(19)国家知识产权局

(12)发明专利申请

(10)申请公布号 CN 118147141 A
(43)申请公布日 2024.06.07

(21)申请号 202410269446.8

(22)申请日 2016.11.30

(30)优先权数据
62/260712 2015.11.30 US
62/330336 2016.05.02 US

(62)分案原申请数据
201680080439.7 2016.11.30

(71)申请人 杜克大学
地址 美国北卡罗来纳州

(72)发明人 C·A·格尔斯巴赫
J·N·罗宾逊哈姆

(74)专利代理机构 中国专利代理(香港)有限公司 72001
专利代理师 彭昶

(51)Int.Cl.
C12N 15/113(2010.01)
C12N 9/22(2006.01)
C12N 15/864(2006.01)
A61K 48/00(2006.01)
A61P 21/00(2006.01)

权利要求书7页 说明书84页
序列表（电子公布） 附图42页

(54)发明名称

用于通过基因编辑修正人肌营养不良蛋白基因的治疗靶标和使用方法

(57)摘要

本文披露了用于通过基因编辑修正人肌营养不良蛋白基因的治疗靶标和使用方法。

CN 118147141 A

1.一种包含第一gRNA和第二gRNA的DNA靶向组合物,其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:

(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:1中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:2中列出的核苷酸序列编码;

(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码;和

(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码。

2.如权利要求1所述的DNA靶向组合物,进一步包含成簇规律间隔短回文重复序列相关(Cas)蛋白。

3.如权利要求2所述的DNA靶向组合物,其中该Cas蛋白包含识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的Cas9分子。

4.如权利要求2所述的DNA靶向组合物,其中该Cas蛋白包含具有SEQ ID NO:45的氨基酸序列的金黄色葡萄球菌Cas9分子。

5.如权利要求1所述的DNA靶向组合物,其中该DNA靶向组合物包含SEQ ID NO:83的核苷酸序列、SEQ ID NO:84的核苷酸序列、SEQ ID NO:37的核苷酸序列和/或SEQ ID NO:38的核苷酸序列。

6.一种分离的多核苷酸,其编码如权利要求1所述的DNA靶向组合物。

7.一种载体,编码:

(a)第一指导RNA(gRNA)分子,

(b)第二gRNA分子,以及

(c)至少一种识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的Cas9分子,

其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:

(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:1中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:2中列出的核苷酸序列编码;

(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码;和

(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码。

8.如权利要求7所述的载体,其中该载体被配置为形成在人DMD基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂。

9.如权利要求7所述的载体,其中该载体是病毒载体。

10.如权利要求9所述的载体,其中该载体是腺相关病毒(AAV)载体。

11.如权利要求10所述的载体,其中该AAV载体是AAV8载体或AAV9载体。

12.如权利要求7所述的载体,其中该载体包含可操作地连接至编码该第一gRNA分子、

该第二gRNA分子和/或该Cas9分子的核苷酸序列的组织特异性启动子。

13.如权利要求12所述的载体,其中该组织特异性启动子是肌肉特异性启动子。

14.如权利要求7所述的载体,其中该载体包含SEQ ID NO:39或SEQ ID NO:40中列出的核苷酸序列。

15.一种用于缺失包含外显子51的肌营养不良蛋白基因区段的组合物,该组合物包含:

(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或

NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第一Cas9分子的多核苷酸序列,以及

(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:

25)的原型间隔子邻近基序(PAM)的第二Cas9分子的多核苷酸序列,

其中该第一载体和第二载体被配置为分别形成在人DMD基因的外显子51侧翼的第一内含子和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段,并且其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:

(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:1中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:2中列出的核苷酸序列编码;

(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码;和

(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构

域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码。

16.如权利要求15所述的组合物,其中该区段具有50个碱基对的长度。

17.如权利要求16所述的组合物,其中该区段具有118个碱基对、233个碱基对、326个碱基对、766个碱基对、805个碱基对或1611个碱基对的长度。

18.如权利要求15所述的组合物,其中该第一Cas9分子和该第二Cas9分子是相同的。

19.如权利要求18所述的组合物,其中该第一Cas9分子和该第二Cas9分子是金黄色葡萄球菌Cas9分子。

20.如权利要求19所述的组合物,其中该第一Cas9分子和该第二Cas9分子是突变型金黄色葡萄球菌Cas9分子。

21.如权利要求15所述的组合物,其中该第一Cas9分子和该第二Cas9分子是不同的。

22.如权利要求21所述的组合物,其中该第一Cas9分子或该第二Cas9分子是金黄色葡萄球菌Cas9分子。

23.如权利要求15所述的组合物,其中该第一Cas9分子和/或该第二Cas9分子包含具有SEQ ID NO:45的氨基酸序列的SaCas9分子。

24.如权利要求15所述的组合物,其中该第一载体和/或该第二载体是病毒载体。

25.如权利要求24所述的组合物,其中该第一载体和/或该第二载体是腺相关病毒(AAV)载体。

26.如权利要求25所述的组合物,其中该AAV载体是AAV8载体或AAV9载体。

27.如权利要求15所述的组合物,其中该肌营养不良蛋白基因是人肌营养不良蛋白基因。

28.如权利要求15所述的组合物,其中该第一载体包含SEQ ID NO:39中列出的核苷酸序列,并且该第二载体包含SEQ ID NO:40中列出的核苷酸序列。

29.组合物在制备用于修正细胞中突变型肌营养不良蛋白基因的药物中的用途,所述组合物包含:

(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第一Cas9分子的多核苷酸序列,以及

(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第二Cas9分子的多核苷酸序列,

其中该载体被配置为分别形成在人肌营养不良蛋白基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段,并且其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:

(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:1中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:2中列出的核苷酸序列编码;

(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核

苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码;和

(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码。

30.如权利要求29所述的用途,其中该突变型肌营养不良蛋白基因包含提前终止密码子、被破坏的阅读框、异常剪接受体位点或异常剪接供体位点。

31.如权利要求29所述的用途,其中该突变型肌营养不良蛋白基因包含导致提前终止密码子和截短的基因产物的移码突变。

32.如权利要求29所述的用途,其中该突变型肌营养不良蛋白基因包含破坏阅读框的一个或多个外显子的缺失。

33.如权利要求29所述的用途,其中该突变型肌营养不良蛋白基因的修正包括提前终止密码子的缺失、被破坏的阅读框的修正或通过破坏剪接受体位点或破坏剪接供体序列对剪接进行的调节。

34.如权利要求29所述的用途,其中该突变型肌营养不良蛋白基因的修正包括外显子51的缺失。

35.如权利要求29所述的用途,其中该突变型肌营养不良蛋白基因的修正包括同源定向修复。

36.如权利要求35所述的用途,进一步包括向该细胞给予供体DNA。

37.如权利要求29所述的用途,其中该突变型肌营养不良蛋白基因的修正包括核酸酶介导的非同源末端连接。

38.如权利要求29所述的用途,其中该细胞是成肌细胞。

39.如权利要求29所述的用途,其中该细胞来自患有杜氏肌营养不良的受试者。

40.如权利要求29所述的用途,其中该细胞是来自患有杜氏肌营养不良的人类受试者的成肌细胞。

41.组合物在制备用于治疗具有突变型肌营养不良蛋白基因的有需要的受试者的药物中的用途,所述组合物包含:

(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第一Cas9分子的多核苷酸序列,以及

(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第二Cas9分子的多核苷酸序列,

其中该载体被配置为分别形成在人肌营养不良蛋白基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段,并且其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:

(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:1中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:2中列出的核苷酸序列编码;

(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码;

(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:4中列出的核苷酸序列编码;

(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:19中列出的核苷酸序列编码;

(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构

域由SEQ ID NO:14中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:15中列出的核苷酸序列编码;和

(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域由SEQ ID NO:11中列出的核苷酸序列编码,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域由SEQ ID NO:18中列出的核苷酸序列编码。

42.如权利要求41所述的用途,其中该受试者患有杜氏肌营养不良。

43.如权利要求41所述的用途,其中将该第一载体和第二载体肌内、静脉内或其组合给予该受试者。

用于通过基因编辑修正人肌营养不良蛋白基因的治疗靶标和使用方法

相关申请的交叉引用

[0001] 本申请要求2015年11月30日提交的美国临时申请号62/260,712和2016年5月2日提交的美国临时申请号62/330,336的优先权,将其全部通过引用以其全部内容并入本文。

政府利益的声明

[0002] 本发明是在国家科学基金会研究生研究奖学金项目的奖励下在政府支持下完成的。美国政府对本发明享有一定的权利。

技术领域

[0003] 本披露涉及使用基于成簇规律间隔短回文重复序列(CRISPR)/CRISPR相关(Cas)9的系统和病毒递送系统的基因表达改变、基因组工程和基因的基因组改变的领域。本披露还涉及肌肉(如骨骼肌和心肌)中基因的基因组工程和基因组改变的领域。

背景技术

[0004] 合成转录因子已被工程化为控制基因表达用于哺乳动物系统中许多不同医学和科学应用,包括刺激组织再生、药物筛选、补偿遗传缺陷、激活沉默肿瘤抑制因子、控制干细胞分化、进行遗传筛选、以及创造合成基因电路。这些转录因子可以靶向内源基因的启动子或增强子,或有目的地设计为识别与哺乳动物基因组正交的序列以进行转基因调控。用于工程化靶向用户定义序列的新颖转录因子的最常见策略是基于锌指蛋白和转录激活因子样效应子(TALE)的可编程DNA结合结构域。这两种方法都涉及应用这些结构域的蛋白质-DNA相互作用的原理来工程化具有独特DNA结合特异性的新蛋白。尽管这些方法在许多应用中已经取得了广泛的成功,但操纵蛋白质-DNA相互作用所必需的蛋白质工程可能很费力并且需要专业知识。

[0005] 此外,这些新蛋白并不总是有效。造成这种情况的原因尚不清楚,但可能与表观遗传修饰和染色质状态对蛋白质与基因组靶位点结合的影响有关。此外,在确保将这些新蛋白以及其他组分递送至每个细胞方面存在挑战。用于递送这些新蛋白及其多种组分的现有方法包括在单独的质粒或载体上递送至细胞,由于拷贝数的差异这导致每个细胞中高度变化的表达水平。此外,转染后的基因激活由于质粒DNA的稀释而是短暂的,并且临时基因表达可能不足以诱导治疗效果。此外,这种方法不适用于不容易转染的细胞类型。因此,这些新蛋白的另一个限制是转录激活的效力。

[0006] 基于CRISPR/Cas9的基因编辑系统可以用于在靶向基因座处引入位点特异性双链断裂。这种DNA裂解刺激天然DNA修复机器,导致两种可能的修复途径之一。在不存在供体模板的情况下,断裂将通过非同源末端连接(NHEJ)修复,这是一种导致DNA的小插入或缺失的易错的修复途径。此方法可以用于有意地破坏、缺失或改变靶基因序列的阅读框。然而,如果供体模板与核酸酶一起提供,则细胞机器将通过同源重组修复断裂,在存在DNA裂解的情况下该修复增强几个数量级。此方法可以用于在DNA序列中在靶位点处引入特定变化。工程

化的核酸酶已被用于多种人干细胞和细胞系中的基因编辑以及小鼠肝脏中的基因编辑。然而,实施这些技术的主要障碍是以有效、高效并有助于成功进行基因组修饰的方式体内递送至特定组织。

[0007] 遗传性的遗传疾病对美国的儿童具有破坏性影响。这些疾病目前无法治愈,并且只能通过试图缓解症状来进行管理。几十年来,基因治疗领域已经承诺治愈这些疾病。然而,关于安全且高效地将治疗基因递送至细胞和患者的技术障碍限制了这种方法。杜氏肌营养不良(DMD)是一种致命的遗传疾病,临床特征为由于功能性肌营养不良蛋白的丧失,典型地在生命的第三个十年中出现肌肉萎缩、无法步行和死亡。DMD的致因是在肌营养不良蛋白(dystrophin)基因中的遗传或自发突变。导致DMD的大多数突变是一个或多个外显子缺失的结果,推动翻译阅读框脱框。

[0008] 肌营养不良蛋白是负责调控肌细胞完整性和功能的蛋白质复合体的关键组分。DMD患者典型地在童年时期失去了自我身体支撑的能力,在青少年时期逐渐变得更弱,并且在二十多岁时死亡。当前用于DMD的实验性基因治疗策略需要重复给予瞬时基因递送运载体或依赖于将外源遗传物质永久整合到基因组DNA中。这两种方法都有严重的安全问题。此外,这些策略受到无法递送大且复杂的肌营养不良蛋白基因序列的限制。仍需要更精确且有效的基因编辑工具来修正和治疗在肌营养不良蛋白基因中具有突变的患者。

发明内容

[0009] 本发明涉及包含靶向结构域的指导RNA(gRNA),该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:41、SEQ ID NO:42中列出的核苷酸序列或其互补序列。

[0010] 本发明还涉及包含第一gRNA和第二gRNA的DNA靶向组合物。第一gRNA分子和第二gRNA分子包含靶向结构域,该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ IDNO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:41、SEQ ID NO:42中列出的核苷酸序列或其互补序列。第一gRNA分子和第二gRNA分子包含不同的靶向结构域。

[0011] 本发明还涉及包含上述gRNA分子或上述DNA靶向组合物的分离的多核苷酸。

[0012] 本发明涉及包含上述gRNA、上述DNA靶向组合物或上述分离的多核苷酸的载体。

[0013] 本发明还涉及包含上述DNA靶向组合物的载体。

[0014] 本发明还涉及如下载体,其编码:(a)第一指导RNA(gRNA)分子,(b)第二gRNA分子和(c)至少一种识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的Cas9分子。第一gRNA分子和第二gRNA分子包含靶向结构域,该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:41、SEQ ID NO:42中列出的核苷酸序列或其互补序列。第一gRNA分子和

第二gRNA分子包含不同的靶向结构域。

[0015] 本发明还涉及包含上述gRNA、上述DNA靶向组合物、上述分离的多核苷酸或上述载体的细胞。

[0016] 本发明还涉及包含上述gRNA、上述DNA靶向系统、上述分离的多核苷酸、上述载体或上述细胞和任选地使用说明书的试剂盒。

[0017] 本发明还涉及修正细胞中突变型肌营养不良蛋白基因的方法。该方法包括向细胞给予上述gRNA、上述DNA靶向系统、上述分离的多核苷酸或上述载体。

[0018] 本发明还涉及基因组编辑受试者中突变型肌营养不良蛋白基因的方法。该方法包括向该受试者给予包含上述gRNA、上述DNA靶向系统、上述分离的多核苷酸、上述载体或上述细胞的基因组编辑组合物。

[0019] 本发明还涉及治疗具有突变型肌营养不良蛋白基因的有需要的受试者的方法。该方法包括向该受试者给予上述gRNA、上述DNA靶向系统、上述分离的多核苷酸、上述载体或上述细胞。

[0020] 本发明还涉及用于基因组编辑受试者中突变型肌营养不良蛋白基因的经修饰腺相关病毒载体,其包含编码上述gRNA的第一多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的Cas9分子的第二多核苷酸序列。

[0021] 本发明还涉及用于缺失包含外显子51的肌营养不良蛋白基因区段的组合物,该组合物包含:(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第一Cas9分子的多核苷酸序列,以及(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第二Cas9分子的多核苷酸序列。第一和第二gRNA分子各自具有长度为19至24个核苷酸的靶向结构域,并且其中第一载体和第二载体被配置为分别形成在人DMD基因的外显子51侧翼的第一内含子和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段。

[0022] 本发明还涉及包含上述组合物的细胞。

[0023] 本发明还涉及修正细胞中突变型肌营养不良蛋白基因的方法,该方法包括向该细胞给予:(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第一Cas9分子的多核苷酸序列,以及(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第二Cas9分子的多核苷酸序列。第一gRNA和第二gRNA分子各自具有长度为19至24个核苷酸的靶向结构域,并且其中载体被配置为分别形成在人肌营养不良蛋白基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段。

[0024] 本发明还涉及治疗具有突变型肌营养不良蛋白基因的有需要的受试者的方法。该方法包括向该受试者给予:(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序

(PAM)的第一Cas9分子的多核苷酸序列,以及(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第二Cas9分子的多核苷酸序列。第一gRNA和第二gRNA分子各自具有长度为19至24个核苷酸的靶向结构域,并且其中载体被配置为分别形成在人肌营养不良蛋白基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段。

[0025] 本发明还涉及产生具有外显子52缺失(Δ52)的人肌营养不良蛋白基因(hDMD)的转基因啮齿动物胚胎的方法。该方法包括向啮齿动物胚胎给予上述gRNA、上述DNA靶向系统、上述分离的多核苷酸、上述载体、上述经修饰腺相关病毒载体或上述组合物,从而缺失人肌营养不良蛋白基因的外显子52,并选择具有人肌营养不良蛋白基因外显子52缺失的转基因啮齿动物胚胎,其中该啮齿动物胚胎包含正常的人肌营养不良蛋白基因。

[0026] 本发明还涉及通过上述方法产生的转基因啮齿动物胚胎。

[0027] 本发明还涉及由上述转基因啮齿动物胚胎产生的转基因啮齿动物。

附图说明

[0028] 图1显示了靶向HEK293T细胞(野生型肌营养不良蛋白基因)和DMD患者成肌细胞系(DMD 8036和DMD 6594,其各自具有肌营养不良蛋白基因的突变形式)中的人肌营养不良蛋白基因的单独gRNA JCR89和JCR91的活性,如通过Surveyor测定确定的。

[0029] 图2A和2B显示了通过用SaCas9和gRNA JCR89和JCR91共处理,HEK293T细胞和DMD成肌细胞(DMD 8036和DMD6594)的基因组DNA(图2A)和来自DMD成肌细胞的cDNA(图2B)中的外显子51的缺失。

[0030] 图3显示了用于将两种病毒载体上的SaCas9和gRNA JCR89和JCR91共同递送至肌肉组织的基于AAV的体内系统。

[0031] 图4显示了在将携带SaCas9和gRNA的病毒载体局部AAV8递送至胫骨前肌(TA)肌肉之后,携带人DMD基因的转基因小鼠(hDMD/mdx小鼠)中人外显子51缺失的检测。

[0032] 图5显示了在通过尾静脉注射进行的系统性AAV8递送之后,携带hDMD基因的转基因小鼠中人外显子51缺失的检测。

[0033] 图6显示了人和恒河猴基因组之间保守的不同gRNA靶标(参见表2中的gRNA的序列)。每种gRNA的位置都关于人肌营养不良蛋白基因的外显子51来指示。

[0034] 图7显示了如通过Surveyor测定确定的在转染人HEK293T细胞后单独gRNA的活性。

[0035] 图8显示了如使用CasOFFinder程序(Bae et al.(2014)Bioinformatics[Bae等人(2014)生物信息学]30:1473-1475)预测的候选gRNA的特异性。

[0036] 图9显示了如通过基因组DNA的PCR确定的HEK293T细胞和DMD 6594细胞中gRNA JCR157和JCR160引起的外显子51的缺失。

[0037] 图10显示了如通过Surveyor测定确定的各种靶标长度的gRNA JCR157的活性:19、20、21、22和23个核苷酸。

[0038] 图11显示了如通过Surveyor测定确定的各种靶标长度的gRNA JCR160的活性:19、20、21、22和23个核苷酸。

[0039] 图12显示了如通过基因组DNA的PCR确定的通过组合各种长度(21、22或23个核苷

酸)的JCR157和JCR160产生的缺失。

[0040] 图13显示了如通过Surveyor测定测得的体外中靶核酸酶活性。

[0041] 图14显示了基因组DNA中外显子51的体外缺失。

[0042] 图15显示了分化7天的人DMD成肌细胞的cDNA中外显子51的体外缺失。

[0043] 图16显示了DMD患者成肌细胞的cDNA中体外外显子47至52的连接。

[0044] 图17显示了从健康hDMD/mdx小鼠开始的Δ52/mdx小鼠的设计。

[0045] 图18显示了体外指导物验证:单独的(Surveyor测定)。

[0046] 图19显示了体外指导物验证:成对的:使用gRNA对产生HEK293T细胞的基因组DNA中外显子51的缺失。

[0047] 图20显示了DNA显微注射方案的示意图。

[0048] 图21示出了小鼠繁殖的示意图。

[0049] 图22显示了首建小鼠基因分型结果。

[0050] 图23显示了首建小鼠7、63和76的测序结果的一部分。

[0051] 图24显示了进一步的小鼠繁殖的示意图。

[0052] 图25显示了来自首建雄鼠76+mdx/mdx繁殖结果的第5窝(仅雄鼠)的基因分型。“63”是首建雄鼠(但在这种情况下不是亲本)。“293”代表HEK293T细胞基因组DNA对照。

[0053] 图26显示了来自首建雄鼠63+mdx/mdx繁殖结果的第1窝的基因分型。

[0054] 图27显示了幼鼠54497和54498的392bp测序读数的一部分。

[0055] 图28显示了来自幼鼠54497和54498的心脏和TA的免疫组织化学染色。

[0056] 图29显示Δ52/mdx小鼠缺乏肌营养不良蛋白。

[0057] 图30显示了蛋白质印迹,指示Δ52/mdx小鼠缺乏与DMD基因型一致的肌营养不良蛋白,而健康的hDMD/mdx小鼠表达肌营养不良蛋白。

[0058] 图31显示了如通过运动和探索所示的,Δ52/mdx小鼠与mdx小鼠和hDMD/mdx小鼠相比的整体活动。

[0059] 图32显示了使用SaCas9和gRNA,通过靶向要去除的内含子区域中的外显子51的上游和下游的gRNA来跳过外显子51的针对Δ52/mdx小鼠的修正策略。

[0060] 图33显示了使用SaCas9以及gRNA JCR179和JCR183,在DMD患者成肌细胞(DMD 6594细胞)中从外显子51缺失体外恢复肌营养不良蛋白。

[0061] 图34显示了使用gRNA和SaCas9系统处理Δ52/mdx小鼠的实验设计。

[0062] 图35显示了右侧TA肌肉中的体内外显子51缺失。

[0063] 图36显示了右侧TA肌肉中的体内外显子51缺失。

[0064] 图37显示了经处理的TA肌肉中的体内肌营养不良蛋白恢复。

[0065] 图38显示了经处理的TA肌肉中的体内肌营养不良蛋白恢复。

[0066] 图39显示了总移动距离的所有时间点的平均值。

[0067] 图40显示了总饲养姿势(rearing posture)的所有时间点的平均值。

[0068] 图41显示了16周的未经处理和经处理的小鼠的握力。

[0069] 图42显示了心脏组织的cDNA PCR结果。

[0070] 图43显示了来自图42的经扩增的cDNA PCR带的测序。

具体实施方式

[0071] 如本文所述,已经发现某些方法和工程化的gRNA可以与基于CRISPR/CRISPR相关(Cas)9的基因编辑系统一起用于改变遗传疾病(例如,DMD)中涉及的肌营养不良蛋白基因的表达、基因组工程和修正或减少其突变的影响。产生所披露的gRNA以靶向更适于临床翻译的位点。例如,编码化脓链球菌Cas9(SpCas9)的基因太大而无法通过腺相关病毒(AAV)递送,腺相关病毒是当包括所有其他必需调控序列时用于向肌肉递送系统性基因的载体。作为替代,选择并筛选所披露的gRNA与比SpCas9小约1kb的金黄色葡萄球菌Cas9(SaCas9)一起使用。针对在人和恒河猴基因组之间保守的序列上为SaCas9相容靶标来筛选靶标选择,这极大地限制了可能的基因靶标的数目。此选择标准被选取为允许在人和恒河猴中都可能有活性的gRNA候选物,以便于在非人灵长类动物模型中进行临床前测试。所披露的靶向人和恒河猴两者的肌营养不良蛋白基因序列的gRNA可以与基于CRISPR/Cas9的系统一起用于靶向人肌营养不良蛋白基因的外显子51周围的内含子区域,导致此区域的基因组缺失以恢复来自DMD患者的细胞中功能性肌营养不良蛋白的表达。

