[例如BLAST、ExPASy；Clustal0；FASTA；使用例如尼德曼-翁施算法(Needleman-Wunsch algorithm)、史密斯-沃特曼算法(Smith-Waterman algorithm)]来确定。使用多种公开或可商购的多序列比对程序中的任一种进行比对。序列比对程序可供用于氨基酸序列，例如“Clustal Omega”、“Clustal X”、“MAP”、“PIMA”、“MSA”、“BLOCKMAKER”、“MEME”以及“Match-Box”程序。通常，以默认设置使用这些程序中的任何程序，尽管本领域技术人员可以根据需要改变这些设置。或者，本领域技术人员可以利用另一种算法或计算机程序，所述算法或程序提供至少与通过参考算法和程序所提供的一样水平的一致性或比对。参见例如J.D.Thomson等人，《核酸研究(Nucl.Acids.Res.)》，“多序列比对的全面比较(A comprehensive comparison of multiple sequence alignments)”，27 (13) :2682-2690 (1999)。

[0082] 多序列比对程序也可用于核酸序列。此类程序的实例包括“Clustal Omega”、“Clustal W”、“CAP Sequence Assembly”、“BLAST”、“MAP”和“MEME”，所述程序可通过因特网上的Web服务器进行访问。此类程序的其它来源是本领域技术人员已知的。或者，也使用Vector NTI实用程序。还存在多种本领域已知的可以用于测量核苷酸序列一致性的算法，包括上述程序中包含的那些。作为另一实例，可以使用GCG 6.1版本的程序Fasta<sup>TM</sup>比较多核苷酸序列。Fasta<sup>TM</sup>提供了查询序列与搜索序列之间最佳重叠区的比对和序列一致性百分比。例如，核酸序列之间的序列一致性百分比可使用如GCG 6.1版本中所提供的采用其默认参数(字号6和评分矩阵的NPMAM系数)的Fasta<sup>TM</sup>确定，所述程序以引用的方式并入本文中。

[0083] B.rAAV载体和组合物

[0084] 在另一方面，本文提供了利用本文所述的AAV衣壳序列(包括其片段)来产生适用

于将异源基因或其它核酸序列递送至靶细胞的病毒载体的分子。在一个实施例中,适用于本文所述的组合物和方法的载体至少含有编码AAVrh.90衣壳或其片段的序列。在另一实施例中,适用的载体至少含有编码AAVrh.90rep蛋白或其片段的序列。任选地,此类载体可含有AAV cap和rep蛋白。在提供AAV rep和cap的载体中,AAV rep和AAV cap序列可以是一种血清型来源,例如都是AAVrh.90来源。或者,可以使用rep序列来自与提供cap序列的野生型AAV不同的AAV的载体。在一个实施例中,rep和cap序列由单独来源(例如单独的载体,或宿主细胞和载体)表达。在另一实施例中,这些rep序列与不同AAV血清型的cap序列同框融合以形成嵌合AAV载体,如以引用的方式并入本文中的美国专利第7,282,199号中所描述的AAV2/8。任选地,载体进一步含有小基因,所述小基因包含两侧是AAV 5'ITR和AAV 3'ITR的所选转基因。在另一实施例中,AAV是自互补AAV(sc-AAV)(参见以引用的方式并入本文中的US 2012/0141422)。自互补载体包装一个反向重复基因组,它可以折叠成dsDNA,而无需DNA合成或多个载体基因组之间的碱基配对。由于scAAV不需要在表达之前将单股DNA(ssDNA)基因组转化为双股DNA(dsDNA),因此其为更有效的载体。然而,这种效率的代价是损失了载体一半的编码能力,ScAAV适用于小蛋白质编码基因(至多约55kd)和任何当前可用的基于RNA的疗法。

[0085] 假型载体适用于本文,其中一个AAV的衣壳被异源衣壳蛋白替换。出于说明性目的,在下文所述的实例中使用了利用本文所述的AAVrh.90衣壳和AAV2 ITR的AAV载体。参见上文所引用的Mussolino等人。除非另外规定,否则本文所述的AAV ITR和其它所选AAV成分可个别地选自任何AAV血清型,包括但不限于AAV1、AAV2、AAV3、AAV4、AAV5、AAV6、AAV7、AAV8、AAV9或其它已知和未知的AAV血清型。在一个理想的实施例中,使用AAV血清型2的ITR。然而,可选择来自其它合适血清型的ITR。这些ITR或其它AAV成分可以使用本领域技术人员可用的技术从AAV血清型中容易地分离。此类AAV可以从学术、商业或公共来源(例如弗吉尼亚州马纳萨斯的美国典型培养物保藏中心(American Type Culture Collection, Manassas, VA))分离。或者,可以通过合成或其它合适的手段,通过参考公开的序列,如在文献中或在如GenBank、PubMed等数据库中可获得的公开的序列获得AAV序列。

[0086] 本文所述的rAAV还包含载体基因组。载体基因组至少由如下所述的非AAV或异源核酸序列(转基因)和其调节序列以及5'和3'AAV反向末端重复序列(ITR)组成。这种小基因被包装到衣壳蛋白中并且递送到所选靶细胞。

[0087] 转基因是与转基因两侧的载体序列异源的核酸序列,其编码所关注的多肽、蛋白质或其它产物。核酸编码序列以允许转基因在靶细胞中转录、翻译和/或表达的方式与调节成分可操作地连接。异源核酸序列(转基因)可以来源于任何生物体。AAV可以包含一种或多种转基因。

[0088] 在某些实施例中,本文提供包括转基因的rAAV.90载体,所述转基因包含编码红细胞生成素(EPO)的序列。在某些实施例中,转基因编码犬或猫EPO基因。此类重组载体适用于例如治疗受试者的慢性肾病和其它以循环红细胞的量减少为特征的病状的方案。

[0089] 在某些实施例中,本文提供包括转基因的rAAV.90载体,所述转基因包含编码抗神经生长因子(NGF)抗体的序列。在某些实施例中,转基因编码犬或猫抗NGF抗体。此类重组载体适用于例如治疗受试者的骨关节炎疼痛的方案。

[0090] 在某些实施例中,本文提供包括转基因的rAAV.90载体,所述转基因包含编码抗神

经生长因子(NGF)抗体的序列。在某些实施例中,转基因编码犬或猫抗NGF抗体。此类重组载体适用于例如治疗受试者的骨关节炎疼痛的方案。

[0091] 在某些实施例中,本文提供包括转基因的rAAV.90载体,所述转基因包含编码胰高血糖素样肽1(GLP-1)的序列。在某些实施例中,转基因编码犬或猫GLP-1。此类重组载体适用于例如治疗受试者的II型糖尿病的方案。

[0092] 在某些实施例中,本文提供包括转基因的rAAV.90载体,所述转基因包含编码胰高血糖素样肽1(GLP-1)的序列。在某些实施例中,转基因编码犬或猫GLP-1。此类重组载体适用于例如治疗受试者的II型糖尿病的方案。

[0093] 在某些实施例中,本文提供包括转基因的rAAV.90载体,所述转基因包含编码胰岛素的序列。在某些实施例中,转基因编码犬或猫胰岛素。此类重组载体适用于例如治疗受试者的I型糖尿病或II型糖尿病的方案。

[0094] 在某些实施例中,本文提供包括转基因的rAAVrh.90载体,所述转基因包含编码IgE、IL-32或IL-4/IL-13受体(包括例如抗体和受体-IgG融合蛋白)的白介素-4受体 $\alpha$ (IL-4Ra)亚单位的拮抗剂的序列。在某些实施例中,转基因编码犬或猫IgE、IL-32或IL-4Ra亚单位的拮抗剂。此类重组载体适用于例如治疗受试者的异位性皮肤炎的方案。

[0095] 转基因序列的组成将取决于所得载体的用途。例如,一种类型的转基因序列包括报告序列,其在表达时产生可检测信号。此类报告序列包括但不限于编码以下者的DNA序列: $\beta$ -内酰胺酶、 $\beta$ -半乳糖苷酶(LacZ)、碱性磷酸酶、胸苷激酶、绿色荧光蛋白(GFP)、增强型GFP(EGFP)、氯霉素乙酰转移酶(CAT)、荧光素酶、膜结合蛋白(包括例如CD2、CD4、CD8)、流感血凝素蛋白和本领域熟知的其它蛋白,针对其存在或可通过常规方法产生高亲和力抗体,以及融合蛋白,包含适当地与尤其来自血凝素或Myc的抗原标签域融合的膜结合蛋白。

[0096] 这些编码序列当与驱动其表达的调节元件相关联时提供可通过常规方法检测的信号,所述常规方法包括酶促、放射照相、比色、荧光或其它光谱分析、荧光激活细胞分选分析和免疫分析,包括酶联免疫吸附分析(ELISA)、放射免疫分析(RIA)和免疫组织化学。例如,当标记序列为LacZ基因时,通过针对 $\beta$ -半乳糖苷酶活性的分析来检测携带信号的载体的存在。当转基因为绿色荧光蛋白质或荧光素酶时,携带信号的载体可以通过光度计中的颜色或光产生在视觉上测量。

[0097] 但是,理想地,转基因是编码适用于生物学和医学的产物,如蛋白质、肽、RNA、酶、显性失活突变体或催化RNA的非标记序列。理想的RNA分子包括tRNA、dsRNA、核糖体RNA、催化RNA、siRNA、小发夹RNA、反式剪接RNA和反义RNA。适用的RNA序列的一个实例是抑制或消除经治疗动物中的靶核酸序列表达的序列。通常,合适的靶序列包括肿瘤靶标和病毒性疾病。关于此类靶标的实例,参见下文与免疫原相关的部分中鉴别的肿瘤靶标和病毒。

[0098] 转基因可用于纠正或改善基因缺陷,所述基因缺陷可包括正常基因以低于正常水平表达的缺陷或功能基因产物不表达的缺陷。或者,转基因可以向细胞提供不天然表达于所述细胞类型或宿主中的产物。优选类型的转基因序列编码在宿主细胞中表达的治疗性蛋白质或多肽。本发明进一步包括使用多种转基因。在某些情形下,不同转基因可用于编码蛋白质的每个亚单位,或编码不同肽或蛋白质。当编码蛋白质亚单位的DNA的大小较大时,例如对于免疫球蛋白、血小板衍生生长因子或肌营养不良蛋白,这是理想的。为了使细胞产生多亚单位蛋白质,细胞用含有不同亚单位中的每一个的重组病毒感染。或者,蛋白质的不同

亚单位可以由相同转基因编码。在此情况下，单一转基因包括编码亚单位中的每一个的DNA，其中每个亚单位的DNA由内部核酶进入位点(IRES)分离。当编码亚单位中的每一个的DNA的大小较小，例如编码亚单位的DNA和IRES的总大小为小于五千碱基时，这是理想的。作为IRES的替代方案，DNA可由编码2A肽的序列分离，所述肽在翻译后事件中自裂解。参见例如M.L.Donnelly等人，《普通病毒学杂志(J.Gen.Viro.)》，78(Pt 1):13-21(1997年1月)；Furler,S.等人，《基因疗法》，8(11):864-873(2001年6月)；Klump H.等人，《基因疗法》，8(10):811-817(2001年5月)。这种2A肽显著小于IRES，因此非常适合在空间为限制因素时使用。更常见的是，当转基因较大，由多个亚单位组成，或两个转基因共同递送时，携带所需转基因或亚单位的rAAV被共同施用以允许其在体内连环化以形成单个载体基因组。在此类实施例中，第一AAV可以携带表达单一转基因的表达盒，并且第二AAV可以携带表达不同转基因以在宿主细胞中共表达的表达盒。然而，所选转基因可以编码任何生物活性产物或其它产物，例如研究所需的产物。

[0099] 合适的转基因或基因产物的实例包括与家族性高胆固醇血症、肌营养不良、囊性纤维化以及罕见病或孤儿病相关的那些。此类罕见疾病的实例可包括脊髓性肌萎缩(SMA)、亨廷顿氏病(Huntington's Disease)、瑞特综合征(Rett Syndrome)(例如，甲基-CpG结合蛋白2(MeCP2)；UniProtKB-P51608)、肌肉萎缩性侧索硬化(ALS)、杜氏型肌肉营养不良(Duchenne Type Muscular dystrophy)、弗雷德里希共济失调(Friedrichs Ataxia)(例如共济蛋白)、与2型脊髓小脑共济失调(SCA2)/ALS相关的ATXN2；与ALS、颗粒蛋白前体(PRGN)(与非阿尔茨海默氏病脑变性相关，包括额颞叶痴呆(FTD)、进行性非流畅性失语症(PNFA)和语义性痴呆)相关的TDP-43等。参见例如www.orpha.net/consor/cgi-bin/Disease\_Search\_List.php;rarediseases.info.nih.gov/diseases。

[0100] 由转基因编码的适用的治疗产物包括激素以及生长和分化因子，包括但不限于胰岛素、胰高血糖素、胰高血糖素样肽1(GLP-1)、生长激素(GH)、甲状旁腺激素(PTH)、生长激素释放因子(GRF)、促卵泡激素(FSH)、促黄体激素(LH)、人绒毛膜促性腺激素(hCG)、血管内皮生长因子(VEGF)、血管生成素、血管生长抑素、粒细胞集落刺激因子(GCSF)、红细胞生成素(EPO)、结缔组织生长因子(CTGF)碱性成纤维细胞生长因子(bFGF)、酸性成纤维细胞生长因子(aFGF)、表皮生长因子(EGF)、转型生长因子 $\alpha$ (TGF $\alpha$ )、血小板衍生生长因子(PDGF)、胰岛素生长因子I和II(IGF-I和IGF-II)、转型生长因子 $\beta$ 超家族中的任一种(包括TGF $\beta$ )、激活素、抑制素、或骨形态发生蛋白(BMP)BMP 1-15中的任一种、生长因子的调蛋白/神经调节蛋白/ARIA/neu分化因子(NDF)家族中的任一种、神经生长因子(NGF)、脑源性神经营养因子(BDNF)、神经营养素NT-3和NT-4/5、睫状神经营养因子(CNTF)、神经胶质细胞系衍生神经营养因子(GDNF)、神经营养因子、集聚蛋白、信号素(semaphorin)/脑衰蛋白(collapsin)家族中的任一种、轴突引导因子-1和轴突引导因子-2、肝细胞生长因子(HGF)、肝配蛋白(ephrin)、头蛋白(noggin)、音猬因子和酪氨酸羟化酶。

[0101] 其它适用的转基因产物包括调节免疫系统的蛋白质，包括但不限于细胞因子和淋巴因子，如血小板生成素(TPO)、白介素(IL)IL-1至IL-25(包括IL-2、IL-4、IL-12和IL-18)、单核细胞趋化蛋白、白血病抑制因子、粒细胞-巨噬细胞集落刺激因子、Fas配体、肿瘤坏死因子 $\alpha$ 和 $\beta$ 、干扰素( $\alpha$ 、 $\beta$ 和 $\gamma$ )、干细胞因子、f1k-2/f1t3配体。由免疫系统产生的基因产物也可用于本发明。这些包括但不限于免疫球蛋白IgG、IgM、IgA、IgD和IgE、嵌合免疫球蛋白、人

源化抗体、单链抗体、T细胞受体、嵌合T细胞受体、单链T细胞受体、MHC I类和II类分子以及经过工程化的免疫球蛋白和MHC分子。适用的基因产物还包括补体调节蛋白,如补体调节蛋白、膜辅因子蛋白(MCP)、衰变加速因子(DAF)、CR1、CF2和CD59。

[0102] 其它适用的基因产物包括用于激素、生长因子、细胞因子、淋巴因子、调节蛋白和免疫系统蛋白的受体中的任一种。本发明涵盖用于胆固醇调节的受体,包括低密度脂蛋白(LDL)受体、高密度脂蛋白(HDL)受体、极低密度脂蛋白(VLDL)受体和清道夫受体。本发明还涵盖基因产物,如类固醇激素受体超家族的成员,包括糖皮质激素受体和雌激素受体、维生素D受体和其它核受体。另外,适用的基因产物包括转录因子,如jun、fos、max、mad、血清应答因子(SRF)、AP-1、AP2、myb、MyoD和肌生成素、含ETS盒的蛋白质、TFE3、E2F、ATF1、ATF2、ATF3、ATF4、ZF5、NFAT、CREB、HNF-4、C/EBP、SP1、CCAAT盒结合蛋白、干扰素调节因子(IRF-1)、威尔姆斯肿瘤蛋白、ETS结合蛋白、STAT、GATA盒结合蛋白(例如,GATA-3)和带翼螺旋蛋白的叉头家族。

[0103] 其它适用的基因产物包括氨甲酰基合成酶I、鸟氨酸转氨甲酰酶、精氨酸琥珀酸合成酶、精氨酸琥珀酸裂解酶、精氨酸酶、富马酸乙酰乙酸水解酶、苯丙氨酸羟化酶、 $\alpha$ -1抗胰蛋白酶、葡萄糖-6-磷酸酶、胆色素原脱氨酶、胱硫醚 $\beta$ -合成酶、支链酮酸脱羧酶、白蛋白、异戊酰CoA脱氢酶、丙酰CoA羧化酶、甲基丙二酰CoA变位酶、戊二酰CoA脱氢酶、胰岛素、 $\beta$ -葡萄糖苷酶、丙酮酸羧化酶、肝磷酸化酶、磷酸化酶激酶、甘氨酸脱羧酶、H蛋白、T蛋白、囊性纤维化跨膜调节(CFTR)序列和肌营养不良蛋白序列或其功能片段。其它适用的基因产物包括如可适用于酶替代疗法的酶,所述酶替代疗法适用于由于酶活性不足而导致的多种病状。例如,可以将含有甘露糖-6-磷酸的酶用于溶酶体贮积病的疗法中(例如,合适的基因包括对 $\beta$ -葡萄糖醛酸酶(GUSB)进行编码的基因)。在另一实例中,基因产物为泛素蛋白质连接酶E3A(UBE3A)。仍适用的基因产物包括UDP葡萄糖醛酸基转移酶家族1成员A1(UGT1A1)。

[0104] 其它适用的基因产物包括非天然存在的多肽,如具有含有插入、缺失或氨基酸取代的非天然存在的氨基酸序列的嵌合或杂合多肽。例如,单链工程化的免疫球蛋白可能适用于某些免疫功能不全患者。其它类型的非天然存在的基因序列包括反义分子和催化核酸,如核酶,其可以用于减少靶标的过度表达。

[0105] 减少和/或调节基因表达对于治疗以过度增殖细胞为特征的过度增殖性病状(如癌症和牛皮癣)是特别期望的。靶多肽包括与正常细胞相比在过度增殖性细胞中专门产生或以更高水平产生的那些多肽。靶抗原包括由癌基因如myb、myc、fyn和易位基因bcr/abl、ras、src、P53、neu、trk和EGFR编码的多肽。除作为靶抗原的癌基因产物之外,用于抗癌治疗和保护方案的靶多肽包括由B细胞淋巴瘤产生的抗体的可变区和T细胞淋巴瘤的T细胞受体的可变区,在一些实施例中,所述可变区还被用作自身免疫疾病的靶抗原。其它肿瘤相关多肽可用作靶多肽,如在肿瘤细胞中以较高水平存在的多肽,包括由单克隆抗体17-1A识别的多肽和叶酸结合多肽。

[0106] 其它合适的治疗性多肽和蛋白质包括可适用于通过为针对与自身免疫相关的靶标赋予广泛基础的保护性免疫反应而治疗患有自身免疫性疾病和病症的个体的多肽和蛋白质,所述靶标包括细胞受体和产生自身定向抗体的细胞。T细胞介导的自身免疫疾病包括类风湿关节炎(RA)、多发性硬化症(MS)、休格连氏综合征(Sjögren's syndrome)、类肉瘤病、胰岛素依赖型糖尿病(IDDM)、自身免疫甲状腺炎、反应性关节炎、强直性脊柱炎、硬皮病、多

发性肌炎、皮肌炎、牛皮癣、韦格纳氏肉芽肿病 (Wegner's granulomatosis)、克罗恩氏病 (Crohn's disease) 和溃疡性结肠炎。这些疾病中的每一种以与内源性抗原结合并引发与自身免疫疾病相关的发炎级联的T细胞受体 (TCR) 为特征。

[0107] 其它适用的基因产物包括用于治疗血友病的基因产物，所述血友病包括血友病B (包括因子IX) 和血友病A (包括因子VIII和其变体，如异二聚体和B缺失结构域的轻链和重链；美国专利第6,200,560号和美国专利第6,221,349号)。在一些实施例中，小基因包含因子VIII重链的前57个碱基对，所述重链对10个氨基酸信号序列以及人生长激素 (hGH) 聚腺苷酸化序列进行编码。在替代性实施例中，小基因进一步包含A1和A2域，以及来自B域的N端的5个氨基酸和/或B域的C端的85个氨基酸，以及A3、C1和C2域。在其它实施例中，编码因子VIII重链和轻链的核酸在由42个编码B结构域的14个氨基酸的核酸分隔的单个小基因中提供 [美国专利第6,200,560号]。

