



(12) 发明专利申请

(10) 申请公布号 CN 113474460 A

(43) 申请公布日 2021.10.01

(21) 申请号 201980092298.4

(74) 专利代理机构 北京品源专利代理有限公司
11332

(22) 申请日 2019.12.19

代理人 刘明海 宁涛

(30) 优先权数据

18214415.4 2018.12.20 EP

(51) Int.Cl.

C12N 15/86 (2006.01)

(85) PCT国际申请进入国家阶段日

2021.08.17

(86) PCT国际申请的申请数据

PCT/EP2019/086454 2019.12.19

(87) PCT国际申请的公布数据

W02020/127831 EN 2020.06.25

(71) 申请人 维格内罗有限责任公司

地址 德国施塔恩贝格

(72) 发明人 斯蒂利亚诺斯·米查拉基斯

马丁·比尔

埃尔维尔·贝西洛维奇

权利要求书5页 说明书67页

序列表27页 附图37页

(54) 发明名称

用于生物学应用和生物技术应用的优化的
受体剪接位点模块

(57) 摘要

本发明涉及新的受体剪接区及其用途和应用。

1. 一种pre-mRNA反式剪接分子,其包括
 - (i) 受体剪接区,其包括
 - (ia) 嘧啶区,其中所述嘧啶区包括
 - (iaa) 5至25个核苷酸;
 - (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
 - (ib) 受体剪接位点,
 - (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
 - (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
 - (ii) 目的核苷酸序列或其部分,其中所述受体剪接区位于目的核苷酸序列或其部分的3'或5';
 - (iii) 靶向pre-mRNA的结合域,其位于目的核酸序列或其部分的3'或5'端;和
 - (iv) 任选的间隔序列,其中所述间隔序列位于结合域和受体剪接区之间。
 2. 权利要求1所述的pre-mRNA反式剪接分子,其包括,
 - (ii) 所述目的核苷酸序列或其部分,其中所述受体剪接区位于目的核苷酸序列或其部分的5';
 - (iii) 供体剪接位点,其中所述供体剪接位点位于所述目的核苷酸序列或其部分的3'端;
 - (iv) 第一结合域,其靶向位于所述目的核苷酸序列或其部分的5'的pre-mRNA;
 - (v) 第二结合域,其靶向位于所述目的核苷酸序列或其部分的3'的pre-mRNA;
 - (vi) 任选的第一间隔序列,其中所述第一间隔位于第一结合域和受体剪接区之间;和
 - (vii) 任选的第二间隔序列,其中所述第二间隔位于所述第二结合域和所述受体剪接位点之间。
 3. 权利要求1或2所述的pre-mRNA反式剪接分子,其中所述受体剪接区位于目的核苷酸序列或其部分的5'端,并且所述结合域位于受体剪接区的5'端。
 4. 权利要求1-3中任一项所述的pre-mRNA反式剪接分子,其还包含终止序列,优选polyA序列。
 5. 权利要求1-4中任一项所述的pre-mRNA反式剪接分子,其中
 - (a) 所述嘧啶区的5至25个核苷酸包含由序列TTTTTT或TCTTTT编码的序列;
 - (b) 嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基;和/或
 - (c) 受体剪接位点具有CAGG序列。
 6. 权利要求1-5中任一项所述的pre-mRNA反式剪接分子,其中所述受体剪接区进一步包括
 - (a) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;
 - (b) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;
 - (c) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核

昔酸是CAA;或

(d) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC;

7. 一种DNA分子,其包含启动子和编码根据权利要求1-6中任一项所述的pre-mRNA反式剪接分子的序列。

8. 一种用于产生核酸序列的方法,该方法包括

(A) 提供包含一个或多个供体剪接位点序列的第一核酸序列;

(B) 提供第二核酸序列,其中所述第二核酸序列包括

(i) 受体剪接区序列,其包括

(ia) 嘧啶区,其中所述嘧啶区包括

(iaa) 5至25个核苷酸;

(iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

(ib) 受体剪接位点,

(iba) 其中所述受体剪接位点位于嘧啶区的3'端;和

(ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和

(C) 在一个或多个供体剪接位点序列切割第一核酸序列并在受体剪接位点切割第二核酸序列,

(D) 将第一切割核酸序列与第二切割核酸序列连接,

从而获得核酸序列。

9. 根据权利要求8所述的方法,其中所述第一核酸序列进一步包括目的核苷酸序列或其一部分,其中所述目的核苷酸序列的至少一部分位于供体剪接位点的5'端,并且其中所述第二核酸序列进一步包括目的核苷酸序列或其一部分,其中所述目的核苷酸序列的至少一部分位于剪接受体剪接区的3'端。

10. 权利要求8或9所述的方法,其中将第一和第二核酸序列引入宿主细胞,优选其中所述第一和第二核酸序列是重组核酸序列。

11. 权利要求8-10中任一项所述的方法,其包括

(A) 将第一核酸序列引入宿主细胞,所述第一核酸序列包含pre-mRNA反式剪接分子序列的核酸序列或编码所述pre-mRNA反式剪接分子的核酸序列,其中第一pre-mRNA反式剪接分子从5'至3'包含

(a) 目的核苷酸序列的5'部分;

(b) 供体剪接位点;

(c) 任意的间隔序列;

(d) 第一结合域;和

(e) 任意的终止序列,优选polyA序列,和

(B) 将第二核酸序列引入宿主细胞,所述第二核酸序列包含pre-mRNA反式剪接分子序列或编码所述pre-mRNA反式剪接分子的核酸序列,其中第二pre-mRNA反式剪接分子从5'至3'包含

(i) 第二结合域,其与第一核酸序列的第一靶结构域互补;

(ii) 受体剪接区序列,其包括
(iia) 嘧啶区,其中所述嘧啶区包括
(iiaa) 5至25个核苷酸;
(iiab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
(iib) 受体剪接位点,
(iiba) 其中所述受体剪接位点位于嘧啶区的3'端;和
(iibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;
(iii) 目的核苷酸序列的3'部分,和
(iv) 终止序列,优选polyA序列,
(C) 在供体剪接位点序列切割第一核酸序列,并在受体剪接位点切割第二核酸序列;和
(D) 将包含目的核苷酸序列5'部分的第一切割核酸序列与包含目的核苷酸序列3'部分的第二切割核酸序列连接,从而获得目的核酸序列。

12. 一种腺相关病毒(AAV)载体,其包含至少两个反向末端重复,所述反向末端重复包含这两个反向末端重复之间的核酸序列,其中所述核酸序列从5'至3'包含:

(i) 启动子;
(ii) 结合域;
(iii) 任意的间隔序列;
(iv) 受体剪接区序列,其包括
(a) 嘧啶区,其中所述嘧啶区包括
(aa) 5至25个核苷酸;
(ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
(b) 受体剪接位点,
ba) 其中所述受体剪接位点位于嘧啶区的3'端;和
bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;
(v) 目的核苷酸序列或其一部分;和
(vi) 任意的poly A序列。

13. 一种腺相关病毒(AAV)载体系统,其包括

(I) 第一AAV载体,其包含至少两个反向末端重复,所述反向末端重复包含这两个反向末端重复之间的核酸序列,其中这两个反向末端重复之间的所述核酸序列从5'至3'包含:

(a) 启动子,
(b) 编码目的多肽的N末端部分的核苷酸序列;
(c) 供体剪接位点;
(d) 任意的间隔序列;
(e) 第一结合域;和
(f) 任意的终止序列,优选polyA序列。

(II) 第二AAV载体,其包含含有至少两个反向末端重复的核酸序列,所述反向末端重复包含这两个反向末端重复之间的核酸序列,其中这两个反向末端重复之间的所述核酸序列

从5′至3′包含：

- (i) 启动子；
- (ii) 第二结合域，其与所述第一AAV载体的所述第一结合域互补；
- (ii) 任意的间隔序列；
- (iii) 受体剪接区序列，其包括
 - (a) 嘧啶区，其中所述嘧啶区包括
 - (aa) 5至25个核苷酸；
 - (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基，如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U)；
 - (b) 受体剪接位点，
 - (ba) 其中所述受体剪接位点位于嘧啶区的3′端；和
 - (bb) 其中所述受体剪接位点包含5′至3′的NAGG序列，其中N是A、C、T/U或G；
 - (iv) 编码目的多肽的C末端部分的核苷酸序列；
 - (ivb) 其中所述目的多肽的C末端部分和所述目的多肽的N末端部分重构所述目的多肽；和
 - (v) 终止序列，优选polyA序列。

14. 权利要求13所述的AAV载体系统，其中所述多肽是全长多肽，并且所述第一AAV载体包含目的全长多肽的N-末端部分，并且所述第二AAV载体包含目的全长多肽的C-末端部分。

15. 一种核酸序列，其包括

- (i) 受体剪接区序列，其包括
 - (ia) 嘧啶区，其中所述嘧啶区包括
 - (aa) 5至25个核苷酸；
 - (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基，如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U)；
 - (ib) 受体剪接位点，
 - ba) 其中所述受体剪接位点位于嘧啶区的3′端；和
 - bb) 其中所述受体剪接位点包含5′至3′的NAGG序列，其中N是A、C、T/U或G；和
 - (ii) 目的核苷酸序列，其中所述目的核苷酸序列
 - (iia) 位于受体剪接区的3′或5′。

16. 权利要求8-11中任一项所述的方法、权利要求12所述的AAV载体、权利要求13或14所述的AAV载体系统和权利要求15所述的核酸，其中

- (a) 嘧啶区的5至25个核苷酸包含序列TTTTTT或TCTTTT；
- (b) 嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个，优选少于5个，更优选少于3个碱基；
- (c) 受体剪接位点具有CAGG序列；
- (d) 所述受体剪接区进一步包含所述嘧啶区5′的7个核苷酸，其具有序列CAACGAG的至少4个核苷酸，其中5′第一个核苷酸是C；和/或
- (e) 其中所述剪接受体区包含SEQ ID NO:3或4的序列。

17. 权利要求1-6中任一项所述的pre-mRNA反式剪接分子、权利要求12所述的腺相关病毒载体或用于治疗的权利要求13或14所述的腺相关病毒载体系统。

用于生物学应用和生物技术应用的优化的受体剪接位点模块

技术领域

[0001] 本发明涉及新的受体剪接区及其用途和应用。

背景技术

[0002] 大多数真核基因都含有非编码内含子,在称为“剪接”的过程中,必须从前体信使RNA (pre-mRNA) 中去除非编码内含子,才能产生可翻译的成熟信使RNA (mRNA)。剪接是通过一个大的核糖核蛋白复合物(即剪接体)介导的,它由五个保守的小核核糖核蛋白(snRNP)(即,U1、U2、U4、U5和U6 snRNP),以及300多种蛋白质组成。剪接体在一个高度动态的过程中为每个内含子组装和拆卸。为此,需要识别特定的剪接序列,并因此,外显子和内含子的边界由不同的序列基序定义,这些基序是参与调节mRNA剪接的核糖核蛋白的结合位置。

[0003] 最相关的剪接基序是规范的供体和受体剪接位点,它们定义了外显子-内含子的边界。5'供体剪接位点(DSS)位于外显子的3'端和下游内含子的5'端,3'受体剪接位点(ASS)位于内含子的3'端和下游外显子的5'端。一个功能性受体剪接位点(ASS)需要三个不同的元件(其通常位于大约50bp的范围内):一个分支点(BP),一个多嘧啶区(PPT)和规范的ASS序列。

[0004] 所谓的GT-AG剪接是迄今为止哺乳动物中最常见的mRNA剪接类型,在DNA水平上定义了内含子的前两个和后两个核苷酸。因此,规范的DSS序列的一部分代表5'内含子序列的前两个碱基,其是DNA序列中的一个鸟嘌呤,然后是胸腺嘧啶(GT)(对应于RNA序列中的GU),并且内含子的最后两个碱基(代表共有ASS最保守的部分)是一个腺嘌呤接一个鸟嘌呤(AG)。GT-AG核苷酸对于有效的剪接反应至关重要,并且它们的破坏或取代会导致功能性剪接位点的丢失。

[0005] 每个剪接循环由两个酯交换反应组成。对于称为分支的第一个反应,所述分支点核苷(通常是腺苷)攻击5'外显子-内含子连接处的磷酸键。这导致内含子套索-3'外显子中间体(包括剪接位点下游的外显子和内含子)和游离5'外显子末端(包括剪接位点上游的外显子)的形成。在称为外显子连接的第二个反应中,5'外显子的自由端攻击内含子-3'外显子连接处的磷酸,导致两个外显子连接和内含子套索结构的释放。剪接反应由pre-mRNA的辅助顺式作用剪接调节元件调节,该元件由最高10个核苷酸组成。根据它们的功能和位置,它们可以分为外显子剪接增强子(ESEs)、外显子剪接沉默子(ESSs)、内含子剪接增强子(ISEs)和内含子剪接沉默子(ISSs)。这些元件能够招募反式作用蛋白来促进或预防外显子包含,因此被认为在选择性剪接中发挥重要作用。

[0006] 给定外显子的剪接效率预期取决于供体和受体剪接位点的强度。这种DSS或ASS强度又取决于GT或AG上游或下游的内含子或外显子序列元件。使用标准的计算机预测软件(例如NNSplice:http://www.fruitfly.org/seq_tools/splice.html或Human Splice Finder:<http://www.umd.be/HSF3/>)可以可靠地预测功能性DSS的强度。相比之下,由于其复杂性,对ASS强度的计算机预测导致不太可靠的结果(例如Koller et al., (2011) “A novel screening system improves genetic correction by internal exon

replacement.”Nucleic acids research.39:e108;Lorain et al., (2013) “Dystrophin rescue by trans-splicing:a strategy for DMD genotypes not eligible for exon skipping approaches.”Nucleic acids research.41:8391-8402)。因此,实际的ASS强度需要实验验证。许多生物学、生物技术和治疗应用依赖于经典的GT-AG mRNA剪接的效率,因此依赖于强剪接位点的使用。

[0007] 除了通过去除非编码内含子从pre-mRNA产生可翻译的成熟mRNA的规则顺式剪接事件之外,剪接也可以反式发生,从而结合两个独立的pre-mRNA分子产生非共线嵌合RNA (Lei et al., (2016) “Evolutionary insights into RNA trans-splicing in vertebrates”,Genome Biol.Evol.8 (3) :562-577)。这一过程首先在锥虫中被发现。从那以后,反式剪接事件也在更多的物种中被发现,包括小鼠 (Hirano M and Noda T., (2004) “Genomic organization of the mouse Msh4 gene producing bicistronic chimeric and antisense mRNA”,Gene 342:165-177),以及在人类细胞中 (Chuang et al., (2018) “Integrative transcriptome sequencing reveals extensive alternative trans-splicing and cis-back splicing in human cells”.Nucleic Acids Research,46 (7) :3671-3691),尽管反式剪接似乎只在高等脊椎动物中发生得不太频繁。

[0008] 已经尝试利用反式剪接进行基因治疗。病毒载体是基因治疗的一种有吸引力的载体,然而,它们通常具有有限的装载能力,因此对于哪些基因可以被替换存在限制。例如,腺相关病毒载体具有最高约5.0kb的包装容量,这允许约4kb或更小的转基因。反式拼接(即,将两个物理上分离的pre-mRNA连接在一起形成成熟的mRNA)可能是克服这些限制的一种方法。例如,剪接体介导的RNA反式剪接(SMaRT)可以用作基因治疗的工具,包括外源性pre-mRNA反式剪接分子,该分子被引入靶细胞以仅替换突变的内源性pre-mRNA的一部分。这可以允许在病毒载体中递送较短的编码序列。

[0009] 解决病毒载体(特别是基于腺相关病毒(AAV)的载体)大小限制的另一种方法是使用重组AAV(rAAV)双载体技术。AAV是一种单链DNA病毒。对于rAAV双载体技术,转基因(目的基因)的编码序列被分成至少两部分,并包装成两个或多个独立的rAAV载体。在通过分裂的基因组载体共转导靶细胞后,全长编码序列被重组。由于许多细胞,如视网膜的感光细胞,显示出>90%的高共转导效率,因此两种rAAV向同一靶细胞的有效递送似乎并不受限。然而,两个转基因半体的有效重组仍然具有挑战性。在rAAV双载体系统开发后的几年里,重构仅在DNA水平上得到解决,并且已经探索了几种策略来改进这种方法 (McClements and MacLaren, (2017) “Adeno-associated virus (AAV) dual vector strategies for gene therapy encoding large transgenes,Yale Journal of Biology and Medicine 90: 611-623)。

[0010] 这些rAAV双载体系统通常被误导性地称为“反式剪接双载体”,尽管这些方法中的mRNA剪接实际上从未反式发生。相反,在先前描述的rAAV双载体系统中的重构依赖于ITR结构的串联化和/或重叠序列的同源重组,产生可以顺式剪接的单个pre-mRNA,以去除串联化的ITR元件和/或人工重组元件。这种rAAV双载体系统是,例如,由Trapani et al., (“Effective delivery of large genes to the retina by dual AAV vectors”, (2014) EMBO Molecular Medicine,6 (2) :194-211)公开的。通过在第一AAV载体中存在驱动编码序列5’部分表达的启动子,以及在第二AAV载体中不存在驱动编码序列3’部分表达的启动子,

可以识别在DNA水平上重构rAAV双载体系统。报道的重构这种rAAV双载体系统的体内效率相对较低,即低于10% (Carvalho et al., “Evaluating efficiencies of dual AAV approaches for retinal targeting”, (2017) Frontiers in Neuroscience, 11 (503) :1-8)。相反,为了使pre-mRNA剪接以反式构型发生,rAAV双载体系统的两个载体都需要启动子来产生两个独立的pre-mRNA分子。

[0011] 为了实现基因剪接,特别是基因反式剪接的最佳性能,需要鉴定优化的和实验验证的强ASS。特别是,需要强ASS区,其尤其可用于开发进一步的AAV载体,例如可提供高度特异性和高效的pre-mRNA反式剪接分子的下一代rAAV双载体系统或AAV,其可用于基因治疗。

发明内容

[0012] 本发明符合如本文中所描述的且如在实施例、附图和权利要求书中所描述的此需要。

[0013] 本文提供了一种pre-mRNA反式剪接分子,其包含:(i)受体剪接区,其包含(ia)嘧啶区,其中所述嘧啶区包含(iaa)5至25个核苷酸;(iab)其中这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);(ib)受体剪接位点,(iba),其中所述受体剪接位点位于所述嘧啶区的3'端;和(ibb)其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;(ii)目的核苷酸序列或其部分;其中所述受体剪接区位于目的核苷酸序列或其部分的3'或5';(iii)靶向pre-mRNA的结合域,其位于目的核苷酸序列或其部分的3'或5';和(iv)任意的间隔序列,其中所述间隔序列位于所述结合域和所述受体剪接区之间。在一个实施方案中,权利要求1的所述pre-mRNA反式剪接分子包含(ii)所述目的核苷酸序列或其部分,其中所述受体剪接区位于目的核苷酸序列或其部分的5'端;(iii)供体剪接位点,其中所述供体剪接位点位于目的核苷酸序列或其部分的3'端;(iv)第一结合域,其靶向位于目的核苷酸序列或其部分的5'的pre-mRNA;(v)第二结合域,其靶向位于目的核苷酸序列或其部分的3'的pre-mRNA;(vi)任意的第一间隔序列,其中所述第一间隔位于所述第一结合域和所述受体剪接区之间;和(vii)任意的第二间隔序列,其中所述第二间隔位于所述第二结合域和所述供体剪接位点之间。

[0014] 优选地,所述受体剪接区位于目的核苷酸序列或其部分的5'端,并且所述结合域位于受体剪接区的5'端。所述pre-mRNA反式剪接分子可以进一步包含终止序列,优选polyA序列。优选地,根据本发明所述的pre-mRNA反式剪接分子包含嘧啶区,其中所述嘧啶区的5至25个核苷酸包含由TTTTTT或TCTTTT序列编码的序列。另外地或任选地,所述嘧啶区的最后一个嘧啶和所述受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基。另外地或任选地,所述受体剪接位点具有CAGG序列。

[0015] 所述受体剪接区可以包含(a)嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;(b)嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;(c)嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAA;或者(d)嘧啶区5'的7个核苷酸,具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC。在一个实施方案中,所述受体剪接区由核苷酸序列编码,所述核苷酸序列具有SEQ ID NO:3或4的序列,或者与SEQ ID NO:3或4的序列至少85%、90%、95%、97%、98%、99%或100%相同。还提供了一种DNA分子,其

包含启动子和编码根据本发明所述的pre-mRNA反式剪接分子的序列,其中所述DNA分子优选为载体或质粒。

[0016] 另一方面,本发明涉及产生核酸序列的方法,所述方法包括(A)提供第一核酸序列,所述第一核酸序列包含一个或多个供体剪接位点序列;(B)提供第二核酸序列,其中所述第二核酸序列包含(i)受体剪接区,所述受体剪接区包含(ia)嘧啶区,其中所述嘧啶区包含(iaa)5至25个核苷酸;(iab)其中这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);(ib)受体剪接位点,(iba)其中所述受体剪接位点位于嘧啶区的3'端;和(ibb)其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;(C)在一个或多个供体剪接位点序列切割所述第一核酸序列,并在所述受体剪接位点切割所述第二核酸序列,(D)将第一切割核酸序列与第二切割核酸序列连接,从而获得核酸序列。所述第一核酸序列还包含目的核苷酸序列或其一部分,其中目的核苷酸序列的至少一部分位于所述供体剪接位点的5'端,并且其中所述第二核酸序列还包含目的核苷酸序列或其一部分,其中目的核苷酸序列的至少一部分位于所述受体剪接区的3'端。优选地,将所述第一和第二核酸序列引入宿主细胞,优选地,其中所述第一和第二核酸序列是重组核酸序列。

[0017] 在本发明方法实施方案的另一面,所述方法包括步骤(A)将第一核酸序列导入宿主细胞,所述第一核酸序列包含pre-mRNA反式剪接分子序列或编码所述pre-mRNA反式剪接分子的核酸序列,其中第一pre-mRNA反式剪接分子从5'至3'包含,(a)目的核苷酸序列的5'部分;(b)供体剪接位点;(c)任意的间隔序列;(d)第一结合域;和(e)任意的终止序列,优选polyA序列;步骤(B)将第二核酸序列引入宿主细胞,所述第二核酸序列包含pre-mRNA反式剪接分子序列或编码所述pre-mRNA反式剪接分子的核酸序列,其中第二pre-mRNA反式剪接分子从5'至3'包含;(i)第二结合域,其与第一核酸序列的第一靶结构域互补;(ii)受体剪接区序列,其包括:(iia)嘧啶区,其中所述嘧啶区包括:(iiaa)5至25个核苷酸;(iiab)其中这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);(iib)受体剪接位点,(iiba)其中所述受体剪接位点位于所述嘧啶区的3'端;和(iibb)其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;(iii)目的核苷酸序列的3'部分,和(iv)终止序列,优选polyA序列;步骤(C)在所述供体剪接位点序列切割所述第一核酸序列,并在所述受体剪接位点切割所述第二核酸序列;以及步骤(D)将包含目的核苷酸序列5'部分的第一切割核酸序列与包含目的核苷酸序列3'部分的第二切割核酸序列连接,从而获得目的核酸序列。

[0018] 在另一方面,本发明涉及包含至少两个反向末端重复序列的腺相关病毒(AAV)载体,所述反向末端重复序列包含在这两个反向末端重复序列之间的核酸序列,其中所述核酸序列包含5'至3':(i)启动子;(ii)结合域;(iii)任意的间隔序列;(iv)受体剪接区序列,所述受体剪接区序列包含(a)嘧啶区,其中所述嘧啶区包含(aa)5至25个核苷酸;(ab)其中这5-25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);(b)受体剪接位点,其中所述受体剪接位点位于所述嘧啶区的3'端;并且其中所述受体剪接位点包含从5'到3'的NAGG序列,其中N是A、C、T/U或G;(v)目的核苷酸序列或其部分;和(vi)任意的多聚腺苷酸序列。所述AAV载体也可以是AAV载体系统的一部分,该系统包括(I)第一AAV载体,所述第一AAV载体包含至少两个反向末端重复,所述反向末端重复包含这两

个反向末端重复之间的核酸序列,其中这两个反向末端重复之间的所述核酸序列从5'至3'包含:(a)启动子,(b)编码目的多肽的N末端部分的核苷酸序列;(c)供体剪接位点;(d)任意的间隔序列;(e)第一结合域;和(f)任意的终止序列,优选polyA序列;(II)第二AAV载体,所述第二AAV载体包含含有至少两个反向末端重复的核酸序列,所述反向末端重复包含这两个反向末端重复之间的核酸序列,其中这两个反向末端重复之间的所述核酸序列从5'至3'包含(i)启动子;(ii)第二结合域,其与第一AAV载体的第一结合域互补;(iii)任意的间隔序列;(iii)受体剪接区序列,所述受体剪接区序列包含(a)嘧啶区,其中所述嘧啶区包含(aa)5至25个核苷酸;(ab)其中这5-25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);(b)受体剪接位点,其中所述受体剪接位点位于所述嘧啶区的3'端;并且其中所述受体剪接位点包含从5'到3'的NAGG序列,其中N是A、C、T/U或G;(iv)编码目的多肽的C末端部分的核苷酸序列;(iva)其中目的多肽的C末端部分和目的多肽的N末端部分重构所述目的多肽;和(v)终止序列,优选polyA序列。在一个实施方案中,所述多肽是全长多肽,并且所述第一AAV载体包含全长目的多肽的N末端部分,并且所述第二AAV载体包含全长目的多肽的C末端部分。

[0019] 还提供了包含受体剪接区序列的核酸序列,所述受体剪接区序列包含(ia)嘧啶区,其中所述嘧啶区包含(aa)5至25个核苷酸;(ab)其中这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);(ib)受体剪接位点,ba),其中所述受体剪接位点位于所述嘧啶区的3';和bb)其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和(ii)目的核苷酸序列,其中所述目的核苷酸序列(ia)位于所述受体剪接区的3'或5'。

[0020] 在根据本发明所述的某些实施方案中,本文描述的受体剪接区包含嘧啶区,其中所述嘧啶区的5至25个核苷酸包含由TTTTTT或TCTTTT序列编码的序列。所述嘧啶区的最后一个嘧啶和所述受体剪接位点之间的序列优选少于10个,更优选少于5个,甚至更优选少于3个碱基。此外,所述受体剪接位点优选具有序列CAGG。所述受体剪接区可以进一步包含嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C。在一个优选的实施方案中,所述剪接受体区包含SEQ ID NO:3或4的序列或由SEQ ID NO:3或4的序列组成,或者由包含SEQ ID NO:3或4的序列或由SEQ ID NO:3或4的序列组成的核酸序列编码。

附图说明

[0021] 图1RH0微小基因在HEK293细胞和转导的光受体中的mRNA剪接。

[0022] A, RH0基因的真实比例视图示意图。方框代表外显子,中间的粗线代表内含子,标出了各自的起始密码子和终止密码子。星号象征外显子3中的c.620T>G突变。B和C,包含外显子的编码部分和侧翼内含子的视紫红质微小基因由CMV启动子驱动以在HEK293细胞中表达(B),或由人视紫红质启动子驱动以在光受体中表达(C)。为了能够将RH0微小基因包装到AAV载体中,内含子1如所示被缩短。为了蛋白质的可视化和检测,RH0在C端与柠檬黄蛋白(citrine)(在B中)或myc标签(在C中)融合。在B和C中,箭头所示的引物用于特异性检测来源于HEK293细胞(D)或来源于转导的光受体(E)的野生型或突变型剪接产物。D,瞬时转染野生型和突变型RH0微小基因的HEK293细胞的RT-PCR分析。E,左图,在出生后第14天视网膜下

RHO微小基因递送至小鼠视网膜的示意图(P14)。右图,来自注射的鼠视网膜的RT-PCR,其中含有相应的野生型和突变型RHO微小基因。注射后4周进行RT-PCR。所有实验重复一次。

[0023] 图2最有效的ASS_620序列的鉴定。

[0024] A,功能性ASS的单个元件的组成。分支点(BP)用下划线表示出人类分支点的共有序列,分支点核苷腺苷用粗体突出显示。多嘧啶区(Poly-C/T)表示为一个空框,跨越内含子-外显子边界的ASS表示为一个灰色填充框。B,人类RHO基因的外显子-内含子组织和野生型(WT,上图)和c.620T>G突变型(下图)的人类RHO基因外显子3的DNA序列放大图。内含子序列以小写显示,外显子序列以大写(粗体)字母显示。单个ASS元件按照A所示的方案高亮显示,BP序列加下划线(粗线),Poly-C/T序列用空方框标记,ASS序列用实心灰色方框标记。WT中的ATG序列被转换成c.620T>G突变型中的AGG序列,用细下划线标出。注意,c.620T>G突变产生了一个新的规范的ASS序列。功能性ASS所需的另外两个元件(BP和多聚C/T)已经存在于WT RHO外显子3中。C,用于剪接实验的人核糖体蛋白27(RPS27)微小基因的结构。(D)中显示的用于RT-PCR的引物用箭头表示。如下所示,将包含称为RHO_E3a-g的ASS_620潜在元件的单个序列插入到RPS27基因的外显子3中。RHO_E3a是缺少功能性ASS的WT RHO序列。潜在的BP、多聚C/T或ASS序列如上所示。D,来自用所示的单个嵌合RPS27微小基因转染的HEK293细胞的RT-PCR。CS,正确剪接的RPS27微小基因(即使用天然外显子3ASS)。AS,异常剪接的RPS27微小基因(即使用ASS_620)。

[0025] 图3ASS_620是强受体剪接位点。

[0026] A,使用两种常用预测工具进行硅受体剪接位点(ASS)强度预测,NNSplice(0.9版;1997年1月)(http://www.fruitfly.org/seq_tools/splice.html)或human splicing finder(3.1版)(HSF,<http://www.umd.be/HSF3/>)。B,被称为vgASS_620的26bp序列的单个元件的组成和序列,包括ASS、多聚C/T和多聚C/T的5'的另外7个核苷酸。C,用于确定vgASS_620强度的不同微小基因的示意图。单个微小基因中vgASS_620插入的位置用星号和虚线表示。D中所示的用于RT-PCR的引物的结合位置用箭头表示。D,来自用仅含有天然(nat)受体剪接位点或同时含有天然受体剪接位点和vgASS_620(620)的相应微小基因转染的HEK293细胞的RT-PCR。通过测序检查所有条带。E,来自用天然受体剪接位点或vgASS_620的RT-PCR产物的长度。

[0027] 图4测试不同结合域的天蓝蛋白重构试验。

[0028] A,RHO内含子2序列的示意图,用作通过mRNA反式剪接重构天蓝蛋白的结合域。测试整个RHO内含子2序列(a)以及不同的5'(b,d,f,h)和3'部分(c,e,g,i),或融合到内含子2的小3'部分(h+i)的小5'部分,提供B-G所示的结果。B和C,来自转染的HEK293细胞的对照qRT-PCR(n=3),以比较包含A所示不同结合域的单个构建体的mRNA水平。显示了与管家氨基乙酰丙酸合酶(ALAS)相关的delta CT(Δ CT)值。引物的结合位置(B中的p1+p2和C中的p3+p4)显示在D中。通过单向ANOVA进行统计分析(每次转染n=3),然后进行Tukey检验进行多重比较。D,使用h+i结合域、供体剪接位点(DSS)和受体剪接位点ASS_620的实例的天蓝蛋白重构分析的原理。如虚线所示,天蓝蛋白在全长天蓝蛋白序列(c1)的起始密码子下游的核苷酸位置154处被裂解。在对照构建体(c2)中,在指定位置154处将人工内含子引入到天蓝蛋白编码序列中以产生两个人工天蓝蛋白外显子。用于蛋白质印迹的抗体(α -cerNT)与N端天蓝蛋白半部分结合。对照构建体(c2)用作共聚焦成像和蛋白质印迹实验中定量的参考。

使用双载体方法测试(A)中所示的不同序列作为结合域。针对h+i结合域与vgASS_620(c3)或天然RHO外显子3(RHO_E3)受体剪接位点(c4)的组合,示例性地示出了用于天蓝蛋白重构的双载体方法的方案。E,从三次独立转染产生的蛋白质印迹带强度计算不同结合域的天蓝蛋白重构效率。为了定量,首先将条带强度归一化为内部微管蛋白对照。归一化值以参考结构强度(c2)的百分比给出。F,用由CMV启动子驱动c1、c2、c3或c4构建体转染的活HEK293细胞的代表性共聚焦成像。为了成像,对于成像,使用了天蓝蛋白特定的激光和过滤器设置。比例尺,50 μ m。G,来自转染的HEK293细胞的蛋白质裂解物的代表性蛋白质印迹(O,未转染细胞;IB,免疫印迹; α -Tub, β -微管蛋白特异性抗体)。

[0029] 图5:包含ABCA4的5'编码序列的第一双AAV的序列(SEQ ID NO:19):5'端和3'端的ITR序列用灰色突出显示,没有下划线(NNN)。间隔序列用小写字母表示。CMV启动子用灰色突出显示,并用实线下划线标出(NNN)。ABCA4蛋白的5'编码序列用虚线下划线标出(NNN)。DSS用斜体字母表示,并带有下划线(Nnn,大写字母代表编码序列,小写字母代表非编码序列)。结合域以灰色突出显示,并用间断下划线标出(NNN)。

[0030] 图6:包含ABCA4的3'编码序列的第二双AAV的序列(SEQ ID NO:20):5'端和3'端的ITR序列用灰色突出显示,没有下划线(NNN)。间隔序列用小写字母表示。CMV启动子用灰色突出显示,并用实线下划线标出(NNN)。结合域用灰色突出显示,并用虚线下划线(NNN)。受体剪接位点用斜体表示并加下划线(NNN)。ABCA4蛋白的3'编码序列用虚线下划线标出(NNN)。polyA序列用深灰色突出显示,并用波浪下划线标出(NNN)。

[0031] 图7:包含ABCA4的5'编码序列和ABCA4启动子的第一双AAV的序列(SEQ ID NO:19):5'端和3'端的ITR序列用灰色突出显示,没有下划线(NNN)。间隔序列用小写字母表示。ABCA4启动子用灰色突出显示,并用实线下划线标出(NNN)。ABCA4蛋白的5'编码序列用虚线下划线标出(NNN)。DSS用斜体字母表示,并带有下划线(Nnn,大写字母代表编码序列,小写字母代表非编码序列)。结合域用间断下划线标出(NNN)。

[0032] 图8:包含ABCA4的3'编码序列和ABCA4启动子的第二双AAV的序列(SEQ ID NO:22):5'端和3'端的ITR序列用灰色突出显示,没有下划线(NNN)。间隔序列用小写字母表示。ABCA4启动子用灰色突出显示,并用实线下划线标出(NNN)。结合域用间断下划线标出(NNN)。ASS用斜体表示并加下划线(NNN)。ABCA4蛋白的3'编码序列用灰色突出显示,并用虚线下划线标出(NNN)。polyA序列用深灰色突出显示,并用波浪下划线标出(NNN)。

[0033] 图9:受体剪接位点与结合域对天蓝蛋白重构效率的影响。A,本实验测试了结合域(BD)和受体剪接位点(ASS)。所有BD序列都来源于人类RHO基因。预期会产生高效率的BD和ASS以粗体和斜体标示。B,用包含A中所示的三种BD和三种ASS的不同组合的构建体转染的HEK293细胞的共聚焦活体图像。显示了相应BD和ASS的强度。比例尺,50 μ m。C,上图,不同BD和ASS组合的RT-PCR。下图,不同BD和ASS组合的蛋白质印迹。GAPDH和 β 微管蛋白作为上样对照。D,通过与cis-ctrl(n=3-8)相关的天蓝蛋白带的辐射分析来量化重构效率。在定量前,将所有蛋白质条带归一化为 β 微管蛋白。

[0034] 图10:体内mRNA反式剪接rAAV双载体途径。A,用于体内表达的5'载体和3'载体构建体。作为表达对照,柠檬黄蛋白和mCherry序列分别融合在天蓝蛋白5'编码序列(CDS)的5

‘和天蓝蛋白3‘CDS的3‘上。B,注射后两周视网膜切片的代表性共聚焦图像,表达含BD_h+i的构建体。荧光团表达存在于视网膜色素上皮(RPE)中。ONL,外核层。比例尺,20 μ m。使用514nm激光对柠檬黄蛋白和mCherry荧光团进行选择性的漂白之前(上图)和之后(下图)的RPE细胞的共聚焦图像。比例尺,2 μ m。

[0035] 图11:鉴定来源于lacZ基因的合适BD。A,结合域(BD)取自细菌lacZ基因,并经过修饰以与人类基因组没有可检测的同源性。B,用含有如A中所示的BD的构建体瞬时共转染的HEK293细胞的共聚焦活体图像。比例尺,50 μ m。从转染的HEK293细胞裂解物获得的蛋白质印迹。D,基于蛋白质印迹带强度的比率分析的天蓝蛋白重构效率的量化(n=3-8)。BD_g效率用作迄今为止获得的最高重构的度量(参见图9)。

[0036] 图12:调控元件缺失及其对mRNA剪接效率的影响。A,HEK293细胞瞬时共转染所示含BD_k的天蓝蛋白构建体的共聚焦活体图像。5‘^{polyA}del,没有多聚腺苷酸信号的5‘载体。3‘^{promdel},无启动子序列的3‘载体。比例尺,50 μ m。B,从转染的HEK293细胞中获得的蛋白质印迹如图A所示。 β 微管蛋白作为上样对照。

[0037] 图13:SpCas9-VPR的重构。A,用于在HEK293细胞中进行SpCas9-VPR的mRNA反式剪接的质粒(5‘和3‘载体)。含有全长(FL)SpCas9-VPR CDS的质粒(FL载体)作为阳性对照。B,来自与含有BD_k(n=3)或FL载体(n=3)的5‘和3‘载体共转染的HEK293细胞的RT-PCR。C,重构的SpCas9-VPR产物的代表性测序结果。D,来自各自转染的蛋白质裂解物的蛋白质印迹。 \emptyset ,未转染的细胞。 β 微管蛋白作为上样对照。

[0038] 图14:ABCA4的重构。A,用于在HEK293细胞中进行ABCA4的mRNA反式剪接的质粒(5‘和3‘载体)。此外,六个不同基因的六个短内含子分别被整合到ABCA4的CDS中,即三个进入5‘CDS(带内含子的5‘载体,5‘wi)和三个进入3‘CDS(带内含子的3‘载体,3‘wi)。跨连接引物对用于RT-PCR(黑色箭头)。myc,myc标签。B;来自所示相应的构建体共转染的HEK293细胞以不同的循环次数进行RT-PCR。GAPDH作为上样对照。 \emptyset ,未转染的细胞。C,重构的ABCA4的代表性测序结果。

[0039] 图15:体内ABCA4的rAAV双载体mRNA反式剪接。A,用于体内表达的含BD_k的5‘载体和3‘载体构建体。hRho,人类视紫红质启动子。B,来自所示相应的构建体共转染的C54B1/6J野生型小鼠的视网膜裂解物以不同的循环次数进行RT-PCR。未注射的野生型(WT)小鼠作为阴性对照。NN/GL,AAV2的衣壳变体。-RT,阴性对照,无逆转录酶。GAPDH用作上样对照。C,重构的ABCA4的代表性测序结果。D,来自与图B所示相同的视网膜进行的定量RT-PCR的初步结果。共转染获得的ABCA4表达被归一化为未注射的WT视网膜。E,来自于B-D中显示的相同小鼠的视网膜裂解物的蛋白质印迹。抗myc抗体检测到人ABCA4蛋白,并用箭头指示。 β 微管蛋白作为上样对照。

[0040] 图16:编码天蓝蛋白5‘编码序列和结合域BD_g的微小基因序列(SEQ ID NO:35)。CMV启动子用灰色突出显示。天蓝蛋白的5‘编码序列(粗体字母)用灰色突出显示,并用粗实线下划线(NNN)。DSS用波浪线下划线(NNN)。结合域用斜体字母标记,并用虚线下划线(NNN),多聚腺苷酸化信号用双线下划线(NNN)。

[0041] 图17:编码天蓝蛋白3‘编码序列和结合域BD_g的微小基因序列(SEQ ID NO:36)。CMV启动子用灰色突出显示。结合域用斜体字母标记,并用虚线下划线(NNN)。ASS用波浪

线下划线(NNN)和粗体字母。天蓝蛋白的3'编码序列用灰色突出显示,并用粗实线下划线(NNN),多聚腺苷酸化信号用双线下划线(NNN)。

[0042] 图18:编码SpCas9-VPR的5'编码序列和结合域BD_k的微小基因序列(SEQ ID NO:37)。CMV启动子用灰色突出显示。SpCas9-VPR蛋白的5'编码序列用粗体字母突出显示,并用粗实线下划线(NNN)。DSS用波浪线下划线(NNN)。结合域用斜体字母标记,并用虚线下划线(NNN),多聚腺苷酸化信号用双线下划线(NNN)。

[0043] 图19:编码SpCas9-VPR的3'编码序列和结合域BD_k的微小基因序列(SEQ ID NO:38)。CMV启动子用灰色突出显示。结合域用斜体字母标出,并用虚线下划线(NNN)。ASS用波浪线下划线(NNN)和粗体字母。SpCas9-VPR蛋白的3'编码序列用灰色突出显示,并用粗实线下划线(NNN),多聚腺苷酸化信号用双线下划线(NNN)。

[0044] 图20:包含带内含子的ABCA4的5'编码序列的第一双AAV序列(SEQ ID NO:39):5'端和3'端的ITR序列用灰色突出显示,没有下划线(NNN)。间隔序列用小写字母表示。CMV启动子用灰色突出显示,并用实线下划线(NNN)。ABCA4蛋白的5'编码序列用虚线下划线其中用波浪下划线表示内含子(NNNNNNNNNN)。DSS用斜体字母表示,并带有下划线(Nnn,大写字母代表编码序列,小写字母代表非编码序列)。结合域以灰色突出显示,并用间断线下划线(NNN)。polyA序列用深灰色突出显示,并用波浪下划线(NNN)。

[0045] 图21:包含ABCA4的3'编码序列的第二双AAV的序列(SEQ ID NO:40):5'端和3'端的ITR序列用灰色突出显示,没有下划线(NNN)。间隔序列用小写字母表示。CMV启动子用灰色突出显示,并用实线下划线标出(NNN)。结合域用灰色突出显示,并用间断下划线(NNN)。受体剪接位点用斜体表示并加下划线(NNN)。ABCA4蛋白的3'编码序列用虚线下划线其中用波浪下划线表示内含子(NNNNNNNNNN)。polyA序列用深灰色突出显示,并用波浪下划线标出(NNN)。

具体实施方式

[0046] 发明人发现了如实施例中所描述的新型受体剪接区。该受体剪接区非常有效,可用于利用剪接的各种应用中。因此,本发明的受体剪接区尤其可用于使用pre-mRNA反式剪接分子的反式剪接、顺式剪接、SMaRT技术(剪接体介导的RNA反式剪接),用于重构分裂的AAV载体、包含2个AAV载体和/或陷阱载体的AAV载体系统。然而,本文所述的受体剪接区也可用于目的剪接的任何其他应用。

[0047] 本发明涉及核酸序列,其包括

[0048] (i) 受体剪接区序列,其包括

[0049] (ia) 嘧啶区,其中所述嘧啶区包括

[0050] (aa) 5至25个核苷酸;

[0051] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

- [0052] (ib) 受体剪接位点,
- [0053] ba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0054] bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
- [0055] (ii) 目的核苷酸序列,其中所述目的核苷酸序列
- [0056] (iia) 位于所述受体剪接区的3'或5',优选位于所述受体剪接区的3';
- [0057] 其中根据本发明所述的核酸序列优选在所述受体剪接区被切割,从而将目的核苷酸序列与受体剪接区序列分离。
- [0058] 本发明还涉及所述核酸序列用于在受体剪接区切割从而将目的核苷酸序列与受体剪接区序列分离的用途。
- [0059] 术语“核酸分子”、“核酸序列”或“核苷酸序列”在本文中同义使用,并包括具有核苷酸序列的任何核酸分子,所述核苷酸序列包含由所述核酸分子/序列包含的嘌呤和嘧啶碱基,其中所述碱基代表核酸分子的一级结构。核酸序列可包括DNA、cDNA、基因组DNA、RNA,既包括正义链也包括反义链。所述RNA例如可以是pre-mRNA、mRNA、tRNA或rRNA。本发明的多核苷酸可以由任何多核糖核苷酸或多脱氧核糖核苷酸组成,其可以是未修饰的RNA或DNA,或经修饰的RNA或DNA。本领域技术人员会理解,多脱氧核糖核苷酸中的胸腺嘧啶转录成多核糖核苷酸中的尿嘧啶。本文所指的序列通常是以可被转录成相应的RNA序列的DNA序列提供(在任何剪接事件之前)。因此,包含T的核酸序列也公开了相应的(转录的)RNA序列,优选pre-mRNA序列,包含U。
- [0060] 可以对DNA和RNA进行各种各样的修饰;因此,术语“核酸分子”或“核苷酸”可以包括化学、酶或代谢修饰的形式。“修饰的”碱基/核苷酸包括,例如,三苯甲基碱基和异常碱基(如肌苷)。
- [0061] 本发明的受体剪接区序列包括两个特征,即受体剪接位点和嘧啶区。所述嘧啶区包含5至25个核苷酸,其中在这些(共)5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U)。还可以预见,所述嘧啶区包含5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24或25个核苷酸,优选10至18个核苷酸,更优选12至16个核苷酸。另外地或替代地,这些(共)5至25个核苷酸中至少60%、65%、70%、75%、80%、85%、90%、95%、97%、98%、99%或100%的核苷酸是嘧啶碱基,例如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U)。例如,所述嘧啶区可以包含6个核苷酸。进一步设想所述嘧啶区包含序列TTTTTT或由序列TTTTTT组成。还预期所述嘧啶区包含序列TCTTTT或由序列TCTTTT组成。还预期所述嘧啶区包含序列TTTTTGTCA TTT (SEQ ID NO.11) 或由序列TTTTTTGTCA TTT (SEQ ID NO.11) 组成。还预期所述嘧啶区包含序列TCTTTTGTCA TCTA (SEQ ID NO.12) 或由序列TCTTTTGTCA TCTA (SEQ ID NO.12) 组成。还预期所述嘧啶区前面有7个核苷酸,一起包含序列CAACGAGTCTTTTGTCA TCTA (SEQ ID NO.13) 或由序列CAACGAGTCTTTTGTCA TCTA (SEQ ID NO.13) 组成。所述嘧啶区也可以包含TCTTTTGTCA TCT (SEQ ID NO.1) 的序列或由TCTTTTGTCA TCT (SEQ ID NO.1) 的序列组成。还预期本文所述的嘧啶区具有与SEQ ID NO.1、11、12或13的序列或TTTTTT或TCTTTT的序列至少70%、75%、80%、85%、90%、95%、97%、98%、99%或100%相同的核苷酸序列。在一个实施方案中,所述嘧啶区的5-25个核苷酸包含一个TTTTTT或TCTTTT序列。优选地,所述嘧啶区的最后一个嘧啶和所述受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基。术语“嘧啶区”在本文中可与术

语“多聚嘧啶区”互换使用,并可缩写为PPT或PolyC/T。“嘧啶区的最后一个嘧啶”是指嘧啶区5-25个核苷酸中最靠3'的嘧啶。

[0062] 根据本发明,在两个或多个核酸分子的上下文中,术语“相同”或“百分比同一性”是指两个或多个序列或亚序列是相同的,或者具有特定百分比的相同核苷酸(例如,至少95%、96%、97%、98%或99%同一性),当在比较窗口上进行比较和比对以获得最大一致性时,或者通过本领域已知的序列比对算法测量指定区域,或者通过手动比对和视觉检查。具有例如80%至95%或更高序列同一性的序列被认为是基本上相同的。这样的定义也适用于测试序列的互补序列。本领域技术人员将知道如何使用本领域已知的例如基于CLUSTALW计算机程序(Thompson *Nucl. Acids Res.* 2 (1994), 4673-4680)或FASTDB(Brutlag *Comp. App. Biosci.* 6 (1990), 237-245)的算法来确定序列之间的百分比同一性。

[0063] 本领域技术人员还可以使用BLAST和BLAST 2.0算法(Altschul *Nucl. Acids Res.* 25 (1977), 3389-3402)。核酸序列的BLASTN程序使用默认的字长(W)为28、期望阈值(E)为10、匹配/错配得分为1、-2,间隙成本为线性且为两条链的比较。对于氨基酸序列,BLASTP程序使用默认的字长(W)为6,期望阈值(E)为10,间隙成本为Existence:11和Extension:1。此外,可以使用BLOSUM62评分矩阵(Henikoff *Proc. Natl. Acad. Sci., USA*, 89, (1989), 10915)。

[0064] 例如,BLAST2.0,其代表Basic Local Alignment Search Tool(Altschul, *Nucl. Acids Res.* 25 (1997), 3389-3402; Altschul, *J. Mol. Evol.* 36 (1993), 290-300; Altschul, *J. Mol. Biol.* 215 (1990), 403-410),可用于搜索局部序列比对。

[0065] 如本文所述,除了嘧啶区之外,本发明的受体剪接区序列还包含受体剪接位点。如本文所用,“受体剪接位点”具有本领域技术人员已知的含义,尤其是如Alberts B, Johnson A, Lewis J, et al. (2002) “Molecular Biology of the Cell. 4th edition.” New York: Garland Science under the headline “from DNA to RNA”所述。本发明的受体剪接位点包含核苷酸NAGG,优选CAGG或CAGGT(或RNA中的CAGG或CAGGU)或由其组成。所述受体剪接位点通常位于被定义为“受体剪接区”的序列中。本文使用的术语“受体剪接位点”(缩写为ASS)指共有受体剪接序列,也可称为“剪接受体位点”(缩写为SAS)。在本发明的上下文中,术语“受体剪接区”和“受体剪接位点”用作不同的术语,以便区分该区域和共有受体剪接位点,其中所述受体剪接区包括所述受体剪接位点。在文献中,所述受体剪接区以及所述受体剪接位点可称为ASS或SAS。

[0066] 受体剪接位点被称为受体剪接位点是因为在该位点核酸被所谓的剪接体或其人工变体切割。剪接体及其工作原理也为技术人员所知。人类剪接体的结构例如在Zhang et al. (2017) “An atomic structure of the human spliceosome” *Cell* 169, 918-926中描述。人工剪接体包括,例如介导与剪接体相同的剪接位点切割的酶。人工剪接体可以包含如在Coppins和Silvermann (2005) “Mimicking the First Step of RNA Splicing: An Artificial DNA Enzyme Can Synthesize Branched RNA Using an Oligonucleotide Leaving Group as a 5'-Exon Analogue” *Biochemistry*, 44 (41), pp 13439-13446,以及Müller (2017) “Design and Experimental Evolution of trans-Splicing Group I Intron Ribozymes.” *Molecules*. 22 (1) .pii:E75. doi:10.3390/molecules22010075中所述额DNA酶。本文所述的受体剪接位点具有NAGG序列,其中N可以是任何核苷酸。例如,N是选自A、C、

T、G或U的核苷酸。本文所述的受体剪接位点也可以具有CAGG序列。本文所述的受体剪接位点也可以具有CAGGT (或RNA的CAGGU) 序列。

[0067] 本发明受体剪接位点的切割产生两个片段,一个包含NAGG和/或NAGGT剪接位点的NAG或CAGG剪接位点和/或CAGGT剪接位点的CAG。另一个片段包含NAGG或CAGG剪接位点的最后一个G,或NAGGT或CAGGT剪接位点的最后一个GT。因此,可以设想,在本文所述的用途、核酸序列、AAV载体、AAV载体系统和方法中,最终获得的(切割/剪接后的)核苷酸序列(目的)包含NAGG或CAGG序列的最后一个G或NAGGT或CAGGT序列的GT。本领域技术人员会理解,本文所指的剪接发生在RNA水平,并因此最终获得的(切割/剪接后的)RNA核苷酸序列(目的)包含NAGG或CAGG序列的最后一个G或pre-mRNA核苷酸序列的CAGGU序列的GU。然而,也预期在方法和用途中,最终核酸序列(目的)包含NAGG剪接位点序列的NAG,和/或CAGG或CAGGT剪接位点序列的CAG。因此,设想受体剪接位点的切割包括NAGG、CAGG和/或CAGGT (或CAGGU) 的受体剪接位点序列的G和G之间的切割,从而将核苷酸序列(目的)与内含子受体剪接区序列分离。