[0072] 本文还描述了用于递送基于CRISPR/Cas9的基因编辑系统和多种gRNA以靶向肌营养不良蛋白基因的基因构建体、组合物和方法。本发明披露的主题还提供了用于将基因构建体(例如,载体)或包含其的组合物递送至骨骼肌和心肌的方法。载体可以是AAV,包括经修饰的AAV载体。本发明披露的主题描述了将这类治疗剂的活性形式递送至骨骼肌或心肌的方式,其是有效、高效并有助于成功进行基因组修饰,连同提供了重写人类基因组用于治疗应用并且靶向模型物种用于基础科学应用的手段。

[0073] 如在本节和本文的整个披露中所用的节标题只是出于组织的目的,而不旨在进行限制。

1.定义

[0074] 除非另外定义,否则本文使用的所有技术和科学术语具有与本领域普通技术人员通常所理解的相同的含义。在矛盾的情况下,应以本文件(包括定义)为准。虽然与本文描述的那些方法和材料类似或等效的方法和材料可以用于本发明的实践或测试中,但以下描述优选的方法和材料。本文提及的所有出版物、专利申请、专利以及其他参考文献通过引用以其全部内容并入本文。本文披露的材料、方法以及实例仅是说明性的并且不旨在是限制性的。

[0075] 如本文所用,术语“包含(comprise(s))”、“包括(include(s))”、“具有(having)”、“有(has)”、“可以(can)”、“含有(contain(s))”及其变体旨在是不排除另外的行为或结构的可能性的开放性的过渡短语、术语或词语。单数形式“一个/一种(a/an)”和“该(the)”包括复数个指示物,除非上下文中另外明确指明。本披露还涵盖“包括本文所呈现的实施例或要素”、“由本文所呈现的实施例或要素组成”以及“基本上由本文所呈现的实施例或要素组成”的其他实施例,无论是否明确地提出。

[0076] 对于本文数值范围的叙述,明确考虑了具有相同精度的介于两者之间的每个介于中间的数字。例如,对于范围6-9,除了6和9之外还考虑了数字7和8,并且对于范围6.0-7.0,明确考虑了数字6.0、6.1、6.2、6.3、6.4、6.5、6.6、6.7、6.8、6.9和7.0。

[0077] 如本文所用,术语“约(about)”或“大约(approximately)”意指由本领域普通技术人员确定的特定值在可接受的误差范围之内,这将部分地取决于该值是怎样测定或确定

的,即,受到测量系统的限制。例如,“约”可以意指根据本领域的实践在3个或大于3个标准差之内。可替代地,“约”可以意指给定值的高达20%、优选高达10%、更优选高达5%并且还更优选高达1%的范围。可替代地,特别是对于生物学系统和方法,该术语可以意指在某一值的数量级内,优选在5倍内并且更优选在2倍内。

[0078] 如在本文中可互换使用的“腺相关病毒”或“AAV”是指感染人及其他一些灵长类物种的属于细小病毒科(Parvoviridae)依赖病毒属(Dependovirus)的小病毒。当前尚不知AAV引起疾病并且因此该病毒引起非常轻微的免疫应答。

[0079] 如本文所用,“结合区”是指在核酸靶区域内被核酸酶识别并结合的区域。

[0080] 如在本文中可互换使用的“心肌(cardiac muscle)”或“心肌(heart muscle)”意指心脏的壁和组织学基础中发现的一种类型的不随意横纹肌(involuntary striated muscle),即心肌(myocardium)。心肌由心肌细胞(cardiomyocyte或myocardiocyte)组成。心肌细胞显示与骨骼肌细胞相似的条纹,但只含有一个独特的核,与多核骨骼细胞不同。在某些实施例中,“心肌病症”是指与心肌有关的病症,如心肌病、心力衰竭、心律失常和炎症性心脏病。

[0081] 如本文所用,“编码序列”或“编码核酸”意指包含编码蛋白质的核苷酸序列的核酸(RNA或DNA分子)。编码序列可以进一步包括可操作地连接至调控元件的起始信号和终止信号,这些调控元件包括能够在向其给予核酸的个体或哺乳动物的细胞中指导核酸的表达的启动子和聚腺苷酸化信号。编码序列可以是密码子优化的。

[0082] 如本文所用,意指核酸的“互补”或“互补的”可以意指核酸分子的核苷酸或核苷酸类似物之间的沃森-克里克(例如,AT/U和C-G)或胡斯坦(Hoogsteen)碱基配对。“互补性”是指两个核酸序列之间共有的特性,以使得当它们彼此反平行比对时,每个位置上的核苷酸碱基将是互补的。

[0083] 如本文所用,“修正”、“基因组编辑”和“恢复”是指改变编码截短蛋白或根本不编码蛋白的突变基因,以使得获得全长功能性或部分全长功能性蛋白的表达。修正或恢复突变基因可以包括利用修复机制如同源定向修复(HDR)用不具有突变的基因拷贝替换具有突变的基因区域或替换整个突变基因。修正或恢复突变基因还可以包括通过以下方式修复引起提前终止密码子、异常剪接受体位点或异常剪接供体位点的移码突变:在基因中产生双链断裂,然后使用非同源末端连接(NHEJ)进行修复。NHEJ可以在修复过程中添加或缺失至少一个碱基对,从而可以恢复正确的阅读框并消除提前终止密码子。修正或恢复突变基因还可以包括破坏异常剪接受体位点或剪接供体序列。修正或恢复突变基因还可以包括通过两种核酸酶在同一DNA链上的同时作用来缺失非必需基因区段,以便通过去除两个核酸酶靶位点之间的DNA并修复由NHEJ产生的DNA断裂来恢复正确的阅读框。

[0084] 如在本文中可互换使用的“供体DNA”、“供体模板”和“修复模板”是指包括感兴趣的基因的至少一部分的双链DNA片段或分子。供体DNA可以编码全功能性蛋白或部分功能性蛋白。

[0085] 如在本文中可互换使用的“杜氏肌营养不良”或“DMD”是指一种隐性遗传的、致命的、X连锁病症,其导致肌肉变性和最终的死亡。DMD是一种常见的遗传性单基因疾病,并且在3500名男性中发生1例。DMD的致因是在肌营养不良蛋白基因中引起无义或移码突变的遗传或自发突变。引起DMD的大多数肌营养不良蛋白突变是破坏阅读框并且引起肌营养不良

蛋白基因的提前翻译终止的外显子缺失。DMD患者典型地在童年时期失去了自我身体支撑的能力,在青少年时期逐渐变得更弱,并且在二十多岁时死亡。

[0086] 如本文所用,“肌营养不良蛋白”是指棒状胞质蛋白,其是通过细胞膜将肌纤维细胞骨架连接到周围的细胞外基质的蛋白质复合体的一部分。肌营养不良蛋白提供负责调控肌细胞完整性和功能的细胞膜肌营养不良蛋白聚糖复合体的结构稳定性。如在本文中可互换使用的肌营养不良蛋白基因或“DMD基因”是在基因座Xp21处的2.2兆碱基。初级转录测量了约2,400kb,其中成熟mRNA是约14kb。79个外显子编码超过3500个氨基酸的蛋白质。

[0087] 如本文所用,“外显子51”是指肌营养不良蛋白基因的第51个外显子。在DMD患者中,外显子51常常邻近破坏框的缺失并且已在临床试验中被靶向用于基于寡核苷酸的外显子跳跃。外显子51跳跃化合物依替利森(eteplirsen)的临床试验最近报道了跨48周的显著功能益处,与基线相比具平均47%的肌营养不良蛋白阳性纤维。外显子51中的突变理想地适合于通过基于NHEJ的基因组编辑进行永久性修正。

[0088] 如在本文中可互换使用的“移码”或“移码突变”是指一种类型的基因突变,其中一个或多个核苷酸的添加或缺失引起mRNA中密码子的阅读框移位。阅读框的移位可能导致蛋白质翻译时氨基酸序列改变,如错义突变或提前终止密码子。

[0089] 如本文所用,“功能性”和“全功能性”描述了具有生物活性的蛋白质。“功能性基因“是指转录成mRNA的基因,mRNA被翻译成功能性蛋白。

[0090] 如本文所用,“融合蛋白”是指通过最初编码单独蛋白质的两个或更多个基因的连接产生的嵌合蛋白。融合基因的翻译产生具有来源于每种原始蛋白质的功能特性的单一多肽。

[0091] 如本文所用,“基因构建体”是指包含编码蛋白质的核苷酸序列的DNA或RNA分子。编码序列包括可操作地连接至调控元件的起始信号和终止信号,这些调控元件包括能够在向其给予核酸分子的个体的细胞中指导核酸的表达的启动子和聚腺苷酸化信号。如本文所用,术语“可表达形式”是指如下基因构建体,该基因构建体含有可操作连接至编码蛋白质的编码序列的必需调控元件,使得当存在于个体的细胞中时,编码序列将被表达。

[0092] 如本文所用,“遗传疾病”是指部分或完全、直接或间接地由基因组中的一个或多个异常引起的疾病,特别是从出生就存在的病状。异常可以是突变、插入或缺失。异常可以影响基因或其调控序列的编码序列。遗传疾病可以是但不限于DMD、贝克肌营养不良(BMD)、血友病、囊性纤维化、亨廷顿氏舞蹈病、家族性高胆固醇血症(LDL受体缺陷)、肝母细胞瘤、威尔逊氏病、先天性肝性卟啉症、遗传性肝性代谢病症、莱施尼汉综合征(Lesch Nyhan syndrome)、镰状细胞性贫血、地中海贫血、着色性干皮病、范可尼氏贫血(Fanconi′s anemia)、色素性视网膜炎、共济失调毛细血管扩张、布卢姆综合征(Bloom′s syndrome)、成视网膜细胞瘤以及泰-萨二氏病(Tay-Sachs disease)。

[0093] 如在本文中可互换使用的“同源定向修复”或“HDR”是指在细胞中当DNA的同样片段存在于细胞核时(主要在细胞周期的G2和S期)修复双链DNA损伤的机理。HDR使用供体DNA模板来指导修复,并且可以用于产生基因组的特定序列变化,包括整个基因的靶向添加。如果供体模板与基于CRISPR/Cas9的基因编辑系统一起提供,则细胞机构将通过同源重组修复断裂,在存在DNA裂解的情况下该修复增强几个数量级。当不存在同源DNA片段时,可以作为替代发生非同源末端连接。

[0094] 如本文所用,“基因组编辑”是指改变基因。基因组编辑可以包括修正或恢复突变基因。基因组编辑可以包括敲除基因,如突变基因或正常基因。基因组编辑可以通过改变感兴趣的基因用于治疗疾病或增强肌肉修复。

[0095] 如本文所用,“相同”或“同一性”在两个或更多个核酸或多肽序列的背景下意指序列在指定区域上具有指定百分比的相同残基。百分比可以通过以下方式计算:最佳比对两个序列,在指定区域上比较两个序列,确定两个序列中出现相同残基的位置数以产生匹配位置数,将匹配位置数除以指定区域中的位置总数,并且将结果乘以100而得出序列同一性百分比。在两个序列具有不同长度或者比对产生一个或多个交错末端并且比较的指定区域仅包括单个序列的情况下,单个序列的残基包括在计算的分母中,而不包括在计算的分子中。当比较DNA和RNA时,可以认为胸腺嘧啶(T)和尿嘧啶(U)是等同的。可以手动或通过使用计算机序列算法如BLAST或BLAST2.0来获得同一性。

[0096] 如在本文中可互换使用的“突变基因“或“突变的基因“是指已经经历可检测的突变的基因。突变基因经历了改变,如遗传物质的丢失、获得或交换,该改变影响基因的正常传递和表达。如本文所用,“破坏的基因”是指具有引起提前终止密码子的突变的突变基因。破坏的基因产物相对于全长未破坏的基因产物是截短的。

[0097] 如本文所用,“非同源末端连接(NHEJ)途径”是指通过在不需要同源模板的情况下直接连接断裂末端来修复双链断裂的途径。NHEJ对DNA末端的不依赖模板的重新连接是随机的、易错的修复过程,该过程在DNA断点处引入随机微插入和微缺失(插入缺失(indel))。此方法可以用于有意地破坏、缺失或改变靶基因序列的阅读框。NHEJ通常使用短同源DNA序列(称为微同源物)指导修复。这些微同源物常常存在于双链断裂末端上的单链突出端中。当突出端完全相容时,NHEJ通常准确地修复断裂,然而导致核苷酸丢失的不精确修复也可能发生,但是当突出端不相容时不精确修复更普遍得多。

[0098] 如本文所用,“正常基因”是指未经历改变(如遗传物质的丢失、获得或交换)的基因。正常基因经历正常的基因传递和基因表达。

[0099] 如本文所用,“核酸酶介导的NHEJ”是指核酸酶如cas9分子切割双链DNA后开始的NHEJ。

[0100] 如本文所用,“核酸”或“寡核苷酸”或“多核苷酸”意指共价连接在一起的至少两个核苷酸。单链的描绘也定义了互补链的序列。因此,核酸也涵盖所描绘的单链的互补链。核酸的许多变体可以用于与给定核酸相同的目的。因此,核酸也涵盖基本相同的核酸及其互补序列。单链提供可以在严格杂交条件下与靶序列杂交的探针。因此,核酸还涵盖在严格杂交条件下杂交的探针。

[0101] 核酸可以是单链或双链的,或者可以含有双链和单链序列两者的部分。核酸可以是基因组和cDNA的DNA、RNA或杂交体,其中核酸可以含有脱氧核糖核苷酸和核糖核苷酸的组合,以及碱基的组合,这些碱基包括尿嘧啶、腺嘌呤、胸腺嘧啶、胞嘧啶、鸟嘌呤、肌苷、黄嘌呤、次黄嘌呤、异胞嘧啶和异鸟嘌呤。核酸可以通过化学合成方法或通过重组方法来获得。

[0102] 如本文所用,“可操作地连接”意指基因的表达在与之在空间上连接的启动子的控制下。启动子可以位于在其控制下的基因的5’(上游)或3’(下游)。启动子与基因之间的距离可以与该启动子与它控制的基因(在启动子所来源的基因中)之间的距离大致相同。正如

本领域已知,可以调节此距离的变化而不丧失启动子功能。

[0103] 如本文所用,“部分功能性”描述由突变基因编码并具有比功能性蛋白低但比非功能性蛋白高的生物活性的蛋白质。

[0104] 如在本文中可互换使用的“提前终止密码子”或“框外终止密码子”是指DNA序列的无义突变,其导致通常不存在于野生型基因中的位置处的终止密码子。提前终止密码子可能引起蛋白质被截短或短于蛋白质的全长型式。

[0105] 如本文所用,“启动子”是指合成的或天然来源的分子,其能够赋予、激活或增强细胞中核酸的表达。启动子可以包含一个或多个特异性转录调控序列以进一步增强表达和/或改变该序列的空间表达和/或时间表达。启动子还可以包含远端增强子或阻遏物元件,其可以定位于距转录起始位点多达几千个碱基对。启动子可以来源于以下来源:包括病毒、细菌、真菌、植物、昆虫和动物。就其中发生表达的细胞、组织或器官而言,或者就发生表达时的发育阶段而言,或者在响应于外部刺激如生理应激、病原体、金属离子或诱导剂时,启动子可以组成型地或差异性地调控基因组分的表达。启动子的代表性实例包括噬菌体T7启动子、噬菌体T3启动子、SP6启动子、乳糖操纵基因启动子、tac启动子、SV40晚期启动子、SV40早期启动子、RSV-LTR启动子、CMV IE启动子、SV40早期启动子或SV40晚期启动子、人U6(hU6)启动子和CMV IE启动子。

[0106] 如本文所用,“骨骼肌”是指一种类型的横纹肌,其在躯体神经系统的控制下,并通过称为肌腱的胶原纤维束与骨连接。骨骼肌由称为肌细胞或“肌肉细胞”(有时通称为“肌纤维”)的各组分组成。肌细胞在称为肌生成的过程中由发育的成肌细胞(一种类型的产生肌肉细胞的胚胎祖细胞)融合形成。这些长的圆柱形多核细胞也称为肌纤维。

[0107] 如本文所用,“骨骼肌病状”是指与骨骼肌有关的病状,如肌营养不良、老化、肌肉变性、伤口愈合和肌无力或萎缩。

[0108] 如在本文中可互换使用的“受试者”和“患者”是指任何脊椎动物,包括但不限于哺乳动物(例如,牛、猪、骆驼、美洲驼、马、山羊、兔、绵羊、仓鼠、豚鼠、猫、狗、大鼠和小鼠、非人灵长类动物(例如,猴,如食蟹猴(cynomolgous)或恒河猴、黑猩猩等)和人)。在一些实施例中,受试者可以是人或非人。受试者或患者可能正在接受其他形式的治疗。

[0109] 如本文所用,“靶基因”是指编码已知或推定的基因产物的任何核苷酸序列。靶基因可以是遗传疾病中涉及的突变的基因。在某些实施例中,靶基因是人肌营养不良蛋白基因。在某些实施例中,靶基因是突变型人肌营养不良蛋白基因。

[0110] 如本文所用,“靶区域”是指被设计为基于CRISPR/Cas9的基因编辑系统结合和裂解的靶基因的区域。

[0111] 如本文所用,“转基因“是指含有从一种生物体分离并引入不同生物体中的基因序列的基因或遗传物质。DNA的这种非天然区段可以保留在转基因生物体中产生RNA或蛋白质的能力,或可以改变转基因生物体的遗传密码的正常功能。转基因的引入具有改变生物体的表型的潜力。

[0112] 本文关于核酸使用的“变体”意指(i)参比核苷酸序列的一部分或片段;(ii)参比核苷酸序列的互补序列或其部分;(iii)与参比核酸基本上相同的核酸或其互补序列;或(iv)在严格条件下与参比核酸、其互补序列或与其基本上相同的序列杂交的核酸。

[0113] 有关肽或多肽的“变体”因氨基酸的插入、缺失或保守取代而在氨基酸序列上不同

但保留至少一种生物活性。变体还可以意指具有如下氨基酸序列的蛋白质,该氨基酸序列与具有保留至少一种生物活性的氨基酸序列的参比蛋白质基本相同。氨基酸的保守取代(即,将氨基酸用具有相似特性(例如,亲水性、带电荷区域的程度和分布)的不同氨基酸替换)在本领域中被视为典型涉及微小变化。正如本领域所了解的,通过考虑氨基酸的亲水指数可以部分地鉴定这些微小变化。Kyte et al.,J.Mol.Biol.[Kyte等人,分子生物学杂志]157:105-132(1982)。氨基酸的亲水指数是基于其疏水性和电荷的考量。本领域已知,具有类似亲水指数的氨基酸可以被取代并仍保留蛋白质功能。在一方面,具有±2的亲水指数的氨基酸是取代的。氨基酸的亲水性还可以用于揭示将产生保留生物功能的蛋白质的取代。在肽的背景下考虑氨基酸的亲水性允许计算该肽的最大局部平均亲水性。取代可以用彼此亲水性值在±2内的氨基酸进行。氨基酸的疏水性指数和亲水性值两者受氨基酸的特定侧链影响。应了解,与该观察结果一致,与生物功能相容的氨基酸取代取决于如通过疏水性、亲水性、电荷、大小和其他性质揭示的氨基酸、并且特别是那些氨基酸的侧链的相对相似性。

[0114] 如本文所用,“载体”意指含有复制起点的核酸序列。载体可以是病毒载体、噬菌体、细菌人工染色体或酵母人工染色体。载体可以是DNA或RNA载体。载体可以是自我复制的染色体外载体,并且优选地是DNA质粒。例如,载体可以编码Cas9蛋白和至少一种gRNA分子,如包含SEQ ID NO:1-19、41、42中任一个或其互补序列的靶向结构域的gRNA。在一些实施例中,Cas9蛋白可以具有SEQ ID NO:27、SEQ ID NO:33或SEQ ID NO:45的氨基酸序列。在一些实施例中,Cas9蛋白可以是金黄色葡萄球菌Cas9,如具有SEQ ID NO:33或45的氨基酸序列的SaCas9。在一些实施例中,Cas9蛋白是由如下核酸序列编码,该核酸序列包含SEQ ID NO:26、SEQ ID NO:28、SEQ ID NO:29、SEQ ID NO:30、SEQ ID NO:31、SEQ ID NO:32、SEQ ID NO:34、SEQ ID NO:35、SEQ ID NO:43或SEQ ID NO:44的核酸序列。

[0115] 除非本文另外定义,否则结合本披露使用的科学和技术术语应具有本领域的普通技术人员通常所理解的含义。例如,与本文所描述的细胞和组织培养、分子生物学、免疫学、微生物学、遗传学以及蛋白质和核酸化学及杂交学结合使用的任何命名法是本领域众所周知并且常用的那些。术语的含义和范围应是清楚的;然而在任何潜在歧义的情况下,本文提供的定义优先于任何字典或外在定义。此外,除非上下文另外要求,单数术语应该包括复数含义并且复数术语应该包括单数含义。

2.用于肌营养不良蛋白基因的基因组编辑的基因构建体

[0116] 本发明涉及用于基因组编辑、基因组改变或改变肌营养不良蛋白基因(例如,人肌营养不良蛋白基因)的基因表达的基因构建体。基因构建体包括靶向人和恒河猴两者的肌营养不良蛋白基因序列(如SaCas9相容靶标)的至少一种gRNA。所披露的gRNA可以被包括在基于CRISPR/Cas9的基因编辑系统中,包括使用SaCas9的系统,来靶向人肌营养不良蛋白基因的外显子51周围的内含子区域,导致此区域的基因组缺失以恢复来自DMD患者的细胞中功能性肌营养不良蛋白的表达。

a.肌营养不良蛋白基因

[0117] 肌营养不良蛋白是指棒状胞质蛋白,其是通过细胞膜将肌纤维细胞骨架连接到周围的细胞外基质的蛋白质复合体的一部分。肌营养不良蛋白提供细胞膜肌营养不良蛋白聚糖复合体的结构稳定性。肌营养不良蛋白基因是在基因座Xp21处的2.2兆碱基。初级转录测

量了约2,400kb,其中成熟mRNA是约14kb。79个外显子编码超过3500个氨基酸的蛋白质。正常骨骼肌组织仅含有少量肌营养不良蛋白,但不存在异常表达会导致严重且无法治愈的症状的发展。肌营养不良蛋白基因中的一些突变导致受影响患者中产生缺陷型肌营养不良蛋白和严重肌营养不良表型。肌营养不良蛋白基因中的一些突变导致受影响患者中的部分功能性肌营养不良蛋白和轻微得多的肌营养不良表型。

[0118] DMD的致因是在肌营养不良蛋白基因中引起无义或移码突变的遗传或自发突变。对于DMD,天然存在的突变及其后果相对较好地理解。已知在棒结构域内包含的外显子45-55区域(例如,外显子51)中发生的框内缺失可产生高度功能性肌营养不良蛋白,并且许多携带者无症状或显示轻度症状。此外,超过60%的患者理论上可以通过靶向肌营养不良蛋白基因的此区域的外显子(例如,靶向外显子51)进行治疗。已经做出努力通过在mRNA剪接期间跳过非必需外显子(例如,外显子51跳跃)以产生内部缺失但为功能性的肌营养不良蛋白,来恢复DMD患者中被破坏的肌营养不良蛋白阅读框。一个或多个内部肌营养不良蛋白外显子的缺失(例如,外显子51的缺失)保留了正确的阅读框,但导致不那么严重的贝克肌营养不良或BMD。贝克肌营养不良或BMD基因型与DMD类似的地方在于肌营养不良蛋白基因中存在缺失。然而,这些缺失使得阅读框完好。因此,创建了内部截短但为部分功能性的肌营养不良蛋白。BMD具有广泛的表型,但通常如果缺失在肌营养不良蛋白的外显子45-55之间,则与DMD相比,表型轻微得多。因此,将DMD基因型改为BMD基因型是修正肌营养不良蛋白的常用策略。修正肌营养不良蛋白有许多策略,其中许多依赖于恢复内源肌营养不良蛋白的阅读框。这将疾病基因型从DMD转换为贝克肌营养不良。许多BMD患者具有维持翻译阅读框的基因内缺失,导致较短但很大程度上为功能性的肌营养不良蛋白。