[0108] 可以通过rAAV递送的其它示例性基因包括但不限于与糖原贮积病或1A型缺陷相关的葡萄糖-6-磷酸酶 (GSD1)、与PEPCK缺陷相关的磷酸烯醇丙酮酸羧激酶 (PEPCK)；与癫痫发作和严重的神经发育障碍相关的细胞周期蛋白依赖性激酶样5 (CDKL5)，也称为丝氨酸/苏氨酸激酶9 (STK9)；与半乳糖血症相关的半乳糖-1磷酸尿苷转移酶；与苯丙酮尿症 (PKU) 相关的苯丙氨酸羟化酶 (PAH)；与原发性高草酸尿症1型相关的基因产物，包括羟基酸氧化酶1 (GO/HAO1) 和AGXT，与枫糖浆尿病相关的支链 $\alpha$ -酮酸脱氢酶，包括BCKDH、BCKDH-E2、BAKDH-E1a和BAKDH-E1b；与1型酪氨酸血症相关的延胡索乙酰乙酸水解酶；与甲基丙二酸血症相关的甲基丙二酰辅酶A变位酶；与中链乙酰辅酶A缺乏相关的中链酰基辅酶A脱氢酶；与鸟氨酸转氨甲酰酶缺乏相关的鸟氨酸转氨甲酰酶 (OTC)；与瓜氨酸血症相关的精氨酸琥珀酸合成酶 (ASS1)；卵磷脂-胆固醇酰基转移酶 (LCAT) 缺乏症；无甲基丙二酸血症 (MMA)；与尼曼匹克病 (Niemann-Pick disease)，C1型相关的NPC1；丙酸血症 (PA)；与转甲状腺素蛋白 (TTR) 相关的遗传性淀粉样变性相关的TTR；与家族性高胆固醇血症 (FH) 相关的低密度脂蛋白受体 (LDLR) 蛋白，LDLR变体，例如WO 2015/164778中描述的那些；PCSK9；与痴呆症相关的ApoE和ApoC蛋白；与克里格勒-纳贾尔病 (Crigler-Najjar disease) 相关的UDP-葡萄糖醛酸转移酶；与严重的联合免疫缺陷病相关的腺苷脱氨酶；与痛风和勒什-纳阳综合征 (Lesch-Nyhan syndrome) 相关的次黄嘌呤鸟嘌呤磷酸核糖转移酶；与生物素酶缺乏相关的生物素酶；与法布里病 (Fabry disease) 相关的 $\alpha$ -半乳糖苷酶A ( $\alpha$ -Gal A)；与GM1神经节苷脂病相关的 $\beta$ -半乳糖苷酶 (GLB1)；与威尔逊病 (Wilson's Disease) 相关的ATP7B；与戈谢病 (Gaucher disease) 2型和3型相关的 $\beta$ -葡萄糖脑苷脂酶；与齐薇格综合征 (Zellweger syndrome) 相关的过氧化物酶体膜蛋白70kDa；与异染性脑白质营养不良相关的芳基硫酸酯酶A (ARSA)、与克拉培病 (Krabbe disease) 相关的半乳糖脑苷脂酶 (GALC)、与庞贝病 (Pompe disease) 相关的 $\alpha$ -葡萄糖苷酶 (GAA)；与尼曼匹克病A型相关的鞘磷脂酶 (SMPD1) 基因；与成人发病的II型瓜氨酸血症 (CTLN2) 相关的精氨酸琥珀酸合酶；与尿素循环障碍相关的氨基甲酰磷酸合酶1 (CPS1)；与脊髓性肌萎缩症相关的存活运动神经元 (SMN) 蛋白；与法伯脂肪肉芽肿 (Farber lipogranulomatosis) 相关的神经酰胺酶；与GM2神经节苷脂病和泰-萨病 (Tay-Sachs disease) 和桑德霍夫病 (Sandhoff disease) 相关的 $\beta$ -氨基己糖苷酶；与天冬氨酰葡萄糖胺相关的天冬氨酰氨基葡萄糖苷酶；与岩藻糖苷中毒相关的 $\alpha$ -岩藻糖苷酶；与 $\alpha$ -甘露糖苷中毒相关的 $\alpha$ -甘露糖苷酶；与急性间歇性卟啉症 (AIP) 相关的胆色素原脱氨

酶，；用于治疗 $\alpha$ -1抗胰蛋白酶缺乏症(肺气肿)的 $\alpha$ -1抗胰蛋白酶；用于治疗地中海贫血或肾功能衰竭引起的贫血的红细胞生成素；用于治疗缺血性疾病的血管内皮生长因子、血管生成素-1和成纤维细胞生长因子；血栓调节蛋白和组织因子通路抑制剂，用于治疗闭塞的血管，例如动脉粥样硬化、血栓形成或栓塞中所见；用于治疗帕金森病(Parkinson's disease)的芳香族氨基酸脱羧酶(AADC)和酪氨酸羟化酶(TH)；用于治疗充血性心力衰竭的 $\beta$ 肾上腺素能受体、磷蛋白、肌(内)质网腺苷三磷酸酶-2(SERCA2)和心脏腺苷酸环化酶的反义或突变形式；用于治疗各种癌症的肿瘤抑制基因，如p53；细胞因子，例如用于治疗炎症和免疫疾病和癌症的各种白介素中的一种；用于治疗肌营养不良的肌营养不良蛋白或小肌营养不良蛋白和用于治疗肌营养不良的肌营养相关蛋白或小肌营养相关蛋白；以及用于治疗糖尿病的胰岛素或GLP-1。

[0109] 替代地或另外，本发明的载体可含有本发明的AAV序列和转基因，所述转基因编码诱导针对所选免疫原的免疫反应的肽、多肽或蛋白质。例如，免疫原可以选自多种病毒科。需要针对其产生免疫反应的理想病毒科的实例包括小核糖核酸病毒科，其包括鼻病毒属，其造成大约50%的普通感冒病例；肠道病毒属，包括脊髓灰质炎病毒、柯萨奇病毒、埃可病毒和人类肠道病毒，如甲型肝炎病毒；和主要在非人类动物中引起口蹄疫的口蹄疫病毒属。在病毒的小核糖核酸病毒家族内，靶抗原包括VP1、VP2、VP3、VP4和VPG。另一病毒家族包括杯状病毒家族，所述家族涵盖诺瓦克(Norwalk)病毒群，其为流行性胃肠炎的重要病原体。另一个期望用于靶向抗原以在人类和非人类动物中诱导免疫反应的病毒家族是披膜病毒科，其包括甲病毒属，其包括辛德毕斯病毒、罗斯河病毒和委内瑞拉、东方和西方马脑炎，以及风疹病毒属，包括风疹病毒。黄病毒科包括登革热、黄热病、日本脑炎、圣路易斯脑炎和蜱传脑炎病毒。其它靶抗原可能从丙型肝炎或冠状病毒科产生，包括非人类病毒，例如传染性支气管炎病毒(家禽)、猪传染性胃肠病毒(猪)、猪血凝性脑脊髓炎病毒(猪)、猫传染性腹膜炎病毒(猫)、猫肠道冠状病毒(猫)、犬冠状病毒(狗)和人类呼吸道冠状病毒，其可能导致普通感冒和/或非甲、乙或丙型肝炎。在冠状病毒科中，靶抗原包括E1(也称为M或基质蛋白)、E2(也称为S或纤突蛋白)、E3(也称为HE或血凝素-依尔替糖(elterose))糖蛋白(并非在所有冠状病毒中都存在)或N(核衣壳)。其它抗原可以靶向弹状病毒科，其包括水疱病毒属(例如水泡性口炎病毒)和一般狂犬病毒属(例如狂犬病)。在弹状病毒科中，合适的抗原可以来源于G蛋白或N蛋白。包括马堡病毒和埃博拉病毒等出血热病毒的丝状病毒科可能是合适的抗原来源。副粘病毒科包括副流感病毒1型、副流感病毒3型、牛副流感病毒3型、腮腺炎病毒属(腮腺炎病毒)、副流感病毒2型、副流感病毒4型、新城疫病毒(鸡)、牛瘟、麻疹病毒(其包括麻疹和犬瘟热)和肺病毒属，其包括呼吸道合胞病毒。流感病毒分类在正粘病毒科内，并且是合适的抗原来源(例如HA蛋白、N1蛋白)。布尼亚病毒科包括布尼亚病毒属(加利福尼亚脑炎、拉克罗斯(La Crosse)、白蛉病毒属(裂谷热)、汉坦病毒(普马拉病毒(puremala)是一种出血热病毒)、内罗病毒(内罗毕绵羊病)和各种未指定的布尼亚病毒。沙粒病毒科提供了抗LCM和拉沙热病毒的抗原来源。呼肠孤病毒家族包括呼肠孤病毒属、轮状病毒(其可引起儿童急性肠胃炎)、轮状病毒和科罗拉多蜱传热病毒属(cultivirus)(科罗拉多蜱传热、莱邦博病(Lebombo)(人类)、马脑病、蓝舌病)。

[0110] 逆转录病毒科包括肿瘤病毒(oncorivinal)亚科，其涵盖人类和兽医疾病，如猫白血病病毒、HTLV-I和HTLV-II、慢病毒(包括人类免疫缺陷病毒(HIV)、猿猴免疫缺陷病毒

(SIV)、猫免疫缺陷病毒(FIV)、马传染性贫血病毒和泡沫病毒)。在HIV与SIV之间,已描述许多合适的抗原并且可以容易地选择。合适的HIV和SIV抗原的实例包括但不限于gag、pol、Vif、Vpx、VPR、Env、Tat和Rev蛋白以及其各种片段。另外,已描述对这些抗原的多种修饰。出于此目的的合适抗原是本领域的技术人员已知的。例如,可选择编码gag、pol、Vif和Vpr、Env、Tat和Rev的序列,以及其他蛋白质。参见例如美国专利5,972,596中所描述的经修饰的gag蛋白。另外参见D.H.Barouch等人,《病毒学杂志》,75(5):2462-2467(2001年3月)和R.R.Amara等人,《科学》,292:69-74(2001年4月6日)中描述的HIV和SIV蛋白。这些蛋白质或其亚单位可单独递送,或经由单独载体或从单一载体组合递送。

[0111] 乳多泡病毒科包括亚科多瘤病毒(BKU和JCU病毒)和亚科乳头状瘤病毒(与癌症或乳头状瘤的恶性进展相关)。腺病毒科包括引起呼吸道疾病和/或肠炎的病毒(EX、AD7、ARD、O.B.)。细小病毒家科猫细小病毒(猫肠炎)、猫泛白细胞减少症病毒、犬细小病毒和猪细小病毒。疱疹病毒科包括亚科 $\alpha$ 疱疹病毒,其涵盖单纯疱疹病毒属(HSVI、HSVII)、水痘病毒属(假性狂犬病、水痘带状疱疹);和亚科 $\beta$ 疱疹病毒,其包括巨细胞病毒属(HCMV、鼠巨细胞病毒属);和亚科 $\gamma$ 疱疹病毒亚科,其包括淋巴隐病毒属、EBV(伯基特氏淋巴瘤)、传染性鼻气管炎、马立克氏病(Marek's disease)病毒和细长病毒属(rhadinovirus)。痘病毒科包括脊痘病毒亚科,其涵盖正痘病毒属(天花(Variola/Smallpox)和牛痘(Vaccinia/Cowpox))、副痘病毒、禽痘病毒、山羊痘病毒、麻痘病毒、水痘病毒和昆虫痘病毒亚科。肝炎病毒科包括乙型肝炎病毒。一种可能是合适的抗原来源的未分类病毒是丁型肝炎病毒。其它病毒来源可包括禽传染性法氏囊病病毒和猪呼吸和生殖综合征病毒。甲病毒科包括马动脉炎病毒和各种脑炎病毒。

[0112] 本发明还可以涵盖适用于使人类或非人类动物免疫其它病原体的免疫原,所述病原体包括感染人类和非人类脊椎动物的细菌、真菌、寄生微生物或多细胞寄生虫,或来自癌细胞或肿瘤细胞。细菌病原体的实例包括致病性革兰氏阳性球菌,包括肺炎球菌;葡萄球菌;和链球菌。致病性革兰氏阴性球菌包括脑膜炎球菌;淋球菌。致病性肠道革兰氏阴性杆菌包括肠杆菌科;假单胞菌、不动杆菌和艾肯氏菌;类鼻疽;沙门氏菌;志贺氏菌;嗜血杆菌;莫拉氏菌;杜克雷嗜血杆菌(*H. ducreyi*) (导致软下疳);布鲁氏菌;土拉弗朗西斯菌(*Franisella tularensis*) (导致土拉菌病);耶尔森氏菌(巴氏杆菌);念珠状链杆菌和螺旋菌;革兰氏阳性杆菌包括单核细胞增生李斯特菌;猪红斑丹毒丝菌;白喉棒状杆菌(白喉);霍乱;炭疽芽孢杆菌(*B. anthracis*) (炭疽);杜诺凡病(donovanosis) (腹股沟肉芽肿);和巴尔通体病(bartonellosis)。由致病性厌氧菌引起的疾病包括破伤风;肉毒杆菌;其它梭菌;结核;麻风;和其它分枝杆菌。致病性螺旋体病包括梅毒;密螺旋体病:雅司病(yaws)、斑点病和地方性梅毒;以及钩端螺旋体病。其它由高等致病性细菌和致病性真菌引起的感染包括放线菌病;诺卡氏菌病;隐球菌病、芽生菌病、组织胞浆菌病和球孢子菌病;念珠菌病、曲霉病和毛霉菌病;孢子丝菌病;副球孢子菌病、球孢子菌病、圆环孢子菌病、足菌肿和染色体病;和皮肤癣菌病。立克次体感染包括斑疹伤寒、落基山斑疹热、Q热和立克次体痘。支原体和衣原体感染的实例包括:肺炎支原体;性病淋巴肉芽肿;鹦鹉热;和围产期衣原体感染。致病性真核生物涵盖致病性原生动物和蠕虫,且由此产生的感染包括:阿米巴病;疟疾;利什曼病;锥虫病;弓形体病;卡氏肺孢子虫(*Pneumocystis carinii*);特里坎斯(*Trichans*);刚地弓形虫(*Toxoplasma gondii*);巴贝斯虫病(babesiosis);贾第鞭毛虫病(giardiasis);旋毛

虫病；丝虫病；血吸虫病；线虫；吸虫(trematodes)或吸虫(flukes)；和绦虫(cestode/tapeworm)感染。

[0113] 许多这些生物体和/或由此产生的毒素已被疾病控制中心[(CDC),美国卫生和公共服务部]确定为具有用于生物攻击的潜力的病原体。例如,这些生物病原体中的一些,包括炭疽芽孢杆菌(*Bacillus anthracis*) (炭疽)、肉毒杆菌(*Clostridium botulinum*) 和其毒素(肉毒中毒)、鼠疫耶尔森菌(*Yersinia pestis*) (鼠疫)、大天花(天花)、土拉热弗朗西斯氏菌(*Francisella tularensis*) (土拉菌病) 和病毒性出血热,所有这些目前都归类为A类病原体;立克次体(*Coxiella burnetti*) (Q热);布鲁氏菌属(布鲁氏菌病)、鼻疽伯克霍尔德氏菌(*Burkholderia mallei*) (鼻疽病)、蓖麻(*Ricinus communis*) 和其毒素(蓖麻毒素)、产气荚膜芽胞梭菌(*Clostridium perfringens*) 和其毒素(ε毒素)、葡萄球菌属和其毒素(肠毒素B),所有这些目前都归类为B类病原体;以及尼潘病毒(*Nipan virus*) 和汉坦病毒(hantavirus),目前被归类为C类病原体。此外,被如此分类或不同分类的其它生物体可以在未来被识别和/或用于这样的目的。将容易理解的是,本文所述的病毒载体和其它构建体可用于从这些生物体、病毒、其毒素或其它副产物递送抗原,这将预防和/或治疗感染或与这些生物制剂有关的其它不良反应。

[0114] 施用本发明的载体以递送针对T细胞可变区的免疫原引发包括CTL在内的免疫反应以消除那些T细胞。在类风湿性关节炎(RA)中,已经表征了与所述疾病相关的T细胞受体(TCR)的若干特定可变区。这些TCR包括V-3、V-14、V-17和Va-17。因此,编码这些多肽中的至少一种的核酸序列的递送将引发免疫反应,所述免疫反应将靶向涉及RA的T细胞。在多发性硬化症(MS)中,已表征涉及所述疾病的TCR的若干特定可变区。这些TCR包括V-7和Va-10。因此,编码这些多肽中的至少一种的核酸序列的递送将引发免疫反应,所述免疫反应将靶向涉及MS的T细胞。在硬皮病中,已表征涉及所述疾病的TCR的若干特定可变区。这些TCR包括V-6、V-8、V-14和Va-16、Va-3C、Va-7、Va-14、Va-15、Va-16、Va-28和Va-12。因此,编码这些多肽中的至少一种的核酸分子的递送将引发免疫反应,所述免疫反应将靶向涉及硬皮病的T细胞。

[0115] 在一个实施例中,选择转基因以提供光遗传疗法。在光遗传疗法中,人工光感受器是通过将光激活通道或泵基因递送到剩余视网膜回路中的存活细胞类型来构建的。这特别适用于已经失去大量光感受器功能,但其双极细胞回路至神经节细胞和视神经保持完整的患者。在一个实施例中,异源核酸序列(转基因)为视蛋白。视蛋白序列可以来源于任何合适的单细胞或多细胞生物,包括人类、藻类和细菌。在一个实施例中,视蛋白为视紫红质、光视蛋白、L/M波长(红色/绿色)-视蛋白或短波长(S)视蛋白(蓝色)。在另一实施例中,视蛋白为通道视紫红质或盐细菌视紫红质。

[0116] 在另一实施例中,选择转基因用于基因增强疗法,即提供缺失或缺陷基因的替代拷贝。在此实施例中,本领域技术人员可以容易地选择转基因以提供必要的替代基因。在一个实施例中,缺失/缺陷的基因与眼部病症有关。在另一实施例中,转基因为NYX、GRM6、TRPML或GPR179,并且眼部病症为先天性静止性夜盲症。参见例如Zeitz等人,《美国人类遗传学杂志(Am J Hum Genet.)》2013年1月10日;92(1):67-75.电子版2012年12月13日,其以引用的方式并入本文中。在另一实施例中,转基因为RPGR。

[0117] 在另一实施例中,选择转基因用于基因抑制疗法,即,一种或多种天然基因的表达

在转录或翻译水平下被中断或抑制。这可以使用短发夹RNA(shRNA)或所属领域中众所周知的其它技术实现。参见例如Sun等人,《国际癌症杂志(Int J Cancer.)》2010年2月1日;126(3):764-74和O'Reilly M等人《美国人类遗传学杂志》2007年7月;81(1):127-35,其以引用的方式并入本文中。在此实施例中,本领域的技术人员可以基于需要沉默的基因容易地选择转基因。

[0118] 在另一实施例中,转基因包含多于一个转基因。这可以使用携带两个或更多个异源序列的单一载体,或使用各自携带一个或多个异源序列的两个或更多个AAV来实现。在一个实施例中,AAV用于基因抑制(或敲低)和基因增强协同疗法。在敲低/增强协同疗法中,所关注的基因的缺陷拷贝被沉默并提供非突变拷贝。在一个实施例中,这使用两个或更多个共同施用的载体实现。参见Millington-Ward等人,《分子疗法(Molecular Therapy)》,2011年4月,19(4):642-649,其以引用的方式并入本文中。转基因可由本领域的技术人员基于所需结果容易地选择。

[0119] 在另一实施例中,选择转基因用于基因校正疗法。这可以使用例如锌指核酸酶(ZFN)诱导的DNA双股断裂结合外源DNA供体底物来实现。参见例如Ellis等人,《基因疗法》(电子版2012年1月)20:35-42,其以引用的方式并入本文中。转基因可由本领域的技术人员基于所需结果容易地选择。

[0120] 在一个实施例中,本文所述的衣壳可用于美国临时专利申请号61/153,470、62/183,825、62/254,225和62/287,511中描述的CRISPR-Cas双载体系统,这些申请各自以引用的方式并入本文中。衣壳还可用于递送归巢核酸内切酶或其它大范围核酸酶。

[0121] 在另一实施例中,适用于本文的转基因包括报告序列,其在表达时产生可检测信号。此类报告序列包括但不限于编码以下者的DNA序列: $\beta$ -内酰胺酶、 $\beta$ -半乳糖苷酶(LacZ)、碱性磷酸酶、胸苷激酶、绿色荧光蛋白(GFP)、红色荧光蛋白(RFP)、氯霉素乙酰转移酶(CAT)、荧光素酶、膜结合蛋白(包括例如CD2、CD4、CD8)、流感血凝素蛋白和本领域熟知的其它蛋白,针对其存在或可通过常规方法产生高亲和力抗体,以及融合蛋白,包含适当地与尤其来自血凝素或Myc的抗原标签域融合的膜结合蛋白。

[0122] 这些编码序列当与驱动其表达的调节元件相关联时提供可通过常规方法检测的信号,所述常规方法包括酶促、放射照相、比色、荧光或其它光谱分析、荧光激活细胞分选分析和免疫分析,包括酶联免疫吸附分析(ELISA)、放射免疫分析(RIA)和免疫组织化学。例如,当标记序列为LacZ基因时,通过针对 $\beta$ -半乳糖苷酶活性的分析来检测携带信号的载体的存在。当转基因为绿色荧光蛋白质或荧光素酶时,携带信号的载体可以通过光度计中的颜色或光产生在视觉上测量。

[0123] 理想地,转基因编码适用于生物学和医学的产品,如蛋白质、肽、RNA、酶或催化RNA。理想的RNA分子包括shRNA、tRNA、dsRNA、核糖体RNA、催化RNA和反义RNA。适用的RNA序列的一个实例是消除经治疗动物中的靶核酸序列表达的序列。

[0124] 调节序列包括常规控制元件,所述控制元件以允许转基因在用载体转染或用如本文所述产生的病毒感染的细胞中转录、翻译和/或表达的方式与转基因可操作地连接。如本文所用,“可操作地连接的”序列包括与所关注的基因邻接的表达控制序列和以反式或相隔一定距离起作用以控制所关注的基因的表达控制序列。

[0125] 表达控制序列包括适当的转录起始、终止、启动子和增强子序列;有效的RNA加工

信号,例如剪接和聚腺苷酸化(polyA)信号;稳定细胞质mRNA的序列;增强翻译效率的序列(即,Kozak共有序列);增强蛋白质稳定性的序列;以及当需要时,增强所编码产物的分泌的序列。大量表达控制序列(包括启动子)在本领域中已知并且可以被利用。

[0126] 可用于本文提供的构建体中的调节序列还可含有内含子,理想地位于启动子/增强子序列与基因之间。一种理想的内含子序列源自SV-40,并且是100bp的微型内含子剪接供体/剪接受体,称为SD-SA。另一合适的序列包括土拔鼠肝炎病毒转录后元件。(参见例如L.Wang和I.Verma,1999《美国国家科学院院刊(Proc.Natl.Acad.Sci.,USA)》,96:3906-3910)。polyA信号可以源自许多合适的物种,包括但不限于SV-40、人类和牛。