[0068] 本发明的受体剪接区尤其可用于包括反式剪接但也包括顺式剪接或任何其他剪接的应用中。拼接的发生可以被测量,如例如在Berger et al. (2016) “mRNA trans-splicing in gene therapy for genetic diseases” WIREs RNA 7:487-498中所描述的。例如,剪接的核苷酸序列可以通过使用特异性引物和探针的终点定量RT-PCR来定量,以区分目的剪接产物。剪接的其他测量可以如本文实施例中所述进行。例如,剪接可以通过使用标记基因来测量,如在Orenge et al. (2006) “A bichromatic fluorescent reporter for cell-based screens of alternative splicing.” Nucleic Acids Research 34(22): e148中所描述的。可以使用限制性酶分析来检测进一步的剪接,如在Berger et al., (2016) “Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing: a new approach for autosomal dominant retinitis pigmentosa.”, Mol Ther.; 23(5): 918-930中所描述的。

[0069] 受体剪接位点存在于pre-mRNA中时,通常对应于内含子的3'端和下一个外显子的5'端。当人工引入受体剪接位点(优选受体剪接区)时,并不绝对要求所述受体剪接位点位于内含子-外显子边界,但也可以位于开放阅读框内、内含子内或位于完整或部分开放阅读框的5'端。还设想受体剪接位点位于内含子的5'端。进一步设想所述受体剪接位点不位于pre-mRNA分子中,而是位于人工分子如任意核酸分子中。

[0070] 本发明的受体剪接位点包含在受体剪接区中。值得注意的是,本发明的受体剪接位点位于嘧啶区的3'端。

[0071] 受体剪接区(或受体剪接位点模块)可以进一步包含分支点和/或分支点序列。原则上,本发明考虑了任何合适的分支点或分支点序列。示例性的分支点或分支点序列尤其是如在Gao et al. (2008) “Human branch point consensus sequence is yUnAy” Nucleic Acid Research, vol. 36, no. 7, pp. 2257-2267; Mercer et al. (2016) “Genome-wide discovery of human splicing branch points” Genome Research 25:290-303中所描述的。

[0072] 例如,分支点序列可以包括序列UACUAAC或由序列UACUAAC组成。另外地或替代地,分支点序列可以包括序列YNYURAY或由序列YNYURAY组成,其中Y是U或C,并且其中R是A或G。

另外地或替代地,分支点序列可以包括序列YNCURAC或由序列YNCURAC组成,其中Y是U或C,并且其中R是A或G。另外地或替代地,分支点序列可以包括序列YNCURAY或由序列YNCURAY组成,其中Y是U或C,并且其中R是A或G。另外地或替代地,分支点序列可以包括序列CURAY或由序列CURAY组成,其中Y是U或C,并且其中R是A或G。另外地或替代地,分支点序列可以包括序列YUVAY或由序列YUVAY组成,其中Y是U或C,其中R是A或G,并且其中V是A、C或G。另外地或替代地,分支点序列可以包括序列CUSAY或由序列CUSAY组成,其中Y是U或C,其中S是G或C。另外地或替代地,分支点序列可以包括序列CUGAC或由序列CUGAC组成。另外地或替代地,分支点序列可以包括序列CUAAC或由序列CUAAC组成。分支点可以存在于C/U-U-N-A-C/U、CAACGA或CUCAA或GUCAA的分支点序列中(其中所述分支点序列由C/T-T-N-A-C/T、CAACGA或CTCAA或GTCAA各自的DNA序列编码)。所有相应序列中带下划线的腺苷代表分支点核苷酸。

[0073] 本发明的受体剪接区可以进一步包含分支点核苷酸序列(c),其中所述分支点核苷酸序列

[0074] (ca) 包括1-15个核苷酸;

[0075] (cb) 包括分支点核苷酸,优选腺苷(A);和

[0076] (cc) 位于嘧啶区和受体剪接位点的5'端。

[0077] 还预期分支点序列包含约3、4、5、6、7、8、9、10、11、12、13、14、15个或更多个核苷酸。进一步预期分支点序列包含6至8个核苷酸。还设想分支点序列包含8个核苷酸。然而,值得注意的是,分支点序列的核苷酸可以串联存在,这就是为什么它们也可以包含多于15个,例如总共30(15+15)、40、50、100、200或更多个核苷酸。

[0078] 分支点核苷酸可以是任何核苷酸。因此,分支点核苷酸可以是A(腺嘌呤)、T(胸腺嘧啶)、G(鸟嘌呤)、C(胞嘧啶)和U(尿嘧啶)中的任何一种。优选地,分支点核苷酸是A(腺嘌呤)。分支点核苷酸位于分支点序列内。

[0079] 受体剪接区可以另外地或替代地包含内含子剪接增强子。这种剪接增强子是本领域技术人员已知的,尤其记载在Wang et al.(2012)“Intronic splicing enhancers, cognate splicing factors and context-dependent regulation rules”Nature Structural&Molecular Biology,vol.19,no.10,pp.1044-1053中。两个例子是剪接增强子序列“AACG”(组F)和“CGAG”(组D)。然而,该术语包括任何合适的内含子剪接区。

[0080] 示例性内含子剪接增强子具有TGGGGGGAGG(SEQ ID NO:2)的序列。另一个示例性内含子剪接增强子具有GTAACGGC的序列。进一步设想内含子剪接增强子具有AACG的序列。还预期本文所述的受体剪接区具有与SEQ ID NO:2、GTAACGGC或AACG的序列至少70%、75%、80%、85%、90%、95%、97%、98%、99%或100%相同的核苷酸序列。

[0081] 本发明的受体剪接区优选在多聚嘧啶区的5'进一步包含约7个核苷酸(例如5-12个核苷酸,优选6-10个,更优选6-8个)。在一个优选的实施方案中,受体剪接区还包含(a)嘧啶区5'端的7个核苷酸,具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;(b)嘧啶区5'的7个核苷酸,具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;(c)嘧啶区5'的7个核苷酸,具有序列CAACGAG的至少6个核苷酸,其中5'第一个核苷酸是C;(d)嘧啶区5'的7个核苷酸,具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAA;(e)嘧啶区5'的7个核苷酸,具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAAC;(f)嘧啶区5'的7个核苷酸,具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷

酸是CAA; (g) 嘧啶区5'的7个核苷酸,具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC; (h) 嘧啶区5'的7个核苷酸,具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAA; (i) 嘧啶区5'的7个核苷酸,具有序列CACAGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAAC,或(j) 嘧啶区5'的7个核苷酸,具有序列CACACAGAG。不受理论限制,该序列可以作为剪接增强子起作用。

[0082] 进一步设想受体剪接区包含1、2、3、4、5个或更多不同/相同的分支点序列。进一步设想受体剪接区包含1、2、3、4、5个或更多不同或相同的内含子剪接增强子(序列)。

[0083] 可以预见,嘧啶区、受体剪接位点以及任选的分支点序列和/或内含子剪接增强子,因此整个受体剪接区总共包括约1000、500、250、100、50、45、40、35、30、25、20、15个或更少的核苷酸,优选26个核苷酸。

[0084] 在一些实施方案中,特定受体剪接位点使用的分支点序列或分支点核苷酸可以是灵活的。这意味着可以使用不同的分支点,例如3个或更多、5个或更多、7个或更多、9个或更多、或11个或更多。在一个实施方案中,每个分支点的使用频率<30%或<20%。在特定受体剪接位点用的分支点序列或分支点核苷酸的选择是灵活的情况下,受体剪接区的强度不源于也不依赖特定包含的分支点序列或分支点核苷酸的存在。使用任何分支点序列或分支点核苷酸的灵活性可使受体剪接区高效和通用,并且特别适合用于不同的核酸序列以不依赖于序列的方式诱导有效剪接。

[0085] 本文所述以及存在于核酸、pre-mRNA反式剪接分子、AVV载体、AVV载体系统中的受体剪接区,以及用于本发明方法中的受体剪接区,除了受体剪接位点和多聚嘧啶区外,还包括多聚嘧啶区的5'核苷酸。在一个优选的实施方案中,所述受体剪接区在所述多聚嘧啶区5'包含约7个核苷酸(例如5-12个核苷酸,优选6-10个,更优选6-8个)。在特别优选的实施方案中,所述受体剪接区包含多聚嘧啶区5'的序列CACACAGAG。包含受体剪接位点、多聚嘧啶区和多聚嘧啶区5'的另外的核苷酸(优选多聚嘧啶区5'的另外的大约7个核苷酸)的受体剪接区被用作最小的受体剪接区,并且可以插入任何核酸序列中以引入非常有效的功能性受体剪接区。

[0086] 进一步设想,本发明的受体剪接区包含CAACGAGTCTTTTGTCATCTACAGGT (SEQ ID NO:3)的序列或由CAACGAGTCTTTTGTCATCTACAGGT (SEQ ID NO:3)的序列组成。进一步考虑本发明的受体剪接位点包含CAACGAGTTTTTGTGTCATCTACAGGT (SEQ ID NO:4)的序列或由CAACGAGTTTTTGTGTCATCTACAGGT (SEQ ID NO:4)的序列组成。还可以设想,受体剪接区包括以下任意一种或由以下任意一种组成:CTACACGCTCAAGCCGAGGTCAACAACGAGTCTTTTGTCATCTACAGGT (SEQ ID NO:5),GCCGGAGGTCAACAACGAGTCTTTTGTCATCTACAGGT (SEQ ID NO:6),GTCTTTTGTCATCTACAGGT (SEQ ID NO:7),GTCTTTTGTCATCTACAGGTGTTTCGTG (SEQ ID NO:26) or GTCTTTTGTCATCTACAGGTGTTTCGTGGTTCGTGGTCCA (SEQ ID NO:8)。优选地,本发明的受体剪接区包含或由CAACGAGTCTTTTGTCATCTACAGGT (SEQ ID NO:3)或CAACGAGTTTTTGTGTCATCTACAGGT (SEQ ID NO:4)的序列组成(或由包含或由CAACGAGTCTTTTGTCATCTACAGGT (SEQ ID NO:3)或CAACGAGTTTTTGTGTCATCTACAGGT (SEQ ID NO:4)的序列组成的核酸序列编码)。

[0087] 还考虑本文所述的受体剪接区具有与SEQ ID NO:3、4、5、6、7、26或8中任一个的任何序列至少70%、75%、80%、85%、90%、95%、97%、98%、99%或100%相同的核苷酸序列。优选地,所述受体剪接区包含与SEQ ID NO:3、4、5、6、7、26或8中任一个的任何序列至少

85%、90%、95%、97%、98%、99%或100%相同的核苷酸序列,更优选与SEQ ID NO:3或4中的任何序列至少85%、90%、95%、97%、98%、99%或100%相同的核苷酸序列,甚至更优选与SEQ ID NO:3或4中的任何序列至少90%、95%、97%、98%、99%或100%相同的核苷酸序列。

[0088] 根据本发明所述的核酸序列除了受体剪接位点之外,还可以包含目的核苷酸序列。所述目的核苷酸序列可以是任何合适的核苷酸序列。术语“目的核苷酸序列”和“目的核酸序列”在本文中可互换使用。通常将其称为“目的核苷酸序列”,以更好地将该术语与本发명의核酸序列或核酸分子区分开来。例如,目的核苷酸序列可以是内含子或外显子核苷酸序列,或者同时包含内含子和外显子(即非编码和编码)序列。进一步预期目的核苷酸序列是cDNA、mRNA、rRNA或tRNA。目的核苷酸序列或其一部分可以位于受体剪接区的3'或5',优选3'。在某些实施方案中,包含在任何核酸序列、pre-mRNA反式剪接分子、AAV载体、AAV载体中或用于根据本发明所述的任何方法中的目的核苷酸序列或其部分可以是转基因或其部分。已经发现,剪接效率,特别是反式剪接效率,可以通过目的核苷酸序列或其一部分中内含子的存在来提高。特别是,对于目的核苷酸序列或其大于1000bp的部分,在序列中引入或保持内含子可能是有利的。优选地,内含子存在于目的核苷酸序列或其一部分的大约每200至1000bp处,但是也可以以更低或更高的序列间隔出现。内含子可以来自目的核苷酸序列和/或来自不同的基因和/或人工序列。

[0089] 本文所用的目的核苷酸序列(也称为目的核酸序列)也可以是编码序列,例如编码目的多肽或其部分的序列。在一个优选的实施方案中,目的核苷酸序列是编码序列,更优选转基因,编码目的多肽或其一部分的序列。

[0090] 本发明进一步设想,目的核苷酸序列编码治疗性多肽、治疗性核酸(如治疗性RNA)或其一部分。

[0091] 治疗性核酸或治疗性多肽可用于治疗眼部疾病,如常染色体隐性严重早发性视网膜变性(Leber氏先天性黑蒙症)、先天性色盲、Stargardt病、Best病(卵黄样黄斑变性)、Doyme病、色素性视网膜炎(特别是常染色体显性、常染色体隐性、X连锁、二基因或多基因色素性视网膜炎)、(X连锁)视网膜劈裂、黄斑变性(AMD)、年龄相关性黄斑变性、萎缩性年龄相关性黄斑变性、新生血管性AMD、糖尿病性黄斑病变、增殖性糖尿病视网膜病变(PDR)、囊样黄斑水肿、中心性浆液性视网膜病变、视网膜脱离、眼内炎症、青光眼、后葡萄膜炎、先天性静止性夜盲症、脉络膜视网膜病变、早发性视网膜营养不良、视锥、视杆-视锥或视锥-视杆营养不良、模式营养不良、Usher综合征和其他综合征性睫状体炎,如Bardet-Biedl综合征、Joubert综合征、Senior-Løken综合征或Alström综合征。

[0092] 核酸、治疗性核酸、治疗性蛋白质/多肽或治疗性分子可用于治疗影响感光细胞(如视杆和/或视锥细胞)的病症(感光细胞疾病)。感光细胞疾病的非限制性例子包括色盲、年龄相关性黄斑变性、视网膜变性、视网膜营养不良、视网膜色素变性、视锥细胞营养不良、视杆-视锥营养不良、色素失明、黄斑变性、夜盲症、视网膜分裂、脉络膜出血、糖尿病性视网膜病变、遗传性视神经病变、Oguchi疾病I型、点状视网膜炎(RPA)、进行性视网膜萎缩(PRA)、白点状眼底(FA)或先天性静止性夜盲(CSNB)。

[0093] 治疗性核酸或治疗性多肽可用于治疗遗传性视网膜疾病(IRD)。IRD是一组遗传和表型异质性疾病。它们是15至45岁人群失明的主要原因,估计患病率为1/1500-1/3000。可

以根据主要受影响的视网膜细胞类型(即,视杆感光细胞和视锥感光细胞)和根据疾病状态对IRD进行分组。IRD的非限制性例子是色盲、年龄相关性黄斑变性、视网膜变性、视网膜营养不良、色素性视网膜炎、视锥营养不良、视杆-视锥营养不良、色盲、黄斑变性、视网膜劈裂、脉络膜视网膜炎、糖尿病性视网膜炎、遗传性视神经病变、Oguchi病I型、点状视网膜炎(RPA)、进行性视网膜炎(PRA)、眼底白斑病(FA)或先天性静止性夜盲症(CSNB)。

[0094] 主要影响视锥细胞的最常见的进行性IRD病是Stargardt黄斑营养不良。它以常染色体隐性方式遗传,主要由编码视网膜转运蛋白ABCR的ABCA4基因突变引起。在一个实施方案中,治疗性核酸编码ABCA4或其一部分,其中治疗性核酸用于治疗Stargardt黄斑营养不良。

[0095] 主要影响视杆感光细胞的最常见的IRD病是色素性视网膜炎(RP)。RP是一种进行性疾病,可导致夜盲症和隧道视觉。在后期,继发性视锥感光细胞死亡被诱导,最终导致完全的视力丧失。色素性视网膜炎的一些形式与视紫红质编码基因的突变有关。在一个实施方案中,治疗性核酸编码视紫红质或其一部分,其中治疗性核酸用于治疗色素性视网膜炎。

[0096] 根据本发明所述的核酸序列还可以包含供体剪接位点,其中所述供体剪接位点位于目的核苷酸序列或其一部分的3'端。术语“供体剪接区”在本文中可与“供体剪接位点”(缩写为DSS)或“剪接供体位点”(缩写为SDS)互换使用。因此,本发明的核酸序列也可以包含供体剪接位点。供体剪接位点是本领域技术人员已知的,并且尤其记载在Buckley et al. (2009) “A method for identifying alternative or cryptic donor splice sites within gene and mRNA sequences. Comparisons among sequences from vertebrates, echinoderms and other groups” BMC Genomics 10:318和Qu et al. (2017) “A Bioinformatics-Based Alternative mRNA Splicing Code that May Explain Some Disease Mutations Is Conserved in Animals” Front. Genet., vol. 8 Art. 38中。本文,该术语涵盖任何合适的供体剪接位点。供体剪接位点通常位于外显子的3'端和内含子侧翼的5'端。然而,在本发明的上下文中,供体剪接位点也可以位于目的核苷酸序列内的其他地方。与受体剪接位点类似,供体剪接位点允许剪接体在其位点切割。

[0097] 原则上,本文所述的供体剪接位点可以位于核酸分子的任何位置。还设想供体剪接位点位于(第一)核酸序列的3'端。例如,供体剪接位点序列可以包含或由以下序列组成:AAGGTAAGT、AAGGTGAGT、CAGGTAAGT、CAGGTGAGT、AAGGTAAG、AAGGTGAG、CAGGTAAG或CAGGTGAG。还设想供体剪接位点序列与以下序列具有至少70%、80%、85%、90%、95%、98%、99%或100%的序列同一性:AAGGTAAGT、AAGGTGAGT、CAGGTAAGT、CAGGTGAGT、AAGGTAAG、AAGGTGAG、CAGGTAAG或CAGGTGAG。

[0098] 类似地,本文所述的受体剪接区可以位于核酸分子的任何位置。还设想受体剪接位点位于(第二)核酸序列的5'端。

[0099] 目的核苷酸序列可以附加地或可选地包含本文所述的多聚腺苷酸化信号和/或启动子。

[0100] 本发明还涉及产生核酸序列的方法,所述方法包括

[0101] (A) 提供第一核酸序列,其包含一个或多个供体剪接位点序列;

[0102] (B) 提供第二核酸序列,其中所述第二核酸序列包括

[0103] (i) 受体剪接区序列,其包括

[0104] (ia) 嘧啶区,其中所述嘧啶区包括

[0105] (iaa) 5至25个核苷酸;

[0106] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0107] (ib) 受体剪接位点,

[0108] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0109] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;

[0110] 从而获得核酸序列。所述方法优选为体外方法。所述方法可以进一步包括在一个或多个供体剪接序列处剪接第一核酸序列和在受体剪接位点处剪接第二核酸序列以获得剪接核酸序列的步骤。所述方法还可以包括步骤(C)在一个或多个供体剪接位点序列切割第一核酸序列,并在受体剪接位点切割第二核酸序列,和(D)将第一切割核酸序列连接到第二切割核酸序列。

[0111] 这种方法反映了所谓的反式剪接。特别地,反式剪接是一种剪接类型,其中来自两个分离的核酸分子(优选pre-mRNA分子)的外显子连接在一起形成一个核酸分子。例如,当pre-RNA剪接为最终的mRNA、tRNA或rRNA时。该方法也是技术人员已知的,并且尤其记载在Alberts B,Johnson A,Lewis J,et al.(2002)“Molecular Biology of the Cell.4th edition.New York:Garland Science under the headline“from DNA to RNA”中。因此,剪接的核酸序列优选是mRNA、rRNA或tRNA序列。因此,目的核苷酸序列可以是pre-mRNA。

[0112] 第一和第二核酸序列可以是DNA或RNA序列。在第一和第二核酸序列都是DNA序列的情况下,第一和第二核酸序列还包含启动子,并且至少第二核酸序列,优选两者都还包含转录终止序列,例如polyA序列。在第一和第二核酸序列是RNA序列的情况下,至少第二核酸序列(优选两种核酸序列)包含终止序列,例如polyA序列。优选地,所述核酸序列是DNA序列,更优选地是包含所述DNA序列的载体或质粒。

[0113] 在一个实施方案中,所述第一和第二核酸序列进一步包含结合域。在又一个实施方案中,所述第一和第二核酸序列是pre-mRNA反式剪接分子或编码如本文所述的pre-mRNA反式剪接分子的DNA序列。

[0114] 不受理论的限制,在一个或多个供体剪接位点序列切割第一核酸序列和在受体剪接区序列中切割第二核酸序列并将第一切割核酸序列连接到第二切割核酸序列是在顺式剪接和反式剪接中使用两个酯交换反应来实现的,其在本文中被称为“剪接”。本文使用的术语“反式剪接”指的是RNA剪接,特别是两个分离的RNA或pre-RNA分子的pre-RNA剪接,以形成成熟的RNA,如最终的mRNA、tRNA或rRNA。

[0115] 术语“体外方法”指体外的方法,即,人体或动物体,但包括体外的细胞系或原代细胞。

[0116] 本文所述的方法和用途可以在任何合适的无细胞系统如细胞裂解物或任何合适的宿主细胞中进行。

[0117] 这种合适的宿主细胞是技术人员已知的。例如,合适的宿主细胞含有剪接机制,即,剪接体。本领域普通技术人员可以选择调节元件用于合适的宿主细胞,例如哺乳动物或人类宿主细胞。调节元件包括例如启动子、终止序列(如转录终止序列、翻译终止序列)、增强子和多聚腺苷酸化元件。此外,转基因(目的核酸序列)的表达水平可以通过使用调节元

件如增强子序列或土拨鼠肝炎病毒转录后应答元件 (WPRE) 来增强。然而, 这些元件进一步限制了 rAAV 的包装能力。因此, 在某些情况下, 特别是在大小限制很关键的情况下, 建议省略这些任选元件, 并优选使用小的强制性调节元件, 例如短的多聚腺苷酸化信号作为 AAV 载体 (特别是根据本发明所述的双 rAAV 载体系统) 的终止序列。宿主细胞可以是真核细胞或哺乳动物细胞, 例如人或啮齿动物细胞系。宿主细胞可以是 HEK 细胞 (例如 HEK293 (T) 细胞)、Cos7 细胞、CHO 细胞、如人或鼠来源的成纤维细胞、视网膜母细胞瘤细胞、661W 细胞、诱导的多能干细胞 (iPSC) (如人 iPSC)、感光细胞 (如来自脊椎动物的)、神经元细胞或神经胶质细胞。

[0118] 在一个实施方案中, 所述方法在体外宿主细胞中进行 (即体外), 包括引入所述第一核酸序列和引入所述第二核酸序列。所述第一核酸序列和所述第二核酸序列可以通过转染或转导 () 更具体地说是共转染或共转导) 引入。

[0119] 因此, 在一个实施方案中, 所述方法包括

[0120] (A) 提供包含一个或多个供体剪接位点序列的第一核酸序列;

[0121] (B) 提供第二核酸序列, 其中所述第二核酸序列包括

[0122] (i) 受体剪接区序列, 其包括

[0123] (ia) 嘧啶区, 其中所述嘧啶区包括

[0124] (iaa) 5 至 25 个核苷酸;

[0125] (iab) 其中在这 5 至 25 个核苷酸中至少 60% 的核苷酸是嘧啶碱基, 如胞嘧啶 (C)、胸腺嘧啶 (T) 和/或尿嘧啶 (U);

[0126] (ib) 受体剪接位点,

[0127] (iba) 其中所述受体剪接位点位于嘧啶区的 3' 端; 和

[0128] (ibb) 其中所述受体剪接位点包含 5' 至 3' 的 NAGG 序列, 其中 N 是 A、C、T/U 或 G; 和

[0129] (ii) 目的核苷酸序列, 其中所述目的核苷酸序列

[0130] (iia) 位于受体剪接区的 3' ;

[0131] 从而获得核酸序列。所述方法可以进一步包括在一个或多个供体剪接位点序列处剪接第一核酸序列和在受体剪接位点处剪接第二核酸序列以获得剪接核酸序列的步骤。所述方法还可以包括步骤 (C) 在一个或多个供体剪接位点序列切割第一核酸序列, 并在受体剪接位点切割第二核酸序列, 和 (D) 将第一切割核酸序列连接到第二切割核酸序列, 从而获得核酸序列。

[0132] 在两个核酸序列都包含目的核苷酸序列的情况下, 该方法还包括以下特征: 第一核苷酸序列包含 (ii) 目的核苷酸序列, 其中所述目的核苷酸序列位于供体剪接位点的 5'。

[0133] 第二和第一核酸序列都可以包含如本文所述的目的核苷酸序列。这种目的核苷酸序列可以位于供体剪接位点的 5' 端。目的核苷酸序列可以另外地或替代地位于受体剪接区的 3' 端。在本发明的另一个实施方案中, 多核苷酸包含以下排列顺序: 优选地, 5' - (目的序列) - (供体剪接位点) - 3' 结合以 5' - (受体剪接区) - (目的序列) - 3'。替代地, 5' - (目的序列) - (受体剪接区) - 3' 结合以 5' - (供体剪接位点) - (目的序列) - 3'。

[0134] 在一个优选的实施方案中, 目的核苷酸序列是编码多肽的序列, 并且第一核酸序列包含编码多肽的序列的 5' 部分, 第二核酸序列包含编码多肽的序列的 3' 部分, 并且反式剪接重构编码多肽的序列的 5' 部分和 3' 部分。

[0135] 切割的核酸序列的连接是技术人员已知的过程。例如,该连接过程可以通过连接分子进行。例如,连接分子可以是RNA连接酶或具有RNA连接酶功能的蛋白质。这种RNA连接酶是本领域技术人员已知的,尤其记载在例如Chambers和Patrick (2015) “Archaeal Nucleic Acid Ligases and Their Potential in Biotechnology” Hindawi Publishing Corporation Archaea Volume 2015, Article ID 170571, 10页中。一些(t) RNA连接酶例如记载在Popow et al. (2012) “Diversity and roles of (t) RNA ligases” Cell Mol Life Sci. 69 (16) :2657-70中。因此,包括连接步骤的方法可以进一步包括添加RNA连接酶的步骤。所述RNA连接酶例如可以是mRNA连接酶、tRNA连接酶或rRNA连接酶。描述例如连接的mRNA分子可如何分析的方法显示在实施例中。

[0136] 在优选实施方案中,该方法在宿主细胞中进行,更优选在体外进行(即体外)。因此,将第一和第二核酸序列引入宿主细胞,其中所述第一和第二核酸序列是重组核酸序列和/或对宿主细胞是异源的。将第一和第二核酸引入宿主细胞包括转染或转导。转染可以是DNA或RNA转染,转染DNA或RNA的方法是本领域技术人员已知的。优选地,核酸序列是DNA序列,并且DNA序列以包含所述DNA序列的质粒或载体的形式转染。本文所指的转导是指使用病毒载体导入核酸,其中病毒载体可以包含DNA或RNA。在优选的实施方案中,病毒载体是DNA病毒载体,可以包含单链DNA(如AAV)或双链DNA。

[0137] 在一个实施方案中,根据本发明所述的方法包括(A)将第一核酸序列导入宿主细胞,所述第一核酸序列包含pre-mRNA反式剪接分子序列或编码所述pre-mRNA反式剪接分子的核酸序列,其中第一pre-mRNA反式剪接分子从5'至3'包含,(a)目的核苷酸序列的5'部分;(b)供体剪接位点;(c)任意的间隔序列;(d)第一结合域;和(e)任意的终止序列,优选polyA序列;和(B)将第二核酸序列引入宿主细胞,所述第二核酸序列包含pre-mRNA反式剪接分子序列或编码所述pre-mRNA反式剪接分子的核酸序列,其中第二pre-mRNA反式剪接分子从5'至3'包含;(i)第二结合域,其与第一核酸序列的第一结合域互补;(ii)受体剪接区序列,其包括:(iia)嘧啶区,其中所述嘧啶区包括:(iiaa)5至25个核苷酸;(iiab)其中这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);(iib)受体剪接位点,(iiba)其中所述受体剪接位点位于所述嘧啶区的3'端;和(iibb)其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和(iii)目的核苷酸序列的3'部分。该方法可以进一步包括(C)在所述供体剪接位点序列切割所述第一核酸序列,并在所述受体剪接位点切割第二核酸序列;(D)将包含目的核苷酸序列5'部分的第一切割核酸序列与包含目的核苷酸序列3'部分的第二切割核酸序列连接,从而获得目的核酸序列。在一个优选的实施方案中,第一和第二核酸序列是编码所述pre-mRNA反式剪接分子的DNA序列,并且第一核酸序列进一步包含目的核酸序列的5'部分5'的启动子,第二核酸序列进一步包含第二结合域5'的启动子。

[0138] 本发明还涉及核酸序列,其包括

[0139] (i)受体剪接区序列,其包括

[0140] (ia)嘧啶区,其中所述嘧啶区包括

[0141] (iaa)5至25个核苷酸;

[0142] (iab)其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

- [0143] (ib) 受体剪接位点,
- [0144] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0145] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
- [0146] (ii) 目的核苷酸序列,其中所述目的核苷酸序列
- [0147] (iia) 位于受体剪接区的3'或5';
- [0148] 其中目的核苷酸序列不是SEQ ID NO:9的视紫红质基因的外显子3,或者其中目的核苷酸序列不包含SEQ ID No:9和/或10中定义的序列,或者其中目的核苷酸序列不包含SEQ ID NO:9的视紫红质基因的外显子3。
- [0149] 本发明还涉及核酸序列,其包括
- [0150] (i) 受体剪接区序列,其包括
- [0151] (ia) 嘧啶区,其中所述嘧啶区包括
- [0152] (iaa) 5至25个核苷酸;
- [0153] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0154] (ib) 受体剪接位点,
- [0155] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0156] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
- [0157] (ii) 目的核苷酸序列,其中所述目的核苷酸序列
- [0158] (iia) 位于受体剪接区的3'或5';
- [0159] 其中目的核苷酸序列不是SEQ ID NO.10的视紫红质mRNA。
- [0160] 设想本发明的核酸序列具有至多5800个核苷酸的长度。进一步设想本发明的核酸序列具有至多5500、5000、4500、4000、3500、3000、2500、2000、1500、1000、500、400、300、200或至多150个核苷酸的长度。
- [0161] 本发明的受体剪接区可用于核酸分子的另一个应用是用于Smart技术的所谓外源pre-mRNA反式剪接分子。可以在其中引入本发明的受体剪接区的此类pre-mRNA反式剪接分子尤其记载于WO 2011/042556、WO2013/025461和Berger et al. (2016) “mRNA trans splicing in gene therapy for genetic diseases”WIREs RNA,7:487-498,Puttaraju et al. (1999) “Spliceosome-mediated RNA trans-splicing as a tool for gene therapy”Nature Biotechnology,vol.17,pp.246-252;以及Mansfield et al. (2003) “5' Exon replacement and repair by spliceosome-mediated RNA trans-splicing”RNA, vol.9:1290-1297中。从这些参考文献中,技术人员也知道如何构建这种pre-mRNA反式剪接分子。本文还描述了一些示例性的pre-mRNA反式剪接分子核酸分子。
- [0162] 本发明还涉及一种pre-mRNA反式剪接分子,其包括
- [0163] (i) 受体剪接区,其包括
- [0164] (ia) 嘧啶区,其中所述嘧啶区包括
- [0165] (iaa) 5至25个核苷酸;
- [0166] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)或尿嘧啶(U);
- [0167] (ib) 受体剪接位点,

- [0168] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0169] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、U或G;
- [0170] (ii) 目的核苷酸序列;其中所述受体剪接区位于目的核苷酸序列的5'端;
- [0171] (iii) 靶向pre-mRNA的结合域,其位于目的核酸序列的5'端;和
- [0172] (iv) 任选的间隔序列,其中所述间隔序列位于结合域和受体剪接区之间。
- [0173] 在一个优选的实施方案中,所述受体剪接区位于目的核苷酸序列的5'端,并且所述结合域位于受体剪接区的5'端。优选地,所述核酸序列或所述pre-mRNA反式剪接分子还包含终止序列,优选polyA序列。
- [0174] 本发明还涉及一种pre-mRNA反式剪接分子,其包括
- [0175] (i) 受体剪接区,其包括
- [0176] (ia) 嘧啶区,其中所述嘧啶区包括
- [0177] (iaa) 5至25个核苷酸;
- [0178] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0179] (ib) 受体剪接位点,
- [0180] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0181] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、U或G;
- [0182] (ii) 目的核苷酸序列,其中所述受体剪接区位于目的核苷酸序列的5'端;
- [0183] (iii) 供体剪接位点,其中所述供体剪接位点位于目的核苷酸序列的3'端;
- [0184] (iv) 第一结合域,其靶向位于目的核苷酸序列5'的pre-mRNA;
- [0185] (v) 第二结合域,其靶向位于目的核苷酸序列3'的pre-mRNA;
- [0186] (vi) 任选的第一间隔序列,其中所述第一间隔位于第一结合域和受体剪接区之间;
- [0187] (vii) 任选的第二间隔序列,其中所述第二间隔位于所述第二结合域和供体剪接位点之间。
- [0188] 在一个实施方案中,所述第一结合域位于受体剪接区5',所述第二结合域位于供体剪接位点3'。本领域技术人员会理解,编码pre-mRNA反式剪接分子的DNA序列包含启动子,以便转录pre-mRNA反式剪接分子(其为RNA分子)。术语“靶向pre-mRNA的结合域”也可以称为“靶向pre-mRNA的结合域”,其可以位于目的核苷酸序列的5'或3'。
- [0189] 本文所述的核酸序列可以反映pre-mRNA反式剪接分子。本领域技术人员会理解,本文所述的核酸序列和pre-mRNA反式剪接分子是重组序列或分子。这些pre-mRNA反式剪接分子在本领域中是已知的,并且尤其记载在WO 2011/042556、WO2013/025461和Berger et al. (2016) “mRNA trans splicing in gene therapy for genetic diseases”WIREs RNA, 7:487-498,Puttaraju et al. (1999) “Spliceosome-mediated RNA trans-splicing as a tool for gene therapy”Nature Biotechnology, vol.17, pp.246-252, Mansfield et al. (2003) “5' Exon replacement and repair by spliceosome-mediated RNA trans-splicing”RNA, vol.9:1290-1297以及Berger et al. (2016) “mRNA trans-splicing in gene therapy for genetic diseases”WIREs RNA 7:487-498中。因此,技术人员知道如何构建那些pre-mRNA反式剪接分子。设想pre-mRNA反式剪接分子与靶pre-mRNA结合,其中

pre-mRNA可以是天然的pre-mRNA,特别是宿主细胞内源性的pre-mRNA,或重组的pre-mRNA,特别是另一种pre-mRNA反式剪接分子。还考虑pre-mRNA反式剪接分子比顺式剪接反应更有效地优先诱导反式剪接反应。

[0190] 本文使用的术语“靶向pre-mRNA的结合域”可以是任何合适的“靶向pre-mRNA的结合域”,这意味着它与位于目的核苷酸序列5'或3'的靶pre-mRNA序列互补。本文所用的“靶向pre-mRNA的结合域”也可称为“靶结合域”、“结合域”(缩写为BD)或“结合序列”。例如,结合域可以通过碱基配对识别靶pre-mRNA或mRNA。例如,mRNA上的靶可以是内含子。本文所述的pre-mRNA反式剪接分子的靶结合域可包含一个或两个至少15至30个核苷酸的结合域,优选80-120个核苷酸,更优选约100个核苷酸;或者具有如美国专利公开号US 2006-0194317 A1所述的长靶结合域,其由多达几百个核苷酸组成,这些核苷酸与所选择的(例如内源性)pre-mRNA的被靶向区域互补并处于反义方向。这赋予了结合的特异性,并将内源性pre-mRNA紧密地锚定在空间中,以便剪接体(例如宿主细胞核的)可以将pre-mRNA反式剪接分子的一部分反式剪接到(例如内源性)pre-mRNA的一部分。替代地,结合域可以与另一种重组pre-mRNA(例如另一种pre-mRNA反式剪接分子)的结合域反向互补。这赋予了两种pre-mRNA反式剪接分子在空间紧密结合和锚定的特异性。可以将一种pre-mRNA反式剪接分子的一部分反式剪接到另一种pre-mRNA反式剪接分子的一部分上。因此,设想结合域包含约15至250个核苷酸、约15至200个核苷酸、约100至200个核苷酸或少于500、400、300或200个核苷酸。在一个实施方案中,结合域包含50-150个核苷酸,优选80-120个核苷酸,甚至更优选90-110个核苷酸或约100个核苷酸。替代地或更优选地,有效的结合域具有45-65%的GC含量和/或来源于内含子真核序列或细菌序列和/或在最后30bp包含至少一个分支点(例如,使用human splicing finder 3.1(<http://www.umd.be/HSF/index.html>)进行评估)。

[0191] 在宿主细胞内,所述结合域可以结合内源性pre-mRNA或第二异源性或重组的pre-mRNA(提供或引入宿主细胞),只要所述内源性pre-mRNA或第二异源性或重组的pre-mRNA包含与靶结合域反向互补的序列。因此,对于第二异源性或重组的pre-mRNA,其包含与第一异源性pre-mRNA的结合域反向互补的结合域。术语“内源性”和“异源性”在本文中相对于宿主细胞使用。靶向内源性pre-mRNA也被称为剪接体介导的RNA反式剪接,并且通常用于替代内源性pre-mRNA的一部分。靶向第二异源性或重组的pre-mRNA也被称为反式剪接,并产生新的重组的mRNA,将来自一个pre-mRNA反式剪接分子的第一核苷酸序列或其一部分(5')连接到另一个pre-mRNA反式剪接分子的第二核苷酸序列或其一部分(3')上。如本文所用,术语“重组的”是指通过实验室遗传重组方法如分子克隆形成的DNA分子或序列,以及由所述重组的DNA分子或序列编码的RNA或多肽分子/序列。

[0192] 术语“pre-mRNA”或“pre-RNA”是指反式剪接前的RNA,通常描述顺式剪接前的RNA分子(例如,包含内含子和外显子),但也可以指顺式剪接后的RNA分子(例如,仅包含外显子),例如在结合域靶向跨越外显子-外显子边界的序列的情况下。

[0193] 本文使用的术语“异源性”是指通过实验转入或引入细胞的蛋白质或核酸分子或序列,因此不是该细胞内源的。其中异源性蛋白质或核酸分子或序列可以来自与受体不同的细胞类型或不同的物种,它也可以来自与受体相同的细胞类型或物种,只要它被引入受体宿主细胞。如本文所指,术语“异源性”和“重组的”可互换使用。

[0194] 适用于本发明但不限于此的示例性结合域是具有由SEQ ID NO:27、28、32、33或18

的序列编码的RNA序列的结合域。进一步的靶结合域可以使用报告重构测定来鉴定,例如用于确定重构效率的天蓝蛋白重组测定,如实施例2和5中所记载和使用的或如Riedmayr LM. (2020, SMaRT for Therapeutic Purposes. *Methods Mol Biol* 2079:219-32. doi: 10.1007/978-1-4939-9904-0_17) 或Dallinger G. et al. (2003, Development of spliceosome-mediated RNA trans-splicing (SMaRT) for the correction of inherited skin diseases. *Exp Dermatol* 12:37-46) 中所记载的。

[0195] 第二结合区可以位于分子的3'端,并且可以被引入到本发明的pre-mRNA反式剪接分子中。绝对互补虽然是优选的,但不是必需的。如本文所述,与内源性pre-mRNA的一部分“互补”的序列是指具有足够互补性的序列,能够与内源性pre-mRNA杂交,形成稳定的双链体。杂交的能力将取决于核酸的互补程度和长度(例如,见Sambrook et al., 1989, *Molecular Cloning, A Laboratory Manual*, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.)。

[0196] 因此,关于靶向pre-mRNA的结合域所使用的互补性意味着结合域包含与pre-mRNA上的靶序列至少60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%互补,优选与pre-mRNA上的靶序列至少90%、95%、98%、99%或100%互补的序列。

[0197] 将剪接位点与结合域分隔的间隔区优选也包含在pre-mRNA反式剪接分子中。所述间隔序列是pre-mRNA反式剪接分子的一个区域,它可以通过相对弱的互补性覆盖pre-mRNA反式剪接分子的3'和/或5'剪接位点的元件,从而防止非特异性反式剪接。

[0198] 分隔5'供体剪接位点和3'末端结合域的间隔区可包含约10至100个核苷酸,优选约10至70个核苷酸,约20至70个核苷酸,约20至50个核苷酸,更优选约30至50个核苷酸。该间隔区可以包含下游内含子剪接增强子(DISE)。将3'受体剪接区与5'末端结合域分隔的间隔序列可包含约2至100个核苷酸或约10至100个核苷酸,优选约2至50个核苷酸,更优选约5至20个核苷酸。

[0199] 所述间隔区可以是非编码序列,但是可以设计成包括诸如终止密码子之类的特征,其将阻断剪接的pre-mRNA反式剪接分子的任何翻译。如上所述,本领域技术人员已知可以向pre-mRNA反式剪接分子添加额外的特征。进一步设想,本文所述的核酸序列或pre-mRNA反式剪接分子包含在(重组)载体中。这种载体也是技术人员已知的。包含本发明核苷酸序列(例如目的pre-mRNA反式剪接分子)的载体可以是质粒、病毒或本领域已知的其他载体,用于在哺乳动物细胞中复制和表达。

[0200] 目的核苷酸序列或本文所述的核苷酸序列(如pre-mRNA反式剪接分子)的表达可以通过本领域已知的任何启动子/增强子序列来调节,以在哺乳动物(优选人细胞)中起作用。这种启动子/增强子可以是诱导型的或组成型的。此类启动子也在本文其他地方描述。一种示例性启动子是人或鼠视紫红质或短波长或中波长或长波长敏感的视蛋白启动子等。任何类型的质粒、粘粒、YAC或病毒载体都可用于制备可直接导入组织部位的重组DNA构建体。或者,可以使用选择性感染所需靶细胞的病毒载体。用于实施本发明的载体包括任何真核表达载体,包括但不限于病毒表达载体,例如衍生自逆转录病毒、腺病毒或腺相关病毒的载体。在一个优选的实施方式中,本发明的重组载体是真核表达载体。

[0201] 在另一个具体实施方式中,本发明包括将核酸序列(如本发明的pre-mRNA反式剪接分子或编码本发明的pre-mRNA反式剪接分子的核酸序列)递送至靶细胞。各种递送系统

是已知的,并且可以用于将本发明的组合物转移到细胞中,例如,包封在脂质体、微粒、微囊、能够表达该组合物的重组细胞、受体介导的内吞作用、构建作为逆转录病毒、腺病毒、腺相关病毒或其他载体的一部分的核酸、注射DNA、电穿孔、磷酸钙介导的转染等。本发明还涉及包含核酸序列的细胞,所述核酸序列例如本发明的pre-mRNA反式剪接分子、编码pre-mRNA反式剪接分子的核酸序列或包含编码本发明的pre-mRNA反式剪接分子的核酸序列的重组载体。

[0202] 目的核酸序列或本文公开的核酸序列可用于向患有遗传性遗传病症的个体的细胞提供编码功能性生物活性分子的基因,其中缺失或突变基因产物的表达产生正常表型。这尤其可以通过本文公开的腺相关病毒载体来实现。

[0203] 本发明还涉及脱氧核糖核酸(DNA)分子,其包含启动子和编码本文所述的pre-mRNA反式剪接分子的序列。优选地,所述DNA分子是载体或质粒,其中所述载体可以是病毒载体,例如AAV病毒、腺病毒或慢病毒载体或质粒。所述pre-mRNA反式剪接分子或所述DNA分子(如病毒载体)可用于治疗,特别是基因治疗,更特别是用于治疗眼部病症的基因治疗。

[0204] 可以使用本发明的受体剪接位点的进一步应用是在任何AAV载体或AAV载体系统中。这种载体系统及其构建方式是本领域技术人员已知的,尤其记载在Carvalho et al. (2017) "Evaluating efficiencies of dual AAV approaches for retinal targeting" *Frontiers in Neuroscience*, vol.11, Article 503, US 2014/0256802或Trapani et al. (2013) "Effective delivery of large genes to the retina by dual AAV vectors" *EMBO Molecular Medicine*, vol.6, no.2, pp.194-211中。本文还描述了一些示例性的AAV载体和AAV载体系统。

[0205] 本发明还涉及包含至少两个反向末端重复序列的腺相关病毒(AAV)载体,所述反向末端重复序列包含这两个反向末端重复序列之间的核酸序列,其中所述核酸序列从5'至3'包含:

[0206] (i) 启动子;

[0207] (ii) 结合域;

[0208] (iii) 任意的间隔序列;

[0209] (iv) 受体剪接区序列,其包括

[0210] (a) 嘧啶区,其中所述嘧啶区包括

[0211] (aa) 5至25个核苷酸;

[0212] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0213] (b) 受体剪接位点,

[0214] ba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0215] bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G(优选A、C、T或G);

[0216] (v) 目的核苷酸序列

[0217] (vi) 任意的终止序列,如poly A序列。

[0218] 结构上,AAV是有二十面体衣壳的小的(25nm)、单链DNA无包膜病毒。在其衣壳(cap)蛋白的组成和结构上不同的天然存在的或工程化的AAV变体(也是AAV血清型)具有不

同的嗜性,即转导不同(视网膜的)细胞类型的能力(Boye et al.(2013)“A comprehensive review of retinal gene therapy”Molecular therapy 21 509-519)。当与泛活性启动子结合时,这种嗜性决定了基因表达的位点。然而与细胞类型特异性启动子结合时,位点特异性的程度(即转基因仅在视杆或视锥感光细胞中表达)由AAV血清型的嗜性和启动子的特异性二者共同来决定(Schön et al.(2015)“Retinal gene delivery by adeno-associated virus (AAV) vectors:Strategies and applications”Eur J Pharm Biopharm.95(Pt B): 343-52)。

[0219] 术语“腺相关病毒载体”或“重组AAV”或“rAAV”都可以互换使用,意在包括在其病毒基因组中包含异源性多核苷酸序列(本文也称为目的核苷酸序列)的任何AAV病毒。通常,目的异源性多核苷酸/核苷酸序列的侧翼是至少一个,并且通常是两个天然存在的或变异的AAV反向末端重复序列(ITR)。术语rAAV载体包括rAAV载体颗粒和rAAV载体质粒。因此,例如,包含异源多核苷酸序列的rAAV将是包含通常不包含在天然存在的野生型AAV中的核酸序列的rAAV,例如转基因(例如,非AAV RNA编码多核苷酸序列、非AAV蛋白编码多核苷酸序列)、非AAV启动子序列、非AAV多腺苷酸化序列等。

[0220] 这种重组AAV载体是本领域的公知常识,因此技术人员也知道如何构建这种重组AAV。

[0221] “rAAV载体基因组”或“rAAV基因组”是包含一个或多个异源性核酸序列的AAV基因组(即vDNA)。rAAV载体通常只需要末端重复序列(TR)就能产生病毒。所有其他病毒序列都被认为是可有可无的,可以反式形式提供(Muzyczka,(1992)Curr Topics Microbiol.Immunol.158:97)。典型地,rAAV载体基因组将仅保留一个或多个TR序列,以便最大化可被载体/衣壳有效包装的转基因/目的核苷酸序列/异源性核酸序列的大小。结构和非结构蛋白质编码序列可以反式(例如,来自载体,如质粒,或通过将序列稳定整合到包装细胞中)提供。在本发明的实施方案中,rAAV载体基因组包含至少一个TR序列(例如,AAV TR序列),任选地两个TR(例如,两个AAV TR),其通常位于载体基因组的5'和3'末端,并且在异源性核酸的侧翼,但是不需要与其邻接。TR可以相同,也可以不同。

[0222] 术语“反向末端重复”、“末端重复”或“TR”都可以互换使用,包括形成发夹结构并作为反向末端重复起作用的任何病毒末端重复或合成的序列(即,如果它介导期望的功能,例如复制、病毒包装、整合和/或原病毒拯救等)。所述TR可以是AAV TR或非AAV TR。例如,诸如其他细小病毒(例如犬细小病毒(CPV)、小鼠细小病毒(MVM)、人细小病毒B-19)或任何其他合适的病毒序列(例如,用作SV40复制起点的SV40发夹)的非AAV TR序列可以用作TR,其可以通过截短、取代、缺失、插入和/或添加而被进一步修饰。此外,TR可以是部分或完全合成的,例如美国专利号5,478,745中描述的“双D序列”。所述末端重复可以具有大约50、100、150、200、250、300个或更多核苷酸的长度。例如,末端重复约有145个核苷酸。例如,反向末端重复序列可以具有SEQ ID NO.14和/或15的序列。还设想,反向末端重复序列可以具有与SEQ ID 14和/或15的序列具有至少60%、70%、75%、80%、85%、90%、95%、98%或100%序列同一性的序列。优选地,AAV载体包括SEQ ID NO.14和15或与SEQ ID NO.14和15具有至少60%序列同一性的序列。

[0223] “AAV末端重复”或“AAV TR”可以来自任何AAV,包括但不限于血清型1、2、3、3B、4、5、6、7、8、9、10、11、12或13或现在已知或以后发现的任何其他AAV。只要末端重复介导所需

的功能(例如复制、病毒包装、整合和/或前病毒拯救等),AAV末端重复不需要具有天然末端重复序列(例如,天然AAV TR序列可以通过插入、缺失、截短和/或错义突变而被改变)。本发明的病毒载体还可以是“靶向”病毒载体(例如,具有特定嗜性)和/或如国际公开专利WO 00/28004中所描述的“杂交”细小病毒(即,其中病毒TR和病毒衣壳来自不同的细小病毒)。

[0224] 还设想,所述重组AAV或AAV载体包含或具有与天然存在的AAV1型(AAV-1)、AAV2型(AAV-2)、AAV3型(AAV-3)、AAV3B、AAV4型(AAV-4)、AAV5型(AAV-5)、AAV6型(AAV-6)、AAV7型(AAV-7)、AAV8型(AAV-8)、AAV9、AAV10、AAV11、AAV12、rh10、禽流感AAV、牛AAV、犬AAV、马AAV、灵长类AAV、非灵长类AAV和绵羊AAV具有至少70%、75%、80%、85%、90%、95%、97%、99%或100%序列同一性的序列。

[0225] 根据本发明,两个反向末端重复序列之间的核酸序列还包含本文所述的启动子。在AAV中使用的启动子的选择可以从大量组成型或诱导型启动子中进行,所述启动子可以在期望的靶细胞中表达所选择的目的转基因/核苷酸序列。靶细胞可以是感光细胞。启动子可以来自任何物种,包括人。

[0226] 本文所述的启动子可以是“细胞特异性”。术语“细胞特异性”是指为重组载体选择的特定启动子可以在特定细胞或眼部细胞类型中指导所选转基因/目的核苷酸序列的表达。有用的启动子包括但不限于视杆视蛋白启动子、红绿视蛋白启动子、蓝色视蛋白启动子、cGMP-P-磷酸二酯酶启动子、小鼠视蛋白启动子(Beltran et al. (2010) “rAAV2/5gene-targeting to rods:dose-dependent efficiency and complications associated with different promoters.” *Gene Ther.* 17 (9) :1162-74)、视紫红质启动子(Mussolino et al, *Gene Ther*, July 2011, 18 (7) :637-45);视锥转导素的 α 亚单位(Morrissey et al, *BMC Dev, Biol*, Jan 2011, 11:3); β 磷酸二酯酶(PDE)启动子;视网膜色素变性(RP1)启动子(Nicord et al, *J. Gene Med*, Dec 2007, 9 (12) :1015-23);NXNL2/NXNL 1启动子(Lambard et al, *PLoS One*, Oct. 2010, 5 (10) :e13025)、PE65启动子;视网膜变性慢/外周蛋白2(Rds/perph2)启动子(Cai et al, *Exp Eye Res.* 2010 Aug; 91 (2) :186-94);和VMD2启动子(Achi et al, *Human Gene Therapy*, (2009) 20:31-9)。