[0119] 在某些实施例中,修饰外显子51(例如,通过例如NHEJ进行的外显子51的缺失或切除)以恢复阅读框改善了表型DMD受试者,包括具有缺失突变的DMD受试者。在某些实施例中,肌营养不良蛋白基因的外显子51是指肌营养不良蛋白基因的第51个外显子。在DMD患者中,外显子51常常邻近破坏框的缺失并且已在临床试验中被靶向用于基于寡核苷酸的外显子跳跃。外显子51跳跃化合物依替利森的临床试验报道了跨48周的显著功能益处,与基线相比具平均47%的肌营养不良蛋白阳性纤维。外显子51中的突变理想地适合于通过基于NHEJ的基因组编辑进行永久性修正。

[0120] 本发明披露的载体可以在肌营养不良蛋白基因(例如,人肌营养不良蛋白基因)中产生缺失。在某些实施例中,载体被配置为形成在肌营养不良蛋白基因的靶位置侧翼的两个内含子(第一内含子和第二内含子)中的两个双链断裂(第一双链断裂和第二双链断裂),从而缺失包含肌营养不良蛋白靶位置的肌营养不良蛋白基因区段。“肌营养不良蛋白靶位置”可以是肌营养不良蛋白外显子靶位置或肌营养不良蛋白外显子内靶位置,如本文所述。肌营养不良蛋白外显子靶位置的缺失可以优化患有杜氏肌营养不良的受试者的肌营养不良蛋白序列,例如它可以增加所编码的肌营养不良蛋白的功能或活性,或导致受试者疾病状态的改进。在某些实施例中,肌营养不良蛋白外显子靶位置的切除恢复阅读框。肌营养不良蛋白外显子靶位置可以包含肌营养不良蛋白基因的一个或多个外显子。在某些实施例中,肌营养不良蛋白靶位置包含肌营养不良蛋白基因(例如,人肌营养不良蛋白基因)的外显子51。

[0121] 本发明披露的基因构建体(例如,载体)可以介导肌营养不良蛋白基因(例如,人肌

营养不良蛋白基因)的外显子51处的高效基因编辑。本发明披露的基因构建体(例如,载体)恢复来自DMD患者的细胞中的肌营养不良蛋白表达。

[0122] 在DMD中,外显子51常常邻近破坏框的缺失。通过外显子跳跃消除肌营养不良蛋白的外显子51可以用于治疗大约15%的所有DMD患者。这类肌营养不良蛋白突变理想地适合通过基于NHEJ的基因组编辑和HDR进行的永久性修正。本文所述的基因构建体(例如,载体)已经开发用于人肌营养不良蛋白基因中外显子51的靶向修饰。将本发明披露的基因构建体(例如,载体)转染到人DMD细胞中并介导有效的基因修饰和向正确阅读框的转化。蛋白质修复伴随着阅读框修复,并在基于CRISPR/Cas9的基因编辑系统处理的细胞的大群体中检测到。

b.CRISPR系统

[0123] 本发明披露的基因构建体(例如,载体)编码对肌营养不良蛋白基因(例如,人肌营养不良蛋白基因)有特异性的基于CRISPR/Cas9的基因编辑系统。如在本文中可互换使用的“成簇规律间隔短回文重复序列”和“CRISPR”是指含有多个短的正向重复序列的基因座,这些重复序列存在于大约40%的测序细菌和90%的测序古细菌的基因组中。CRISPR系统是一种微生物核酸酶系统,其参与防御提供获得性免疫形式的入侵噬菌体和质粒。微生物宿主中的CRISPR基因座含有CRISPR相关(Cas)基因以及能够编程CRISPR介导的核酸裂解特异性的非编码RNA元件的组合。将短的外来DNA区段(称为间隔子)在CRISPR重复序列之间整合到基因组中,并用作暴露后的“记忆”。Cas9与sgRNA(在本文中也可互换地称为“gRNA”)的3’端形成复合体,并且该蛋白质-RNA对通过sgRNA序列的5’端与预先确定的20bp DNA序列(称为原型间隔子)之间的互补碱基配对识别其基因组靶标。此复合体经由crRNA内编码的区域(即,原型间隔子)和病原体基因组内的原型间隔子邻近基序(PAM)指导至病原体DNA的同源基因座。同向重复序列内的非编码CRISPR阵列被转录并裂解成含有单独间隔子序列的短的crRNA,这些crRNA将Cas核酸酶指导至靶位点(原型间隔子)。通过简单交换所表达的sgRNA的20bp识别序列,Cas9核酸酶可以被指导至新的基因组靶标。使用CRISPR间隔子以类似于真核生物的RNAi的方式识别外源遗传元件并使其沉默。

[0124] 已知CRISPR系统的3个类别(I、II和III型效应器系统)。II型效应器系统使用单一效应酶Cas9以四个相继步骤进行靶向DNA双链断裂以裂解dsDNA。与需要多个不同的效应器作为复合体起作用的I型和III型效应器系统相比,II型效应器系统可以在替代性情况(如真核细胞)中起作用。II型效应器系统由长的前crRNA(其自含有间隔子的CRISPR基因座转录)、Cas9蛋白和参与前crRNA加工的tracrRNA组成。tracrRNA与分隔前crRNA的间隔子的重复区杂交,因此通过内源RNA酶III开始dsRNA裂解。此裂解之后是通过Cas9在各间隔子内的第二裂解事件,产生保持与tracrRNA和Cas9缔合的成熟crRNA,从而形成Cas9:crRNA-tracrRNA复合体。

[0125] Cas9:crRNA-tracrRNA复合体解开DNA双链体,并搜寻匹配crRNA的序列以裂解。靶标识别在检测靶DNA中的“原型间隔子”序列与crRNA中的其余间隔子序列之间的互补性时发生。如果正确的原型间隔子邻近基序(PAM)也存在于原型间隔子的3’端,则Cas9介导靶DNA的裂解。对于原型间隔子靶向,序列必须紧接原型间隔子邻近基序(PAM)(一种被为DNA裂解所必需的Cas9核酸酶识别的短序列)之后。不同的II型系统具有不同的PAM要求。化脓链球菌CRISPR系统可以具有作为5’-NRG-3’的这种Cas9(SpCas9)的PAM序列,其中R为A或G,

并以此系统在人细胞中的特异性为特征。基于CRISPR/Cas9的基因编辑系统的独特能力是通过共表达单一Cas9蛋白与两种或更多种sgRNA同时靶向多个不同基因座的直接能力。例如,化脓链球菌II型系统天然更喜欢使用“NGG”序列,其中“N”可以是任何核苷酸,但在工程化系统中还可接受其他PAM序列,如“NAG”(Hsu et al.,Nature Biotechnology[Hsu等人,自然生物技术](2013)doi:10.1038/nbt.2647)。类似地,来源于脑膜炎奈瑟球菌(Neisseria meningitidis)的Cas9(NmCas9)通常具有NNNNGATT的天然PAM,但具有多种PAM间的活性,包括高度简并的NNNNGNNN PAM(Esvelt et al.Nature Methods[Esvelt等人自然方法](2013)doi:10.1038/nmeth.2681)。

[0126] 金黄色葡萄球菌的Cas9分子识别序列基序NNGRR(R＝A或G)(SEQ ID NO:22)并指导裂解靶核酸序列的在该序列上游的1至10(例如,3至5)bp。在某些实施例中,金黄色葡萄球菌的Cas9分子识别序列基序NNGRRN(R＝A或G)(SEQ ID NO:23)并指导裂解靶核酸序列的在该序列上游的1至10(例如,3至5)bp。在某些实施例中,金黄色葡萄球菌的Cas9分子识别序列基序NNGRRT(R＝A或G)(SEQ ID NO:24)并指导裂解靶核酸序列的在该序列上游的1至10(例如,3至5)bp。在某些实施例中,金黄色葡萄球菌的Cas9分子识别序列基序NNGRRV(R＝A或G)(SEQ ID NO:25)并指导裂解靶核酸序列的在该序列上游的1至10(例如,3至5)bp。在前述实施例中,N可以是任何核苷酸残基,例如A、G、C或T中的任一个。可以工程化Cas9分子以改变Cas9分子的PAM特异性。

(1)基于CRISPR/Cas9的基因编辑系统

[0127] 化脓链球菌的II型效应器系统的一种工程化形式显示在人细胞中发挥作用而用于基因组工程。在此系统中,通过合成再造的“指导RNA”(“gRNA”,在本文中也可互换地用作嵌合单指导RNA(“sgRNA”))(该指导RNA是一般避免RNA酶III和crRNA加工的需要的crRNA-tracrRNA融合物)将Cas9蛋白指导至基因组靶位点。本文提供了用于基因组编辑和治疗遗传疾病的基于CRISPR/Cas9的工程化系统。基于CRISPR/Cas9的工程化系统可以被设计为靶向任何基因,包括遗传疾病、老化、组织再生或伤口愈合中涉及的基因。基于CRISPR/Cas9的基因编辑系统可以包括Cas9蛋白或Cas9融合蛋白和至少一种gRNA。在某些实施例中,该系统包含两种gRNA分子。Cas9融合蛋白可以包括例如具有不同活性、对Cas9是内源的结构域,如反式激活结构域。

[0128] 靶基因(例如,肌营养不良蛋白基因,例如人肌营养不良蛋白基因)可参与细胞的分化或其中可能希望激活基因的任何其他过程,或可具有突变如移码突变或无义突变。如果靶基因具有引起提前终止密码子、异常剪接受体位点或异常剪接供体位点的突变,则基于CRISPR/Cas9的基因编辑系统可以被设计为识别并结合提前终止密码子、异常剪接受体位点或异常剪接供体位点的上游或下游的核苷酸序列。基于CRISPR-Cas9的系统也可以用来通过靶向剪接受体和供体破坏正常基因剪接以诱导提前终止密码子跳跃或恢复被破坏的阅读框。基于CRISPR/Cas9的基因编辑系统可以介导或可以不介导脱靶改变成为基因组的蛋白质编码区。

(a)Cas9分子和Cas9融合蛋白

[0129] 基于CRISPR/Cas9的基因编辑系统可以包括Cas9蛋白或Cas9融合蛋白。Cas9蛋白是一种裂解核酸的内切核酸酶,并且由CRISPR基因座编码并参与II型CRISPR系统。Cas9蛋白可以来自任何细菌或古细菌物种,包括但不限于化脓链球菌、金黄色葡萄球菌

(S.aureus)、燕麦食酸菌(Acidovorax avenae)、胸膜肺炎放线杆菌(Actinobacillus pleuropneumoniae)、琥珀酸放线杆菌(Actinobacillus succinogenes)、猪放线杆菌(Actinobacillus suis)、放线菌属(Actinomyces sp.)、cycliphilus denitrificans、少食氨基单胞菌(Aminomonas paucivorans)、蜡样芽孢杆菌(Bacillus cereus)、史氏芽孢杆菌(Bacillus smithii)、苏云金芽孢杆菌(Bacillus thuringiensis)、拟杆菌属(Bacteroides sp.)、Blastopirellula marina、慢生根瘤菌属(Bradyrhizobium sp.)、侧孢短芽孢杆菌(Brevibacillus laterosporus)、结肠弯曲杆菌(Campylobacter coli)、空肠弯曲杆菌(Campylobacter jejuni)、红嘴鸥弯曲杆菌(Campylobacter lari)、Candidatus Puniceispirillum、解纤维梭菌(Clostridium cellulolyticum)、产气荚膜梭菌(Clostridium perfringens)、拥挤棒状杆菌(Corynebacterium accolens)、白喉棒状杆菌(Corynebacterium diphtheria)、马氏棒状杆菌(Corynebacterium matruchotii)、Dinoroseobacter shibae、细长真杆菌(Eubacterium dolichum)、γ-变形菌(gamma proteobacterium)、重氮营养葡糖酸醋杆菌(Gluconacetobacter diazotrophicus)、副流感嗜血杆菌(Haemophilus parainfluenzae)、生痰嗜血杆菌(Haemophilus sputorum)、加拿大螺杆菌(Helicobacter canadensis)、同性恋螺杆菌(Helicobacter cinaedi)、鼬鼠螺杆菌(Helicobacter mustelae)、多营养泥杆菌(Ilyobacter polytropus)、金氏金氏菌(Kingella kingae)、卷曲乳酸杆菌(Lactobacillus crispatus)、伊氏李斯特菌(Listeria ivanovii)、单核细胞增生李斯特菌(Listeria monocytogenes)、李斯特氏菌科菌(Listeriaceae bacterium)、甲基孢囊菌属(Methylocystis sp.)、甲烷氧化菌(Methylosinus trichosporium)、羞怯动弯杆菌(Mobiluncus mulieris)、杆状奈瑟球菌(Neisseria bacilliformis)、灰色奈瑟球菌(Neisseria cinerea)、浅黄色奈瑟球菌(Neisseria flavescens)、乳糖奈瑟球菌(Neisseria lactamica)、奈瑟球菌属(Neisseria sp.)、瓦茨瓦尔西奈瑟球菌(Neisseria wadsworthii)、亚硝化单胞菌属(Nitrosomonas sp.)、食清洁剂细小棒菌(Parvibaculum lavamentivorans)、多杀巴氏杆菌(Pasteurella multocida)、Phascolarctobacterium succinatutens、Ralstonia syzygii、沼泽红假单胞菌(Rhodopseudomonas palustris)、小红卵菌属(Rhodovulum sp.)、米氏西蒙斯氏菌(Simonsiella mnuelleri)、鞘氨醇单胞菌属(Sphingomonas sp.)、Sporolactobacillus vineae、路邓葡萄球菌(Staphylococcus lugdunensis)、链球菌属(Streptococcus sp.)、Subdoligranulum sp.、运动替斯崔纳菌(Tistrella mobilis)、密螺旋体属(Treponema sp.)或Verminephrobacter eiseniae。在某些实施例中,Cas9分子是化脓链球菌Cas9分子(在本文中也称为“SpCas9”)。在某些实施例中,Cas9分子是金黄色葡萄球菌Cas9分子(在本文中也称为“SaCas9”)。

[0130] Cas9分子或Cas9融合蛋白可以与一种或多种gRNA分子相互作用,并且与该一种或多种gRNA分子一起定位至包含靶结构域(并且在某些实施例中为PAM序列)的位点。Cas9分子或Cas9融合蛋白识别PAM序列的能力可以例如使用如前所述的转化测定(Jinek 2012)来确定。

[0131] 在某些实施例中,Cas9分子或Cas9融合蛋白与靶核酸相互作用并裂解靶核酸的能力是PAM序列依赖性的。PAM序列是靶核酸中的序列。在某些实施例中,靶核酸的裂解发生在PAM序列的上游。来自不同细菌物种的Cas9分子可以识别不同的序列基序(例如,PAM序列)。

在某些实施例中,化脓链球菌的Cas9分子识别序列基序NGG并指导裂解靶核酸序列的在该序列上游的1至10(例如,3至5)bp(参见例如,Mali 2013)。在某些实施例中,嗜热链球菌的Cas9分子识别序列基序NGGNG(SEQ ID NO:36)和/或NNAGAAW(W=A或T)(SEQ ID NO:20)并指导裂解靶核酸序列的这些序列上游的1至10(例如,3至5)bp(参见例如,Horvath 2010;Deveau 2008)。在某些实施例中,变异链球菌的Cas9分子识别序列基序NGG和/或NAAR(R=A或G)(SEQ ID NO:21)并指导裂解靶核酸序列的此序列上游的1至10(例如,3至5)bp(参见例如,Deveau 2008)。在某些实施例中,金黄色葡萄球菌的Cas9分子识别序列基序NNGRR(R=A或G)(SEQ ID NO:22)并指导裂解靶核酸序列的在该序列上游的1至10(例如,3至5)bp。在某些实施例中,金黄色葡萄球菌的Cas9分子识别序列基序NNGRRN(R=A或G)(SEQ ID NO:23)并指导裂解靶核酸序列的在该序列上游的1至10(例如,3至5)bp。在某些实施例中,金黄色葡萄球菌的Cas9分子识别序列基序NNGRRT(R=A或G)(SEQ ID NO:24)并指导裂解靶核酸序列的在该序列上游的1至10(例如,3至5)bp。在某些实施例中,金黄色葡萄球菌的Cas9分子识别序列基序NNGRRV(R=A或G;V=A或C或G)(SEQ ID NO:25)并指导裂解靶核酸序列的在该序列上游的1至10(例如,)3至5bp。在前述实施例中,N可以是任何核苷酸残基,例如A、G、C或T中的任一个。可以工程化Cas9分子以改变Cas9分子的PAM特异性。

[0132] 在某些实施例中,载体编码至少一种识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的Cas9分子。在某些实施例中,该至少一种Cas9分子是金黄色葡萄球菌Cas9分子。在某些实施例中,该至少一种Cas9分子是突变型金黄色葡萄球菌Cas9分子。

[0133] 可以使Cas9蛋白突变,使得核酸酶活性失活。没有内切核酸酶活性的失活Cas9蛋白(iCas9,也称为“dCas9”)最近通过gRNA靶向细菌、酵母和人细胞中的基因以通过位阻使基因表达沉默。关于化脓链球菌Cas9序列的示例性突变包括:D10A、E762A、H840A、N854A、N863A和/或D986A。关于金黄色葡萄球菌Cas9序列的示例性突变包括D10A和N580A。在某些实施例中,Cas9分子是突变型金黄色葡萄球菌Cas9分子。在某些实施例中,突变型金黄色葡萄球菌Cas9分子包含D10A突变。编码此突变型金黄色葡萄球菌Cas9的核苷酸序列在下文提供的SEQ ID NO:34中列出:

```
atgaaaagga actacattct ggggctggcc atcgggatta caagcgtggg
gtatgggatt attgactatg aaacaaggga cgtgatcgac gcaggcgtca
gactgttcaa ggaggccaac gtggaaaaca atgagggacg gagaagcaag
aggggagcca ggcgcctgaa acgacggaga aggcacagaa tccagagggt
```

gaagaaactg ctgttcgatt acaacctgct gaccgaccat tctgagctga
gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg
agtgcataac gtcaatgagg tggaagagga caccggcaac gagctgtcta
caaaggaaca gatctcacgc aatagcaaag ctctggaaga gaagtatgtc
gcagagctgc agctggaacg gctgaagaaa gatggcgagg tgagagggtc
aattaatagg ttcaagacaa gcgactacgt caaagaagcc aagcagctgc
tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga
agggagcccc ttcggatgga aagacatcaa ggaatggtac gagatgctga
tgggacattg cacctatttt ccagaagagc tgagaagcgt caagtacgct
tataacgcag atctgtacaa cgccctgaat gacctgaaca acctggtcat
caccagggat gaaaacgaga aactggaata ctatgagaag ttccagatca
tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct
aaggagatcc tggtcaacga agaggacatc aagggctacc gggtgacaag
cactggaaaa ccagagttca ccaatctgaa agtgtatcac gatattaagg
acatcacagc acggaaagaa atcattgaga acgccgaact gctggatcag
attgctaaga tcctgactat ctaccagagc tccgaggaca tccaggaaga
gctgactaac ctgaacagcg agctgaccca ggaagagatc gaacagatta
gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat
ctttaaccgg ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga
aagagatccc aaccacactg gtggacgatt tcattctgtc acccgtggtc
aagcggagct tcatccagag catcaaagtg atcaacgcca tcatcaagaa
gtacggcctg cccaatgata tcattatcga gctggctagg gagaagaaca
gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc
aaagtacctg attgaaaaaa tcaagctgca cgatatgcag gagggaaagt
gtctgtattc tctggaggcc atccccctgg aggacctgct gaacaatcca
ttcaactacg aggtcgatca tattatcccc agaagcgtgt ccttcgacaa
ttcctttaac aacaaggtgc tggtcaagca ggaagagaac tctaaaaagg
gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg
catcagcaag accaaaaagg agtacctgct ggaagagcgg gacatcaaca
gattctccgt ccagaaggat tttattaacc ggaatctggt ggacacaaga
tacgctactc gcggcctgat gaatctgctg cgatcctatt tccgggtgaa
caatctggat gtgaaagtca agtccatcaa cggcgggttc acatcttttc
tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac

catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga
gtggaaaaag ctggacaaag ccaagaaagt gatggagaac cagatgttcg
aagagaagca ggccgaatct atgcccgaaa tcgagacaga acaggagtac
aaggagattt tcatcactcc tcaccagatc aagcatatca aggatttcaa
ggactacaag tactctcacc gggtggataa aaagcccaac agagagctga
tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa
aaagctgatc aacaaaagtc ccgagaagct gctgatgtac caccatgatc
ctcagacata tcagaaactg aagctgatta tggagcagta cggcgacgag
aagaacccac tgtataagta ctatgaagag actgggaact acctgaccaa
gtatagcaaa aaggataatg gccccgtgat caagaagatc aagtactatg
ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt
cgcaacaagg tggtcaagct gtcactgaag ccatacagat tcgatgtcta
tctggacaac ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca
tcaaaaagga gaactactat gaagtgaata gcaagtgcta cgaagaggct
aaaaagctga aaaagattag caaccaggca gagttcatcg cctcctttta
caacaacgac ctgattaaga tcaatggcga actgtatagg gtcatcgggg
tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact
taccgagagt atctggaaaa catgaatgat aagcgccccc ctcgaattat
caaaacaatt gcctctaaga ctcagagtat caaaaagtac tcaaccgaca
ttctgggaaa cctgtatgag gtgaagagca aaaagcaccc tcagattatc
aaaaagggc [SEQ ID NO: 34]。

[0134] 在某些实施例中,突变型金黄色葡萄球菌Cas9分子包含N580A突变。编码此突变型金黄色葡萄球菌Cas9分子的核苷酸序列在下文提供的SEQ ID NO:35中列出:

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg
gtatgggatt attgactatg aaacaaggga cgtgatcgac gcaggcgtca
gactgttcaa ggaggccaac gtggaaaaca atgagggacg gagaagcaag
aggggagcca ggcgcctgaa acgacggaga aggcacagaa tccagagggt
gaagaaactg ctgttcgatt acaacctgct gaccgaccat tctgagctga
gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg
agtgcataac gtcaatgagg tggaagagga caccggcaac gagctgtcta
caaaggaaca gatctcacgc aatagcaaag ctctggaaga gaagtatgtc
gcagagctgc agctggaacg gctgaagaaa gatggcgagg tgagagggtc
aattaatagg ttcaagacaa gcgactacgt caaagaagcc aagcagctgc

tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga
agggagcccc ttcggatgga aagacatcaa ggaatggtac gagatgctga
tgggacattg cacctatttt ccagaagagc tgagaagcgt caagtacgct
tataacgcag atctgtacaa cgccctgaat gacctgaaca acctggtcat
caccagggat gaaaacgaga aactggaata ctatgagaag ttccagatca
tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct
aaggagatcc tggtcaacga agaggacatc aagggctacc gggtgacaag
cactggaaaa ccagagttca ccaatctgaa agtgtatcac gatattaagg
acatcacagc acggaaagaa atcattgaga acgccgaact gctggatcag
attgctaaga tcctgactat ctaccagagc tccgaggaca tccaggaaga
gctgactaac ctgaacagcg agctgaccca ggaagagatc gaacagatta
gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat
ctttaaccgg ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga
aagagatccc aaccacactg gtggacgatt tcattctgtc acccgtggtc
aagcggagct tcatccagag catcaaagtg atcaacgcca tcatcaagaa
gtacggcctg cccaatgata tcattatcga gctggctagg gagaagaaca
gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc
aaagtacctg attgaaaaaa tcaagctgca cgatatgcag gagggaaagt
gtctgtattc tctggaggcc atccccctgg aggacctgct gaacaatcca
ttcaactacg aggtcgatca tattatcccc agaagcgtgt ccttcgacaa
ttcctttaac aacaaggtgc tggtcaagca ggaagaggcc tctaaaaagg
gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg
catcagcaag accaaaaagg agtacctgct ggaagagcgg gacatcaaca
gattctccgt ccagaaggat tttattaacc ggaatctggt ggacacaaga
tacgctactc gcggcctgat gaatctgctg cgatcctatt tccgggtgaa
caatctggat gtgaaagtca agtccatcaa cggcgggttc acatcttttc
tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac
catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga
gtggaaaaag ctggacaaag ccaagaaagt gatggagaac cagatgttcg
aagagaagca ggccgaatct atgcccgaaa tcgagacaga acaggagtac
aaggagattt tcatcactcc tcaccagatc aagcatatca aggatttcaa
ggactacaag tactctcacc gggtggataa aaagcccaac agagagctga
tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa

aaagctgatc aacaaaagtc ccgagaagct gctgatgtac caccatgatc
ctcagacata tcagaaactg aagctgatta tggagcagta cggcgacgag
aagaacccac tgtataagta ctatgaagag actgggaact acctgaccaa
gtatagcaaa aaggataatg gccccgtgat caagaagatc aagtactatg
ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt
cgcaacaagg tggtcaagct gtcactgaag ccatacagat tcgatgtcta
tctggacaac ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca
tcaaaaagga gaactactat gaagtgaata gcaagtgcta cgaagaggct
aaaaagctga aaaagattag caaccaggca gagttcatcg cctcctttta
caacaacgac ctgattaaga tcaatggcga actgtatagg gtcatcgggg
tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact
taccgagagt atctggaaaa catgaatgat aagcgccccc ctcgaattat
caaaacaatt gcctctaaga ctcagagtat caaaaagtac tcaaccgaca
ttctgggaaa cctgtatgag gtgaagagca aaaagcaccc tcagattatc
aaaaagggc [SEQ ID NO: 35]。

[0135] 编码Cas9分子的核酸可以是合成核酸序列。例如,合成核酸分子可以经化学修饰。可以对合成核酸序列进行密码子优化,例如,至少一个非常见密码子或较不常见密码子已被常见密码子替换。例如,合成核酸可以指导优化的信使mRNA的合成,例如针对哺乳动物表达系统中的表达优化的,例如本文所述的。

[0136] 此外或可替代地,编码Cas9分子或Cas9多肽的核酸可以包含核定位序列(NLS)。核定位序列在本领域是已知的。

[0137] 编码化脓链球菌的Cas9分子的示例性密码子优化的核酸序列在下文提供的SEQ ID NO:26中列出:

atggataaaa agtacagcat cgggctggac atcggtacaa actcagtggg
gtgggccgtg attacggacg agtacaaggt accctccaaa aaatttaaag
tgctgggtaa cacggacaga cactctataa agaaaaatct tattggagcc
ttgctgttcg actcaggcga gacagccgaa gccacaaggt tgaagcggac
cgccaggagg cggtatacca ggagaaagaa ccgcatatgc tacctgcaag
aaatcttcag taacgagatg gcaaaggttg acgatagctt tttccatcgc
ctggaagaat cctttcttgt tgaggaagac aagaagcacg aacggcaccc
catctttggc aatattgtcg acgaagtggc atatcacgaa aagtacccga
ctatctacca cctcaggaag aagctggtgg actctaccga taaggcggac
ctcagactta tttatttggc actcgcccac atgattaaat ttagaggaca

tttcttgatc gagggcgacc tgaacccgga caacagtgac gtcgataagc
tgttcatcca acttgtgcag acctacaatc aactgttcga agaaaaccct
ataaatgctt caggagtcga cgctaaagca atcctgtccg cgcgcctctc
aaaatctaga agacttgaga atctgattgc tcagttgccc ggggaaaaga
aaaatggatt gtttggcaac ctgatcgccc tcagtctcgg actgacccca
aatttcaaaa gtaacttcga cctggccgaa gacgctaagc tccagctgtc
caaggacaca tacgatgacg acctcgacaa tctgctggcc cagattgggg
atcagtacgc cgatctcttt ttggcagcaa agaacctgtc cgacgccatc
ctgttgagcg atatcttgag agtgaacacc gaaattacta aagcacccct
tagcgcatct atgatcaagc ggtacgacga gcatcatcag gatctgaccc
tgctgaaggc tcttgtgagg caacagctcc ccgaaaaata caaggaaatc
ttctttgacc agagcaaaaa cggctacgct ggctatatag atggtggggc
cagtcaggag gaattctata aattcatcaa gcccattctc gagaaaatgg
acggcacaga ggagttgctg gtcaaactta acagggagga cctgctgcgg
aagcagcgga cctttgacaa cgggtctatc ccccaccaga ttcatctggg
cgaactgcac gcaatcctga ggaggcagga ggatttttat ccttttctta
aagataaccg cgagaaaata gaaaagattc ttacattcag gatcccgtac
tacgtgggac ctctcgcccg gggcaattca cggtttgcct ggatgacaag
gaagtcagag gagactatta caccttggaa cttcgaagaa gtggtggaca
agggtgcatc tgcccagtct ttcatcgagc ggatgacaaa ttttgacaag
aacctcccta atgagaaggt gctgcccaaa cattctctgc tctacgagta
ctttaccgtc tacaatgaac tgactaaagt caagtacgtc accgagggaa
tgaggaagcc ggcattcctt agtggagaac agaagaaggc gattgtagac
ctgttgttca agaccaacag gaaggtgact gtgaagcaac ttaaagaaga
ctactttaag aagatcgaat gttttgacag tgtggaaatt tcaggggttg
aagaccgctt caatgcgtca ttggggactt accatgatct tctcaagatc
ataaaggaca aagacttcct ggacaacgaa gaaaatgagg atattctcga
agacatcgtc ctcaccctga ccctgttcga agacagggaa atgatagaag
agcgcttgaa aacctatgcc cacctcttcg acgataaagt tatgaagcag
ctgaagcgca ggagatacac aggatgggga agattgtcaa ggaagctgat
caatggaatt agggataaac agagtggcaa gaccatactg gatttcctca
aatctgatgg cttcgccaat aggaacttca tgcaactgat tcacgatgac
tctcttacct tcaaggagga cattcaaaag gctcaggtga gcgggcaggg
agactccctt catgaacaca tcgcgaattt ggcaggttcc cccgctatta
aaaagggcat ccttcaaact gtcaaggtgg tggatgaatt ggtcaaggta
atgggcagac ataagccaga aaatattgtg atcgagatgg cccgcgaaaa
ccagaccaca cagaagggcc agaaaaatag tagagagcgg atgaagagga
tcgaggaggg catcaaagag ctgggatctc agattctcaa agaacacccc

gtagaaaaca cacagctgca gaacgaaaaa ttgtacttgt actatctgca
gaacggcaga gacatgtacg tcgaccaaga acttgatatt aatagactgt
ccgactatga cgtagaccat atcgtgcccc agtccttcct gaaggacgac
tccattgata acaaagtctt gacaagaagc gacaagaaca ggggtaaaag
tgataatgtg cctagcgagg aggtggtgaa aaaaatgaag aactactggc
gacagctgct taatgcaaag ctcattacac aacggaagtt cgataatctg
acgaaagcag agagaggtgg cttgtctgag ttggacaagg cagggtttat
taagcggcag ctggtggaaa ctaggcagat cacaaagcac gtggcgcaga
ttttggacag ccggatgaac acaaaatacg acgaaaatga taaactgata
cgagaggtca aagttatcac gctgaaaagc aagctggtgt ccgattttcg
gaaagacttc cagttctaca aagttcgcga gattaataac taccatcatg
ctcacgatgc gtacctgaac gctgttgtcg ggaccgcctt gataaagaag
tacccaaagc tggaatccga gttcgtatac ggggattaca aagtgtacga
tgtgaggaaa atgatagcca agtccgagca ggagattgga aaggccacag
ctaagtactt cttttattct aacatcatga atttttttaa gacggaaatt
accctggcca acggagagat cagaaagcgg ccccttatag agacaaatgg
tgaaacaggt gaaatcgtct gggataaggg cagggatttc gctactgtga
ggaaggtgct gagtatgcca caggtaaata tcgtgaaaaa aaccgaagta
cagaccggag gattttccaa ggaaagcatt ttgcctaaaa gaaactcaga
caagctcatc gcccgcaaga aagattggga ccctaagaaa tacgggggat
ttgactcacc caccgtagcc tattctgtgc tggtggtagc taaggtggaa
aaaggaaagt ctaagaagct gaagtccgtg aaggaactct tgggaatcac
tatcatggaa agatcatcct ttgaaaagaa ccctatcgat ttcctggagg
ctaagggtta caaggaggtc aagaaagacc tcatcattaa actgccaaaa
tactctctct tcgagctgga aaatggcagg aagagaatgt tggccagcgc
cggagagctg caaaagggaa acgagcttgc tctgccctcc aaatatgtta
attttctcta tctcgcttcc cactatgaaa agctgaaagg gtctcccgaa
gataacgagc agaagcagct gttcgtcgaa cagcacaagc actatctgga
tgaaataatc gaacaaataa gcgagttcag caaaagggtt atcctggcgg
atgctaattt ggacaaagta ctgtctgctt ataacaagca ccgggataag
cctattaggg aacaagccga gaatataatt cacctcttta cactcacgaa
tctcggagcc cccgccgcct tcaaatactt tgatacgact atcgaccgga
aacggtatac cagtaccaaa gaggtcctcg atgccaccct catccaccag
tcaattactg gcctgtacga aacacggatcgacctctctc aactgggcgg
cgactag [SEQ ID NO: 26]。

[0138] 化脓链球菌Cas9分子的相应氨基酸序列在下文提供的SEQ ID NO:27中列出：

MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGALLF
DSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFL
VEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALA
HMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAIL
SARLSKSRRLENLIAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQ
LSKDTYDDDLDNLLAQIGDQYADLFLAAKNLSDAILLSDILRVNTEITKAPLS
ASMIKRYDEHHQDLTLLKALVRQQLPEKYKEIFFDQSKNGYAGYIDGGASQEE
FYKFIKPILEKMDGTEELLVKLNREDLLRKQRTFDNGSIPHQIHLGELHAILR
RQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWN
FEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYV
TEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGV
EDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERL
KTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFA
NRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVK
VVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQI
LKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLK
DDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLT
KAERGGLSELDKAGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVK
VITLKSKLVSDFRKDFQFYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESE
FVYGDYKVYDVRKMIAKSEQEIGKATAKYFFYSNIMNFFKTEITLANGEIRKR
PLIETNGETGEIVWDKGRDFATVRKVLSMPQVNIVKKTEVQTGGFSKESILPK
RNSDKLIARKKDWDPKKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGI
TIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGE
LQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQ
ISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAFK
YFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD [SEQ ID NO: 27]

[0139] 编码金黄色葡萄球菌的Cas9分子并且任选含有核定位序列(NLS)的示例性密码子优化的核酸序列在下文提供的SEQ ID NO:28-32、43和44中列出。编码金黄色葡萄球菌的Cas9分子的另一示例性密码子优化的核酸序列包含SEQ ID NO:83的核苷酸1293-4451。

[0140] 下文列出了SEQ ID NO:28:

atgaaaagga actacattct ggggctggac atcgggatta caagcgtggg
gtatgggatt attgactatg aaacaaggga cgtgatcgac gcaggcgtca
gactgttcaa ggaggccaac gtggaaaaca atgagggacg gagaagcaag
aggggagcca ggcgcctgaa acgacggaga aggcacagaa tccagagggt

gaagaaactg ctgttcgatt acaacctgct gaccgaccat tctgagctga
gtggaattaa tccttatgaa gccagggtga aaggcctgag tcagaagctg
tcagaggaag agttttccgc agctctgctg cacctggcta agcgccgagg
agtgcataac gtcaatgagg tggaagagga caccggcaac gagctgtcta
caaaggaaca gatctcacgc aatagcaaag ctctggaaga gaagtatgtc
gcagagctgc agctggaacg gctgaagaaa gatggcgagg tgagagggtc
aattaatagg ttcaagacaa gcgactacgt caaagaagcc aagcagctgc
tgaaagtgca gaaggcttac caccagctgg atcagagctt catcgatact
tatatcgacc tgctggagac tcggagaacc tactatgagg gaccaggaga
agggagcccc ttcggatgga aagacatcaa ggaatggtac gagatgctga
tgggacattg cacctatttt ccagaagagc tgagaagcgt caagtacgct
tataacgcag atctgtacaa cgccctgaat gacctgaaca acctggtcat
caccagggat gaaaacgaga aactggaata ctatgagaag ttccagatca
tcgaaaacgt gtttaagcag aagaaaaagc ctacactgaa acagattgct
aaggagatcc tggtcaacga agaggacatc aagggctacc gggtgacaag
cactggaaaa ccagagttca ccaatctgaa agtgtatcac gatattaagg
acatcacagc acggaaagaa atcattgaga acgccgaact gctggatcag
attgctaaga tcctgactat ctaccagagc tccgaggaca tccaggaaga
gctgactaac ctgaacagcg agctgaccca ggaagagatc gaacagatta
gtaatctgaa ggggtacacc ggaacacaca acctgtccct gaaagctatc
aatctgattc tggatgagct gtggcataca aacgacaatc agattgcaat
ctttaaccgg ctgaagctgg tcccaaaaaa ggtggacctg agtcagcaga
aagagatccc aaccacactg gtggacgatt tcattctgtc acccgtggtc
aagcggagct tcatccagag catcaaagtg atcaacgcca tcatcaagaa
gtacggcctg cccaatgata tcattatcga gctggctagg gagaagaaca
gcaaggacgc acagaagatg atcaatgaga tgcagaaacg aaaccggcag
accaatgaac gcattgaaga gattatccga actaccggga aagagaacgc
aaagtacctg attgaaaaaa tcaagctgca cgatatgcag gagggaaagt
gtctgtattc tctggaggcc tccccctgg aggacctgct gaacaatcca
ttcaactacg aggtcgatca tattatcccc agaagcgtgt ccttcgacaa
ttcctttaac aacaaggtgc tggtcaagca ggaagagaac tctaaaaagg
gcaataggac tcctttccag tacctgtcta gttcagattc caagatctct
tacgaaacct ttaaaaagca cattctgaat ctggccaaag gaaagggccg
catcagcaag accaaaaagg agtacctgct ggaagagcgg gacatcaaca
gattctccgt ccagaaggat tttattaacc ggaatctggt ggacacaaga
tacgctactc gcggcctgat gaatctgctg cgatcctatt tccgggtgaa
caatctggat gtgaaagtca agtccatcaa cggcgggttc acatcttttc
tgaggcgcaa atggaagttt aaaaaggagc gcaacaaagg gtacaagcac

catgccgaag atgctctgat tatcgcaaat gccgacttca tctttaagga
gtggaaaaag ctggacaaag ccaagaaagt gatggagaac cagatgttcg
aagagaagca ggccgaatct atgcccgaaa tcgagacaga acaggagtac
aaggagattt tcatcactcc tcaccagatc aagcatatca aggatttcaa
ggactacaag tactctcacc gggtggataa aaagcccaac agagagctga
tcaatgacac cctgtatagt acaagaaaag acgataaggg gaataccctg
attgtgaaca atctgaacgg actgtacgac aaagataatg acaagctgaa
aaagctgatc aacaaaagtc ccgagaagct gctgatgtac caccatgatc
ctcagacata tcagaaactg aagctgatta tggagcagta cggcgacgag
aagaacccac tgtataagta ctatgaagag actgggaact acctgaccaa
gtatagcaaa aaggataatg gccccgtgat caagaagatc aagtactatg
ggaacaagct gaatgcccat ctggacatca cagacgatta ccctaacagt
cgcaacaagg tggtcaagct gtcactgaag ccatacagat tcgatgtcta
tctggacaac ggcgtgtata aatttgtgac tgtcaagaat ctggatgtca
tcaaaaagga gaactactat gaagtgaata gcaagtgcta cgaagaggct
aaaaagctga aaaagattag caaccaggca gagttcatcg cctcctttta
caacaacgac ctgattaaga tcaatggcga actgtatagg gtcatcgggg
tgaacaatga tctgctgaac cgcattgaag tgaatatgat tgacatcact
taccgagagt atctggaaaa catgaatgat aagcgccccc ctcgaattat
caaaacaatt gcctctaaga ctcagagtat caaaaagtac tcaaccgaca
ttctgggaaa cctgtatgag gtgaagagca aaaagcaccc tcagattatc
aaaaagggc [SEQ ID NO: 28]

[0141] 下文列出了SEQ ID NO:29。

atgaagcgga actacatcct gggcctggac atcggcatca ccagcgtggg
ctacggcatc atcgactacg agacacggga cgtgatcgat gccggcgtgc
ggctgttcaa agaggccaac gtggaaaaca acgagggcag gcggagcaag
agaggcgcca gaaggctgaa gcggcggagg cggcatagaa tccagagagt
gaagaagctg ctgttcgact acaacctgct gaccgaccac agcgagctga
gcggcatcaa cccctacgag gccagagtga agggcctgag ccagaagctg
agcgaggaag agttctctgc cgccctgctg cacctggcca agagaagagg
cgtgcacaac gtgaacgagg tggaagagga caccggcaac gagctgtcca
ccaaagagca gatcagccgg aacagcaagg ccctggaaga gaaatacgtg
gccgaactgc agctggaacg gctgaagaaa gacggcgaag tgcggggcag
catcaacaga ttcaagacca gcgactacgt gaaagaagcc aaacagctgc
tgaaggtgca gaaggcctac caccagctgg accagagctt catcgacacc
tacatcgacc tgctggaaac ccggcggacc tactatgagg gacctggcga
gggcagcccc ttcggctgga aggacatcaa agaatggtac gagatgctga

tgggccactg cacctacttc cccgaggaac tgcggagcgt gaagtacgcc
tacaacgccg acctgtacaa cgccctgaac gacctgaaca atctcgtgat
caccagggac gagaacgaga agctggaata ttacgagaag ttccagatca
tcgagaacgt gttcaagcag aagaagaagc ccaccctgaa gcagatcgcc
aaagaaatcc tcgtgaacga agaggatatt aagggctaca gagtgaccag
caccggcaag cccgagttca ccaacctgaa ggtgtaccac gacatcaagg
acattaccgc ccggaaagag attattgaga acgccgagct gctggatcag
attgccaaga tcctgaccat ctaccagagc agcgaggaca tccaggaaga
actgaccaat ctgaactccg agctgaccca ggaagagatc gagcagatct
ctaatctgaa gggctatacc ggcacccaca acctgagcct gaaggccatc
aacctgatcc tggacgagct gtggcacacc aacgacaacc agatcgctat
cttcaaccgg ctgaagctgg tgcccaagaa ggtggacctg tcccagcaga
aagagatccc caccaccctg gtggacgact tcatcctgag ccccgtcgtg
aagagaagct tcatccagag catcaaagtg atcaacgcca tcatcaagaa
gtacggcctg cccaacgaca tcattatcga gctggcccgc gagaagaact
ccaaggacgc ccagaaaatg atcaacgaga tgcagaagcg gaaccggcag
accaacgagc ggatcgagga aatcatccgg accaccggca aagagaacgc
caagtacctg atcgagaaga tcaagctgca cgacatgcag gaaggcaagt
gcctgtacag cctggaagcc atccctctgg aagatctgct gaacaacccc
ttcaactatg aggtggacca catcatcccc agaagcgtgt ccttcgacaa
cagcttcaac aacaaggtgc tcgtgaagca ggaagaaaac agcaagaagg
gcaaccggac cccattccag tacctgagca gcagcgacag caagatcagc
tacgaaacct tcaagaagca catcctgaat ctggccaagg gcaagggcag
aatcagcaag accaagaaag agtatctgct ggaagaacgg gacatcaaca
ggttctccgt gcagaaagac ttcatcaacc ggaacctggt ggataccaga
tacgccacca gaggcctgat gaacctgctg cggagctact tcagagtgaa
caacctggac gtgaaagtga agtccatcaa tggcggcttc accagctttc
tgcggcggaa gtggaagttt aagaaagagc ggaacaaggg gtacaagcac
cacgccgagg acgccctgat cattgccaac gccgatttca tcttcaaaga
gtggaagaaa ctggacaagg ccaaaaaagt gatggaaaac cagatgttcg
aggaaaagca ggccgagagc atgcccgaga tcgaaaccga gcaggagtac
aaagagatct tcatcacccc ccaccagatc aagcacatta aggacttcaa
ggactacaag tacagccacc gggtggacaa gaagcctaat agagagctga
ttaacgacac cctgtactcc acccggaagg acgacaaggg caacaccctg
atcgtgaaca atctgaacgg cctgtacgac aaggacaatg acaagctgaa
aaagctgatc aacaagagcc ccgaaaagct gctgatgtac caccacgacc
cccagaccta ccagaaactg aagctgatta tggaacagta cggcgacgag
aagaatcccc tgtacaagta ctacgaggaa accgggaact acctgaccaa

gtactccaaa aaggacaacg gccccgtgat caagaagatt aagtattacg
gcaacaaact gaacgcccat ctggacatca ccgacgacta ccccaacagc
agaaacaagg tcgtgaagct gtccctgaag ccctacagat tcgacgtgta
cctggacaat ggcgtgtaca agttcgtgac cgtgaagaat ctggatgtga
tcaaaaaaga aaactactac gaagtgaata gcaagtgcta tgaggaagct
aagaagctga agaagatcag caaccaggcc gagtttatcg cctccttcta
caacaacgat ctgatcaaga tcaacggcga gctgtataga gtgatcggcg
tgaacaacga cctgctgaac cggatcgaag tgaacatgat cgacatcacc
taccgcgagt acctggaaaa catgaacgac aagaggcccc ccaggatcat
taagacaatc gcctccaaga cccagagcat taagaagtac agcacagaca
ttctgggcaa cctgtatgaa gtgaaatcta agaagcaccc tcagatcatc
aaaaagggc [SEQ ID NO: 29]