[0127] 适用于本文所述的方法的rAAV的另一调节成分为内部核糖体进入位点(IRES)。IRES序列或其它合适的系统可用于从单一基因转录物产生多于一种多肽。IRES(或其它合适的序列)用于产生含有多于一条多肽链的蛋白质或从相同细胞或在相同细胞内表达两种不同蛋白质。示例性IRES是脊髓灰质炎病毒内部核糖体进入序列,其支持光感受器、RPE和神经节细胞中的转基因表达。优选地,IRES位于rAAV载体的转基因的3'。

[0128] 在一个实施例中,AAV包含启动子(或启动子的功能片段)。用于rAAV中的启动子可选自多种可在所需靶细胞中表达所选转基因的组成性或诱导性启动子。在一个实施例中,靶细胞是眼细胞。启动子可以来源于任何物种,包括人类。理想地,在一个实施例中,启动子具有“细胞特异性”。术语“细胞特异性”意指为重组载体选择的特定启动子可以引导所选转基因在特定细胞组织中的表达。在一个实施例中,启动子对转基因在肌肉细胞中的表达具有特异性。在另一实施例中,启动子对肺中的表达具有特异性。在另一实施例中,启动子对转基因在肝细胞中的表达具有特异性。在另一实施例中,启动子对转基因在气道上皮中的表达具有特异性。在另一实施例中,启动子对转基因在神经元中的表达具有特异性。在另一实施例中,启动子对转基因在心脏中的表达具有特异性。

[0129] 表达盒通常含有作为表达控制序列的一部分的启动子序列,例如位于所选5'ITR序列与免疫球蛋白构建体编码序列之间。在一个实施例中,肝脏中的表达是合乎需要的。因此,在一个实施例中,使用肝特异性启动子。可以在本文所述的载体中使用组织特异性启动子、组成型启动子、可调控启动子[参见例如WO 2011/126808和WO 2013/04943]或对生理学线索有应答的启动子。在另一实施例中,肌肉中的表达是合乎需要的。因此,在一个实施例中,使用肌肉特异性启动子。在一个实施例中,启动子是基于MCK的启动子,如dMCK(509-bp)或tMCK(720-bp)启动子(参见例如Wang等人,《基因疗法》2008年11月;15(22):1489-99.doi:10.1038/gt.2008.104.电子版2008年6月19日,其以引用的方式并入本文中)。另一适用的启动子是SPc5-12启动子(参见Rasowo等人,《欧洲科学杂志(European Scientific Journal)》2014年6月版第10卷,第18期,其以引用的方式并入本文中)。在一个实施例中,启动子为CMV启动子。在另一实施例中,启动子为TBG启动子。在另一实施例中,使用CB7启动子或CAG启动子。CB7是具有巨细胞病毒增强子元件的鸡β-肌动蛋白启动子。或者,可以使用其它肝特异性启动子[参见例如,肝特异性基因启动子数据库(The Liver Specific Gene Promoter Database),Cold Spring Harbor,rulai.sch1.edu/LSPD,α-1抗胰蛋白酶(A1AT);人白蛋白,Miyatake等人,《病毒学杂志》,71:5124-32(1997),humAlb;以及乙型肝炎病毒核心启动子,Sandig等人,《基因疗法》,3:1002-9(1996)]。TTR最小增强子/启动子、α-抗胰蛋白酶启动子,LSP(845nt)25(需要无内含子scAAV)。

[0130] 一种或多种启动子可以选自不同的来源,例如人巨细胞病毒(CMV)立即早期增强子/启动子、SV40早期增强子/启动子、JC多瘤病毒启动子、髓鞘碱性蛋白(MBP)或神经胶质原纤维酸性蛋白(GFAP)启动子、单纯疱疹病毒(HSV-1)潜伏期相关启动子(LAP)、劳氏肉瘤病毒(RSV)长末端重复LTR启动子、神经元特异性启动子(NSE)、血小板源性生长因子(PDGF)启动子、hSYN、黑色素浓缩激素(MCH)启动子、CBA、基质金属蛋白启动子(MPP)和鸡 $\beta$ -肌动蛋白启动子。

[0131] 表达盒可含有至少一种增强子,即CMV增强子。其它增强子元件可包括例如载脂蛋白增强子、斑马鱼增强子、GFAP增强子元件和脑特异性增强子(如WO 2013/1555222中所描述)、土拔鼠肝炎后转录后调节元件。另外或替代地,可以选择其它,例如杂交人类巨细胞病毒(HCMV)立即早期(IE)-PDGR启动子或其它启动子-增强子元件。本文适用的其它增强子序列包括IRBP增强子(Nicoud 2007,《基因医学杂志》2007年12月;9(12):1015-23)、立即早期巨细胞病毒增强子、源自免疫球蛋白基因或SV40增强子的一种、小鼠近端启动子中鉴别的顺式作用元件等。

[0132] 除启动子之外,表达盒和/或载体还可以含有一个或多个其它合适的转录起始、终止、增强子序列、如剪接和聚腺苷酸化(polyA)信号等的有效RNA加工信号;稳定胞质mRNA的序列;增强翻译效率的序列(即,Kozak共有序列);增强蛋白稳定性的序列;以及当需要时,增强所编码的产物的分泌的序列。多种合适的polyA是已知的。在一个实例中,polyA是兔 $\beta$ 珠蛋白,如127bp兔 $\beta$ 珠蛋白聚腺苷酸化信号(GenBank编号V00882.1)。在其它实施例中,选择SV40 polyA信号。可选择其它合适的polyA序列。在某些实施例中,包括内含子。一个合适的内含子是鸡 $\beta$ -肌动蛋白内含子。在一个实施例中,内含子为875bp(GenBank编号X00182.1)在另一实施例中,使用可购自Promega的嵌合内含子。然而,可选择其它合适的内含子。在一个实施例中,包括间隔子以使得载体基因组与天然AAV载体基因组的大小大致相同(例如,在4.1与5.2kb之间)。在一个实施例中,包括间隔子以使得载体基因组为约4.7kb。参见Wu等人,基因组大小对AAV载体包装的影响(Effect of Genome Size on AAV Vector Packaging),《分子治疗学(Mol Ther.)》2010年1月;18(1):80-86,其以引用的方式并入本文中。

[0133] 这些和其它常见载体和调节元件的选择是常规的,并且许多此类序列是可用的。参见例如Sambrook等人和其中所引用的参考文献,例如第3.18-3.26和16.17-16.27页,以及Ausubel等人,现代分子生物学实验技术(Current Protocols in Molecular Biology),John Wiley&Sons,New York,1989。当然,并非所有载体和表达控制序列都将同样好地起作用以表达如本文所述的所有转基因。然而,本领域的技术人员可以在不脱离本发明的范围的情况下在这些和其它表达控制序列中进行选择。

[0134] 在某些实施例中,表达盒含有作为miR-183靶序列的至少一个miRNA靶序列。在某些实施例中,载体基因组或表达盒含有miR-183靶序列,其包括AGTGAATTCTACCAGTGCCATA(SEQ ID NO:9),其中与miR-183种子序列互补的序列加下划线。在某些实施例中,载体基因组或表达盒含有与miR-183种子序列100%互补的序列的多于一个拷贝(例如,两个或三个拷贝)。在某些实施例中,miR-183靶序列的长度为约7个核苷酸到约28个核苷酸并且包括与miR-183种子序列至少100%互补的至少一个区。在某些实施例中,miR-183靶序列含有与SEQ ID NO:9部分互补的序列,且因此当与SEQ ID NO:9比对时,存在一个或多个错配。在某

些实施例中,当与SEQ ID NO:9比对时,miR-183靶序列包含具有至少1、2、3、4、5、6、7、8、9或10个错配的序列,其中所述错配可以是不连续的。在某些实施例中,miR-183靶序列包括具有100%互补性的区,所述区还包含miR-183靶序列的长度的至少30%。在某些实施例中,具有100%互补性的区域包括与miR-183种子序列具有100%互补性的序列。在某些实施例中,miR-183靶序列的其余部分与miR-183具有至少约80%至约99%的互补性。在某些实施例中,表达盒或载体基因组包括miR-183靶序列,所述靶序列包含截短的SEQ ID NO:9,即在SEQ ID NO:9的5'端或3'端中的任一端或两端处缺乏至少1、2、3、4、5、6、7、8、9或10个核苷酸的序列。在某些实施例中,表达盒或载体基因组包含转基因和一个miR-183靶序列。在其它实施例中,表达盒或载体基因组包含至少两个、三个或四个miR-183靶序列。

[0135] 在某些实施例中,表达盒含有作为miR-182靶序列的至少一个miRNA靶序列。在某些实施例中,载体基因组或表达盒含有miR-182靶序列,其包括AGTGTGAGTTCTACCATTGCCAAA (SEQ ID NO:10)。在某些实施例中,载体基因组或表达盒含有与miR-182种子序列100%互补的序列的多于一个拷贝(例如,两个或三个拷贝)。在某些实施例中,miR-182靶序列的长度为约7个核苷酸到约28个核苷酸并且包括与miR-182种子序列至少100%互补的至少一个区。在某些实施例中,miR-182靶序列含有与SEQ ID NO:10部分互补的序列,且因此当与SEQ ID NO:10比对时,存在一个或多个错配。在某些实施例中,当与SEQ ID NO:10比对时,miR-183靶序列包含具有至少1、2、3、4、5、6、7、8、9或10个错配的序列,其中所述错配可以是不连续的。在某些实施例中,miR-182靶序列包括具有100%互补性的区,所述区还包含miR-182靶序列的长度的至少30%。在某些实施例中,具有100%互补性的区包括与miR-182种子序列具有100%互补性的序列。在某些实施例中,miR-182靶序列的其余部分与miR-182具有至少约80%至约99%的互补性。在某些实施例中,表达盒或载体基因组包括miR-182靶序列,所述靶序列包含截短的SEQ ID NO:10,即在SEQ ID NO:10的5'端或3'端中的任一端或两端处缺乏至少1、2、3、4、5、6、7、8、9或10个核苷酸的序列。在某些实施例中,表达盒或载体基因组包含转基因和一个miR-182靶序列。在其它实施例中,表达盒或载体基因组包含至少两个、三个或四个miR-182靶序列。

[0136] 本文所使用的术语“串联重复序列”是指存在两个或更多个连续miRNA靶序列。这些miRNA靶序列可以是连续的,即一个接一个地直接定位,使得一个靶序列的3'端直接位于下一个靶序列的5'端的上游,没有中间序列,或者反之亦然。在另一个实施例中,miRNA靶序列中的两个或更多个miRNA靶序列由短间隔序列隔开。

[0137] 如本文所用,“间隔子”是任何所选核酸序列,例如,长度为1个、2个、3个、4个、5个、6个、7个、8个、9个或10个核苷酸的定位在两个或更多个连续miRNA靶序列之间的核酸序列。在某些实施例中,间隔子的长度为1个到8个核苷酸、长度为2个到7个核苷酸、长度为3个到6个核苷酸、长度为4个核苷酸、4个到9个核苷酸、3个到7个核苷酸或更大的值。合适地,间隔子是非编码序列。在某些实施例中,间隔子可以具有四(4)个核苷酸。在某些实施例中,间隔子是GGAT。在某些实施例中,间隔子是六(6)个核苷酸。在某些实施例中,间隔子是CACGTG或GCATGC。

[0138] 在某些实施例中,串联重复序列含有相同miRNA靶序列中的两个、三个、四个或更多个。在某些实施例中,串联重复序列含有至少两个不同miRNA靶序列、至少三个不同miRNA靶序列或至少四个不同miRNA靶序列等。在某些实施例中,串联重复序列可含有相同miRNA

靶序列中的两个或三个和不同的第四miRNA靶序列。

[0139] 在某些实施例中,表达盒中可存在至少两组不同的串联重复序列。例如,3'UTR可含有紧接在转基因下游的串联重复序列、UTR序列和两个或更多个更接近UTR的3'端的串联重复序列。在另一实例中,5'UTR可含有一个、两个或更多个miRNA靶序列。在另一实例中,3'可含有串联重复序列,并且5'UTR可含有至少一个miRNA靶序列。

[0140] 在某些实施例中,表达盒含有两个、三个、四个或更多个串联重复序列,所述串联重复序列在转基因的终止密码子的约0个至20个核苷酸内开始。在其它实施例中,表达盒含有距转基因的终止密码子至少100个至约4000个核苷酸的miRNA串联重复序列。

[0141] 参见2019年12月20日提交的PCT/US19/67872,其以引用的方式并入本文中并且要求2018年12月21日申请的美国临时专利申请第62/783,956号的优先权,所述临时申请以引用的方式并入本文中。

[0142] 在另一实施例中,提供一种产生重组腺相关病毒的方法。合适的重组腺相关病毒(AAV)通过培养宿主细胞产生,所述宿主细胞含有编码如本文所述的AAV衣壳蛋白质或其片段的核酸序列;功能性rep基因;至少由AAV反向末端重复序列(ITR)和编码所需转基因的异源核酸序列构成的小基因;和允许将小基因包装至AAV衣壳蛋白质中的足够辅助功能。在宿主细胞中培养以将AAV小基因包装至AAV衣壳中所需的组分可以反式形式提供至宿主细胞。或者,可通过稳定宿主细胞提供所需组分(例如小基因、rep序列、cap序列和/或辅助功能)中的任一种或多种,所述宿主细胞已使用本领域的技术人员已知的方法工程化为含有所需组分中的一种或多种。

[0143] 本文还提供用如本文所述的AAV转染的宿主细胞。最合适的是,此类稳定的宿主细胞将含有在诱导型启动子的控制下的所需组分。然而,所需组分可在组成型启动子的控制下。合适的诱导型和组成型启动子的实例提供于下文适用于转基因的调节元件的论述中。在另一替代方案中,所选稳定宿主细胞可含有在组成型启动子控制下的所选成分和在一或多个诱导型启动子控制下的其它所选成分。例如,可产生稳定宿主细胞,所述宿主细胞来源于293细胞(其含有在组成型启动子控制下的E1辅助功能),但含有在诱导型启动子控制下的rep和/或cap蛋白。其它稳定的宿主细胞可由本领域的技术人员产生。在另一实施例中,宿主细胞包含如本文所述的核酸分子。

[0144] 产生本文所述的rAAV所需的小基因、rep序列、cap序列和辅助功能可以转移其上携带的序列的任何遗传元件的形式递送到包装宿主细胞。所选遗传元件可通过包括本文所述的方法在内的任何合适方法递送。用于构建本发明的任何实施例的方法对核酸操纵技术人员是已知的并且包括基因工程、重组工程以及合成技术。参见例如Sambrook等人,《分子克隆实验指南(Molecular Cloning:A Laboratory Manual)》,Cold Spring Harbor Press,Cold Spring Harbor,NY。类似地,产生rAAV病毒粒子的方法是众所周知的并且对合适的方法的选择不是对本发明的限制。参见例如K.Fisher等人,1993,《病毒学杂志(J.Virol.)》,70:520-532和美国专利5,478,745等。这些出版物以引用的方式并入本文中。

[0145] 本文还提供用于产生本文所述的载体的质粒。此类质粒描述于实例部分中。

[0146] C.药物组合物和施用

[0147] 在一个实施例中,任选地通过常规方法评估如上所详述用于靶细胞的含有所需转基因和启动子的重组AAV的污染,且接着配制成意图向有需要的受试者施用的药物组合物。

此类调配物涉及使用药学上和/或生理学上可接受的媒剂或载剂,如缓冲生理盐水或其它缓冲剂,例如HEPES,以将pH维持在适当生理水平,和任选地其它药剂、医药剂、稳定剂、缓冲剂、载剂、佐剂、稀释剂等。对于注射,载剂将通常为液体。示例性生理上可接受的载剂包括无菌、无热原水和无菌、无热原磷酸盐缓冲盐水。美国专利公开案第7,629,322号中提供了各种此类已知载剂,所述公开案以引用的方式并入本文中。在一个实施例中,载剂是等渗氯化钠溶液。在另一实施例中,载剂是平衡盐溶液。在一个实施例中,载剂包括tween。如果病毒要长期存储,则其可以在甘油或Tween20存在下冷冻。在另一实施例中,药学上可接受的载剂包含表面活性剂,例如全氟辛烷(Perfluoron液体)。在适合人类受试者输注的缓冲液/载剂中配制载体。缓冲液/载剂应包括防止rAAV粘附到输液管道上但不干扰rAAV体内结合活性的成分。

[0148] 在本文所述的方法的某些实施例中,向受试者肌内施用上文所述的药物组合物。在其它实施例中,通过静脉内施用药物组合物。在其它实施例中,通过脑室内注射施用药物组合物。在其它实施例中,药物组合物通过小脑延髓池内(ICM)注射给药。可用于本文所述的方法的其它施用形式包括但不限于直接递送至所需器官(例如,眼睛),包括视网膜下或玻璃体内递送、口服、吸入、鼻内、气管内、静脉内、肌肉内、皮下、皮内和其它肠胃外施用途径。必要时,可以组合施用途径。

[0149] 如本文所用,术语“鞘内递送”或“鞘内施用”是指通过注射到椎管中,更具体地注射到蛛网膜下腔中使得其到达脑脊液(CSF)的施用途径。鞘内递送可包括腰椎穿刺、室内(包括脑室内(ICV))、枕骨下/脑池内和/或C1-2穿刺。例如,可以通过腰椎穿刺引入材料以在整个蛛网膜下腔中扩散。在另一实例中,可以向小脑延髓池中注射。

[0150] 如本文所用,术语“脑池内递送”或“脑池内施用”是指直接进入小脑延髓池(cisterna magna cerebellomedularis)的脑脊液中,更具体地通过枕骨下穿刺或通过直接注射到小脑延髓池(cisterna magna)中或通过永久定位的管的施用途径。

[0151] 取决于待治疗区域的大小、所用病毒滴度、施用途径和所述方法的期望效果,组合物可以约0.1 $\mu$ L至约10mL体积递送,包括在所述范围内的所有数字。在一个实施例中,体积为约50 $\mu$ L。在另一实施例中,体积为约70 $\mu$ L。在另一实施例中,体积为约100 $\mu$ L。在另一实施例中,体积为约125 $\mu$ L。在另一实施例中,体积为约150 $\mu$ L。在另一实施例中,体积为约175 $\mu$ L。在另一实施例中,体积为约200 $\mu$ L。在另一实施例中,体积为约250 $\mu$ L。在另一实施例中,体积为约300 $\mu$ L。在另一实施例中,体积为约450 $\mu$ L。在另一实施例中,体积为约500 $\mu$ L。在另一实施例中,体积为约600 $\mu$ L。在另一实施例中,体积为约750 $\mu$ L。在另一实施例中,体积为约850 $\mu$ L。在另一实施例中,体积为约1000 $\mu$ L。在另一实施例中,体积为约1.5mL。在另一实施例中,体积为约2mL。在另一实施例中,体积为约2.5mL。在另一实施例中,体积为约3mL。在另一实施例中,体积为约3.5mL。在另一实施例中,体积为约4mL。在另一实施例中,体积为约5mL。在另一实施例中,体积为约5.5mL。在另一实施例中,体积为约6mL。在另一实施例中,体积为约6.5mL。在另一实施例中,体积为约7mL。在另一实施例中,体积为约8mL。在另一实施例中,体积为约8.5mL。在另一实施例中,体积为约9mL。在另一实施例中,体积为约9.5mL。在另一实施例中,体积为约10mL。

[0152] 携带在调节序列控制下编码所需转基因的核酸序列的重组腺相关病毒的有效浓度理想地在约10<sup>7</sup>至10<sup>14</sup>个载体基因组/毫升(vg/mL)(也称为基因组拷贝/mL(GC/mL))范

围内。在一个实施例中，rAAV载体基因组通过实时PCR测量。在另一实施例中，rAAV载体基因组通过数字PCR测量。参见Lock等人，通过液滴数字PCR绝对测定单股和自互补腺相关病毒载体基因组滴度 (Absolute determination of single-stranded and self-complementary adeno-associated viral vector genome titers by droplet digital PCR) ,《人类基因治疗方法(Hum Gene Ther Methods.)》2014年4月;25 (2) :115-25.doi: 10.1089/hgtb.2013.131.电子版2014年2月14日,所述文献以引用的方式并入本文中。在另一实施例中,如以引用的方式并入本文中的S.K. McLaughlin等人,1988《病毒学杂志(J.Virol.)》,62:1963中所述地测量rAAV感染单位。

[0153] 优选地,浓度为约 $1.5 \times 10^9$  vg/mL至约 $1.5 \times 10^{13}$  vg/mL,并且更优选为约 $1.5 \times 10^9$  vg/mL至约 $1.5 \times 10^{11}$  vg/mL。在一个实施例中,有效浓度为约 $1.4 \times 10^8$  vg/mL。在一个实施例中,有效浓度为约 $3.5 \times 10^{10}$  vg/mL。在另一实施例中,有效浓度为约 $5.6 \times 10^{11}$  vg/mL。在另一实施例中,有效浓度为约 $5.3 \times 10^{12}$  vg/mL。在另一实施例中,有效浓度为约 $1.5 \times 10^{12}$  vg/mL。在另一实施例中,有效浓度为约 $1.5 \times 10^{13}$  vg/mL。本文中所述的所有范围均包括端点。

[0154] 在一个实施例中,剂量为约 $1.5 \times 10^9$  vg/kg体重至约 $1.5 \times 10^{13}$  vg/kg,并且更优选为约 $1.5 \times 10^9$  vg/kg至约 $1.5 \times 10^{11}$  vg/kg。在一个实施例中,剂量为约 $1.4 \times 10^8$  vg/kg。在一个实施例中,剂量为约 $3.5 \times 10^{10}$  vg/kg。在另一实施例中,剂量为约 $5.6 \times 10^{11}$  vg/kg。在另一实施例中,剂量为约 $5.3 \times 10^{12}$  vg/kg。在另一实施例中,剂量为约 $1.5 \times 10^{12}$  vg/kg。在另一实施例中,剂量为约 $1.5 \times 10^{13}$  vg/kg。在另一实施例中,剂量为约 $3.0 \times 10^{13}$  vg/kg。在另一实施例中,剂量为约 $1.0 \times 10^{14}$  vg/kg。本文所述的所有范围均包括端点。