[0227] 用于本发明的有用启动子还包括但不限于视杆视蛋白启动子(RHO)、红绿视蛋白启动子、蓝色视蛋白启动子、cGMP-磷酸二酯酶启动子、SWS启动子(蓝色短波长敏感(SWS)视蛋白启动子)、小鼠视蛋白启动子(Beltran et al 2010上文引用)、视紫红质启动子(Mussolino et al, *Gene Ther*, July 2011, 18 (7) :637-45);视锥转导素的 α 亚基(Morrissey et al, *BMC Dev, Biol*, Jan 2011, 11:3);视锥抑制素(ARR3)启动子(Kahle NA et al., *Hum Gene Ther Clin Dev*, September 2018, 29 (3) :121-131), β 磷酸二酯酶(PDE)启动子;视网膜炎色素(RP1)启动子(Nicord et al, *J. Gene Med*, Dec 2007, 9 (12) :1015-23);NXNL2/NXNL 1启动子(Lambard et al, *PLoS One*, Oct. 2010, 5 (10) :e13025),RPE65启动子;视网膜变性缓慢/外周蛋白2(Rds/perph2)启动子(Cai et al, *Exp Eye Res.* 2010 Aug; 91 (2) :186-94);VMD2启动子(Achi et al, *Human Gene Therapy*, 2009 (20: 31-9)),以及ABCA4启动子或由至少两种不同启动子组成的任何杂交启动子。

[0228] 还考虑两个反向末端重复序列之间的核酸序列进一步包含终止信号,例如位于目的核酸序列3'端的多聚腺苷酸化信号。位于目的核酸序列3'端的终止信号(如多聚腺苷酸化信号)是任选的,因为pre-RNA可以反式剪接到内源性pre-RNA或另一重组pre-RNA上,以形

成成熟RNA (如成熟mRNA) 的5'端。然而, 已经表明, 在存在多聚腺苷酸化信号的情况下, 反式剪接的功效更高。

[0229] 转录pre-RNA需要启动子, 然而, 核酸序列不一定需要ATG (翻译起点) 和/或Kozak共有序列。如果pre-RNA可以反式剪接到内源性pre-RNA或另一种重组pre-RNA以形成成熟RNA (例如成熟mRNA) 的3'端, 则缺乏ATG和/或Kozak共有序列可能有利于避免反式剪接之前从pre-RNA的不希望的翻译。

[0230] 这些和其他常见载体和调控元件的选择是常规的, 并且许多这样的序列是可用的。参见, 例如Ausubel et al., Current Protocols in Molecular Biology, John Wiley & Sons, New York, 1989)。当然, 并不是所有的载体和表达控制序列都可以同样好地表达本文所述的所有转基因。然而, 本领域技术人员在不脱离本发明的范围的情况下可以在这些以及其他表达控制序列中进行选择。

[0231] 本发明还涉及腺相关病毒 (AAV) 载体系统, 其包括

[0232] (I) 第一AAV载体, 其包含具有至少两个反向末端重复的核酸序列, 所述反向末端重复包含这两个反向末端重复之间的核酸序列, 其中这两个反向末端重复之间的所述核酸序列从5'至3'包含:

[0233] (a) 启动子,

[0234] (b) 编码目的多肽的N末端部分的核苷酸序列;

[0235] (c) 供体剪接位点;

[0236] (d) 任意的间隔序列;

[0237] (e) 第一结合域; 和

[0238] (f) 任选地, 终止信号, 优选polyA信号:

[0239] (II) 第二AAV载体, 其包含含有至少两个反向末端重复的核酸序列, 所述反向末端重复包含这两个反向末端重复之间的核酸序列, 其中这两个反向末端重复之间的所述核酸序列从5'至3'包含:

[0240] (i) 启动子;

[0241] (ii) 第二结合域, 其与所述第一AAV载体的所述第一结合域互补;

[0242] (iii) 任意的间隔序列;

[0243] (iv) 受体剪接区序列, 其包括

[0244] (a) 嘧啶区, 其中所述嘧啶区包括

[0245] (aa) 5至25个核苷酸;

[0246] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基, 如胞嘧啶 (C)、胸腺嘧啶 (T) 和/或尿嘧啶 (U);

[0247] (b) 受体剪接位点,

[0248] (ba) 其中所述受体剪接位点位于嘧啶区的3'端; 和

[0249] (bb) 其中所述受体剪接位点包含5'至3'的NAGG序列, 其中N是A、C、T/U或G;

[0250] (v) 编码目的多肽的C末端部分的核苷酸序列;

[0251] (vb) 其中所述目的多肽的C末端部分和所述目的多肽的N末端部分重构目的多肽; 和

[0252] (vi) 终止信号, 优选polyA信号。

[0253] 因此,目的多肽的C末端部分对应于第一AAV载体中包含的多肽的N末端部分缺失的目的多肽部分。在一个实施方案中,多肽是全长多肽,并且所述第一AAV载体包含目的全长多肽的N末端部分,并且所述第二AAV载体包含目的全长多肽的C末端部分,并且目的全长多肽的C末端部分和目的全长多肽的N末端部分重构目的多肽。因此,目的全长多肽的C末端部分对应于第一AAV载体中包含的全长多肽的N末端部分缺失的目的全长多肽部分。更具体地说,在反式剪接之后,该mRNA包含编码目的多肽的N末端部分和目的多肽的C末端部分的序列(在框架中),其中该mRNA编码目的多肽(全长),并因此所述目的多肽被重构。

[0254] 值得注意的是,rAAV或AAV载体基因组包装容量存在约5kb的大小限制。由于许多治疗性蛋白质的cDNA很大,设计使用rAAV载体递送大型转基因的策略可以显著扩展rAAV介导的基因治疗的临床应用。因此,在一个实施方案中,目的多肽是由转基因编码的多肽。为了递送已经超过AAV包装容量的转基因,开发了反式剪接方法。简而言之,两个独立的rAAV载体将转基因的两部分递送到靶细胞中,一部分在转基因5'部分的3'端含有剪接供体信号,另一部分在转基因3'部分的5'端含有剪接受体信号。尽管在现有技术系统中,两个载体基因组之间的分子间重组产生插入的ITR连接或包含重组基因序列的插入序列,其在下一步中被细胞顺式剪接机制从单个pre-mRNA分子中切除以形成完整的全长转基因盒,但是在根据本发明所述的AAV载体系统中,两个载体基因组被分别转录,pre-mRNA通过它们的互补结合域相互作用,并且两个pre-mRNA被反式剪接以形成完整的(全长)转基因转录物。

[0255] 载体系统的效率可以通过测量目的(全长)蛋白质或编码目的蛋白质的mRNA的表达来评估,所述表达由双AAV载体方法提供,例如通过如本文所记载的或本领域技术人员已知的措施和技术。

[0256] 所述第二结合域与所述第一AAV载体的第一结合域互补。绝对互补虽然是优选的,但不是必需的。如本文所述,与第一结合域的一部分“互补”的序列是指具有足够互补性以能够与第一结合域杂交的序列,从而形成稳定的双链体。杂交的能力将取决于核酸的互补程度和长度(例如,见Sambrook et al,1989,Molecular Cloning,A Laboratory Manual,2d Ed.,Cold Spring Harbor Laboratory Press,Cold Spring Harbor,N.Y.)。因此,关于第二结合域使用的互补性意味着第二结合域包含与包含在第一AAV中的第一结合域至少60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%互补,优选与包含在第一AAV中的第一结合域至少90%、95%、98%、99%或100%互补的序列。

[0257] 第一结合域可以具有或包含如SEQ ID NO:17所示的序列,或者与SEQ ID NO:17的序列具有60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%序列同一性的序列。第二结合域可以具有或包含如SEQ ID NO:18所示的序列,或者与SEQ ID NO:18的序列具有60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%序列同一性的序列。本领域技术人员会理解,只要第二结合域与第一结合域互补,则第一和第二结合域可以交换。因此,第一结合域可以具有或包含如SEQ ID NO:18所示的序列,或与SEQ ID NO:17的序列具有60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%序列同一性的序列。第二结合域可以具有或包含如SEQ ID NO:17所示的序列,或与SEQ ID NO:18的序列具有60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%序列同一性的序列。

[0258] 任选地,第二结合域具有或包含SEQ ID NO:18、27、28、32和33中任一个的序列,或与SEQ ID NO:18、27、28、32和33中任一个的序列具有60%、65%、70%、75%、80%、85%、

90%、95%、98%、99%或100%序列同一性的序列,并且第一结合域具有或包含与第二结合域互补的序列,优选与第二结合域至少80%、85%、90%、95%、98%99%或100%互补,更优选与第二结合域至少90%、95%、98%99%或100%互补,甚至更优选与第二结合域98%99%或100%互补。本领域技术人员会理解,第一和第二结合域可以交换,只要所述结合域彼此互补。因此,第一结合域可以具有或包含SEQ ID NO:18、27、28、32和33中任一个的序列,或者与SEQ ID NO:18、27、28、32和33中任一个的序列具有60%、65%、70%、75%、80%、85%、90%、95%、98%、99%或100%序列同一性的序列,并且第二结合域具有或包含与第一结合域互补的序列,优选与第一结合域至少80%、85%、90%、95%、98%99%或100%互补,更优选与第一结合域至少90%、95%、98%99%或100%互补,甚至更优选与第一结合域98%99%或100%互补。

[0259] 进一步表明,SEQ ID NO:27或28的第50至100位核苷酸作为结合域是有效的。因此,在一个实施方案中,第二或第一结合域包含SEQ ID NO:27或28的第50至100位核苷酸的序列,或与SEQ ID NO:27或28的第50至100位核苷酸的序列具有80%、85%、90%、95%、98%、99%或100%序列同一性的序列。优选地,结合域具有至少80个核苷酸,其包含SEQ ID NO:27或28的第50至100位核苷酸的序列,或者与SEQ ID NO:27或28的第50至100位核苷酸的序列具有80%、85%、90%、95%、98%、99%或100%序列同一性的序列。本领域技术人员会理解,另一个结合域具有或包含分别与第一或第二结合域互补的序列,优选分别与第一或第二结合域至少80%、85%、90%、95%、98%99%或100%互补,更优选分别与第一或第二结合域至少90%、95%、98%99%或100%互补,甚至更优选分别与第一或第二结合域98%、99%或100%互补。

[0260] 在另一个实施方案中,第二或第一结合域包含SEQ ID NO:18、32或33的第1至50位核苷酸的序列,或与SEQ ID NO:18、32或33的第1至50位核苷酸的序列具有80%、85%、90%、95%、98%、99%或100%序列同一性的序列。优选地,结合域具有至少80个核苷酸,其包含SEQ ID NO:18、32或33的第1至50位核苷酸的序列,或者与SEQ ID NO:18、32或33的第1至50位核苷酸的序列具有80%、85%、90%、95%、98%、99%或100%序列同一性的序列。本领域技术人员会理解,另一个结合域具有或包含分别与第一或第二结合域互补的序列,优选分别与第一或第二结合域至少80%、85%、90%、95%、98%99%或100%互补,更优选分别与第一或第二结合域至少90%、95%、98%99%或100%互补,甚至更优选分别与第一或第二结合域98%、99%或100%互补。

[0261] 结合域可具有约15至250个核苷酸、约15至200个核苷酸、约100至200个核苷酸或少于500、400、300或200个核苷酸。在一个实施方案中,结合域包含50-150个核苷酸,优选80-120个核苷酸,甚至更优选90-110个核苷酸或约100个核苷酸。本文所述的结合域也可用于pre-mRNA反式剪接分子或根据本发明所述的方法。

[0262] 根据本发明使用的第一和第二结合域可以来源于人类序列或非人类序列,例如细菌序列。对于人类使用,例如在治疗(特别是基因治疗)中,优选非人序列以避免人类细胞中的脱靶效应。

[0263] 本文所述的第二AAV载体和任选的第一AAV载体可以包含终止信号,例如位于核酸序列3'的多聚腺苷酸化(polyA)信号。PolyA信号/序列是技术人员已知的,并且可以来自许多合适的物种,包括但不限于SV-40、人和牛。“PolyA”(A=腺苷酸)是指包含多个腺苷单磷

酸的核酸序列,例如包含AAUAAA共有序列的核酸序列,其能够使加工的转录物多聚腺苷酸化。在基因破坏或选择盒(GDSC)中,polyA序列位于报告基因和/或选择标记基因的下流,并向RNA聚合酶发出转录终止信号。可以预见,AAV载体可以包含SEQ ID 16的polyA序列或与SEQ ID No.16中记载的序列具有至少60%、65%、70%、75%、80%、85%、90%、95%、98%或100%序列同一性的序列。

[0264] 本发明的AAV载体或其他载体可以进一步任选地包括一个或多个转录终止序列、一个或多个翻译终止序列、一个或多个信号肽序列、一个或多个内部核糖体进入位点(IRES)和/或一个或多个增强子元件,或其任意组合。转录终止区通常可以从真核或病毒基因序列的3'非翻译区获得。转录终止序列可以位于编码序列的下流,以提供有效的终止。信号肽序列是氨基端肽序列,其编码负责将可操作连接的多肽定位到一个或多个翻译后细胞目的地的信息,所述目的地包括例如特定的细胞器区室,或者蛋白质合成和/或活性的位点,甚至细胞外环境。

[0265] 增强子——增加基因转录的顺式作用调节元件——也可以包含在公开的AAV载体或载体之一中。相关领域的普通技术人员已知多种增强子元件,包括但不限于CaMV 35S增强子元件、巨细胞病毒(CMV)早期启动子增强子元件、SV40增强子元件及其组合和/或衍生物。一个或多个指导或调节由目的结构基因编码的mRNA的多聚腺苷酸化的核酸序列也可以任选地包含在本发明的一个或多个载体中。

[0266] 可以使用基因治疗和/或病毒领域普通技术人员已知的任何一种或多种方法将所公开的双载体系统引入到一种或多种选定的哺乳动物细胞中。这些方法包括但不限于转染、显微注射、电穿孔、脂转染、细胞融合和磷酸钙沉淀,以及生物枪法。在一个实施方案中,本发明的载体可以在体内引入,包括例如通过脂转染(即,通过由一种或多种阳离子脂质制备的脂质体的DNA转染)。合成阳离子脂质(LIPOFECTIN,Invitrogen Corp.,La Jolla, Calif.,USA)可用于制备脂质体,该脂质体将包封所述载体以促进其引入一个或多个选定的细胞中。本发明的载体系统也可以使用本领域普通技术人员已知的方法作为“裸”DNA在体内被引入。

[0267] 本发明还涉及包含本发明的核酸序列和/或AAV载体和/或AAV载体系统的试剂盒。

[0268] 本发明还涉及产生目的核酸序列的方法,所述方法包括

[0269] (A) 将核酸序列与宿主细胞接触,其中所述核酸序列包括

[0270] (i) 一个或多个供体剪接位点序列;

[0271] (ii) 受体剪接区序列,其包括

[0272] (a) 嘧啶区,其中所述嘧啶区包括

[0273] (aa) 5至25个核苷酸;

[0274] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0275] (b) 受体剪接位点,

[0276] ba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0277] bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;

[0278] (iii) 目的核苷酸序列或其部分,其中所述目的核苷酸序列或其部分;

[0279] (a) 位于供体剪接位点的3'端和受体剪接区的5'端;和

[0280] (B) 在 (i) 的一个或多个供体剪接位点序列和 (ii) 的受体剪接区序列中切割核酸序列,从而将目的核苷酸序列或其部分与供体剪接位点和受体剪接区分离。

[0281] 如技术人员所知,剪接可以以不同的方式进行。技术人员已知的是例如顺式剪接和反式剪接。顺式剪接是将内含子序列从顺式转录物中切除的过程。宿主细胞中的这一过程发生在细胞核中。顺式剪接是本领域技术人员已知的,尤其记载在Alberts B, Johnson A, Lewis J, et al. (2002) “Molecular Biology of the Cell. 4th edition.” New York: Garland Science under the headline “from DNA to RNA”) 中。

[0282] 通常,疾病的疗法或治疗或预防包括给有此需要的哺乳动物受试者施用有效量的组合物,该组合物包含AAV载体、AAV载体系统、载体或核酸分子(如本文所述的pre-mRNA反式剪接分子),例如在调节序列的控制下携带编码目的转基因/核苷酸序列或其片段的核酸序列,所述调节序列在受试者的靶细胞(如眼细胞)中表达基因产物,并且任选地还携带药学上可接受的载体。

[0283] 因此,本发明还涉及包含本文所述的AAV载体、AAV载体系统、载体或核酸序列(例如pre-mRNA反式剪接分子)的药物组合物。这种药物组合物可以进一步包含载体,优选药学上可接受的载体。

[0284] 药物组合物可以是可注射溶液的形式。可注射溶液或混悬液可根据已知技术配制,使用合适的无毒、药学上可接受的稀释剂或溶剂(例如甘露醇、1,3-丁二醇、水、林格氏(Ringer's)溶液或等渗氯化钠溶液),或合适的分散剂或湿润剂和悬浮剂(例如无菌的、无刺激性、不挥发性油(包括合成的甘油一酯或甘油二酯,以及脂肪酸(包括油酸)))。

[0285] 对于可注射制剂,药物组合物可以是在合适的小瓶或试管中与合适的赋形剂混合的冻干粉。在临床使用之前,所述药物可通过将冻干粉溶解在合适的溶剂体系中以形成适于静脉内或肌内注射或用于视网膜下、玻璃体内或结膜下注射的组合物来重构。

[0286] 还设想本发明的药物组合物被配制/施用为滴眼剂。

[0287] 本发明的AAV载体、AAV载体系统、载体或核酸序列(如pre-mRNA反式剪接分子)或本发明的药物组合物可以治疗有效量施用。AAV载体、AAV载体系统、核酸序列(如pre-mRNA反式剪接分子或载体)的“治疗有效量”可随多种因素而变化,包括但不限于活性化合物在患者体内的稳定性、待缓解病况的严重程度、所治疗患者的总重量、施用途径、活性化合物被身体吸收、分布和排泄的容易程度、待治疗患者的年龄和敏感性、不良事件等,这对技术人员来说是显而易见的。所述施用量可以随着各种因素随时间的变化而调整。

[0288] 原则上,本发明的AAV载体、AAV载体系统、载体或核酸序列(如本发明的pre-mRNA反式剪接分子)或药物组合物可以任何合适的方式施用。本发明的AAV载体、AAV载体系统、载体、核酸序列或药物组合物例如包含用于靶向感光细胞的所需转基因(即目的核苷酸序列)可以配制成用于视网膜下或玻璃体内注射的药物组合物。可用于本文所述方法的其他施用形式包括但不限于直接递送至所需器官(例如,眼睛,例如滴眼剂)、口服、吸入、鼻内、气管内、静脉内、肌内、皮下、皮内和其他肠胃外施用途径。如果需要,可以组合施用途径。

[0289] 此外,可能需要进行非侵入性视网膜成像和功能研究,以确定特定眼细胞区域作为治疗目的。

[0290] 本发明的AAV载体、AAV载体系统、载体或核酸序列(如本文所述的pre-mRNA反式剪接分子)或药物组合物,如本文所述,可以在生理学上可接受的载体中向受试者施用。药物

组合中或施用 AAV 的浓度可以在每 μl 10^8 和 10^{12} 总载体基因组之间, 优选在每 μl 6×10^8 和 6×10^{10} 之间。AAV 载体也可以以每 μl 约 10^9 载体基因组的浓度施用。

[0291] 本发明的 AAV 载体、AAV 载体系统、载体或核酸序列 (如 pre-mRNA 反式剪接分子) 或药物组合可以单独施用或与其他治疗联合施用。因此, 药物组合还可另外地或替代地包含一种或多种其他活性成分。

[0292] 用于本发明的药物组合、AAV 载体、AAV 载体系统、载体、pre-mRNA 反式剪接分子或核酸序列可以对受试者施用。本文所述的 AAV 载体、AAV 载体系统、载体、pre-mRNA 反式剪接分子、核酸序列或药物组合可用于人类治疗和兽医应用, 优选用于治疗眼部病症, 特别是眼部病症的基因治疗。合适的眼部病症的例子不限于常染色体隐性严重早发性视网膜变性 (Leber 氏先天性黑蒙症)、先天性色盲、Stargardt 病、Best 病 (卵黄样黄斑变性)、Doyme 病、色素性视网膜炎 (特别是常染色体显性、常染色体隐性、X 连锁、二基因或多基因色素性视网膜炎)、(X 连锁) 视网膜劈裂、黄斑变性 (AMD)、年龄相关性黄斑变性、萎缩性年龄相关性黄斑变性、新生血管性 AMD、糖尿病性黄斑病变、增殖性糖尿病视网膜病变 (PDR)、囊样黄斑水肿、中心性浆液性视网膜病变、视网膜脱离、眼内炎症、青光眼、后葡萄膜炎、先天性静止性夜盲症、脉络膜视网膜病变、早发性视网膜营养不良、视锥、视杆-视锥或视锥-视杆营养不良、模式营养不良、Usher 综合征和其他综合征性睫状体炎, 如 Bardet-Biedl 综合征、Joubert 综合征、Senior-Løken 综合征或 Alström 综合征。

[0293] 受试者可以是哺乳动物或任何其他脊椎动物。合适的哺乳动物的实例包括但不限于小鼠、大鼠、牛、山羊、绵羊、猪、狗、猫、马、豚鼠、犬、仓鼠、水貂、海豹、鲸鱼、骆驼、黑猩猩、恒河猴和人类, 优选人类。其他脊椎动物的例子包括但不限于火鸡、鸡、鹅、鸭、短颈野鸭、野鸭、八哥、北方平尾鸟、海鸥、天鹅、珍珠鸡或水禽等。

[0294] 本发明还涉及用于治疗感光细胞疾病的本发明的 AAV 载体、AAV 载体系统、载体、核酸序列 (例如 pre-mRNA 反式剪接分子) 或药物组合。在这些实施方案中, AAV 载体 (例如 rAAV) AAV 载体、AAV 载体系统、载体、核酸序列 (例如 pre-mRNA 反式剪接分子) 或药物组合可包含目的核苷酸序列, 该核苷酸序列是编码治疗性多肽或其一部分、治疗性核酸或其一部分、治疗性蛋白质/多肽或其一部分或治疗性分子或其一部分的异源性核酸。

[0295] 与 mRNA 剪接 (其进化为去除非编码 RNA 序列) 类似, 可有可无的蛋白质序列可以在称为蛋白质剪接的过程中被切除。在这种情况下, 被称为外显肽 (外显子的类似物) 的蛋白质片段可以在去除所谓的内含肽 (内含子的类似物) 后剪接在一起。与需要复杂剪接机制的 mRNA 剪接不同, 蛋白质剪接是一种自催化化学过程。蛋白质剪接也可以发生在两种不同的蛋白质之间, 这一过程被称为蛋白质反式剪接。为此, 内含肽被分成两部分, 每部分都被标记到要融合的蛋白质上。这种方法导致两种蛋白质或多肽或其部分的无缝融合, 而不需要额外的因子。分裂内含肽技术可以进一步与 mRNA 反式剪接结合, 以进一步提高蛋白质重构效率。因此, 在一个实施方案中, 第一核酸序列或第一 AAV 载体进一步包含编码目的多肽的 N 末端部分 (或目的核酸序列的 5' 部分) 的核苷酸序列和供体剪接位点之间的内含肽的 N 末端部分的核酸序列, 并且第二核酸序列或第二 AAV 载体进一步包含编码目的多肽的 C 末端部分 (或目的核酸序列的 3' 部分) 的核苷酸序列和受体剪接区之间的内含肽的 C 末端部分的核酸序列。

[0296] 显然, 对于本发明的 ASS 所述的所有可能的实施方案都可以比照用于本文所述的

方法、核酸序列(例如pre-mRNA反式剪接分子)、试剂盒、AAV载体、AAV载体系统和用途中。

[0297] 本发明的特征还在于以下项:

[0298] 1. 一种核酸序列的用途,其包括

[0299] (i) 受体剪接区序列,其包括

[0300] (ia) 嘧啶区,其中所述嘧啶区包括

[0301] (aa) 5至25个核苷酸;

[0302] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0303] (ib) 受体剪接位点,

[0304] ba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0305] bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和

[0306] (ii) 目的核苷酸序列,其中所述目的核苷酸序列

[0307] (iia) 位于剪接区的3'或5';

[0308] 用于在所述受体剪接区切割,从而将目的核苷酸序列与受体剪接区序列分离。

[0309] 2. 一种核酸序列,其包括

[0310] (i) 受体剪接区序列,其包括

[0311] (ia) 嘧啶区,其中所述嘧啶区包括

[0312] (aa) 5至25个核苷酸;

[0313] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0314] (ib) 受体剪接位点,

[0315] ba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0316] bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G,

[0317] 任选地用于在所述受体剪接区切割,从而将目的核苷酸序列与受体剪接区序列分离。

[0318] 3. 第1项的核酸序列或第2项的核酸序列的用途,其中受体剪接区进一步包含分支点核苷酸,优选腺苷和/或内含子剪接增强子。

[0319] 4. 前述任一项的用途或核酸序列,其中受体剪接区进一步包含分支点核苷酸序列(c),其中分支点核苷酸序列

[0320] (ca) 包括1-15个核苷酸;

[0321] (cb) 包括所述分支点核苷酸,优选腺苷(A);和

[0322] (cc) 位于嘧啶区和受体剪接位点的5'端。

[0323] 5. 前述任一项所述的核酸序列的用途或核酸序列,其中嘧啶区、受体剪接区和任选的另外的分支点序列和/或内含子剪接增强子总共包括约200、150、100、50、45、40、35、30、25、20、15或更少的核苷酸,优选26个核苷酸。

[0324] 6. 或前述任一项所述的核酸序列的用途或核酸序列,其中

[0325] (a) 嘧啶区的5至25个核苷酸包含序列TTTTTT或TCTTTT;

[0326] (b) 嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基;和/或

- [0327] (c) 受体剪接位点具有CAGG序列。
- [0328] 7. 或前述任一项所述的核酸序列的用途或核酸序列,其中所述受体剪接区进一步包括
- [0329] (a) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;
- [0330] (b) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;
- [0331] (c) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'第一个核苷酸是C;
- [0332] (d) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAA;
- [0333] (e) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0334] (f) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAA;
- [0335] (g) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0336] (h) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAA;或
- [0337] (i) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0338] 8. 前述任一项所述的核酸序列的用途或核酸序列,其中所述受体剪接区具有SEQ ID NO:3或4的序列。
- [0339] 9. 前述任一项所述的核酸序列的用途或核酸序列,其中所述核酸序列进一步包含供体剪接位点。
- [0340] 10. 前述任一项所述的核酸序列的用途或核酸序列,其中所述核酸序列进一步包含启动子,优选其中所述核酸序列是DNA序列,并且进一步包含启动子。
- [0341] 11. 前述任一项所述的核酸序列的用途或核酸序列,其中所述核酸序列包含
- [0342] (i) 受体剪接区,其包括
- [0343] (iia) 嘧啶区,其中所述嘧啶区包括
- [0344] (iiaa) 5至25个核苷酸;
- [0345] (iiab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0346] (iib) 受体剪接位点,
- [0347] (iiba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0348] (iibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;
- [0349] (ii) 目的核苷酸序列,其中所述受体剪接区位于目的核苷酸序列的3'或5';
- [0350] (iii) 靶向pre-mRNA的结合域,其位于目的核酸序列的3'或5';和
- [0351] (iv) 任选的间隔序列,其中所述间隔序列位于结合域和受体剪接区之间。

- [0352] 12. 前述任一项所述的核酸序列的用途或核酸序列,其中所述核酸序列包含
- [0353] (i) 受体剪接区,其包括
- [0354] (ia) 嘧啶区,其中所述嘧啶区包括
- [0355] (iiiaa) 5至25个核苷酸;
- [0356] (iiiab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0357] (ib) 受体剪接位点,
- [0358] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0359] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
- [0360] (ii) 目的核苷酸序列,其中所述受体剪接区位于目的核苷酸序列的5'端;
- [0361] (iii) 供体剪接位点,其中所述供体剪接位点位于目的核苷酸序列的3'端;
- [0362] (iv) 第一结合域,其靶向位于目的核苷酸序列5'的pre-mRNA;
- [0363] (v) 第二结合域,其靶向位于目的核苷酸序列3'的pre-mRNA;
- [0364] (vi) 任选的第一间隔序列,其中所述第一间隔位于所述第一结合域和受体剪接区之间;和
- [0365] (vii) 任选的第二间隔序列,其中所述第二间隔位于所述第二结合域和供体剪接位点之间。
- [0366] 13. 前述任一项的核酸序列的用途或核酸序列,其中所述核酸序列包含腺相关病毒(AAV)载体,所述载体包含至少两个反向末端重复序列,所述反向末端重复序列包含在这两个反向末端重复序列之间的核酸序列,其中所述核酸序列从5'至3'包含
- [0367] (i) 启动子;
- [0368] (ii) 任选地结合域;
- [0369] (iii) 受体剪接区序列,其包括
- [0370] (a) 嘧啶区,其中所述嘧啶区包括
- [0371] (aa) 5至25个核苷酸;
- [0372] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0373] (b) 受体剪接位点,
- [0374] ba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0375] bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;
- [0376] (iv) 目的核苷酸序列
- [0377] (v) 任选地终止序列,优选polyA序列。
- [0378] 14. 前述任一项所述的核酸序列的用途或核酸序列,其中所述受体剪接区位于载体内、AAV载体内和/或pre-mRNA反式剪接分子内。
- [0379] 15. 前述任一项所述的核酸序列的用途或核酸序列,其中所述受体剪接区用于反式剪接。
- [0380] 16. 一种用于产生核酸序列的方法,该方法包括
- [0381] (A) 提供包含一个或多个供体剪接位点序列的第一核酸序列;
- [0382] (B) 提供第二核酸序列,其中所述第二核酸序列包括

- [0383] (i) 受体剪接区序列,其包括
- [0384] (ia) 嘧啶区,其中所述嘧啶区包括
- [0385] (iaa) 5至25个核苷酸;
- [0386] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0387] (ib) 受体剪接位点,
- [0388] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0389] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
- [0390] (ii) 目的核苷酸序列,其中所述目的核苷酸序列
- [0391] (iia) 位于受体剪接区的3';和
- [0392] (C) 获得核酸序列。
- [0393] 17. 根据项目16所述的方法,其中所述方法还包括以下步骤
- [0394] (D) 在一个或多个供体剪接位点序列切割第一核酸序列并在受体剪接位点切割第二核酸序列,
- [0395] (E) 将第一切割核酸序列与第二切割核酸序列连接,
- [0396] 从而获得核酸序列。
- [0397] 18. 根据第16或17项所述的方法,其中所述第一核酸序列进一步包括目的核苷酸序列或其一部分,其中所述目的核苷酸序列的至少一部分位于供体剪接位点的5'端,并且其中所述第二核酸序列进一步包括目的核苷酸序列或其一部分,其中所述目的核苷酸序列的至少一部分位于剪接受体剪接区的3'端。
- [0398] 19. 第16至18项中任一项所述的方法,其中将所述第一和第二核酸序列引入宿主细胞,优选其中所述第一和第二核酸序列是重组核酸序列。
- [0399] 20. 一种用于产生核酸序列的方法,该方法包括
- [0400] (A) 提供包含一个或多个供体剪接位点序列的第一核酸序列;
- [0401] (B) 提供第二核酸序列,其中所述第二核酸序列包括
- [0402] (i) 受体剪接区序列,其包括
- [0403] (ia) 嘧啶区,其中所述嘧啶区包括
- [0404] (iaa) 5至25个核苷酸;
- [0405] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0406] (ib) 受体剪接位点,
- [0407] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0408] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
- [0409] (C) 在一个或多个供体剪接位点序列切割第一核酸序列并在受体剪接位点切割第二核酸序列,
- [0410] (D) 将第一切割核酸序列与第二切割核酸序列连接,
- [0411] 从而获得核酸序列。
- [0412] 21. 根据第16-20项任一项所述的方法,其中所述第一核酸序列进一步包括目的核苷酸序列或其一部分,其中所述目的核苷酸序列的至少一部分位于供体剪接位点的5'端,

并且其中所述第二核酸序列进一步包括目的核苷酸序列或其一部分,其中所述目的核苷酸序列的至少一部分位于剪接受体剪接区的3'端。

[0413] 22.一种生产核酸的方法或第16至21项中任一项所述的方法,包括

[0414] (A) 将第一核酸序列引入宿主细胞,所述第一核酸序列包含pre-mRNA反式剪接分子序列或编码所述pre-mRNA反式剪接分子的核酸序列,其中第一pre-mRNA反式剪接分子从5'至3'包含

[0415] (a) 目的核苷酸序列的5'部分;

[0416] (b) 供体剪接位点;

[0417] (c) 任意的间隔序列;

[0418] (d) 第一结合域;和

[0419] (e) 任选地终止序列,优选polyA序列,和(B) 将第二核酸序列引入宿主细胞,所述第二核酸序列包含pre-mRNA反式剪接分子序列或编码所述pre-mRNA反式剪接分子的核酸序列,其中第二pre-mRNA反式剪接分子从5'至3'包含

[0420] (i) 第二结合域,其与第一核酸序列的第一靶结构域互补;

[0421] (ii) 受体剪接区序列,其包括

[0422] (iia) 嘧啶区,其中所述嘧啶区包括

[0423] (iiaa) 5至25个核苷酸;

[0424] (iiab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0425] (iib) 受体剪接位点,

[0426] (iiba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0427] (iibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和

[0428] (iii) 目的核苷酸序列的3'部分,和

[0429] 从而获得目的核酸序列,任选地进一步包括

[0430] (C) 在供体剪接位点序列切割第一核酸序列,并在受体剪接位点切割第二核酸序列;和

[0431] (D) 将包含目的核苷酸序列5'部分的第一切割核酸序列与包含目的核苷酸序列3'部分的第二切割核酸序列连接,从而获得目的核酸序列。

[0432] 23. 根据16至22项中任一项所述的方法,其中

[0433] (a) 嘧啶区的5至25个核苷酸包含序列TTTTTT或TCTTTT;

[0434] (b) 嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基;和/或

[0435] (c) 受体剪接位点具有CAGG序列。

[0436] 24. 根据16至23项中任一项所述的方法,其中所述受体剪接区进一步包括

[0437] (a) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;

[0438] (b) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;

[0439] (c) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'第一个

核苷酸是C;

[0440] (d) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAA;

[0441] (e) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAAC;

[0442] (f) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAA;

[0443] (g) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC;

[0444] (h) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAA;或

[0445] (i) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAAC;

[0446] 25. 一种核酸序列,其包括

[0447] (i) 受体剪接区序列,其包括

[0448] (ia) 嘧啶区,其中所述嘧啶区包括

[0449] (iaa) 5至25个核苷酸;

[0450] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0451] (ib) 受体剪接位点,

[0452] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0453] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和

[0454] (ii) 目的核苷酸序列,其中所述目的核苷酸序列

[0455] (iia) 位于受体剪接区的3'或5'。

[0456] 26. 一种核酸序列,其包括

[0457] (i) 受体剪接区序列,其包括

[0458] (ia) 嘧啶区,其中所述嘧啶区包括

[0459] (iaa) 5至25个核苷酸;

[0460] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0461] (ib) 受体剪接位点,

[0462] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0463] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和

[0464] (ii) 目的核苷酸序列,其中所述目的核苷酸序列

[0465] (iia) 位于受体剪接区的3'或5';

[0466] 并且任选地,其中目的核苷酸序列不是SEQ ID NO:10的视紫红质基因和/或不是SEQ ID NO:9的视紫红质基因的外显子3和/或不是SEQ ID NO:9的视紫红质基因的外显子3和/或其中目的核苷酸序列不包含SEQ ID No:9和/或10所示的序列。

[0467] 27. 25或26项所述的核酸序列,其中该核酸序列具有至多150个核苷酸的长度。

- [0468] 28. 25-27项中任一项所述的核酸序列,其中该核酸序列具有至多5500个核苷酸的长度。
- [0469] 29. 25-28项中任一项所述的核酸序列,其中(a)嘧啶区的5至25个核苷酸包含序列TTTTTT或TCTTTT和/或(b)嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基。
- [0470] 30. 受体剪接区序列,其包括
- [0471] (ia) 嘧啶区,其中所述嘧啶区包括
- [0472] (iaa) 5至25个核苷酸;
- [0473] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0474] (ib) 受体剪接位点,
- [0475] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0476] (ibb) 其中所述受体剪接位点包含5'至3'的CAGG序列。
- [0477] 31. 25至30项中任一项所述的核酸或受体剪接区序列,其中
- [0478] (a) 嘧啶区的5至25个核苷酸包含序列TTTTTT或TCTTTT;
- [0479] (b) 嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基;和/或
- [0480] (c) 受体剪接位点具有CAGG序列。
- [0481] 32. 25至31项中任一项所述的核酸或受体剪接区序列,其中所述受体剪接区进一步包括
- [0482] (a) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;
- [0483] (b) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;
- [0484] (c) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'第一个核苷酸是C;
- [0485] (d) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAA;
- [0486] (e) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0487] (f) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAA;
- [0488] (g) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0489] (h) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAA;或
- [0490] (i) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0491] 33. 一种pre-mRNA反式剪接分子,其包括

- [0492] (i) 受体剪接区,其包括
- [0493] (iia) 嘧啶区,其中所述嘧啶区包括
- [0494] (iiaa) 5至25个核苷酸;
- [0495] (iiab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0496] (iib) 受体剪接位点,
- [0497] (iiba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0498] (iibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;
- [0499] (ii) 目的核苷酸序列或其部分,其中所述受体剪接区位于目的核苷酸序列或其部分的3'或5';
- [0500] (iii) 靶向pre-mRNA的结合域,其定位于目的核酸序列或其部分的5'端;和
- [0501] (iv) 任选的间隔序列,其中所述间隔序列位于结合域和受体剪接区之间。
- [0502] 34. 第33项所述的pre-mRNA反式剪接分子,其还包含位于核酸分子和受体剪接区3'端的供体剪接位点。
- [0503] 35. 第33或34项所述的pre-mRNA反式剪接分子包括,
- [0504] (i) 受体剪接区,其包括
- [0505] (ia) 嘧啶区,其中所述嘧啶区包括
- [0506] (iiaa) 5至25个核苷酸;
- [0507] (iiab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0508] (ib) 受体剪接位点,
- [0509] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0510] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
- [0511] (ii) 目的核苷酸序列,其中所述受体剪接区位于目的核苷酸序列的5'端;
- [0512] (iii) 供体剪接位点,其中所述供体剪接位点位于目的核苷酸序列的3'端;
- [0513] (iv) 第一结合域,其靶向位于目的核苷酸序列5'端的pre-mRNA;
- [0514] (v) 第二结合域,其靶向位于目的核苷酸序列3'端的pre-mRNA;
- [0515] (vi) 任选的第一间隔序列,其中所述第一间隔位于所述第一结合域和受体剪接区之间;
- [0516] (vii) 任选的第二间隔序列,其中所述第二间隔定位于所述第二结合域和供体剪接位点之间。
- [0517] 36. 第33至35项中任一项所述的pre-mRNA反式剪接分子,其中所述受体剪接区位于目的核苷酸序列或其部分的5'端,并且所述结合域位于受体剪接区的5'端。
- [0518] 37. 第33至36项中任一项所述的pre-mRNA反式剪接分子还包含终止序列,优选polyA序列。
- [0519] 38. 第33至37项中任一项所述的pre-mRNA反式剪接分子,其中
- [0520] (a) 嘧啶区的5至25个核苷酸包含序列TTTTTT或TCTTTT;
- [0521] (b) 嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基;和/或

[0522] (c) 受体剪接位点具有CAGG序列。

[0523] 39. 第33至38项中任一项所述的pre-mRNA反式剪接分子,其中所述受体剪接区进一步包括

[0524] (a) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;

[0525] (b) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;

[0526] (c) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'第一个核苷酸是C;

[0527] (d) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAA;

[0528] (e) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAAC;

[0529] (f) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAA;

[0530] (g) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC;

[0531] (h) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAA;或

[0532] (i) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAAC;

[0533] 40. 一种DNA分子,其包含启动子和编码根据第33至39项中任一项所述的pre-mRNA反式剪接分子的序列。

[0534] 41. 一种腺相关病毒(AAV)载体系统,其包括

[0535] (I) 第一AAV载体,其包含至少两个反向末端重复,所述反向末端重复包含这两个反向末端重复之间的核酸序列,其中这两个反向末端重复之间的所述核酸序列从5'至3'包含:

[0536] (a) 启动子,

[0537] (b) 编码目的多肽的N末端部分的核苷酸序列;

[0538] (c) 供体剪接位点;

[0539] (d) 任意的间隔序列;

[0540] (e) 第一结合域;和

[0541] (f) 任意地终止序列,优选polyA序列。

[0542] (II) 第二AAV载体,其包含含有至少两个反向末端重复的核酸序列,所述反向末端重复包含这两个反向末端重复之间的核酸序列,其中这两个反向末端重复之间的所述核酸序列从5'至3'包含:

[0543] (i) 启动子;

[0544] (ii) 第二结合域,其与所述第一AAV载体的所述第一结合域互补;

[0545] (iii) 受体剪接区序列,其包括

- [0546] (a) 嘧啶区,其中所述嘧啶区包括
- [0547] (aa) 5至25个核苷酸;
- [0548] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0549] (b) 受体剪接位点,
- [0550] (ba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0551] (bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;
- [0552] (iv) 编码目的多肽的C末端部分的核苷酸序列;
- [0553] (ivb) 其中所述目的多肽的C末端部分和所述目的多肽的N末端部分重构所述目的多肽;和
- [0554] (v) 终止序列,优选polyA序列。
- [0555] 42. 第41项所述的AAV载体系统,其中所述多肽是全长多肽,并且所述第一AAV载体包含目的全长多肽的N-末端部分,并且所述第二AAV载体包含目的全长多肽的C-末端部分。
- [0556] 43. 第41或42项所述的AAV载体系统,其中目的多肽的C末端部分对应于目的多肽的部分,任选地对应于从包含在第一AAV载体中的多肽的N末端部分缺失的目的全长蛋白质。
- [0557] 44. 第41至43项中任一项所述的AAV载体系统,其中所述第一和第二AAV载体包含终止序列,优选polyA序列。
- [0558] 45. 一种腺相关病毒(AAV)载体,其包含至少两个反向末端重复,所述反向末端重复包含这两个反向末端重复之间的核酸序列,其中所述核酸序列从5'至3'包含:
- [0559] (i) 启动子;
- [0560] (ii) 任选的间隔序列;
- [0561] (iii) 结合域;
- [0562] (iv) 受体剪接区序列,其包括
- [0563] (a) 嘧啶区,其中所述嘧啶区包括
- [0564] (aa) 5至25个核苷酸;
- [0565] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0566] (b) 受体剪接位点,
- [0567] (ba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0568] (bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;
- [0569] (iv) 目的核苷酸序列
- [0570] (v) 任选地终止序列,优选polyA序列。
- [0571] 46. 41至45项中任一项所述的AAV载体或AAV载体系统,其中
- [0572] (a) 嘧啶区的5至25个核苷酸包含序列TTTTTT或TCTTTT;
- [0573] (b) 嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基;和/或
- [0574] (c) 受体剪接位点具有CAGG序列。
- [0575] 47. 41至46项中任一项所述的AAV载体或AAV载体系统,其中所述受体剪接区进一

步包括

[0576] (a) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;

[0577] (b) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;

[0578] (c) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'第一个核苷酸是C;

[0579] (d) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAA;

[0580] (e) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAAC;

[0581] (f) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAA;

[0582] (g) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC;

[0583] (h) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAA;或

[0584] (i) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAAC;

[0585] 48. 一种受体剪接区,其包括

[0586] (ia) 嘧啶区,其中所述嘧啶区包括

[0587] (iaa) 5至25个核苷酸;

[0588] (iab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);

[0589] (ib) 受体剪接位点,

[0590] (iba) 其中所述受体剪接位点位于嘧啶区的3'端;和

[0591] (ibb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G。

[0592] 49. 前述任一项所述的受体剪接区用于根据本发明所述的载体、AAV载体或pre-mRNA反式剪接分子中。

[0593] 50. 48至49项中任一项所述的受体剪接区序列,其中

[0594] (a) 嘧啶区的5至25个核苷酸包含序列TTTTTT或TCTTTT;

[0595] (b) 嘧啶区的最后一个嘧啶和受体剪接位点之间的序列具有少于10个,优选少于5个,更优选少于3个碱基;和/或

[0596] (c) 受体剪接位点具有CAGG序列。

[0597] 51. 48至50项中任一项所述的受体剪接区序列,其中所述受体剪接区进一步包括

[0598] (a) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'第一个核苷酸是C;

[0599] (b) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'第一个核苷酸是C;

- [0600] (c) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'第一个核苷酸是C;
- [0601] (d) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAA;
- [0602] (e) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少4个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0603] (f) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAA;
- [0604] (g) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少5个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0605] (h) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAA;或
- [0606] (i) 嘧啶区5'的7个核苷酸,其具有序列CAACGAG的至少6个核苷酸,其中5'最前面的核苷酸是CAAC;
- [0607] 52. 一种试剂盒,其包括根据本发明所述的受体剪接区、AAV载体、AAV载体系统、载体、DNA、核酸序列或pre-mRNA反式剪接分子。
- [0608] 53. 一种用于产生目的核酸序列的方法,所述方法包括
- [0609] (A) 向宿主细胞提供核酸序列,其中所述核酸序列包括
- [0610] (i) 一个或多个供体剪接位点序列;
- [0611] (ii) 受体剪接区序列,其包括
- [0612] (a) 嘧啶区,其中所述嘧啶区包括
- [0613] (aa) 5至25个核苷酸;
- [0614] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0615] (b) 受体剪接位点,
- [0616] ba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0617] bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;
- [0618] (iii) 目的核苷酸序列,其中所述目的核苷酸序列
- [0619] (a) 位于供体剪接位点的3'端和受体剪接区的5'端;和
- [0620] (B) 在(i)的一个或多个供体剪接位点序列处以及在NAGG的剪接受体序列(位点)的G和G之间切割核酸序列,从而从供体剪接位点和受体剪接区分离目的核苷酸序列。
- [0621] 54. 一种用于产生/切割目的核酸序列的方法,所述方法包括
- [0622] (a) 提供核酸序列,其包括
- [0623] (i) 受体剪接区序列,其包括
- [0624] (ia) 嘧啶区,其中所述嘧啶区包括
- [0625] (aa) 5至25个核苷酸;
- [0626] (ab) 其中在这5至25个核苷酸中至少60%的核苷酸是嘧啶碱基,如胞嘧啶(C)、胸腺嘧啶(T)和/或尿嘧啶(U);
- [0627] (ib) 受体剪接位点,

- [0628] ba) 其中所述受体剪接位点位于嘧啶区的3'端;和
- [0629] bb) 其中所述受体剪接位点包含5'至3'的NAGG序列,其中N是A、C、T/U或G;和
- [0630] (ii) 目的核苷酸序列,其中所述目的核苷酸序列
- [0631] (iia) 位于剪接区的3'或5';
- [0632] (b) 切割受体剪接区,从而将目的核苷酸序列与受体剪接区序列分离。
- [0633] 55. 药物组合物,其包含前述任一项所述的核酸分子、pre-mRNA反式剪接分子、AAV载体、AAV载体系统、载体。
- [0634] 56. 前述任一项所述的核酸分子、pre-mRNA反式剪接分子、AAV载体、AAV载体系统、载体、DNA分子或所述药物组合物,用于治疗,优选用于治疗眼部病症。
- [0635] 57. 一种针对目的核酸序列治疗受试者的方法,所述方法包括
- [0636] (a) 将(治疗有效量的)前述任一项的核酸分子、所述pre-mRNA反式剪接分子、AAV载体、AAV载体系统、载体、DNA分子或药物组合物施用于(有此需要的)受试者,其中所述受试者优选患有眼部病症。
- [0637] 58. 前述任一项所述的核酸分子、pre-mRNA反式剪接分子、AAV载体、AAV载体系统、载体、DNA分子或药物组合物用于制造药物。
- [0638] 59. 核酸分子、pre-mRNA反式剪接分子、AAV载体、AAV载体系统、载体、DNA分子、所述药物组合物的应用或前述任一项所述的方法,其中所述眼部病症选自自由如下组成的组:常染色体隐性严重早发性视网膜变性(Leber氏先天性黑蒙症)、先天性色盲、Stargardt病、Best病(卵黄样黄斑变性)、Doyne病、色素性视网膜炎(特别是常染色体显性、常染色体隐性、X连锁、二基因或多基因色素性视网膜炎)、(X连锁)视网膜劈裂、黄斑变性(AMD)、年龄相关性黄斑变性、萎缩性年龄相关性黄斑变性、新生血管性AMD、糖尿病性黄斑病变、增殖性糖尿病视网膜病变(PDR)、囊样黄斑水肿、中心性浆液性视网膜病变、视网膜脱离、眼内炎症、青光眼、后葡萄膜炎、先天性静止性夜盲症、脉络膜视网膜病变、早发性视网膜营养不良、视锥、视杆-视锥或视锥-视杆营养不良、模式营养不良、Usher综合征和其他综合征性睫状体炎,如Bardet-Biedl综合征、Joubert综合征、**Senior-Løken**综合征或**Alström**综合征。
- [0639] 必须指出,本文所使用的单数形式“一个(a)”、“一个(an)”和“所述(the)”包括复数引用,除非上下文另有明确说明。因此,例如,对于“一种试剂”包括一种或多种这样的不同试剂,以及对于“所述方法”包括本领域普通技术人员已知的等效步骤和方法,这些步骤和方法可以被修改或替代本文所述的方法。
- [0640] 此外,本文所用术语“约”指的是可测量的值,例如多核苷酸或多肽序列的量或长度、剂量、时间、温度等的量,意在包括特定量的±20%、±10%、±5%、±1%、±0.5%或±0.1%的变化。本文所用术语“和/或”指的是并包括相关列出项的一个或多个的任何和所有可能的组合,以及当以替代方式(“或”)解释时组合的缺乏。
- [0641] 除非上下文另有说明,否则明确的意图是可以以任何组合使用本文所述的本发明的各种特征。
- [0642] 此外,本发明还预期在本发明的一些实施方案中,可以排除或省略本文所阐述的任何特征或特征的组合。
- [0643] 本公开中引用的所有出版物和专利通过引用整体并入。被引入以供参考之数据范围与本说明书抵触或不符时本说明书将取代任何此类资料。

[0644] 除非另有说明,否则在一系列元素之前的术语“至少”应理解为意指该系列中的每个元素。本领域技术人员将认识到,或者仅使用常规实验就能够确定本文描述的本发明的具体实施方案的许多等同方案。这样的等同方案旨在被本发明所包含。

[0645] 在整个说明书和随后的权利要求书中,除非上下文另有要求,否则词语“包括”以及诸如“包含”和“含有”的变体将被理解为暗示包括所述整体或步骤或整体或步骤的组,但不排除任何其他整体或步骤或整体或步骤的组。当在本文使用时,术语“包括”可以用术语“包含”代替,或者有时当在本文使用时,用术语“具有”代替。然而,本文使用的术语“包括”或等同物包括如下定义的“由...组成”或“基本由...组成”,因此可以用术语“由...组成”或“基本由...组成”代替。

[0646] 本文中使用的“由...组成”排除权利要求要素中未指明的任何要素、步骤或成分。如本文所用,“基本上由...组成”不排除实质上不影响权利要求的基本和新颖特征的材料或步骤。

[0647] 用于本发明的序列:

[0648] 具有超过10个核苷酸的以下序列在本文中被提及:

[0649] 表1:

[0650]

Seq	名称	序列
1	嘧啶区	<i>TCTTTTGTCATCT</i>
2	内含子剪接 增强子	TGGGGGGAGG
3	vgASS_620 (图 2 中的 RHOE3d) (ASS 区)	<i>CAACGAGTCTTTTGTCATCTACAGGT</i>
4	vgASS_620 mod (ASS 区)	<i>CAACGAGTTTTTTGTCATCTACAGGT</i>
5	ASS 区(图 2 中的 RHOE3b)	<i>CTACACGCTCAAGCCGGAGGTCAACAACGAG</i> <i>TCTTTTGTCATCTACAGGT</i>
6	ASS 区(图 2)	<i>GCCGGAGGTCAACAACGAGTCTTTTGTCATCTACA</i>

[0651]