[0142] 下文列出了SEQ ID NO:30。

atgaagcgca actacatcct cggactggac atcggcatta cctccgtggg
atacggcatc atcgattacg aaactaggga tgtgatcgac gctggagtca
ggctgttcaa agaggcgaac gtggagaaca acgaggggcg gcgctcaaag
agggggccc gccggctgaa gcgccgccgc agacatagaa tccagcgcgt
gaagaagctg ctgttcgact acaaccttct gaccgaccac tccgaacttt
ccggcatcaa cccatatgag gctagagtga agggattgtc ccaaaagctg
tccgaggaag agttctccgc cgcgttgctc cacctcgcca agcgcagggg
agtgcacaat gtgaacgaag tggaagaaga taccggaaac gagctgtcca
ccaaggagca gatcagccgg aactccaagg ccctggaaga gaaatacgtg
gcggaactgc aactggagcg gctgaagaaa gacggagaag tgcgcggctc
gatcaaccgc ttcaagacct cggactacgt gaaggaggcc aagcagctcc
tgaaagtgca aaaggcctat caccaacttg accagtcctt tatcgatacc
tacatcgatc tgctcgagac tcggcggact tactacgagg gtccagggga
gggctcccca tttggttgga aggatattaa ggagtggtac gaaatgctga
tgggacactg cacatacttc cctgaggagc tgcggagcgt gaaatacgca
tacaacgcag acctgtacaa cgcgctgaac gacctgaaca atctcgtgat
cacccgggac gagaacgaaa agctcgagta ttacgaaaag ttccagatta
ttgagaacgt gttcaaacag aagaagaagc cgacactgaa gcagattgcc
aaggaaatcc tcgtgaacga agaggacatc aagggctatc gagtgacctc
aacgggaaag ccggagttca ccaatctgaa ggtctaccac gacatcaaag
acattaccgc ccggaaggag atcattgaga acgcggagct gttggaccag
attgcgaaga ttctgaccat ctaccaatcc tccgaggata ttcaggaaga
actcaccaac ctcaacagcg aactgaccca ggaggagata gagcaaatct
ccaacctgaa gggctacacc ggaactcata acctgagcct gaaggccatc

aacttgatcc tggacgagct gtggcacacc aacgataacc agatcgctat
tttcaatcgg ctgaagctgg tccccaagaa agtggacctc tcacaacaaa
aggagatccc tactaccctt gtggacgatt tcattctgtc ccccgtggtc
aagagaagct tcatacagtc aatcaaagtg atcaatgcca ttatcaagaa
atacggtctg cccaacgaca ttatcattga gctcgcccgc gagaagaact
cgaaggacgc ccagaagatg attaacgaaa tgcagaagag gaaccgacag
actaacgaac ggatcgaaga aatcatccgg accaccggga aggaaaacgc
gaagtacctg atcgaaaaga tcaagctcca tgacatgcag gaaggaaagt
gtctgtactc gctggaggcc attccgctgg aggacttgct gaacaaccct
tttaactacg aagtggatca tatcattccg aggagcgtgt cattcgacaa
ttccttcaac aacaaggtcc tcgtgaagca ggaggaaaac tcgaagaagg
gaaaccgcac gccgttccag tacctgagca gcagcgactc caagatttcc
tacgaaacct tcaagaagca catcctcaac ctggcaaagg ggaagggtcg
catctccaag accaagaagg aatatctgct ggaagaaaga gacatcaaca
gattctccgt gcaaaaggac ttcatcaacc gcaacctcgt ggatactaga
tacgctactc ggggtctgat gaacctcctg agaagctact ttagagtgaa
caatctggac gtgaaggtca agtcgattaa cggaggtttc acctccttcc
tgcggcgcaa gtggaagttc aagaaggaac ggaacaaggg ctacaagcac
cacgccgagg acgccctgat cattgccaac gccgacttca tcttcaaaga
atggaagaaa cttgacaagg ctaagaaggt catggaaaac cagatgttcg
aagaaaagca ggccgagtct atgcctgaaa tcgagactga acaggagtac
aaggaaatct ttattacgcc acaccagatc aaacacatca aggatttcaa
ggattacaag tactcacatc gcgtggacaa aaagccgaac agggaactga
tcaacgacac cctctactcc acccggaagg atgacaaagg gaataccctc
atcgtcaaca accttaacgg cctgtacgac aaggacaacg ataagctgaa
gaagctcatt aacaagtcgc ccgaaaagtt gctgatgtac caccacgacc
ctcagactta ccagaagctc aagctgatca tggagcagta tggggacgag
aaaaacccgt tgtacaagta ctacgaagaa actgggaatt atctgactaa
gtactccaag aaagataacg gccccgtgat taagaagatt aagtactacg
gcaacaagct gaacgcccat ctggacatca ccgatgacta ccctaattcc
cgcaacaagg tcgtcaagct gagcctcaag ccctaccggt ttgatgtgta
ccttgacaat ggagtgtaca agttcgtgac tgtgaagaac cttgacgtga
tcaagaagga gaactactac gaagtcaact ccaagtgcta cgaggaagca
aagaagttga agaagatctc gaaccaggcc gagttcattg cctccttcta
taacaacgac ctgattaaga tcaacggcga actgtaccgc gtcattggcg
tgaacaacga tctcctgaac cgcatcgaag tgaacatgat cgacatcact
taccgggaat acctggagaa tatgaacgac aagcgcccgc cccggatcat
taagactatc gcctcaaaga cccagtcgat caagaagtac agcaccgaca

tcctgggcaa cctgtacgag gtcaaatcga agaagcaccc ccagatcatc aagaaggga [SEQ ID NO: 30]

[0143] 下文列出了SEQ ID NO:31。

ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCAA
GCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCA
TCATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAA
GAGGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAG
GCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCG
ACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAG
GCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGC
CCTGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAG
AGGACACCGGCAACGAGCTGTCCACCAGAGAGCAGATCAGCCGGAACAGCAAG
GCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGA
CGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAG
AAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGC
TTCATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGG
ACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGA
TGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTAC
GCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGAT
CACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCG
AGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAA
ATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAA
GCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCC
GGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTG
ACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTC
CGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCG
GCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGG
CACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAA
GAAGGTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGACT
TCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCATCAAAGTGATC
AACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGC
CCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGC
GGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAA
GAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGG
CAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACAACC
CCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAAC
AGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAA

CCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAA
CCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAG
ACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCA
GAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCC
TGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTG
AAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAA
GAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCATTG
CCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAA
GTGATGGAAAACCAGATGTTCGAGGAAAGGCAGGCCGAGAGCATGCCCGAGAT
CGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGC
ACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAGCCT
AATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGG
CAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACA
AGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCAC
GACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGA
GAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCTGACCAAGT
ACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAAC
AAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAA
GGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATG
GCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAAC
TACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGAT
CAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGA
TCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGG
ATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAA
CGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCCAGAGCA
TTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAG
AAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGC
CGGCCAGGCAAAAAAGAAAAAG [SEQ ID NO: 31]

[0144] 下文列出了SEQ ID NO:32。

ACCGGTGCCA CCATGTACCC ATACGATGTT CCAGATTACG CTTCGCCGAA
GAAAAAGCGC AAGGTCGAAG CGTCCATGAA AAGGAACTAC ATTCTGGGGC
TGGACATCGG GATTACAAGC GTGGGGTATG GGATTATTGA CTATGAAACA
AGGGACGTGA TCGACGCAGG CGTCAGACTG TTCAAGGAGG CCAACGTGGA
AAACAATGAG GGACGGAGAA GCAAGAGGGG AGCCAGGCGC CTGAAACGAC
GGAGAAGGCA CAGAATCCAG AGGGTGAAGA AACTGCTGTT CGATTACAAC
CTGCTGACCG ACCATTCTGA GCTGAGTGGA ATTAATCCTT ATGAAGCCAG
GGTGAAAGGC CTGAGTCAGA AGCTGTCAGA GGAAGAGTTT TCCGCAGCTC

TGCTGCACCT GGCTAAGCGC CGAGGAGTGC ATAACGTCAA TGAGGTGGAA
GAGGACACCG GCAACGAGCT GTCTACAAAG GAACAGATCT CACGCAATAG
CAAAGCTCTG GAAGAGAAGT ATGTCGCAGA GCTGCAGCTG GAACGGCTGA
AGAAAGATGG CGAGGTGAGA GGGTCAATTA ATAGGTTCAA GACAAGCGAC
TACGTCAAAG AAGCCAAGCA GCTGCTGAAA GTGCAGAAGG CTTACCACCA
GCTGGATCAG AGCTTCATCG ATACTTATAT CGACCTGCTG GAGACTCGGA
GAACCTACTA TGAGGGACCA GGAGAAGGGA GCCCCTTCGG ATGGAAAGAC
ATCAAGGAAT GGTACGAGAT GCTGATGGGA CATTGCACCT ATTTTCCAGA
AGAGCTGAGA AGCGTCAAGT ACGCTTATAA CGCAGATCT TACAACGCCC
TGAATGACCT GAACAACCTG GTCATCACCA GGGATGAAAA CGAGAAACTG
GAATACTATG AGAAGTTCCA GATCATCGAA AACGTGTTTA AGCAGAAGAA
AAAGCCTACA CTGAAACAGA TTGCTAAGGA GATCCTGGTC AACGAAGAGG
ACATCAAGGG CTACCGGGTG ACAAGCACTG GAAAACCAGA GTTCACCAAT
CTGAAAGTGT ATCACGATAT TAAGGACATC ACAGCACGGA AAGAAATCAT
TGAGAACGCC GAACTGCTGG ATCAGATTGC TAAGATCCTG ACTATCTACC
AGAGCTCCGA GGACATCCAG GAAGAGCTGA CTAACCTGAA CAGCGAGCTG
ACCCAGGAAG AGATCGAACA GATTAGTAAT CTGAAGGGGT ACACCGGAAC
ACACAACCTG TCCCTGAAAG CTATCAATCT GATTCTGGAT GAGCTGTGGC
ATACAAACGA CAATCAGATT GCAATCTTTA ACCGGCTGAA GCTGGTCCCA
AAAAAGGTGG ACCTGAGTCA GCAGAAAGAG ATCCCAACCA CACTGGTGGA
CGATTTCATT CTGTCACCCG TGGTCAAGCG GAGCTTCATC CAGAGCATCA
AAGTGATCAA CGCCATCATC AAGAAGTACG GCCTGCCCAA TGATATCATT
ATCGAGCTGG CTAGGGAGAA GAACAGCAAG GACGCACAGA AGATGATCAA
TGAGATGCAG AAACGAAACC GGCAGACCAA TGAACGCATT GAAGAGATTA
TCCGAACTAC CGGGAAAGAG AACGCAAAGT ACCTGATTGA AAAAATCAAG
CTGCACGATA TGCAGGAGGG AAAGTGTCTG TATTCTCTGG AGGCCATCCC
CCTGGAGGAC CTGCTGAACA ATCCATTCAA CTACGAGGTC GATCATATTA
TCCCCAGAAG CGTGTCCTTC GACAATTCCT TTAACAACAA GGTGCTGGTC
AAGCAGGAAG AGAACTCTAA AAAGGGCAAT AGGACTCCTT TCCAGTACCT
GTCTAGTTCA GATTCCAAGA TCTCTTACGA AACCTTTAAA AAGCACATTC
TGAATCTGGC CAAAGGAAAG GGCCGCATCA GCAAGACCAA AAAGGAGTAC
CTGCTGGAAG AGCGGGACAT CAACAGATTC TCCGTCCAGA AGGATTTTAT
TAACCGGAAT CTGGTGGACA CAAGATACGC TACTCGCGGC CTGATGAATC
TGCTGCGATC CTATTTCCGG GTGAACAATC TGGATGTGAA AGTCAAGTCC
ATCAACGGCG GGTTCACATC TTTTCTGAGG CGCAAATGGA AGTTTAAAAA
GGAGCGCAAC AAAGGGTACA AGCACCATGC CGAAGATGCT CTGATTATCG
CAAATGCCGA CTTCATCTTT AAGGAGTGGA AAAAGCTGGA CAAAGCCAAG
AAAGTGATGG AGAACCAGAT GTTCGAAGAG AAGCAGGCCG AATCTATGCC

CGAAATCGAG ACAGAACAGG AGTACAAGGA GATTTTCATC ACTCCTCACC
AGATCAAGCA TATCAAGGAT TTCAAGGACT ACAAGTACTC TCACCGGGTG
GATAAAAAGC CCAACAGAGA GCTGATCAAT GACACCCTGT ATAGTACAAG
AAAAGACGAT AAGGGGAATA CCCTGATTGT GAACAATCTG AACGGACTGT
ACGACAAAGA TAATGACAAG CTGAAAAAGC TGATCAACAA AAGTCCCGAG
AAGCTGCTGA TGTACCACCA TGATCCTCAG ACATATCAGA AACTGAAGCT
GATTATGGAG CAGTACGGCG ACGAGAAGAA CCCACTGTAT AAGTACTATG
AAGAGACTGG GAACTACCTG ACCAAGTATA GCAAAAAGGA TAATGGCCCC
GTGATCAAGA AGATCAAGTA CTATGGGAAC AAGCTGAATG CCCATCTGGA
CATCACAGAC GATTACCCTA ACAGTCGCAA CAAGGTGGTC AAGCTGTCAC
TGAAGCCATA CAGATTCGAT GTCTATCTGG ACAACGGCGT GTATAAATTT
GTGACTGTCA AGAATCTGGA TGTCATCAAA AAGGAGAACT ACTATGAAGT
GAATAGCAAG TGCTACGAAG AGGCTAAAAA GCTGAAAAAG ATTAGCAACC
AGGCAGAGTT CATCGCCTCC TTTTACAACA ACGACCTGAT TAAGATCAAT
GGCGAACTGT ATAGGGTCAT CGGGGTGAAC AATGATCTGC TGAACCGCAT
TGAAGTGAAT ATGATTGACA TCACTTACCG AGAGTATCTG GAAAACATGA
ATGATAAGCG CCCCCCTCGA ATTATCAAAA CAATTGCCTC TAAGACTCAG
AGTATCAAAA AGTACTCAAC CGACATTCTG GGAAACCTGT ATGAGGTGAA
GAGCAAAAAG CACCCTCAGA TTATCAAAAA GGGCTAAGAA TTC [SEQ ID NO: 32]

[0145] 下文列出了SEQ ID NO:43。

*ATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCC*AA
GCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCA
TCATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAA
GAGGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGAAG
GCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTTCG
ACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACGAG
GCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCCGC
CCTGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGAAG
AGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGAACAGCAAG
GCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGA
CGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAG
AAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGC
TTCATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGG
ACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGAGA
TGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTAC
GCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGAT

CACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCG
AGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAA
ATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAA
GCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCC
GGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCCTG
ACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAACTC
CGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATACCG
GCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGTGG
CACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAA
GAAGGTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGACT
TCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCATCAAAGTGATC
AACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGC
CCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGC
GGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAA
GAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAAGG
CAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACAACC
CCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACAAC
AGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAA
CCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAA
CCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAG
ACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCA
GAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGGCC
TGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAGTG
AAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTTAA
GAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCATTG
CCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAAAA
GTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGAT
CGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGC
ACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAGCCT
AATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGG
CAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACA
AGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCAC
GACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGACGA
GAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCTGACCAAGT
ACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAAC
AAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAA
GGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATG
GCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAAC

TACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAGAT
CAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAAGA
TCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACCGG
ATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGGAAAACATGAA
CGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCCAGAGCA
TTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTAAG
AAGCACCCTCAGATCATCAAAAAGGGCAAAAGGCCGGCGGCCACGAAAAAGGC
CGGCCAGGCAAAAAAGAAAAAG [SEQ ID NO: 43]

[0146] 在一些实施例中,编码金黄色葡萄球菌Cas9分子的核苷酸序列包括下文提供的SEQ ID NO:44的核苷酸序列:

AAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGG
CATCATCGACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCA
AAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGA
AGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCTGTT
CGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACCCCTACG
AGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGAGTTCTCTGCC
GCCCTGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGTGAACGAGGTGGA
AGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAGATCAGCCGGAACAGCA
AGGCCCTGGAAGAGAAATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAA
GACGGCGAAGTGCGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAA
AGAAGCCAAACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGA
GCTTCATCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAG
GGACCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGA
GATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGT
ACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAATCTCGTG
ATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCAT
CGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAG
AAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGC
AAGCCCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGC
CCGGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATCC
TGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAATCTGAAC
TCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGGGCTATAC
CGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGGACGAGCTGT
GGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTGAAGCTGGTGCCC
AAGAAGGTGGACCTGTCCCAGCAGAAAGAGATCCCCACCACCCTGGTGGACGA
CTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCATCCAGAGCATCAAAGTGA
TCAACGCCATCATCAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTG

GCCCGCGAGAAGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAA
GCGGAACCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCA
AAGAGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAA
GGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAACAA
CCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTCGACA
ACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGC
AACCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGA
AACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCA
AGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTG
CAGAAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGG
CCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGAAAG
TGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGAAGTTT
AAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGCCCTGATCAT
TGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGACAAGGCCAAAA
AAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAG
ATCGAAACCGAGCAGGAGTACAAAGAGATCTTCATCACCCCCCACCAGATCAA
GCACATTAAGGACTTCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAGC
CTAATAGAGAGCTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAG
GGCAACACCCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGA
CAAGCTGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACC
ACGACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGAC
GAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCTGACCAA
GTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCA
ACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAAC
AAGGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAA
TGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAA
ACTACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAG
ATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATCAA
GATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCTGAACC
GGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGGAAAACATG
AACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCCAAGACCCAGAG
CATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTATGAAGTGAAATCTA
AGAAGCACCCTCAGATCATCAAAAAGGGC [SEQ ID NO: 44]

[0147] 金黄色葡萄球菌Cas9分子的氨基酸序列在下文提供的SEQ ID NO:33中列出。

MKRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGA
RRLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFS

AALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLK
KDGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYY
EGPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNL
VITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTST
GKPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNL
NSELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLV
PKKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIE
LAREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQ
EGKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKK
GNRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFS
VQKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWK
FKKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMP
EIETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDD
KGNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYG
DEKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSR
NKVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLK
KISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLEN
MNDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG [SEQ ID NO: 33]

[0148] 金黄色葡萄球菌Cas9分子的氨基酸序列在下文提供的SEQ ID NO:45中列出。

KRNYILGLDIGITSVGYGIIDYETRDVIDAGVRLFKEANVENNEGRRSKRGAR
RLKRRRRHRIQRVKKLLFDYNLLTDHSELSGINPYEARVKGLSQKLSEEEFSA
ALLHLAKRRGVHNVNEVEEDTGNELSTKEQISRNSKALEEKYVAELQLERLKK
DGEVRGSINRFKTSDYVKEAKQLLKVQKAYHQLDQSFIDTYIDLLETRRTYYE
GPGEGSPFGWKDIKEWYEMLMGHCTYFPEELRSVKYAYNADLYNALNDLNNLV
ITRDENEKLEYYEKFQIIENVFKQKKKPTLKQIAKEILVNEEDIKGYRVTSTG
KPEFTNLKVYHDIKDITARKEIIENAELLDQIAKILTIYQSSEDIQEELTNLN
SELTQEEIEQISNLKGYTGTHNLSLKAINLILDELWHTNDNQIAIFNRLKLVP
KKVDLSQQKEIPTTLVDDFILSPVVKRSFIQSIKVINAIIKKYGLPNDIIIEL
AREKNSKDAQKMINEMQKRNRQTNERIEEIIRTTGKENAKYLIEKIKLHDMQE
GKCLYSLEAIPLEDLLNNPFNYEVDHIIPRSVSFDNSFNNKVLVKQEENSKKG
NRTPFQYLSSSDSKISYETFKKHILNLAKGKGRISKTKKEYLLEERDINRFSV
QKDFINRNLVDTRYATRGLMNLLRSYFRVNNLDVKVKSINGGFTSFLRRKWKF
KKERNKGYKHHAEDALIIANADFIFKEWKKLDKAKKVMENQMFEEKQAESMPE
IETEQEYKEIFITPHQIKHIKDFKDYKYSHRVDKKPNRELINDTLYSTRKDDK
GNTLIVNNLNGLYDKDNDKLKKLINKSPEKLLMYHHDPQTYQKLKLIMEQYGD

EKNPLYKYYEETGNYLTKYSKKDNGPVIKKIKYYGNKLNAHLDITDDYPNSRN
KVVKLSLKPYRFDVYLDNGVYKFVTVKNLDVIKKENYYEVNSKCYEEAKKLKK
ISNQAEFIASFYNNDLIKINGELYRVIGVNNDLLNRIEVNMIDITYREYLENM
NDKRPPRIIKTIASKTQSIKKYSTDILGNLYEVKSKKHPQIIKKG [SEQ ID NO: 45]

[0149] 可替代地或此外,基于CRISPR/Cas9的基因编辑系统可以包括融合蛋白。融合蛋白可以包含两个异源多肽结构域,其中第一多肽结构域包含Cas蛋白,并且第二多肽结构域具有诸如转录激活活性、转录抑制活性、转录释放因子活性、组蛋白修饰活性、核酸酶活性、核酸结合活性、甲基化酶活性或脱甲基酶活性等活性。融合蛋白可以包括与第二多肽结构域融合的Cas9蛋白或突变的Cas9蛋白,该第二多肽结构域具有诸如转录激活活性、转录抑制活性、转录释放因子活性、组蛋白修饰活性、核酸酶活性、核酸结合活性、甲基化酶活性或脱甲基酶活性等活性。

(a)转录激活活性

[0150] 第二多肽结构域可以具有转录激活活性,即反式激活结构域。例如,内源哺乳动物基因(如人类基因)的基因表达可以通过将iCas9和反式激活结构域的融合蛋白经由gRNA的组合靶向哺乳动物启动子来实现。反式激活结构域可以包括一种VP 16蛋白、多种VP 16蛋白,如VP48结构域或VP64结构域,或NFκB转录激活物活性的p65结构域。例如,融合蛋白可以是iCas9-VP64。

(b)转录抑制活性

[0151] 第二多肽结构域可以具有转录抑制活性。第二多肽结构域可以具有Kruppel相关盒活性如KRAB结构域、ERF阻遏物结构域活性、Mxi1阻遏物结构域活性、SID4X阻遏物结构域活性、Mad-SID阻遏物结构域活性或TATA盒结合蛋白活性。例如,融合蛋白可以是dCas9-KRAB。

(c)转录释放因子活性

[0152] 第二多肽结构域可以具有转录释放因子活性。第二多肽结构域可以具有真核释放因子1(ERF1)活性或真核释放因子3(ERF3)活性。

(d)组蛋白修饰活性

[0153] 第二多肽结构域可以具有组蛋白修饰活性。第二多肽结构域可以具有组蛋白脱乙酰酶、组蛋白乙酰转移酶、组蛋白脱甲基酶或组蛋白甲基转移酶活性。组蛋白乙酰转移酶可以是p300或CREB结合蛋白(CBP)或其片段。例如,融合蛋白可以是dCas9-p300。

(e)核酸酶活性

[0154] 第二多肤结构域可以具有不同于Cas9蛋白的核酸酶活性的核酸酶活性。核酸酶或具有核酸酶活性的蛋白质是能够裂解核酸的核苷酸亚单位之间的磷酸二酯键的酶。核酸酶通常进一步分为内切核酸酶和外切核酸酶,但是一些酶可落入两个类别内。众所周知的核酸酶是脱氧核糖核酸酶和核糖核酸酶。

(f)核酸结合活性

[0155] 第二多肽结构域可以具有核酸结合活性,或者核酸结合蛋白DNA结合结构域(DBD)是独立折叠的蛋白结构域,其含有至少一个识别双链或单链DNA的基序。DBD可以识别特定

的DNA序列(识别序列)或对选自下组的DNA核酸结合区具有一般亲和力,该组由以下组成:螺旋-转角-螺旋区域、亮氨酸拉链区域、有翼螺旋区域、有翼螺旋转角螺旋区域、螺旋-环-螺旋区域、免疫球蛋白折叠、B3结构域、锌指、HMG盒、Wor3结构域、TAL效应子DNA结合结构域。

(g)甲基化酶活性

[0156] 第二多肽结构域可以具有甲基化酶活性,其涉及将甲基转移至DNA、RNA、蛋白质、小分子、胞嘧啶或腺嘌呤。第二多肽结构域可以包括DNA甲基转移酶。

(h)脱甲基酶活性

[0157] 第二多肽结构域可以具有脱甲基酶活性。第二多肽结构域可以包括从核酸、蛋白(特别是组蛋白)和其他分子中去除甲基(CH3-)基团的酶。可替代地,在使DNA脱甲基化的机制中,第二多肽可以将甲基基团转化为羟甲基胞嘧啶。第二多肽可以催化此反应。例如,催化此反应的第二多肽可以是Tet1。

(b)靶向肌营养不良蛋白基因的gRNA

[0158] 基于CRISPR/Cas9的基因编辑系统包括至少一种gRNA分子,例如两种gRNA分子。gRNA提供了基于CRISPR/Cas9的基因编辑系统的靶向。gRNA是以下两种非编码RNA的融合物:crRNA和tracrRNA。sgRNA可以通过交换编码20bp原型间隔子的序列(该序列通过与所需DNA靶标的互补碱基配对赋予靶向特异性)而靶向任何所需的DNA序列。gRNA模拟参与II型效应器系统的天然存在的crRNA:tractRNA双链体。可以包括例如42个核苷酸crRNA和75个核苷酸tracrRNA的这种双链体起指导Cas9裂解靶核酸的作用。如在本文中可互换使用的“靶区域”、“靶序列”或“原型间隔子”是指基于CRISPR/Cas9的基因编辑系统所靶向的靶基因(例如,肌营养不良蛋白基因)的区域。基于CRISPR/Cas9的基因编辑系统可以包括至少一种gRNA,其中这些gRNA靶向不同的DNA序列。靶DNA序列可以是重叠的。靶序列或原型间隔子后面是在原型间隔子3’端处的PAM序列。不同的II型系统具有不同的PAM要求。例如,化脓链球菌II型系统使用“NGG”序列,其中“N”可以是任何核苷酸。在一些实施例中,PAM序列可以是“NGG”,其中“N”可以是任何核苷酸。在一些实施例中,PAM序列可以是NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)。