[0155] 在一个实施例中,有效剂量(递送的总基因组复本)为约 $10^7$ 至 $10^{13}$ 个载体基因组。在一个实施例中,总剂量为约 $10^8$ 个基因组复本。在一个实施例中,总剂量为约 $10^9$ 个基因组复本。在一个实施例中,总剂量为约 $10^{10}$ 个基因组复本。在一个实施例中,总剂量为约 $10^{11}$ 个基因组复本。在一个实施例中,总剂量为约 $10^{12}$ 个基因组复本。在一个实施例中,总剂量为约 $10^{13}$ 个基因组复本。在一个实施例中,总剂量为约 $10^{14}$ 个基因组复本。在一个实施例中,总剂量为约 $10^{15}$ 个基因组复本。

[0156] 需要利用最低有效病毒浓度以便降低不合需要的影响(例如毒性)的风险。在这些范围内的其它剂量和施用体积可以由主治医师考虑正在治疗的受试者(优选地人类)的身体状态、所述受试者的年龄、特定病症以及所述病症(如果是进行性的)的发展程度来选择。例如,静脉内递送可能需要约 $1.5 \times 10^{13}$  vg/kg的剂量。

#### [0157] D.方法

[0158] 在另一方面,提供了一种转导靶组织的方法。在一个实施例中,所述方法包括施用具有如本文所述的AAVrh.90衣壳的AAV。如下实例中所展示,本发明人已展示被称为AAVrh.90的AAV有效地转导CNS(脑)、肝脏、心脏和肌肉组织。因此,本文提供一种转导肌肉的方法,其包含施用具有AAVrh.90衣壳的rAAV。在另一实施例中,本文提供一种转导脑的方法,其包含施用具有AAVrh.90衣壳的rAAV。在另一实施例中,本文提供一种转导肝脏的方法,其包含施用具有AAVrh.90衣壳的rAAV。在另一实施例中,本文提供一种转导心脏的方法,其包含施用具有AAVrh.90衣壳的rAAV。在一个实施例中,采用静脉内施用。在另一实施例中,采用ICV施用。在另一实施例中,采用ICM施用。

[0159] 如本文所论述,包含本文所述的AAV衣壳的载体能够以高水平转导靶组织。因此,

本文提供一种将转基因递送至肝细胞的方法。所述方法包括使细胞与具有AAVrh.90衣壳的rAAV接触,其中所述rAAV包含转基因。在另一方面,提供具有AAVrh.90衣壳的rAAV的用途,其用于将转基因递送至肝脏。在某些实施例中,方法包含施用具有AAVrh.90衣壳的rAAV以将转基因递送至肝细胞,其中所述转基因选自前蛋白转化酶枯草杆菌蛋白酶/枯草溶菌素9型(PCSK9)(胆固醇相关疾病)、转甲状腺素蛋白(TTR)(转甲状腺素蛋白淀粉样变性)、HAO、载脂蛋白C-III(APOC3)、因子VIII、因子IX、低密度脂蛋白受体(LDLr)、脂蛋白脂肪酶(LPL)(脂蛋白脂肪酶缺乏症)、卵磷脂-胆固醇酰基转移酶(LCAT)、鸟氨酸转氨甲酰酶(OTC)、肌肽酶(CN1)、鞘磷脂磷酸二酯酶(SMPD1)(尼曼-皮克病(Niemann-Pick disease))、次黄嘌呤-鸟嘌呤磷酸核糖转移酶(HGPRT)、分支链 $\alpha$ -酮酸脱氢酶复合物(BCKDC)(枫糖浆尿病)和红细胞生成素(EPO)。

[0160] 本文还提供一种将转基因递送至肌肉细胞的方法。所述方法包括使细胞与具有AAVrh.90衣壳的rAAV接触,其中所述rAAV包含转基因。在另一方面,提供具有AAVrh.90衣壳的rAAV的用途,其用于将转基因递送至肌肉。

[0161] 本文还提供一种将转基因递送至心脏细胞的方法。所述方法包括使细胞与具有AAVrh.90衣壳的rAAV接触,其中所述rAAV包含转基因。在另一方面,提供具有AAVrh.90衣壳的rAAV的用途,其用于将转基因递送至心脏。

[0162] 本文还提供一种将转基因递送至肾细胞的方法。所述方法包括使细胞与具有AAVrh.90衣壳的rAAV接触,其中所述rAAV包含转基因。在另一方面,提供具有AAVrh.90衣壳的rAAV的用途,其用于将转基因递送至心脏。

[0163] 本文还提供一种将转基因递送至胰腺细胞的方法。所述方法包括使细胞与具有AAVrh.90衣壳的rAAV接触,其中所述rAAV包含转基因。在另一方面,提供具有AAVrh.90衣壳的rAAV的用途,其用于将转基因递送至胰腺。

[0164] 本文还提供一种将转基因递送至脑细胞的方法。所述方法包括使细胞与具有AAVrh.90衣壳的rAAV接触,其中所述rAAV包含转基因。在另一方面,提供具有AAVrh.90衣壳的rAAV的用途,其用于将转基因递送至脑。在某些实施例中,rAAV使用ICM递送来递送。

#### [0165] 单基因组扩增

[0166] AAV基因组传统上使用基于PCR的方法与全哺乳动物基因组DNA分离:引物用于检测在大部分不同VP1(衣壳)基因两侧的保守区。然后将PCR产物克隆到质粒主链中并且使用桑格方法对单独的克隆进行测序。传统的基于PCR和分子克隆的病毒分离方法对于恢复新的AAV基因组有效,但所恢复的基因组可能受PCR介导的重组和聚合酶错误的影响。此外,与先前使用的桑格技术相比,当前可用的下一代测序技术允许我们以空前未有的准确度对病毒基因组测序。本文提供一种从病毒群体内准确分离单个AAV基因组的新型、更高通量、基于PCR和下一代测序的方法。此方法,AAV-单基因组扩增(AAV-SGA)可用于提高我们对哺乳动物宿主体内的AAV多样性的了解。此外,它允许我们鉴别适用作基因疗法的载体的新型衣壳。

[0167] AAV-SGA已经过验证和优化,以有效地从包含基因组群体的样本中恢复单个AAV序列。此技术先前已用于从人类和非人类灵长类动物宿主内分离单个HIV和HCV基因组。通过对衣壳检测PCR筛选出AAV阳性的基因组DNA样本进行终点稀释。根据泊松分布(置信度为80%),PCR扩增产生小于30%阳性反应的稀释度含有单个可扩增AAV基因组。此程序允许病

毒基因组的PCR扩增,降低由聚合酶的模板转换引起的PCR介导的重组的机率。AAV-SGA PCR扩增子使用Illumina MiSeq平台使用 $2\times 150$ 或 $2\times 250$ 双端测序进行测序。此方法允许对全长AAV VP1序列进行准确的从头组装,而无需担心来自包含具有高度同源性区域的多个病毒的单个样本的测序读段的收敛。

[0168] AAV-SGA技术已成功从恒河猴组织中分离出多种新型AAV衣壳序列。已经从单个样本中鉴定出来自不同AAV进化枝的多种病毒;这表明宿主组织中可以存在AAV的群体。例如,从单个肝组织样本中分离出与进化枝D、E和外围“边缘”病毒具有序列相似性的衣壳。

[0169] 先前尚未描述过SGA在AAV发现中的应用。所述方法解决了可能导致无效AAV基因组序列的模板转换和聚合酶错误问题。此外,当从与单个分离株相同的宿主样本中重复回收相同的序列时,分离出的基因组的质量是不言而喻的。

[0170] 提供以下实例以说明本发明的各种实施例。实例并不打算以任何方式限制本发明

[0171] E. 实例

[0172] 实例1:材料和方法

[0173] AAV序列的检测和分离

[0174] 非人灵长类动物组织来源

[0175] 来自宾夕法尼亚大学群落的恒河猴是人工饲养的,并且源自中国或印度。恒河猴的肝组织样本由宾夕法尼亚大学的基因疗法计划和Timothy H.Lucas实验室友情提供。

[0176] 新型AAV分离

[0177] 提取基因组DNA (QIAamp DNA Mini Kit,QIAGEN) 并通过使用PCR策略从NHP肝组织标本中扩增3.1-kb全长Cap片段来分析AAV DNA的存在。使用AAV Rep基因保守区内的5'引物 (AV1NS, 5'-GCTCGTCAACTGGACCAATGAGAAC-3') (SEQ ID N0:5) 与位于AAV Cap下游保守区的3'引物 (AV2CAS, 5'-CGCAGAGACCAAAGTTC AACTGAAACGA-3') (SEQ ID N0:6) 的组合来扩增全长AAV Cap扩增子的基因。Q5High-Fidelity Hot Start DNA聚合酶 (New England Biolabs) 用于使用以下循环条件扩增AAV DNA: 98°C持续30s; 98°C持续10s, 59°C持续10s, 72°C持续93s, 50个循环;以及72°C延伸持续120s。

[0178] 对PCR反应呈阳性的模板基因组DNA样品进行AAV-单基因组扩增(AAV-SGA)。在96孔板中对基因组DNA进行终点稀释,使得使用上述相同引物的96次PCR反应中的少于29次产生扩增产物。根据泊松分布,在不超过30%的孔中产生PCR产物的DNA稀释液在超过80%的时间内每个阳性PCR含有一个可扩增的AAV DNA模板。使用Illumina MiSeq  $2\times 150$ 或 $2\times 250$ 配对末端测序平台对来自阳性PCR反应的AAV DNA扩增子进行测序,并使用SPAdes组装程序 ([cab.spbu.ru/software/spades](http://cab.spbu.ru/software/spades)) 从头组装所得读段。使用NCBI BLASTn ([blast.ncbi.nlm.nih.gov](http://blast.ncbi.nlm.nih.gov)) 和Vector NTI AlignX软件 (Thermo Fisher) 进行序列分析。

[0179] 使用新型AAV衣壳生产载体

[0180] 来自所关注的PCR产物的AAV衣壳基因DNA序列被TOP0克隆和扩增(Invitrogen)。扩增的衣壳基因被进一步克隆到包含AAV2 Rep基因和其它相关质粒元件的AAV转质粒骨架中。

[0181] 如前所述,AAV载体由Penn Vector Core生产和滴定(参见例如Lock,M.等人(2010)《人类基因疗法(Hum.Gene Ther.)》21:1259-71)。HEK293细胞进行三次转染,然后收集细胞培养上清液,浓缩,并用碘克沙醇梯度纯化。如前所述,使用针对兔β-珠蛋白polyA序

列的引物,通过液滴数字PCR滴定纯化的载体(参见例如Lock,M.等人(2014)《人类基因治疗方法(Hum.Gene Ther.Methods)》25:115-125)。

[0182] 噬齿动物中新型AAV衣壳的体内表征

[0183] 动物

[0184] 所有动物方案均经宾夕法尼亚大学的机构动物护理和使用委员会(the Institutional Animal Care and Use Committee of the University of Pennsylvania)批准。从杰克逊实验室(Jackson Laboratory)购买C56BL/6J小鼠。对于GFP报告基因实验,对成年(6-8周大)雄性进行注射。动物被圈养在每笼二到五只动物的标准笼中。笼子、水瓶和垫料基材在屏障设施中进行高压灭菌,且每周更换一次笼子。保持自动控制的12小时光暗循环。每个黑暗期从晚上7:00(±30分钟)开始。随意提供辐射的实验室啮齿动物食物。

[0185] 测试品和研究设计

[0186] 小鼠通过侧尾静脉以静脉内(IV)0.1mL接受每只小鼠 $1 \times 10^{12}$ GC的每种载体,或以每只小鼠含 $1 \times 10^{11}$ GC的5μL剂量脑室内(ICV)注射到脑的侧脑室。每组向三或五只小鼠给药。

[0187] 注射后14天通过吸入CO<sub>2</sub>对小鼠实施安乐死。收集组织,在干冰上快速冷冻以进行生物分布分析,或在10%中性福尔马林中浸泡固定,在蔗糖中冷冻保存,在OCT中冷冻,并用低温恒温器切片用于GFP直接观察。用于内皮细胞转导分析的组织在尸体剖检后进行石蜡包埋。

[0188] 载体生物分布

[0189] 使用QIAamp DNA Mini Kit (QIAGEN) 提取组织基因组DNA,并使用Taqman试剂(Applied Biosystems,Life Technologies)与靶向载体的EGFP序列的引物/探针通过实时PCR对AAV载体基因组进行定量。

[0190] 报告基因可视化

[0191] 为了观察直接GFP荧光,将组织样品在福尔马林中固定约24小时,在PBS中短暂洗涤,在含15%和30%蔗糖的PBS中依次平衡直至达到最大密度,且接着在OCT包埋培养基中冷冻以制备冷冻切片。切片安装在含有DAPI(Electron Microscopy Sciences,Hatfield,PA)作为核复染剂的Fluoromount G中。

[0192] 对石蜡包埋的组织样本进行GFP免疫组织化学。切片用乙醇和二甲苯脱蜡,在10mM柠檬酸盐缓冲液(pH 6.0)中煮沸6分钟以进行抗原修复,依次用2%H<sub>2</sub>O<sub>2</sub>处理15分钟,用抗生素蛋白/生物素封闭试剂各处理15分钟(Vector Laboratories),以及用封闭缓冲液(含1%驴血清+0.2%Triton的PBS)处理10分钟。然后将其与一级抗体一起培育1小时,并在封闭缓冲液中与生物素标记的二级抗体一起培育45分钟(Jackson Immunoresearch)。使用一级抗体鸡抗GFP (Abcam ab13970) 和兔抗CD31 (Abcam ab28364) 内皮细胞标记。按照制造商说明书使用Vectastain Elite ABC试剂盒(Vector Laboratories),以DAB作为底物,将结合的抗体可视化为棕色沉淀物。

[0193] 对于免疫荧光,在用含1%驴血清的PBS+0.2%Triton进行抗原修复后,石蜡切片脱蜡并封闭15分钟,然后与在封闭缓冲液中稀释的一级(1小时)和荧光标记的二级抗体(45分钟,Jackson Immunoresearch)连续培育。使用的抗体是鸡抗GFP (Abcam ab13970)、兔抗

CD31 (Abcam ab28364) 和小鼠抗NF-200 (克隆RT97, Millipore CBL212)。将一级抗体混合在一起,且分别通过FITC和TRITC标记的二抗检测GFP和NF-200抗体。根据制造商的方案 (Vector Labs), 使用VectaFluor<sup>TM</sup> Excel AmplifiedDyLight<sup>®</sup> 488Anti-Rabbit IgG试剂盒增强了针对CD31的兔抗体的信号。荧光和明场显微镜图像是用Nikon Eclipse TiE显微镜拍摄的。

[0194] 带条形码的载体转基因的非人灵长类动物转导评估

[0195] 测试品和研究设计

[0196] 五个新型衣壳和五个对照衣壳 (AAVrh90、AAVrh91、AAVrh92、AAVrh93、AAVrh91.93、AAV8、AAV6.2、AAVrh32.33、AAV7和AAV9) 用于包装改良的ATG耗尽自互补eGFP (dGFP) 转基因。每个独特的衣壳制剂都含有dGFP转基因,在载体基因组的多聚腺苷酸化序列之前具有相应的独特6bp条形码。转基因含有CB8启动子和SV40聚腺苷酸化序列 (AAVsc.CB8.dGFP.barcode.SV40)。如前所述,AAV载体由Penn Vector Core生产和滴定(参见例如Lock,M.等人 (2010)《人类基因疗法 (Hum.Gene Ther.)》21:1259-71)。HEK293细胞进行三次转染,然后收集细胞培养上清液,浓缩,并用碘克沙醇梯度纯化。如前所述,使用靶向 SV40 polyA序列的引物,通过液滴数字PCR滴定纯化的载体(参见例如Lock,M.等人 (2014)《人类基因治疗方法 (Hum.Gene Ther.Methods)》25:115-25)。

[0197] 将十个纯化的载体以相等的基因组拷贝量汇集,用于注射到两只不同的动物中:通过IV递送递送的总剂量为2e13 GC/kg,且通过小脑延髓池内 (ICM) 递送递送到鞘内空间的总剂量为3e13 GC/动物。在注射后30天处死动物,并在RNAlater (QIAGEN) 中收集所有组织用于下游转基因RNA表达分析。

[0198] 动物

[0199] 所有动物程序均经宾夕法尼亚大学的机构动物护理和使用委员会 (the Institutional Animal Care and Use Committee of the University of Pennsylvania) 批准。猕猴(食蟹猴)由Bristol Meyers Squibb (USA) 捐赠。动物被安置在宾夕法尼亚州费城市的费城儿童医院的实验室动物护理评估和认证协会认可的非人类灵长类动物研究计划设施的不锈钢挤压式后笼 (squeeze back cage) 中。动物接受了各种丰富活动,如食物款待、视觉和听觉刺激、操纵和社会互动。

[0200] 一只10岁的雄性8kg动物被用于ICM研究。一只6岁的雄性6.98kg动物被用于IV研究。对此动物进行了AAV中和抗体的存在的筛选,且基线时AAV6、AAV8和AAVrh32.33呈血清阴性。在基线时,此动物针对AAV7和AAV9的中和抗体滴度分别为1:5和1:10。

[0201] ICM注射程序

[0202] 将麻醉的猕猴以侧卧位放置在X射线台上,头部向前弯曲。使用无菌技术将21G-27G、1至1.5英寸Quincke脊髓针 (Becton Dickinson, Franklin Lakes, NJ, USA) 推进至枕下空间,直至观察到CSF流动。收集1mL CSF用于基线分析。针的正确放置通过荧光透视法 (OEC 9800C形臂; GE Healthcare, Little Chalfont, UK) 进行验证,以避免潜在的脑干损伤。CSF收集后,将Luer通路延长装置或小口径T端口延长装置导管连接到脊髓针上,以促进180mg/mL碘海醇造影剂 (GE Healthcare, Little Chalfont, UK) 的给药。在验证针头位置后,将装有测试品的注射器(体积等于1mL加上注射器体积和接头死区)连接到柔性接头并注射30±5秒。取下针头,且直接对穿刺部位施加压力。

[0203] IV注射程序

[0204] 通过输液泵 (Harvard Apparatus, Holliston, MA) 以1mL/min的速率将10mL载体测试品施用至猕猴的外周静脉。

[0205] 转基因表达分析

[0206] 根据制造商说明书 (Life Technologies) , 使用TRIzol从所有RNALater处理的组织中提取全组织RNA。根据制造商的方案 (Roche, Basel, Switzerland) 用DNase I处理提取的RNA。使用RNeasy Mini Kit (QIAGEN) 纯化RNA。使用Applied Biosystems High Capacity cDNA Reverse Transcriptase Kit (Life Technologies) 进行cDNA的逆转录合成。靶向位于6 bp 独特条形码两侧的区域的引物用于PCR扩增117 bp 扩增子 (正向引物 : GGCGAACAGCGGACACCGATATGAA (SEQ ID NO:7) , 反向引物:GGCTCTCGTCGCGTGAGAATGAGAA (SEQ ID NO:8)) , 且Q5 High-Fidelity Hot Start DNA聚合酶 (New England Biolabs) 用于使用以下循环条件进行反应:98°C持续30s;98°C持续10s,72°C持续17s,25个循环;以及72°C延伸持续120s。使用MiSeq Standard2×150bp测序平台 (Illumina) 对扩增子进行测序。使用来自Expression Analysis包 ([github.com/ExpressionAnalysis/ea-utils](https://github.com/ExpressionAnalysis/ea-utils)) 、cutadapt ([cutadapt.readthedocs.io/en/stable/](https://cutadapt.readthedocs.io/en/stable/)) 、fastx工具包 ([hannonlab.cshl.edu/fastx\\_toolkit/](https://hannonlab.cshl.edu/fastx_toolkit/)) 和R版本3.3.1. ([cran.r-project.org/bin/windows/base/old/3.3.1/](https://cran.r-project.org/bin/windows/base/old/3.3.1/)) 的fastq-join程序分析条形码读段。来自组织样本的条形码表达计数数据被标准化为来自每个动物的测序注射载体材料的条形码计数,并且来自每个组织样本的条形码比例使用GraphPad Prism版本7.04绘制。

[0207] NHP中的ICM AAVrh90转导表征研究

[0208] 动物和研究设计

[0209] 所有动物程序均经宾夕法尼亚大学的机构动物护理和使用委员会 (the Institutional Animal Care and Use Committee of the University of Pennsylvania) 批准。六只成年恒河猴 (Macaca mulatta) 通过PreLabs从Orient Biologics (Alice, TX) 获得。动物被安置在宾夕法尼亚州费城的费城儿童医院的实验室动物护理评估和认证协会认可的非人类灵长类动物研究计划设施的不锈钢挤压式后笼中。动物接受了各种丰富活动,如食物款待、视觉和听觉刺激、操纵和社会互动。

[0210] 使用先前描述的方法用AAV.CB7.CI.eGFP.WPRE.rBG转基因包装AAVrh90、AAV8和AAV9衣壳 (参见例如Lock, M. 等人 (2010)《人类基因疗法 (Hum. Gene Ther.)》21:1259-71和Lock, M. 等人 (2014)《人类基因治疗方法 (Hum. Gene Ther. Methods)》25:115-25) 。将3e13 GC的剂量ICM注射到每只动物中。ICM注射方法描述于上文。注射后14天处死动物,并在干冰上收集组织用于DNA载体生物分布研究。在非临床一般毒性研究期间,根据神经系统 (大脑、脊髓、神经和眼睛) 采样和处理的推荐实践,使用脑模具收集整个脑、修剪并切片。Pardo等人 (2012) .STP Position Paper。还收集组织,用福尔马林固定,并用石蜡包埋以用于组织病理学分析。

[0211] 载体转导的组织学分析

[0212] 对于GFP免疫组织化学 (IHC) , 切片用乙醇和二甲苯脱蜡,在10mM柠檬酸盐缓冲液 (pH 6.0) 中煮沸6分钟以进行抗原修复,依次用2% H2O2处理15分钟,用抗生素蛋白/生物素封闭试剂各处理15分钟 (Vector Laboratories) , 以及用封闭缓冲液 (含1% 驴血清的PBS

+0.2% Triton) 处理10分钟。随后将其与抗GFP的山羊抗体 (Novus Biologicals, NB100-1770, 1:500) 一起在4°C下在封闭缓冲液中培育过夜，并在PBS中洗涤后，与生物素标记的二级抗山羊抗体一起在封闭缓冲液中培育45分钟 (Jackson ImmunoResearch, 1:500)。在PBS中洗涤后，按照制造商说明书应用Vectastain Elite ABC试剂盒 (Vector Laboratories)，以DAB作为底物，将结合的抗体可视化为棕色沉淀物。