	中的 RHOE3c)	GGT
7	ASS 区(图 2 中的 RHOE3e)	GTCTTTTGTCATCTACAGGT
8	ASS 区(图 2 中的 RHOE3f)	GTCTTTTGTCATCTACAGGTGTTTCGTGGTTCGTGGT CCA
9	视紫红质基 因的突变外 显子 3 (c.620T>G 用下划线标 出)	GTACATCCCCGAGGGCCTGCAGTGCTCGTGTGGAA TCGACTACTACACGCTCAAGCCGGAGGTCAACAA CGAGTCTTTTGTCATCTACAG <u>G</u> TTCGTGGTCCACT TCACCATCCCCATGATTATCATCTTTTTCTGCTATGG GCAGCTCGTCTTCACCGTCAAGGAG
10	突变的视紫 红质基因 GenBank: BC112104.1 智人视紫红 质, mRNA (cDNA 克隆	CCAGCTGGAGCCCTGAGTGGCTGAGCTCAGGCCT TCGCAGCATTCTTGGGTGGGAGCAGCCACGGGTC AGCCACAAGGGCCACAGCCATGAATGGCACAGAA GGCCCTAACTTCTACGTGCCCTTCTCCAATGCGAC GGGTGTGGTACGCAGCCCCTTCGAGTACCCACAGT ACTACCTGGCTGAGCCATGGCAGTTCTCCATGCTG GCCGCCTACATGTTTCTGCTGATCGTGCTGGGCTT

[0652]

MGC:13830 9 IMAGE:832 7572), 完整 的 cds (终止密码 子和起始密 码子以及 c.620T>G 突 变用下划线 标出)	CCCCATCAACTTCCTCACGCTCTACGTCACCGTCC AGCACAAGAAGCTGCGCACGCCTCTCAACTACAT CCTGCTCAACCTAGCCGTGGCTGACCTCTTCATGG TCCTAGGTGGCTTCACCAGCACCTCTACACCTCT CTGCATGGATACTTCGTCTTCGGGCCACAGGATG CAATTTGGAGGGCTTCTTTGCCACCCTGGGCGGTG AAATTGCCCTGTGGTCCTTGGTGGTCCTGGCCATC GAGCGGTACGTGGTGGTGTGTAAGCCCATGAGCA ACTTCCGCTTCGGGGAGAACCATGCCATCATGGGC GTTGCCTTCACCTGGGTCATGGCGCTGGCCTGCGC CGCACCCCCACTCGCCGGCTGGTCCAGGTACATCC CCGAGGGCCTGCAGTGCTCGTGTGGAATCGACTA CTACACGCTCAAGCCGGAGGTCAACAACGAGTCT TTTGTCATCTACAGGTTTCGTGGTCCACTTCACCATC CCCATGATTATCATCTTTTTCTGCTATGGGCAGCTC GTCTTCACCGTCAAGGAGGCCGCTGCCAGCAGC AGGAGTCAGCCACCACACAGAAGGCAGAGAAGG AGGTCACCCGCATGGTCATCATCATGGTCATCGCTT TCCTGATCTGCTGGGTGCCCTACGCCAGCGTGGCA TTCTACATCTTCACCCACCAGGGCTCCA ACTTCGG TCCCATCTTCATGACCATCCCAGCGTTCTTTGCCAA
---	---

[0653]

		GAGCGCCGCCATCTACAACCCTGTCATCTATATCAT GATGAACAAGCAGTTCCGGAAGTGCATGCTCACC ACCATCTGCTGCGGCAAGAACCCTGGGTGACG ATGAGGCCTCTGCTACCGTGTCCAAGACGGAGAC GAGCCAGGTGGCCCCGGCCTAAGACCTGCCTAGG ACTCTGTGGCCGACTATAGGCGTCTCCCATCCCCT ACACCTTCCCCCAGCCACAGCCATCCCACCAG
11	嘧啶区	<i>TTTTTTGTCATTT</i>
12	嘧啶区	<i>TCTTTTGTCATCTA</i>
13	嘧啶区+7nt	<i>CAACGAGTCTTTTGTCATCTA</i>
14	反向重复(5)	CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGG CAAAGCCCAGGCGTCGGGCGACCTTTGGTCGCCC GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGG AGTGGCCAACCTCCATCACTAGGGGTTCT
15	反向重复(3)	AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCT CTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGAC CAAAGGTCGCCCAGCGCCCGGGCTTTGCCCGGGC GGCCTCAGTGAGCGAGCGAGCGCGCAG
16	Poly A 序列	GCAGTGAAAAAATGCTTTATTTGTGAAATTTGTG ATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAA

		TAAACAAGTT
17	第一个结合域序列	AGTGCATCAAGGCGATCACATCAGTGAAAAAAG CCAGACAGGCGGTTAAACCAACGCAGATTAAACA GCAGGATGCAAAAATTCGCAGGTGGTCAGATG
18	第二结合域序列 (BD_k)	CATCTGACCACCTGCGAATTTTTGCATCCTGCTGTT TAATCTGCGTTGGTTTAACCGCCTGTCTGGCTTTTT TTC ACTGATGTGATCGCCTTGATGCACT
19	ABCA4 的双 AAV 5' 编码序列 (CMV 启动子)	如图 5 所示的序列
20	ABCA4 的双 AAV 3' 编码序列 (CMV 启动子)	如图 6 所示的序列
21	包含 ABCA4 的 5' 编码序列	如图 7 所示的序列

[0654]

[0655]

	的 双 AAV(ABCA 4 启动子)	
22	包 含 ABCA4 的 3' 编码序列 的 双 AAV(ABCA 4 启动子)	如图 8 所示的序列
23	图 2 中的 WT ASS 序 列	ctcggcagccaccttggtgttcccaagtccctcacaggcagggtctccctacctg cctgtcctcaggtacatccccgagggcctgcagtgtcgtgtggaatcgactacta cacgctcaagccggagggtcaacaacgagtcttttgcctctacatgttcgtggcca cttcaccatecccatgattatcatcttttctgctatgggcagctcgtcttcaccgtcaa ggaggtacggg
24	图 2 中的 c.620T>G ASS 序列(突 变用下划线 标出)	ctcggcagccaccttggtgttcccaagtccctcacaggcagggtctccctacctg cctgtcctcaggtacatccccgagggcctgcagtgtcgtgtggaatcgactacta cacgctcaagccggagggtcaacaacgagtcttttgcctctacag <u>gttc</u> gtggcca acttcaccatecccatgattatcatcttttctgctatgggcagctcgtcttcaccgtca aggaggtacggg
25	ASS 区(图 2 中 的	ctacacgctcaagccggagggtcaacaacgagtcttttgcctctacatgt

[0656]

	RHOE3a)	
26	ASS 区(图 2 中的 RHOE3f)	gtcttttgcacatctacaggtgttcgtg
27	BD_g	GTTCCAGGGAGGGAATGTGAAGCCCCAGAAAGGG CCAGCGCCAGGTGGAATTCGCTAGCTCGGCAGCC ACCTTGGCTGTTCCCAAGTCCCTCACAGGCAGGG
28	BD_h+i	GTAAACCCTCCAGTCAGGACTCAAACCCAGTAGTG TCTGGTTCCAGGCACTGACCTGCTAGCTCGGCAGC CACCTTGGCTGTTCCCAAGTCCCTCACAGGCAGG G
29	从 <i>RHO</i> 内 舍子 1 获得 的 BD 序列 (PTM1 , B2)(Berger et al., 2015)	CACCATTCATGGTGATAGCCGGGCTGCTGTTTGTG CAGGGCTGGCACTGAACACTGCCTTGATCTTATTT GGAGCAATATGCGCTTGTCTAATTCACAGCAAGA AAACTGAGCTGAGGCTCAAAGAAGTCAAGCGCCC TGCTGGGGCG
30	ITR2 正向引 物	GGAACCCCTAGTGATGGAGTT
31	ITR2 反向引 物引物	CGGCCTCAGTGAGCGA

32	BD_j	CATCTGACCACCAGCGAATTTTTGCATCCAGCTGT TTAATCAGCGTTGGTTTAACCGCCAGTCAGGCTTT CTTTCAAAGATGTGATCGCCTTGCAGCACT
33	BD_l	CCATCCCGCATCTGACCACCAGCGAAATGGATTTT TGCATCCAGCTGGGTAACGTTGGCAATTTAACCGC CAGTCAGGCTTTCTTTCAAAGATGTGGATT
34	BD_m	CCATCCCGCATCTGACCACCTGCGAAATGGATTTT TGCATCCTGCTGGGTAACGTTGGCAATTTAACCGC CTGTCTGGCTTTCTTTCAATGATGTGGATT
35	5' 天蓝蛋白 + BD_g	序列如图 16 所示
36	3' 天蓝蛋白 + BD_g	序列如图 17 所示
37	5' SpCas9-VP R + BD_k	序列如图 18 所示
38	3' SpCas9-VP R + BD_k	序列如图 19 所示
39	双 AAV 载体 -5' ABCA4_	序列如图 20 所示

[0657]

	w 内含子	
[0658]	40 双 AAV 载体 -3'ABCA4_ w 内含子	序列如图 21 所示

[0659] 表1:假定的分支点序列用下划线标出,分支点加两次下划线并用粗体字母表示,嘧啶区序列用斜体字母表示,受体剪接位点用粗体字母表示。SEQ ID NO:9和10中加下划线的是来自620T>G的突变以及起始和终止密码子。

[0660] 实施例

[0661] 以下实施例说明本发明。这些实施例不应被解释为限制本发明的范围。实施例被包括以用于说明的目的,并且本发明仅由权利要求限制。

[0662] 材料和方法

[0663] 细胞培养和转染

[0664] 在37°C和10%CO₂的CO₂培养箱(Heraeus,Thermo Fisher Scientific)中,将人胚胎肾293(HEK293)细胞(DMSZ)维持在DMEM培养基+GlutaMAX+1g/1葡萄糖+丙酮酸盐+10%FBS(Biochrom)+1%青霉素/链霉素(Biochrom)中。在相同条件下在DMEM培养基+GlutaMAX+4.5g/1葡萄糖+10%FBS+1%青霉素/链霉素中培养源自HEK293的Lenti-X 293T(HEK293T)细胞(Clontech,Takara)。两种细胞系每周传代两次,融合率约为90%。

[0665] 使用磷酸钙方法进行瞬时转染。为此,将细胞接种到6cm的细胞培养板上。当转染含SpCas9质粒用于蛋白质印迹时,使用10cm细胞培养板。将细胞孵育过夜,直到它们达到期望的约70%的融合。转染混合物组分按指定顺序加入到15ml Falcon试管中。涡旋过程中滴加2x BBS。

[0666] 转染混合物在室温下孵育3-4分钟,并滴加到培养基中。对于初始实验(实施例2,图4),细胞已经在5%的CO₂环境中孵育24小时,并且在不更换培养基的情况下收获细胞。对于优化的方案(实施例3、5-8;图9和11-14),细胞在5%CO₂环境中孵育3-4小时,更换培养基,细胞在10%CO₂下保持约48小时。当转染含荧光团的质粒时,通过EVOS®FL细胞成像系统(Life Technologies,Thermo Fisher Scientific)评估成功的转染和表达。

[0667] 661W细胞由Muayyad Al-Ubaidi教授(休斯顿大学)惠赠。该细胞系是从鼠视网膜肿瘤克隆的,并被发现表现出视锥感光细胞的分子特征(al-Ubaidi et al.,1992,Tan et al.,2004)。将661W细胞保存在37°C和10%CO₂的CO₂培养箱(Heraeus,Thermo Fisher Scientific)中的DMEM培养基+GlutaMAX+1g/1葡萄糖+丙酮酸盐+10%FBS(Biochrom)+1%抗生素-抗真菌剂中。所述细胞以大约90%的融合率每周传代两次。使用如上所述的磷酸钙方法进行瞬时转染。

[0668] 如所述产生小鼠胚胎成纤维细胞(MeFs)(Jat et al.,1986,Xu,2005)。将细胞保存在37°C和5%CO₂的CO₂培养箱(Heraeus,Thermo Fisher Scientific)中的DMEM培养基+GlutaMAX+1g/1葡萄糖+丙酮酸盐+10%FBS(Biochrom)+1%青霉素/链霉素(Biochrom)中。所述细胞以大约90%的融合率每周传代一次。使用TurboFect™转染试剂(Thermo Fisher

Scientific) 瞬时转染MEF。将细胞接种到6cm的细胞培养板上,并孵育至70-90%融合。按以下顺序制备反应混合物。

[0669] 加入每种组分后,通过涡旋剧烈混合溶液。转染混合物在室温下孵育15-20分钟,然后滴加到培养板上。3小时后更换培养基,转染后48小时收获细胞。当转染含荧光团的质粒时,通过EVOS®FL细胞成像系统(Life Technologies,Thermo Fisher Scientific)评估成功的转染和表达。

[0670] 重组腺相关病毒的生产

[0671] 重组腺相关病毒(rAAV)是通过三重磷酸钙转染含有目的基因的pAAV2.1质粒、pAD辅助质粒和编码所需衣壳的质粒产生的。对于鼠视网膜的视网膜下注射,选择2/8Y733F衣壳变体是因为其在转导感光细胞和RPE方面的高效率(Petrs-Silva et al.,2009, Mol. Ther.,17,463-71)。将HEK293T细胞接种到15×15cm的细胞培养板上,过夜培养,直到达到60-80%的融合。转染前,用无血清培养基代替含FBS的培养基。将转染试剂按所示顺序加入到50ml Falcon试管中:pAAV2.1质粒270μg、pAD辅助质粒Xμg、衣壳质粒Yμg、H₂O 11.85ml、聚芳烃(8mg/ml) 15μl、葡聚糖(10mg/ml) 1.5ml、CaCl₂(2.5M) 1.5ml、2x BBS 15ml。pAD Helper和衣壳质粒所需的量使用以下公式计算: $X\mu\text{g} = 270\mu\text{g} \times \text{pAD辅助质粒MM pAAV2.1MM}$ $Y\mu\text{g} = 270\mu\text{g} \times \text{衣壳质粒MM pAAV2.1MM}$ 。在涡旋过程中滴加CaCl₂和2x BBS。将2ml转染混合物滴加到15个培养板的每一个中。轻轻摇动平板,随后将其置于5%CO₂环境中24小时。随后更换培养基,并将板置于10%CO₂环境中孵育另外48小时。

[0672] 含病毒的培养基收获两次。第一次收获发生在转染后72小时,通过收集所有平板的全部培养基并加入新鲜培养基。第二次收获发生在另一个72小时的孵育期之后。将培养基收集到500ml离心管中。通过在4,000rpm和4℃下离心15分钟(JA-10转子,J2-MC高速离心机,Beckman Coulter)并通过0.45μm PES过滤单元(Nalgene,Thermo Fisher Scientific)过滤上清液,从培养基中除去残余细胞。将40%聚乙二醇(PEG)溶液添加到流通液中至最终浓度为8%,并在4℃保持过夜以沉淀病毒颗粒。随后,溶液在4,000rpm和4℃下离心15分钟(JA-10转子,J2-MC高速离心机,Beckman Coulter)。弃去上清液,并将含病毒的沉淀储存在-20℃直至进一步处理。

[0673] 对于碘克沙醇密度梯度离心,将沉淀重新悬浮在7.5ml无菌过滤的PBS中,并与最终浓度为50U/ml的Benzonase®(VWR)在37℃的水浴(Haake)中孵育30分钟,以除去残留的未包装的DNA。接下来,将病毒悬液移至Quick-Seal聚丙烯管(39ml,Beckman Coulter)中,并通过在病毒悬液之下以如下顺序加入不同碘克沙醇浓度的溶液来建立密度梯度:7ml 15%碘克沙醇溶液、6ml 25%碘克沙醇溶液、5ml 40%碘克沙醇溶液和6ml 60%碘克沙醇溶液。为此,使用了MINIPULS 3蠕动泵(Gilson)和长玻璃移液管。随后,用Beckman Tube Topper密封试管,并在Optima L-80K超速离心机(70Ti转子,Beckman Coulter)中以70,000rpm和18℃离心1小时45分钟。然后,用套管在管子顶部打孔,以确保空气流动。使用20G套管和20ml注射器,通过在40%和60%相之间的边界处横向刺穿管,从梯度中收集富含病毒颗粒的40%碘克沙醇相。含病毒的溶液储存在-80℃直到进一步处理。

[0674] 为了进一步纯化病毒,使用ÄKTaprime plus色谱系统(GE Healthcare)、5ml HiTrap™ Q FF阴离子交换色谱柱(GE Healthcare)和PrimeView 5.31软件(GE

Healthcare)进行了阴离子交换色谱分析。开始前,用缓冲液A(20mM Tris,15mM NaCl,pH 8.5)平衡柱子,用该缓冲液以1:1的比例稀释含病毒溶液。溶液通过环形注射器(50ml Superloop,GE Healthcare)装载到色谱柱上。监测收集部分的紫外光吸收和电导特性,提供关于所含病毒量的信息。使用2.5M氯化钠溶液从柱中除去剩余的结合分子。将所有含病毒的部分汇集在一起,用于后续处理。3.5.4rAAV浓度的增加

[0675] 为了提高病毒浓度,使用了分子量截止值为100kDa的Amicon®Ultra-4离心过滤单元(Merck)。将含病毒的溶液装载到过滤装置的顶部,并以20分钟的间隔在4,000rpm(JA-10转子、J2-MC高速离心机、Beckman Coulter)和4°C下离心,直到体积减少到500μl。随后,过滤单元用1ml 0.014%吐温/PBS-MK(10xPBS 50ml,1M MgCl 500μl,2.5M KCl 500μl,加入500ml水)洗涤。溶液在相同条件下进一步离心,直到体积减少到100μl浓缩病毒溶液。制备10μl等分试样,并在-80°C下储存直至使用。

[0676] 为了确定产生的rAAV的滴度,使用StepOnePlus实时PCR系统(应用生物系统,Thermo Fisher Scientific)进行qPCR。制作了一条标准曲线作为参考。为此,使用以下引物通过PCR扩增含有部分ITR的片段。

[0677] ITR2正向:5'GGAACCCCTAGTGATGGAGTT 3'(SEQ ID NO:30)

[0678] ITR2反向:5'CGGCCTCAGTGAGCGA 3'(SEQ ID NO:31)

[0679] 接下来,纯化扩增子,并使用Nanodrop™ 2000c分光光度计(Thermo Fisher Science)测定浓度。在制备浓度为c的标准溶液后,制备 10^{10} 至 10^1 拷贝的稀释系列。为了获得标准曲线,用标准稀释系列的三个技术复制品进行qPCR。为此,使用了MicroAmp™ Fast Optical 96孔反应板(Applied Biosystems,Thermo Fisher Scientific)和PowerUp™ SYBRTM Green Master Mix(Thermo Fisher Scientific)。病毒溶液在H₂O稀释100倍,并在同一个反应板上进行三次技术复制。如下制备反应混合物。使用StepOnePlus实时PCR系统软件(Applied Biosystems,Thermo Fisher Scientific)分析获得的数据。如有必要,手动调整基线设置和循环阈值位置。通过将所得循环阈值(Ct)值与稀释度的对数作图来获得标准曲线。从标准曲线可以推断出每μl产生的rAAVs的病毒基因组数量(vg/μl)。

[0680] 视网膜下注射

[0681] 对于视网膜下注射,出生后第21天(P21)的C57B16/J小鼠通过腹膜内注射氯胺酮(40mg/kg体重)和赛拉嗪(20mg/kg体重)进行麻醉。在完全没有爪子缩回反射后,通过施用阿托品(1%)和含托吡卡胺(0.5%)的滴眼液(Mydriaticum Stulln,Pharma Stulln GmbH)来扩张瞳孔。使用手术显微镜(OPMI 1FR pro,蔡司)聚焦眼底。使用NANOFIL 10μl注射器(World Precision Instruments)和34G斜面针(World Precision Instruments)进行单次注射,将1μl含1010rAAV颗粒注射到视网膜下。注射的眼睛用含有5mg/g庆大霉素和0.3mg/g地塞米松的眼膏治疗。将小鼠放在37°C的加热板上(Leica HI1120,Leica Biosystems),直到完全从麻醉中恢复。注射后2-4周,收集所有注射的视网膜并进行RT-PCR分析或免疫组织化学处理。

[0682] 免疫组织化学

[0683] 对于免疫组织化学,视网膜下注射的小鼠通过颈椎脱位实施安乐死。取出眼睛并置于0.1M磷酸盐缓冲液(PB)中。随后,用21G套管在锯齿缘刺穿眼球,并在4%多聚甲醛(PFA,Sigma Aldrich,pH调节至7.4)中固定5分钟。接下来,将眼睛置于体视显微镜(Stemi

2000, Zeiss) 下方, 置于浸有0.1M PB的滤纸上。通过使用手术剪刀(SuperFine Vannas, World Precision Instruments) 沿着锯齿缘切割, 角膜、晶状体和玻璃体被移除。含有视网膜的眼球的剩余部分在室温下固定在4%的PFA中45分钟, 随后在0.1M PB中洗涤3次5分钟。为了冷冻保存, 将眼球置于30%蔗糖溶液(w/v) 中, 在4℃过夜。

[0684] 第二天, 将眼球包埋在组织冷冻培养基(Sakura) 中, 在干冰上冷冻, 直到培养基凝固。使用低温恒温器(Leica CM3050 S, Leica Biosystems) 将视网膜切片, 切成10μm厚的切片, 收集在涂层玻璃载玻片(Superfrost Plus显微载玻片, Thermo Fisher Scientific) 上并储存在-20℃。

[0685] 对于免疫组织化学染色, 将视网膜切片在室温下解冻, 并使用Super PAP Pen Liquid Blocker(Science Services) 包围。随后, 切片用0.1M PB再水合5分钟, 并用4%PFA固定10分钟。在使用0.1M磷酸盐缓冲液pH 7.4(PB) 将切片洗涤3次, 每次5分钟后, 应用0.1M PB中含有一抗、5%ChemiBLOCKER(Merck) 和0.3%Triton X-100的溶液。冷冻切片与一抗溶液在4℃孵育过夜。第二天, 用0.1M PB洗涤视网膜三次, 每次5分钟, 并在室温下与0.1M PB中含有二抗和2%ChemiBLOCKER的溶液一起孵育1.5h。随后用0.1M PB洗涤三次, 每次5分钟后, 用5μg/ml Hoechst 33342溶液(Invitrogen) 对细胞核进行染色。最后, 切片用0.1M PB洗涤, 包埋在Fluoromount-G Mounting Medium(Thermo Fisher Scientific) 中, 用盖玻片覆盖, 并储存在4℃。

[0686] 共聚焦显微镜

[0687] 染色视网膜的图像是使用徕卡TCS SP8倒置共焦激光扫描显微镜(Leica Microsystems) 获得的, 该显微镜配备有405nm二极管以及分别适用于激发Hoechst 33342、Cy3和Cy5的552nm和633nm光泵半导体激光器。根据相应染料的发射光谱选择滤光器设置。使用LAS X软件(Leica Microsystems) 通过HC PL APO 40x/1.30油CS2物镜(Leica Microsystems) 和F型浸没液体(Leica Microsystems) 以z堆栈(1个微米步长) 采集图像。使用相同的软件, 通过应用最大强度投影, z堆栈被压缩成2D图像。这些图像用ImageJ 1.48v软件(美国国立卫生研究院) 进一步处理。

[0688] 使用徕卡TCS SP8光谱共聚焦激光扫描显微镜(Leica Microsystems) 获得瞬时转染活细胞的图像, 该显微镜配备有分别适用于激发天蓝、柠檬黄蛋白和mCherry的448nm、514nm和552nm光泵半导体激光器。根据相应荧光团的发射光谱选择滤光器设置。图像是用HCX APO 20x/1.00W物镜(Leica Microsystems) 获得的。所有图像都是用ImageJ 1.48v软件处理的。

[0689] RNA提取

[0690] 为了从注射的视网膜中提取RNA, 通过颈椎脱位对小鼠实施安乐死。通过将钝镊子放在眼睛下方, 使用无菌手术刀切开眼球(Swann-Morton) 并逐渐向上移动镊子来收集视网膜。每个构建体汇集三个视网膜, 根据制造商的说明, 使用RNeasy迷你试剂盒(Qiagen) 提取RNA。为了进行破碎, 通过添加350μl的RLT缓冲液(Qiagen, 随试剂盒提供) +3.5μlβ-巯基乙醇(β-ME, Sigma Aldrich) 并使其通过安装在无菌注射器上的20G针至少5次, 对组织进行匀浆。其余步骤按照方案执行。在30μl无RNA酶的H₂O中洗脱RNA。

[0691] 对于从瞬时转染的细胞中提取RNA, 使用了RNeasy Mini Kit Plus(Qiagen)。为此, 将培养基从6cm的培养板上取出, 用16cm的细胞刮刀(Sarstedt) 刮去细胞。以500μl培养

基将细胞收集到2ml安全锁管 (Eppendorf) 中,并在 $3,000 \times g$ 和 4°C 下离心10分钟。弃去上清液,将沉淀重新悬浮在 $600\mu\text{l}$ RLT Plus缓冲液 (Qiagen,随试剂盒提供)+ $6\mu\text{l}$ β -ME (Sigma Aldrich) 中。在每个试管中放入钢球,使用混合研磨机MM400 (Retsch) 在30Hz下破碎细胞1分钟。然后,取出球,将悬浮液在 $21,000 \times g$ 和室温下离心5分钟。剩余的步骤根据方案进行,包括通过gDNA消除器旋转柱除去基因组DNA的任选步骤。在 $30\mu\text{l}$ 无RNA酶的H₂O中洗脱RNA。

[0692] 使用Nanodrop™ 2000c分光光度计 (Thermo Fisher Scientific) 测量RNA浓度。RNA保存在冰上,直到进一步使用,或者短期储存在 -20°C 或长期储存在 -80°C 。

[0693] cDNA合成

[0694] 对于cDNA合成,根据制造商的说明使用了Revert Aid First Strand cDNA Synthesis Kit (Thermo Fisher Scientific)。每个实验使用等量的RNA。将cDNA反应混合物在Mastercycler®连续梯度中孵育。将cDNA保存在冰上,直到进一步使用,或短期储存在 -20°C 或长期储存在 -80°C 。

[0695] 反转录PCR

[0696] 反转录PCR (RT-PCR) 使用Herculase II融合DNA聚合酶 (Agilent Technologies) 或VWR Taq DNA聚合酶 (VWR) 进行。

[0697] 分支点分析

[0698] 对于巢式套索RT-PCR,如上所述提取RNA。按照制造商的说明,将 $10\mu\text{g}$ RNA与RNA酶R (Lucigen) 一起孵育,以去除任何非圆形RNA。随后的cDNA合成 (Revert Aid First Strand cDNA Synthesis Kit,Life Technologies) 如上所述进行。对于该反应,仅使用随机六聚体引物。接下来,套索通过巢式RT-PCR,使用Herculase II融合DNA聚合酶 (Agilent Technologies) 进行扩增。对于第一轮扩增,按照3.12所述制备反应混合物。对于第二轮扩增,向反应混合物中加入 $5\mu\text{l}$ 的第一次PCR,而不是cDNA。此外,使用了结合第一对引物下游25-30bp的第二对引物。表11列出了应用的循环条件。

[0699] 为了进行套索的TOPO克隆和分析,巢式RT-PCR的产物被亚克隆到质粒中。为此,通过在 72°C 下将1个单位的Taq聚合酶 (VWR) 与PCR反应一起孵育10分钟,将3'-腺嘌呤突出端添加到扩增后的DNA片段中。随后,根据制造商的说明,将套索亚克隆到TOPO载体 (TOPO TA克隆试剂盒,Thermo Fisher Scientific) 中。将质粒转化到细菌中,并进行小规模质粒制备。对得到的质粒进行测序 (Eurofins Genomics),并使用DNAMAN软件 (Lynnon Biosoft) 通过与研究的内含子比对来分析获得的套索序列。

[0700] 蛋白质提取

[0701] 为了提取蛋白质,将培养基从6或10cm的培养板中取出。使用16cm的细胞刮除器 (Sarstedt) 刮除细胞,并以 $500\mu\text{l}$ 培养基收集到2ml安全锁管 (Eppendorf) 中。将细胞悬浮液在 $3,000 \times g$ 和 4°C 下离心10分钟,弃去上清液并将沉淀重新悬浮在 $150\mu\text{l}$ 和 $250\mu\text{l}$ Triton X-100 (TX) 裂解缓冲液中 (2.5ml Triton X-100, 15ml 5mM NaCl ml , 0.4ml 2.5M CaCl_2 ,加入 500ml 水;cocomplete™ ULTRA protease Inhibitor Cocktail片剂 (Roche) 在使用前加入 (1片/ 10ml))。在每个安全锁管中加入钢球,并使用混合研磨机MM400 (Retsch) 在30Hz下破碎细胞1分钟。然后,装有球的管在 4°C 下颠倒旋转 (VWR™管旋转器) 20分钟。随后,取出球并将裂解物在 4°C 下以 $5000 \times g$ 离心10分钟。将含蛋白质的上清液转移到新的反应管中,并储存在 -20°C 。

[0702] 总蛋白浓度用Bradford测定法测定。将5 μ l蛋白质裂解物与95 μ l 0.15M NaCl溶液混合,并转移到PMMA标准一次性比色皿(BRAND)中。随后,加入1ml考马斯蓝溶液,通过移液彻底混合,并在室温下孵育2分钟。使用BioPhotometer (Eppendorf),针对含有5 μ l TX裂解缓冲液的空白对照测量溶液的吸收。获得的值代表5 μ l裂解物中包含的蛋白质总量。

[0703] 实施例1:优化的ASS模块的识别

[0704] 在HEK293细胞和转导的鼠感光细胞中,使用人视紫红质(RHO)微小基因分析了疾病相关视紫红质突变对mRNA剪接的影响。其中,我们在RHO基因的外显子3中发现了一个突变(c.620T>G),其产生了一个新的ASS(图1)。ASS和预测的ASS元件的顺序如图2A和B所示。

[0705] 计算机预测显示,由c.620T>G突变形成的受体剪接位点(以下称为ASS_620)与外显子3的天然视紫红质受体剪接位点具有相似的剪接分数(参见图3A)。这表明剪接机器可以交替使用两个剪接位点。然而,然而,实验数据表明ASS_620专门用于HEK293细胞(图1D)和鼠光感受器(图1E)。这表明ASS_620是一个强受体剪接位点。

[0706] 对于生物技术应用中的潜在用途,ASS的强度和功能应该独立于基因环境。为了测试ASS_620在另一个非天然环境中是否有功能,并确定最有效剪接所需的ASS_620元件,将c.620T>G突变两侧的可变长度序列引入RPS27基因的外显子3(图2C)。单个序列导致在ASS_620位置可变的剪接效率(图2D)。当使用包含RPS27微小基因的RHO-E3d时,获得了最高的剪接效率(接近100%),即,该微小基因包含26bp序列(CAACGAGTCTTTTGTTCATCTACAGGT;SEQ ID NO:3),所述序列由5'端的7bp序列、多嘧啶区和典型的ASS组成。令人惊讶的是,该序列不包含RHO外显子3中任何预测的分支点。5'端的7bp序列(CAACGAG)可能含有目前尚未确定的有效分支点序列,或者,任选地,有效利用ASS所需的内含子剪接增强子识别位点。所述26bp的序列(SEQ ID NO:3)在本文被称为“vgASS_620”。

[0707] 为了进一步确认vgASS_620的剪接效率,将其引入另外四个随机选择的基因(HBQ1、S100A12、CLRN1和CNGB1)中,这些基因具有可变的预测受体剪接位点强度(图3C)。与vgASS_620相比,天然外显子受体剪接位点的预测强度在RPS27的情况下更高,在S100A12、CLRN1和HBQ1的情况下相似,在CNGB1的情况下更低(图3A)。在随后的RT-PCR实验中,显示vgASS_620不仅专门用于RPS27,而且用于本文测试的所有其他基因(图3D-E)。这些结果表明,vgASS_620是一个不依赖基因环境的非常强的受体剪接位点。因此,假设vgASS_620可用于改善SMaRT或双腺相关病毒(AAV)载体杂交技术中目前相当低的反式剪接效率。

[0708] 实施例2:体外应用的天蓝蛋白报告分析

[0709] 为了验证这一假设,通过将天蓝蛋白的编码序列分成被人工内含子打断的两个人工外显子,产生了剪接报告子分析(图4D)。该内含子包含vgASS_620和一个强供体剪接位点,通过人工145bp相互分离。转染的HEK293细胞的共聚焦成像显示,当两个剪接位点都以顺式方式提供时,有强烈而稳健的天蓝蛋白荧光(图4F(c1))。RT-PCR实验和相应条带的测序证实,在这种构型中,两个天蓝蛋白外显子被有效地拼接在一起(数据未显示)。如图4G所示,使用针对天蓝蛋白N端部分的特异性抗体的蛋白质印迹实验检测到预期大小(27kDa)的特异性免疫信号。综上所述,这些发现表明报告基因检测在顺式构型中具有功能性且非常有效。因此,在随后的实验中,当两个天蓝蛋白剪接片段以反式(即在单独的质粒上)提供时,该构建体用作确定天蓝蛋白重组效率的参考。本实验中使用的DSS具有AAGGTAAG序列。

[0710] 接下来,使用报道基因测试分析结合域和受体剪接位点强度对天蓝蛋白重构效率

的影响,其中报道基因的重构需要mRNA反式剪接。

[0711] 为此,开发了一种基于荧光报告子的测试方法,其中,青色荧光蛋白“天蓝蛋白”的编码序列在第154位再次被分成两半(图4D)。如图4D(c3)所示,包含在CMV启动子控制下的5'部分的第一DNA构建体进一步配备了强DSS(AAGGTAAG),随后是结合域,而包含在CMV启动子控制下的3'部分的第二DNA构建体配备了vgASS_620作为ASS,随后是与第一DNA构建体的结合域互补的结合域。转录后,BD的碱基配对使剪接元件非常接近,这有助于反式剪接,并产生全长的成熟天蓝蛋白mRNA。天蓝蛋白的重构可以用光学方法检测,例如,通过显微术(如图4F)或流式细胞术,在mRNA水平上(例如,使用RT-PCR)或在蛋白质水平上(例如,使用如图4G所示的蛋白质印迹)。

[0712] 选择人类视紫红质基因的内含子2作为结合域优化的示例性模板有两个原因:(1)优化类型的RHO内含子2结合域可用于基于SMaRT的c.620T>G剪接突变的替换,因为该突变位于外显子3(图1)。RHO内含子2序列与小鼠基因组中的任何序列都不同源(数据未显示),因此预计不会引起脱靶效应。因此,优化的RHO内含子2结合域也可以用于使用双AAV载体方法在小鼠视网膜中重构其他大型人类基因。

[0713] 将HEK 293细胞与如上所述的第一和第二DNA构建体瞬时共转染,并且在图4D(c3)中显示了BD_h+i的示例,并且通过共聚焦活细胞成像来评估天蓝蛋白荧光的存在。包含两个顺式天蓝蛋白部分和一个包含相同剪接元件的中间人工内含子的构建体被用作顺式剪接参考对照(cis-ctrl)。当将两半分开转染时,没有检测到荧光。当共转染这两种构建体时,可以观察到荧光细胞,表明成功的转剪接和天蓝蛋白编码序列的重构。

[0714] RHO内含子2结合域通过改变其大小和位置进行了优化(图4A)。此外,人工BD是通过融合来自内含子5'和3'末端的序列而产生和测试的,如同BD_h+i一样。

[0715] 在共转染HEK293细胞后,使用引物p1和p2针对5'构建体,引物p3和p4针对3'构建体,将两种构建体的mRNA表达与看家基因ALAS进行比较分析,证明在pre-mRNA水平上仅存在微小的表达差异(图4B和C)。重构效率是通过相对于顺式ctrl的天蓝蛋白条带的辐射分析来确定的,并且使用不同的BD差异很大(图4E)。在定量前,将所有蛋白质条带标准化为 β 微管蛋白。两个结合域BD_g(SEQ ID NO:27)和BD_h+i(SEQ ID NO:28),长度均约为100bp,导致天蓝蛋白重构高效率达到>30%。重要的是,这种高效率远远优于先前研究中已知的基于基因组水平上分裂基因的重构的替代方法(杂交、重叠和基因组“反式”剪接方法)(Carvalho et al.,2017,Frontiers in Neurosciences,11,Article 503)。使用报告基因lacZ在体外测试了所有三种策略。在这种情况下,报告的最高重构效率为17.7%。

[0716] 最近的一项研究假设反式剪接的效率不受剪接位点强度的影响(Lorain et al.,2013)。为了检验该假设,将vgASS_620的天蓝蛋白重构效率与RHO外显子3的天然受体剪接位点结合最有效的结合域BD_h+i(SEQ ID NO:28)进行比较。从图4E可以看出,与vgASS_620相比,源自天然ASS的重构效率明显较低($35.7 \pm 4.6\%$ 对 $0.7 \pm 0.1\%$)。这一发现提供了明确的证据,证明双载体方法中的反式剪接效率强烈依赖于受体剪接位点的强度(图4D-G)。

[0717] 在随后的实验中,重构效率可增加(例如,实施例3,图9D)。首先,在最初的实验中,独立样本的数量(n)仍然很低。通过增加独立样品的数量(n)和优化转染方案,可以更可靠地量化重构效率并且可以达到更高的值,例如60%,或者随着共转染效率的提高,甚至可以获得更高的重构效率(图9)。

[0718] 除了HEK293细胞重构效率外,还在661W细胞和MEF细胞中测试了反式剪接。使用结合域BD_g,已经观察到在661W细胞中>40%和在MEF细胞中>50%的重构效率(数据未显示)。当比较两种细胞系时,没有检测到反式剪接效率的显著差异。这表明mRNA反式剪接方法是细胞类型独立的。

[0719] 总之,体外结果表明,不仅可以通过改变结合域的序列和长度,而且可以通过优化受体剪接位点的强度来显著提高反式剪接效率。这些有希望的发现为进一步优化基于反式剪接的技术开辟了新的途径。

[0720] 结合vgASS₆₂₀,图4所示的结合域“g”可用于建立外显子3中或其下游视紫红质突变的基于SMaRT的基因治疗方法。此外,vgASS₆₂₀还可以与多种其他结合域结合,通过SMaRT或通过双AAV载体方法治疗其他遗传性视网膜疾病(IRD)基因。本文提供的公开内容适用于除其天然存在的情况之外的包含vgASS₆₂₀剪接模块的任何DNA或RNA序列的使用(即除了携带c.620T>G突变的患者的视紫红质基因)。SEQ ID NO:3的vgASS₆₂₀或与SEQ ID NO:3具有至少70%序列同一性的核苷酸序列可用于以下应用:

[0721] 1) 微小基因的设计(例如,用于分析mRNA剪接上的突变)

[0722] 2) 在生物或治疗环境中靶向内源性mRNA(例如使用SMaRT技术)

[0723] 3) 反式重构两个mRNA片段。这在克服腺相关病毒的有限基因组包装能力的背景(AAV)载体(约5.0kb,优选<4.7kb)下特别重要。图5至8绘出了可用于本文所述的双AAV载体系统的AAV载体的序列。

[0724] 4) 设计基因表达或靶向盒,以选择性地增加含有弱ASS剪接位点的可变剪接基因中某些剪接产物的存在。

[0725] 实施例3:受体剪接位点相对于结合域(BD)对重构效率的影响。

[0726] BD长度和序列代表重构效率的关键决定因素。BD最有可能影响mRNA的紧密结合和潜在的折叠,但预计它们不会直接促进后续剪接过程的效率或精确度。如上所述,DSS具有良好的特性,并且对其强度的预测与其实验性能非常匹配。因此,于并无明显的需要在分裂荧光团重构测定的框架中优化该剪接位点。相比之下,由于其复杂性,ASS的强度无法可靠预测。结果表明,vgASS₆₂₀是一个特别强的受体剪接位点。考虑到剪接位点强度会影响分裂荧光团分析的重构效率,与其他受体剪接位点相比,vgASS₆₂₀应该会导致更高的值。

[0727] 为了分析这一点,比较了存在vgASS₆₂₀或其他两种ASS(即RHO外显子3(S3,图9)的天然ASS和杂交ASS(S2),所述杂交ASS是通过用天然RHO外显子3ASS的PPT代替vgASS₆₂₀的多聚嘧啶区(PPT)而产生的)的重构效率。此外,每个ASS与三个不同的结合域结合:强结合域(BD_g,B1;SEQ ID NO:27)和源自该研究的弱结合域(BD_f,B3),以及从RHO内含子1(SEQ ID NO::29)获得的已发表的BD序列(PTM1,B2)。该序列显示出通过剪接体介导的mRNA反式剪接对突变的RHO转录物进行高效修复(Berger et al.,“Repair of rhodopsin mRNA by spliceosome-mediated RNA trans-splicing:A new approach for autosomal dominant retinitis pigmentosa”,(2015)Mol.Ther.23(5):918-930)。通过共聚焦活细胞成像、RT-PCR和蛋白质印迹分析所有组合(图9)。这个实验产生了几项重要的发现。首先,它揭示了BD和ASS都是决定重组效率的关键成分。第二,本研究中确定的最有效的BD(BD_g,B1)优于已发表的RHO结合域(B2)。第三,ASS强度似乎比BD更能影响重组效率,因为最弱的BD与最强的ASS(B3+S1)结合仍能产生可检测的反式剪接,而最强的BD与最弱的ASS(B1+S3)

结合不能导致编码序列的可检测的重构。

[0728] vgASS_620由ASS、PPT和上游额外的7bp序列组成,缺失该序列会削弱ASS_620的识别能力。因此,据推测,这个7bp的序列可能包含一个非常有效的分支点,潜在地解释了vgASS_620不受基因环境影响的普遍而有效的识别。然而,在这个序列中没有预测到有效的分支点。相反,一些其他序列被预测用作位于c.620T>G ASS上游的40bp处的分支点。然而,当同时突变包含在PPT上游7bp序列中的所有预测分支点核苷酸和所有潜在分支点腺嘌呤时,c.620T>G突变体的剪接不能改变(数据未显示)。这一发现表明分支点要么位于别处,要么c.620T>G ASS在分支点选择上具有很高的灵活性。为了更直接地鉴定用于c.620T>G ASS剪接的分支点,使用瞬时表达RHO c.620T>G微小基因的HEK293细胞进行巢式套索RT-PCR。转染RHO WT微小基因的HEK293细胞作为参照。当进行套索RT-PCR时,分别为野生型和突变型微小基因获得了一个条带,两者的大小不同。两条条带看起来都有些分散,表明相应的套索具有不同的大小。为了鉴定这些扩散带中包含的单个序列,将套索RT-PCR产物亚克隆到TOP0载体中,并通过测序对获得的克隆进行单独分析。在调查RHO WT套索时,确定了两个主要分支点(在42%和33%的情况下使用)和三个次要分支点(各占8%的情况)(表2)。这两个主要的分支点非常类似于一致序列,并产生了很高的预测分数。所有RHO WT分支点都在内含子-外显子连接的上游57-184bp处被发现。由于90%以上的人类分支点位于ASS序列上游50bp内,RHO外显子3的mRNA剪接似乎非常规(Corvelo et al.,2010,PLoS Comput.Biol.,6(11):e1001016)。

[0729] 表2:检测到的分支点和相应的预测分数

<i>RHO</i> WT 微小基因			
分支点	相对位置	HSF 分数	使用频次
cactgAt	-184 bp	84.3	41.6 %
tccacAa	-177 bp	76.8	8.3 %
tccccAt	-158 bp	85.6	33.3 %
[0730] ttatgAt	-113 bp	60.6	8.3 %
cagccAc	-57 bp	58.4	8.3 %
<i>RHO</i> c.620T>G 微小基因			
atcgAt	-245 bp	83.6	16.7 %
attccAt	-236 bp	55.6	8.3 %
tgcagAc	-213 bp	73.9	8.3 %
cgtttAt	-208 bp	55.9	8.3 %
atcccCt	-200 bp	未预测	16.7 %
cagggAg	-189 bp	未预测	8.3 %
[0731] gaagcCc	-174 bp	未预测	8.3 %
ttcccAa	-132 bp	79.5	8.3 %
caggcAg	-118 bp	未预测	8.3 %
tcctAc	-107 bp	77.6	8.3 %

[0732] 然而,当与野生型微小基因相比时,获得的*RHO* c.620T>G的分支点分布图有显著差异。首先,已经为c.620T>G突变体鉴定了各种分支点。然而,没有检测到主要的分支点,并且没有一个与获得的WT微小基因相同。第二,分支点甚至位于所使用的ASS的更上游,即107-245bp对应于天然外显子3ASS上游的21-159bp。第三,几乎一半的检测到的分支点不太像共有序列,因此无法使用人类剪接发现器(HSF)剪接预测工具进行预测。总之,这些数据

表明,vgASS_620的强度并不是源于一个包含的、非常有效的分支点的存在,而是可能部分地由分支点选择的高度灵活性引起的。这可以解释其异常高效的性能,并使其成为需要高效剪接的生物技术的非常有吸引力的工具。

[0733] 实施例4:基于mRNA反式剪接的rAAV双载体的体内研究

[0734] 在前几节中评估的基于mRNA反式剪接的分析最有力的应用是在双rAAV载体的框架中重构大基因。因此,使用rAAV在小鼠视网膜中测试了mRNA反式剪接方法。为此,使用了稍微修改的分裂荧光团测定法。为了控制单个病毒转导的细胞中rAAV载体的表达,两个双rAAV载体盒都配有荧光团序列,即5'载体编码序列的5'末端的柠檬黄蛋白和3'载体编码序列3'末端的mCherry。一种在体外重构效率最高的BD,即BD_h+i,用于体内(图10A)。在此实验设置中,表达柠檬黄蛋白和mCherry的细胞中应存在天蓝色荧光。

[0735] 将效价匹配的病毒在出生后第21天(P21)视网膜下注射到WT C57B16/J小鼠中。注射后两周收获视网膜后,可以在视网膜色素上皮中检测到固体荧光团表达(图10B)。此外,在表达柠檬黄蛋白和mCherry的所有区域都可以观察到天蓝色荧光,表明在与两种AAV共转导的细胞中成功的mRNA反式剪接。然而,柠檬黄蛋白和天蓝色具有部分重叠的激发和发射光谱。为了排除天蓝色荧光是由例如柠檬黄蛋白或mCherry的渗色引起的假象的可能性,通过用高强度514nm激光激发荧光团在RPE的小区域选择性漂白两种荧光团。通过这一程序,柠檬黄蛋白和mCherry荧光可以完全去除(图10C)。然而,天蓝色荧光保持不变,表明它仅来源于反式剪接的天蓝蛋白基因。该实验为mRNA反式剪接用于体内重组由AAV表达的基因提供了原理证明。

[0736] 实施例5:鉴定适用于人类基因治疗的有效结合域

[0737] 到目前为止,所有的结合域(BD)序列都是从人类内含子区域获得的。因此,当用于人类基因治疗时,它们还可能与内源性mRNAs结合,并诱导与这些转录物的反式剪接。为了在人类基因治疗中应用mRNA反式剪接,因此有必要鉴定不包含与人类基因组同源序列的BD。为此,从细菌lacZ基因中提取随机的100bp序列,并通过随机插入、缺失和替换进行修饰,以获得与人类基因组没有任何同源性的四个序列(图11A)。

[0738] 当用含有各自结合域的5'载体和3'载体构建体共转染HEK293细胞时,可以观察到其中一种BD(BD_k,图11B-D)的极高的天蓝蛋白重构效率为78.3%±2.1%。因此,BD_k比图4所示的表现最佳的人类BD更有效,因此被用于评估通过mRNA反式剪接重构大基因的初步实验。

[0739] 实施例6:多聚腺苷酸化信号对重构效率的影响以及体外试验中基于DNA的重构的参与

[0740] 由于来源于5'和3'载体的两个pre-mRNA分子的重构应该发生在细胞核中,理论上在5'载体中可以省略对于稳定和翻译成熟mRNA所必需的聚腺苷酸化信号(pA)。这种缺失的优点是任何残留的未剪接的5'pre-mRNA都不会导致截短蛋白的翻译。因此,研究了5'载体中pA缺失对天蓝蛋白重构效率的影响。为此,将缺乏pA的5'载体与常规的3'载体共转染。与含有pA的共转染载体相比,重构似乎略有减少(图12A和12B)。然而,该实验表明,如果必要,可以省略pA信号。此外,迄今为止观察到的成功重构理论上也可以通过在DNA水平上的同源重组来介导,如现有技术的杂交双载体方法,因为所有必要的组分,即重组基因序列和用于通过顺式剪接去除该序列的剪接元件存在于mRNA反式剪接载体中(Carvalho et al.,

2017)。为了排除这种可能性,删除了3'载体中的启动子,以防止其转录成pre-mRNA,因此类似于迄今已知的用于杂交双载体方法的3'载体。将该构建体与常规5'载体共转染后,未观察到重构。这一结果证实了所有观察到的天蓝蛋白重构都是完全通过mRNA反式剪接介导的。

[0741] 实施例7:通过mRNA反式剪接重构大基因的原理证明

[0742] 除了用于评估mRNA反式剪接的AAV天蓝蛋白分裂报告测定,还测试了一种具有治疗潜力的蛋白质。转录激活因子SpCas9-VPR,一种与转录激活因子域VP64-p65-Rta (VPR) 融合的无催化活性的核酸酶,是最近开发的基因治疗新工具。由于其体积较大(5.8kb),需要通过双载体进行体内应用。因此,SpCas9-VPR代表了通过mRNA反式剪接重构的合适候选物。SpCas9-VPR的编码序列在c.2185被分成两半,并且编码序列的两半配备有BD_k或其互补序列和适当的剪接元件,即DSS或vgASS_620。全长(FL) SpCas9-VPR构建体用作阳性对照。来自与分裂构建体共转染的HEK293细胞的RT-PCR显示SpCas9-VPR基因被成功重构,并且没有不需要的副产物产生(图13B)。PCR产物的测序证实了阅读框的准确恢复(图13C)。此外,FL SPCA 9-VPR蛋白可以通过蛋白质印迹法检测(图13D)。SpCas9-VPR的重构效率为13.2%±0.9%。这为大基因重构的mRNA反式剪接方法的适用性提供了原理证明。

[0743] 实施例8:ABCA4的重构(图14)

[0744] 最后,为了研究人类来源的大基因的重构,选择了编码视网膜ATP结合盒转运蛋白的ABCA4基因(6.8kb)。该基因是基因治疗的合适候选基因,因为其突变导致遗传性视网膜营养不良Stargardt黄斑营养不良。ABCA4在c.3243分裂成两半,并配备BD_k、DSS和vgASS_620(图14A)。两半都包含一个长的无内含子的ABCA4编码序列(CDS),它不像天然的由外显子和内含子组成的pre-mRNA。因此,这可能会阻碍剪接因子向pre-mRNA的募集,因此也可能会降低mRNA反式剪接的效率。为了使分裂的基因更接近内源性人类pre-mRNA,设计了额外的构建体,在CDS的两半中包含三个短的(80bp)插入内含子。将HEK293细胞与无内含子或内含子的构建体瞬时共转染,以检测体外重构能力。RT-PCR显示,对于有和没有插入内含子的分裂构建体,ABCA4在mRNA水平上成功重构(图14B)。有趣的是,含有内含子的分裂构建体似乎能更有效地进行反式剪接。同样,没有检测到非特异性剪接产物,阅读框被正确恢复(图14C)。

[0745] 实施例9:ABCA4的体内rAAV双载体mRNA反式剪接(图15)

[0746] 在最后的实验中,还在体内测试了ABCA4的重构。为此,用人视紫红质(hRH0)启动子替换CMV启动子,以确保感光细胞特异性表达(图15A)。如WO 2019/076856中所述,使用源自野生型AAV2衣壳的内部优化的NN和GL衣壳变体生产AAV。将含有ABCA4的5'或3'CDS的滴度匹配的病毒共注射到1月龄C57B16/J野生型小鼠的视网膜下。4周后,收获视网膜,并在mRNA和蛋白质水平上评估成功重构。使用跨接点引物对进行的RT-PCR显示,两种衣壳变体,即NN和GL,均导致成功重构,其中NN衣壳产生更高水平的重构mRNA,这很可能是由于更高的(共)转导效率(图15B)。通过测序证实了两个分离的pre-mRNA分子的无缝连接(图15C)。为了进行更精确的定量,进行了qRT-PCR。这些初步结果(n=1)表明,相对于未注射的C57B16/J野生型视网膜,相对ABCA4表达增加了10倍至42倍,再次证实了在mRNA水平上的成功和有效重构(图15D)。最后,为了在蛋白质水平上研究ABCA4的重构,将注射的视网膜的蛋白质裂解物用于蛋白质印迹。为了确保转基因衍生蛋白的特异性检测,使用了抗myc抗体。结果显