[0159] 由本发明披露的基因构建体(例如,AAV载体)编码的gRNA分子的数目可以是至少1种gRNA、至少2种不同的gRNA、至少3种不同的gRNA、至少4种不同的gRNA、至少5种不同的gRNA、至少6种不同的gRNA、至少7种不同的gRNA、至少8种不同的gRNA、至少9种不同的gRNA、至少10种不同的gRNA、至少11种不同的gRNA、至少12种不同的gRNA、至少13种不同的gRNA、至少14种不同的gRNA、至少15种不同的gRNA、至少16种不同的gRNA、至少17种不同的gRNA、至少18种不同的gRNA、至少18种不同的gRNA、至少20种不同的gRNA、至少25种不同的gRNA、至少30种不同的gRNA、至少35种不同的gRNA、至少40种不同的gRNA、至少45种不同的gRNA或至少50种不同的gRNA。由本发明披露的载体编码的gRNA的数目可以是在至少1种gRNA到至少50种不同的gRNA、至少1种gRNA到至少45种不同的gRNA、至少1种gRNA到至少40种不同的gRNA、至少1种gRNA到至少35种不同的gRNA、至少1种gRNA到至少30种不同的gRNA、至少1种gRNA到至少25种不同的gRNA、至少1种gRNA到至少20种不同的gRNA、至少1种gRNA到至少16种不同的gRNA、至少1种gRNA到至少12种不同的gRNA、至少1种gRNA到至少8种不同的gRNA、至少1种gRNA到至少4种不同的gRNA、至少4种gRNA到至少50种不同的gRNA、至少4种不同的

gRNA到至少45种不同的gRNA、至少4种不同的gRNA到至少40种不同的gRNA、至少4种不同的gRNA到至少35种不同的gRNA、至少4种不同的gRNA到至少30种不同的gRNA、至少4种不同的gRNA到至少25种不同的gRNA、至少4种不同的gRNA到至少20种不同的gRNA、至少4种不同的gRNA到至少16种不同的gRNA、至少4种不同的gRNA到至少12种不同的gRNA、至少4种不同的gRNA到至少8种不同的gRNA、至少8种不同的gRNA到至少50种不同的gRNA、至少8种不同的gRNA到至少45种不同的gRNA、至少8种不同的gRNA到至少40种不同的gRNA、至少8种不同的gRNA到至少35种不同的gRNA、8种不同的gRNA到至少30种不同的gRNA、至少8种不同的gRNA到至少25种不同的gRNA、8种不同的gRNA到至少20种不同的gRNA、至少8种不同的gRNA到至少16种不同的gRNA或8种不同的gRNA到至少12种不同的gRNA之间。在某些实施例中,基因构建体(例如,AAV载体)编码一种gRNA分子(即,第一gRNA分子)和任选地Cas9分子。在某些实施例中,第一基因构建体(例如,第一AAV载体)编码一种gRNA分子(即,第一gRNA分子)和任选地Cas9分子,并且第二基因构建体(例如,第二AAV载体)编码一种gRNA分子(即,第二gRNA分子)和任选地Cas9分子。

[0160] gRNA分子包含靶向结构域,即后跟PAM序列的靶DNA序列的互补多核苷酸序列。gRNA可以在靶向结构域或互补多核苷酸序列的5'端包含“G”。gRNA分子的靶向结构域可以包含后跟PAM序列的靶DNA序列的至少10个碱基对、至少11个碱基对、至少12个碱基对、至少13个碱基对、至少14个碱基对、至少15个碱基对、至少16个碱基对、至少17个碱基对、至少18个碱基对、至少19个碱基对、至少20个碱基对、至少21个碱基对、至少22个碱基对、至少23个碱基对、至少24个碱基对、至少25个碱基对、至少30个碱基对或至少35个碱基对互补多核苷酸序列。在某些实施例中,gRNA分子的靶向结构域的长度为19-25个核苷酸。在某些实施例中,gRNA分子的靶向结构域的长度为20个核苷酸。在某些实施例中,gRNA分子的靶向结构域的长度为21个核苷酸。在某些实施例中,gRNA分子的靶向结构域的长度为22个核苷酸。在某些实施例中,gRNA分子的靶向结构域的长度为23个核苷酸。

[0161] gRNA可以靶向肌营养不良蛋白基因(DMD)的区域。在某些实施例中,gRNA可以靶向肌营养不良蛋白基因的外显子、内含子、启动子区、增强子区、转录区中的至少一种。在某些实施例中,gRNA分子靶向人肌营养不良蛋白基因的内含子50。在某些实施例中,gRNA分子靶向人肌营养不良蛋白基因的内含子51。在某些实施例中,gRNA分子靶向人肌营养不良蛋白基因的外显子51。gRNA可以包括靶向结构域,该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19中列出的核苷酸序列或其互补序列。

[0162] 可以设计单一或多重gRNA,以通过靶向外显子51处的突变热点或和引入外显子51的外显子内小插入和缺失或切除来恢复肌营养不良蛋白阅读框。用本发明披露的载体处理后,在体外在杜氏肌营养不良患者肌肉细胞中的肌营养不良蛋白表达可以恢复。在将基因修正的患者细胞移植到免疫缺陷小鼠中之后,在体内检测到人肌营养不良蛋白。值得注意的是,基于CRISPR/Cas9的基因编辑系统的独特的多重基因编辑能力使得能够高效地产生此突变热点区域的大量缺失,其通过通用或患者特异性基因编辑方法可以修正多达62%的患者突变。在一些实施例中,基于如通过surveyor测量的脱靶活性、中靶活性和距外显子的

距离来评估和选取候选gRNA。

3.DNA靶向组合物

[0163] 本发明还涉及包含此类基因构建体的DNA靶向组合物。如上所述,DNA靶向组合物包括至少一种靶向肌营养不良蛋白基因(例如,人肌营养不良蛋白基因)的gRNA分子(例如,两种gRNA分子)。该至少一种gRNA分子可以结合并识别靶区域。可以在可能的框外终止密码子的临上游选取靶区域,使得修复过程中的插入或缺失通过框转换恢复肌营养不良蛋白阅读框。靶区域也可以是剪接受体位点或剪接供体位点,使得修复过程中的插入或缺失破坏剪接并通过剪接位点破坏和外显子排除来恢复肌营养不良蛋白阅读框。靶区域也可以是异常终止密码子,使得修复过程中的插入或缺失通过消除或破坏终止密码子来恢复肌营养不良蛋白阅读框。

[0164] 在某些实施例中,本发明披露的DNA靶向组合物包括第一gRNA和第二gRNA,其中第一gRNA分子和第二gRNA分子包含靶向结构域,该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19中列出的核苷酸序列或其互补序列。在某些实施例中,第一gRNA分子和第二gRNA分子包含不同的靶向结构域。在某些实施例中,第一gRNA分子是SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14或SEQ ID NO:15,并且第二gRNA分子是SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18或SEQ ID NO:19。在某些实施例中,第一gRNA分子选自下组,该组由以下组成:SEQ ID NO:1、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14和SEQ ID NO:15,并且第二gRNA分子选自下组,该组由以下组成:SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18和SEQ ID NO:19。

[0165] 在某些实施例中,第一gRNA分子和第二gRNA分子选自下组,该组由以下组成:(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:1中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:2中列出的核苷酸序列;(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID

NO:18中列出的核苷酸序列;(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(x)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列;(xi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;以及(xii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:41中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:42中列出的核苷酸序列。在一些实施例中,DNA靶向组合物包括SEQ ID NO:37中列出的核苷酸序列和/或SEQ ID NO:38中列出的核苷酸序列。

[0166] 在某些实施例中,DNA靶向组合物可以进一步包括至少一种识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的PAM的Cas9分子或Cas9融合蛋白。在一些实施例中,DNA靶向组合物包括SEQ ID NO:83或SEQ ID NO:84中列出的核苷酸序列。在某些实施例中,载体被配置为分别形成在人肌营养不良蛋白基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段。

[0167] 本发明披露的载体的缺失效率可以与缺失大小(即,通过载体缺失的区段的大小)相关。在某些实施例中,特定缺失的长度或大小由被靶向的基因(例如,肌营养不良蛋白基因)中PAM序列之间的距离决定。在某些实施例中,肌营养不良蛋白基因区段的根据其长度和其包含的序列(例如,外显子51)定义的特定缺失是与靶基因(例如,肌营养不良蛋白基因)内与特定PAM序列相邻的断裂的结果。

[0168] 在某些实施例中,缺失大小为约50至约2,000个碱基对(bp),例如约50至约1999bp、约50至约1900bp、约50至约1800bp、约50至约1700bp、约50至约1650bp、约50至约1600bp、约50至约1500bp、约50至约1400bp、约50至约1300bp、约50至约1200bp、约50至约1150bp、约50至约1100bp、约50至约1000bp、约50至约900bp、约50至约850bp、约50至约800bp、约50至约750bp、约50至约700bp、约50至约600bp、约50至约500bp、约50至约400bp、约50至约350bp、约50至约300bp、约50至约250bp、约50至约200bp、约50至约150bp、约50至约100bp、约100至约1999bp、约100至约1900bp、约100至约1800bp、约100至约1700bp、约100至约1650bp、约100至约1600bp、约100至约1500bp、约100至约1400bp、约100至约1300bp、约100至约1200bp、约100至约1150bp、约100至约1100bp、约100至约1000bp、约100至约900bp、约100至约850bp、约100至约800bp、约100至约750bp、约100至约700bp、约100至约600bp、约100至约1000bp、约100至约400bp、约100至约350bp、约100至约300bp、约100至约250bp、约100至约200bp、约100至约150bp、约200至约1999bp、约200至约1900bp、约200至约1800bp、约200至约1700bp、约200至约1650bp、约200至约1600bp、约200至约1500bp、约200至约1400bp、约200至约1300bp、约200至约1200bp、约200至约1150bp、约200至约1100bp、约200至约1000bp、约200至约900bp、约200至约850bp、约200至约800bp、约200至约750bp、约200至约700bp、约200至约600bp、约200至约2000bp、约200至约400bp、约200至约350bp、约200

至约300bp、约200至约250bp、约300至约1999bp、约300至约1900bp、约300至约1800bp、约300至约1700bp、约300至约1650bp、约300至约1600bp、约300至约1500bp、约300至约1400bp、约300至约1300bp、约300至约1200bp、约300至约1150bp、约300至约1100bp、约300至约1000bp、约300至约900bp、约300至约850bp、约300至约800bp、约300至约750bp、约300至约700bp、约300至约600bp、约300至约3000bp、约300至约400bp或约300至约350bp。在某些实施例中,缺失大小可以是约118个碱基对、约233个碱基对、约326个碱基对、约766个碱基对、约805个碱基对或约1611个碱基对。

4.用于肌肉中基因组编辑的组合物

[0169] 本发明涉及用于基因组编辑受试者的骨骼肌或心肌中的靶基因的基因构建体(例如,载体)或其组合物。组合物包括经修饰的AAV载体和编码基于CRISPR/Cas9的基因编辑系统(例如,gRNA分子和Cas9分子)的核苷酸序列。组合物向骨骼肌或心肌递送基于CRISPR/Cas9的基因编辑系统的活性形式。本发明披露的基因构建体(例如,载体)可以用于修正或减少遗传疾病和/或其他骨骼肌或心肌病症(例如,DMD)中涉及的肌营养不良蛋白基因中的突变的影响。组合物可以进一步包含供体DNA或转基因。这些组合物可以用于基因组编辑、基因组工程以及修正或减小遗传疾病和/或其他骨骼肌或心肌病症中涉及的基因突变的影响。

a.用于靶向肌营养不良蛋白的基于CRISPR/Cas9的基因编辑系统

[0170] 本文披露了对肌营养不良蛋白基因有特异性的基于CRISPR/Cas9的基因编辑系统。基于CRISPR/Cas9的基因编辑系统可以包括Cas9和至少一种靶向肌营养不良蛋白基因的gRNA。基于CRISPR/Cas9的基因编辑系统可以结合并识别靶区域。可以在可能的框外终止密码子的临上游选取靶区域,使得修复过程中的插入或缺失通过框转换恢复肌营养不良蛋白阅读框。靶区域也可以是剪接受体位点或剪接供体位点,使得修复过程中的插入或缺失破坏剪接并通过剪接位点破坏和外显子排除来恢复肌营养不良蛋白阅读框。靶区域也可以是异常终止密码子,使得修复过程中的插入或缺失通过消除或破坏终止密码子来恢复肌营养不良蛋白阅读框。

[0171] gRNA可以靶向选自SEQ ID NO:1-19、41、42的核苷酸序列或其互补序列。例如,所披露的基于CRISPR/Cas9的基因编辑系统被工程化为介导肌营养不良蛋白基因的外显子51处的高效基因编辑。这些基于CRISPR/Cas9的基因编辑系统恢复来自DMD患者的细胞中的肌营养不良蛋白表达。在一些实施例中,DNA靶向组合物包括SEQ ID NO:37中列出的核苷酸序列、SEQ ID NO:38中列出的核苷酸序列、SEQ ID NO:83中列出的核苷酸序列和/或SEQ ID NO:84中列出的核苷酸序列。例如,DNA靶向组合物包括SEQ ID NO:37中列出的核苷酸序列、SEQ ID NO:38中列出的核苷酸序列和SEQ ID NO:83中列出的核苷酸序列,或者DNA靶向组合物包括SEQ ID NO:37中列出的核苷酸序列、SEQ ID NO:38中列出的核苷酸序列和SEQ ID NO:84中列出的核苷酸序列。

b.腺相关病毒载体

[0172] 组合物还可以包括病毒递送系统。在某些实施例中,载体是腺相关病毒(AAV)载体。AAV载体是感染人及其他一些灵长类物种的属于细小病毒科依赖病毒属的小病毒。AAV载体可以用于使用不同构建体配置来递送基于CRISPR/Cas9的基因编辑系统。例如,AAV载体可以在单独的载体上或在相同的载体上递送Cas9和gRNA表达盒。可替代地,如果使用来源于如金黄色葡萄球菌或脑膜炎奈瑟球菌等物种的小Cas9蛋白,则Cas9和最多两种gRNA的

表达盒可以组合在4.7kb包装限制内的单一AAV载体中。

[0173] 在某些实施例中,AAV载体是经修饰的AAV载体。经修饰的AAV载体可以具有增强的心肌和骨骼肌组织嗜性。经修饰的AAV载体可以能够在哺乳动物的细胞中递送基于CRISPR/Cas9的基因编辑系统并且使该系统在该细胞中表达。例如,经修饰的AAV载体可以是AAV-SASTG载体(Piacentino et al.(2012)Human Gene Therapy[Piacentino等人(2012)人类基因治疗]23:635-646)。经修饰的AAV载体可以在体内将核酸酶递送至骨骼肌和心肌。经修饰的AAV载体可以基于若干种衣壳类型中的一种或多种,这些衣壳类型包括AAV1、AAV2、AAV5、AAV6、AAV8和AAV9。经修饰的AAV载体可以基于具有替代性肌肉向性AAV衣壳的AAV2假型,如AAV2/1、AAV2/6、AAV2/7、AAV2/8、AAV2/9、AAV2.5和AAV/SASTG载体,这些载体通过系统性和局部递送高效转导骨骼肌或心肌(Seto et al.Current Gene Therapy[Seto等人当今基因治疗](2012)12:139-151)。经修饰的AAV载体可以是AAV2i8G9(Shen et al.J.Biol.Chem.[Shen等人生物化学杂志](2013)288:28814-28823)。在一些实施例中,组合物包括SEQ ID NO:39中列出的核苷酸序列和/或SEQ ID NO:40中列出的核苷酸序列。在一些实施例中,组合物包括包含SEQ ID NO:39中列出的核苷酸序列的第一载体和包含SEQ ID NO:40中列出的核苷酸序列的第二载体。

5.肌肉中基因组编辑的方法

[0174] 本披露涉及在受试者的骨骼肌或心肌中进行基因组编辑的方法。该方法包括向该受试者的骨骼肌或心肌给予如上所述的用于在骨骼肌或心肌中进行基因组编辑的组合物。基因组编辑可以包括修正突变基因或插入转基因。修正突变基因可以包括缺失、重排或替换突变基因。修正突变基因可以包括核酸酶介导的NHEJ或HDR。

6.修正突变基因和治疗受试者的方法

[0175] 本发明披露的主题提供了修正细胞中的突变基因(例如,突变型肌营养不良蛋白基因,例如突变型人肌营养不良蛋白基因)和治疗患有遗传疾病(如DMD)的受试者的方法。该方法可以包括向细胞或受试者给予如上所述的本发明披露的基因构建体(例如,载体)或包含其的组合物。该方法可以包括向该受试者的骨骼肌或心肌给予如上所述的本发明披露的基因构建体(例如,载体)或包含其的用于在骨骼肌或心肌中进行基因组编辑的组合物。使用本发明披露的基因构建体(例如,载体)或包含其的组合物将基于CRISPR/Cas9的基因编辑系统递送至骨骼肌或心肌,用可以替换整个基因或含有突变的区域的修复模板或供体DNA,可以使全功能性或部分功能性蛋白的表达恢复。基于CRISPR/Cas9的基因编辑系统可以用于在靶向基因组座处引入位点特异性双链断裂。当使基于CRISPR/Cas9的基因编辑系统与靶DNA序列结合时,产生位点特异性双链断裂,从而允许裂解靶DNA。这种DNA裂解可以刺激天然DNA修复机器,导致两种可能的修复途径之一:同源定向修复(HDR)或非同源末端连接(NHEJ)途径。

[0176] 本披露涉及在没有修复模板的情况下用基于CRISPR/Cas9的基因编辑系统进行基因组编辑,这可以高效地修正阅读框并恢复遗传疾病中涉及的功能性蛋白的表达。所披露的基于CRISPR/Cas9的基因编辑系统可以涉及使用同源定向修复或核酸酶介导的非同源末端连接(NHEJ)修正方法,其使得在可能不适于同源重组或基于选择的基因修正的增殖受限制的原代细胞系中的高效修正成为可能。此策略将活性的基于CRISPR/Cas9的基因编辑系统的快速且稳健的组织与高效的基因编辑方法整合用于治疗由引起移码、提前终止密码

子、异常剪接供体位点或异常剪接受体位点的非必需编码区中的突变所引起的遗传疾病。

a.核酸酶介导的非同源末端连接

[0177] 从内源突变的基因恢复蛋白质表达可以通过无模板NHEJ介导的DNA修复。与靶向靶基因RNA的瞬时方法相对比,通过瞬时表达基于CRISPR/Cas9的基因编辑系统修正基因组中的靶基因阅读框可以导致各个经修饰细胞及其所有子代永久性地恢复靶基因表达。在某些实施例中,NHEJ是核酸酶介导的NHEJ,其在某些实施例中是指启动Cas9分子、切割双链DNA的NHEJ。该方法包括将本发明披露的基因构建体(例如,载体)或包含其的组合物给予至受试者的骨骼肌或心肌以在骨骼肌或心肌中进行基因组编辑。

[0178] 核酸酶介导的NHEJ基因修正可以修正突变的靶基因,并提供优于HDR途径的若干潜在优势。例如,NHEJ不需要可能引起非特异性插入诱变的供体模板。与HDR相对比,NHEJ在细胞周期的所有阶段都高效地操作,因此可以有效用于周期中的细胞和有丝分裂后细胞(如肌纤维)两者中。相对于基于寡核苷酸的外显子跳跃或终止密码子的药理学强制通读,这提供了替代性的稳健的永久性基因恢复,并且在理论上可以将需要少至一种药物治疗。除本文所述的质粒电穿孔方法以外,使用基于CRISPR/Cas9的基因编辑系统以及包括大范围核酸酶和锌指核酸酶的其他工程化核酸酶的基于NHEJ的基因修正可以与用于基于细胞和基因的疗法的其他现有离体和体内平台联合。例如,基于CRISPR/Cas9的基因编辑系统通过基于mRNA的基因转移或作为纯化的细胞可渗透蛋白质的递送可以使不含DNA的基因组编辑方法成为可能,该方法将克服插入诱变的任何可能性。

b.同源定向修复

[0179] 从内源突变的基因恢复蛋白质表达可以涉及同源定向修复。如上所述的方法进一步包括将供体模板给予至细胞。供体模板可以包括编码全功能性蛋白或部分功能性蛋白的核苷酸序列。例如,供体模板可以包括小型化肌营养不良蛋白构建体,称为微小肌营养不良蛋白(“minidys”),即一种用于恢复突变型肌营养不良蛋白基因或肌营养不良蛋白基因的片段的全功能性肌营养不良蛋白构建体,其在同源定向修复后导致突变型肌营养不良蛋白基因的恢复。

c.使用CRISPR/Cas9修正突变基因和治疗受试者的方法

[0180] 本披露还涉及用基于CRISPR/Cas9的基因编辑系统进行基因组编辑以用修复模板或供体DNA恢复全功能性或部分功能性蛋白的表达,从而可以替换整个基因或含有突变的区域。基于CRISPR/Cas9的基因编辑系统可以用于在靶向基因组座处引入位点特异性双链断裂。当使用gRNA使基于CRISPR/Cas9的基因编辑系统与靶DNA序列结合时,产生位点特异性双链断裂,从而允许裂解靶DNA。基于CRISPR/Cas9的基因编辑系统由于其高速率的成功且高效的遗传修饰而具有高级基因组编辑的优势。这种DNA裂解可以刺激天然DNA修复机器,导致两种可能的修复途径之一:同源定向修复(HDR)或非同源末端连接(NHEJ)途径。例如,针对肌营养不良蛋白基因的基于CRISPR/Cas9的基因编辑系统可以包括具有SEQ ID NO:1-19、41、42中任一个的核酸序列或其互补序列的gRNA。

[0181] 本披露涉及在没有修复模板的情况下用基于CRISPR/Cas9的基因编辑系统进行基因组编辑,这可以高效地修正阅读框并恢复遗传疾病中涉及的功能性蛋白的表达。所披露的基于CRISPR/Cas9的基因编辑系统和方法可以涉及使用同源定向修复或核酸酶介孚的非同源末端连接(NHEJ)修正方法,其使得在可能不适于同源重组或基于选择的基因修正的增

殖受限制的原代细胞系中的高效修正成为可能。此策略将活性的基于CRISPR/Cas9的基因编辑系统的快速且稳健的组织与高效的基因编辑方法整合用于治疗由引起移码、提前终止密码子、异常剪接供体位点或异常剪接受体位点的非必需编码区中的突变所引起的遗传疾病。

[0182] 本披露提供了修正细胞中的突变基因和治疗患有遗传疾病(如DMD)的受试者的方法。该方法可以包括向细胞或受试者给予如上所述的基于CRISPR/Cas9的基因编辑系统、编码所述基于CRISPR/Cas9的基因编辑系统的多核苷酸或载体、或所述基于CRISPR/Cas9的基因编辑系统的组合物。该方法可以包括给予基于CRISPR/Cas9的基因编辑系统,如给予含有具有核酸酶活性的第二结构域的Cas9蛋白或Cas9融合蛋白、编码所述Cas9蛋白或Cas9融合蛋白的核苷酸序列,和/或至少一种gRNA,其中这些gRNA靶向不同的DNA序列。靶DNA序列可以是重叠的。如上所述,给予至细胞的gRNA的数目可以是至少1种gRNA、至少2种不同的gRNA、至少3种不同的gRNA、至少4种不同的gRNA、至少5种不同的gRNA、至少6种不同的gRNA、至少7种不同的gRNA、至少8种不同的gRNA、至少9种不同的gRNA、至少10种不同的gRNA、至少15种不同的gRNA、至少20种不同的gRNA、至少30种不同的gRNA或至少50种不同的gRNA。gRNA可以包括SEQ ID NO:1-19、41、42中的至少一个的核酸序列或其互补序列。该方法可以涉及同源定向修复或非同源末端连接。