[0213] AAV衣壳上氨基酸修饰的质谱 (MS) 分析

[0214] 试剂

[0215] 碳酸氢铵、二硫苏糖醇 (DTT)、碘乙酰胺 (IAM) 购自Sigma (St. Louis, MO)。乙腈、甲酸和三氟乙酸 (TFA)、8M盐酸胍 (GndHCl) 和胰蛋白酶购自Thermo Fisher Scientific (Rockford, IL)。

[0216] 胰蛋白酶消化

[0217] 制备1M DTT和1.0M碘乙酰胺的储备溶液。使衣壳蛋白变性，并在存在10mM DTT和2M GndHCl的情况下在90°C下还原10分钟。允许样品冷却至室温，接着在黑暗中在室温下用30mM IAM使其烷基化30分钟。通过添加1mL DTT淬灭烷基化反应。以将最终GndHCl浓度稀释至200mM的体积向变性的蛋白质溶液中添加20mM碳酸氢铵，pH 7.5-8。添加胰蛋白酶溶液，使胰蛋白酶与蛋白质的比率达到1:20，并在37°C下培育4小时。在消化之后，将TFA添加到最终0.5%，以淬灭消化反应。

[0218] LC-MS/MS

[0219] 使用Acclaim PepMap柱 (长15cm, 内径300μm) 和与具有NanoFlex源的Q Exactive HF (赛默飞世尔科技公司) 耦合的Thermo UltiMate 3000RSLC系统 (赛默飞世尔科技公司) 进行在线色谱。在线分析期间，将柱温调节至35°C。使用流动相A (含0.1%甲酸的MilliQ水) 和流动相B (含0.1%甲酸的乙腈) 的梯度分离肽。梯度在15分钟内从4%B运行到6%B，接着在25分钟内运行到10%B (共40分钟)，接着在46分钟内运行到30%B (共86分钟)。将样品直接装载到柱。柱尺寸为75cm×15um I.D.，并装有2微米的C18培养基 (Acclaim PepMap)。由于装载、导入和洗涤步骤，LC-MS/MS运行的总时间约为2小时。

[0220] MS数据是使用用于Q Exactive HF的数据依赖性前20种方法获取的，所述方法从调查扫描 (200-2000m/z) 中动态选择最丰富的尚未测序的前体离子。测序通过高能碰撞解离片段化进行，其中通过预测性自动增益控制确定的目标值为1e5离子，并且以4m/z的窗口进行前体分离。在m/z 200下以120,000的分辨率获取调查扫描。HCD光谱的分辨率在m/z 200下设置为30,000，其中最大离子注入时间为50毫秒，并且归一化碰撞能量为30。S-透镜RF水平设置为50，以使来自我们消化的肽所占据的m/z区达到最佳传输。从片段化选择中排除具有单个、未分配或六个和更高电荷状态的前体离子。

[0221] 数据处理

[0222] BioPharma Finder 1.0软件 (Thermo Fisher Scientific) 用于分析所有获得的数据。对于肽作图，使用单进入蛋白FASTA数据库进行搜索，其中脲基甲基化设置为固定的修饰；并将氧化、脱酰胺化和磷酸化设置为可变修饰、10ppm质量准确度、高蛋白酶特异性和置信水平为0.8的MS/MS光谱。通过将修饰肽的质量面积除以修饰肽和天然肽的面积总和来确定肽的修饰百分比。考虑到可能的修饰位点的数量，在不同位点处被修饰的同量异位物质可以在单个峰中共迁移。因此，源自具有多个潜在的修饰位点的肽的片段离子可以用于

定位或区分多个修饰位点。在这些情况下,观察到的同位素图案内的相对强度可以用于特异性确定不同的修饰肽异构体的相对丰度。此方法假设所有异构物种的片段化效率是相同的,并且在修饰位点上是独立的。此方法允许定义特定的修饰位点以及所涉及的潜在组合。

[0223] 实例2:AAV-SGA

[0224] 腺相关病毒(AAV)是单股DNA细小病毒,无致病性且免疫原性较弱,这使其成为基因疗法的有效候选载体。自从发现第一代AAV(AAV1-6)以来,我们的实验室一直在努力从各种高等灵长类动物物种中分离出大量病毒。此处鉴定的第二代AAV是使用基于批量PCR的技术分离的,所述技术使用针对特异于灵长类动物源性AAV基因组的保守区的引物。我们使用AAV-SGA探索了AAV在其天然哺乳动物宿主中的遗传变异(图1)。

[0225] AAV-SGA是一种强大的技术,可用于从混合群体中高精度分离单个病毒基因组。在本研究中,我们使用AAV-SGA从恒河猴组织标本中鉴定了新的AAV基因组。新病毒分离株具有遗传多样性,并且可分类成进化枝D、E和边缘进化枝(图2)。

[0226] 使用新型衣壳和先前鉴定的对照衣壳产生含有增强型GFP(eGFP)基因的载体。通过静脉内(IV)(图5A)和脑室内(ICV)(图5C)递送途径在小鼠中测试具有各种衣壳的载体。在心脏、骨骼肌、肝脏和脑组织中分析载体基因组的生物分布(图5B和图5D)。我们的小鼠研究表明,新型衣壳通常表现出特定于进化枝的转导模式(进化枝D衣壳除外)。确切地说,新型进化枝E载体AAVrh.90在IV和ICV递送后显示出外周器官的稳固转导。这通过肝切片的组织学证实,显示出高水平的GFP检测,在AAV8中也很明显。肌肉的组织学分析显示IV注射AAVrh.90载体后血管系统中的内皮细胞转导(通过用抗GFP、抗CD31(内皮细胞)和神经丝(神经元)染色的小鼠脑切片的免疫荧光成像确定)。

[0227] 进行额外研究以评估在IM递送之后肌肉组织的转导。具有各种衣壳并表达LacZ(图6A)或mAb的载体被IM递送并且通过肌纤维染色(对于LacZ)或在血清中检测(对于mAb)来分析转基因的表达。图6B示出通过检测LacZ对肌肉转导的比较。虽然进化枝A载体(AAV1、AAV6和AAVrh.91)高效转导肌肉纤维(染色较深),但观察到AAVrh.90的染色很少或没有染色。通过AAVrh.90进行IM递送导致血清中可检测的mAb水平与AAV8相当(图7)。图8示出mAb和LacZ载体的各种制剂的产率。对于两种转基因,与AAV8相比,AAVrh.90的产率均更高。

[0228] 实例3:使用带条形码的转基因系统对非人类灵长类动物的新型AAV天然分离物进行转导评估

[0229] 腺相关病毒(AAV)载体已被证明是临床应用中安全且有效的基因转移载体,但它们可能会受到预先存在的对病毒的免疫力的阻碍,并且可能具有受限的组织向性。我们证明了带条形码的转基因方法可以有效地同时比较多种AAV血清型对单一动物中各种组织的转导。这种技术减少了使用的动物数量,并防止了与外来转基因相关的免疫反应。因此,在转录本的polyA信号之前,将新型衣壳和其各自的原型进化枝成员对照(AAV6.2、AAV7、AAV8、AAVrh32.33和AAV9)制成包含修饰的eGFP转基因和独特的六碱基对条形码的载体(图9)。通过删除ATG序列基序来修饰转基因,以防止多肽翻译和随后对外来蛋白质的免疫反应。载体以等量汇集并在食蟹猴体内IV或ICM注射(总剂量:2e13 GC/kg IV和3e13 GC ICM)以评估新型衣壳的全身和中枢神经系统转导模式。IV注射的动物在基线时对AAV6、AAV8和AAVrh32.33呈血清阴性,并且针对AAV7和AAV9的中和抗体滴度分别为1:5和1:10。

[0230] AAVrh.90和AAVrh.91在IV递送后均以高效率转导外周器官(图10A)。IV递送后的

组织分析表明,AAVrh.90衣壳在NHP肝脏、心脏、骨骼肌、肾脏和胰腺中的表达水平高于所有其它测试的衣壳(新型和对照)(图10A)。AAVrh.90还以比IV递送后测试的其它衣壳更高的效率转导CNS组织(图10B)。

[0231] AAVrh.90在ICM注射后未显示NHP脑中转导水平的增加(相对于其它测试的载体,包括AAV8)。组织学研究还表明,AAVrh.90以低于AAV9的水平转导神经元和星形胶质细胞,并且不以3e13 GC/动物的剂量转导排列在脑室中的室管膜细胞。这与新型进化枝A变体AAVrh.91形成对比,与AAV9相比,AAVrh.91以高水平转导神经元和星形胶质细胞,并在脊髓中显示出稳固的运动神经元转导。IHC染色还显示AAVrh.91和AAV1都显示出对排列在脑室中的室管膜细胞的有效转导。

[0232] 对新型和对照衣壳的小规模制剂的产率的评估表明,AAVrh.90显示出与AAV8相当的产率。(图11)。如先前所述地分析AAVrh.90衣壳的脱酰胺化和其它修饰(参见PCT/US19/019804和PCT/US19/2019/019861)。如图12A和图12B所示,结果表明AAVrh.90具有四个高度脱酰胺化的氨基酸(N57、N263、N385、N514),其对应于天冬酰胺-甘氨酸对中的天冬酰胺(SEQ ID NO:2的AAVrh.90的编号)。在残基N94、N305、N499和N599以及S149处的磷酸化中始终观察到较低的脱酰胺化百分比。

[0233] (序列表自由文本)

[0234] 针对含有在数字标识符<223>下的自由文本的序列提供以下信息。

SEQ ID NO	<223>下的自由文本
5	<223>引物序列
6	<223>引物序列
7	<223>引物序列
8	<223>引物序列
9	<223>miRNA靶序列
10	<223>miRNA靶序列

[0236] 本说明书中引用的所有文献均以引用的方式并入本文中。于2019年10月21日提交的美国临时专利申请第62/924,095号、于2019年10月10日提交的美国临时专利申请第62/913,314号以及于2019年4月29日提交的美国临时专利申请第62/840,184号连同其序列表以全文引用的方式并入。随此提交的名称为“19-8901PCT1\_ST25.txt”的序列表和其中的序列和文本以引用的方式并入。虽然已经参考特定实施例描述了本发明,但应当理解,可以在不脱离本发明的精神的情况下进行修改。此类修改意图属于所附权利要求的范围内。

## 序列表

<110>	宾夕法尼亚大学信托人		
<120>	新型 AAV 衣壳和含有其的组合物		
<130>	19-8901PCT1		
<150>	US 62/924, 095		
<151>	2019-10-21		
<150>	US 62/913, 314		
<151>	2019-10-10		
<150>	US 62/840, 184		
<151>	2019-04-29		
<160>	10		
<170>	PatentIn 版本 3.5		
<210>	1		
<211>	2217		
<212>	DNA		
<213>	腺相关病毒 rh. 90		
<220>			
<221>	CDS		
<222>	(1)..(2217)		
<400>	1		
	atg gct gcc gat ggt tat ctt cca gat tgg ctc gag gac aac ctc tct		48
[0001]	Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser		
	1 5 10 15		
	gag ggc att cgc gag tgg tgg gac ctg aaa cct gga gcc cca aaa ccc		96
	Glu Gly Ile Arg Glu Trp Trp Asp Leu Lys Pro Gly Ala Pro Lys Pro		
	20 25 30		
	aaa gcc aac cag caa aag cag gac gac ggc cgg ggt ctg gtg ctt cct		144
	Lys Ala Asn Gln Gln Lys Gln Asp Asp Gly Arg Gly Leu Val Leu Pro		
	35 40 45		
	ggc tac aag tac ctc gga ccc ttc aac gga ctc gac aag ggg gag ccc		192
	Gly Tyr Lys Tyr Leu Gly Pro Phe Asn Gly Leu Asp Lys Gly Glu Pro		
	50 55 60		
	gtc aac gcg gcg gac gca gcg gcc ctc gag cac gac aag gcc tac gac		240
	Val Asn Ala Ala Asp Ala Ala Leu Glu His Asp Lys Ala Tyr Asp		
	65 70 75 80		
	cag cag ctg cag gcg ggt gat aat ccg tac ctg cgg tat aac cac gcc		288
	Gln Gln Leu Gln Ala Gly Asp Asn Pro Tyr Leu Arg Tyr Asn His Ala		
	85 90 95		
	gac gcc gag ttt cag gag cgt ctg caa gaa gat acg tca ttt ggg ggc		336
	Asp Ala Glu Phe Gln Glu Arg Leu Gln Glu Asp Thr Ser Phe Gly Gly		
	100 105 110		
	aac ctc ggg cga gca gtc ttc cag gcc aag aag cgg gtt ctc gaa cct		384
	Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Val Leu Glu Pro		
	115 120 125		

	ctc ggt ctg gtt gag gaa ggc gct aag acg gct cct gga aag aag aga		432
	Leu Gly Leu Val Glu Glu Gly Ala Lys Thr Ala Pro Gly Lys Lys Arg		
130	135	140	
	ccg gta gag cca tca cca cag cgt tcc ccc gac tcc tcc acg ggc atc		480
	Pro Val Glu Pro Ser Pro Gln Arg Ser Pro Asp Ser Ser Thr Gly Ile		
145	150	155	160
	ggc aag aaa ggc cag cag ccc gcc aga aag aga ctc aat ttc ggt cag		528
	Gly Lys Lys Gly Gln Gln Pro Ala Arg Lys Arg Leu Asn Phe Gly Gln		
	165	170	175
	act ggc gac tca gag tca gtc ccc gac cct caa cct ctc gga gaa cct		576
	Thr Gly Asp Ser Glu Ser Val Pro Asp Pro Gln Pro Leu Gly Glu Pro		
	180	185	190
	cca gca gcg ccc tct agt gtg gga tct ggt aca atg gct gca ggc ggt		624
	Pro Ala Ala Pro Ser Ser Val Gly Ser Gly Thr Met Ala Ala Gly Gly		
	195	200	205
	ggc gca cca atg gca gac aat aac gaa ggt gcc gac gga gtg ggt agt		672
	Gly Ala Pro Met Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser		
	210	215	220
	tcc tcg gga aat tgg cat tgc gat tcc aca tgg ctg ggc gac aga gtc		720
	Ser Ser Gly Asn Trp His Cys Asp Ser Thr Trp Leu Gly Asp Arg Val		
225	230	235	240
	atc acc acc agc acc cga acc tgg gcc ctg ccc acc tac aac aac cac		768
	Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His		
	245	250	255
	ctc tac aag caa atc tcc aac ggg acc tcg gga ggc agc acc aac gac		816
	Leu Tyr Lys Gln Ile Ser Asn Gly Thr Ser Gly Ser Thr Asn Asp		
	260	265	270
	aac acc tac ttc ggc tac agc acc ccc tgg ggg tat ttt gac ttt aac		864
	Asn Thr Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn		
	275	280	285
	aga ttc cac tgc cac ttc tca cca cgt gac tgg cag cga ctt atc aac		912
	Arg Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn		
290	295	300	
	aac aac tgg gga ttc cgg ccc aag aga ctc agc ttc aag ctc ttc aac		960
	Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu Ser Phe Lys Leu Phe Asn		
305	310	315	320
	atc cag gtc aag gag gtc acg caa aat gaa ggc acc aag acc atc gcc		1008
	Ile Gln Val Lys Glu Val Thr Gln Asn Glu Gly Thr Lys Thr Ile Ala		
	325	330	335
	aat aac ctc acc agc acc atc cag gtg ttt acg gac tcg gaa tac cag		1056
	Asn Asn Leu Thr Ser Thr Ile Gln Val Phe Thr Asp Ser Glu Tyr Gln		
	340	345	350
	ctg cca tac gtc ctc ggc tct gcc cac cag ggc tgc ctg cct ccg ttc		1104
	Leu Pro Tyr Val Leu Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe		

	355	360	365	
	ccg gcg gac gtc ttc atg att cct cag tat ggc tac ctg acg ctg aac			1152
	Pro Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn			
	370	375	380	
	aac gga agt cag gcc gtg ggc cgt tcc tcc tac tgc ctg gag tac			1200
	Asn Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr			
	385	390	395	400
	ttt ccc tct cag atg cta aga acg ggc aac aac ttc tcc ttc agc tat			1248
	Phe Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Ser Phe Ser Tyr			
	405	410	415	
	acc ttc gag gac gtg cct ttc cac agc agc tac gcg cac agc cag agc			1296
	Thr Phe Glu Asp Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser			
	420	425	430	
	ctg gac cgg ctg atg aat ccc ctc att gac cag tac ctg tac tac ctg			1344
	Leu Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu			
	435	440	445	
	tcg cgg aca caa tcc aca gga ggc aca gcg gga act cag cag ttg ctg			1392
	Ser Arg Thr Gln Ser Thr Gly Gly Thr Ala Gly Thr Gln Gln Leu Leu			
	450	455	460	
	ttt tct cag gcc ggg cct aac aac atg tct gct cag gcc aaa aac tgg			1440
	Phe Ser Gln Ala Gly Pro Asn Asn Met Ser Ala Gln Ala Lys Asn Trp			
[0003]	465	470	475	480
	cta ccc gga cct tgt tat cgg cag caa cgt gtt tcc acg aca ctg tcg			1488
	Leu Pro Gly Pro Cys Tyr Arg Gln Gln Arg Val Ser Thr Thr Leu Ser			
	485	490	495	
	caa aac aac aac agc aac ttt gcc tgg acc ggt gcc acc aaa tac cac			1536
	Gln Asn Asn Ser Asn Phe Ala Trp Thr Gly Ala Thr Lys Tyr His			
	500	505	510	
	ctg aac gga aga gac tct ctg gta aat ccg ggt gtc gcc atg gca acc			1584
	Leu Asn Gly Arg Asp Ser Leu Val Asn Pro Gly Val Ala Met Ala Thr			
	515	520	525	
	aac aag gac gac gag gac cgc ttc ttc cca tcc agc ggc atc ctc atg			1632
	Asn Lys Asp Asp Glu Asp Arg Phe Phe Pro Ser Ser Gly Ile Leu Met			
	530	535	540	
	ttt ggc aag cag ggg gct gga aaa gac aac gtg gac tat agc aac gtg			1680
	Phe Gly Lys Gln Gly Ala Gly Lys Asp Asn Val Asp Tyr Ser Asn Val			
	545	550	555	560
	atg cta acc agc gag gaa gaa atc aag acc act aac cct gtg gcc aca			1728
	Met Leu Thr Ser Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr			
	565	570	575	
	gaa cag tat ggt gtg gtg gcg gat aac ctg cag cag caa aac aca gct			1776
	Glu Gln Tyr Gly Val Val Ala Asp Asn Leu Gln Gln Asn Thr Ala			
	580	585	590	
	cct att gtg ggg gcc gtc aac agc caa gga gcc tta cct ggc atg gtt			1824

Pro Ile Val Gly Ala Val Asn Ser Gln Gly Ala Leu Pro Gly Met Val			
595	600	605	
tgg cag aac cgg gac gtc tac ctg cag ggt ccc atc tgg gcc aag att			1872
Trp Gln Asn Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile			
610	615	620	
cct cac acg gat ggt aac ttt cac ccg tct ctc atg ggc ggc ttt			1920
Pro His Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe			
625	630	635	640
gga ctt aaa cat ccg cct cct cag atc ctg atc aag aac act ccc gtt			1968
Gly Leu Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val			
645	650	655	
cct gcg gat cct cca acg gcg ttc aac cag gcc aag ctg aac tct ttc			2016
Pro Ala Asp Pro Pro Thr Ala Phe Asn Gln Ala Lys Leu Asn Ser Phe			
660	665	670	
atc acg cag tac agc acc gga caa gtc agc gtc gag atc gag tgg gag			2064
Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu			
675	680	685	
ctg cag aag gag aac agc aag cgc tgg aac cca gag att cag tat acc			2112
Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr			
690	695	700	
tcc aac tac tac aaa tct aca aat gtc gac ttt gct gtt aat act gag			2160
Ser Asn Tyr Tyr Lys Ser Thr Asn Val Asp Phe Ala Val Asn Thr Glu			
705	710	715	720
ggt gtt tac tct gag cct cgc ccc att ggc acc cgt tac ctc acc cgt			2208
Gly Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg			
725	730	735	
aat ctg taa			2217
Asn Leu			
<210> 2			
<211> 738			
<212> PRT			
<213> 腺相关病毒 rh.90			
<400> 2			
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asn Leu Ser			
1	5	10	15
Glu Gly Ile Arg Glu Trp Trp Asp Leu Lys Pro Gly Ala Pro Lys Pro			
20	25	30	
Lys Ala Asn Gln Gln Lys Gln Asp Asp Gly 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 Ala Ala Asp Ala Ala Leu Glu His Asp Lys Ala Tyr Asp			
65	70	75	80

Gln Gln Leu Gln Ala Gly Asp Asn Pro Tyr Leu Arg Tyr Asn His Ala  
                   85                     90                 95  
 Asp Ala Glu Phe Gln Glu Arg Leu Gln 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 Gly Ala Lys Thr Ala Pro Gly Lys Lys Arg  
                   130                135                140  
 Pro Val Glu Pro Ser Pro Gln Arg Ser Pro Asp Ser Ser Thr Gly Ile  
                   145                150                155                160  
 Gly Lys Lys Gly Gln Gln Pro Ala Arg Lys Arg Leu Asn Phe Gly Gln  
                   165                170                175  
 Thr Gly Asp Ser Glu Ser Val Pro Asp Pro Gln Pro Leu Gly Glu Pro  
                   180                185                190  
 Pro Ala Ala Pro Ser Ser Val Gly Ser Gly Thr Met Ala Ala Gly Gly  
                   195                200                205  
 Gly Ala Pro Met Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser  
                   210                215                220  
 Ser Ser Gly Asn Trp His Cys Asp Ser Thr Trp Leu Gly Asp Arg Val  
                   225                230                235                240  
 Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His  
                   245                250                255  
[0005] Leu Tyr Lys Gln Ile Ser Asn Gly Thr Ser Gly Gly Ser Thr Asn Asp  
                   260                265                270  
 Asn Thr Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn  
                   275                280                285  
 Arg Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn  
                   290                295                300  
 Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu Ser Phe Lys Leu Phe Asn  
                   305                310                315                320  
 Ile Gln Val Lys Glu Val Thr Gln Asn Glu Gly Thr Lys Thr Ile Ala  
                   325                330                335  
 Asn Asn Leu Thr Ser Thr Ile Gln Val Phe Thr Asp Ser Glu Tyr Gln  
                   340                345                350  
 Leu Pro Tyr Val Leu Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe  
                   355                360                365  
 Pro Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn  
                   370                375                380  
 Asn Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr  
                   385                390                395                400  
 Phe Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Ser Phe Ser Tyr  
                   405                410                415  
 Thr Phe Glu Asp Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser  
                   420                425                430