示,在这两种情况下,重构的mRNA导致蛋白质在体内成功表达(图15E)。

序列表

<110> 维格内罗有限责任公司

<120> 用于生物学应用和生物技术应用的优化的受体剪接位点模块

<130> 116313P1274PC

<150> EP18214415.4

<151> 2018-12-20

<160> 40

<170> BiSSAP 1.3.6

<210> 1

<211> 13

<212> DNA

<213> 未知

<220>

<223> 嘧啶区

<400> 1

tcttttgtca tct 13

<210> 2

<211> 10

<212> DNA

<213> 未知

<220>

<223> 内含子剪接增强子

<400> 2

tggggggagg 10

<210> 3

<211> 26

<212> DNA

<213> 未知

<220>

<223> ASS区

<400> 3

caacgagtct tttgtcatct acaggt 26

<210> 4

<211> 26

<212> DNA

<213> 未知

<220>

<223> ASS区

<400> 4
caacgagttt tttgtcatct acaggt 26
<210> 5
<211> 50
<212> DNA
<213> 未知
<220>
<223> ASS区
<400> 5
ctacacgctc aagccggagg tcaacaacga gtcttttgtc atctacaggt 50
<210> 6
<211> 38
<212> DNA
<213> 未知
<220>
<223> ASS区
<400> 6
gccggaggtc aacaacgagt cttttgtcat ctacaggt 38
<210> 7
<211> 20
<212> DNA
<213> 未知
<220>
<223> ASS区
<400> 7
gtcttttgtc atctacaggt 20
<210> 8
<211> 39
<212> DNA
<213> 未知
<220>
<223> ASS区
<400> 8
gtcttttgtc atctacaggt gttcgtggtt cgtggtcca 39
<210> 9
<211> 166
<212> DNA
<213> 未知
<220>

<223> 视紫红质基因的外显子3(含有 c.620T>G突变)

<400> 9

gtacatcccc gagggcctgc agtgctcgtg tggaatcgac tactacacgc tcaagccgga 60
 ggtcaacaac gaggcttttg tcatctacag gttcgtggtc cacttcacca tccccatgat 120
 tatcatctttt ttctgctatg ggcagctcgt cttcacctgc aaggag 166

<210> 10

<211> 1213

<212> DNA

<213> 未知

<220>

<223> 突变的视紫红质

<400> 10

ccagctggag ccctgagtgg ctgagctcag gccttcgcag cattcttggg tgggagcagc 60
 cacgggtcag ccacaagggc cacagccatg aatggcacag aaggccctaa cttctacgtg 120
 cccttctcca atgacgacggg tgtggtacgc agccccctcg agtaccaca gtactacctg 180
 gctgagccat ggcagttctc catgctggcc gcctacatgt ttctgctgat cgtgctgggc 240
 ttccccatca acttctctac gctctacgtc accgtccagc acaagaagct gcgcacgcct 300
 ctcaactaca tcttctctca cctagccgtg gctgacctct tcatggtcct aggtggcttc 360
 accagcacc tctacacctc tctgcatgga tacttctctc tcgggccccac aggatgcaat 420
 ttggagggct tctttgccac cctgggcggt gaaattgcc tgtggtcctt ggtggtcctg 480
 gccatcgagc ggtacgtggt ggtgtgtaag cccatgagca acttccgctt cggggagaac 540
 catgccatca tgggcggtgc cttcacctgg gtcatggcgc tggcctgcgc cgcaccccc 600
 ctcgccggct ggtccagta catccccgag ggctgcagc gctcgtgtgg aatcgactac 660
 tacacgctca agccggaggt caacaacgag tcttttgtca tctacaggtt cgtggtccac 720
 ttcacatcc ccatgattat catctttttc tgctatgggc agctcgtctt caccgtcaag 780
 gaggccgctg cccagcagca ggagtcagc accacacaga aggcagagaa ggaggtcacc 840
 cgcattggtca tcatcatggt catcgtttc ctgatctgct ggggtgcccta cgcagcgtg 900
 gcattctaca tcttcacca ccagggtcc aacttcggtc ccatcttcat gaccatccca 960
 gcgttctttg ccaagagcgc cgccatctac aacctgtca tctatatcat gatgaacaag 1020
 cagttccgga actgcatgct caccaccatc tgctgcggca agaaccact gggtgacgat 1080
 gaggcctctg ctaccgtgtc caagacggag acgagccagg tggccccggc ctaagacctg 1140
 cctaggactc tgtggccgac tataggcgtc tccatecccc tacaccttc cccagccaca 1200
 gccatccccac cag 1213

<210> 11

<211> 13

<212> DNA

<213> 未知

<220>

<223> 嘧啶区

<400> 11
ttttttgtca ttt 13
<210> 12
<211> 14
<212> DNA
<213> 未知
<220>
<223> 嘧啶区
<400> 12
tcttttgtca tcta 14
<210> 13
<211> 21
<212> DNA
<213> 未知
<220>
<223> 嘧啶区
<400> 13
caacgagtct tttgtcatct a 21
<210> 14
<211> 130
<212> DNA
<213> 未知
<220>
<223> 反向重复序列
<400> 14
ctgcgcgctc gctcgtcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60
ggtcgccccgg cctcagtgag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120
aggggttcct 130
<210> 15
<211> 130
<212> DNA
<213> 未知
<220>
<223> 反向重复序列
<400> 15
aggaaccct agtgatggag ttggcactc cctctctgcg cgctcgctcg ctactgagg 60
ccgggcgacc aaaggtcgcc cgacgccccg gctttgccc ggcggcctca gtgagcgagc 120
gagcgcgcag 130
<210> 16

<211> 82
 <212> DNA
 <213> 未知
 <220>
 <223> poly A序列
 <400> 16
 gcagtgaaaa aaatgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaacat 60
 tataagctgc aataaacaag tt 82
 <210> 17
 <211> 100
 <212> DNA
 <213> 未知
 <220>
 <223> 第一结合域
 <400> 17
 agtgcacaa ggcgatcaca tcagtgaaaa aaagccagac aggcggttaa accaacgcag 60
 attaaacagc aggatgcaaa aattcgcagg tggtcagatg 100
 <210> 18
 <211> 100
 <212> DNA
 <213> 未知
 <220>
 <223> 第二结合域
 <400> 18
 catctgacca cctgcgaatt tttgcatcct gctgtttaat ctgcgttggt ttaaccgcct 60
 gtctggccttt ttttactga tgtgatcgcc ttgatgcact 100
 <210> 19
 <211> 4455
 <212> DNA
 <213> 未知
 <220>
 <223> 图5所示的序列
 <400> 19
 ctgcgcgctc gctcgetcac tgaggccgcc cgggcaaagc ccggcgctcg ggcgaccttt 60
 ggtcgccccg cctcagtgag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120
 aggggttctt tgtagttaat gattaaccgg ccatgctact tatctacgta gccatgctct 180
 aggaagatcg gaattcgccc ttaagggcgc gccgtttaa tagctagcga cattgattat 240
 tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagccca tatatggagt 300
 tccgcgttac ataacttac gtaaattgcc cgctggctg accgccaac gacccccgcc 360

cattgacgtc aataatgacg tatgttccca tagtaacgcc aatagggact ttccattgac 420
gtcaatgggt ggagtattta cggtaaactg cccacttggc agtacatcaa gtgtatcata 480
tgccaagtac gccccctatt gacgtcaatg acggtaaagt gcccgcctgg cattatgccc 540
agtacatgac cttatgggac tttcctactt ggagctacat ctacgtatta gtcacgcta 600
ttaccatggt gatgcggttt tggcagtaca tcaatgggcg tggatagcgg tttgactcac 660
ggggatttcc aagtctccac cccattgacg tcaatgggag tttgttttgg caccaaaatc 720
aacgggactt tccaaaatgt cgtaacaact ccgccccatt gacgcaaagt ggcggtaggc 780
gtgtacgggt ggaggtctat ataagcagag ctggtaccac cggtgccacc atgggcttcg 840
tgagacagat acagcttttg ctctggaaga actggaccct gcgaaaagg caaaagattc 900
gctttgtggt ggaactcgtg tggcctttat ctttatttct ggtcttgatc tggttaagga 960
atgccaacc actctacagc catcatgaat gccatttccc caacaaggcg atgccctcag 1020
caggaatgct gccgtggetc caggggatct tctgcaatgt gaacaatccc tgttttcaa 1080
gccccacccc aggagaatct cctggaattg tgtcaacta taacaactcc atcttgca 1140
gggtatatcg agattttcaa gaactcctca tgaatgcacc agagagccag caccttgccc 1200
gtatttggac agagctacac atcttgccc aattcatgga caccctccgg actcaccgg 1260
agagaattgc aggaagagga atacgaataa gggatatctt gaaagatgaa gaaactga 1320
cactatttct cattaaaaac atcggcctgt ctgactcagt ggtctacctt ctgatcaact 1380
ctcaagtccg tccagagcag ttcgctcatg gactcccga cctggcctg aaggacatcg 1440
cctgcagcga ggccctcctg gagcgttca tcatcttcag ccagagacgc ggggcaaaga 1500
cgggtgcgcta tgccctgtgc tcctctccc agggcaccct acagtggata gaagacactc 1560
tgtatgcaa cgtggacttc ttcaagctct tccgtgtget tcccactc ctagacagcc 1620
gttctcaagg tatcaatctg agatcttggg gaggaatatt atctgatatg tcaccaagaa 1680
ttcaagagtt tatccatcgg ccgagtatgc aggacttget gtgggtgacc aggccctca 1740
tgcagaatgg tggctcagag acctttaca agctgatggg catcctgtct gacctcctgt 1800
gtggctacc caggggaggt ggctctcggg tgctctcctt caactggtat gaagacaata 1860
actataaggc ctttctgggg attgactcca caaggaagga tcctatctat tcttatgaca 1920
gaagaacaac atccttttgt aatgcattga tccagagcct ggagtcaaat ctttaacca 1980
aaatcgcttg gagggcggca aagccttgc tgatgggaaa aatcctgtac actcctgatt 2040
cacctgcagc acgaaggata ctgaagaatg ccaactcaac ttttgaagaa ctggaacacg 2100
ttaggaagtt ggtcaaagcc tgggaagaag tagggccca gatctggtac ttctttgaca 2160
acagcacaca gatgaacatg atcagagata cctggggaa cccaacagta aaagactttt 2220
tgaataggca gcttgggtgaa gaaggtatta ctgctgaagc catcctaac ttctctaca 2280
agggccctcg ggaaagccag gctgacgaca tggccaactt cgactggagg gacatattta 2340
acatcaactga tcgaccctc cgcttggtca atcaatacct ggagtgcttg gtcttgata 2400
agtttgaaag ctacaatgat gaaactcagc tcaccaacg tgccctctct ctactggagg 2460
aaaacatggt ctgggcccga gtggtattcc ctgacatgta tccttgacc agctctctac 2520
caccacacgt gaagtataag atccgaatgg acatagacgt ggtggagaaa accaataaga 2580
ttaagacag gtattgggat tctggtccca gagctgatcc cgtggaagat ttccggtaca 2640
tctggggcgg gtttgcctat ctgcaggaca tggttgaaca ggggatcaca aggagccagg 2700

tgcaggcgga ggctccagtt ggaatctacc tccagcagat gccctacccc tgcttcgtgg 2760
 acgattcttt catgatcatt ctgaaccgct gtttcctat cttcatggtg ctggcatgga 2820
 tctactctgt ctccatgact gtgaagagca tcgtcttggga gaaggagttg cgactgaagg 2880
 agaccttgaa aatcagggt gtctccaatg cagtgatttg gtgtacctgg ttcctggaca 2940
 gcttctccat catgtcgat agcatcttcc tctgacgat attcatcatg catggaagaa 3000
 tcctacatta cagcgacca ttcattctct tctgttctt gttggctttc tccactgcca 3060
 ccatcatgct gtgctttctg ctgagcact tcttctcaa gccagcttg gcagcagcct 3120
 gtagtggtgt catctatttc accctctacc tgccacacat cctgtgcttc gcctggcagg 3180
 accgcatgac cgctgagctg aagaaggctg tgagcttact gtctccggtg gcatttggat 3240
 ttggcactga gtacctggtt cgctttgaag agcaaggcct ggggctgcag tggagcaaca 3300
 tcgggaacag tcccacggaa ggggacgaat tcagcttctt gctgtccatg cagatgatgc 3360
 tccttgatgc tgctgtctat ggcttactcg cttggtacct tgatcagggt tttccaggag 3420
 actatggaac cccacttctt tggtacttct ttctacaaga gtcgtattgg cttggcggtg 3480
 aagggtgttc aaccagagaa gaaagagccc tggaaaagac cgagccccta acagaggaaa 3540
 cggaggatcc agagcaccca gaaggaatac acgactcctt ctttgaactg gagcatccag 3600
 ggtgggttcc tggggtatgc gtgaagaatc tggtaaagat ttttgagccc tgtggccggc 3660
 cagctgtgga ccgtctgaac atcaccttct acgagaacca gatcaccgca ttctgggcc 3720
 acaatggagc tgggaaaacc accaccttgt ccatcctgac gggctctgtt ccaccaacct 3780
 ctgggactgt gctcgttggg ggaagggaca ttgaaaccag cctggatgca gtccggcaga 3840
 gccttggcat gtgtccacag cacaacatcc tgttcacca cctcacggtg gctgagcaca 3900
 tgctgttcta tgcccagctg aaaggaaagt cccaggagga ggcccagctg gagatggaag 3960
 ccatgttggga ggacacaggc ctccaccaca agcggaatga agaggctcag gacctatcag 4020
 gtggcatgca gagaaagctg tcggttgcca ttgccttgtt gggagatgcc aaggtaaagg 4080
 cactgagcag aagggaagaa gctccggggg ctctttgtag ggtaagctta gtgcatcaag 4140
 gcgatcacat cagtgaaaaa aagccagaca ggcggttaa ccaacgcaga ttaaacagca 4200
 ggatgcaaaa attcgcaggt ggtcagatga agcttattct cgagttaagg gcgaattccc 4260
 gattaggatc ttcttagagc atggctacgt agataagtag catggcgggt taatcattaa 4320
 ctacaaggaa cccctagtga tggagttggc cactcctctt ctgcgcgctc gctcgctcac 4380
 tgaggccggg cgaccaaagg tcgcccagc cccgggcttt gccggggcg cctcagtgag 4440
 cgagcgagcg cgag 4455
 <210> 20
 <211> 4855
 <212> DNA
 <213> 未知
 <220>
 <223> 图6所示的序列
 <400> 20
 ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctg ggcgaccttt 60
 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120

aggggttcct ttagttaa gattaacccg ccatgctact tatctacgta gccatgctct 180
 aggaagatcg gaattcgccc ttaagggcgc gccgtttaa tagctagcga cattgattat 240
 tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagccca tatatggagt 300
 tccgcgttac ataacttac gtaaatggcc cgcctggctg accgccaac gacccccgcc 360
 cattgacgtc aataatgacg tatgttccca tagtaacgcc aataggact ttccattgac 420
 gtcaatgggt ggagtattta cggtaaactg cccacttggc agtacatcaa gtgtatcata 480
 tgccaagtac gccccctatt gacgtcaatg acggtaaatg gccgcctgg cattatgcc 540
 agtacatgac cttatgggac tttctactt ggcagtacat ctacgtatta gtcacgcta 600
 ttaccatggt gatgcggttt tggcagtaca tcaatggcg tggatagcgg ttgactcac 660
 ggggatttcc aagtctccac cccattgacg tcaatgggag tttgttttgg caccaaatc 720
 aacgggactt tccaaaatgt cgtaacaact ccgcccatt gacgcaatg ggcggtaggc 780
 gtgtacggtg ggaggtctat ataagcagag ctggtaccac cggtagaaag cttcatctga 840
 ccacctgca atttttgcat cctgtgttt aatctgcgtt ggtttaaccg cctgtctggc 900
 tttttttcac tgatgtgat gccttgatgc actaagctt aacgagtctt ttgtcatcta 960
 caggtggtga ttctggacga acccacctct ggggtggacc cttactcgag acgtcaatc 1020
 tgggatctgc tctgaagta tcgctcaggc agaaccatca tcatgtccac tcaccacatg 1080
 gacgaggccg acctccttgg ggaccgcatt gccatcattg cccaggaag gctctactgc 1140
 tcaggcacc cactcttct gaagaactgc tttggcacag gcttgtactt aaccttggtg 1200
 cgcaagatga aaaacatcca gagccaaagg aaaggcagtg aggggacctg cagctgctcg 1260
 tctaagggtt tctccaccac gtgtccagcc cacgtcgatg acctaactcc agaacaagtc 1320
 ctggatgggg atgtaaata gctgatggat gtagttctcc accatgttcc agaggcaaag 1380
 ctggtggagt gcattggtca agaacttatc ttcttcttc caataagaa cttcaagcac 1440
 agagcatatg ccagcctttt cagagagctg gaggagacgc tggctgacct tggctcagc 1500
 agttttggaa tttctgacac tcccctggaa gagatttttc tgaaggcac ggaggattct 1560
 gattcaggac ctctgtttgc ggggtggcgt cagcagaaa gagaaaact caacccccga 1620
 caccctgct tgggtcccag agagaaggct ggacagacac cccaggactc caatgtctgc 1680
 tccccaggg cgccggctgc tcaccagag ggccagcctc cccagagcc agagtgccca 1740
 ggcccgcagc tcaacacggg gacacagctg gtctccagc atgtgcaggc gctgctggtc 1800
 aagagattcc aacacacat ccgcagccac aaggacttcc tggcgcagat cgtgctccc 1860
 gctacctttg tgtttttggt tctgatgctt tctattgta tccctcctt tggcgaatac 1920
 cccgctttga ccctcacc ctggatata gggcagcagt acaccttctt cagcatggat 1980
 gaaccaggca gtgagcagtt cacggtactt gcagacgtcc tctgaataa gccaggcttt 2040
 ggcaaccgct gcctgaagga aggttggtt ccggagtacc cctgtggcaa ctcaacacc 2100
 tggaaactc cttctgtgtc cccaaacatc acccagctgt tccagaagca gaaatggaca 2160
 caggtcaacc cttcaccatc ctgcaggtgc agcaccaggg agaagctcac catgctgcca 2220
 gagtgcccc aggtgcccg gggcctccc cccccaga gaacacagc cagcacggaa 2280
 attctacaag acctgacgga caggaacatc tccgacttct tggtaaaaac gtatctgct 2340
 cttataagaa gcagcttaa gagcaaattc tgggtcaatg aacagaggta tggaggaatt 2400
 tccattggag gaaagctccc agtctcccc atcacgggg aagcacttgt tgggttttta 2460

agcgaccttg gccgcatcat gaatgtgagc gggggcccta tcactagaga ggcctctaaa 2520
 gaaataacctg atttccttaa acatctagaa actgaagaca acattaaggt gtggtttaat 2580
 aacaaaggct ggcatgccct ggtcagcttt ctcaatgtgg cccacaacgc catcttacgg 2640
 gccagcctgc ctaaggacag gagccccgag gagtatggaa tcaccgtcat tagccaacce 2700
 ctgaacctga ccaaggagca gctctcagag attacagtgc tgaccacttc agtggatgct 2760
 gtggttgcca tctgctgat tttctccatg tccttcgtcc cagccagctt tgtcctttat 2820
 ttgatccagg agcgggtgaa caaatccaag cacctccagt ttatcagtgg agtgagcccc 2880
 accacctact ggggtgaccaa cttcctctgg gacatcatga attattccgt gagtgctggg 2940
 ctgggtgggg gcactctcat cgggtttcag aagaaagcct acacttctcc agaaaacctt 3000
 cctgcccttg tggcactgct cctgctgtat ggatggggcg tcattcccat gatgtacca 3060
 gcactcctcc tgtttgatgt ccccagcaca gcctatgtgg ctttatcttg tgctaactctg 3120
 ttcatcggca tcaacagcag tgetattacc ttcatcttgg aattatttga gaataaccgg 3180
 acgctgctca ggttcaacgc cgtgctgagg aagctgctca ttgtcttccc ccacttctgc 3240
 ctgggcccgg gcctcattga ccttgcactg agccaggctg tgacagatgt ctatgcccgg 3300
 tttggtgagg agcactctgc aaatccgttc cactgggacc tgattgggaa gaacctgttt 3360
 gccatgggtg tggaaggggt ggtgtacttc ctctgacct tgctggtcca gcgccacttc 3420
 ttctctccc aatggattgc cgagcccact aaggagccca ttgttgatga agatgatgat 3480
 gtggctgaag aaagacaaag aattattact ggtggaaata aaactgacat cttaggcta 3540
 catgaactaa ccaagattta tccaggcacc tccagcccag cagtggacag gctgtgtgtc 3600
 ggagttcgcc ctggagagt ctttggcctc ctgggagtga atggtgccgg caaacaacc 3660
 acattcaaga tgctcactgg ggacaccaca gtgacctcag gggatgccac cgtagcaggc 3720
 aagagtattt taaccaatat ttctgaagtc catcaaaata tgggctactg tcctcagttt 3780
 gatgcaattg atgagctgct cacaggacga gaacatcttt acctttatgc ccggcttcga 3840
 ggtgtaccag cagaagaaat cgaaaagggt gcaactgga gtattaagag cctgggcctg 3900
 actgtctacg ccgactgcct ggctggcacg tacagtgggg gcaacaacgc gaaactctcc 3960
 acagccatcg cactcattgg ctgcccaccg ctggtgctgc tggatgagcc caccacaggg 4020
 atggaccccc aggcacgccg catgctgtgg aacgtcatcg tgagcatcat cagagaaggg 4080
 aggctgtgg tcctcacatc ccacagcatg gaagaatgtg aggcactgtg taccggctg 4140
 gccatcatgg taaagggcgc cttctgatgt atgggacca ttcagcatct caagtccaa 4200
 tttggagatg gctatatcgt cacaatgaag atcaaatccc cgaaggacga cctgcttctc 4260
 gacctgaacc ctgtggagca gttcttccag gggaacttcc caggcagtgt gcagagggag 4320
 aggcactaca acatgctcca gttccaggtc tectectect ccctggcgag gatcttccag 4380
 ctctctctct cccacaagga cagctgctc atcgaggagt actcagtcac acagaccaca 4440
 ctggaccagg tgtttgtaaa ttttgetaaa cagcagactg aaagtcatga cctccctctg 4500
 caccctcgag ctgctggagc cagtcgacaa gcccaggact gactcgacgc ggccgcgcag 4560
 tgaaaaaaat gctttatttg tgaattttgt gatgctattg ctttatttgt aaccattata 4620
 agctgcaata aacaagttct cgagttaagg gcgaattccc gattaggatc ttctagagc 4680
 atggctacgt agataagtag catggcgggt taatcattaa ctacaaggaa ccctagtga 4740
 tggagttggc cactccctct ctgctgctc gctcgtcac tgaggccggg cgaccaaagg 4800

tcgcccgcacg cccgggcttt gcccgggcgg cctcagttag cgagcgagcg cgcag 4855
 <210> 21
 <211> 4390
 <212> DNA
 <213> 未知
 <220>
 <223> 图7所示的序列
 <400> 21
 ctgcgcgctc gctcgtcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgacctt 60
 ggtcgccccg cctcagttag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120
 aggggttccct ttagttaa gattaaccg ccatgctact tatctacgta gccatgctct 180
 aggaagatcg gaattcgccc ttaagggcgc gccgtttaa tagctagcaa aagactaaaa 240
 gagggagggg tcaactcaga tctgccgagt gactcgattg gacttaaagg gccagtcaaa 300
 ccctgactgc cggctcatgg caggctcttg ccgaggaaa atgccagcc tatatttatg 360
 caaagagatt ttgttccaaa ctttaaggtca aagataccta aagacatccc ctcaggaac 420
 ccctctcatg gaggagagt cctgagggtc ttggtttccc attgcatccc ccacctcaat 480
 ttccctgggtg cccagccact tgtgtcttta gggttctctt tctctccata aaaggagacc 540
 aacacagtgt cggcctcctc tccccaaacta agggcttatg tgtaattaaa agggattatg 600
 ctttgaaggg gaaaagtagc ctttaatcac caggagaagg acacagcgtc cggagccaga 660
 ggcgctctta acggcgttta tgcctttgc tgtctgaggg gcctcagctc tgaccaatct 720
 ggtcttcgtg tggtcattag cggtagcggg accaccggg ccaccatggg cttcgtgaga 780
 cagatacagc ttttgctctg gaagaactgg accctgcgga aaaggcaaaa gattcgttt 840
 gtgggtggaac tcgtgtggcc tttatcttta tttctggtct tgatctgggt aaggaatgcc 900
 aaccactct acagccatca tgaatgcat ttccccaaac aggcgatgcc ctcagcagga 960
 atgctgccgt ggctccaggg gatcttctgc aatgtgaaca atccctgttt tcaaagcccc 1020
 accccaggag aatctcctgg aattgtgtca aactataaca actccatctt ggcaagggtta 1080
 tatcgagatt ttcaagaact cctcatgaat gcaccagaga gccagcacct tggccgtatt 1140
 tggacagagc tacacatctt gtccaattc atggacacc tccggactca cccggagaga 1200
 attgcaggaa gaggaatacg aataagggat atcttgaag atgaagaaac actgacacta 1260
 tttctcatta aaaacatcgg cctgtctgac tcagtgtctt accttctgat caactctcaa 1320
 gtccgtccag agcagttcgc tcatggagtc ccggacctgg cgtgaagga catcgctgc 1380
 agcgaggccc tcctggagcg cttcatcctc ttcagccaga gacgcggggc aaagacgggtg 1440
 cgctatgccc tgtgtccct ctcccagggc acctacagt ggatagaaga cactctgtat 1500
 gccaacgtgg acttcttcaa gettctcgt gtgttccca cactctaga cagccgttct 1560
 caaggatatca atctgagatc ttggggagga atattatctg atatgtcacc aagaattcaa 1620
 gaggttatcc atcgccgag tatgcaggac ttgctgtggg tgaccaggcc cctcatgcag 1680
 aatgggtggtc cagagacctt tacaagctg atgggcatcc tgtctgacct cctgtgtggc 1740
 taccocgagg gaggtggctc tcgggtgctc tcttcaact ggtatgaaga caataactat 1800
 aaggcctttc tggggattga ctccacaagg aagatccta tctattctta tgacagaaga 1860

acaacatcct tttgtaatgc attgatccag agcctggagt caaatccttt aaccaaatac 1920
 gcttggaggc cggcaaagcc tttgctgat ggaaaaatcc tgtacactcc tgattcacct 1980
 gcagcacgaa ggatactgaa gaatgccaac tcaacttttg aagaactgga acacgtagg 2040
 aagttggtca aagcctggga agaagtaggg cccagatct ggtacttctt tgacaacagc 2100
 acacagatga acatgatcag agataccctg gggaaccaa cagtaaaaga ctttttgaat 2160
 aggcagcttg gtgaagaagg tattactgct gaagccatcc taaacttctt ctacaagggc 2220
 cctcgggaaa gccaggctga cgacatggcc aacttcgact ggaggacat atttaacatc 2280
 actgatcgca ccctccgctt ggtcaatcaa tacctggagt gcttggctct ggataagttt 2340
 gaaagctaca atgatgaaac tcagctcacc caacgtgcc tctctctact ggaggaaaac 2400
 atgttctggg ccggagtggg attcctgac atgtatecct ggaccagctc tctaccacc 2460
 cacgtgaagt ataagatccg aatggacata gacgtggtgg agaaaaccaa taagattaaa 2520
 gacaggtatt gggattctgg tcccagagct gatcccgtgg aagatttccg gtacatctgg 2580
 ggcgggtttg cctatctgca ggacatggtt gaacagggga tcacaaggag ccagggtcag 2640
 gcggaggctc cagttggaat ctacctccag cagatgcctt accctgctt cgtggacgat 2700
 tctttcatga tcactctgaa ccgctgtttc cctatcttca tgggtctggc atggatctac 2760
 tctgtctcca tgactgtgaa gagcatcgtc ttggagaagg agttgcgact gaaggagacc 2820
 ttgaaaaatc aggggtctc caatgcagtg atttggtgta cctggttcct ggacagcttc 2880
 tccatcatgt cgatgagcat cttcctcctg acgatattca tcatgcatgg aagaatccta 2940
 cattacagcg acccattcat cctcttctg ttcttgttgg ctttctccac tgccaccatc 3000
 atgctgtgct ttctgctcag caccttctt tccaaggcca gtctggcagc agcctgtagt 3060
 ggtgtcatct atttcaccct ctacctgcca cacatcctgt gcttcgctg gcaggaccgc 3120
 atgaccgctg agctgaagaa ggctgtgagc ttactgtctc cgggtggcatt tggatttggc 3180
 actgagtacc tggttcgctt tgaagagcaa ggctggggc tgcaagtggag caacatcggg 3240
 aacagtccca cggaaagggga cgaattcagc ttctgctgt ccatgcagat gatgctcctt 3300
 gatgctgctg tctatggctt actcgttgg taccttgatc aggtgttcc aggagactat 3360
 ggaacccac ttccttggta ctttcttca caagatcgt attggcttgg cggtaaggg 3420
 tgttcaacca gagaagaaag agccctggaa aagaccgagc ccctaacaga gaaacggag 3480
 gatccagagc acccagaagg aatacacgac tccttctttg aacgtgagca tccagggtgg 3540
 gttcctgggg tatgctgaa gaatctggtg aagatttttg agccctgtgg ccggccagct 3600
 gtggaccgct tgaacatcac cttctacgag aaccagatca ccgattcct gggccacaat 3660
 ggagctggga aaaccaccac cttgtccatc ctgacgggtc tgttgccacc aacctctggg 3720
 actgtgctcg ttgggggaag ggacattgaa accagcctgg atgcagtccg gcagagcctt 3780
 ggcatgtgtc cacagcacia catctgttc caccacctca cgggtgctga gcacatgctg 3840
 ttctatgccc agctgaaagg aaagtcccag gaggaggccc agctggagat ggaagccatg 3900
 ttggaggaca caggcctcca ccacaagcgg aatgaagagg ctcaaggacct atcaggtggc 3960
 atgcagagaa agctgtcggg tgccattgcc tttgtgggag atgccaaggt aagggcactg 4020
 agcagaaggg aagaagctcc gggggctctt tgtagggtaa gcttagtgca tcaaggcgat 4080
 cacatcagtg aaaaaagcc agacaggcgg ttaaaccaac gcagattaaa cagcaggatg 4140
 caaaaattcg caggtggtca gatgaagctt attctcagat taaggcgaa ttcccatta 4200

ggatcttcct agagcatggc tacgtagata agtagcatgg cgggttaatc attaactaca 4260
 aggaaccctt agtgatggag ttggcactc cctctctgcg cgctcgctcg ctactgagg 4320
 ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcgccctca gtgagcgagc 4380
 gagcgcgcag 4390
 <210> 22
 <211> 4790
 <212> DNA
 <213> 未知
 <220>
 <223> 图8所示的序列
 <400> 22
 ctgcgcgctc gctcgetcac tgaggccgcc cgggcaaagc ccgggcgctc ggcgaccttt 60
 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120
 aggggttctt tgtagttaat gattaaccgg ccatgctact tatctacgta gccatgctct 180
 aggaagatcg gaattcgccc ttaagggcgc gccgtttaa tagctagcaa aagactaaaa 240
 gagggagggg tcacttcaga tctgccgagt gactcgattg gacttaaagg gccagtcaaa 300
 ccctgactgc cggctcatgg caggctcttg ccgaggacaa atgcccagcc tatatttatg 360
 caaagagatt ttgttccaaa ctttaagtca aagataccta aagacatccc cctcaggaac 420
 ccctctcatg gaggagagt cctgagggtc ttggtttccc attgcatccc ccacctcaat 480
 ttccctgggtg cccagccact tgtgtcttta gggttctctt tctctccata aaaggagacc 540
 aacacagtgt cggcctcctc tccccacta agggettatg tgtaattaaa aggattatg 600
 ctttgaaggg gaaaagtagc cttaatacac caggagaagg acacagcgtc cggagccaga 660
 ggcgctctta acggcgttta tgtccttgc tgtctgaggg gcctcagctc tgaccaatct 720
 ggtcttcgtg tggtcattag cggtagcggg accaccggtg gaaagcttca tctgaccacc 780
 tgcgaatctt tgcatcctgc tgtttaatct gcgttggttt aaccgctgt ctggcttttt 840
 ttcactgatg tgatcgctt gatgactaa gcttcaacga gtcttttgtc atctacaggt 900
 ggtgattctg gacgaacca cctctggggg ggaccttac tcgagacgct caatctggga 960
 tctgctcctg aagtatcgtc caggcagaac catcatcatg tccactcacc acatggacga 1020
 ggccgacctc cttggggacc gcattgccat cattgcccag ggaaggctct actgctcagg 1080
 caccctcctc ttctgaaga actgctttgg cacagcttg tacttaacct tgggtgcgcaa 1140
 gatgaaaaac atccagagcc aaaggaaagg cagtgagggg acctgcagct gctcgtctaa 1200
 gggtttctcc accacgtgtc cagcccacgt cgatgaccta actccagaac aagtcttggg 1260
 tgggatgta aatgagctga tggatgtagt tctccaccat gttccagagg caaagctggg 1320
 ggagtgcatt ggtaagaac ttatcttctt tcttccaaat aagaacttca agcacagagc 1380
 atatgccagc cttttcagag agctggagga gacgctggct gaccttggtc tcagcagttt 1440
 tggaaattct gacactcccc tggaagagat ttttctgaag gtcacggagg attctgattc 1500
 aggacctctg tttgcgggtg gcgctcagca gaaaagagaa aacgtcaacc cccgacacc 1560
 ctgcttgggt cccagagaga aggctggaca gacaccccag gactccaatg tctgctcccc 1620
 aggggcgccc gctgctcacc cagagggcca gcctcccca gagccagagt gccagggccc 1680

gcagctcaac acggggacac agctggtcct ccagcatgtg caggcgctgc tggtaagag 1740
attccaacac accatccgca gccacaagga cttcctggcg cagatcgtgc tcccggctac 1800
ctttgtgttt ttggctctga tgctttctat tgttatccct ccttttggcg aataccccgc 1860
tttgaccctt caccctgga tatatgggca gcagtacacc ttcttcagca tggatgaacc 1920
aggcagtgag cagttcacgg tacttgca ga cgtcctcctg aataagccag gctttggcaa 1980
ccgctgcctg aaggaagggt ggcttccgga gtaccctgt ggcaactcaa caccctggaa 2040
gactccttct gtgtcccaa acatcaccca gctgttccag aagcagaaat ggacacaggt 2100
caacccttca ccacctgca ggtgcagcac cagggagaag ctccatgc tgccagagt 2160
ccccgagggt gccgggggccc tcccgcctc ccagagaaca cagcgcagca cggaaattct 2220
acaagacctg acggacagga acatctccga cttcttggtg aaaacgtatc ctgctcttat 2280
aagaagcagc ttaaagagca aattctgggt caatgaacag aggtatggag gaatttccat 2340
tggaggaaaag ctcccagtcg tccccatcac gggggaagca cttgttgggt ttttaagcga 2400
ccttgccggg atcatgaatg tgagcggggg ccctatact agagaggcct ctaaagaaat 2460
acctgatttc cttaaacatc tagaaactga agacaacatt aaggtgtggt ttaataaca 2520
aggctggcat gccctggtca gctttctca tgtggcccac aacgccatct tacgggccag 2580
cctgcctaag gacaggagcc ccgaggagta tggaatcacc gtcattagcc aaccctgaa 2640
cctgaccaag gagcagctct cagagattac agtgcagacc acttcagtgg atgctgtggt 2700
tgccatctgc gtgattttct ccattgcctt cgtcccagcc agctttgtcc tttatttgat 2760
ccaggagcgg gtgaacaaat ccaagcacct ccagtttatc agtggagtga gccccaccac 2820
ctactgggtg accaacttcc tctgggacat catgaattat tccgtgagt ctgggctggt 2880
ggtgggcatc ttcacgggtt ttcagaagaa agcctacact tctccagaaa accttctgc 2940
ccttgtggca ctgctcctgc tgtatggatg ggcggtcatt cccatgatgt acccagcatc 3000
cttctgttt gatgtccca gcacagccta tgtggcttta tcttgtgcta atctgttcat 3060
cggcatcaac agcagtgcta ttacctcat cttggaatta tttgagaata accggacgct 3120
gctcaggttc aacgccgtgc tgaggaagct gctcattgtc tccccact tctgcctggg 3180
ccggggcctc attgacctg cactgagcca ggctgtgaca gatgtctatg cccggtttgg 3240
tgaggagcac tctgcaaatc cgttccactg ggacctgatt gggaagaacc tgtttgcat 3300
ggtggtggaa ggggtggtgt acttctcct gacctgctg gtccagcgc acttcttct 3360
ctcccaatgg attgccgagc ccactaagga gccattgtt gatgaagatg atgatgtggc 3420
tgaagaaaga caaagaatta ttactggtgg aaataaaact gacatcttaa ggctacatga 3480
actaaccaag atttatccag gcacctccag cccagcagtg gacaggctgt gtgtcggagt 3540
tcgccctgga gactgcttg gctcctggg agtgaatggt gccggcaaaa caaccacatt 3600
caagatgctc actggggaca ccacagtac ctcaggggat gccaccgtag caggcaagag 3660
tattttaacc aatatttctg aagtccatca aatatgggc tactgtctc agtttgatgc 3720
aattgatgag ctgctcacag gacgagaaca tctttacct tatgcccggc ttcgaggtgt 3780
accagcagaa gaaatcgaaa aggttgcaaa ctggagtatt aagagcctgg gcctgactgt 3840
ctacgccgac tgcttggtg gcacgtacag tgggggcaac aagcggaaac tctccacagc 3900
catcgcactc attggctgcc caccgctggt gctgctggat gagcccacca caggatgga 3960
ccccaggca cgccgatgc tgtggaacgt catcgtgagc atcatcagag aaggagggc 4020

tgtggctctc acatcccaca gcatggaaga atgtgaggca ctgtgtaccc ggctggccat 4080
 catggtaaag ggcgcctttc gatgtatggg caccattcag catctcaagt ccaaatttgg 4140
 agatggctat atcgtcacia tgaagatcaa atccccgaag gacgacctgc ttcctgacct 4200
 gaaccctgtg gagcagttct tccaggggaa cttcccaggc agtgtgcaga gggagaggca 4260
 ctacaacatg ctccagttcc aggtctcctc ctctccctg gcgaggatct tccagctcct 4320
 cctctcccac aaggacagcc tgctcatcga ggagtactca gtcacacaga ccacactgga 4380
 ccaggtgttt gtaaattttg ctaaacagca gactgaaagt catgacctcc ctctgcacc 4440
 tcgagctgct ggagccagtc gacaagccca ggactgagtc gacgcggccg cgagtgaaa 4500
 aaaatgcttt atttgtgaaa tttgtgatgc tattgcttta tttgtaacca ttataagctg 4560
 caataaacia gttctcgagt taagggcgaa ttcccgatta ggatcttct agagcatggc 4620
 tacgtagata agtagcatgg cgggttaatc attaactaca aggaaccct agtcatggag 4680
 ttggcactc cctctctgcg cgtctgctcg ctactgagg cgggcgacc aaaggtcgcc 4740
 cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag 4790

<210> 23

<211> 240

<212> DNA

<213> 未知

<220>

<223> ASS序列

<400> 23

ctcggcagcc accttggctg ttccaagtc cctcacagge aggtctctcc tacctgcctg 60
 tcctcaggta catccccgag ggctgcagt gctcgtgtgg aatcgactac tacacgctca 120
 agccggagggt caacaacgag tcttttgta tctacatggt cgtgggtccac ttcaccatcc 180
 ccatgattat catcttttct tgctatgggc agctcgtctt caccgtcaag gaggtacggg 240

<210> 24

<211> 240

<212> DNA

<213> 未知

<220>

<223> ASS序列

<400> 24

ctcggcagcc accttggctg ttccaagtc cctcacagge aggtctctcc tacctgcctg 60
 tcctcaggta catccccgag ggctgcagt gctcgtgtgg aatcgactac tacacgctca 120
 agccggagggt caacaacgag tcttttgta tctacaggtt cgtgggtccac ttcaccatcc 180
 ccatgattat catcttttct tgctatgggc agctcgtctt caccgtcaag gaggtacggg 240

<210> 25

<211> 50

<212> DNA

<213> 未知

<220>
<223> ASS序列
<400> 25
ctacacgctc aagccggagg tcaacaacga gtcttttgtc atctacatgt 50
<210> 26
<211> 27
<212> DNA
<213> 未知
<220>
<223> ASS序列
<400> 26
gtcttttgtc atctacaggt gttcgtg 27
<210> 27
<211> 102
<212> DNA
<213> 人工序列
<220>
<223> 结合域BD_g
<400> 27
gttccagga ggaatgtga agccccagaa aggccagcg ccaggtgaa ttcgctagct 60
cggcagccac ctggctgtt cccaagtccc tcacagcag gg 102
<210> 28
<211> 105
<212> DNA
<213> 人工序列
<220>
<223> 结合域BD_h+i
<400> 28
gttaaccctc cagtcaggac taaaccag tagtgtctgg ttccaggcac tgacctgcta 60
gctcggcagc caccttggt gttcccaagt cctcacagg cagg 105
<210> 29
<211> 149
<212> DNA
<213> 人工序列
<220>
<223> 结合域PTM1, B2
<400> 29
caccattcat ggtgatagcc gggctgctgt ttgtgcagg ctggcactga aactgcctt 60
gatcttattt ggagcaatat gcgcttgtct aatttcacag caagaaaact gagctgagc 120

tcaaagaagt caagcgcct gctggggcg 149

<210> 30

<211> 21

<212> DNA

<213> 人工序列

<220>

<223> ITR2正向引物

<400> 30

ggaacccta gtgatggagt t 21

<210> 31

<211> 16

<212> DNA

<213> 人工序列

<220>

<223> ITR2反向引物

<400> 31

cggcctcagt gagcga 16

<210> 32

<211> 100

<212> DNA

<213> 人工序列

<220>

<223> 结合域BD_j

<400> 32

catctgacca ccagcgaatt ttgcatcca gctgtttaat cagcgttggt ttaaccgcca 60

gtcaggcttt ctttcaaaga tgtgatcgcc ttgcagcact 100

<210> 33

<211> 100

<212> DNA

<213> 人工序列

<220>

<223> 结合域BD_I

<400> 33

ccatcccga tctgaccacc agcgaaatgg atttttgcat ccagctgggt aacgttggca 60

atttaaccgc cagtcagget ttctttcaaa gatgtggatt 100

<210> 34

<211> 100

<212> DNA

<213> 人工序列

<220>

<223> 结合域BD_m

<400> 34

ccatccccgca tctgaccacc tgcgaaatgg atttttgcat cctgctgggt aacgttggca 60
athtaaccgc ctgtctggct ttctttcaat gatgtggatt 100

<210> 35

<211> 1324

<212> DNA

<213> 人工序列

<220>

<223> 5' 天蓝蛋白 + BD_g

<400> 35

gacattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
catatatgga gttccgcgtt acataactta cggtaaattgg cccgcctggc tgaccgccc 120
acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
ctttccattg acgtcaatgg gtggactatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaagt acgcccccta ttgacgtcaa tgacggtaaa tggccccct 300
ggcattatgc ccagtacatg accttatggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtgatgcbgg tttggcagta catcaatggg cgtggatagc 420
ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg agtttgtttt 480
ggcaccacaaa tcaacgggac tttccaaaat gtcgtaacaa ctccgcccga ttgacgcaaa 540
tgggcbggtag gcgtgtacgg tgggaggtct atataagcag agctctctgg ctaactagag 600
aacccactgc ttactggcct atcgaaatta atacgactca ctataggag acccaagctg 660
gctagttaag cttatgggtga gcaagggcga ggagctgttc accggggtgg tgcccactct 720
ggtcgagctg gacggcgacg taaacggcca caagttcagc gtgtccggcg agggcgaggg 780
cgatgccacc tacggcaagc tgaccctgaa gttcatctgc accacaggta agggcactga 840
gcagaaggga agaagctccg ggggctcttt gtaggggtga tccccctgcc tgtgagggac 900
ttgggaacag ccaaggtggc tgccgagcta gcgaattcca cctggcgctg gccctttctg 960
gggcttcaca ttccctccct ggaacggatc cctcgaggtc acccattcga acaaaaactc 1020
atctcagaag aggatctgaa tatgcatacc ggtcatcacc accatcacca ttgagtttaa 1080
accgctgat cagcctcgac tgtgcttct agttgccagc catctgttgt ttgcccctcc 1140
cccgtgcctt ccttgaccct ggaaggtgcc actcccactg tcctttccta ataaaatgag 1200
gaaattgcat cgcattgtct gagtaggtgt cattctatc tggggggtgg ggtggggcag 1260
gacagcaagg gggaggattg ggaagacaat agcaggcatg ctggggatgc ggtgggctct 1320
atgg 1324

<210> 36

<211> 1710

<212> DNA

<213> 人工序列

<220>

<223> 3'天蓝蛋白 + BD_g

<400> 36

```
gacattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
catatatgga gttccgcgtt acataactta cggtaaattg cccgcctggc tgaccgcca 120
acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaatagga 180
ctttccattg acgtcaatgg gtggactatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaagt acgcccccta ttgacgtcaa tgacggtaa tgccccgcct 300
ggcattatgc ccagtacatg accttatggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtgatgcggg tttggcagta catcaatggg cgtggatagc 420
ggtttgactc acggggattt ccaagtctcc acccattga cgtcaatggg agtttgtttt 480
ggcaccaaaa tcaacgggac tttccaaaat gtcgtaacaa ctccgcccc ttagcgcaa 540
tgggcgtag gcgtgtacgg tgggaggtct atataagcag agctctctgg ctaactagag 600
aaccactgc ttactggctt atcgaaatta atacgactca ctataggag acccaagctg 660
gctagttaag cttagatccg ttccaggag ggaatgtgaa gccccagaa gggccagcgc 720
cagggtggaat tcgctagctc ggccagccacc ttggctgttc ccaagtcct cacaggcagg 780
gggatcccaa cgagcttttt gtcactaca ggtaagctgc ccgtgccctg gccaccctc 840
gtgaccacc tgacctgggg cgtgcagtgc ttcgccct accccacca catgaagcag 900
cacgacttct tcaagtccgc catgcccga ggctacgtcc aggagcgcac catcttcttc 960
aaggacgacg gcaactacaa gaccgcgcc gaggtgaagt tcgaggcga caccctggtg 1020
aaccgcatcg agctgaaggg catcgaattc aaggaggac gcaacatcct ggggcacaag 1080
ctggagtaca acgcatcag cgacaacgtc tatataccg ccgacaagca gaagaacggc 1140
atcaaggcca acttcaagat ccgccacaac atcgaggac gcagcgtgca gctcgccgac 1200
cactaccagc agaacacccc catcggcgac ggccccgtgc tgctgcccga caaccactac 1260
ctgagcacc agtccaagct gagcaaagac ccaacgaga agcgcgatca catggtcctg 1320
ctggagtctg tgaccgccgc cgggatcact ctggcatgg acgagctgta caagtaactc 1380
gaggtcacc attcgaacaa aaactcatct cagaagagga tctgaatatg cataccggtc 1440
atcatcacca tcaccattga gtttaaacc gctgatcagc ctgactgtg cttctagtt 1500
gccagccatc tgttgtttgc cctcccccg tgcttctctt gaccctgaa ggtgccactc 1560
ccactgtcct ttctaataa aatgaggaaa ttgcatcga ttgtctgagt aggtgtcatt 1620
ctattctggg ggttgggtg gggcaggaca gcaaggggga ggattgggaa gacaatagca 1680
ggcatgctgg ggatgcggtg ggctctatgg 1710
```

<210> 37

<211> 3142

<212> DNA

<213> 人工序列

<220>

<223> 5' SpCas9-VPR + BD_k

<400> 37

gacattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
catatatgga gttccgcgtt acataactta cggtaaattgg cccgcctggc tgaccgcccc 120
acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaagt acgcccccta ttgacgtcaa tgacggtaaa tggccccgct 300
ggcattatgc ccagtacatg accttatggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtgatgcggt tttggcagta catcaatggg cgtggatagc 420
ggtttgactc acggggattt ccaagtctcc accccattga cgtcaatggg agtttgtttt 480
ggcaccaaaa tcaacgggac tttccaaaat gtcgtaacaa ctccgcccc ttagcga 540
tgggcggtag gcgtgtacgg tgggaggtct atataagcag agctcgttta gtgaaccgtc 600
agaaccggtc gactagagga tccatggccc caaagaagaa gcggaaggtc ggtatccacg 660
gagtcccagc agccgacaag aagtactcca ttgggctcgc tatcggcaca aacagcgtcg 720
gctgggccgt cattacggac gagtacaagg tgccgagcaa aaaattcaaa gttctgggca 780
ataccgatcg ccacagcata aagaagaacc tcattggcgc cctcctgttc gactccgggg 840
aaacggccga agccacgcgg ctcaaaagaa cagcacggcg cagatatacc cgcagaaaga 900
atcggatctg ctacctgcag gagatcttta gtaatgagat ggctaagggtg gatgactctt 960
tcttccatag gctggaggag tcctttttgg tggaggagga taaaagcac gagcgccacc 1020
caatctttgg caatatcgtg gacgaggtgg cgtacatga aaagtacca accatatac 1080
atctgaggaa gaagcttcta gacagtactg ataaggctga cttgcggttg atctatctcg 1140
cgctggcgca tatgatcaaa tttcggggac acttctcat cgagggggac ctgaaccag 1200
acaacagcga tgtcgacaaa ctctttatcc aactggtta gacttacaat cagcttttcg 1260
aagagaaccg gatcaacgca tccggagttg acgccaagc aatcctgagc gctaggctgt 1320
ccaaatcccg gcggctcga aacctcatg cacagctccc tggggagaag aagaacggcc 1380
tgtttggtaa tcttatcgcc ctgtcactcg ggctgacccc caactttaa tctaacttcg 1440
acctggccga agatgccaaag cttcaactga gcaaagacac ctacgatgat gatctcgaca 1500
atctgctggc ccagatcggc gaccagtacg cagacctttt tttggcggca aagaacctgt 1560
cagacgccat tctgctgagt gatattctgc gactgaacac ggagatcacc aaagctccgc 1620
tgagcgctag tatgatcaag cgctatgatg agcaccacca agacttgact ttgctgaagg 1680
cccttgctcag acagcaactg cctgagaagt acaaggaaat tttcttcgat cagtctaaaa 1740
atggctacgc cggatacatt gacggcggag caagccagga ggaattttac aaatttatta 1800
agcccatctt ggaaaaaatg gacggcaccg aggagctgct ggtaaagctt aacagagaag 1860
atctgttgcg caaacagcgc actttcgaca atggaagcat cccccaccag attcacctgg 1920
gcgaactgca cgctatcctc aggcggcaag aggatttcta cccctttttg aaagataaca 1980
gggaaaagat tgagaaaatc ctcacatttc ggatacccta ctatgtaggc ccctcgccc 2040
ggggaaatc cagattcgcg tggatgactc gcaaatcaga agagaccatc actccctgga 2100
acttcgagga agtcgtggat aagggggcct ctgcccagtc cttcatcgaa aggatgacta 2160
actttgataa aaatctgcct aacgaaaagg tgcttctaa acactctctg ctgtacgagt 2220
acttcacagt ttataacgag ctaccaagg tcaatacgt cacagaagg atgagaaagc 2280
cagcattcct gtctggagag cagaagaaag ctatcgtgga cctcctcttc aagacgaacc 2340

ggaaagttac cgtgaaacag ctcaaagaag actatttcaa aaagattgaa tgtttcgact 2400
 ctgttgaaat cagcggagtg gaggatcgct tcaacgcatc cctgggaacg tatcacgac 2460
 tcctgaaaat cattaagac aaggacttcc tggacaatga ggagaacgag gacattcttg 2520
 aggacattgt cctcacctt acgttgttt agatagga gatgattgaa gaacgcttga 2580
 aaacttacgc tcatctcttc gacgacaaag tcatgaaaca gctcaagagg cgccgatata 2640
 caggatgggg gcggtgtca agaaaactga tcaatgggat ccgagacaag cagagtggaa 2700
 agacaatcct ggattttctt aagtccgatg gatttgccaa ccggaacttc atgcagttga 2760
 tccatgatga ctctctcacc ttaaggagg acatccagaa agcacaagtt tctggccagg 2820
 gggacagtct tcacgagcac atcgctaate ttgcaggtaa gggcactgag cagaaggga 2880
 gaagctccgg gggctctttg tagggtgcgg ccgcagtga tcaaggcgat cacatcagt 2940
 aaaaaagcc agacaggcgg ttaaaccaac gcagattaa cagcaggatg caaaaattcg 3000
 caggtgtca gatggcgcc gctctagact cgatgagttt ggacaaacca caactagaat 3060
 gcagtgaaaa aatgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaacat 3120
 tataagctgc aataaacaag tt 3142