7.治疗疾病的方法

[0183] 本披露涉及治疗有需要的受试者的方法。该方法包括向受试者的组织给予如上所述的本发明披露的基因构建体(例如,载体)或包含其的组合物。在某些实施例中,该方法可以包括向该受试者的骨骼肌或心肌给予如上所述的本发明披露的基因构建体(例如,载体)或包含其的组合物。在某些实施例中,该方法可以包括向该受试者的静脉给予如上所述的本发明披露的基因构建体(例如,载体)或包含其的组合物。在某些实施例中,该受试者患有引起变性或无力或遗传疾病的骨骼肌或心肌病症。例如,如上所述,该受试者可能患有杜氏肌营养不良。

a.杜氏肌营养不良

[0184] 如上所述,该方法可以用于修正肌营养不良蛋白基因并恢复所述突变的肌营养不良蛋白基因的全功能性或部分功能性蛋白表达。在一些方面和实施例中,本披露提供了用于减少患者中DMD的影响(例如,临床症状/适应症)的方法。在一些方面和实施例中,本披露提供了用于治疗患者的DMD的方法。在一些方面和实施例中,本披露提供了用于预防患者的DMD的方法。在一些方面和实施例中,本披露提供了用于预防患者的DMD进一步发展的方法。

8.具有Δ52 hDMD的转基因啮齿动物的方法

[0185] 本披露涉及产生具有外显子52缺失的人肌营养不良蛋白基因的转基因啮齿动物胚胎的方法。该方法包括向啮齿动物胚胎给予gRNA从而缺失人肌营养不良蛋白基因的外显子52,并选择具有人肌营养不良蛋白基因外显子52缺失的转基因啮齿动物胚胎,其中该啮齿动物胚胎包含正常的人肌营养不良蛋白基因。在一些实施例中,啮齿动物胚胎是小鼠胚胎。在一些实施例中,转基因啮齿动物胚胎是杂合hDMD或杂合hDMD-Δ52。在一些实施例中,将包含包括SEQ ID NO:41中列出的核苷酸序列的靶向结构域的第一gRNA分子和包含包括SEQ ID NO:42中列出的核苷酸序列的靶向结构域的第二gRNA分子给予至啮齿动物胚胎以缺失人肌营养不良蛋白基因的外显子52。在一些实施例中,该方法进一步包括向啮齿动物胚胎给予包含SEQ ID NO:27中列出的氨基酸序列的Cas蛋白。本披露涉及通过此方法产生

的转基因啮齿动物胚胎。本披露还涉及由转基因啮齿动物胚胎产生的转基因啮齿动物。

9.构建体和质粒

[0186] 如上所述的组合物可以包含编码如本文所披露的基于CRISPR/Cas9的基因编辑系统的基因构建体。基因构建体(如质粒)可以包含编码基于CRISPR/Cas9的基因编辑系统如Cas9蛋白和Cas9融合蛋白和/或至少一种gRNA的核酸。如上所述的组合物可以包含编码经修饰的AAV载体的基因构建体和编码如本文所披露的基于CRISPR/Cas9的基因编辑系统的核酸序列。基因构建体(如质粒)可以包含编码基于CRISPR/Cas9的基因编辑系统的核酸。如上所述的组合物可以包含编码如本文所披露的经修饰慢病毒载体的基因构建体。

[0187] 基因构建体(如重组质粒或重组病毒颗粒)可以包含编码Cas9-融合蛋白和至少一种gRNA的核酸。在一些实施例中,基因构建体可以包含编码Cas9融合蛋白和至少两种不同gRNA的核酸。在一些实施例中,基因构建体可以包含编码Cas9融合蛋白和超过两种不同gRNA的核酸。在一些实施例中,基因构建体可以包含可操作地连接至编码至少一种gRNA分子和/或Cas9分子的核苷酸序列的启动子。在一些实施例中,启动子可操作地连接至编码第一gRNA分子、第二gRNA分子和/或Cas9分子的核苷酸序列。基因构建体可以存在于细胞中作为起作用的染色体外分子。基因构建体可以是线性微染色体,包括着丝粒、端粒或质粒或粘粒。

[0188] 基因构建体还可以是重组病毒载体的基因组的一部分,该重组病毒载体包括重组慢病毒、重组腺病毒和重组腺病毒相关病毒。基因构建体可以是生活在细胞中的活的减毒微生物或重组微生物载体中的遗传物质中的一部分。基因构建体可以包含核酸编码序列的基因表达的调控元件。调控元件可以是启动子、增强子、起始密码子、终止密码子或聚腺苷酸化信号。

[0189] 在某些实施例中,基因构建体是载体。载体可以是腺相关病毒(AAV)载体,其编码至少一种Cas9分子和至少一种gRNA分子;载体能够在哺乳动物的细胞中表达至少一种Cas9分子和至少gRNA分子。载体可以是质粒。载体可以用于体内基因疗法。载体可以是重组的。载体可以包含编码融合蛋白(如Cas9-融合蛋白或基于CRISPR/Cas9的基因编辑系统)的异源核酸。载体可以是质粒。载体可以用于用编码Cas9-融合蛋白或基于CRISPR/Cas9的基因编辑系统的核酸转染细胞,在其中发生Cas9-融合蛋白或基于CRISPR/Cas9的基因编辑系统表达的条件下培养并维持转化的宿主细胞。

[0190] 可以针对表达的稳定性和高水平优化编码序列。在一些情况下,选择密码子以减少RNA的二级结构形成,如因分子内键合形成的二级结构。

[0191] 载体可以包含编码基于CRISPR/Cas9的基因编辑系统的异源核酸,并且可以进一步包含起始密码子(其可以在基于CRISPR/Cas9的基因编辑系统编码序列的上游)和终止密码子(其可以在基于CRISPR/Cas9的基因编辑系统编码序列的下游)。起始密码子和终止密码子可以与基于CRISPR/Cas9的基因编辑系统编码序列一起处于阅读框中。载体还可以包含可操作地连接至基于CRISPR/Cas9的基因编辑系统编码序列的启动子。可操作地连接至基于CRISPR/Cas9的基因编辑系统编码序列的启动子可以是来自猿猴病毒40(SV40)的启动子、小鼠乳腺瘤病毒(MMTV)启动子、人免疫缺陷病毒(HIV)启动子如牛免疫缺陷病毒(BIV)长末端重复序列(LTR)启动子、莫洛尼病毒启动子、禽类白血病病毒(ALV)启动子、巨细胞病毒(CMV)启动子如CMV即时早期启动子、EB病毒(EBV)启动子、U6启动子如人U6启动子或劳斯肉瘤病毒(RSV)启动子。启动子还可以是来自以下的启动子:人基因如人泛素C(hUbC)、人肌

动蛋白、人肌球蛋白、人血红蛋白、人肌肉肌酸或人金属硫蛋白。启动子还可以是天然或合成的组织特异性启动子,如肌肉或皮肤特异性启动子。此类启动子的实例描述于美国专利申请公开号US 20040175727和US 20040192593中,将其内容通过引用以其全部内容并入本文。肌肉特异性启动子的实例包括Spc5-12启动子(描述于通过引用以其全部内容并入本文的美国专利申请公开号US20040192593;Hakim et al.Mol.Ther.Methods Clin.Dev.[Hakim等人,分子疗法-方法和临床开发](2014)1:14002;和Lai et al.Hum Mol Genet.[Lai等人,人类分子遗传学](2014)23(12):3189-3199中)、MHCK7启动子(描述于Salva et al.,Mol.Ther.[Salva等人,分子疗法](2007)15:320-329中)、CK8启动子(描述于Park et al.PLoS ONE[Park等人,公共科学图书馆综合](2015)10(4):e0124914中)和CK8e启动子(描述于Muir et al.,Mol.Ther.Methods Clin.Dev.[Muir等人,分子疗法-方法和临床开发](2014)1:14025中)。在一些实施例中,gRNA和/或Cas9蛋白的表达由tRNA驱动。

[0192] 编码gRNA分子和/或Cas9分子的每个多核苷酸序列可以各自可操作地连接至启动子。可操作地连接至gRNA分子和/或Cas9分子的启动子可以是相同的启动子。可操作地连接至gRNA分子和/或Cas9分子的启动子可以是不同的启动子。启动子可以是组成型启动子、诱导型启动子、阻抑型启动子或调控型启动子。

[0193] 载体还可以包含聚腺苷酸化信号,其可以在基于CRISPR/Cas9的基因编辑系统下游。聚腺苷酸化信号可以是SV40聚腺苷酸化信号、LTR聚腺苷酸化信号、牛生长激素(bGH)聚腺苷酸化信号、人生长激素(hGH)聚腺苷酸化信号或人β-珠蛋白聚腺苷酸化信号。SV40聚腺苷酸化信号可以是来自pCEP4载体(英杰公司(Invitrogen),加利福尼亚州圣迭哥)的聚腺苷酸化信号。

[0194] 载体还可以包含在基于CRISPR/Cas9的基因编辑系统(即,Cas9蛋白或Cas9融合蛋白编码序列或sgRNA,或基于CRISPR/Cas9的基因编辑系统)上游的增强子。增强子可能是DNA表达必需的。增强子可以是人肌动蛋白、人肌球蛋白、人血红蛋白、人肌肉肌酸或病毒增强子,如来自CMV、HA、RSV或EBV的增强子。多核苷酸功能增强子描述于美国专利号5,593,972、5,962,428和WO 94/016737中,将其各自的内容通过引用完全并入。载体还可以包含哺乳动物复制起点以在染色体外保持载体,并在细胞中产生多拷贝的载体。载体还可以包含调控序列,其非常适于在其中给予载体的哺乳动物或人细胞中的基因表达。载体还可以包含报告基因,如绿色荧光蛋白(“GFP”)和/或选择性标记,如潮霉素(“Hygro”)。

[0195] 载体可以是表达载体或通过常规技术和容易获得的起始原料产生蛋白质的系统,包括Sambrook et al.,Molecular Cloning and Laboratory Manual,Second Ed.,Cold Spring Harbor[Sambrook等人,分子克隆和实验室手册,第2版,冷泉港出版社](1989),将其通过引用完全并入。在一些实施例中,载体可以包含编码基于CRISPR/Cas9的基因编辑系统的核酸序列,包括编码Cas9蛋白或Cas9融合蛋白的核酸序列和编码至少一种gRNA的核酸序列,该至少一种gRNA包含SEQ ID NO:1-19、41、42中的至少一个的核酸序列或其互补序列。在一些实施例中,Cas9蛋白或Cas9融合蛋白由SEQ ID NO:26中任一个的核酸序列编码。在一些实施例中,载体包含SEQ ID NO:39或SEQ ID NO:40的核酸序列。

10.药物组合物

[0196] 本发明披露的主题提供了包含上述基因构建体的组合物。根据本发明的药物组合物可以根据有待使用的给药方式进行配制。在药物组合物是注射用药物组合物的情况下,

它们是无菌、无致热原且无颗粒的。优选使用等渗配制品。一般而言,用于等渗性的添加剂可以包括氯化钠、右旋糖、甘露糖醇、山梨糖醇和乳糖。在一些情况下,优选等渗溶液,如磷酸盐缓冲盐水。稳定剂包括明胶和白蛋白。在一些实施例中,将血管收缩剂添加到配制品中。

[0197] 组合物可以进一步包含药学上可接受的赋形剂。药学上可接受的赋形剂可以为作为媒介物、辅助剂、载体或稀释剂的功能性分子。药学上可接受的赋形剂可以是转染促进剂,该转染促进剂可以包括表面活性剂(如免疫刺激复合体(ISCOMS))、弗氏不完全佐剂、LPS类似物(包括单磷酰脂质A)、胞壁酰肽、醌类似物、囊泡(如角鲨烯和角鲨烯)、透明质酸、脂质、脂质体、钙离子、病毒蛋白、聚阴离子、聚阳离子或纳米粒子或其他已知的转染促进剂。

[0198] 转染促进剂为聚阴离子、聚阳离子(包括聚-L-谷氨酸(LGS))或脂质。转染促进剂为聚-L-谷氨酸,并且更优选地,聚-L-谷氨酸以小于6mg/ml的浓度存在于用于骨骼肌或心肌中的基因组编辑的组合物中。转染促进剂还可以包括表面活性剂(如免疫刺激复合体(ISCOMS))、弗氏不完全佐剂、LPS类似物(包括单磷酰脂质A)、胞壁酰肽、醌类似物和囊泡(如角鲨烯和角鲨烯),而且透明质酸也可以用来与基因构建体联合给予。在一些实施例中,编码组合物的DNA载体还可以包括转染促进剂如脂质、脂质体(包括卵磷脂脂质体或本领域已知的其他脂质体,像DNA-脂质体混合物(参见例如国际专利公开号WO9324640))、钙离子、病毒蛋白、聚阴离子、聚阳离子或纳米粒子或其他已知的转染促进剂。优选地,转染促进剂为聚阴离子、聚阳离子(包括聚-L-谷氨酸(LGS))或脂质。

11.递送方法

[0199] 本文提供了用于将本发明披露的基因构建体(例如,载体)或其组合物递送至细胞的方法。组合物的递送可以是组合物的转染或电穿孔,作为在细胞中表达并递送至细胞表面的核酸分子。可以使用Biorad Gene Pulser Xcell或Amaxa Nucleofector IIb装置将核酸分子电穿孔。可以使用几种不同的缓冲液,包括BioRad电穿孔溶液、Sigma磷酸盐缓冲盐水制品#D8537(PBS),Invitrogen OptiMEM I(OM)或Amaxa Nucleofector溶液V(N.V.)。转染可以包括转染试剂,如Lipofectamine 2000。

[0200] 在将本发明披露的基因构建体或组合物递送至组织并且随即将载体递送至哺乳动物的细胞中后,转染的细胞将表达该一种或多种gRNA分子和Cas9分子。基因构建体或组合物可以给予哺乳动物以改变基因表达或者重组或改变基因组。例如,可以将基因构建体或组合物给予哺乳动物以修正哺乳动物中的肌营养不良蛋白基因。哺乳动物可以是人、非人灵长类动物、牛、猪、绵羊、山羊、羚羊、野牛、水牛、牛科动物、鹿、刺猬、大象、美洲驼、羊驼、小鼠、大鼠或鸡,并且优选人、牛、猪或鸡。

[0201] 可以在有或没有体内电穿孔,脂质体介导的、纳米粒子促进的和/或重组载体的情况下,通过DNA注射(也称为DNA疫苗接种)将编码该一种或多种gRNA分子和Cas9分子的基因构建体(例如,载体)递送至哺乳动物。重组载体可以通过任何病毒方式递送。病毒方式可以是重组慢病毒、重组腺病毒和/或重组腺相关病毒。

[0202] 可以将本发明披露的基因构建体(例如,载体)或包含其的组合物引入细胞中以基因修正肌营养不良蛋白基因(例如,人肌营养不良蛋白基因)。在某些实施例中,将本发明披露的基因构建体(例如,载体)或包含其的组合物引入来自DMD患者的成肌细胞中。在某些实施例中,将基因构建体(例如,载体)或包含其的组合物引入来自DMD患者的成纤维细胞中,并且经基因修正的成纤维细胞可以用MyoD处理以诱导分化成成肌细胞,其可以被植入受试

者体内(如受试者的受损肌肉),以验证经修正的肌营养不良蛋白是功能性的和/或治疗受试者。经修饰的细胞也可以是干细胞,如诱导多能干细胞、骨髓衍生的祖细胞、骨骼肌祖细胞、来自DMD患者的人骨骼肌成肌细胞、CD 133^+细胞、中成血管细胞(mesoangioblast)和MyoD或Pax7转导的细胞、或其他生肌祖细胞。例如,基于CRISPR/Cas9的基因编辑系统可以导致诱导多能干细胞的神经元或肌源性分化。

12.给药途径

[0203] 可以通过不同的途径将本发明披露的基因构建体(例如,载体)或包含其的组合物给予受试者,这些不同途径包括口服、胃肠外、舌下、经皮、直肠、经粘膜、局部、经由吸入、经颊给药、胸膜内、静脉内、动脉内、腹膜内、皮下、肌内、鼻内、鞘内和关节内或其组合。在某些实施例中,将本发明披露的基因构建体(例如,载体)或组合物肌内、静脉内或其组合给予受试者(例如,患有DMD的受试者)。对于兽用,按照正常的兽医实践,将本发明披露的基因构建体(例如,载体)或组合物作为合适的可接受的配制品给予。兽医可以容易地确定最适合于具体动物的给药方案和给药途径。可以通过传统注射器、无针注射装置、“微弹轰击基因枪”或其他物理方法如电穿孔(“EP”)、“流体动力学方法”或超声给予组合物。

[0204] 可以在有或没有体内电穿孔,脂质体介导的、纳米粒子促进的、重组载体如重组慢病毒、重组腺病毒和重组腺病毒相关病毒的情况下,通过若干种技术包括DNA注射(也称为DNA疫苗接种)将本发明披露的基因构建体(例如,载体)或组合物递送至哺乳动物。可以将组合物注射到骨骼肌或心肌中。例如,可以将组合物注射到胫骨前肌或尾巴中。

[0205] 在一些实施例中,通过以下方式给予本发明披露的基因构建体(例如,载体)或其组合物:1)尾静脉注射(系统性)到成年小鼠中;2)肌内注射(例如,局部注射)到成年小鼠的肌肉如TA或腓肠肌中;3)腹膜内注射到P2小鼠中;或4)面静脉注射(系统性)到P2小鼠中。

13.细胞类型

[0206] 这些递送方法和/或给药途径中的任一种可以与无数种细胞类型一起应用,例如,目前正在研究用于DMD的基于细胞的疗法的那些细胞类型,包括但不限于永生化成肌细胞,如野生型和DMD患者衍生的细胞系,例如Δ48-50 DMD、DMD 6594(del48-50)、DMD 8036(del48-50)、C25C14和DMD-7796细胞系,原代DMD真皮成纤维细胞,诱导多能干细胞,骨髓衍生的祖细胞,骨骼肌祖细胞,来自DMD患者的人骨骼肌成肌细胞,CD 133^+细胞,中成血管细胞,心肌细胞,肝细胞,软骨细胞,间充质祖细胞,造血干细胞,平滑肌细胞和MyoD-或Pax7转导的细胞或其他生肌祖细胞。人生肌细胞的永生化可以用于基因修正的生肌细胞的克隆衍生。细胞可以被离体修饰以分离和扩增永生化DMD成肌细胞的克隆群体,这些DMD成肌细胞包括基因修正的肌营养不良蛋白基因,并且在基因组的蛋白质编码区中不含其他核酸酶引入的突变。可替代地,通过非病毒或非整合病毒基因转移或通过直接递送纯化的蛋白质和含有细胞穿透基序的gRNA进行的基于CRISPR/Cas9的系统的瞬时体内递送可以实现高度特异性的原位修正,有极小或没有外源DNA整合的风险。

14.试剂盒

[0207] 本文提供了可以用于修正突变的肌营养不良蛋白基因的试剂盒。试剂盒至少包含用于修正突变的肌营养不良蛋白基因的gRNA和用于使用基于CRISPR/Cas9的基因编辑系统的说明书。本文还提供了可以用于骨骼肌或心肌中肌营养不良蛋白基因的基因组编辑的试剂盒。试剂盒包含如上所述的用于骨骼肌或心肌中的基因组编辑的基因构建体(例如,载

体)或包含其的组合物以及用于使用所述组合物的说明书。

[0208] 试剂盒中所包括的说明书可以贴在包装材料上,或者可以作为包装插入物而被包括。虽然说明书通常是书面材料或印刷材料,但它们不限于此。任何能够存储此类说明书并将其传送给最终使用者的介质都被本披露所涵盖。这样的介质包括但不限于电子存储介质(例如,磁盘、磁带、盒式磁带(cartridge)、芯片)、光学介质(例如,CD ROM)等。如本文所用,术语“说明书”可以包括提供说明书的互联网网站的地址。

[0209] 用于修正骨骼肌或心肌中突变的肌营养不良蛋白或者肌营养不良蛋白基因的基因组编辑的基因构建体(例如,载体)或包含其的组合物可以包括含如上所述的一种或多种gRNA分子和Cas9分子的经修饰的AAV载体,该Cas9分子特异性结合并裂解肌营养不良蛋白基因的区域。如上所述的基于CRISPR/Cas9的基因编辑系统可以包括在试剂盒中以特异性结合和靶向突变的肌营养不良蛋白基因中的特定区域。如上所述,试剂盒可以进一步包括供体DNA、不同的gRNA或转基因。

15.实例

[0210] 对于本领域技术人员应容易地显而易见的是,本文描述的本披露的方法的其他适合修改和改编是可容易应用和感知的并且可以使用合适的等价方案进行而不脱离本披露或本文披露的方面和实施例的范围。现在已详细描述了本披露,同样将通过参考以下实例更清楚地理解,这些实例仅旨在说明本披露的一些方面和实施例,并且不应被视为对本披露的范围的限制。本文提到的所有期刊参考文献、美国专利和公开的披露通过引用以其全部内容并入本文。

[0211] 本发明具有多个方面,这些方面通过以下非限制性实例说明。

实例1

靶向人肌营养不良蛋白基因

[0212] 使用基于CRISPR/Cas9的基因编辑系统来靶向和缺失人肌营养不良蛋白基因的外显子51。比化脓链球菌Cas9小约1kb的金黄色葡萄球菌Cas9(SaCas9)与腺相关病毒(AAV)一起使用以递送基于CRISPR/Cas9的基因编辑系统。编码金黄色葡萄球菌Cas9分子的密码子优化的核酸序列在SEQ ID NO:43或SEQ ID NO:44中列出。图3显示了基于AAV的体内共同递送两种病毒载体上的SaCas9和两种gRNA至肌肉组织的示意图。每种载体具有分别由CMV和hU6启动子驱动的SaCas9和一种gRNA的拷贝(PT366-179(SEQ ID NO:39)和PT366-183(SEQ ID NO:40))。

[0213] 通过Surveyor测定确定了HEK293T细胞(具有正常型式的肌营养不良蛋白基因)和DMD患者成肌细胞系(DMD 8036和DMD 6594,其各自具有肌营养不良蛋白基因的突变形式)中针对人类基因组设计的靶向人肌营养不良蛋白基因的单独的gRNA JCR89(靶向外显子51的上游)和JCR91(靶向外显子51的下游)的活性(参见图1)。Surveyor测定检测基因组DNA中的错配,这指示着来自基于CRISPR/Cas9的基因编辑系统的插入缺失。对于JCR89,亲本带的大小是555nt,并且所用的引物是:正向引物-aagttacttgtccaggcatga(SEQ ID NO:91);和反向引物-gaaaaacttctgccaacttttatca(SEQ ID NO:92)。预期的切割带的大小为134nt和421nt。对于JCR91,亲本带的大小是632nt,并且所用的引物是:正向引物-tgcaaataacaaaagtagccataca(SEQ ID NO:93);和反向引物-tctttagaaaggcttgaaagctg(SEQ ID NO:94)。预期的切割带的大小为210nt和422nt。