Leu Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu  
           435                  440                  445  
 Ser Arg Thr Gln Ser Thr Gly Gly Thr Ala Gly Thr Gln Gln Leu Leu  
           450                  455                  460  
 Phe Ser Gln Ala Gly Pro Asn Asn Met Ser Ala Gln Ala Lys Asn Trp  
           465                  470                  475                  480  
 Leu Pro Gly Pro Cys Tyr Arg Gln Gln Arg Val Ser Thr Thr Leu Ser  
           485                  490                  495  
 Gln Asn Asn Asn Ser Asn Phe Ala Trp Thr Gly Ala Thr Lys Tyr His  
           500                  505                  510  
 Leu Asn Gly Arg Asp Ser Leu Val Asn Pro Gly Val Ala Met Ala Thr  
           515                  520                  525  
 Asn Lys Asp Asp Glu Asp Arg Phe Phe Pro Ser Ser Gly Ile Leu Met  
           530                  535                  540  
 Phe Gly Lys Gln Gly Ala Gly Lys Asp Asn Val Asp Tyr Ser Asn Val  
           545                  550                  555                  560  
 Met Leu Thr Ser Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr  
           565                  570                  575  
 Glu Gln Tyr Gly Val Val Ala Asp Asn Leu Gln Gln Gln Asn Thr Ala  
           580                  585                  590  
 Pro Ile Val Gly Ala Val Asn Ser Gln Gly Ala Leu Pro Gly Met Val  
           595                  600                  605  
 [0006] Trp Gln Asn Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile  
           610                  615                  620  
 Pro His Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe  
           625                  630                  635                  640  
 Gly Leu Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val  
           645                  650                  655  
 Pro Ala Asp Pro Pro Thr Ala Phe Asn Gln Ala Lys Leu Asn Ser Phe  
           660                  665                  670  
 Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu  
           675                  680                  685  
 Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr  
           690                  695                  700  
 Ser Asn Tyr Tyr Lys Ser Thr Asn Val Asp Phe Ala Val Asn Thr Glu  
           705                  710                  715                  720  
 Gly Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg  
           725                  730                  735  
 Asn Leu

<210> 3  
 <211> 2217  
 <212> DNA  
 <213> 腺相关病毒 8

<220>				
<221> CDS				
<222> (1)..(2217)				
<400> 3				
atg gct gcc gat ggt tat ctt cca gat tgg ctc gag gac aac ctc tct				48
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser				
1 5 10 15				
gag ggc att cgc gag tgg tgg gcg ctg aaa cct gga gcc ccg aag ccc				96
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Lys Pro				
20 25 30				
aaa gcc aac cag caa aag cag gac gac ggc cgg ggt ctg gtg ctt cct				144
Lys Ala Asn Gln Gln Lys Gln Asp Asp Gly Arg Gly Leu Val Leu Pro				
35 40 45				
ggc tac aag tac ctc gga ccc ttc aac gga ctc gac aag ggg gag ccc				192
Gly Tyr Lys Tyr Leu Gly Pro Phe Asn Gly Leu Asp Lys Gly Glu Pro				
50 55 60				
gtc aac gcg gcg gac gca gcg gcc ctc gag cac gac aag gcc tac gac				240
Val Asn Ala Ala Asp Ala Ala Leu Glu His Asp Lys Ala Tyr Asp				
65 70 75 80				
cag cag ctg cag gcg ggt gac aat ccg tac ctg cgg tat aac cac gcc				288
Gln Gln Leu Gln Ala Gly Asp Asn Pro Tyr Leu Arg Tyr Asn His Ala				
85 90 95				
[0007] gac gcc gag ttt cag gag cgt ctg caa gaa gat acg tct ttt ggg ggc				336
Asp Ala Glu Phe Gln Glu Arg Leu Gln Glu Asp Thr Ser Phe Gly Gly				
100 105 110				
aac ctc ggg cga gca gtc ttc cag gcc aag aag cgg gtt ctc gaa cct				384
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Val Leu Glu Pro				
115 120 125				
ctc ggt ctg gtt gag gaa ggc gct aag acg gct cct gga aag aag aga				432
Leu Gly Leu Val Glu Glu Gly Ala Lys Thr Ala Pro Gly Lys Lys Arg				
130 135 140				
ccg gta gag cca tca ccc cag cgt tct cca gac tcc tct acg ggc atc				480
Pro Val Glu Pro Ser Pro Gln Arg Ser Pro Asp Ser Ser Thr Gly Ile				
145 150 155 160				
ggc aag aaa ggc caa cag ccc gcc aga aaa aga ctc aat ttt ggt cag				528
Gly Lys Lys Gly Gln Gln Pro Ala Arg Lys Arg Leu Asn Phe Gly Gln				
165 170 175				
act ggc gac tca gag tca gtt cca gac cct caa cct ctc gga gaa cct				576
Thr Gly Asp Ser Glu Ser Val Pro Asp Pro Gln Pro Leu Gly Glu Pro				
180 185 190				
cca gca gcg ccc tct ggt gtg gga cct aat aca atg gct gca ggc ggt				624
Pro Ala Ala Pro Ser Gly Val Gly Pro Asn Thr Met Ala Ala Gly Gly				
195 200 205				
ggc gca cca atg gca gac aat aac gaa ggc gcc gac gga gtg ggt agt				672

Gly Ala Pro Met Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser			
210	215	220	
tcc tcg gga aat tgg cat tgc gat tcc aca tgg ctg ggc gac aga gtc			720
Ser Ser Gly Asn Trp His Cys Asp Ser Thr Trp Leu Gly Asp Arg Val			
225	230	235	240
atc acc acc agc acc cga acc tgg gcc ctg ccc acc tac aac aac cac			768
Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His			
245	250	255	
ctc tac aag caa atc tcc aac ggg aca tcg gga gga gcc acc aac gac			816
Leu Tyr Lys Gln Ile Ser Asn Gly Thr Ser Gly Gly Ala Thr Asn Asp			
260	265	270	
aac acc tac ttc ggc tac agc acc ccc tgg ggg tat ttt gac ttt aac			864
Asn Thr Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn			
275	280	285	
aga ttc cac tgc cac ttt tca cca cgt gac tgg cag cga ctc atc aac			912
Arg Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn			
290	295	300	
aac aac tgg gga ttc cgg ccc aag aga ctc agc ttc aag ctc ttc aac			960
Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu Ser Phe Lys Leu Phe Asn			
305	310	315	320
atc cag gtc aag gag gtc acg cag aat gaa ggc acc aag acc atc gcc			1008
Ile Gln Val Lys Glu Val Thr Gln Asn Glu Gly Thr Lys Thr Ile Ala			
325	330	335	
aat aac ctc acc agc acc atc cag gtg ttt acg gac tcg gag tac cag			1056
Asn Asn Leu Thr Ser Thr Ile Gln Val Phe Thr Asp Ser Glu Tyr Gln			
340	345	350	
ctg ccg tac gtt ctc ggc tct gcc cac cag ggc tgc cct ccg ttc			1104
Leu Pro Tyr Val Leu Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe			
355	360	365	
ccg gcg gac gtg ttc atg att ccc cag tac ggc tac cta aca ctc aac			1152
Pro Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn			
370	375	380	
aac ggt agt cag gcc gtg gga cgc tcc tcc ttc tac tgc ctg gaa tac			1200
Asn Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr			
385	390	395	400
ttt cct tcg cag atg ctg aga acc ggc aac aac ttc cag ttt act tac			1248
Phe Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Thr Tyr			
405	410	415	
acc ttc gag gac gtg cct ttc cac agc agc tac gcc cac agc cag agc			1296
Thr Phe Glu Asp Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser			
420	425	430	
ttg gac cgg ctg atg aat cct ctg att gac cag tac ctg tac tac ttg			1344
Leu Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu			

	435	440	445		
	tct	cgg	act	caa aca aca gga ggc acg gca aat acg cag act ctg ggc	1392
	Ser	Arg	Thr	Gln Thr Thr Gly Gly Thr Ala Asn Thr Gln Thr Leu Gly	
	450	455	460		
	ttc	agc	caa	ggt ggg cct aat aca atg gcc aat cag gca aag aac tgg	1440
	Phe	Ser	Gln	Gly Gly Pro Asn Thr Met Ala Asn Gln Ala Lys Asn Trp	
	465	470	475	480	
	ctg	cca	gga	ccc tgt tac cgc caa caa cgc gtc tca acg aca acc ggg	1488
	Leu	Pro	Gly Pro Cys Tyr Arg Gln Gln Arg Val Ser Thr Thr Gly		
	485	490	495		
	caa	aac	aac aat	agc aac ttt gcc tgg act gct ggg acc aaa tac cat	1536
	Gln	Asn	Asn	Asn Ser Asn Phe Ala Trp Thr Ala Gly Thr Lys Tyr His	
	500	505	510		
	ctg	aat	gga	aga aat tca ttg gct aat cct ggc atc gct atg gca aca	1584
	Leu	Asn	Gly Arg Asn Ser Leu Ala Asn Pro Gly Ile Ala Met Ala Thr		
	515	520	525		
	cac	aaa	gac	gac gag gag cgt ttt ttt ccc agt aac ggg atc ctg att	1632
	His	Lys	Asp Asp Glu Glu Arg Phe Phe Pro Ser Asn Gly Ile Leu Ile		
	530	535	540		
	ttt	ggc	aaa	caa aat gct gcc aga gac aat gcg gat tac agc gat gtc	1680
	Phe	Gly Lys	Gln Asn Ala Ala Arg Asp Asn Ala Asp Tyr Ser Asp Val		
[0009]	545	550	555	560	
	atg	ctc	acc	agc gag gaa gaa atc aaa acc act aac cct gtg gct aca	1728
	Met	Leu	Thr Ser Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr		
	565	570	575		
	gag	gaa	tac	ggt atc gtg gca gat aac ttg cag cag caa aac acg gct	1776
	Glu	Glu	Tyr Gly Ile Val Ala Asp Asn Leu Gln Gln Asn Thr Ala		
	580	585	590		
	cct	caa	att	gga act gtc aac agc cag ggg gcc tta ccc ggt atg gtc	1824
	Pro	Gln	Ile Gly Thr Val Asn Ser Gln Gly Ala Leu Pro Gly Met Val		
	595	600	605		
	tgg	cag	aac	cgg gac gtg tac ctg cag ggt ccc atc tgg gcc aag att	1872
	Trp	Gln	Asn Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile		
	610	615	620		
	cct	cac	acg	gac ggc aac ttc cac ccg tct ccg ctg atg ggc ggc ttt	1920
	Pro	His	Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe		
	625	630	635	640	
	ggc	ctg	aaa	cat cct ccg cct cag atc ctg atc aag aac acg cct gta	1968
	Gly	Leu	Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val		
	645	650	655		
	cct	gcg	gat	cct ccg acc acc ttc aac cag tca aag ctg aac tct ttc	2016
	Pro	Ala	Asp Pro Pro Thr Thr Phe Asn Gln Ser Lys Leu Asn Ser Phe		
	660	665	670		
	atc	acg	caa	tac agc acc gga cag gtc agc gtg gaa att gaa tgg gag	2064

Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu			
675	680	685	
ctg cag aag gaa aac agc aag cgc tgg aac ccc gag atc cag tac acc			2112
Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr			
690	695	700	
tcc aac tac tac aaa tct aca agt gtg gac ttt gct gtt aat aca gaa			2160
Ser Asn Tyr Tyr Lys Ser Thr Ser Val Asp Phe Ala Val Asn Thr Glu			
705	710	715	720
ggc gtg tac tct gaa ccc cgc ccc att ggc acc cgt tac ctc acc cgt			2208
Gly Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg			
725	730	735	
aat ctg taa			2217
Asn Leu			
<210> 4			
<211> 738			
<212> PRT			
<213> 腺相关病毒 8			
<400> 4			
Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser			
1	5	10	15
[0010] Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Lys Pro			
20	25	30	
Lys Ala Asn Gln Gln Lys Gln Asp Asp Gly 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 Ala Ala Asp Ala Ala Leu Glu His Asp Lys Ala Tyr Asp			
65	70	75	80
Gln Gln Leu Gln Ala Gly Asp Asn Pro Tyr Leu Arg Tyr Asn His Ala			
85	90	95	
Asp Ala Glu Phe Gln Glu Arg Leu Gln 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 Gly Ala Lys Thr Ala Pro Gly Lys Lys Arg			
130	135	140	
Pro Val Glu Pro Ser Pro Gln Arg Ser Pro Asp Ser Ser Thr Gly Ile			
145	150	155	160
Gly Lys Lys Gly Gln Gln Pro Ala Arg Lys Arg Leu Asn Phe Gly Gln			
165	170	175	
Thr Gly Asp Ser Glu Ser Val Pro Asp Pro Gln Pro Leu Gly Glu Pro			
180	185	190	
Pro Ala Ala Pro Ser Gly Val Gly Pro Asn Thr Met Ala Ala Gly Gly			

195                    200                    205  
 Gly Ala Pro Met Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser  
 210                    215                    220  
 Ser Ser Gly Asn Trp His Cys Asp Ser Thr Trp Leu Gly Asp Arg Val  
 225                    230                    235                    240  
 Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His  
 245                    250                    255  
 Leu Tyr Lys Gln Ile Ser Asn Gly Thr Ser Gly Gly Ala Thr Asn Asp  
 260                    265                    270  
 Asn Thr Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn  
 275                    280                    285  
 Arg Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn  
 290                    295                    300  
 Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu Ser Phe Lys Leu Phe Asn  
 305                    310                    315                    320  
 Ile Gln Val Lys Glu Val Thr Gln Asn Glu Gly Thr Lys Thr Ile Ala  
 325                    330                    335  
 Asn Asn Leu Thr Ser Thr Ile Gln Val Phe Thr Asp Ser Glu Tyr Gln  
 340                    345                    350  
 Leu Pro Tyr Val Leu Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe  
 355                    360                    365  
 Pro Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn  
[0011] 370                    375                    380  
 Asn Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr  
 385                    390                    395                    400  
 Phe Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Thr Tyr  
 405                    410                    415  
 Thr Phe Glu Asp Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser  
 420                    425                    430  
 Leu Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu  
 435                    440                    445  
 Ser Arg Thr Gln Thr Thr Gly Gly Thr Ala Asn Thr Gln Thr Leu Gly  
 450                    455                    460  
 Phe Ser Gln Gly Pro Asn Thr Met Ala Asn Gln Ala Lys Asn Trp  
 465                    470                    475                    480  
 Leu Pro Gly Pro Cys Tyr Arg Gln Gln Arg Val Ser Thr Thr Thr Gly  
 485                    490                    495  
 Gln Asn Asn Asn Ser Asn Phe Ala Trp Thr Ala Gly Thr Lys Tyr His  
 500                    505                    510  
 Leu Asn Gly Arg Asn Ser Leu Ala Asn Pro Gly Ile Ala Met Ala Thr  
 515                    520                    525  
 His Lys Asp Asp Glu Glu Arg Phe Phe Pro Ser Asn Gly Ile Leu Ile  
 530                    535                    540  
 Phe Gly Lys Gln Asn Ala Ala Arg Asp Asn Ala Asp Tyr Ser Asp Val

545	550	555	560
Met Leu Thr Ser Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr			
565	570	575	
Glu Glu Tyr Gly Ile Val Ala Asp Asn Leu Gln Gln Gln Asn Thr Ala			
580	585	590	
Pro Gln Ile Gly Thr Val Asn Ser Gln Gly Ala Leu Pro Gly Met Val			
595	600	605	
Trp Gln Asn Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile			
610	615	620	
Pro His Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe			
625	630	635	640
Gly Leu Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val			
645	650	655	
Pro Ala Asp Pro Pro Thr Thr Phe Asn Gln Ser Lys Leu Asn Ser Phe			
660	665	670	
Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu			
675	680	685	
Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr			
690	695	700	
Ser Asn Tyr Tyr Lys Ser Thr Ser Val Asp Phe Ala Val Asn Thr Glu			
705	710	715	720
Gly Val Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg			
725	730	735	
[0012] Asn Leu			
<210> 5			
<211> 26			
<212> DNA			
<213> 人工序列			
<220>			
<223> 引物序列			
<400> 5			
gctgcgtcaa ctggaccaat gagaac			26
<210> 6			
<211> 28			
<212> DNA			
<213> 人工序列			
<220>			
<223> 引物序列			
<400> 6			
cgcagagacc aaagttaac tgaaacga			28
<210> 7			
<211> 25			
<212> DNA			

<213>	人工序列	
<220>		
<223>	引物序列	
<400>	7	
	ggcgaacagc ggacaccgat atgaa	25
<210>	8	
<211>	25	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	引物序列	
<400>	8	
	ggctctcgtc gcgtgagaat gagaa	25
<210>	9	
[0013]	<211> 22	
	<212> DNA	
	<213> 人工序列	
<220>		
<223>	miRNA 靶序列	
<400>	9	
	agtgaattct accagtgcc ta	22
<210>	10	
<211>	24	
<212>	DNA	
<213>	人工序列	
<220>		
<223>	miRNA 靶序列	
<400>	10	
	agtgtgagtt ctaccattgc caaa	24

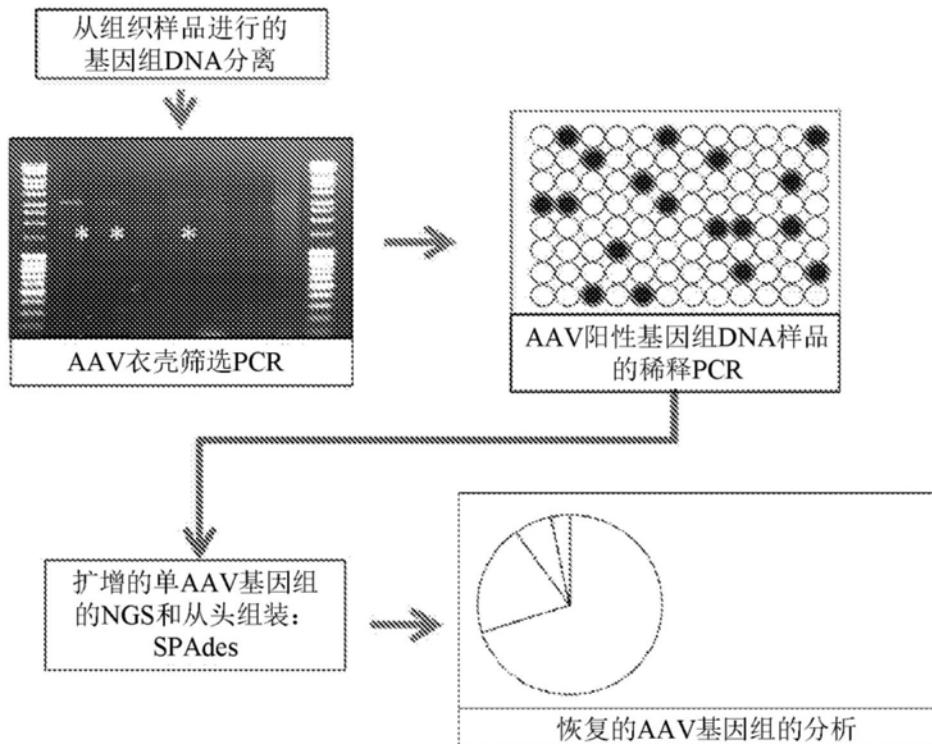


图1

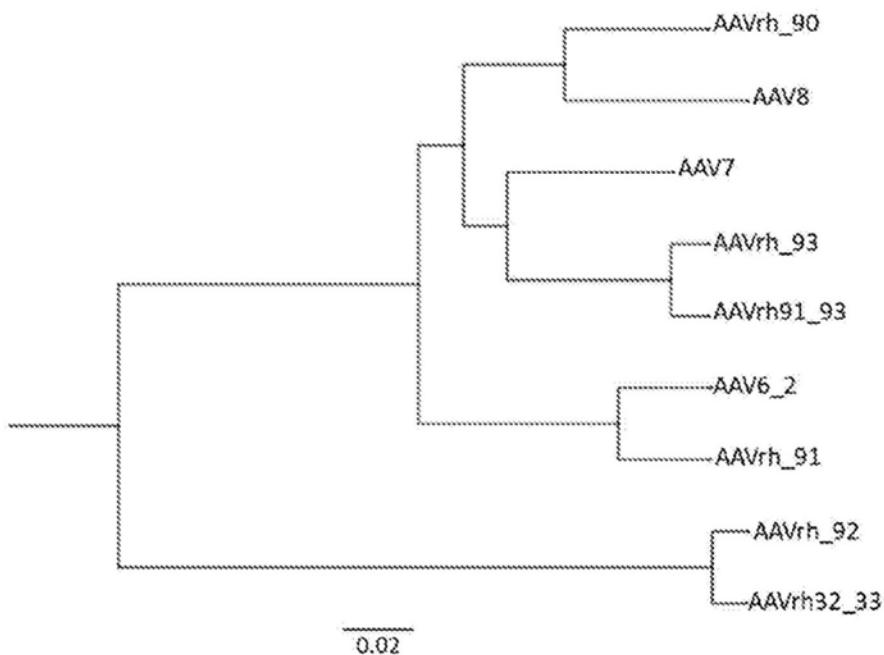


图2

图3A

AAVrh.90	ccatgggggttttgcacutttttaacagattttccactgcacatttttcaccacytgactggcag ccatgggggttttgcacutttttaacagattttccactgcacatttttcaccacgtgactggcag *****	900
AAVrh.90	cgacttatacavcaacaaatggggattttccggccasagagactcaagtttcaagcttttcaac cgacttatacavcaacaaatggggattttccggccasagagactcaagtttcaagcttttcaac *****	960
AAVrh.90	atcccaaggtaaaggagggttacggccaaatgtggggccatggccatggccatggccatggccat atcccaaggtaaaggagggttacggccaaatgtggggccatggccatggccatggccatggccat *****	1020
AAVrh.90	agcacccatccagggttttacggacttccggaggatccaggatgtgggttccgggttttccgggtt agcacccatccagggttttacggacttccggaggatccaggatgtgggttccgggttttccgggtt *****	1080
AAVrh.90	caccagggttgcactgcacccgtttccggggggacgtttttcatgtatccgtatggctac caccagggttgcactgcacccgtttccggggggacgtttttcatgtatccgtatggctac *****	1140
AAVrh.90	ctgacgttgcacaaatggggatccgggggttccggggggatccgtttttcatgtatccgtatggctac ctgacgttgcacaaatggggatccgggggttccggggggatccgtttttcatgtatccgtatggctac ***	1200
AAVrh.90	tttccccatcagatgttgcacaaatggggatccgggggttccggggggatccgtttttcatgtatccgtatgg tttccccatcagatgttgcacaaatggggatccggggggatccgtttttcatgtatccgtatgg *****	1260
AAVrh.90	gttgttttccacagcgttacccgttccggggatccggggatccgtttttcatgtatccgtatgg gttgttttccacagcgttccggggatccggggatccgtttttcatgtatccgtatgg *****	1320
AAVrh.90	attggacccatccgggttccggggatccggggatccggggatccggggatccggggatccggggatcc attggacccatccgggttccggggatccggggatccggggatccggggatccggggatccggggatcc *****	1380
AAVrh.90	caqcaqtgtgttttgcacggccgggttccggggatccggggatccggggatccggggatccggggatcc caqcaqtgtgttttgcacggccgggttccggggatccggggatccggggatccggggatcc ***	1440
AAVrh.90	cttccccggggatccgggttccggggatccggggatccggggatccggggatccggggatccggggatcc cttccccggggatccgggttccggggatccggggatccggggatccggggatccggggatcc **	1500
AAVrh.90	agcaacttttgcacccgttccggggatccggggatccggggatccggggatccggggatccggggatcc agcaacttttgcacccgttccggggatccggggatccggggatccggggatccggggatcc *****	1560
AAVrh.90	aatcccaagggttccggggatccggggatccggggatccggggatccggggatccggggatccggggatcc aatcccaagggttccggggatccggggatccggggatccggggatccggggatccggggatcc ****	1620
AAVrh.90	ggcatccatgttttgcacccgttccggggatccggggatccggggatccggggatccggggatcc ggcatccatgttttgcacccgttccggggatccggggatccggggatccggggatcc **	1680