<210> 38

<211> 4430

<212> DNA

<213> 人工序列

<220>

<223> 3' SpCas9-VPR + BD_k

<400> 38

gacattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
 catatatgga gttccgctt acataacta cggtaaattg cccgcctggc tgaccgcca 120
 acgacccccg cccattgacg tcaataatga cgtatgttc catagtaacg ccaatagga 180
 ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
 aagtgtatca tatgccaagt acgccccta ttgacgtcaa tgacggtaaa tggcccgcct 300
 ggcattatgc ccagtacatg accttatggg actttcctac ttggcagtac atctacgtat 360
 tagtcatcgc tattaccatg gtgatgcggt tttggcagta catcaatggg cgtggatagc 420
 ggtttgactc acggggattt ccaagtctcc acccattga cgtcaatggg agtttgtttt 480
 ggcacaaaaa tcaacgggac tttccaaaat gtcgtaaaa ctccgcccc ttgacgcaa 540
 tggcggttag gcgtgtacgg tgggaggtct atataagcag agctcgttta gtgaaccgtc 600
 agaggtaacca ccggtcgact agaggatccc ggccgccatc tgaccacctg cgaatttttg 660
 catcctgctg ttaaatctgc gttggtttaa ccgctgtct ggctttttt cactgatgtg 720
 atgccttga tgactcggc cgccaacgag tctttgtca tctacaggta gccagctat 780
 caaaaaggga atactgcaga ccgttaaggt cgtggatgaa ctcgtcaaag taatgggaag 840
 gcataagccc gagaatatcg ttatcgagat ggcccagag aaccaaacta ccagaagg 900
 acagaagaac agtagggaaa ggatgaagag gattgaagag ggtataaaag aactggggtc 960
 ccaaatcctt aaggaacacc cagttgaaaa caccagctt cagaatgaga agctctacct 1020
 gtactacctg cagaacggca gggacatgta cgtggatcag gaactggaca tcaatcggct 1080

ctccgactac gacgtggatc atatcgtgcc ccagtctttt ctcaaagatg attctattga 1140
 taataaagtg ttgacaagat ccgataaaaa tagagggag agtgataacg tcccctcaga 1200
 agaagttgtc aagaaaatga aaaattattg gcggcagctg ctgaacgcca aactgatcac 1260
 acaacggaag ttcgataatc tgactaaggc tgaacgaggt ggcctgtctg agttggataa 1320
 agccggcttc atcaaaaggc agcttgttga gacacgccag atcaccaagc acgtggccca 1380
 aattctcgat tcacgcatga acaccaagta cgatgaaaat gacaaactga ttcgagaggt 1440
 gaaagttatt actctgaagt ctaagctggt ctgagatttc agaaaggact ttcagtttta 1500
 taaggtgaga gagatcaaca attaccacca tgcgcatgat gcctacctga atgcagtggg 1560
 aggcaactgca cttatcaaaa aatateccaa gcttgaatct gaatttgttt acggagacta 1620
 taaagtgtac gatgttagga aatgatcgc aaagtctgag caggaaatag gcaaggccac 1680
 cgctaagtac ttcttttaca gcaatattat gaattttttc aagaccgaga ttacactggc 1740
 caatggagag attcggaagc gaccacttat cgaacaaac ggagaaacag gagaaatcgt 1800
 gtgggacaag ggtagggatt tcgcgacagt ccggaagtc ctgtccatgc cgcaggtgaa 1860
 catcgtaaa aagaccgaag tacagaccgg aggcttctcc aaggaaagta tctctccgaa 1920
 aaggaacagc gacaagctga tcgacgcaa aaaagattgg gacccaaga aatacggcgg 1980
 attcgattct cctacagtcg cttacagtgt actggttgtg gccaaagtgg agaaagggaa 2040
 gtctaaaaaa ctcaaaagcg tcaaggaact gctgggcatc acaatcatgg agcgatcaag 2100
 cttcgaaaaa aaccccatcg actttctcga ggcgaaagga tataaagagg tcaaaaaaga 2160
 cctcatcatt aagcttccca agtactctct ctttgagctt gaaaacggcc gaaacgaat 2220
 gctcgctagt gcgggagcgc tcagaaaagg taacgagctg gcaactgcct ctaaatacgt 2280
 taatttcttg tatctggcca gccactatga aaagctcaaa gggctctccg aagataatga 2340
 gcagaagcag ctgttcgtgg aacaacaaa aactacctt gatgagatca tcgagcaaat 2400
 aagcgaattc tcaaaaagag tgatcctcgc cgacgctaac ctcgataagg tgctttctgc 2460
 ttacaataag cacagggata agcccatcag ggagcaggca gaaaacatta tccacttggt 2520
 tactctgacc aacttgggagc gcctgcagc cttcaagtac ttcgacacca ccatagacag 2580
 aaagcggtag acctctacaa aggaggtcct ggacgccaca ctgattcatc agtcaattac 2640
 ggggctctat gaaacaagaa tcgacctc tcagctcggg ggagacagca gggctgacc 2700
 caagaagaag aggaaggtgt cgccagggat ccgtcgactt gacgcgttga tatcaacaag 2760
 tttgtacaaa aaagcaggct acaaagaggc cagcggttcc ggacgggctg acgattgga 2820
 cgattttgat ctggatatgc tgggaagtga cgccctcgat gattttgacc ttgacatgct 2880
 tggttcggat gcccttgatg actttgacct cgacatgctc ggcaagtacg cccttgatga 2940
 tttcgacctg gacatgctga ttaactctag aagtccgga tctccgaaaa agaaacgcaa 3000
 agttggtagc cagtacctgc ccgacaccga cgaccggcac cggatcgagg aaaagcggaa 3060
 gcggacctac gagacattca agagcatcat gaagaagtc cccttcagcg gccccaccga 3120
 ccctagacct ccacctagaa gaatcgccgt gccagcaga tccagcga gctgccc aaa 3180
 acctgcccc cagccttacc ccttaccag cagcctgagc accatcaact acgacagtt 3240
 ccctaccatg gtgttccca gcggccagat ctctcagcc tctgctctgg ctccagcccc 3300
 tctcagggtg ctgcctcagg ctctctctcc tgcaccagct ccagccatgg tgtctgact 3360
 ggctcaggca ccagcaccg tgctgtgct ggctctgga cctccacagg ctgtggctcc 3420

accagcccct aaacctacac aggccggcga gggcacactg tctgaagctc tgctgcagct 3480
gcagttcgac gacgaggatc tgggagccct gctgggaaac agcaccgatc ctgccgtgtt 3540
caccgacctg gccagcgtgg acaacagcga gttccagcag ctgctgaacc agggcatccc 3600
tgtggcccct cacaccaccg agcccatgct gatggaatac cccgaggcca tcaccggct 3660
cgtgacaggc gctcagaggc ctctgatcc agctctgccc cctctgggag caccaggcct 3720
gcctaattgga ctgctgtctg gcgacgagga cttcagctct atcgccgata tggatttctc 3780
agccttgctg ggctctggca gcggcagccc ggattccagg gaagggatgt ttttgccgaa 3840
gcctgaggcc ggctccgcta ttagtgacgt gtttgagggc cgcgagggtg gccagccaaa 3900
acgaatccgg ccatttcac ctcaggaag tccatggccc aaccgcccac tccccgccag 3960
cctcgacca acaccaaccg gtccagtaca tgagccagtc gggtcactga ccccgccacc 4020
agtccctcag ccaactggatc cagcgcggc agtgactccc gaggccagtc acctgttgga 4080
ggatcccgat gaagagacga gccaggctgt caaagcctt cgggagatgg ccgatactgt 4140
gattccccag aaggaagagg ctgcaatctg tggccaaatg gacctttccc atccgcccc 4200
aaggggcat ctggatgagc tgacaaccac acttgagtc atgaccgagg atctgaacct 4260
ggactcacc ctgaccccg aattgaacga gattctggat accttctga acgacgagt 4320
cctcttgcac gccatgcata tcagcacagg actgtccatc ttcgacacat ctctgttttg 4380
acaataaaat atctttatct tcattacatc tgtgtgttgg tttttgtgt 4430

<210> 39

<211> 4775

<212> DNA

<213> 人工序列

<220>

<223> 双AAV载体-5' ABCA4_w内含子

<400> 39

ctgcgcgctc gctcgtcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60
ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gaggggccaa ctccatcact 120
aggggttccct ttagttaa gattaaccg ccatgctact tatctacgta gccatgctct 180
aggaagatcg gaattcgccc ttaagggcgc gccgtttaa tagctagcga cattgattat 240
tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagccca tatatggagt 300
tccgcgttac ataacttac gtaaatggcc cgctggctg accgccaac gacccccgcc 360
cattgacgtc aataatgacg tatgttcca tagtaacgcc aatagggact ttccattgac 420
gtcaatgggt ggagtattta cggtaaactg ccaactggc agtacatcaa gtgtatcata 480
tgccaagtac gccccctatt gacgtcaatg acggtaaag gcccgcctgg cattatgccc 540
agtacatgac ctatgggac tttcctactt ggcagtacat ctacgtatta gtcacgcta 600
ttaccatggt gatgcggttt tggcagtaca tcaatgggcg tggatagcgg tttgactcac 660
ggggatttcc aagtctccac ccattgacg tcaatgggag tttgttttg caccaaaatc 720
aacgggactt tcaaaaatgt cgtaacaact ccgccccatt gacgcaaatg ggcggtaggc 780
gtgtacggtg ggaggtctat ataagcagag ctggtaccac cgggtgccacc atgggcttcg 840
tgagacagat acagcttttg ctctggaaga actggaccct gcgaaaagg caaaagattc 900

gctttgtggt ggaactcgtg tggcctttat ctttatttct ggtcttgatc tggttaagga 960
atgccaaccc actctacagc catcatgaat gccatttccc caacaaggcg atgccctcag 1020
caggaatgct gccgtggctc caggggatct tctgcaatgt gaacaatccc tgttttcaaa 1080
gccccacccc aggagaatct cctggaattg tgtcaacta taacaactcc atcttggcaa 1140
gggtatatcg agattttcaa gaactcctca tgaatgcacc agagagccag caccttggcc 1200
gtatttggac agagctacac atcttgtccc aattcatgga caccctccgg actcaccggg 1260
agagaattgc aggtaagtgg cgaccgtgcc ccacagccct agccgccctc cccactgccc 1320
gggcttaccg tggcctgct cccgcaggaa gaggaatagc aataagggat atcttgaaag 1380
atgaagaaac actgacacta tttctcatta aaaacatcgg cctgtctgac tcagtggctt 1440
accttctgat caactctcaa gtccgtccag agcagttcgc tcatggagtc ccggacctgg 1500
cgctgaagga catcgctgc agcgaggccc tctggagcg ctteatcattc ttcagccaga 1560
gacgcggggc aaagacggtg cgtatgccc tgtgtcctc ctcccagggc accctacagt 1620
ggatagaaga cactctgtat gccaacgtgg acttcttcaa gctcttccgt gtgcttccca 1680
cactcctaga cagccgttct caaggtatca atctgagatc ttggggagga atattatctg 1740
atatgtcacc aagaattcaa gagtttatcc atcgcccgag tatgcaggac ttgctgtggg 1800
tgaccaggcc cctcatgcag aatggtggtc cagagacctt tacaagctg atgggcatcc 1860
tgtctgacct cctgtgtggc taccgccagg gaggtggctc tcgggtgctc tcttcaact 1920
ggtatgaaga caataactat aaggcctttc tggggattga ctccacaagg aaggatccta 1980
tctattctta tgacagaaga acaacatcct tttgtaatgc attgatccag agcctggagt 2040
caaatccttt aacaaaaatc gcttggaggc cggcaaagcc tttgctgatg ggaaaaatcc 2100
tgtacactcc tgattcacct gcagcacgaa ggatactgaa gaatgccaac tcaacttttg 2160
aagaactgga acacgttagg aagttgtca aagcctggga agaagtaggg cccagatct 2220
ggtacttctt tgacaacagc acacagatga acatgatcag agataccctg gggaacccaa 2280
cagtaaaaga ctttttgaat aggcagcttg gtgaagaagg tattactgct gaagccatcc 2340
taaacttctt ctacaagggc cctcgggaaa gccaggctga cgacatggcc aacttcgact 2400
ggagggacat atttaacatc actgatcgca ccctccgctt ggtcaatcaa tacctggagt 2460
gcttggctct ggataagttt gaaagctaca atgatgaaac tcagctcacc caacgtgccc 2520
tctctctact ggaggaaaac atgttctggg ccggagtggg attccctgac atgtatccct 2580
ggaccagctc tctaccacc cactggaagt ataagatccg aatggacata gacgtggtgg 2640
agaaaaacca taagattaaa gacaggtgag tggactggag cctgggcacg aggtgtgggg 2700
tgccccctgc cctgccactt acaccactg cctcttctg caggtattgg gattctggct 2760
ccagagctga tcccgtggaa gatttccggt acatctgggg cgggtttgcc tatctgcagg 2820
acatggttga acaggggatc acaaggagcc aggtgcagge ggaggctcca gttggaatct 2880
acctccagca gatgccctac cctgtcttcg tggacgatc tttcatgatc atctgaacc 2940
gctgtttccc tatcttcatg gtgtggcat ggatctactc tgtctccatg actgtgaaga 3000
gcacgtctt ggagaaggag ttgcactga aggagacctt gaaaaatcag ggtgtctcca 3060
atgcagtgat ttggtgtacc tggttctgga acagcttctc catcatgtcg atgagcatct 3120
tctcctgac gatattcatc atgcatggaa gaatcctaca ttacagcgac ccattcatcc 3180
tcttctggt cttgttggct ttctccactg ccaccatcat gctgtgcttt ctgctcagca 3240

ccttcttctc caaggccagt ctggcagcag cctgtagtgg tgtcatctat ttcaccctct 3300
 acctgccaca catcctgtgc ttgcctggc aggaccgcat gaccgctgag ctgaagaagg 3360
 ctgtgagctt actgtctccg gtggcatttg gatttggcac tgagtacctg gttcgctttg 3420
 aagagcaagg cctggggctg cagtggagca acatcgggaa cagtcccacg gaaggggacg 3480
 aattcagctt cctgctgtcc atgcagatga tgctccttga tgctgctgtc tatggcttac 3540
 tcgcttggtg ccttgatcag gtgtttccag gagactatgg aaccctactt ccttgggtact 3600
 ttcttctaca agagtcgtat tggcttggcg gtgaaggtaa gtagtctgat tatacacaag 3660
 atattgtcta gaacttgatg agactgtgga tatgaatatt tctactcttt ctcagggtgt 3720
 tcaaccagag aagaaagagc cctggaaaag accgagcccc taacagagga aacggaggat 3780
 ccagagcacc cagaaggaat acacgactcc ttctttgaac gtgagcatcc aggggtgggtt 3840
 cctggggatg gcgtgaagaa tcttgtaaag atttttgagc cctgtggccg gccagctgtg 3900
 gaccgtctga acatcacctt ctacgagaac cagatcacg cattcctggg ccacaatgga 3960
 gctgggaaaa ccaccacctt gtccatctg acgggtctgt tgccaccaac ctctgggact 4020
 gtgctcgttg ggggaaggga cattgaaacc agcctggatg cagtccggca gaccttggc 4080
 atgtgtccac agcacaacat cctgttccac cacctcacg tggctgagca catgctgttc 4140
 tatgcccagc tgaaggaaa gtcccaggag gaggcccagc tggagatgga agccatgttg 4200
 gaggacacag gcctccacca caagcggaat gaagaggctc aggacctatc aggtggcatg 4260
 cagagaaaagc tgtcggttgc cattgccttt gtgggagatg ccaaggtaag ggcactgagc 4320
 agaagggaaag aagctccggg ggctctttgt aggtaagct tagtcatca aggcgatcac 4380
 atcagtgaaa aaaagccaga caggcggta aaccaacga gattaaacag caggatgcaa 4440
 aaattcgcag gtggtcagat gaagcttatt gtcgaggcag tgaaaaaaat gctttatttg 4500
 tgaattttgt gatgctattg ctttatttgt aaccattata agctgcaata aacaagttct 4560
 cgagttaagg gcgaattccc gattaggatc ttctagagc atggctacgt agataagtag 4620
 catggcgggt taatcattaa ctacaaggaa ccctagtga tggagtggc cactccctct 4680
 ctgcgcgctc gctcgtcac tgaggccggg cgaccaaagg tcgcccagc cccgggcttt 4740
 gcccgggagg cctcagtgag cgagcgagcg cgcag 4775

<210> 40

<211> 5092

<212> DNA

<213> 人工序列

<220>

<223> 双AAV载体-3' ABCA4_w内含子

<400> 40

ctgcgcgctc gctcgtcac tgaggccgcc cgggcaaagc cgggctgctg ggcgaccttt 60
 ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gactggccaa ctccatcact 120
 aggggttctt tgtagttaat gattaaccgg ccatgctact tatctacgta gccatgctct 180
 aggaagatcg gaattcgccc ttaagggcgc gccgtttaa tagctagcga cattgattat 240
 tgactagtta ttaatagtaa tcaattacgg ggtcattagt tcatagcca tatatggagt 300
 tccgcgttac ataacttac gtaaattgcc cgctggctg accgccaac gacccccgcc 360

cattgacgtc aataatgacg tatgttccca tagtaacgcc aatagggact ttccattgac 420
gtcaatgggt ggagtattta cggtaaactg cccacttggc agtacatcaa gtgtatcata 480
tgccaagtac gccccctatt gacgtcaatg acggtaaatg gcccgcctgg cattatgccc 540
agtacatgac cttatgggac tttcctactt ggacgtacat ctacgtatta gtcatcgcta 600
ttaccatggt gatgcggttt tggcagtaca tcaatgggcg tggatagcgg tttgactcac 660
ggggatttcc aagtctccac cccattgacg tcaatgggag tttgttttgg caccaaatc 720
aacgggactt tccaaaatgt cgtaacaact ccgccccatt gacgcaaatg ggcggtaggc 780
gtgtacgggt ggaggtctat ataagcagag ctggtaccac cggtagaaag cttcatctga 840
ccacctgcga atttttgcat cctgctgttt aatctgcgtt ggtttaaccg cctgtctggc 900
tttttttcac tgatgtgac gcttgatgc actaagcttc aacgagtctt ttgtcatcta 960
caggtggtga ttctggacga acccactctt ggggtggacc cttactcgag acgctcaatc 1020
tgggatctgc tcctgaagta tcgctcaggc agaaccatca tcatgtccac tcaccacatg 1080
gacgaggccg acctccttgg ggaccgcatt gccatcattg cccagggaag gctctactgc 1140
tcaggcacc cactcttctt gaagaactgc tttggcacag gcttgtactt aaccttggtg 1200
cgcaagatga aaaacatcca gagccaaagg aaaggcagtg aggtgagtca ggctgggcgc 1260
ccccgcccc aggggcccct cctccccaaag cccccggac gcgcctcacc cacgttctct 1320
tcgcagggga cctgcagctg ctctgctaag gttttctcca ccacgtgtcc agcccacgtc 1380
gatgacctaa ctccagaaca agtcctggat ggggatgtaa atgagctgat ggatgtagtt 1440
ctccaccatg ttccagaggc aaagctggtg gactgcattg gtcaagaact tatcttctct 1500
cttccaaata agaacttcaa gcacagagca tatgccagcc ttttcagaga gctggaggag 1560
acgctggctg accttggctt cagcagtttt ggaatttctg aactcccctt ggaagagatt 1620
tttctgaagg tcacggagga ttctgattca ggacctctgt ttgcgggtgg cgctcagcag 1680
aaaagagaaa acgtcaacc cgcacacccc tgcttgggtc ccagagagaa ggctggacag 1740
acaccccagg actccaatgt ctgctcccca ggggcgcccg ctgctcacc agagggccag 1800
cctccccag agccagagtg cccaggcccg cagctcaaca cggggacaca gctggtcctc 1860
cagcatgtgc aggcgctgct ggtcaagaga ttccaacaca ccatccgag ccacaaggac 1920
ttcctggcgc agatcgtgct cccggctacc tttgtgtttt tggctctgat gctttctatt 1980
gttatccctc cttttggcga ataccccgtt ttgaccttc acccctggat atatgggcag 2040
cagtacacct tcttcagcat ggatgaacca ggcagtgagc agttcacggt acttgacagc 2100
gtcctcctga ataagccagg ctttggcaac cgctgcctga aggaagggtg gcttccggag 2160
taccctgtg gcaactcaac acctggaag actccttctg tgtccccaaa catcaccag 2220
ctgttccaga agcagaaatg gacacaggtc aaccttcaac catcctgcag gtgcagcacc 2280
agggagaagc tcaccatgct gccagagtgc cccgagggtg ccgggggcct cccgcccc 2340
cagagaacac agcgcagcac ggaaatteta caagacctga cggacaggaa catctccgac 2400
ttcttggtaa aaacgtatcc tgetcttata agaagcagct taaagagcaa attctgggtc 2460
aatgaacaga ggtatggagg aatttccatt ggaggaaagc tcccagtcgt ccccatcacg 2520
ggggaagcac ttgttgggtt ttaagcgac cttggccgga tcatgaatgt gagcgggggc 2580
cctatcacta gagaggcctc taaagaaata cctgatttcc ttaaactctt agaaactgaa 2640
gacaacatta aggtgtggtt taataacaaa ggctggcatg ccctggtcag ctttctcaat 2700

gtggcccaca acgccatcct gtaagtccta cttttttgt tcctttgaaa gcctcctgga 2760
 aagcttttcc tgaagtgttt gttctgtaat ttctttgcag acgggccagc ctgcctaagg 2820
 acaggagccc cgaggagtat ggaatcaccg tcattagcca acccctgaac ctgaccaagg 2880
 agcagctctc agagattaca gtgctgacca cttcagtgga tgctgtggtt gccatctgcg 2940
 tgattttctc catgtccttc gtcccagcca gctttgtcct ttatttgatc caggagcggg 3000
 tgaacaaatc caagcacctc cagtttatca gtggagttag ccccaccacc tactgggtga 3060
 ccaacttcct ctgggacatc atgaattatt ccgtgagtgc tgggctgggt gtgggcatct 3120
 tcatcgggtt tcagaagaaa gctacactt ctccagaaaa cttcctgcc cttgtggcac 3180
 tgctcctgct gtatggatgg gcggtcattc ccatgatgta cccagcatcc ttctgtttg 3240
 atgtccccag cacagcctat gtggctttat cttgtgctaa tctgttcacg gccatcaaca 3300
 gcagtgetat taccttcacg ttggaattat ttgagaataa ccggacgctg ctcaggttca 3360
 acgccgtgct gaggaagetg ctcatgtctt tccccactt ctgectgggc cggggcctca 3420
 ttgacctgac actgagccag gctgtgacag atgtctatgc ccggtttggt gaggagcact 3480
 ctgcaaatcc gttccactgg gacctgattg ggaagaacct gtttgccatg gtggtggaag 3540
 ggggtggtga cttcctcctg accctgctgg tccagcgcca cttcttctc tccaatgga 3600
 ttgccgagcc cactaaggag cccattgttg atgaagatga tgatgtggct gaagaaagac 3660
 aaagaattat tactggtgga aataaaaactg acatcttaag gctacatgaa ctaaccaaga 3720
 tttatccagg cacctccagc ccagcagtgg acaggctgtg tgtcggagtt cgcctggag 3780
 agtgctttgg cctcctggga gtgaatggtg ccggcaaac aaccacattc aagatgctca 3840
 ctggggacac cacagtgacc tcaggggatg ccaccgtagc aggcaagagt attttaacca 3900
 atatttctga agtccatcaa aatatgggct actgtcctca gtttgatgca attgatgagc 3960
 tgctcacagg acgagaacat ctttacctt atgcccggct tcgaggtgta ccagcagaag 4020
 aaatcgaaaa ggtaagtgat tcctagggct ggggaagggt ggtgggaatc ctctcctgct 4080
 cacctcctct ctctgcccc acaggttgca aactggagta ttaagagcct gggcctgact 4140
 gtctacgccg actgcctggc tggcacgtac agtgggggca acaagcggaa actctccaca 4200
 gccatcgcac tcattggctg cccaccgctg gtgctgctgg atgagcccac cacagggatg 4260
 gacccccagg cacgccgat gctgtggaac gtcatcgtga gcatcatcag agaagggagg 4320
 gctgtggtcc tcacatccca cagcatgga gaatgtgagg cactgtgtac ccgctggcc 4380
 atcatggtaa agggcgctt tcgatgtatg ggcaccattc agcatctcaa gtccaaattt 4440
 ggagatggct atatcgtcac aatgaagatc aaatccccga aggacgacct gcttctgac 4500
 ctgaaccctg tggagcagtt cttccagggg aacttcccag gcagtgtgca gagggagagg 4560
 cactacaaca tgetccagtt ccaggtctc tectctctc tggcgaggat cttccagctc 4620
 ctctctctcc acaaggacag cctgctcacc gaggagtact cagtcacaca gaccacactg 4680
 gaccaggtgt ttgtaaattt tgetaaacag cagactgaaa gtcatgacct ccctctgcac 4740
 cctcgagctg ctggagccag tcgacaagcc caggactgag tcgacgcggc cgcgcagtga 4800
 aaaaaatgct ttatttgtga aatgtgtgat gctattgctt tttttgtaac cattataagc 4860
 tgcaataaac aagtctcga gtttaaggcg aattcccgat taggatctc ctagagcatg 4920
 gctacgtaga taagtagcat ggcgggttaa tcattaacta caaggaacc ctagtgatgg 4980
 agttggccac tccctctctg cgcgctcgtc cgctcactga ggccgggcga ccaaggtcg 5040

cccgacgcc gggctttgcc cgggcggcct cagtgagcga gcgagcgcgc ag 5092

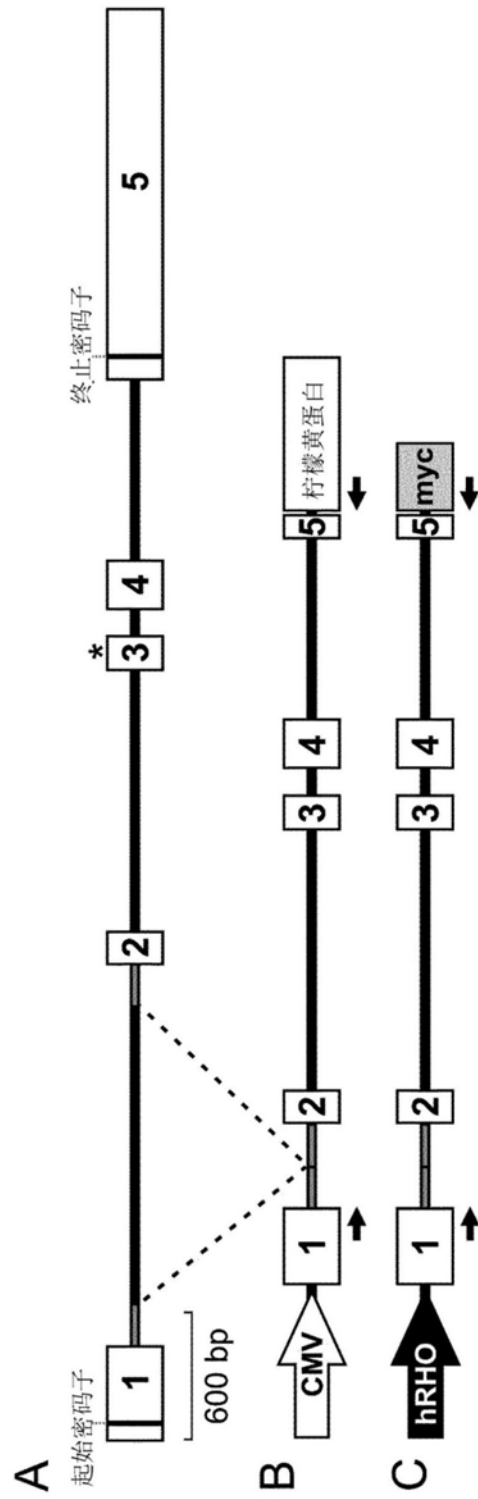


图1

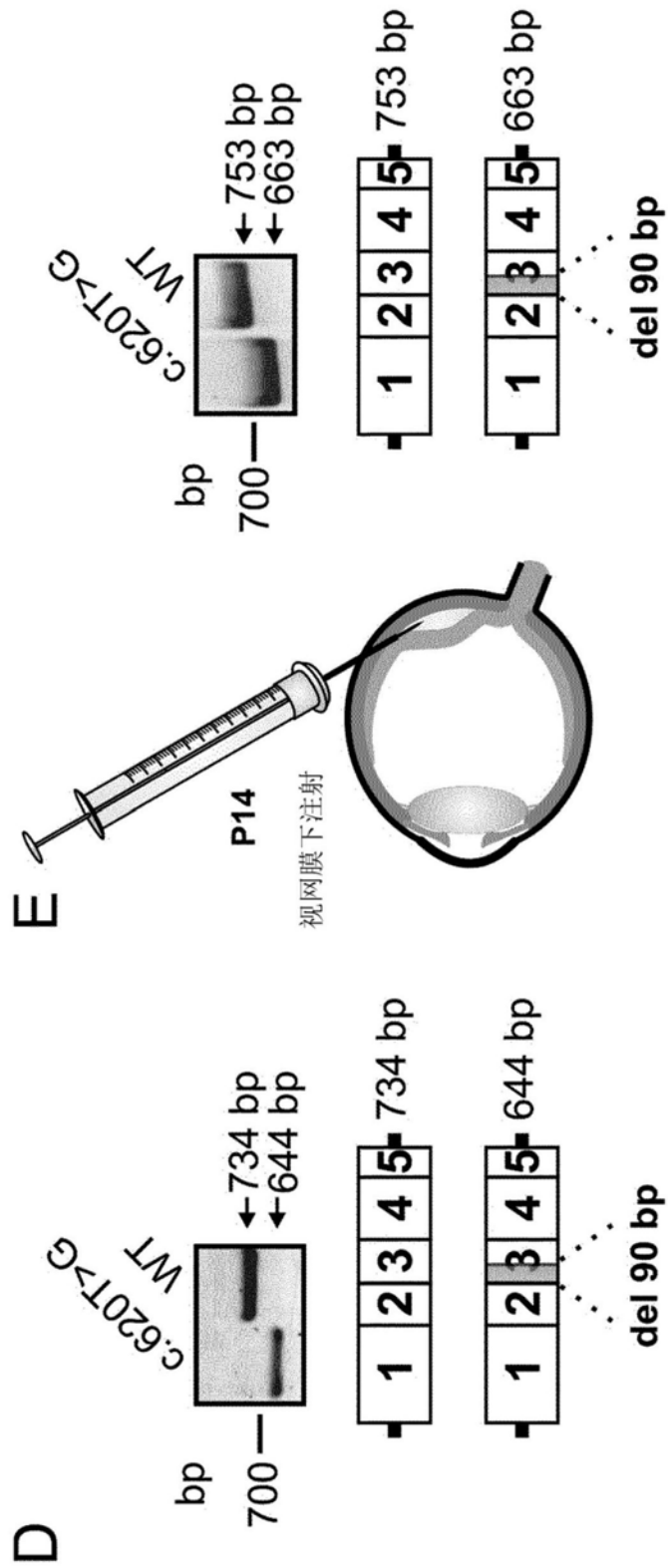


图1(续)

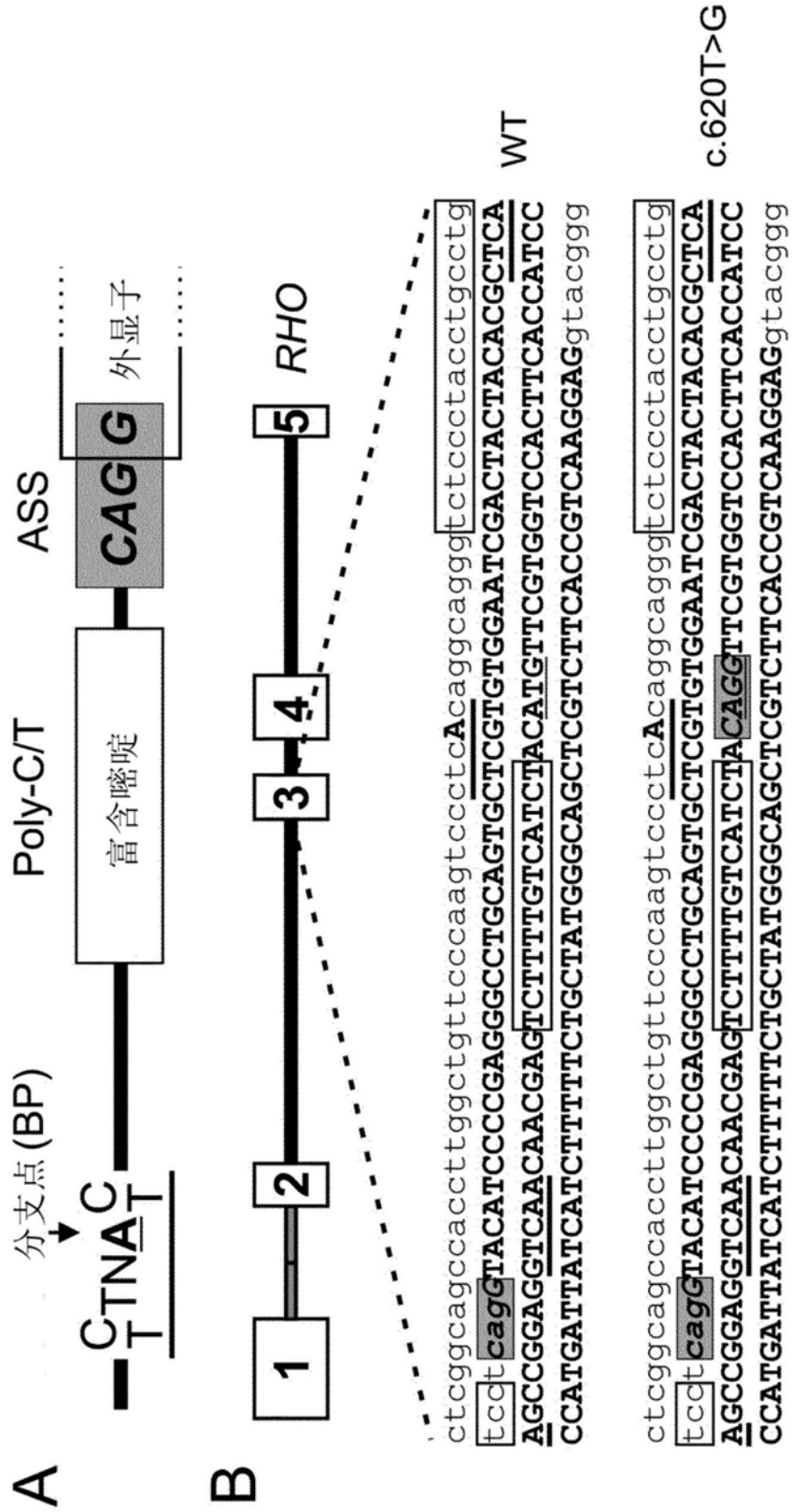


图2

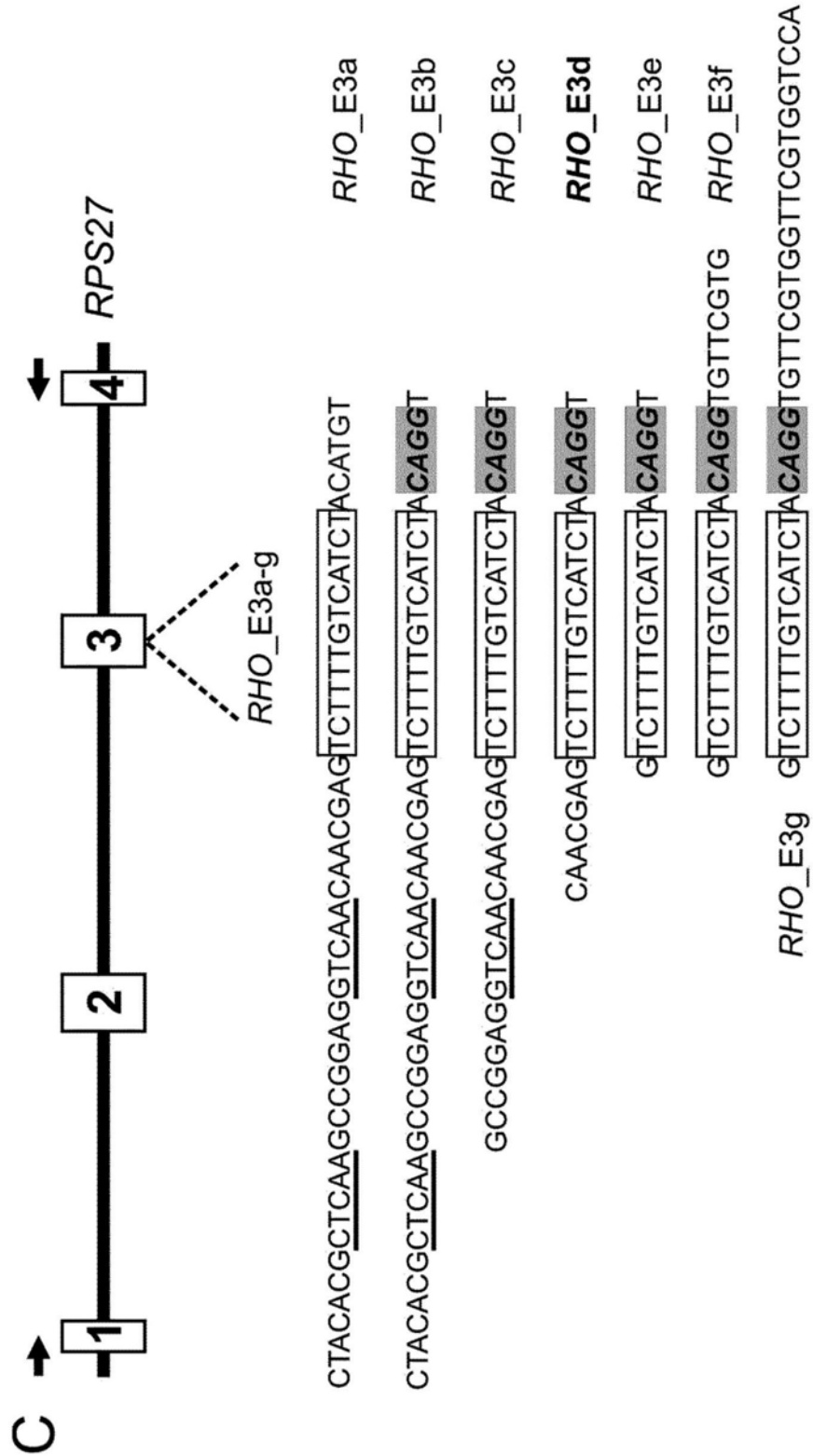


图2 (续)

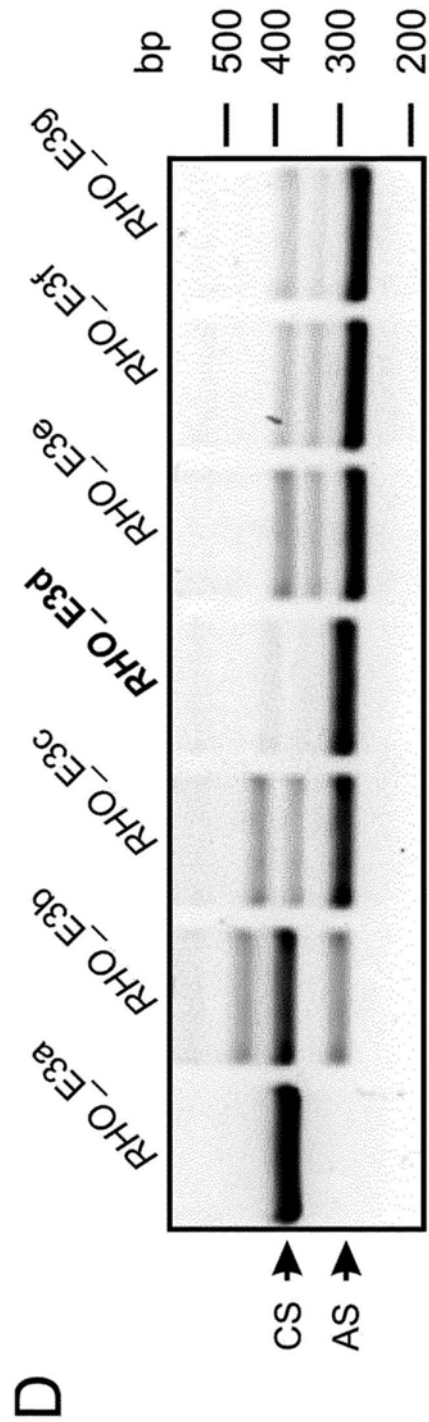


图2(续)

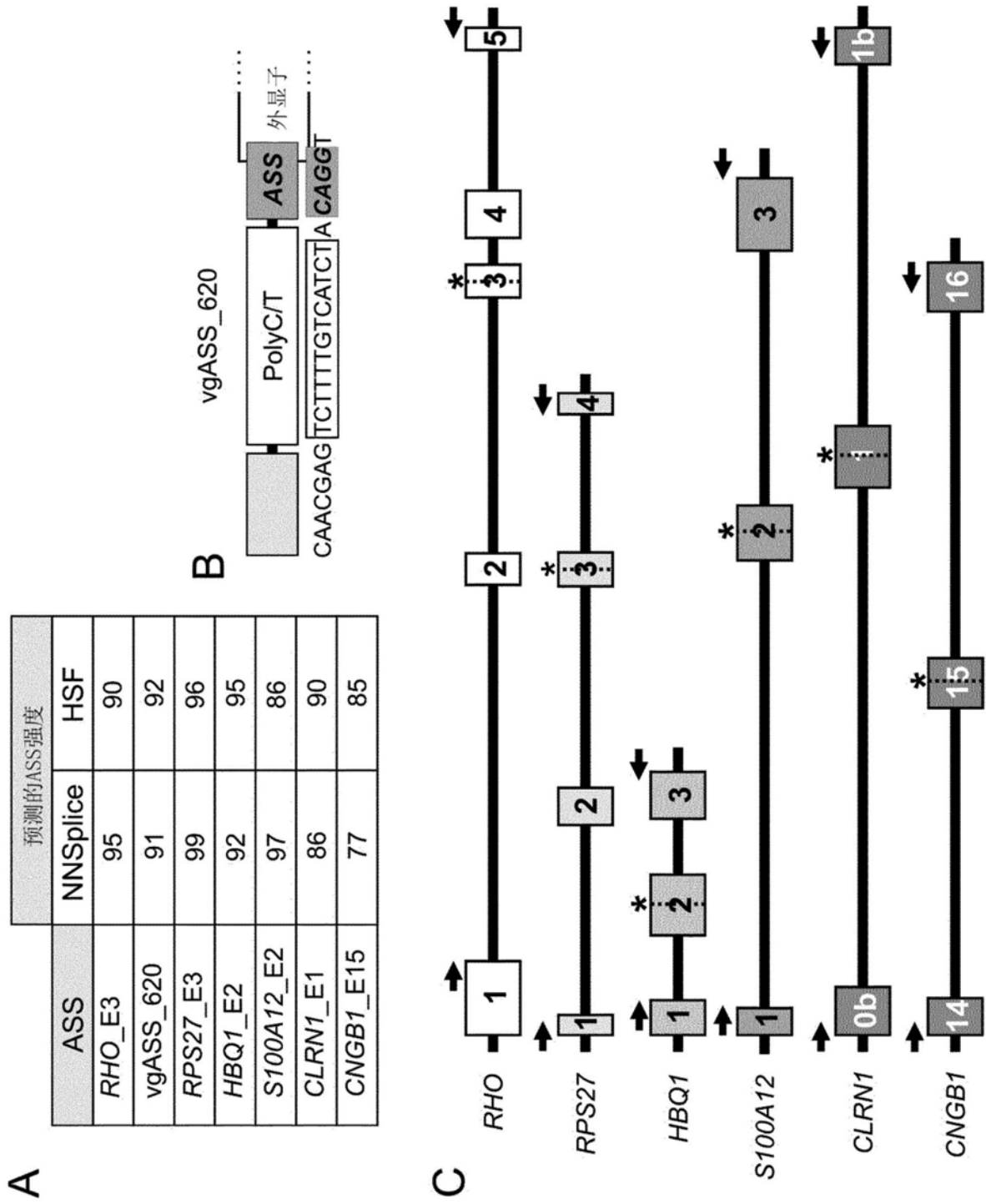


图3

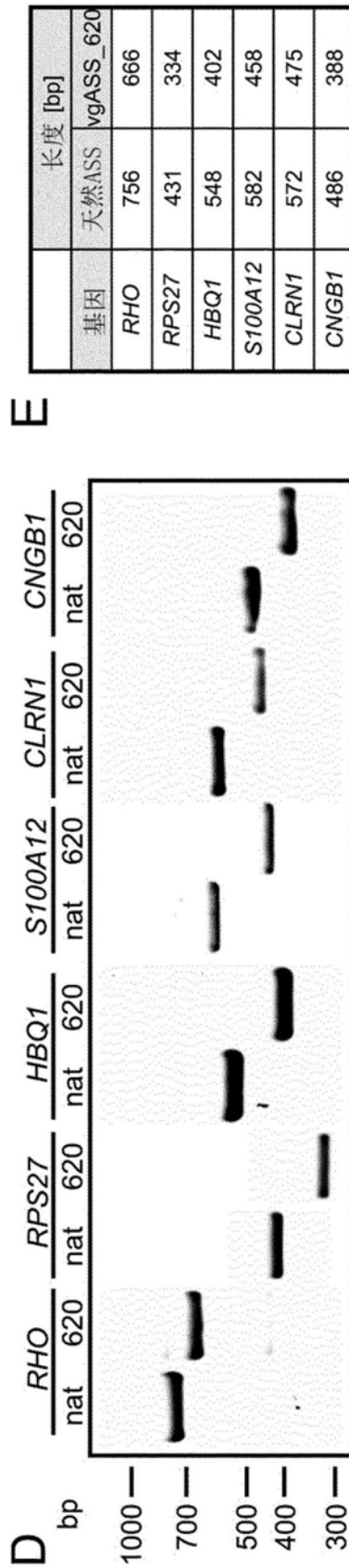


图3(续)

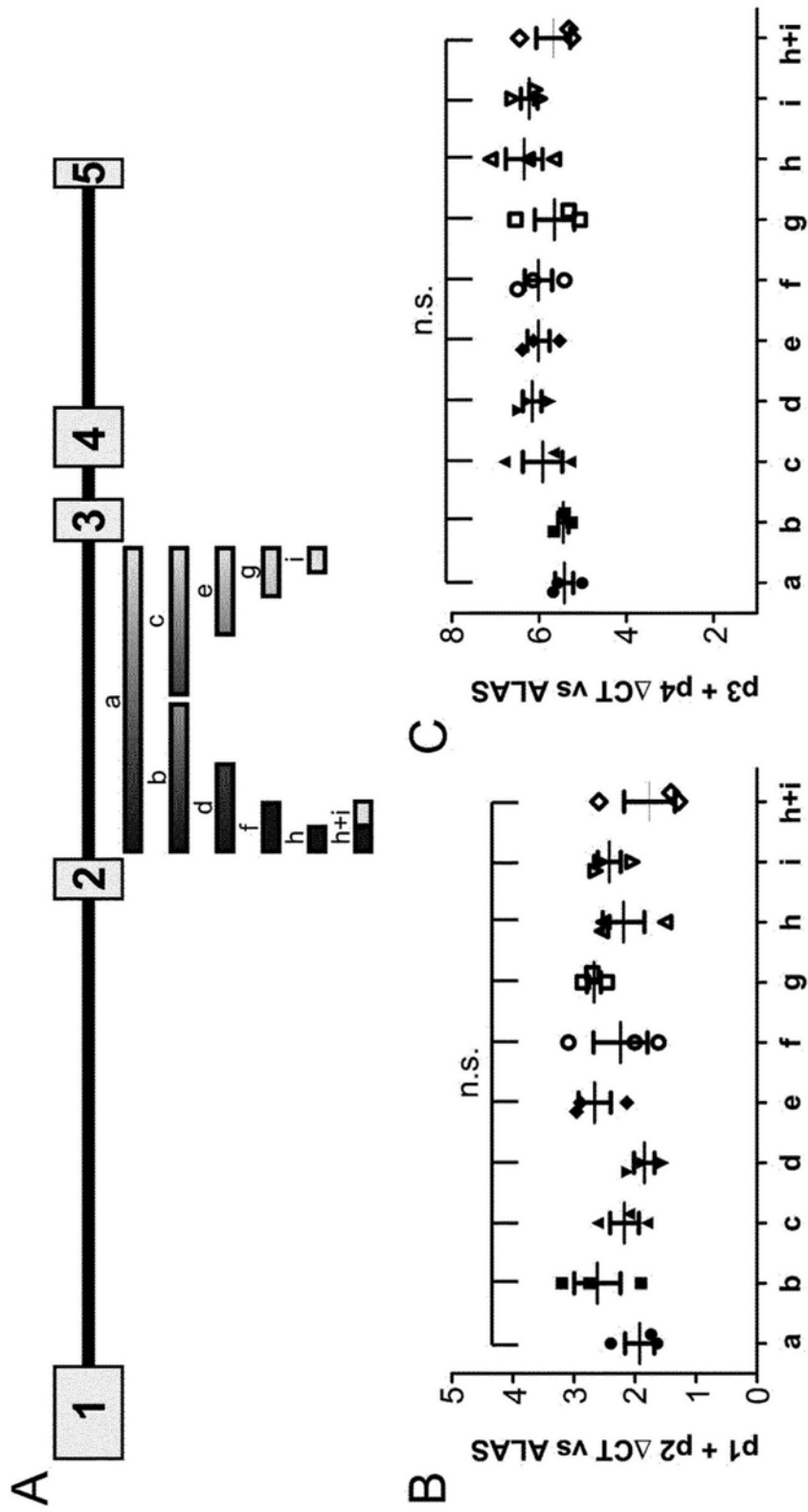


图4

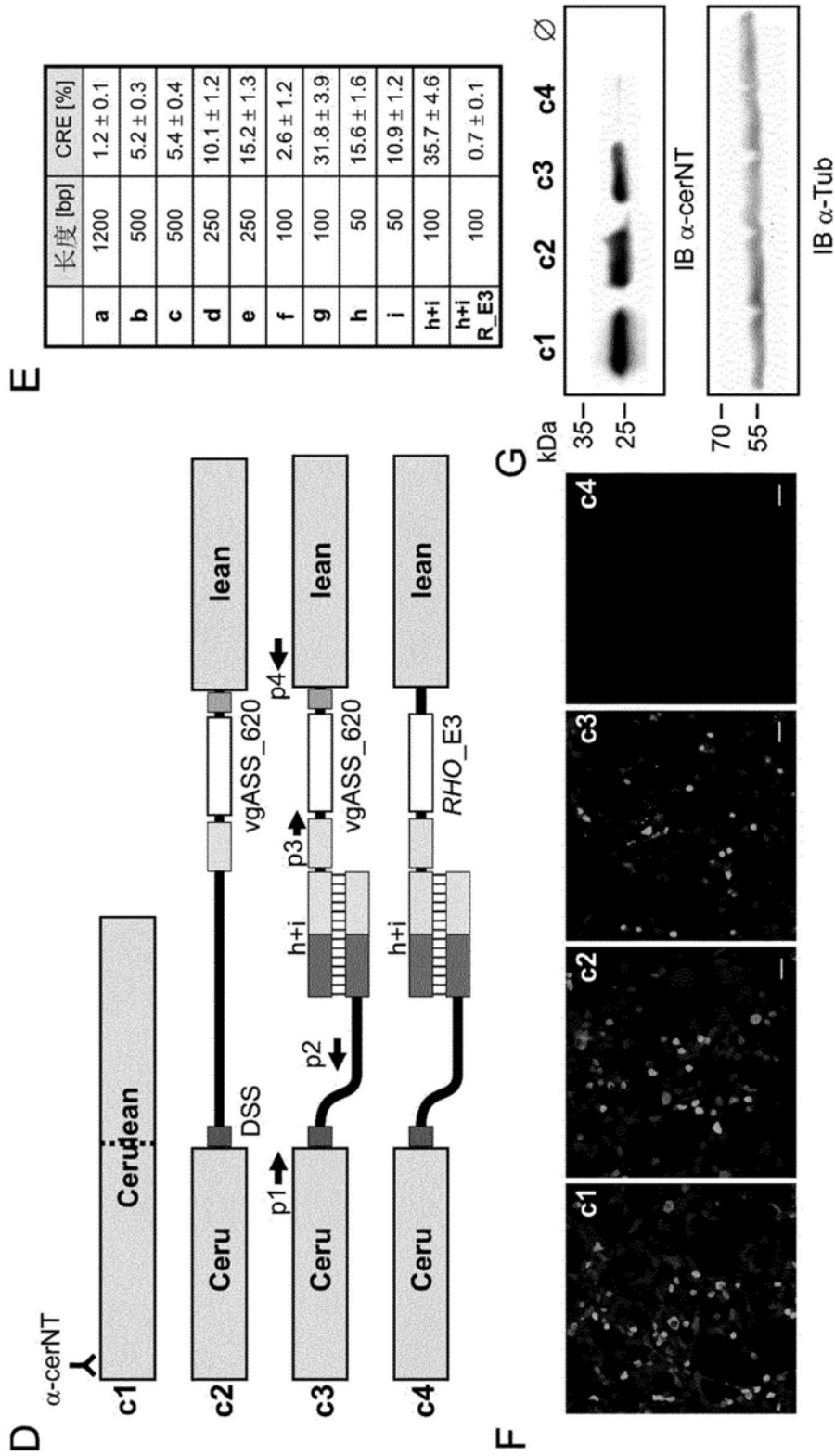


图4(续)

包含ABCA4的5' 编码序列的双AAV

(SEQ ID NO: 19):

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCCGGGCAAAGCCCCGGGCGTCGGGCGACCTTTGGTTCGCCC
 GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTAGGGGTTCCCTtgtagt
 taatgattaaccgcatgctacttatctacgtagccatgctctaggaagatcggaattcgcccttaa
 gggcgcgccgttttaaatagctagcGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGG
 TCATTAGTTCATAGCCCATATATGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCCGCTGGCTG
 ACCGCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA
 CTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
 CATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTA
 CATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA
 TCGGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTGACTCACGGGGATTTCCAAGTCTCCAC
 CCCATTGACGTCAATGGGAGTTTGTFTTGGCACAAAATCAACGGGACTTTCCAAAATGTCGTAACAA
 CTCGCCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTggt
 ccaccggtgccaccATGGGCTTTCGTGAGACAGATACAGCTTTTGCTCTGGAAGAATGGACCCTGCGG
 AAAAGGCAAAGATTTCGCTTTGTGGTGGAACTCGTGTGGCCTTTATCTTTATTTCTGGTCTTGATCTG
 GTTAAGGAATGCCAACCCACTCTACAGCCATCATGAATGCCATTTCCCAACAAGGCGATGCCCTCAG
 CAGGAATGCTGCCGTGGCTCCAGGGGATCTTCTGCAATGTGAACAATCCCTGTTTTCAAAGCCCCACC
 CCAGGAGAATCTCCTGGAATTGTGTCAAACATAACAACCTCCATCTTGGCAAGGGTATATCGAGATTT
 TCAAGAATCCTCATGAATGCACCAGAGAGCCAGCACCTTGGCCGTATTTGGACAGAGCTACACATCT
 TGTCCCAATTCATGGACACCCTCCGGACTCACCCGGAGAGAATTGCAGGAAGAGGAATACGAATAAGG
 GATATCTTGAAGATGAAGAAACACTGACACTATTTCTCATTTAAAACATCGGCCTGTCTGACTCAGT
 GGTCTACCTTCTGATCAACTCTCAAGTCCGTCCAGAGCAGTTCGCTCATGGAGTCCCGGACCTGGCGC
 TGAAGGACATCGCCTGCAGCGAGGCCCTCCTGGAGCGCTTCATCATCTTCAGCCAGAGACGCGGGGCA
 AAGACGGTGCCTATGCCCTGTGCTCCCTCTCCAGGGCACCCCTACAGTGGATAGAAGACTCTGTA
 TGCCAACGTGGACTTCTTCAAGCTCTTCCGTGTGCTTCCCACACTCCTAGACAGCCGTTCTCAAGGTA
 TCAATCTGAGATCTTGGGGAGGAATATTATCTGATATGTCACCAAGAATTCAAGAGTTTATCCATCGG
 CCGAGTATGCAGGACTTGTGTGGGTGACCAGGCCCTCATGCAGAATGGTGGTCCAGAGACCTTTAC
 AAAGCTGATGGGCATCCTGTCTGACCTCCTGTGTGGCTACCCCGAGGGAGGTGGCTCTCGGGTGTCT
 CCTTCAACTGGTATGAAGACAATAACTATAAGGCCTTTCTGGGGATTGACTCCACAAGGAAGGATCCT
 ATCTATTCTTATGACAGAAGAACAACATCCTTTTGTAAATGCATTGATCCAGAGCCTGGAGTCAAATCC
 TTTAACCAAAATCGCTTGGAGGGCGGCAAAGCCTTTGCTGATGGGAAAAATCCTGTACACTCCTGATT
 CACCTGCAGCACGAAGGATACTGAAGAATGCCAACTCAACTTTTGAAGAATGGAACACGTTAGGAAG
 TTGGTCAAAGCCTGGGAAGAAGTAGGGCCCCAGATCTGGTACTTCTTTGACAACAGCACACAGATGAA
 CATGATCAGAGATACCCTGGGGAACCCAACAGTAAAAGACTTTTTGAATAGGCAGCTTGGTGAAGAAG
 GTATTACTGCTGAAGCCATCTAACTTCTCTACAAGGGCCCTCGGGAAAGCCAGGCTGACGACATG
 GCCAACTTCGACTGGAGGGACATATTTAACATCACTGATCGCACCCCTCCGCCTGGTCAATCAATACCT
 GGAGTGTCTGGTCCCTGGATAAGTTTGAAGCTACAATGATGAAACTCAGCTCACCCAACGTGCCCTCT
 CTCTACTGGAGGAAAACATGTTCTGGGCGGAGTGGTATCCCTGACATGTATCCCTGGACCAGCTCT
 CTACCACCCACGTTGAAGTATAAGATCCGAATGGACATAGACGTGGTGGAGAAAACCAATAAGATTAA

图5

AGACAGGTATTGGGATTCTGGTCCCAGAGCTGATCCCGTGGAAGATTTCCGGTACATCTGGGGCGGGT
 TTGCCTATCTGCAGGACATGGTTGAACAGGGGATCACAAGGAGCCAGGTGCAGGCGGAGGCTCCAGTT
 GGAATCTACCTCCAGCAGATGCCCTACCCCTGCTTCGTGGACGATTCTTTCATGATCATCCTGAACCG
 CTGTTTTCCCTATCTTCATGGTGTGGCATGGATCTACTCTGTCTCCATGACTGTGAAGAGCATCGTCT
 TGGAGAAGGAGTTGCGACTGAAGGAGACCTTGAAAAATCAGGGTGTCTCCAATGCAGTGATTTGGTGT
 ACCTGGTTCCCTGGACAGCTTCTCCATCATGTGCGATGAGCATCTTCCTCCTGACGATATTCATCATGCA
 TGGAAGAATCCTACATTACAGCGACCCATTCATCCTCTTCTGTTCTTGTGGCTTTCTCCACTGCCA
 CCATCATGCTGTGCTTTCTGCTCAGCACCTTCTTCTCCAAGGCCAGTCTGGCAGCAGCCTGTAGTGGT
 GTCATCTATTTACCCCTCTACCTGCCACACATCCTGTGCTTCGCCTGGCAGGACCGCATGACCGCTGA
 GCTGAAGAAGGCTGTGAGCTTACTGTCTCCGGTGGCATTGGATTGGCACTGAGTACCTGGTTCGCT
 TTGAAGAGCAAGGCCTGGGGCTGCAGTGGAGCAACATCGGGAACAGTCCCACGGAAGGGGACGAATTC
 AGCTTCCTGCTGTCCATGCAGATGATGCTCCTTGATGCTGCTGTCTATGGCTTACTCGCTTGGTACCT
 TGATCAGGTGTTTCCAGGAGACTATGGAACCCCACTTCTTGGTACTTTCTTCTACAAGAGTCGTATT
 GGCTTGGCGGTGAAGGGTGTTC AACAGAGAAGAAAGAGCCCTGGAAAAGACCGAGCCCTAACAGAG
 GAAACGGAGGATCCAGAGCACCCAGAAGGAATACACGACTCCTTCTTTGAACGTGAGCATCCAGGGTG
 GGTTCCTGGGGTATGCGTGAAGAATCTGGTAAAGATTTTGGAGCCCTGTGGCCGGCCAGCTGTGGACC
 GTCTGAACATCACCTTCTACGAGAACCAGATCACCGCATTCCTGGGCCACAATGGAGCTGGGAAAACC
 ACCACCTTGTCCATCCTGACGGGTCTGTTGCCACCAACCTCTGGGACTGTGCTCGTTGGGGGAAGGGA
 CATTGAAACCAGCCTGGATGCAGTCCGGCAGAGCCTTGGCATGTGTCCACAGCACAACATCCTGTTCC
 ACCACCTCACGGTGGCTGAGCACATGCTGTTCTATGCCAGCTGAAAGGAAAGTCCCAGGAGGAGGCC
 CAGCTGGAGATGGAAGCCATGTTGGAGGACACAGGCCTCCACCACAAGCGGAATGAAGAGGCTCAGGA
 CCTATCAGGTGGCATGCAGAGAAAGCTGTCCGGTTGCCATTGCCTTTGTGGGAGATGCCAAG*gtaaggg*
 cactgagcagaaggyaagaagctccgggggctctttgtagggtaagcttAGTGCATCAAGGCGATCAC
 ATCAGTGAAAAAAGCCAGACAGGCGGTTAAACCAACGCAGATTAACAGCAGGATGCAAAAATTCGC
 AGGTGGTCAGATG*aa*gcttattctcgagttaagggcgaattcccgattaggatcttcctagagcatgg
 ctacgtagataagtagcatggcgggttaatcattaactacaAGGAACCCCTAGTGATGGAGTTGGCCA
 CTCCTCTCTGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTT
 GCCCGGGCGGCCCTCAGTGAGCGAGCGAGCGCGCAG

图5 (续)

包含ABCA4的3' 编码序列的双AAV

(SEQ ID NO:20):

```

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTAGGGGTTCCCTtgtagt
taatgattaaccgcgcatgctacttatctacgtagccatgctctaggaagatcggaattcgcccttaa
ggcgcgccgctttaaatagctagcGACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGG
TCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCTGGCTG
ACCGCCCAACGACCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA
CTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
CATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCCAGTA
CATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGA
TGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTACTCACGGGGATTTCCAAGTCTCCAC
CCCATTGACGTCAATGGGAGTTTTGTTTTGGCACAAAATCAACGGGACTTTCAAAATGTCGTAACAA
CTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTggta
ccaccggtagaaagcttCATCTGACCACCTGCGAATTTTTGCATCCTGCTGTTTAATCTGCGTTGGTT
TAACCGCCTGTCTGGCTTTTTTTTCACTGATGTGATCGCCTTGATGCACTaagcttCAACGAGTCTTTT
GTCATCTACAGGTGGTGATTCTGGACGAACCCACCTCTGGGGTGGACCCTTACTCGAGACGCTCAATC
TGGGATCTGCTCCTGAAGTATCGCTCAGGCAGAACCATCATCATGTCCACTCACCACATGGACGAGGC
CGACCTCCTTGGGGACCGCATTGCCATCATTGCCAGGGAAGGCTCTACTGCTCAGGCACCCCACTCT
TCCTGAAGAACTGCTTTGGCACAGGCTTGTACTTAACCTTGGTGCGCAAGATGAAAAACATCCAGAGC
CAAAGGAAAGGCAGTGAGGGGACCTGCAGCTGCTCGTCTAAGGTTTTCTCCACCACGTGTCCAGCCCA
CGTCGATGACCTAACTCCAGAACAAGTCCGGATGGGGATGTAAATGAGCTGATGGATGTAGTCTCC
ACCATGTTCCAGAGGCAAAGCTGGTGGAGTGCATTGGTCAAGAACTTATCTTCCTTCTTCCAAAATAAG
AACTTCAAGCACAGAGCATATGCCAGCCTTTTTAGAGAGCTGGAGGAGACGCTGGCTGACCTTGGTCT
CAGCAGTTTTTGAATTTCTGACACTCCCCTGGAAGAGATTTTTCTGAAGGTCACGGAGGATTCTGATT
CAGGACCTCTGTTTGGGGTGGCGCTCAGCAGAAAAGAGAAAACGTCAACCCCGACACCCCTGCTTG
GGTCCCAGAGAGAAGGCTGGACAGACACCCAGGACTCCAATGTCTGCTCCCAGGGGCGCCGGCTGC
TCACCCAGAGGGCCAGCCTCCCCAGAGCCAGAGTGCCAGGCCCGCAGCTCAACACGGGGACACAGC
TGGTCCTCCAGCATGTGCAGGCGCTGCTGGTCAAGAGATTCCAACACACCATCCGCAGCCACAAGGAC
TTCTGGCGCAGATCGTGCTCCCGGCTACCTTTGTGTTTTGGCTCTGATGCTTTCTATTGTTATCCC
TCCTTTTGGCGAATACCCCGCTTTGACCCTTACCCCTGGATATATGGGCAGCAGTACACCTTCTTCA
GCATGGATGAACCAGGCAGTGAGCAGTTCACGGTACTTGCAGACGTCCTCCTGAATAAGCCAGGCTTT
GGCAACCGCTGCCTGAAGGAAGGGTGGCTTCCGGAGTACCCCTGTGGCAACTCAACACCCTGGAAGAC
TCCTTCTGTGTCCCCAAACATCACCCAGCTGTTCCAGAAGCAGAAATGGACACAGGTCAACCCCTTCAC
CATCCTGCAGGTGCAGCACCAGGGAGAAGCTCACCATGCTGCCAGAGTGCCCCGAGGGTGCCGGGGGC
CTCCCGCCCCCAGAGAACACAGCGCAGCACGGAATTTCTACAAGACCTGACGGACAGGAACATCTC
CGACTTCTTGGTAAAAACGTATCCTGCTCTTATAAGAAGCAGCTTAAAGAGCAAATTTCTGGGTCAATG
AACAGAGGTATGGAGGAATTTCCATTGGAGGAAAGCTCCAGTCGTCCCCATCACGGGGGAAGCACTT

```

图6

GTTGGGTTTTTAAGCGACCTTGGCCGGATCATGAATGTGAGCGGGGCCCTATCACTAGAGAGGCCTC
TAAAGAAATACCTGATTTCCTTAAACATCTAGAAACTGAAGACAACATTAAGGTGTGGTTTTAATAACA
AAGGCTGGCATGCCCTGGTCAGCTTCTCAATGTGGCCACAACGCCATCTTACGGGCCAGCCTGCCT
AAGGACAGGAGCCCCGAGGAGTATGGAATCACCGTCATTAGCCAACCCCTGAACCTGACCAAGGAGCA
GCTCTCAGAGATTACAGTGCTGACCACTTCACTGGATGCTGTGGTTGCCATCTGCGTGATTTTCTCCA
TGTCCTTCGTCCCAGCCAGCTTTGTCCTTTATTTGATCCAGGAGCGGGTGAACAAATCCAAGCACCTC
CAGTTTATCAGTGGAGTGAGCCCCACCACCTACTGGGTGACCAACTTCCTCTGGGACATCATGAATTA
TTCCGTGAGTGCTGGGCTGGTGGTGGGCATCTTCATCGGGTTTTCAGAAGAAAGCCTACACTTCTCCAG
AAAACCTTCCTGCCCTTGTGGCACTGCTCCTGCTGTATGGATGGGCGGTTCATCCCATGATGTACCCA
GCATCCTTCCTGTTTGTATGTCGCCAGCACAGCCTATGTGGCTTTATCTTGTGCTAATCTGTTTCATCGG
CATCAACAGCAGTGCTATTACCTTCATCTTGAATTATTTGAGAATAACCGGACGCTGCTCAGGTTCA
ACGCCGTGCTGAGGAAGCTGCTCATTGTCTTCCCCACTTCTGCCTGGGCCGGGGCCTCATTGACCTT
GCACTGAGCCAGGCTGTGACAGATGTCTATGCCCGGTTTGGTGAGGAGCACTCTGCAAATCCGTTCCA
CTGGGACCTGATTGGGAAGAACCTGTTTGGCATGGTGGTGGGAAGGGTGGTGTACTTCTCCTGACCC
TGCTGGTCCAGCGCCACTTCTTCTCTCCCAATGGATTGCCGAGCCCACTAAGGAGCCCATTGTTGAT
GAAGATGATGATGTGGCTGAAGAAAGACAAAGAATTATTACTGGTGGAAATAAAACTGACATCTTAAG
GCTACATGAACAAACCAAGATTTATCCAGGCACCTCCAGCCCAGCAGTGGACAGGCTGTGTGTCGGAG
TTCGCCCTGGAGAGTGCTTTGGCCTCCTGGGAGTGAATGGTGCCGGCAAAACAACCACATTCAAGATG
CTCACTGGGGACACCACAGTGACCTCAGGGGATGCCACCGTAGCAGGCAAGAGTATTTTAACCAATAT
TTCTGAAGTCCATCAAAAATATGGGCTACTGTCCTCAGTTTGTATGCAATTGATGAGCTGCTCACAGGAC
GAGAACATCTTTACCTTTATGCCCCGGCTTCAGAGTGTACCAGCAGAAGAAATCGAAAAGGTTGCAAAC
TGGAGTATTAAGAGCCTGGGCCTGACTGTCTACGCCGACTGCCTGGCTGGCACGTACAGTGGGGGCAA
CAAGCGGAAACTCTCCACAGCCATCGCACTCATTGGCTGCCACCGCTGGTGCTGCTGGATGAGCCCA
CCACAGGGATGGACCCCCAGGCACGCCGATGCTGTGGAACGTCATCGTGAGCATCATCAGAGAAGGG
AGGGCTGTGGTCTCACATCCCACAGCATGGAAGAATGTGAGGCACTGTGTACCCGGCTGGCCATCAT
GGTAAAGGGCGCCTTTCGATGTATGGGCACCACTCAGCATCTCAAGTCCAAATTTGGAGATGGCTATA
TCGTCACAATGAAGATCAAAATCCCCGAAGGACGACCTGCTTCTGACCTGAACCCTGTGGAGCAGTTC
TTCCAGGGGAACCTTCCAGGCAGTGTGCAGAGGGAGAGGCACTACAACATGCTCCAGTTCAGGTCTC
CTCCTCCTCCCTGGCGAGGATCTTCCAGCTCCTCCTCTCCACAAGGACAGCCTGCTCATCGAGGAGT
ACTCAGTCACACAGACCACACTGGACCAGGTGTTTGTAAATTTTGTAAACAGCAGACTGAAAGTCAT
GACCTCCCTCTGCACCCTCGAGCTGCTGGAGCCAGTCGACAAGCCCAGGACTGAgtcgagcgggcgcg
GCAGTGAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCT
GCAATAAACAAAGTTctcgagttaagggcgaattcccgattaggatcttctctagagcatggctacgtag
ataagtagcatggcgggttaataactacaAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCT
CTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCCCCGACGCCGGGCTTTGCCCGGGC
GGCCTCAGTGAGCGAGCGAGCGCGAG

图6 (续)

包含ABCA4的5' 编码序列的双AAV

(SEQ ID NO: 21):

```

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGGTTCGCCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTAGGGGTTCCTtgtagt
taatgattaaccgcatgctacttatctacgtagccatgctctaggaagatcggaattcgccttaa
gggcgcgccgtttaaatagctagcAAAAGACTAAAAGAGGGAGGGATCACTTCAGATCTGCCGAGTGA
GTCGATTGGACTTAAAGGGCCAGTCAAACCCTGACTGCCGGCTCATGGCAGGCTCTTGCCGAGGACAA
ATGCCCAGCCTATATTTATGCAAAGAGATTTTGTTCCAAACCTTAAGGTCAAAGATACCTAAAGACATC
CCCCTCAGGAACCCCTCTCATGGAGGAGAGTGCCTGAGGGTCTTGGTTTCCCATTGCATCCCCACCT
CAATTTCCCTGGTGCCCAGCCACTTGTGTCTTTAGGGTTCTCTTTCTCTCCATAAAAAGGGAGCCAACA
CAGTGTTCGGCCTCCTCTCCCAACTAAGGGCTTATGTGTAATTA AAAAGGGATTATGCTTTGAAGGGGA
AAAGTAGCCTTTAATCACCAGGAGAAGGACACAGCGTCCGGAGCCAGAGGCGCTCTTAACGGCGTTTA
TGTCCTTTGCTGTCTGAGGGGCCCTCAGCTCTGACCAATCTGGTCTTCGTGTGGTCATTAGCGGTACCg
gtaccaccggtgccaccATGGGCTTCGTGAGACAGATACAGCTTTTGCTCTGGAAGAACTGGACCCTG
CGGAAAAGGCAAAAAGATTGCTTTGTGGTGGAACTCGTGTGGCCTTTATCTTTATTTCTGGTCTTGAT
CTGGTTAAGGAATGCCAACCCACTCTACAGCCATCATGAATGCCATTTCCCAACAAGGCGATGCCCT
CAGCAGGAATGCTGCCGTGGCTCCAGGGGATCTTCTGCAATGTGAACAATCCCTGTTTTCAAAGCCCC
ACCCAGGAGAATCTCCTGGAATTGTGTCAAACCTATAACAACCTCCATCTTGGCAAGGGTATATCGAGA
TTTTCAAGAACTCCTCATGAATGCACCAGAGAGCCAGCACCTTGGCCGTATTTGGACAGAGCTACACA
TCTTGTCCCAATTCATGGACACCCTCCGGACTCACCCGGAGAGAATTGCAGGAAGAGGAATACGAATA
AGGGATATCTTGAAAGATGAAGAAACACTGACACTATTTCTCATTAAAAACATCGGCCTGTCTGACTC
AGTGGTCTACCTTCTGATCAACTCTCAAGTCCGTCCAGAGCAGTTTCGCTCATGGAGTCCCGGACCTGG
CGCTGAAGGACATCGCCTGCAGCGAGGCCCTCCTGGAGCGCTTCATCATCTTCAGCCAGAGACGCGGG
GCAAAGACGGTGCCTATGCCCTGTGCTCCCTCTCCAGGGCACCTACAGTGGATAGAAGACACTCT
GTATGCCAACGTGGACTTCTTCAAGCTCTTCCGTGTGCTTCCCACACTCCTAGACAGCCGTTCTCAAG
GTATCAATCTGAGATCTTGGGGAGGAATATTATCTGATATGTCACCAAGAATTCAAGAGTTTATCCAT
CGGCCGAGTATGCAGGACTTGCTGTGGGTGACCAGGCCCTCATGCAGAATGGTGGTCCAGAGACCTT
TACAAAGCTGATGGGCATCCTGTCTGACCTCCTGTGTGGCTACCCCGAGGGAGGTGGCTCTCGGGTGC
TCTCCTTCAACTGGTATGAAGACAATAACTATAAGGCCTTTCTGGGGATTGACTCCACAAGGAAGGAT
CCTATCTATTCTTATGACAGAAGAACAACATCCTTTTGTAAATGCATTGATCCAGAGCCTGGAGTCAA
TCCTTTAACCAAAAATCGCTTGGAGGGCGGCAAAGCCTTTGCTGATGGGAAAAATCCTGTACACTCCTG
ATTCACCTGCAGCACGAAGGATACTGAAGAATGCCAACTCAACTTTTGAAGAACTGGAACACGTTAGG
AAGTTGGTCAAAGCCTGGGAAGAAGTAGGGCCCAGATCTGGTACTTCTTTGACAACAGCACACAGAT
GAACATGATCAGAGATACCCTGGGGAACCAACAGTAAAAGACTTTTTGAATAGGCAGCTTGGTGAAG
AAGGTATTACTGCTGAAGCCATCCTAAACTTCCTCTACAAGGGCCCTCGGGAAAGCCAGGCTGACGAC
ATGGCCAACCTTCGACTGGAGGGACATATTTAACATCACTGATCGCACCCCTCCGCCTGGTCAATCAATA
CCTGGAGTGCTTGGTCCCTGGATAAGTTTGAAAGCTACAATGATGAAACTCAGCTCACCCAACGTGCC
TCTCTCTACTGGAGGAAAACATGTTCTGGGCCGGAGTGGTATTCCCTGACATGTATCCCTGGACCAGC
TCTCTACCACCCACGTGAAGTATAAGATCCGAATGGACATAGACGTGGTGGAGAAAACCAATAAGAT
TAAAGACAGGTATTGGGATTCTGGTCCCAGAGCTGATCCCGTGGAAGATTTCCGGTACATCTGGGGCG

```

图7

GGTTTGCCTATCTGCAGGACATGGTTGAACAGGGGATCACAAGGAGCCAGGTGCAGGCCGAGGCTCCA
GTTGGAATCTACCTCCAGCAGATGCCCTACCCCTGCTTCGTGGACGATTCTTTCATGATCATCCTGAA
CCGCTGTTCCCTATCTTCATGGTGTGGCATGGATCTACTCTGTCTCCATGACTGTGAAGAGCATCG
TCTTGGAGAAGGAGTTGCGACTGAAGGAGACCTTGAAAAATCAGGGTGTCTCCAATGCAGTGATTTGG
TGTACCTGGTTCCCTGGACAGCTTCTCCATCATGTGCGATGAGCATCTTCTCCTGACGATATTCATCAT
GCATGGAAGAATCCTACATTACAGCGACCCATTTCATCCTCTTCTGTTCTTGTGGCTTTCTCCACTG
CCACCATCATGCTGTGCTTTCTGCTCAGCACCTTCTTCTCCAAGGCCAGTCTGGCAGCAGCCTGTAGT
GGTGTCACTATTTTACCCTCTACCTGCCACACATCCTGTGCTTCGCCTGGCAGGACCCGCATGACCCG
TGAGCTGAAGAAGGCTGTGAGCTTACTGTCTCCGGTGGCATTGGATTGGCACTGAGTACCTGGTTC
GCTTTGAAGAGCAAGGCCTGGGGCTGCAGTGGAGCAACATCGGGAACAGTCCCACGGAAGGGGACGAA
TTCAGCTTCTGCTGTCCATGCAGATGATGCTCCTTGATGCTGCTGTCTATGGCTTACTCGCTTGGTA
CCTTGATCAGGTGTTTCCAGGAGACTATGGAACCCCACTTCCTTGGTACTTTCTTCTACAAGAGTCGT
ATTGGCTTGGCGGTGAAGGGTGTTCACCAGAGAAGAAAGAGCCCTGGAAAAGACCCGAGCCCCTAACA
GAGGAAACGGAGGATCCAGAGCACCCAGAAGGAATACACGACTCCTTCTTTGAACGTGAGCATCCAGG
GTGGGTTCCTGGGGTATGCGTGAAGAATCTGGTAAAGATTTTTGAGCCCTGTGGCCGGCCAGCTGTGG
ACCGTCTGAACATCACCTTCTACGAGAACCAGATCACCGCATTCCTGGGCCACAATGGAGCTGGGAAA
ACCACCACCTTGTCCATCCTGACGGGTCTGTTGCCACCAACCTCTGGGACTGTGCTCGTTGGGGGAAG
GGACATTGAAACCAGCCTGGATGCAGTCCGGCAGAGCCTTGGCATGTGTCCACAGCACAAATCCTGT
TCCACCACCTCACGGTGGCTGAGCACATGCTGTTCTATGCCAGCTGAAAGGAAAGTCCCAGGAGGAG
GCCAGCTGGAGATGGAAGCCATGTTGGAGGACACAGGCCTCCACCACAAGCGGAATGAAGAGGCTCA
GGACCTATCAGGTGGCATGCAGAGAAAGCTGTCGGTTGCCATTGCCTTGTGGGAGATGCCAAGgtaa
gggactgagcagaaggaagaagctccgggggctctttgtagggttaagcttAGTGCATCAAGGCGAT
CACATCAGTGAAAAAAGCCAGACAGGCGGTTAAACCAACGCAGATTAACAGCAGGATGCAAAAATT
CGCAGGTGGTCAGATGaagcttattctcgagttaagggcgaattcccgattaggatcttcctagagca
tggctacgtagataagtagcatggcgggttaatcattaactacaAGGAACCCCTAGTGATGGAGTTGG
CCTACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGC
TTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCAG

图7(续)

包含ABCA4的3' 编码序列的双AAV

(SEQ ID NO: 22):

```

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCGGGCAAAGCCCGGGCGTTCGGGCGACCTTTGGTTCGCC
GGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACATCCATCACTAGGGGTTCCCTtgtagt
taatgattaacccgccatgctacttatctacgtagccatgctctaggaagatcggaaattcgcccttaa
gggcgcgcgcgttttaaatagctagcAAAAGACTAAAAGAGGGAGGGATCACTTCAGATCTGCCGAGTGA
GTTCGATTGGACTTAAAGGGCCAGTCAAACCCCTGACTGCCGGCTCATGGCAGGCTCTTGCCGAGGACAA
ATGCCAGCCTATATTTATGCAAAGAGATTTTGTTCAAACTTAAGGTCAAAGATACCTAAAGACATC
CCCCTCAGGAACCCCTCTCATGGAGGAGAGTGCCTGAGGGTCTTGGTTTCCCATTCATCCCCACCT
CAATTTCCCTGGTGCCAGCCACTTGTGTCTTTAGGGTTCTCTTTCTCTCCATAAAAGGGAGCCAACA
CAGTGTCGGCCTCCTCTCCCAACTAAGGGCTTATGTGTAATTAAGGGATTATGCTTTGAAGGGGA
AAAGTAGCCTTTAATCACCAGGAGAAGGACACAGCGTCCGGAGCCAGAGGCGCTCTAACGGCGTTTA
TGTCCCTTGCTGTCTGAGGGGCTCAGCTCTGACCAATCTGGTCTTCGTGTGGTCATTAGCGGTACCg
gtaccaccggtagaaagcttCATCTGACCACCTGCGAATTTTGCATCCTGCTGTTAATCTGCGTTG
GTTTAACCGCCTGTCTGGCTTTTTTTCACTGATGTGATCGCCTTGATGCACTaagcttCAACGAGTCT
TTTGTCATCTACAGGTGGTGATTCTGGACGAACCCACCTCTGGGGTGGACCCTTACTCGAGACGCTCA
ATCTGGGATCTGCTCCTGAAGTATCGCTCAGGCAGAACCATCATCATGTCCACTCACCACATGGACGA
GGCCGACCTCCTTGGGGACCGCATTGCCATCATTGCCAGGGAAGGCTCTACTGCTCAGGCACCCAC
TCTTCCTGAAGAACTGCTTTGGCACAGGCTTGTACTTAACCTTGGTGCGCAAGATGAAAAACATCCAG
AGCCAAAGGAAAGGCAGTGAGGGGACCTGCAGCTGCTCGTCTAAGGGTTTCTCCACCACGTGTCCAGC
CCACGTCGATGACCTAACTCCAGAACAAGTCTGGATGGGGATGTAATGAGCTGATGGATGTAGTTC
TCCACCATGTTCCAGAGGCAAAGCTGGTGGAGTGCATTGGTCAAGAACTTATCTTCCTTCTTCAAAT
AAGAACTTCAAGCACAGAGCATATGCCAGCCTTTTTCAGAGAGCTGGAGGAGACGCTGGCTGACCTTGG
TCTCAGCAGTTTTTGAATTTCTGACACTCCCCTGGAAGAGATTTTTTCTGAAGGTCACGGAGGATTCTG
ATTCAGGACCTCTGTTTGGGGTGGCGCTCAGCAGAAAAGAGAAAACGTCAACCCCCGACACCCCTGC
TTGGGTCCCAGAGAGAAGGCTGGACAGACACCCAGGACTCCAATGTCTGCTCCCAGGGGCGCCGGC
TGCTCACCCAGAGGGCCAGCCTCCCCAGAGCCAGAGTGCCAGGCCCGCAGCTCAACACGGGGACAC
AGCTGGTCCCTCCAGCATGTGCAGGCGCTGCTGGTCAAGAGATTCCAACACACCATCCGCAGCCACAAG
GACTTCCTGGCGCAGATCGTGCTCCCAGCTACCTTTGTGTTTTTGGCTCTGATGCTTTCTATTGTTAT
CCCTCCTTTTGGCGAATACCCCGCTTTGACCCTTACCCCTGGATATATGGGCAGCAGTACACCTTCT
TCAGCATGGATGAACCAGGCAGTGAGCAGTTCACGGTACTTGACAGACGTCCCTCCTGAATAAGCCAGGC
TTTGGCAACCGCTGCCTGAAGGAAGGGTGGCTTCCGGAGTACCCCTGTGGCAACTCAACACCCCTGGAA
GACTCCTTCTGTGTCCCAACATCACCCAGCTGTTCCAGAAGCAGAAATGGACACAGGTCAACCCCTT
CACCATCCTGCAGGTGCAGCACCAGGGAGAAGCTCACCATGCTGCCAGAGTGCCCCGAGGGTGCCTGGG
GGCCTCCCGCCCCCCCCAGAGAACACAGCGCAGCACGGAAATTCACAAGACCTGACGGACAGGAACAT
CTCCGACTTCTTGGTAAAAACGTATCCTGCTCTTATAAGAAGCAGCTTAAAGAGCAAATTTCTGGGTCA
ATGAACAGAGGTATGGAGGAATTTCCATTGGAGGAAAGCTCCAGTCGTCCCATCACGGGGGAAGCA
CTTGTGGGTTTTTAAGCGACCTTGGCCGGATCATGAATGTGAGCGGGGGCCCTATCACTAGAGAGGC

```

图8

CTCTAAAGAAATACCTGATTTCCCTTAAACATCTAGAAACTGAAGACAACATTAAGGTGTGGTTAATA
 ACAAAGGCTGGCATGCCCTGGTCAGCTTTCTCAATGTGGCCCACAACGCCATCTTACGGGCCAGCCTG
 CCTAAGGACAGGAGCCCCGAGGAGTATGGAATCACCGTCATTAGCCAACCCCTGAACCTGACCAAGGA
 GCAGCTCTCAGAGATTACAGTGCTGACCCTTCAGTGGATGCTGTGGTTGCCATCTGCGTGATTTTCT
 CCATGTCCTTCGTCCCAGCCAGCTTTGTCCTTTATTTGATCCAGGAGCGGGTGAACAAATCCAAGCAC
 CTCCAGTTTATCAGTGGAGTGAGCCCCACCACCTACTGGGTGACCAACTTCCTCTGGGACATCATGAA
 TTATTCCGTGAGTGCTGGGCTGGTGGTGGGCATCTTCATCGGGTTTCAGAAGAAAGCCTACACTTCTC
 CAGAAAACCTTCCTGCCCTTGTGGCACTGCTCCTGCTGTATGGATGGGCGGTCAATCCCATGATGTAC
 CCAGCATCCTTCCTGTTTGATGTCCCCAGCACAGCCTATGTGGCTTTATCTTGTGCTAATCTGTTTCAT
 CGGCATCAACAGCAGTGCTATTACCTTCATCTTGGAAATTATTTGAGAATAACCGGACGCTGCTCAGGT
 TCAACGCCGTGCTGAGGAAGCTGCTCATTGTCTTCCCCACTTCTGCCTGGGCGGGGGCCTCATTGAC
 CTTGCACTGAGCCAGGCTGTGACAGATGTCTATGCCCGTTTGGTGAGGAGCACTCTGCAAATCCGTT
 CCACTGGGACCTGATTGGGAAGAACCTGTTTGCCATGGTGGTGGAAAGGGTGGTGTACTTCTCCTGA
 CCCTGCTGGTCCAGCGCCACTTCTTCTCTCCAATGGATTGCCGAGCCCCTAAGGAGCCCATTGTT
 GATGAAGATGATGATGTGGCTGAAGAAAGACAAAGAATTATTACTGGTGGAAATAAAACTGACATCTT
 AAGGCTACATGAAC TAACCAAGATTTATCCAGGCACCTCCAGCCAGCAGTGGACAGGCTGTGTGTCG
 GAGTTCGCCCTGGAGAGTGCTTTGGCCTCCTGGGAGTGAATGGTGCCGGCAAACAACCATTTCAAG
 ATGCTCACTGGGGACACCACAGTGACCTCAGGGGATGCCACCGTAGCAGGCAAGAGTATTTTAACCAA
 TATTTCTGAAGTCCATCAAAATATGGGCTACTGTCTCAGTTTGATGCAATTGATGAGCTGCTCACAG
 GACGAGAACATCTTTACCTTTATGCCCGCTTCGAGGTGTACCAGCAGAAGAAATCGAAAAGGTTGCA
 AACTGGAGTATTAAGAGCCTGGGCCTGACTGTCTACGCCGACTGCCTGGCTGGCACGTACAGTGGGGG
 CAACAAGCGGAAACTCTCCACAGCCATCGCACTCATTGGCTGCCACCGCTGGTGCTGCTGGATGAGC
 CCACCACAGGGATGGACCCCCAGGCACGCCGCATGCTGTGGAACGTCATCGTGAGCATCATCAGAGAA
 GGGAGGGCTGTGGTCTCACATCCCACAGCATGGAAGAATGTGAGGCACTGTGTACCCGGCTGGCCAT
 CATGGTAAAGGGCGCCTTTCGATGTATGGGCACCATTACAGCATCTCAAGTCCAAATTTGGAGATGGCT
 ATATCGTCACAATGAAGATCAAAATCCCCGAAGGACGACCTGCTTCCTGACCTGAACCCTGTGGAGCAG
 TTCTTCCAGGGGAACTTCCAGGCAGTGTGCAGAGGGAGAGGCACTACAACATGCTCCAGTTCCAGGT
 CTCTCTCCTCCCTGGCGAGGATCTTCCAGCTCCTCCTCTCCACAAGGACAGCCTGCTCATCGAGG
 AGTACTCAGTCACACAGACCACACTGGACCAGGTGTTTGTAAATTTTGCTAAACAGCAGACTGAAAGT
 CATGACCTCCCTCTGCACCCTCGAGCTGCTGGAGCCAGTCGACAAGCCCAGGACTGAgtcgacgcggc
 CGCGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAA
 GCTGCAATAACAAGTctcgagttaagggcgaattcccgattaggatcttcctagagcatggctacg
 tagataagtagcatggcgggtaatacattaactacaAGGAACCCCTAGTGATGGAGTTGGCCACTCCC
 TCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAAGGTGCCCCGACGCCGGGCTTTGCCCG
 GGCGCCTCAGTGAGCGAGCGAGCGCGCAG

图8(续)

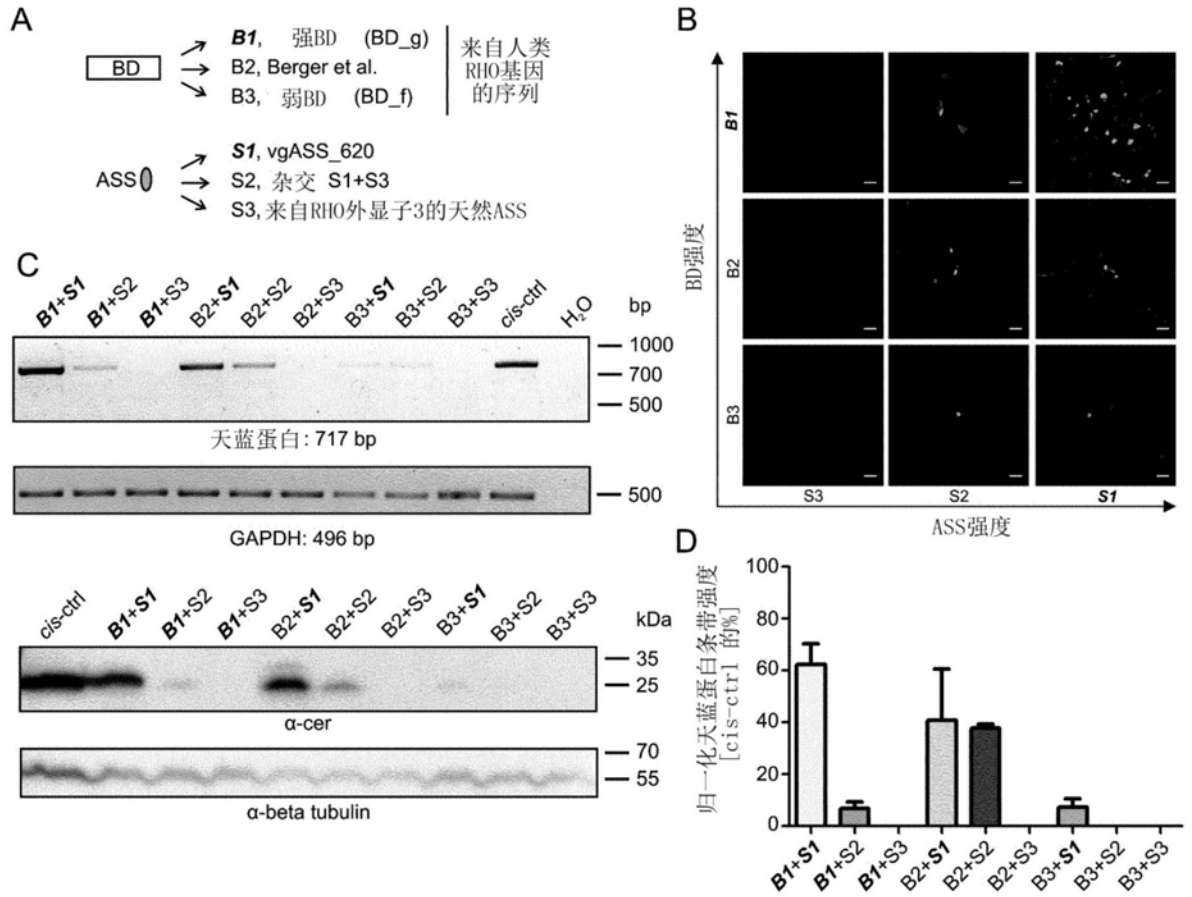
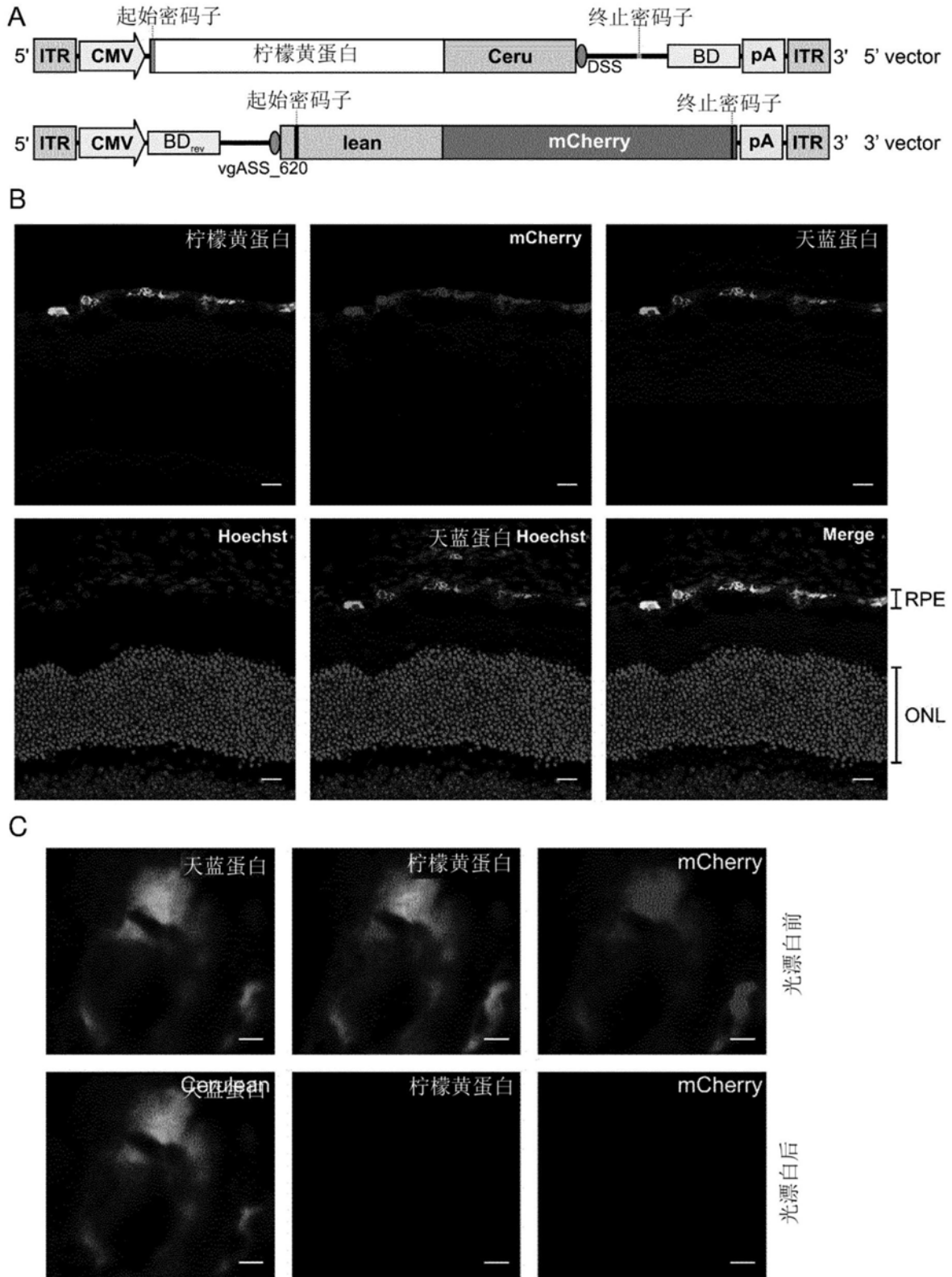


图9



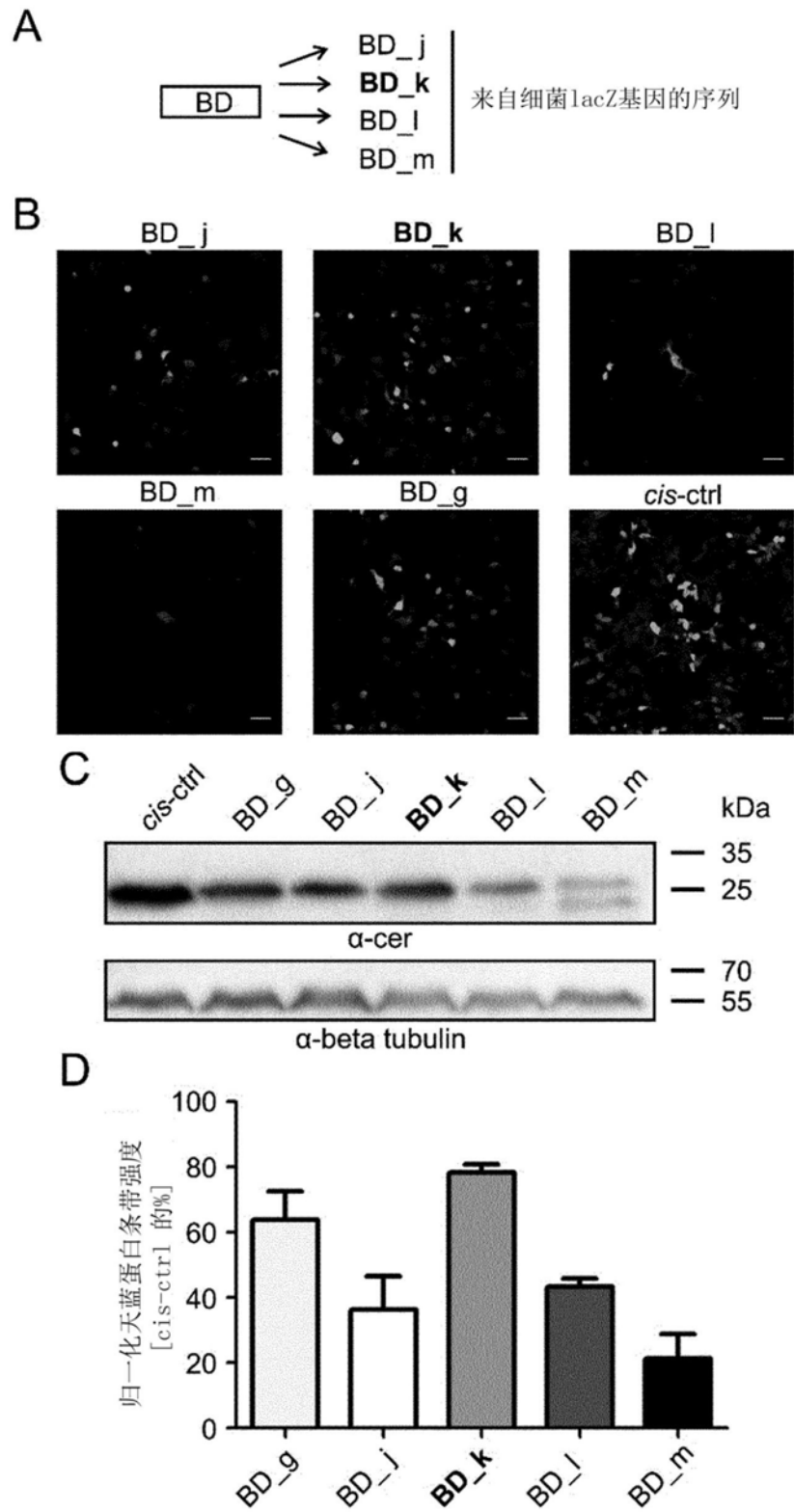


图11

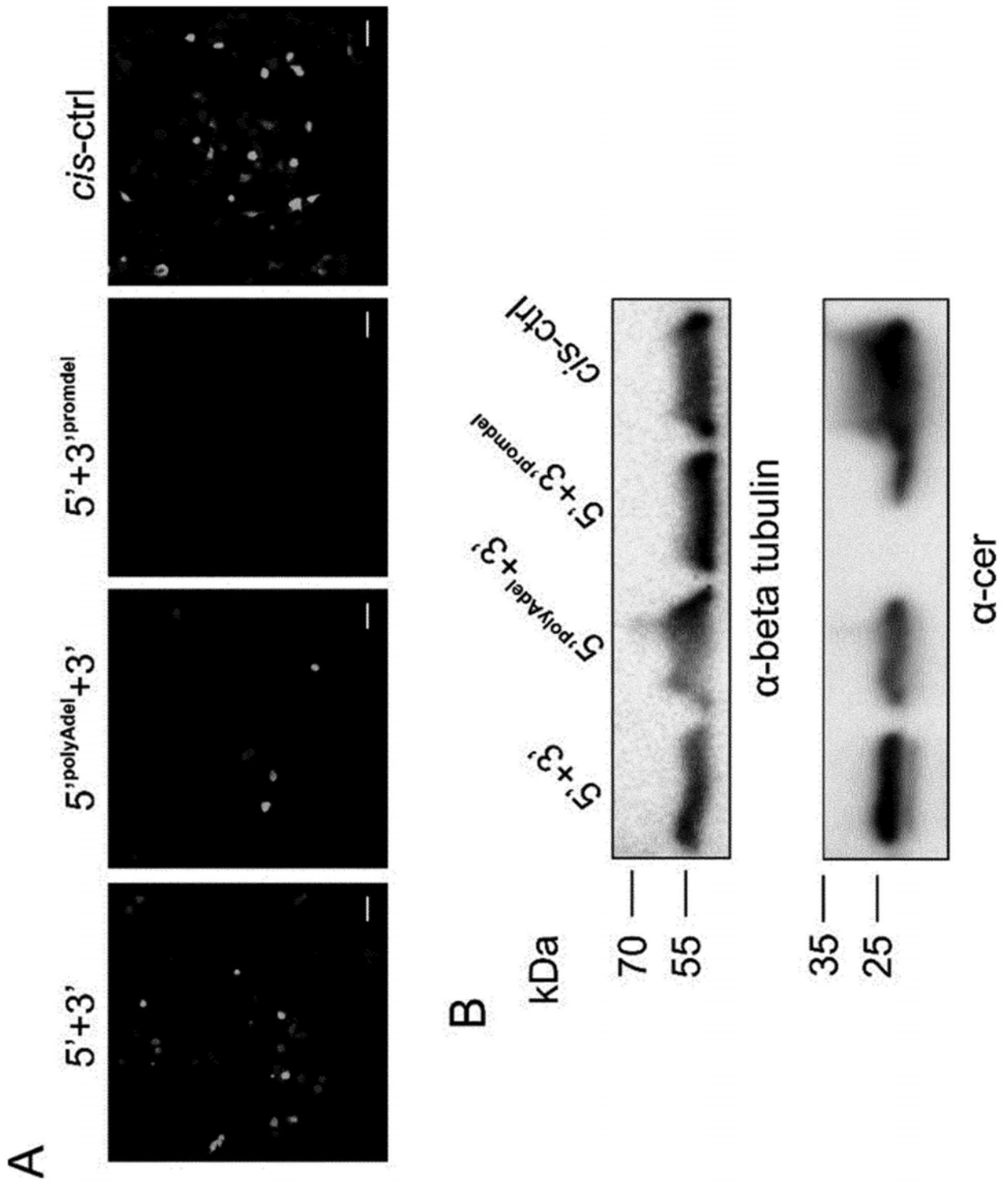


图12

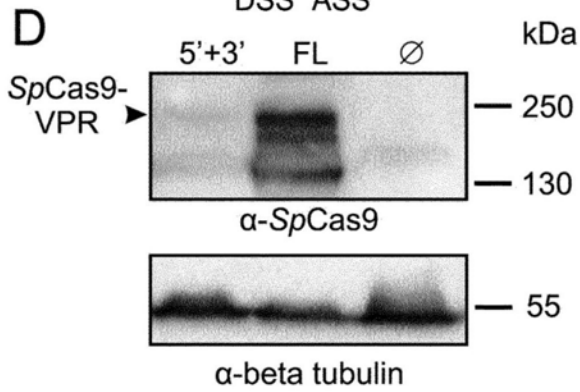
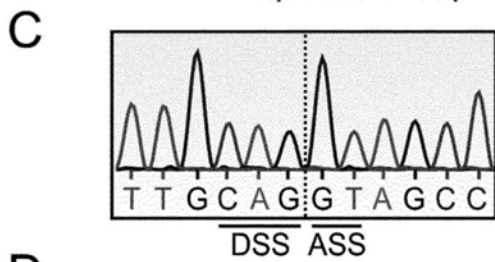
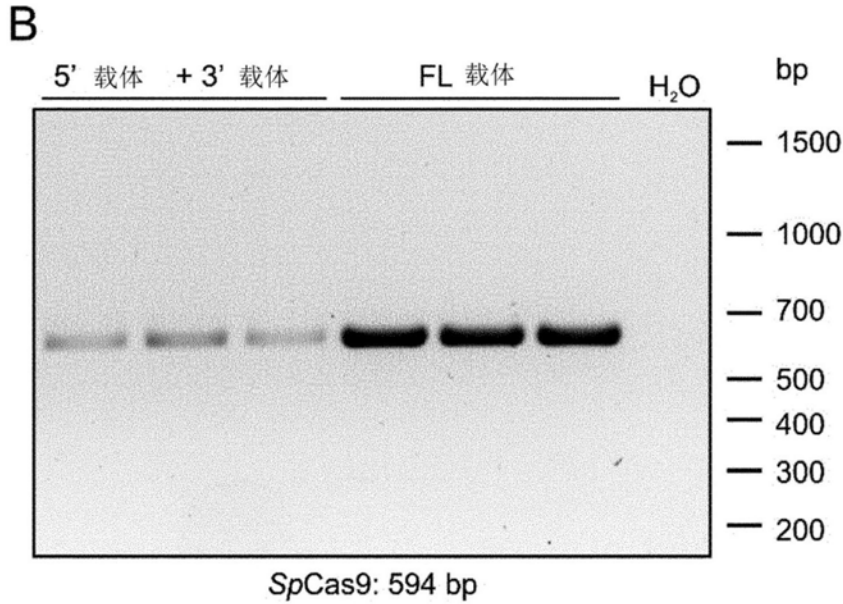
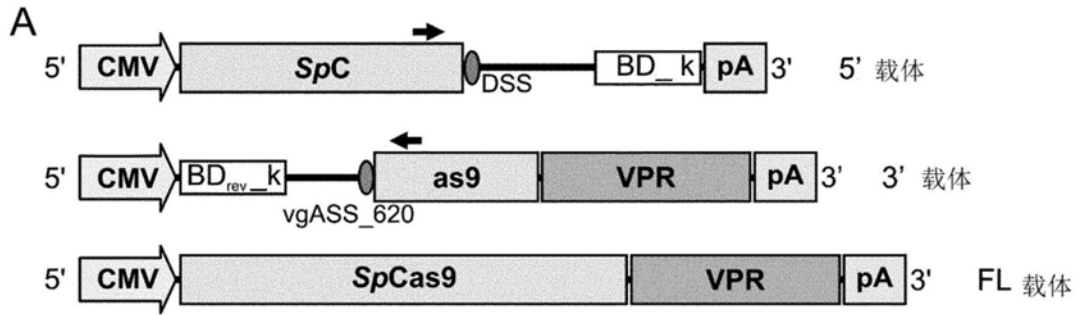