图3B

图3C

vp1起点		
AAVrh.90	MAADGYILPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDGRGLVLPGYKYLGPFFNLD	60
AAV8	MAADGYILPDWLEDNLSEGIREWWALKPGAPKPKANQQKQDGRGLVLPGYKYLGPFFNLD	60
	*****	*****
AAVrh.90	KGEFPVNAAADAALALENDKAYDQQLQAGDNFYLRYNHADAEGQERLQEDTSFGGNLGPAYFQ	120
AAV8	KGEFPVNAAADAALALENDKAYDQQLQAGDNFYLRYNHADAEGQERLQEDTSFGGNLGPAYFQ	120
	*****	*****
AAVrh.90	AKKRVLEPLGLVEEGAKTAEGKKRVEPSPQRSPDSSTGIGKKGGQGPARKPLNPGQTDGS	180
AAV8	AKKRVLEPLGLVEEGAKTAEGKKRVEPSPQRSPDSSTGIGKKGGQGPARKPLNPGQTDGS	180
	*****	*****
AAVrh.90	ESVPDPQPLGEPPAAPS9VGSGTMAAGGGAGPMADNNNEGADGVGS3SGNWHCDSTWLGDGV	240
AAV8	ESVPDPQPLGEPPAAPS9VGSGTMAAGGGAGPMADNNNEGADGVGS3SGNWHCDSTWLGDGV	240
	*****	*****
AAVrh.90	ITTSTRTWALPTYNHILYKQISMNGTSGGATNDNTYFGYSTPWGYFDENRFHCNSFSPRDWQ	300
AAV8	ITTSTRTWALPTYNHILYKQISMNGTSGGATNDNTYFGYSTPWGYFDENRFHCNSFSPRDWQ	300
	*****	*****
AAVrh.90	RLLINNWGFRPKRLSTFKLIFIQVKEVTQNEGKTIANNLTSTIQVFTDSEYQLPFTVLGSA	360
AAV8	RLLINNWGFPKRLSTFKLIFIQVKEVTQNEGKTIANNLTSTIQVFTDSEYQLPFTVLGSA	360
	*****	*****
AAVrh.90	HQGCLPPPPADVFMIPQXGYLTLMNGSQAVGRSSFYCLEYFFPSQMLPTGNNFSFSYTFED	420
AAV8	HQGCLPPPPADVFMIPQXGYLTLMNGSQAVGRSSFYCLEYFFPSQMLPTGNNFSFSYTFED	420
	*****	*****
AAVrh.90	VPFHSSSYAH9QSLDRIMNPLIDQYLYLERTQSTTGTAGTQQLLPSQAGPNMSAQAKNW	480
AAV8	VPFHSSSYAH9QSLDRIMNPLIDQYLYLERTQTTGGTANTQTLGFSQGGPNTMANQAKNW	480
	*****	*****
AAVrh.90	LPGPCYRQQRVSTTLSQNNNSNFANTGATKYHLNGRDSLNVNPVAMATNKDDEURFFFSS	540
AAV8	LPGPCYRQQRVSTTLSQNNNSNFANTGATKYHLNGRNSSLANPGIAMATHKDDEURFFFSS	540
	*****	*****
AAVrh.90	GILMFQKQGMKINVDYSNVMILTSEECIKTTNPVATEQYGVVADNLQGQNTAFIVGAVNS	600
AAV8	GILMFQKQGMKINVDYSNVMILTSEECIKTTNPVATEQYGVVADNLQGQNTAFIVGAVNS	600
	*****	*****
AAVrh.90	QQALPGMVWQNRDVYLQGPWAKIPHTDNFHPSPLMGGFGLKHFFFQILIKNTPVPAQF	660
AAV8	QQALPGMVWQNRDVYLQGPWAKIPHTDNFHPSPLMGGFGLKHFFFQILIKNTPVPAQF	660
	*****	*****
AAVrh.90	PTAFNQAKLNSFITQYSTCQSVSEIEWEIQLKENSKRWNPEIQTTSNYYKSTNVDFAVNTE	720
AAV8	PTTFNQSKLNSFITQYSTCQSVSEIEWEIQLKENSKPWNPEIQTTSNYYKSTSVDFAVNTE	720
	*****	*****
AAVrh.90	GYVSEPRFIGTRYLTFNL	738
AAV8	GYVSEPRFIGTRYLTFNL	738
	*****	

图4

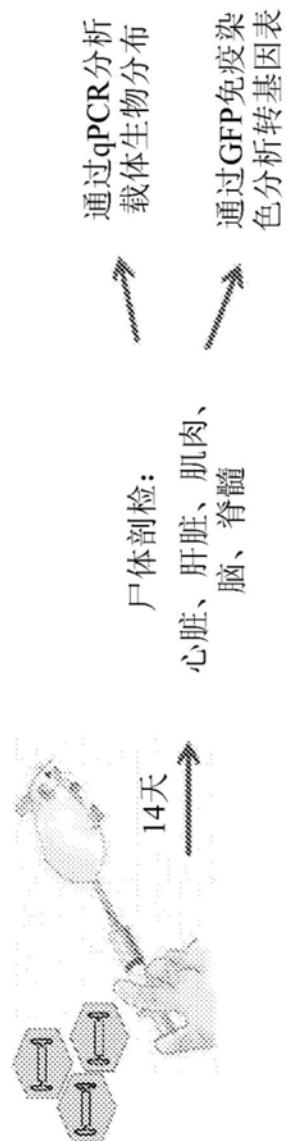


图5A

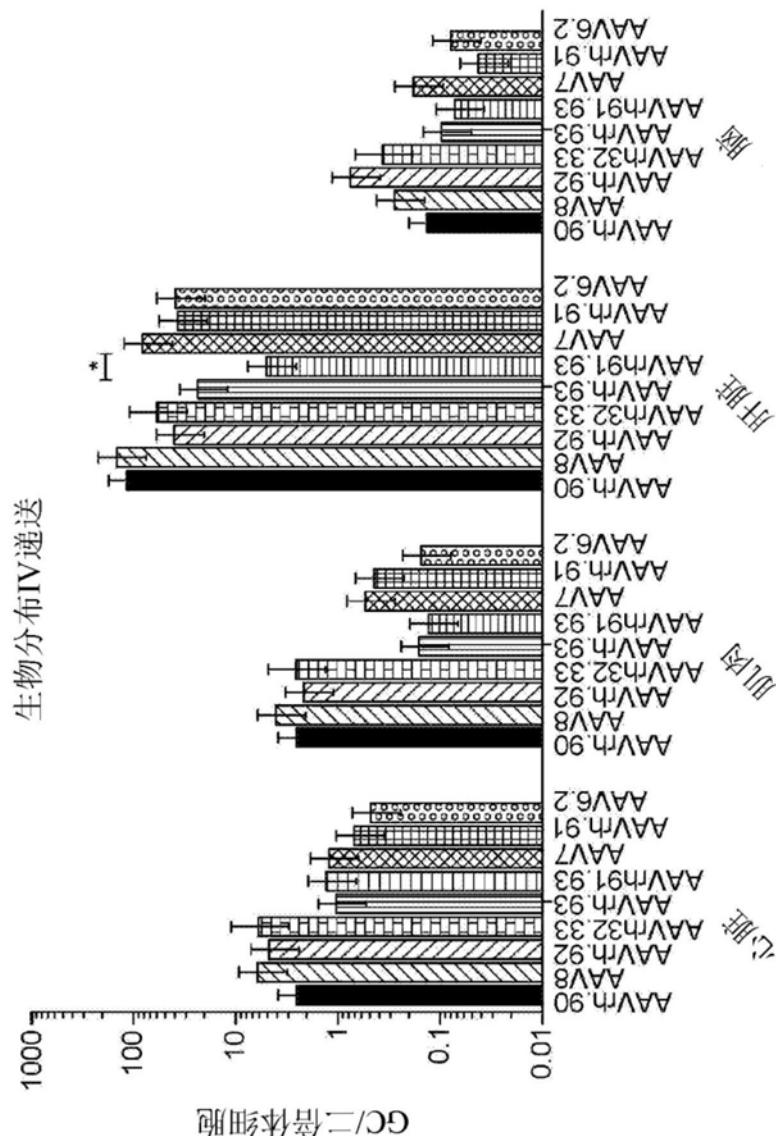


图5B

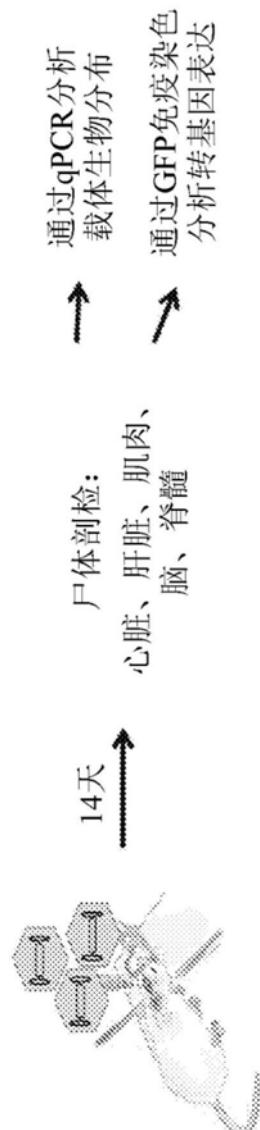


图5C

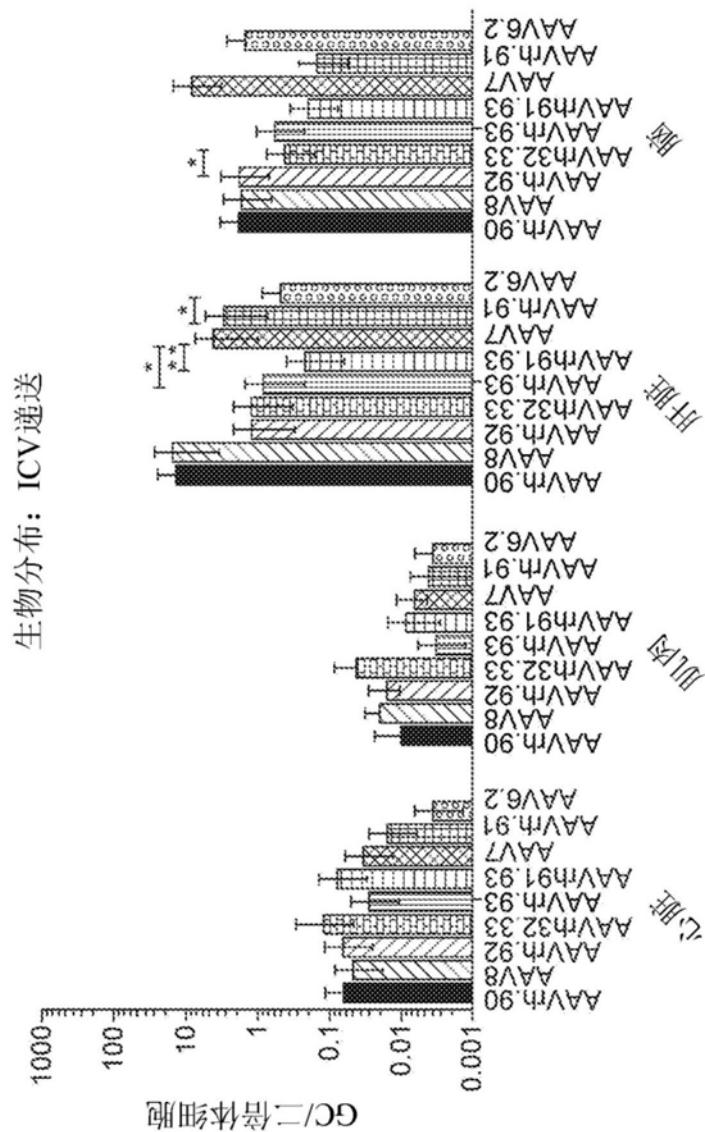


图5D

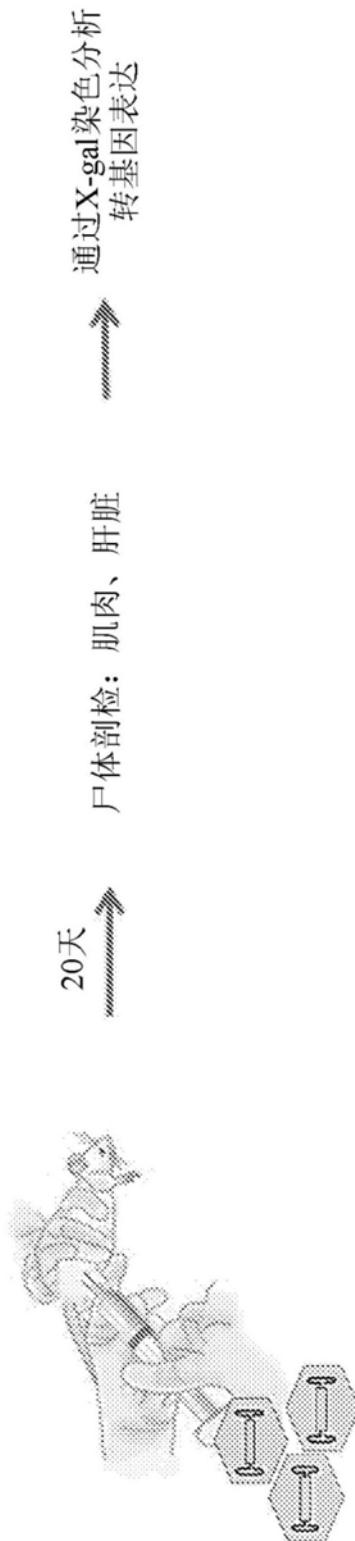


图6A

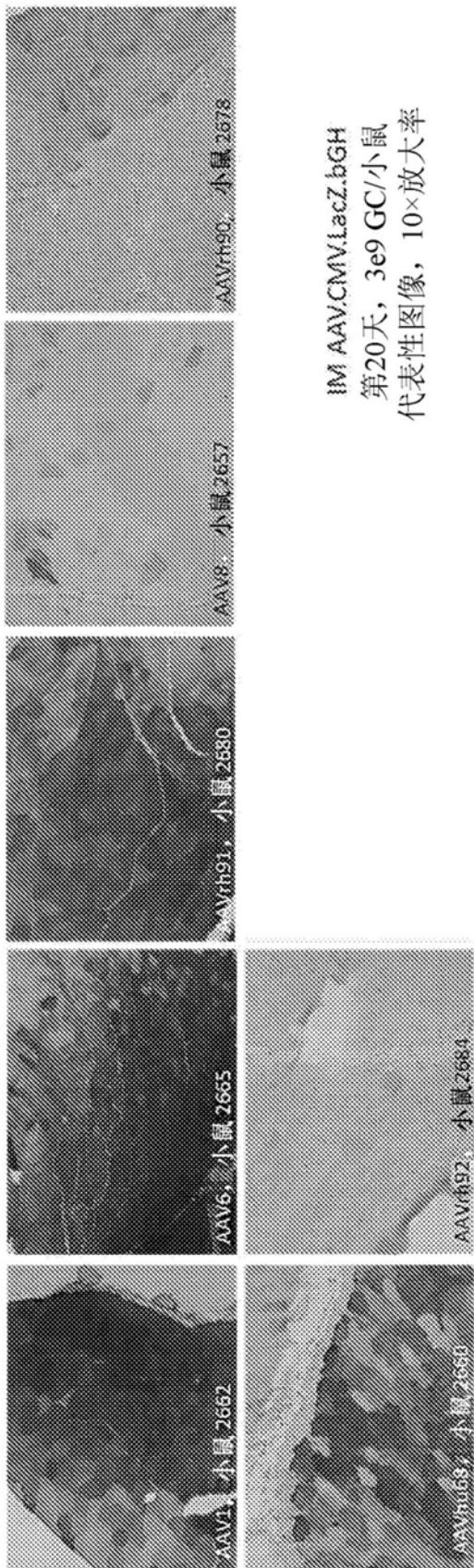
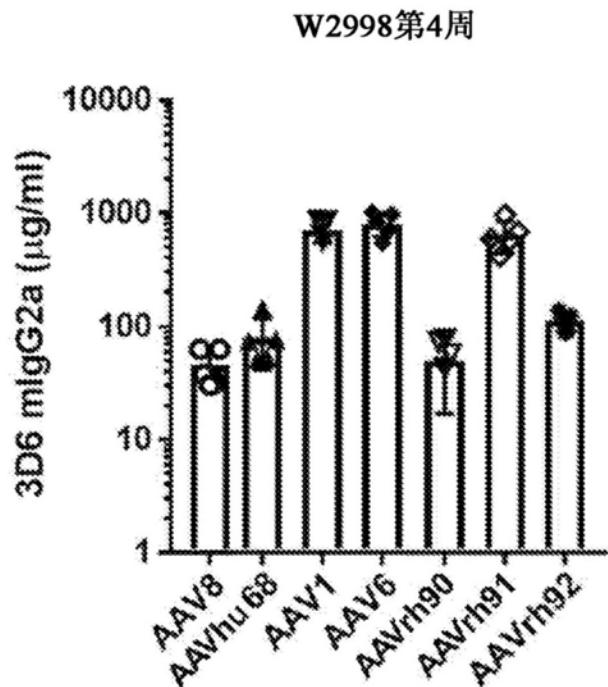


图6B



转基因      3D6 (小鼠mAb)

启动子      tMCK (肌肉选择性)

动物      WT B6小鼠 (5只/衣壳)  
途径      IM  
剂量      1e11 gc/小鼠  
读数      血清3D6 (每周)

图7

衣壳	产率 (% AAV8)	产率 (% AAV8)	来源
	(mAb转基因)	(LacZ转基因)	
AAV8	100%	100%	天然分离物
AAVhu68	76%	89%	天然分离物
AAV1	26%	28%	天然分离物
AAV6	27%	28%	天然分离物
AAVrh90	105%	118%	天然分离物
AAVrh91	53%	74%	天然分离物
AAVrh92	68%	102%	天然分离物