图13

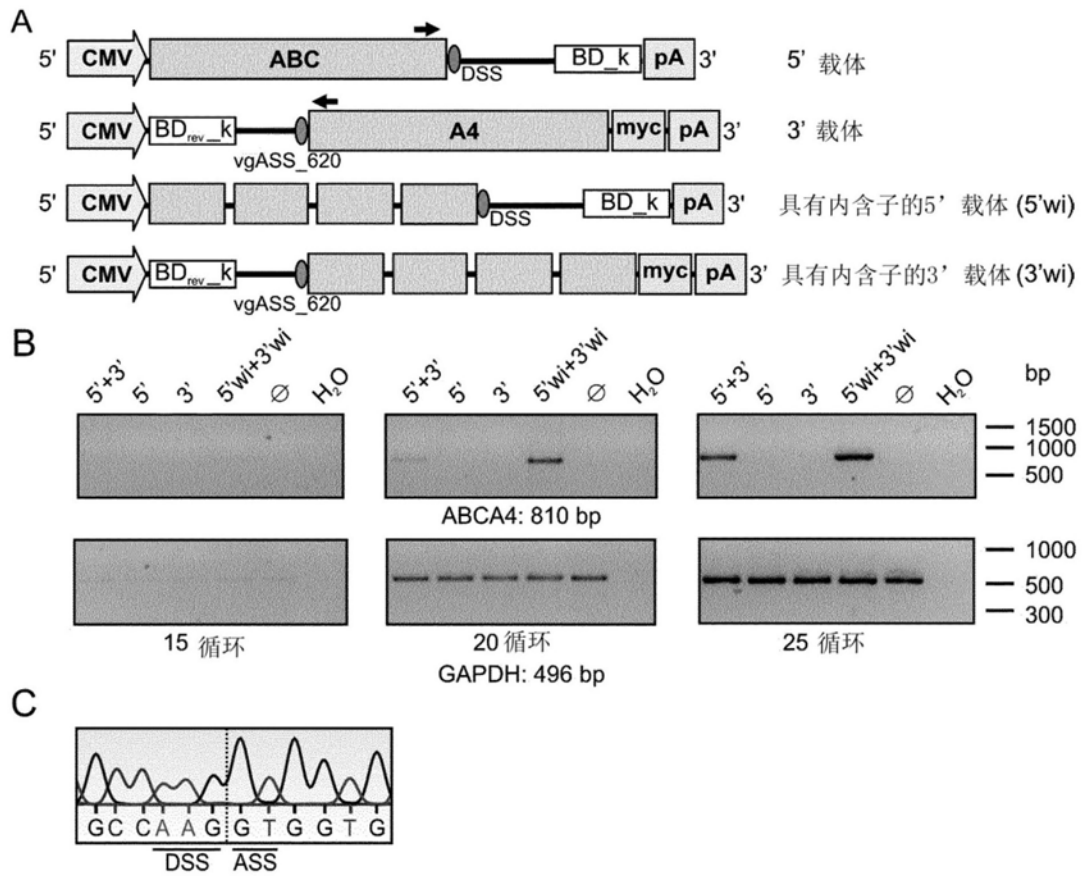


图14

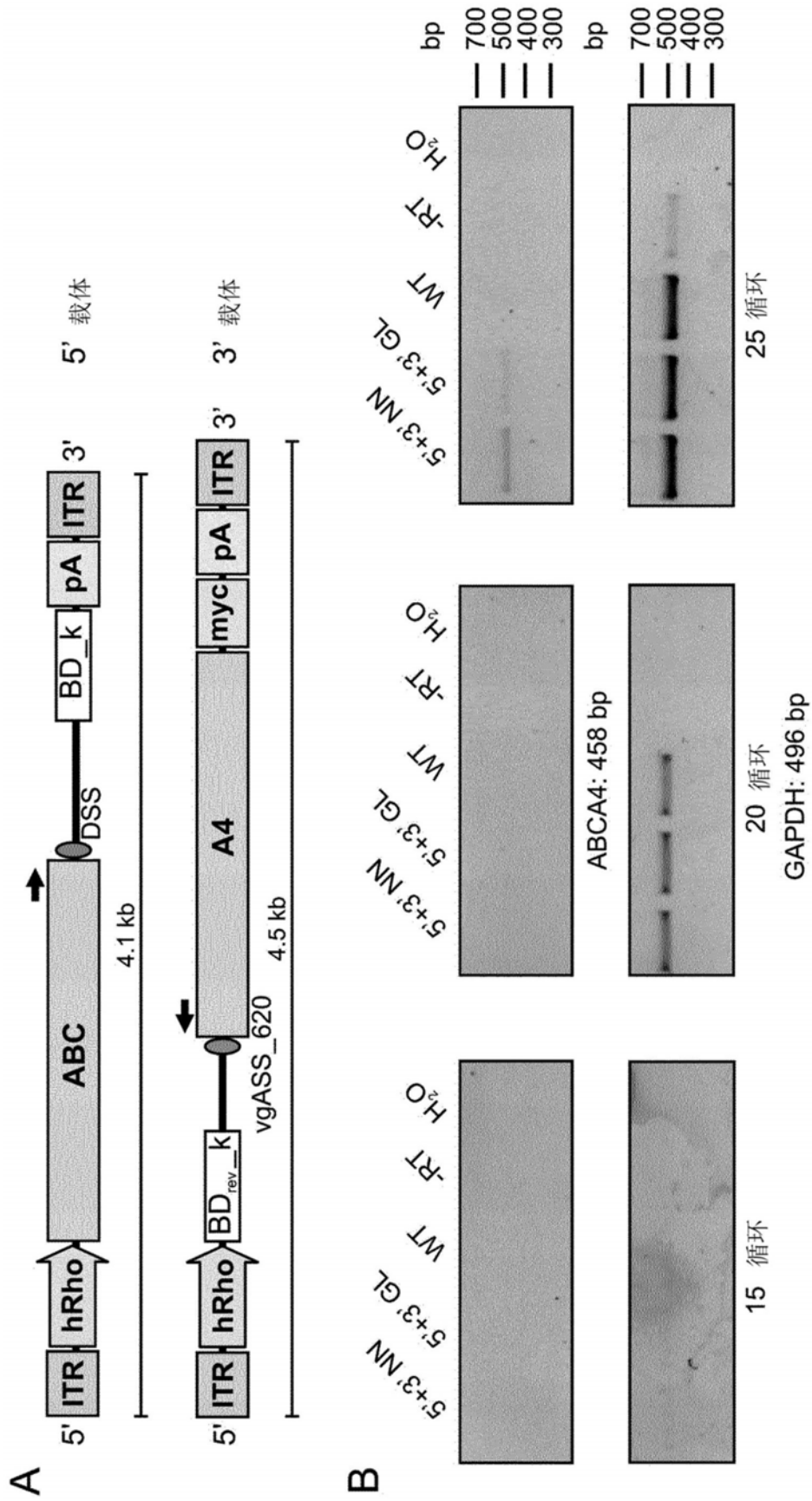
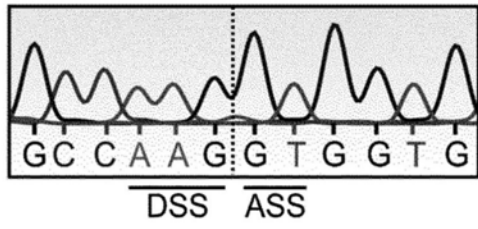
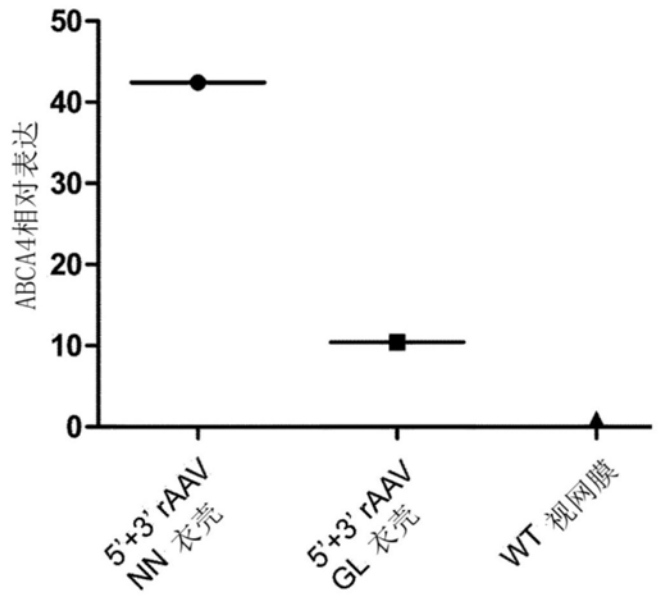


图15

C



D



E

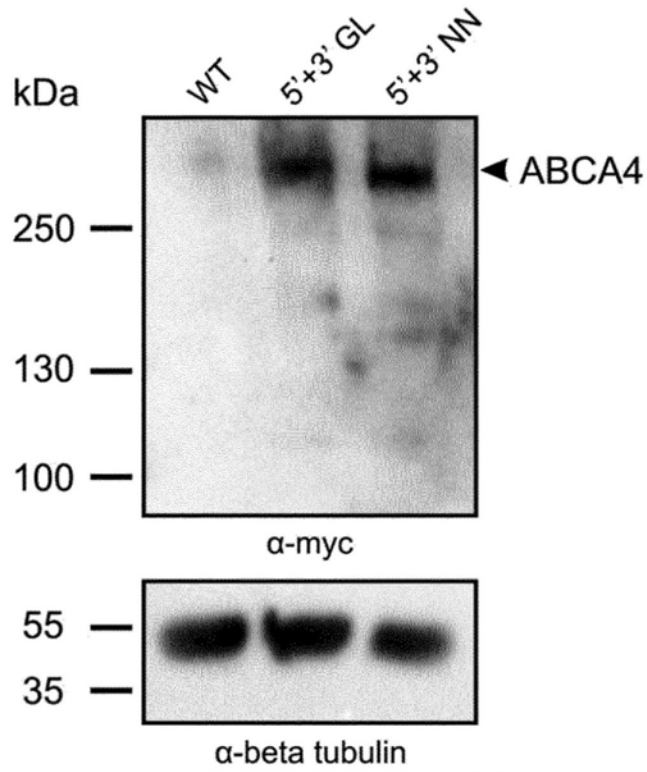


图15(续)

5' 天蓝蛋白 + BD_g (SEQ ID NO: 35):

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
TATATGGAGTTCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGA
CCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
CATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC
CCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTA
TTACCATGGTGTATGCGTTTTTGGCAGTACATCAATGGGCGTGGATAGCGTTTTGACTCACGG
GGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTTGTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTAC
GGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTT
ATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGTTAAGCTT**ATGGTGACC**
AAGGGCGAGGAGCTGTTACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAAA
CGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCC
TGAAGTTCATCTGCACCACAGGTAAGGGCACTGAGCAGAAGGGAAGAAGCTCCGGGGGCTCT
TTGTAGGGTGGATCCCCCTGCCTGTGAGGGACTTGGGAACAGCCAAGGTGGCTGCCGAGCTA
GCGAATTCCACCTGGCGCTGGCCCTTTCTGGGGCTTCACATTCCCTCCCTGGAACGGATCCC
TCGAGGTCACCCATTCGAACAAAACTCATCTCAGAAGAGGATCTGAATATGCATACCGGTC
ATCATCACCATCACCATTGAGTTTAAACCCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGC
CAGCCATCTGTTGTTTGCCCTCCCCGTGCCTTCTTGACCCTGGAAGGTGCCACTCCCAC
TGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTC
TGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCT
GGGGATGCGGTGGGCTCTATGG

图16

3' 天蓝蛋白 + BD_g (SEQ ID NO: 36):

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGA
CCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGACTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
CATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC
CCAGTACATGACCTTATGGGACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTA
TTACCATGGTGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTGACTCACGG
GGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTGGTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACCTCCGCCCATTTGACGCAAATGGGCGGTAGGCGTGTAC
GGTGGGAGGTCTATATAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTT
ATCGAAATTAATACGACTCACTATAGGGAGACCCAAGCTGGCTAGTTAAGCTTGGATCCGTT
CCAGGGAGGGAAATGTGAAGCCCAGAAAGGGCCAGCGCCAGGTGGAATTCGCTAGCTCGGCA
GCCACCTTGGCTGTTCCCAAGTCCCTCACAGGCAGGGGGATCCCAACGAGTCTTTTGTCTC
TACAGGTAAGCTGCCCGTGCCCTGGCCACCCCTCGTGACCACCCCTGACCTGGGGCGTGCAGT
GCTTCGCCCCTACCCGACCACATGAAGCAGCAGACTTCTTCAAGTCCGCCATGCCCGAA
GGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGA
GGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGG
AGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACGCCATCAGCGACAACGTCTATATC
ACCGCCGACAAGCAGAAGAACGGCATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGA
CGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCATCGGCGACGGCCCCGTGC
TGCTGCCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAG
CGCGATCACATGGTCTGCTGGAGTTCGTGACCGCCGCGGGATCACTCTCGGCATGGACGA
GCTGTACAAGTAACTCGAGGTCACCCATTCGAACAAAACTCATCTCAGAAGAGGATCTGAA
TATGCATACCGGTCATCATCACCATCACCATTTAGTTTAAACCCGCTGATCAGCCTCGACTG
TGCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCCGTGCCTTCCTTGACCCTGGAA
GGTGCCACTCCCACTGTCTTTTCTAATAAAAATGAGGAAATTGCATCGCATTGTCTGAGTAG
GTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACA
ATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGG

图17

5' SpCas9-VPR + BD_k (SEQ ID NO: 37):

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGA
CCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
ATTGACGTCAATGGGTGGAGTATTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
CATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC
CCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTA
TTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTTGACTCACGG
GGATTTCCAAGTCTCCACCCCATGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACG
GGACTTTCCAAAATGTCGTAACAACCTCCGCCCATGACGCAAATGGGCGGTAGGCGTGTAC
GGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAACCGGTCGACTAGAGGA
TCCATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCACGAGCCGACAAGAA
GTACTCCATTGGGCTCGCTATCGGCACAAACAGCGTCGGCTGGGCCGTCATTACGGACGAGT
ACAAGGTGCCGAGCAAAAATTCAAAGTTCTGGGCAATACCGATCGCCACAGCATAAAGAAG
AACCTCATTGGCGCCCTCCTGTTCGACTCCGGGGAAACGGCCGAAGCCACGCGGCTCAAAG
AACAGCACGGCGCAGATATACCCGCAGAAAGAATCGGATCTGCTACCTGCAGGAGATCTTTA
GTAATGAGATGGCTAAGGTGGATGACTCTTTCTTCCATAGGCTGGAGGAGTCCTTTTTGGTG
GAGGAGGATAAAAAGCACGAGCGCCACCCAATCTTTGGCAATATCGTGGACGAGGTGGCGTA
CCATGAAAAGTACCCAACCATATATCATCTGAGGAAGAAGCTTGTAGACAGTACTGATAAGG
CTGACTTGCGGTTGATCTATCTCGCGCTGGCGCATATGATCAAATTCGGGGACACTCCTC
ATCGAGGGGGACCTGAACCCAGACAACAGCGATGTCGACAAACTCTTTATCCAACCTGGTTCA
GACTTACAATCAGCTTTTTCGAAGAGAACCCGATCAACGCATCCGGAGTTGACGCCAAAGCAA
TCCTGAGCGCTAGGCTGTCAAATCCGGCGGCTCGAAAACCTCATCGCACAGCTCCCTGGG
GAGAAGAAGAACGGCCCTGTTTGGTAATCTTATCGCCCTGTCACTCGGGCTGACCCCAACTT
TAAATCTAACTTCGACCTGGCCGAAGATGCCAAGCTTCAACTGAGCAAAGACACCTACGATG
ATGATCTCGACAATCTGCTGGCCCAGATCGGCGACCAGTACGCAGACCTTTTTTTGGCGCA
AAGAACCTGTCAGACGCCATTCTGCTGAGTGATATCTGCGAGTGAACACGGAGATCACCAA
AGCTCCGCTGAGCGCTAGTATGATCAAGCGCTATGATGAGCACCACCAAGACTTGACTTTGC
TGAAGGCCCTTGTCAGACAGCAACTGCCTGAGAAGTACAAGGAAATTTCTTCGATCAGTCT
AAAAATGGCTACGCCGGATACATTGACGGCGGAGCAAGCCAGGAGGAATTTTACAAATTTAT
TAAGCCCATCTTGAAAAAATGGACGGCACCGAGGAGCTGCTGGTAAAGCTTAACAGAGAAG
ATCTGTTGCGCAAACAGCGCACTTTCGACAATGGAAGCATCCCCACCAGATTCACCTGGGC
GAACGTCACGCTATCCTCAGGCGGCAAGAGGATTTCTACCCCTTTTTGAAAGATAACAGGGA
AAAGATTGAGAAAATCCTCACATTTTCGGATACCCTACTATGTAGGCCCCCTCGCCGGGGAA
ATTCAGATTTCGCTGGATGACTCGCAAATCAGAAGAGACCATCACTCCCTGGAACCTTCGAG
GAAGTCGTGGATAAGGGGGCCTCTGCCAGTCTTCATCGAAAGGATGACTAACTTTGATAA
AAATCTGCCTAACGAAAAGGTGCTTCTAAACACTCTCTGCTGTACGAGTACTTCACAGTTT
ATAACGAGCTACCAAGGTCAAATACGTCACAGAAGGGATGAGAAAGCCAGCATTCTGTCT
GGAGAGCAGAAGAAAGCTATCGTGGACCTCCTCTTCAAGACGAACCGGAAAGTTACCGTGAA
ACAGCTCAAAGAAGACTATTTCAAAGATTGAATGTTTCGACTCTGTTGAAATCAGCGGAG
TGGAGGATCGCTTCAACGCATCCCTGGGAACGTATCACGATCTCCTGAAAATCATTAAAGAC
AAGGACTTCTGGACAATGAGGAGAACGAGGACATTCTTGAGGACATTGTCCTCACCCCTTAC

图18

GTTGTTTGAAGATAGGGAGATGATTGAAGAACGCTTGAAAACCTTACGCTCATCTCTTCGACG
ACAAAGTCATGAAACAGCTCAAGAGGGCGCCGATATACAGGATGGGGGCGGCTGTCAAGAAAA
CTGATCAATGGGATCCGAGACAAGCAGAGTGGAAAGACAATCCTGGATTTCTTAAGTCCGA
TGGATTTGCCAACCGGAACTTCATGCAGTTGATCCATGATGACTCTCTCACCTTTAAGGAGG
ACATCCAGAAAGCACAAGTTTCTGGCCAGGGGGACAGTCTTCACGAGCACATCGCTAATCTT
GCAGGTAAGGGCACTGAGCAGAAGGGAAGAAGCTCCGGGGGCTCTTTGTAGGGTGCGGCCGC
AGTGCATCAAGGCGATCACATCAGTGAAAAAAGCCAGACAGGCGGTTAAACCAACGCAGAT
TAAACAGCAGGATGCAAAAATTCGCAGGTGGTCAGATGGCGGCCGCTCTAGACTCGATGAGT
TTGGACAAACCACAACCTAGAATGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCT
ATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTT

图18(续)

3' SpCas9-VPR + BD_k (SEQ ID NO: 38):

GACATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCA
 TATATGGAGTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGA
 CCCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCC
 ATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
 CATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGC
 CCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTA
 TTACCATGGTGTATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTACTCACGG
 GGATTTCCAAGTCTCCACCCATTGACGTCAATGGGAGTTTGTTTTGGCACAAAATCAACG
 GGACTTTCCAATAATGTCGTAACAACCTCCGCCCATGACGCAAATGGGCGGTAGGCGTGTAC
 GGTGGGAGGTCTATATAAGCAGAGCTCGTTTTAGTGAACCGTCAGAGGTACCACCGGTGCGACT
 AGAGGATCCCGGCCGCCATCTGACCACCTGCGAATTTTTGCATCCTGCTGTTAATCTGCGT
TGGTTTTAACCGCCTGTCTGGCTTTTTTTCACTGATGTGATCGCCTTGATGCACTCGGCCGCC
AACGAGTCTTTTGTATCTACAGGTAGCCAGCTATCAAAAAGGGAATACTGCAGACCGTTA
AGGTCGTGGATGAACTCGTCAAAGTAATGGGAAGGCATAAGCCCGAGAATATCGTTATCGAG
ATGGCCCGAGAGAACCAACTACCCAGAAGGGACAGAAGAACAGTAGGGAAAGGATGAAGAG
GATTGAAGAGGGTATAAAGAAGTGGGGTCCCAAATCCTTAAGGAACACCCAGTTGAAAACA
CCCAGCTTCAGAATGAGAAGCTCTACCTGTACTACCTGCAGAACGGCAGGGACATGTACGTG
GATCAGGAAGTGGACATCAATCGGCTCTCCGACTACGACGTGGATCATATCGTGCCCCAGTC
TTTTCTCAAAGATGATTCTATTGATAATAAAGTGTGACAAGATCCGATAAAAATAGAGGGA
AGAGTGATAACGTCCCCTCAGAAGAAGTTGTCAAGAAAATGAAAAATTATTGGCGGCAGCTG
CTGAACGCCAACTGATCACACAACGGAAGTTCGATAATCTGACTAAGGCTGAACGAGGTGG
CCTGTCTGAGTTGGATAAAGCCGGCTTCATCAAAAGGCAGCTTGTGAGACACGCCAGATCA
CCAAGCACGTGGCCCAAATCTCGATTACGCATGAACACCAAGTACGATGAAAATGACAAA
CTGATTCGAGAGGTGAAAGTTATTACTCTGAAGTCTAAGCTGGTCTCAGATTTAGAAAAGGA
CTTTCAGTTTTATAAGGTGAGAGAGATCAACAATTACCACCATGCGCATGATGCCTACCTGA
ATGCAGTGGTAGGCACTGCACCTTATCAAAAATATCCCAAGCTTGAATCTGAATTTGTTTAC
GGAGACTATAAAGTGTACGATGTTAGGAAAATGATCGCAAAGTCTGAGCAGGAAATAGGCAA
GGCCACCGCTAAGTACTTCTTTTACAGCAATATTATGAATTTTTTCAAGACCGAGATTACAC
TGGCCAATGGAGAGATTTCGGAAGCGACCATTATCGAAAACAAACGGAGAAACAGGAGAAATC
GTGTGGGACAAGGGTAGGGATTCGCGACAGTCCGGAAGGTCTGTCCATGCCGAGGTGAA
CATCGTTAAAAAGACCGAAGTACAGACCGGAGGCTTCTCCAAGGAAAGTATCCTCCGAAAA
GGAACAGCGACAAGCTGATCGCACGCAAAAAGATTGGGACCCCAAGAAATACGGCGGATTC
GATTCTCTACAGTCGCTTACAGTGTACTGGTTGTGGCCAAAGTGGAGAAAGGGAAGTCTAA
AAAACCTCAAAGCGTCAAGGAAGTCTGGGCATCACAATCATGGAGCGATCAAGCTTCGAAA
AAAACCCATCGACTTCTCGAGGCGAAAGGATATAAAGAGGTCAAAAAGACCTCATCATT
AAGCTTCCAAGTACTCTCTTTGAGCTTGAAAACGGCCGGAAACGAATGCTCGCTAGTGC

图19

GGGCGAGCTGCAGAAAGGTAACGAGCTGGCACTGCCCTCTAAATACGTTAATTTCTTGTATC
TGGCCAGCCACTATGAAAAGCTCAAAGGGTctCCCGAAGATAATGAGCAGAAGCAGCTGTTC
GTGGAACAACACAACACTACCTTGATGAGATCATCGAGCAAATAAGCGAATTCTCCAAAAG
AGTGATCCTCGCCGACGCTAACCTCGATAAGGTGCTTTC TGCTTACAATAAGCACAGGGATA
AGCCCATCAGGGAGCAGGCAGAAAACATTATCCACTTGTTTACTCTGACCAACTTGGGCGCG
CCTGCAGCCTTCAAGTACTTCGACACCACCATAGACAGAAAGCGGTACACCTCTACAAAGGA
GGTCTTGACGCCACACTGATTCATCAGTCAATTACGGGGCTCTATGAAACAAGAATCGACC
TCTCTCAGCTCGGTGGAGACAGCAGGGCTGACCCCAAGAAGAAGAGGAAGGTGTCGCCAGGG
ATCCGTGCACTTGACGCGTTGATATCAACAAGTTTGTACAAAAAAGCAGGCTACAAAGAGGC
CAGCGTTCCGGACGGGCTGACGCATTGGACGATTTTGATCTGGATATGCTGGGAAGTGACG
CCCTCGATGATTTTGACCTTGACATGCTTGGTTCGGATGCCCTTGATGACTTTGACCTCGAC
ATGCTCGGCAGTGACGCCCTTGATGATTCGACCTGGACATGCTGATTAActCTAGAAGTTC
CGGATCTCCGAAAAGAAACGCAAAGTTGGTAGCCAGTACCTGCCCGACACCGACGACCGGC
ACCGGATCGAGGAAAAGCGGAAGCGGACCTACGAGACATTCAAGAGCATCATGAAGAAGTCC
CCCTTCAGCGGCCCCACCGACCCTAGACCTCCACCTAGAAGAATCGCCGTGCCCAGCAGATC
CAGCGCCAGCGTGCCAAAACCTGCCCCAGCCTTACCCCTTACCAGCAGCCTGAGCACCA
TCAACTACGACGAGTTCCTTACCATGGTGTTCCTCCAGCGGCCAGATCTCTCAGGCCTCTGCT
CTGGCTCCAGCCCCTCCTCAGGTGCTGCCTCAGGCTCCTGCTCCTGCACCAGCTCCAGCCAT
GGTGTCTGCACTGGCTCAGGCACCAGCACCCGTGCCTGTGCTGGCTCCTGGACCTCCACAGG
CTGTGGCTCCACCAGCCCCTAAACCTACACAGGCCGGCGAGGGCACACTGTCTGAAGCTCTG
CTGCAGCTGCAGTTCGACGACGAGGATCTGGGAGCCCTGCTGGGAAACAGCACCGATCCTGC
CGTGTTCACCGACCTGGCCAGCGTGGACAACAGCGAGTTCAGCAGCTGCTGAACCAGGGCA
TCCCTGTGGCCCCCTCACACCACCGAGCCATGCTGATGGAATACCCCGAGGCCATCACCCGG
CTCGTGACAGGCGCTCAGAGGCCTCCTGATCCAGCTCCTGCCCTCTGGGAGCACCAGGCCT
GCCTAATGGACTGCTGTCTGGCGACGAGGACTTCAGCTCTATCGCCGATATGGATTTCTCAG
CCTTGCTGGGCTCTGGCAGCGGCAGCCGGGATTCAGGGAAGGGATGTTTTTGCCGAAGCCT
GAGGCCGGCTCCGCTATTAGTGACGTGTTTGGGGCCGCGAGGTGTGCCAGCCAAAACGAAT
CCGGCCATTTTCATCCTCCAGGAAGTCCATGGGCCAACCGCCACTCCCCGCCAGCCTCGCAC
CAACACCAACCGGTCCAGTACATGAGCCAGTCGGGTCAC TGACCCCGGCACCAGTCCCTCAG
CCACTGGATCCAGCGCCCGCAGTGACTCCCGAGGCCAGTCACCTGTTGGAGGATCCCGATGA
AGAGACGAGCCAGGCTGTCAAAGCCCTTCGGGAGATGGCCGATACTGTGATTTCCCAGAAGG
AAGAGGCTGCAATCTGTGGCCAAATGGACCTTTCCCATCCGCCCCCAAGGGGCCATCTGGAT
GAGCTGACAACCACACTTGAGTCCATGACCGAGGATCTGAACCTGGACTCACCCCTGACCCC
GGAATTGAACGAGATTCTGGATACCTTCCCTGAACGACGAGTGCCTCTTGATGCCATGCATA
TCAGCACAGGACTGTCCATCTTCGACACATCTCTGTTTtga cAATAAAATATCTTTATTTTC
ATTACATCTGTGTGTTGGTTTTTTTGTGT

图19(续)

包含带有内含子的 5' ABCA4 的双 AAV 载体 (SEQ ID NO: 39):

```

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCGACCTTTGG
TCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGCCAACTCCATCACTAGGG
GTTCTTgtagttaatgattaaccgcctatgctacttatctacgtagccatgctctaggaag
atcggaattcgccttaagggcgccgctttaaatagctagcGACATTGATTATTGACTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGTTAC
ATAACTTACGGTAAATGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAG
TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCC
TATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTACATGACCTTATGGG
ACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTT
TGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCC
CATTGACGTCAATGGGAGTTTGTGTTTGGCACAAAATCAACGGGACTTTCCAAAATGTGTA
ACAACCTCCGCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGC
AGAGCTggtaccaccggtgccaccATGGGCTTCGTGAGACAGATACAGCTTTTGCTCTGGAA
GAACTGGACCCCTGCGGAAAAGGCAAAGATTTCGCTTTGTGGTGGAACTCGTGTGGCCTTTAT
CTTTATTTCTGGTCTTGATCTGGTTAAGGAATGCCAACCCACTCTACAGCCATCATGAATGC
CATTTCCCAACAAGGCGATGCCCTCAGCAGGAATGCTGCCGTGGCTCCAGGGGATCTTCTG
CAATGTGAACAATCCCTGTTTTCAAAGCCCCACCCAGGAGAATCTCCTGGAATTGTGTCAA
ACTATAACAACCTCCATCTTGGCAAGGTATATCGAGATTTTCAAGAACTCCTCATGAATGCA
CCAGAGAGCCAGCACCTTGGCCGTATTTGGACAGAGCTACACATCTTGTCCCAATTCATGGA
CACCTCCGGACTCACCCGAGAGAATTGCAGGTAAGTGGCGACCGTGCCCCACAGCCCTAG
CCGCCCTCCCCACTGCCCGGGCTTACCCTGGTCTGCTCCCGCAGGAAGAGGAATACGAATA
AGGGATATCTTGAAAGATGAAGAAACACTGACACTATTTCTCATTAAAAACATCGGCCTGTC
TGACTCAGTGGTCTACCTTCTGATCAACTCTCAAGTCCGTCCAGAGCAGTTCGCTCATGGAG
TCCCGGACCTGGCGCTGAAGGACATCGCTGCAGCGAGGCCCTCCTGGAGCGCTTCATCATC
TTCAGCCAGAGACGCGGGGCAAAGACGGTGCCTATGCCCTGTGCTCCCTCTCCAGGGCAC
CCTACAGTGGATAGAAGACACTCTGTATGCCAACGTGGACTTCTTCAAGCTCTTCCGTGTGC
TTCCACACTCCTAGACAGCCGTTCTCAAGGTATCAATCTGAGATCTTGGGGAGGAATATTA
TCTGATATGTCACCAAGAATTCAAGAGTTTATCCATCGGCCGAGTATGCAGGACTTGCTGTG
GGTGACCAGGCCCTCATGCAGAAATGGTGGTCCAGAGACCTTTACAAAGCTGATGGGCATCC
TGTCTGACCTCCTGTGTGGCTACCCCGAGGGAGGTGGCTCTCGGGTGCTCTCCTTCAACTGG
TATGAAGACAATAACTATAAGGCCTTTCTGGGGATTGACTCCACAAGGAAGGATCCTATCTA
TTCTTATGACAGAAGAACAACATCCTTTTGTAAATGCATTGATCCAGAGCCTGGAGTCAAATC
CTTTAACCAAAAATCGCTTGGAGGGCGGCAAAGCCTTTGCTGATGGGAAAAATCCTGTACT
CCTGATTCACCTGCAGCACGAAGATACTGAAGAATGCCAACTCAACTTTTGAAGAACTGGA
ACACGTTAGGAAGTTGGTCAAAGCCTGGGAAGAAGTAGGGCCCCAGATCTGGTACTTCTTTG

```

图20

ACAACAGCACACAGATGAACATGATCAGAGATACCCTGGGGAACCCAACAGTAAAAGACTTT
 TTGAATAGGCAGCTTGGTGAAGAAGGTATTACTGCTGAAGCCATCCTAAACTTCCTCTACAA
 GGGCCCTCGGAAAGCCAGGCTGACGACATGGCCAACCTCGACTGGAGGGACATATTTAACA
 TCACTGATCGCACCCCTCCGCCTGGTCAATCAATACCTGGAGTGCTTGGTCCTGGATAAGTTT
 GAAAGCTACAATGATGAAACTCAGCTCACCCAACGTGCCCTCTCTCTACTGGAGGAAAACAT
 GTTCTGGGCCGGAGTGGTATTCCCTGACATGTATCCCTGGACCAGCTCTCTACCACCCACG
 TGAAGTATAAGATCCGAATGGACATAGACGTGGTGGAGAAAACCAATAAGATTAAAGACAGG
 TGAGTGGACTGGAGCCTGGGCACGAGGTGTGGGGTGGCCCCTGCCCTGCCACTTACACCACC
 TGCCCTCTTCTGCAGGTATTGGGATTCTGGTCCCAGAGCTGATCCCCTGGAAGATTTCGGGT
 ACATCTGGGGCGGGTTTGCCTATCTGCAGGACATGGTTGAACAGGGGATCACAAGGAGCCAG
 GTGCAGGCGGAGGCTCCAGTTGGAATCTACCTCCAGCAGATGCCCTACCCCTGCTTCGTGGA
 CGATTCTTTTCATGATCATCTGAACCGCTGTTTCCCTATCTTCATGGTGCTGGCATGGATCT
 ACTCTGTCTCCATGACTGTGAAGAGCATCGTCTTGGAGAAGGAGTTGCGACTGAAGGAGACC
 TTGAAAATCAGGGTGTCTCCAATGCAGTGATTTGGTGTACCTGGTTCCTGGACAGCTTCTC
 CATCATGTGATGAGCATCTTCCCTGACGATATTCATCATGCATGGAAGAATCCTACATT
 ACAGCGACCCATTTCATCCTCTTCCCTGTTCTTGTGGCTTTCTCCACTGCCACCATCATGCTG
 TGCTTTCTGCTCAGCACCTTCTTCTCAAGGCCAGTCTGGCAGCAGCCTGTAGTGGTGTGAT
 CTATTTACCCCTCTACCTGCCACACATCCTGTGCTTCGCCTGGCAGGACCGCATGACCGCTG
 AGCTGAAGAAGGCTGTGAGCTTACTGTCTCCGGTGGCATTGGATTGGCACTGAGTACCTG
 GTTCGCTTTGAAGAGCAAGGCCTGGGGCTGCAGTGGAGCAACATCGGGAACAGTCCCACGGA
 AGGGGACGAATTCAGCTTCCCTGCTGTCCATGCAGATGATGCTCCTTGATGCTGCTGTCTATG
 GCTTACTCGCTTGGTACCTTGATCAGGTGTTTCCAGGAGACTATGGAACCCCACTTCCCTGG
 TACTTTCTTCTACAAGAGTCGTATTGGCTTGGCGGTGAAGGTAAGTAGTCTGATTATACACA
 AGATATTGTCTAGAACTTGATGAGACTGTGGATATGAATATTTCACTCTTTTCTCAGGGTGT
 TCAACCAGAGAAGAAAGAGCCCTGGAAAAGACCCGAGCCCCTAACAGAGGAAACGGAGGATCC
 AGAGCACCCAGAAGGAATACACGACTCCTTCTTTGAACGTGAGCATCCAGGGTGGGTTCCTG
 GGGTATGCGTGAAGAATCTGGTAAAGATTTTTGAGCCCTGTGGCCGGCCAGCTGTGGACCGT
 CTGAACATCACCTTCTACGAGAACCAGATCACCGCATTCCTGGGCCACAATGGAGCTGGGAA
 AACCACCACCTTGTCCATCCTGACGGGTCTGTTGCCACCAACCTCTGGGACTGTGCTCGTTG
 GGGGAAGGGACATTGAAACCAGCCTGGATGCAGTCCGGCAGAGCCTTGGCATGTGTCCACAG
 CACAACATCCTGTTCCACCACCTCACGGTGGCTGAGCACATGCTGTTCTATGCCAGCTGAA
 AGAAAAGTCCCAGGAGGAGGCCAGCTGGAGATGGAAGCCATGTTGGAGGACACAGGCCTCC
 ACCACAAGCGGAATGAAGAGGCTCAGGACCTATCAGGTGGCATGCAGAGAAAGCTGTCCGTT
 GCCATTGCCTTTGTGGGAGATGCCAAGgt aagggcactgagcagaagggagaagctccggg
 ggctctttgtagggt aagcttAGTGCATCAAGGCGATCACATCAGTGAAAAAAGCCAGACA
 GGCGGTTAAACCAACGCAGATTAACAGCAGGATGCAAAAATTCGCAGGTGGTCAGATGaag
 cttattgtcgagGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTAT
 TTGTAACCATTATAAGCTGCAATAAACAAGTTctcgagttaagggcgaattcccgattagga

图20 (续)

```
tcttcctagagcatggctacgtagataagtagcatggcggggttaatcattaactacaAGGAA  
CCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCG  
ACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCGCA  
G
```

图20(续)

包含带有内含子的 3' ABCA4 的双 AAV 载体 ; (SEQ ID NO: 40):

```
CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCCGGGCGACCTTTGG
TCGCCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCCAACCTCCATCACTAGGG
GTTCTTgtagttaatgattaaccgccatgctacttacctacgtagccatgctctaggaag
atcgggaattcgcccttaagggcgcgccgtttaaatagctagcGACATTGATTATTGACTAGT
TATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCGCGGTTAC
ATAACTTACGGTAAATGGCCCCGCTGGCTGACCGCCCAACGACCCCCGCCATTGACGTCAA
TAATGACGTATGTTCCCATAGTAACGCCAATAGGGACTTTCATTGACGTCAATGGGTGGAG
TATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCC
TATTGACGTCAATGACGGTAAATGGCCCCGCTGGCATTATGCCCAGTACATGACCTTATGGG
ACTTTCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCATGGTGATGCGGTTT
TGGCAGTACATCAATGGGCGTGGATAGCGTTTTGACTCACGGGGATTTCCAAGTCTCCACCC
CATTGACGTCAATGGGAGTTTGTTTTGGCACAAAATCAACGGGACTTTCCAAAATGTGCTA
ACAACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGGTACGGTGGGAGGTCTATATAAGC
AGAGCTggtaccaccggtagaaagcttCATCTGACCACCTGCGAATTTTTGCATCCTGTGT
TTAATCTGCGTTGGTTTAAACCGCCTGTCTGGCTTTTTTTTCACTGATGTGATCGCCTTGATGC
ACTaagcttCAACGAGTCTTTTGTCTACAGGTGGTGATTCTGGACGAACCCACCTCTGG
GGTGGACCCCTACTCGAGACGCTCAATCTGGGATCTGCTCCTGAAGTATCGCTCAGGCAGAA
CCATCATCATGTCCACTCACCACATGGACGAGGCCGACCTCCTTGGGGACCGCATTGCCATC
ATTGCCCAGGGAAGGCTCTACTGCTCAGGCACCCCACTCTTCTGAAGAACTGCTTTGGCAC
AGGCTTGTACTTAACCTTGGTGGCAGATGAAAAACATCCAGAGCCAAAGGAAAGGCAGTG
AGGTGAGTCAGGCTGGGCGCCCCCGCCCCAGGGGCCCTCCCTCCCCAAGCCCCCGGACGC
GCCTCACCCACGTTCTCTCGCAGGGGACCTGCAGCTGCTCGTCTAAGGGTTTCTCCACCAC
GTGTCCAGCCCACGTGATGACCTAACTCCAGAACAAGTCTGGATGGGGATGTAATGAGC
TGATGGATGTAGTTCTCCACCATGTTCCAGAGGCAAAGCTGGTGGAGTGCAATGGTCAAGAA
CTTATCTTCTTCTTCCAAATAAGAACTTCAAGCACAGAGCATATGCCAGCCTTTTTCAGAGA
GCTGGAGGAGACGCTGGCTGACCTTGGTCTCAGCAGTTTTGGAATTTCTGACACTCCCCTGG
AAGAGATTTTTCTGAAGGTCACGGAGGATTCTGATTCAAGACCTCTGTTTGGGGTGGCGCT
CAGCAGAAAAGAGAAAACGTCAACCCCCGACACCCCTGCTTGGGTCCCAGAGAGAAGGCTGG
ACAGACACCCCAAGGACTCCAATGTCTGCTCCCCAGGGGCGCCGGCTGCTCACCCAGAGGGCC
AGCCTCCCCCAGAGCCAGAGTGCCAGGCCCGCAGCTCAACACGGGGACACAGCTGGTCCTC
CAGCATGTGCAGGCGCTGCTGGTCAAGAGATTCCAACACACCATCCGCAGCCACAAGGACTT
CCTGGCGCAGATCGTGCTCCCGGCTACCTTTGTGTTTTTGGCTCTGATGCTTTCTATTGTTA
TCCCTCCTTTTGGCGAATACCCCGCTTTGACCCTTACCCCTGGATATATGGGCAGCAGTAC
ACCTTCTTTCAGCATGGATGAACCAGGCAGTGAGCAGTTCACGGTACTTGACAGACGTCCTCCT
GAATAAGCCAGGCTTTGGCAACCGCTGCCTGAAGGAAGGGTGGCTTCCGGAGTACCCCTGTG
GCAACTCAACACCCTGGAAGACTCCTTCTGTGTCCCCAACATCACCCAGCTGTTCCAGAAG
```

图21

CAGAAATGGACACAGGTCAACCCTTACCATCCTGCAGGTGCAGCACCAGGGAGAAGCTCAC
CATGCTGCCAGAGTGCCCCGAGGGTGCCGGGGGCTCCCGCCCCCAGAGAACACAGCGCA
GCACGGAAATTCTACAAGACCTGACGGACAGGAACATCTCCGACTTCTTGGTAAAAACGTAT
CCTGCTCTTATAAGAAGCAGCTTAAAGAGCAAATTCTGGGTCAATGAACAGAGGTATGGAGG
AATTTCCATTGGAGGAAAGCTCCCAGTCGTCCCCATCACGGGGGAAGCACTTGTGGGTTTT
TAAGCGACCTTGGCCGGATCATGAATGTGAGCGGGGGCCCTATCACTAGAGAGGCCTCTAAA
GAAATACCTGATTTCTTAAACATCTAGAACTGAAGACAACATTAAGGTGTGGTTTAATAA
CAAAGGCTGGCATGCCCTGGTCAGCTTTCTCAATGTGGCCACAACGCCATCTTGTAAGTCC
TACCTTTTTTGTTCCTTTGAAAGCCTCCTGGAAAGCTTTTTCTGAAGTGTTTGTCTGTAAT
TTCTTTGCAGACGGGCCAGCCTGCCTAAGGACAGGAGCCCCGAGGAGTATGGAATCACCGTC
ATTAGCCAACCCCTGAACCTGACCAAGGAGCAGCTCTCAGAGATTACAGTGCTGACCACTTC
AGTGGATGCTGTGGTTGCCATCTGCGTGATTTTCTCCATGTCCTTCGTCCAGCCAGCTTTG
TCCTTTATTTGATCCAGGAGCGGGTGAACAAATCCAAGCACCTCCAGTTTATCAGTGGAGTG
AGCCCCACCACCTACTGGGTGACCAACTTCTCTGGGACATCATGAATTATTCCGTGAGTGC
TGGGCTGGTGGTGGGCATCTTCATCGGGTTTTAGAAGAAAGCCTACACTTCTCCAGAAAACC
TTCTGCCCTTGTGGCACTGCTCCTGCTGTATGGATGGGCGGTTCATCCCATGATGTACCCA
GCATCCTTCTGTTTGTATGTCCCAGCACAGCCTATGTGGCTTTATCTTGTGCTAATCTGTT
CATCGGCATCAACAGCAGTGCTATTACCTTCATCTTGAATTATTTGAGAATAACCGGACGC
TGCTCAGGTTCAACGCCGTGCTGAGGAAGCTGCTCATTGTCTTCCCCACTTCTGCCTGGGC
CGGGCCTCATTGACCTTGCCTGAGCCAGGCTGTGACAGATGTCTATGCCCGGTTTGGTGA
GGAGCACTCTGCAAAATCCGTTCCACTGGGACCTGATTGGGAAGAACCTGTTTGCCATGGTGG
TGGAAAGGGTGGTGTACTTCTCCTGACCCTGCTGGTCCAGCGCCACTTCTTCTCTCCCAA
TGGATTGCCGAGCCCCTAAGGAGCCCATTGTTGATGAAGATGATGATGTGGCTGAAGAAAG
ACAAAGAATTATTACTGGTGGAAATAAACTGACATCTTAAGGCTACATGAACTAACCAAGA
TTTATCCAGGCACCTCCAGCCCAGCAGTGGACAGGCTGTGTGTGCGGAGTTCGCCCTGGAGAG
TGCTTTGGCCTCCTGGGAGTGAATGGTGCCGGCAAACAACCACATTCAGATGCTCACTGG
GGACACCACAGTGACCTCAGGGGATGCCACCGTAGCAGGCAAGAGTATTTTAACCAATATTT
CTGAAGTCCATCAAAATATGGGCTACTGTCCTCAGTTTGTATGCAATTGATGAGCTGCTCACA
GGACGAGAACATCTTTACCTTTATGCCCGCTTCGAGGTGTACCAGCAGAAGAAATCGAAA
GGTAAGTGATTCTTAGGGCTGGGGAAGGTGGGTGGGAATCCTCTCCTGCTCACCTCCTCTCT
CCTGCCCCACAGGTTGCAAACTGGAGTATTAAGAGCCTGGGCCTGACTGTCTACGCCGACTG
CCTGGCTGGCACGTACAGTGGGGGCAACAAGCGGAACTCTCCACAGCCATCGCACTCATTG
GCTGCCACCCTGGTGTGCTGGATGAGCCACCACAGGGATGGACCCCAGGCACGCCGC
ATGCTGTGGAACGTCATCGTGAGCATCATCAGAGAAGGGAGGGCTGTGGTCTCACATCCCA
CAGCATGGAAGAATGTGAGGCACTGTGTACCCGGCTGGCCATCATGGTAAAGGGCGCCTTTC
GATGTATGGGCACCATTACAGCATCTCAAGTCCAAATTTGGAGATGGCTATATCGTCACAATG
AAGATCAAATCCCCGAAGGACGACCTGCTTCTGACCTGAACCCTGTGGAGCAGTCTTCCA
GGGGAACCTCCAGGCAGTGTGCAGAGGGAGAGGCACTACAACATGCTCCAGTTCAGGTCT

图21 (续)

CCTCCTCCTCCCTGGCGAGGATCTTCCAGCTCCTCCTCTCCCACAAGGACAGCCTGCTCATC
GAGGAGTACTCAGTCACACAGACCACACTGGACCAGGTGTTTGTAATTTTGCTAAACAGCA
GACTGAAAGTCATGACCTCCCTCTGCACCCTCGAGCTGCTGGAGCCAGTCGACAAGCCCAGG
ACTGAgtcgacgcggccgcGCAGTGAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATT
GCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGTTctcgagttaagggcgaattcccg
attaggatcttcttagagcatggctacgtagataagtagcatggcgggtaatacattaacta
caAGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCCTCTGCGCGCTCGCTCGCTCACTGAGG
CCGGGCGACCAAAGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGA
GCGCGCAG

图21 (续)