图8

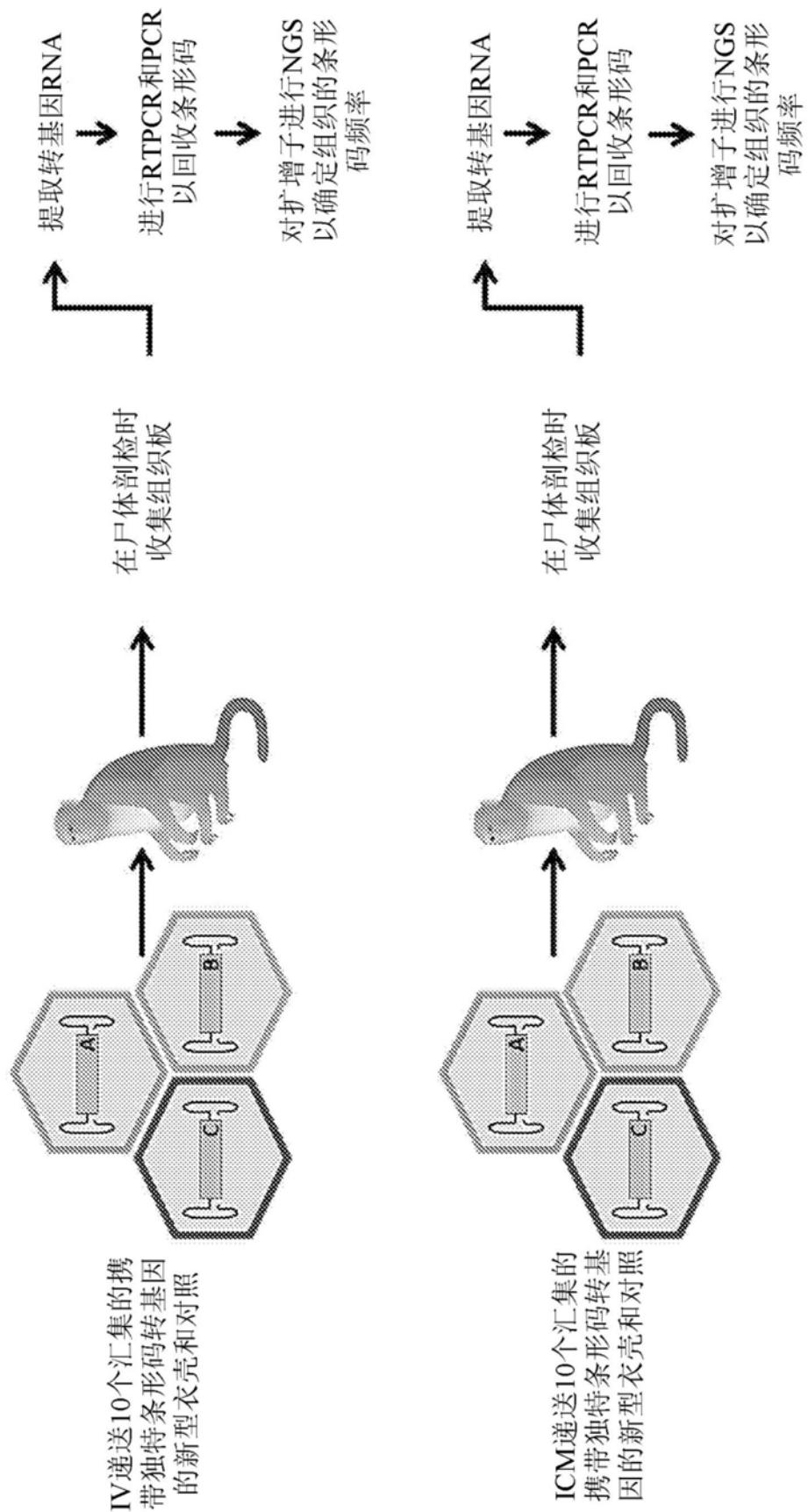


图9

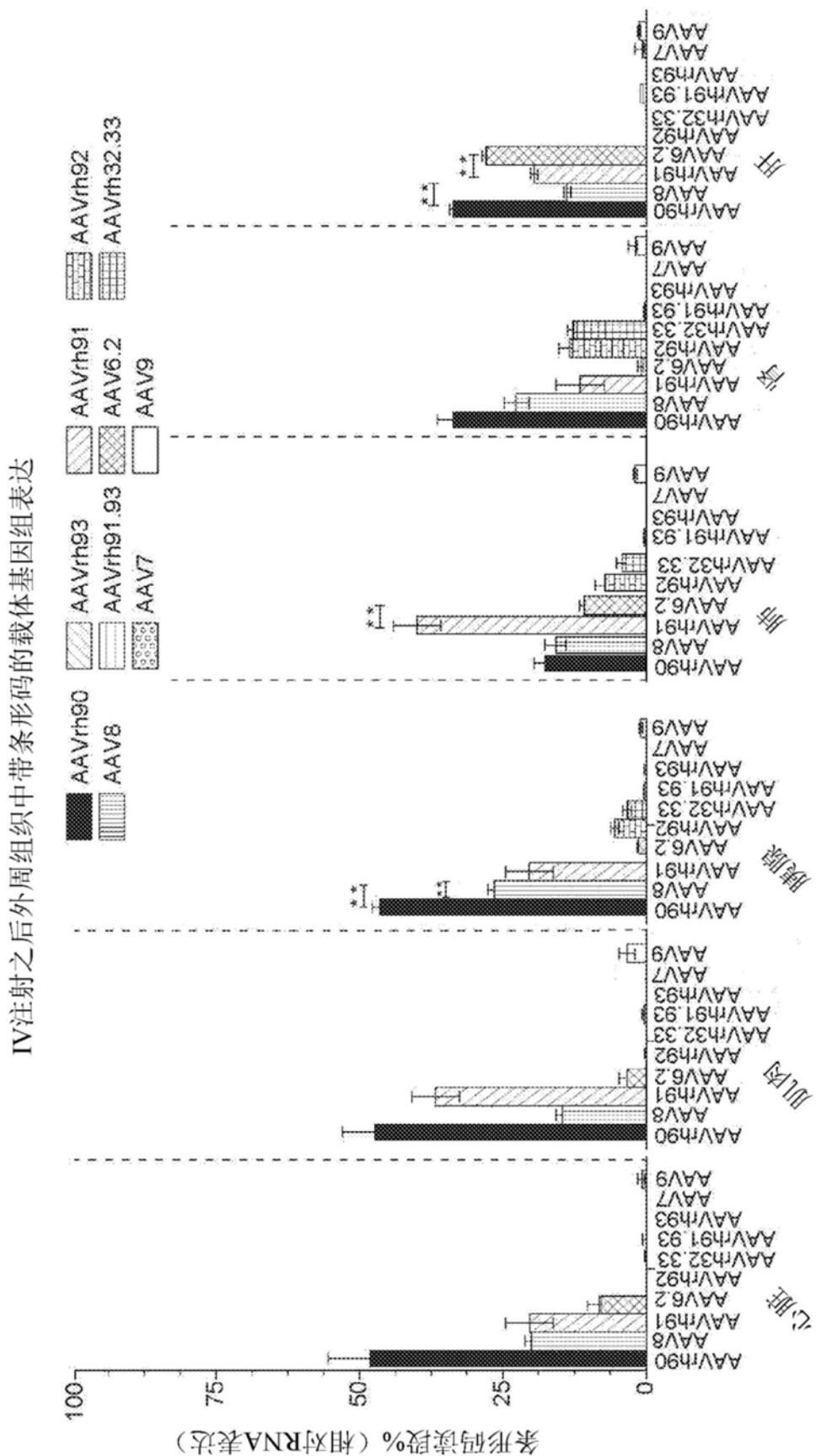


图10A

IV注射之后CNS组织中带条形码的载体基因组表达

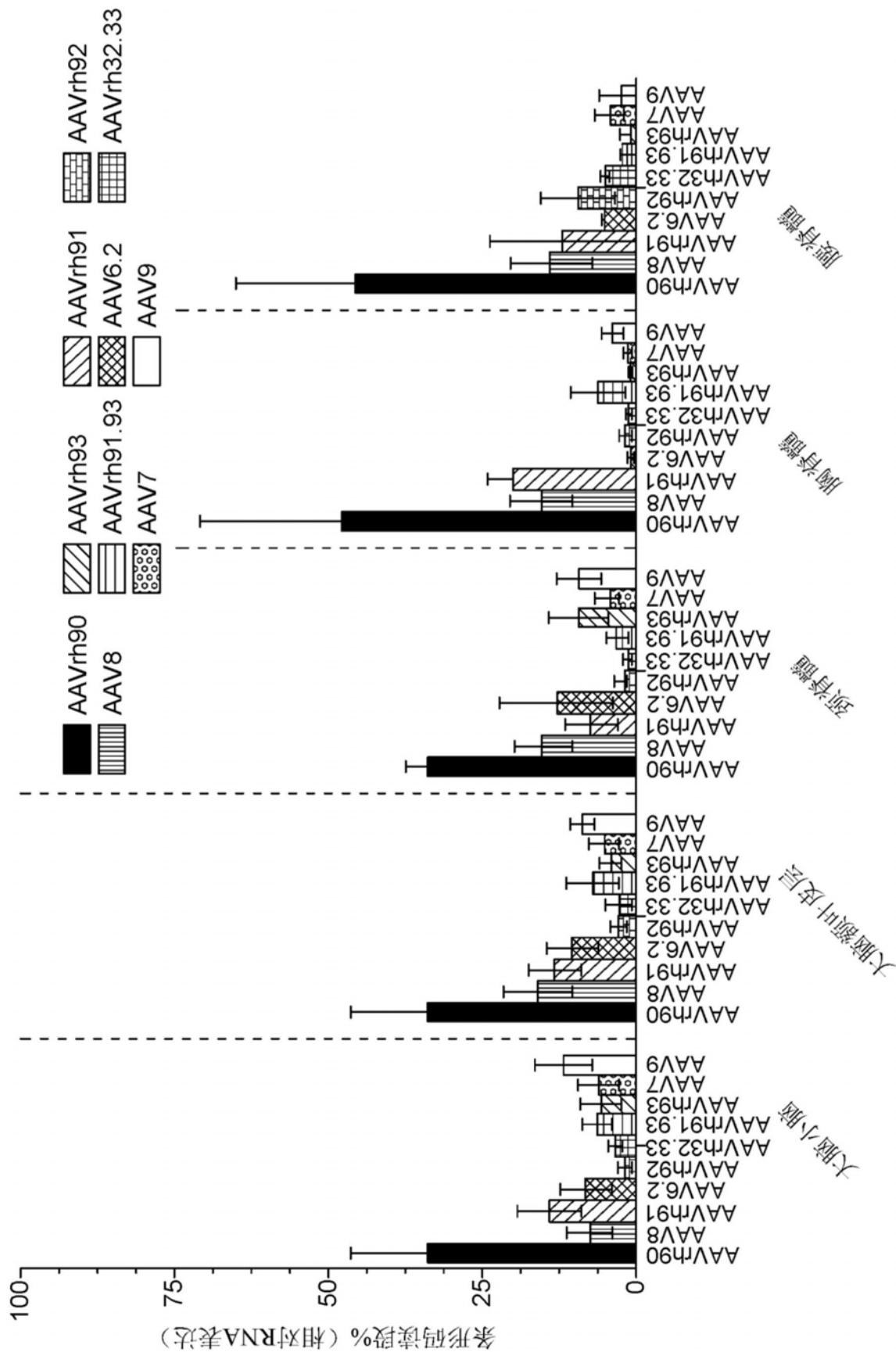


图10B

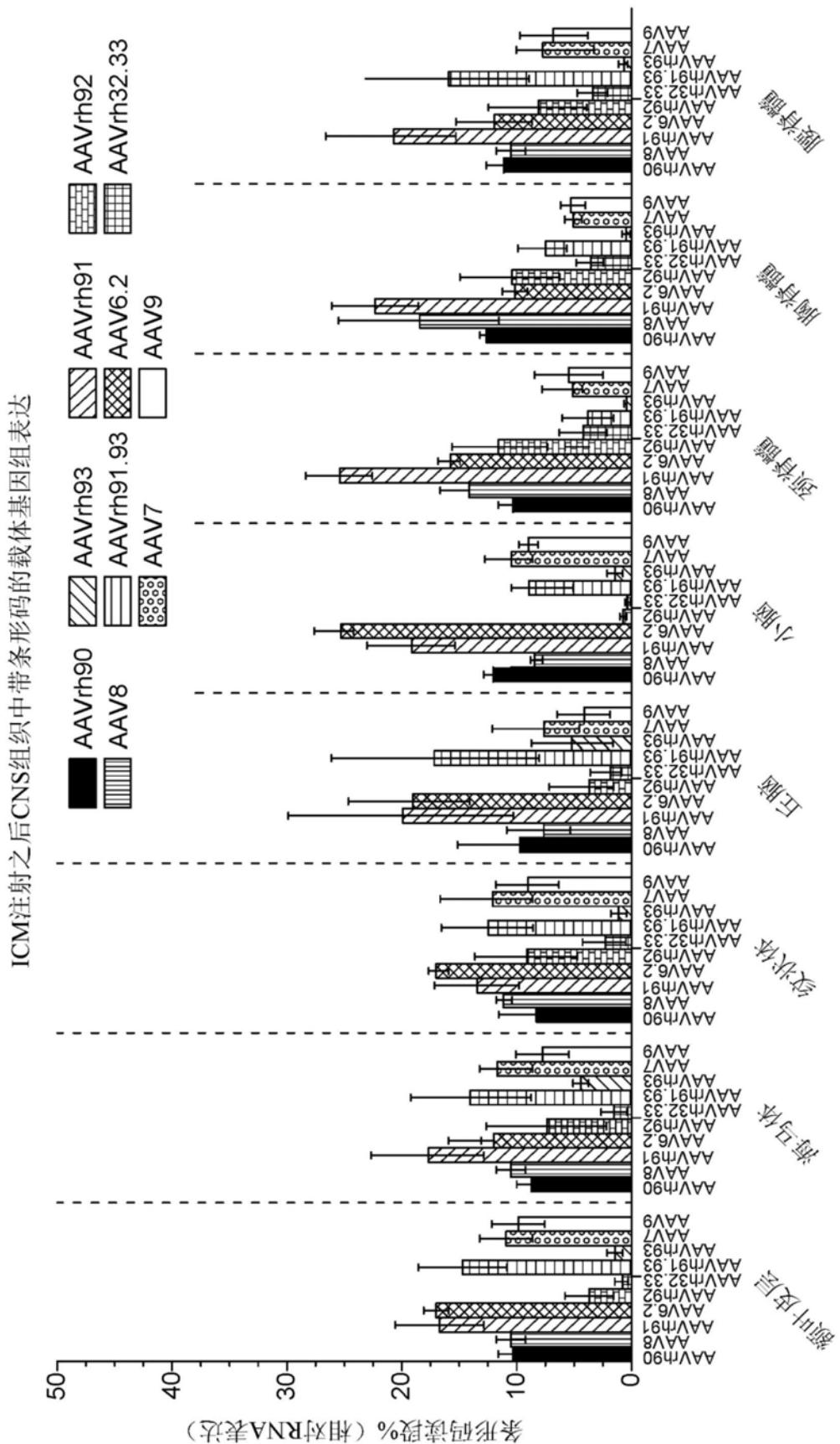


图10C

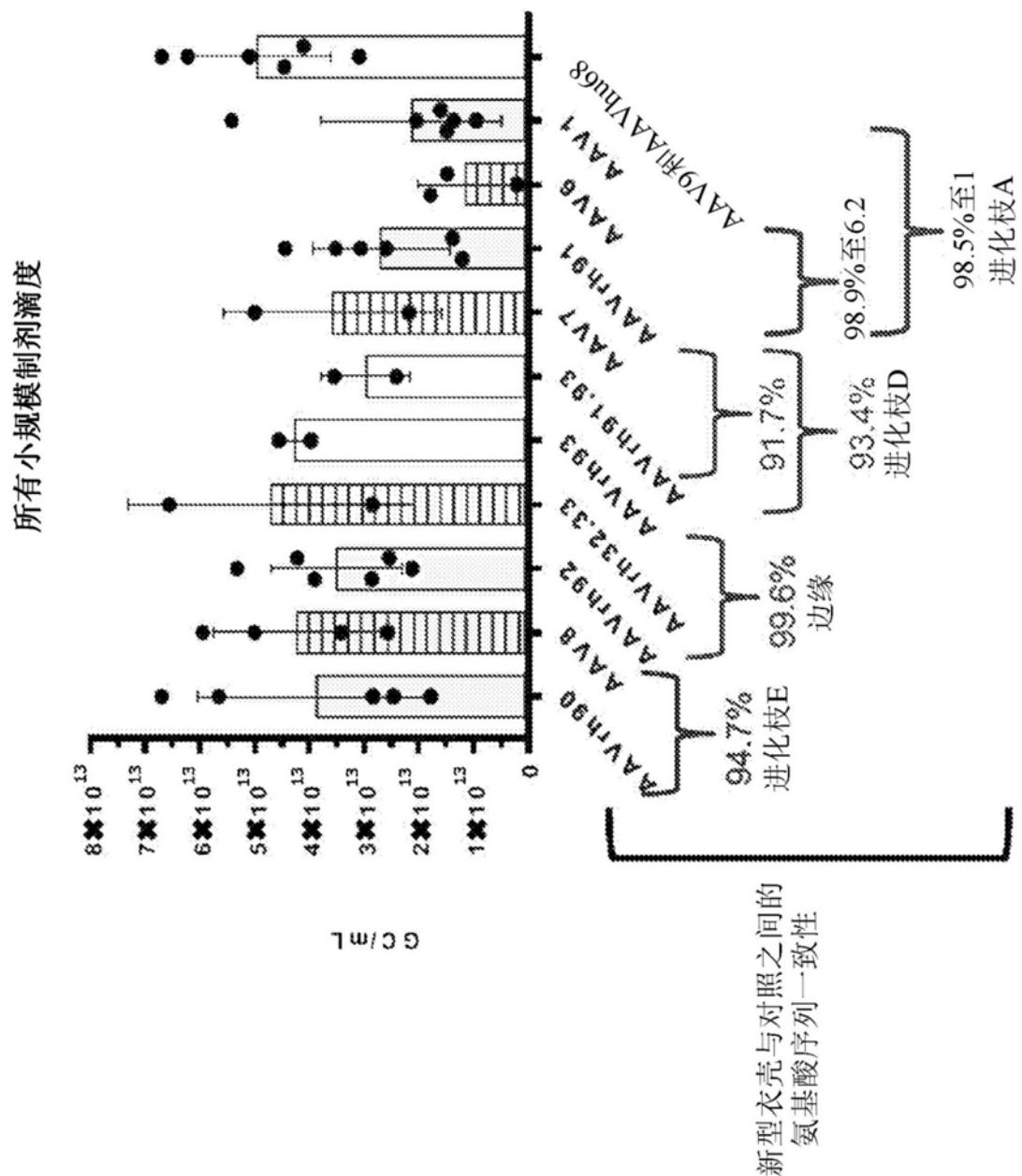


图11

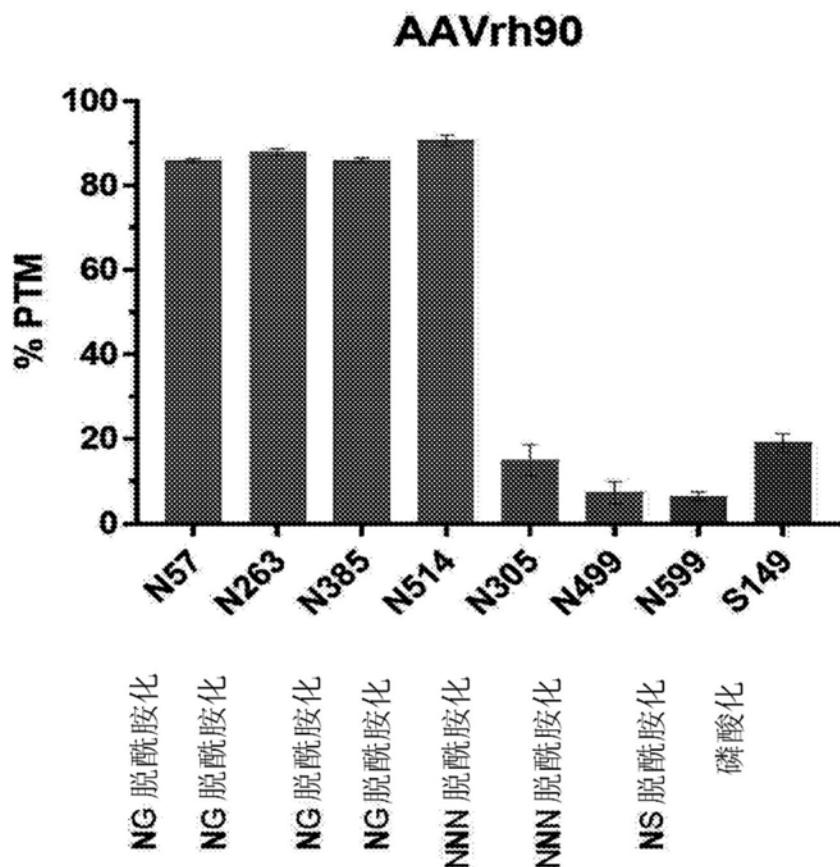


图12A

酶	N+1	WL2882CS 胰蛋白酶-4小时	WL2640S 胰蛋白酶-4小时	WL2258S 胰蛋白酶-4小时	WL2639S 胰蛋白酶-4小时
回收%		100	100	97.4	100.0
N35+脱酰胺化			1.4	0.9	0.8
N57+脱酰胺化	G	87.2	86.0	84.4	85.7
N98+脱酰胺化	H	4.6	5.8	3.7	3.1
~N263+脱酰胺化	G	85.8	87.6	88.9	89.4
~N305+脱酰胺化	NNN	18.7	11.0	13.0	17.4
~N337+脱酰胺化	E			1.6	
~N385+脱酰胺化	G	86.3	85.2	87.4	84.7
~N499+脱酰胺化	NNN	5.0	8.4	10.4	5.6
~N514+脱酰胺化	G	92.3	92.9	90.4	87.3
Q548+脱酰胺化		0.1			
~N599+脱酰胺化	S	6.1	7.6	6.9	5.2
N630+脱酰胺化	F	0.1			
~Q680+脱酰胺化				0.2	
S149+磷酸化		16.6	21.5	19.8	18.6
~W23+氧化		1.2	2.3		1.1
~M204+氧化		1.0	1.2	1.7	1.8
~M212+氧化		0.7	2.9		0.8
W248+氧化		0.3	1.7	2.3	2.0
W282+氧化		0.0	0.0	0.0	
M405+氧化		0.7	0.3	1.1	1.2
M473+氧化		0.4	0.3	0.6	0.58
W480+氧化		2.0	2.1	0.7	1.12
W505+氧化		0.1	0.2	0.2	0.19
M526+氧化		1.0	0.5	1.1	0.20
~M544+氧化		3.1	3.3	3.7	4.44
M561+氧化		2.4	3.1	2.4	3.40
~M607+氧化		1.4	3.1	2.8	2.97
W621+氧化		0.1	1.5	0.9	0.64
M637+氧化		1.9	1.5	2.1	2.72
W697+氧化		0.6	0.8	0.9	0.74

批号	载体名称
WL2882CS(HIT)	AAVrh90.CB7.CI.eGFP.WPRE.rBG(p4698-R2;p1963)
WL2640S	AAVrh90.CMV.LacZ.bGH (p4698;p0102)
WL2639S	AAVrh90.tMCK.PI.3D6 mIgG2a.rBG (p4698;p5229)
WL2258S	AAVrh90.CB7.CI.eGFP.WPRE.rBG(p4698;p1963)

图12B