[0214] 用SaCas9和gRNA JCR89和JCR91(SEQ ID NO:37和SEQ ID NO:38)共同处理

HEK293T细胞和DMD成肌细胞(DMD 8036和DMD 6594)。使用正向引物-cttcactgctggccagttta(SEQ ID NO:95)和反向引物-tctttagaaaggcttgaaagctg(SEQ ID NO:94)扩增基因组DNA。预期的亲本带的大小为1646nt,并且预期的“完美”缺失带为766nt(由于插入缺失的发生,gRNA切割位点之间的实际缺失大小与766nt不同)。图2A显示了HEK293T细胞和DMD成肌细胞的基因组DNA中外显子51的缺失。图2B显示了DMD成肌细胞的cDNA中外显子51的缺失。“无RT”是没有添加逆转录酶的阴性对照。

[0215] 将基于CRISPR/Cas9的基因编辑系统注射到携带人DMD基因的转基因小鼠(hDMD/mdx小鼠)中以缺失外显子51。将携带SaCas9和gRNA的病毒载体的局部AAV8递送应用于胫骨前肌(TA)肌肉。参见表1。3只小鼠注射AAV8:1只小鼠在两侧TA中注射高剂量的AAV8(“HH”),1只小鼠在两侧TA中注射低剂量的AAV8(“LL”),并且1只小鼠在左侧TA中注射低剂量且在右侧TA中注射高剂量(“LH”)。剂量列于表1的第2栏中。处理后8周(“周PT”)将小鼠处死以收获组织用于分析。巢式PCR揭示了HH小鼠两侧肢体、LL小鼠右侧TA和LH小鼠右侧肢体中外显子51的缺失。

表1

实验	AAV8剂量	递送	周PT	巢式PCR-gDNA Δ51
HH, LL, LH	HH: 6.6E11 LL: 1E11	肌内胫骨前肌(TA)	8	HH: 两侧, LL: R, LH:
				R

[0216] 在第一次PCR反应中,使用正向引物:cttcactgctggccagttta(SEQ ID NO:95)和反向引物:tctttagaaaggcttgaaagctg(SEQ ID NO:94)来扩增从小鼠TA肌肉收获的基因组DNA。在使用正向引物-aagttacttgtccaggcatga(SEQ ID NO:91)和反向引物-ttgaacatggcattgcataaA(SEQ ID NO:96)的第二次PCR反应(2X gDNA PCR)中,使用1-3μL PCR产物。这第二次PCR具有1089nt的预期的亲本带和323nt的预期的缺失带(由于插入缺失的发生,gRNA切割位点之间的实际缺失大小与323nt不同)。图4显示了第二次PCR结果。“L”泳道显示了左侧TA肌肉的结果,其被用作对照并接受盐水溶液。“R”泳道显示了右侧TA肌肉的结果,其被注射了等量预先混合的2种病毒载体。基于CRISPR/Cas9的基因编辑系统也通过系统性AAV8递送注射到hDMD/mdx小鼠的尾静脉中(参见图5)。从小鼠肝脏(图5-左图)和心脏(图5-右图)收获的基因组DNA也使用与图4相同的方案进行扩增。大约300个核苷酸的预期带指示外显子51的缺失。

[0217] 产生并选择靶向人和恒河猴肌营养不良蛋白基因序列或人肌营养不良蛋白基因序列的其他gRNA(参见图6和表2)。表2列出了gRNA的一般靶标、识别的基因组链、gRNA序列和与gRNA相关的PAM序列。表3列出了gRNA的靶基因组序列,如基因组+链所示。在培养的人细胞中测试了这些gRNA,以发现gRNA的最佳活性和组合以产生缺失。基于人类基因组中预测的特异性(图8),将选定的gRNA也优先化,并针对从19-23个核苷酸变化的最佳靶序列长度进行筛选(图10和11)。图6显示了表2中列出的人和恒河猴基因组之间保守的各gRNA靶

标。每种gRNA的位置都关于人肌营养不良蛋白基因的外显子51来指示。

表2 gRNA列表

gRNA	一般靶标	链	gRNA序列（5'-3'）	SEQ ID NO:	PAM序列	SEQ ID NO:
JCR89	人DMD外显子51（上游）	+	AAAGATATATAATGTCATGAAT	1	AAGAGT	53
JCR91	人DMD外显子51（下游）	+	GCAGAATCAAATATAATAGTCT	2	GGGAAT	54
JCR94	人DMD外显子52（上游）	-	AACAAATATCCCTTAGTATC	41	AGG	55
JCR99	人DMD外显子52（下游）	-	AATGTATTTCTTCTATTCAA	42	TGG	56
JCR159	人和恒河猴DMD外显子51（上游）	-	AACAATAAGTCAAATTTAATTG	3	AAGAGT	53
JCR160	人和恒河猴DMD外显子51（下游）	-	GAACTGGTGGGAAATGGTCTAG	4	GAGAGT	57
JCR167	人和恒河猴DMD外显子51（下游）	-	TCCTTTGGTAAATAAAAGTCCT	5	GGGAGT	58
JCR166	人和恒河猴DMD外显子51（下游）	-	TAGGAATCAAATGGACTTGGAT	6	TTGAAT	59
JCR168	人和恒河猴DMD外显子51（上游）	+	TAATTCTTTCTAGAAAGAGCCT	7	CAGAGT	60
JCR170	人和恒河猴DMD外显子51	-	CTCTTGCATCTTGCACATGTCC	8	TGGAGT	61

	（上游）					
JCR171	人和恒河猴DMD外显子51（上游）	-	ACTTAGAGGTCTTCTACATACA	9	ATGAGT	62
JCR156	人和恒河猴DMD外显子51（上游）	-	TCAGAGGTGAGTGGTGAGGGGA	10	AGGAAT	63
JCR157	人和恒河猴DMD外显子51（上游）	-	ACACACAGCTGGGTTATCAGAG	11	GAGAGT	57
JCR176	人和恒河猴DMD外显子51（上游）	-	CACAGCTGGGTTATCAGAG	12	GAGAGT	57
JCR177	人和恒河猴DMD外显子51（上游）	-	ACACAGCTGGGTTATCAGAG	13	GAGAGT	57
JCR178	人和恒河猴DMD外显子51（上游）	-	CACACAGCTGGGTTATCAGAG	14	GAGAGT	57
JCR179	人和恒河猴DMD外显子51（上游）	-	AACACACAGCTGGGTTATCAGAG	15	GAGAGT	57
JCR180	人和恒河猴DMD外显子51（下游）	-	CTGGTGGGAAATGGTCTAG	16	GAGAGT	57
JCR181	人和恒河猴DMD外显子51（下游）	-	ACTGGTGGGAAATGGTCTAG	17	GAGAGT	57

JCR182	人和恒河猴DMD外显子51（下游）	-	AACTGGTGGGAAATGGTCTAG	18	GAGAGT	57
JCR183	人和恒河猴DMD外显子51（下游）	-	AGAACTGGTGGGAAATGGTCTAG	19	GAGAGT	57

表3 gRNA的靶序列

gRNA名称	靶序列（+链 5'-3'）	SEQ ID NO:	注解
JCR89	AAAGATATATAATGTCATGAAT	64	
JCR91	GCAGAATCAAATATAATAGTCT	65	
JCR159	CAATTAAATTTGACTTATTGTT	66	外显子51上游101
JCR160	CTAGACCATTTCCCACCAGTTC	67	外显子51下游78
JCR167	AGGACTTTTATTTACCAAAGGA	68	外显子51下游1534
JCR166	ATCCAAGTCCATTTGATTCCTA	69	外显子51下游1266
JCR168	TAATTCTTTCTAGAAAGAGCCT	70	外显子51上游1824
JCR170	GGACATGTGCAAGATGCAAGAG	71	外显子51上游2851
JCR171	TGTATGTAGAAGACCTCTAAGT	72	外显子51上游2947
JCR156	TCCCCTCACCACTCACCTCTGA	73	外显子51上游1300
JCR157	CTCTGATAACCCAGCTGTGTGT	74	外显子51上游1284
JCR176	CTCTGATAACCCAGCTGTG	75	JCR157 - 19核甘酸
JCR177	CTCTGATAACCCAGCTGTGT	76	JCR157 - 20核甘酸
JCR178	CTCTGATAACCCAGCTGTGTG	77	JCR157 - 21核甘酸
JCR179	CTCTGATAACCCAGCTGTGTGTT	78	JCR157 - 23核甘酸
JCR180	CTAGACCATTTCCCACCAG	79	JCR160 - 19核甘酸
JCR181	CTAGACCATTTCCCACCAGT	80	JCR160 - 20核甘酸
JCR182	CTAGACCATTTCCCACCAGTT	81	JCR160 - 21核甘酸
JCR183	CTAGACCATTTCCCACCAGTTCT	82	JCR160 - 23核甘酸

[0218] 用表2中列出的单独的候选gRNA转染人HEK293T细胞。通过Surveyor测定确定候选gRNA的活性(参见图7)。对于JCR160,亲本带的大小是483nt,并且所用的引物是:正向引物-cgggcttggacagaacttac(SEQ ID NO:97);和反向引物-ctgcgtagtgccaaaacaaa(SEQ ID NO:

98)。预期的切割带的大小为192nt和291nt。对于JCR157,亲本带的大小是631nt,并且所用的引物是:正向引物-gagatgtcttttgcagctttcc(SEQ ID NO:99);和反向引物-gggaccttggtaaagccaca(SEQ ID NO:100)。预期的切割带的大小为147nt和484nt。

[0219] 使用CasOFFinder程序预测候选gRNA的特异性(Bae et al.(2014) Bioinformatics[Bae等人(2014)生物信息学]30:1473-1475:参见图8)。基于通过Surveyor测定测量的脱靶活性、中靶活性和距外显子的距离来评估和选取候选gRNA。gRNA JCR157和JCR160具有低的预测脱靶结合并用于进一步测试。

[0220] 用含有gRNA JCR157的经修饰的pDO240质粒、含有gRNA JCR160的经修饰的pDO240质粒和含有SaCas9的质粒(pDO242;SEQ ID NO:83)对HEK293T细胞进行转染并对DMD6594细胞进行电穿孔。预测亲本带为2451nt,并且预测缺失带为约840-850nt。图9显示了使用正向引物-tgcctttcaatcattgtttcg(SEQ ID NO:101)和反向引物-agaaggcaaattggcacaga(SEQ ID NO:102)通过基因组DNA的PCR(大约850个核苷酸)确定的外显子51的缺失。在gRNA切割位点之间产生的缺失为大约1611nt。

[0221] 图10显示了使用图7中针对JCR157的引物和PCR条件,通过Surveyor测定确定HEK293T细胞中各靶标长度的gRNA JCR157(19、20、21、22和23个核苷酸)的活性。图11显示了使用正向引物-cgggcttggacagaacttac(SEQ ID NO:97)和反向引物-ctgcgtagtgccaaaacaaa(SEQ ID NO:98),通过Surveyor测定确定HEK293T细胞中各靶标长度的gRNA JCR160(19、20、21、22和23个核苷酸)的活性。预测亲本带的大小为483nt,并且预期的切割带的大小为209nt和274nt。

[0222] 使用图9中使用的条件,在HEK293T细胞中使用各靶标长度的gRNA JCR157和JCR160(21、22或23个核苷酸)的组合。图12显示了基因组DNA的PCR。各自具有23个核苷酸靶标的JCR157和JCR160的组合具有几乎50%的缺失。

[0223] 使用23nt的靶序列用SaCas9单独执行外显子51侧翼的每种gRNA(上游JCR179和下游JCR183)以证明HEK293T细胞("293s")和DMD6594细胞("DMD6594s")中的中靶核酸酶活性(参见图13)。对于JCR179,亲本带的大小是594nt,并且所用的引物是:正向引物-tgcctttcaatcattgtttcg(SEQ ID NO:101);和反向引物-aaggccccaaaatgtgaaat(SEQ ID NO:103)。预期的切割带的大小为594nt和130nt。对于JCR183,亲本带的大小是731nt,并且所用的引物是:正向引物-gagtttggctcaaattgttactctt(SEQ ID NO:104);和反向引物-ctgcgtagtgccaaaacaaa(SEQ ID NO:98)。预期的切割带的大小为440nt和291nt。图13显示了如通过Surveyor测定测得的体外中靶核酸酶活性。

[0224] 用含有SaCas9以及gRNA JCR179和JCR183(JCR157和JCR160的23nt靶标;SEQ ID NO:37和SEQ ID NO:38)的质粒对人HEK293T细胞进行转染并对DMD成肌细胞(DMD 6594s)进行电穿孔。DMD 6594细胞是已经缺乏外显子48-50的永生化DMD患者成肌细胞。预测亲本带为2451nt,并且预测缺失带为约823nt。图14显示了使用正向引物-tgcctttcaatcattgtttcg(SEQ ID NO:101)和反向引物-agaaggcaaattggcacaga(SEQ ID NO:102)如通过基因组DNA的PCR确定的人HEK293T细胞(左图)和DMD 6594s细胞(右图)的基因组DNA中外显子51的体外缺失。在gRNA切割位点之间产生的缺失为大约1628nt。顶图显示了HEK293T细胞和DMD 6594细胞中上游和下游gRNA的靶基因的示意图,其中紫色指示正常加工的外显子,并且黄色指示突变型外显子。中图显示了跨越基因组缺失区域的PCR结果,其中星号指示缺失。底

图显示了基因组DNA的液滴数字PCR。在HEK293T细胞中,gRNA和SaCas9具有16%的缺失,其中DMD6594细胞具有约10%的编辑。

[0225] 为了确定基因组DNA中的变化是否转录为RNA,从用SaCas9和两种gRNA共转染并分化7天的DMD成肌细胞收获RNA(参见图15)。使用本领域已知的标准方法,将RNA逆转录成cDNA,并对cDNA进行PCR扩增。在图15中,左下图显示了使用正向引物-tggcggcgttttcattat(SEQ ID NO:105)和反向引物-TTCGATCCGTAATGATTGTTCTAGCC(SEQ ID NO:106)从外显子44到外显子52的PCR扩增。预测亲本带为948nt,并且预测缺失带为约715nt。图15显示了用SaCas9和两种gRNA处理的细胞中仅有的缺失带。右下图显示了揭示约14%的cDNA的编辑的ddPCR。对DMD患者成肌细胞cDNA的体外外显子47至52的连接进行测序(参见图16)。在图16中,来自未经处理的细胞(对照细胞Δ48-50)的带序列指示外显子47至51连接,如预期的那样,而经处理的细胞中的缺失带(Δ48-50+Δ51)显示了外显子47至52的连接。因此,外显子51的明显缺乏和针对基因组DNA水平的所披露的系统正在通过转录进行。

实例2

Δ52/mdx小鼠的产生

[0226] 图17显示了Δ52/mdx小鼠的设计。从莱顿大学获得hDMD/mdx小鼠并操作以产生DMD的相关模型,其中外显子52被去除并且该缺失导致阅读框外移和DMD基因型。在mdx背景下,hDMD/mdx小鼠在5号染色体上含有全长野生型人肌营养不良蛋白基因,因此不表达小鼠肌营养不良蛋白。

[0227] 为了产生DMD的相关模型,使用SpCas9 CRISPR/Cas9编辑系统和gRNA靶向和缺失人肌营养不良蛋白基因的外显子52。使用Surveyor测定来测试和验证靶向外显子52上游和下游的各种gRNA(参见图18)。对于上游gRNA,使用正向引物-ctccggaatgtctccatttg(SEQ ID NO:87)和反向引物-TTGTGTGTCCCATGCTTGTT(SEQ ID NO:107),并且亲本带的大小为402nt。对于JCR94(AACAAATATCCCTTAGTATC(SEQ ID NO:41)),预期的切割的大小为243nt和159nt。对于下游gRNA,使用正向引物-CAACGCTGAAGAACCCTGAT(SEQ ID NO:108)和反向引物-atgagggagagactggcatc(SEQ ID NO:88),并且亲本带的大小为509nt。对于JCR99(AATGTATTTCTTCTATTCAA(SEQ ID NO:42)),预期的切割的大小为346nt和163nt。使用正向引物-ctccggaatgtctccatttg(SEQ ID NO:87)和反向引物-atgagggagagactggcatc(SEQ ID NO:88),通过检测HEK293T细胞(包括JCR94和JCR99)的基因组DNA中外显子51的缺失来测试和验证gRNA对(参见图19)。亲本带的大小为718nt,缺失带为392nt,并且gRNA之间的缺失为326nt。在基因组编辑系统中使用这对JCR94和JCR99来产生Δ52/mdx小鼠。确切地说,通过将JCR94 gRNA、JCR99 gRNA和SaCas9 mRNA注射到小鼠胚胎中来创建小鼠。

[0228] 图20显示了包括BAC重组工程服务的DNA显微注射方案。第1天:用血清促性腺激素腹膜内处理孕鼠以诱导排卵。第3天:用人绒毛膜促性腺激素腹膜内处理孕鼠。在第1轮中,产生了471个胚胎,但只有5个塞(plug)可见。使用150个受精的胚胎(14只雌鼠超排卵)。用少于50ng Cas9且每种指导物20ng进行原核注射。图21显示了产生转基因小鼠的小鼠繁殖方案。

[0229] 使用以下基因分型方案对首建小鼠进行基因分型。使用DNEasy血液和组织试剂盒(凯杰公司(QIagen))从小鼠的尾巴剪块中提取基因组DNA(gDNA)。为了对每只幼鼠进行基因分型,如下使用AccuPrime HiFi Taq试剂盒扩增gDNA:i.100ng gDNA;ii.2.5μL

AccuPrime缓冲液II;iii.0.1μL AccuPrime HiFi Taq;iv.1μL JRH261(ctccggaatgtctccatttg(SEQ ID NO:87))(10μM);v.1μL JRH264(atgagggagagactggcatc(SEQ ID NO:88))(10μM);以及vi.加水到25μL总体积。如下在热循环仪上运行反应:i.95度持续4分钟;ii.95度持续30秒;iii.52度持续30秒;iv.68度持续1:00分钟;v.将步骤ii-iv循环35次;和4度恒温。在凝胶上分离PCR反应(图22)。如果不存在缺失(即,外显子52仍然存在),则预期的带的大小为718nt,并且如果存在外显子52的缺失,则预期的带的大小为大约392nt。如图22所示,首建小鼠7、63和76具有外显子52缺失。

[0230] 使用JRH264引物对经扩增的带进行测序(参见图23)以测序靶向区域的重新连接的末端。图23显示了测序区域,其中粗体、加下划线和正常字母指示天然序列,并且斜体字母指示插入或缺失。在预期序列("Δ52")中,粗体字母连接带下划线的字母。在首建小鼠中,在此区域存在插入(斜体字母)和缺失(连字符)。

[0231] 使雄性首建小鼠与mdx/mdx雌鼠交配以繁殖出嵌合体(图24)。使用在图22中使用的条件筛选来自首建雄鼠76或首建雄鼠63与mdx/mdx雌鼠产生的仔,并对外显子52缺失进行基因分型(分别为图25和图26)。如果外显子52缺失,则预期的带的大小为约392nt。如果外显子52存在,则预期的带的大小为约718nt。幼鼠54497和54498(来自首建雄鼠63+mdx/mdx雌鼠繁殖对)具有外显子52缺失并被测序(图27)。在392bp测序读数中,幼鼠54497和54498彼此具有92.86%同一性,并且插入缺失是相同的。

[0232] 使用荧光免疫组织化学比较健康hDMD/mdx小鼠和Δ52/mdx小鼠中肌营养不良蛋白的表达。如图28所示,Δ52/mdx小鼠幼仔54497和54498缺乏肌营养不良蛋白。对于心脏染色,层粘连蛋白探针的暴露为100ms,而肌营养不良蛋白的暴露为900ms。对于TA肌肉样品,层粘连蛋白和肌营养不良蛋白探针的暴露时间为2.0s。还参见图29,其显示Δ52/mdx小鼠缺乏肌营养不良蛋白。在心脏和TA肌肉两者中,Δ52/mdx小鼠中失去了肌营养不良蛋白表达。在TA肌肉中,有一些自发回复突变纤维(随机剪接事件或体细胞突变),但这与mdx小鼠模型一致。蛋白质印迹也指示,Δ52/mdx小鼠缺乏与DMD基因型一致的肌营养不良蛋白,而健康的hDMD/mdx小鼠表达肌营养不良蛋白。参见图30。

[0233] 在旷场试验的前五分钟后,Δ52/mdx小鼠显示出与mdx小鼠相似的活性水平。在旷场试验中,允许小鼠自由地探索旷场竞技场(20x20x30cm)30分钟。使用与计算机运行Fusion Activity软件(版本5.3,欧姆尼技术公司(Omnitech),俄亥俄州哥伦布市)接口的红外二极管(x、y和z轴)来自动监测动物的活动和位置。图31显示了如通过运动和探索所示的,Δ52/mdx小鼠与mdx小鼠和hDMD/mdx小鼠相比的整体活动。行进距离和直立的竖直活动分别显示在左图和右图中。

实例3

通过去除外显子51进行的肌营养不良蛋白恢复

[0234] 在Δ52/mdx小鼠中,外显子51的去除可以产生贝克肌营养不良(BMD)样基因型,并且理论上恢复肌营养不良蛋白表达。使用Δ52/mdx小鼠来证明使用所披露的基于CRISPR/Cas9的基因编辑系统,通过去除外显子51而进行的肌营养不良蛋白表达的恢复。图32显示了使用SaCas9和gRNA,通过靶向内含子区域中的外显子51的上游和下游的gRNA来跳过外显子51的修正策略。

[0235] 标准:将含有gRNA JCR179(上游)或JCR183(下游)(分别为SEQ ID NO:37和SEQ ID

NO:38)和SaCas9的质粒电穿孔到DMD患者成肌细胞中。从分化的细胞中收获蛋白质,并使用肌营养不良蛋白抗体的蛋白质印迹进行分析。图33显示可以编辑基因组DNA以恢复肌营养不良蛋白,因为用全部3种组分(即,gRNA和SaCas9两者)处理的细胞显示出肌营养不良蛋白表达。

[0236] 然后用Δ52/mdx小鼠对系统进行测试。图34显示了使用gRNA和SaCas9系统处理Δ52/mdx小鼠的实验设计,包括在实验设计中使用的2种病毒载体的示意图。使用载体PT366-179(SEQ ID NO:39)和PT366-183(SEQ ID NO:40)和本领域已知的产生病毒颗粒的方法来创建AAV8重组病毒构建体。这些病毒载体(AAV8)作为两种病毒颗粒在体内共同递送。每种病毒颗粒含有SaCas9和一种gRNA(参见图3)。用5E11的AAV8重组病毒构建体处理Δ52/mdx小鼠。将病毒肌内注射到右侧TA肌肉中,而左侧TA肌肉充当对侧对照并注射PBS。处理后,去除左侧和右侧TA肌肉,并取各自的切片进行基因组DNA分析。如图35所示,跨感兴趣的区域进行PCR,并且在左侧凝胶上记录了经处理的右侧TA肌肉中的缺失带,指示一定水平的基因编辑。对缺失带进行测序,并且主要产物是来自每种gRNA的PAM的3个碱基对的预期连接。图35显示了右侧TA肌肉中的体内外显子51缺失。

[0237] 类似地,分析经处理的右侧TA和对照左侧TA肌肉的切片以确定编辑是否进行到RNA。跨cDNA中的外显子50至53进行PCR。如图36的右侧凝胶所示,在两个经处理的样品中缺乏外显子51和52。正如预期的那样,鉴于小鼠已经缺乏外显子52并且基于CRISPR/Cas9的基因编辑系统去除了外显子51,所以对缺失带进行测序并且主要产物是外显子50至53的连接(参见右下序列色谱图)。图36显示了右侧TA肌肉中的体内外显子51缺失。

[0238] 图37显示了代表性荧光免疫组织化学染色,指示在Δ52/mdx小鼠的对照PBS注射的左侧TA肌肉中存在微量肌营养不良蛋白。对照左侧TA上的一定程度的呈绿色的肌营养不良蛋白染色可能归因于回复突变纤维或死细胞(有时也染绿色)。经处理的右侧TA肌肉中绿色肌营养不良蛋白染色有明显增加,如右图所示。图37显示了经处理的TA肌肉中的体内肌营养不良蛋白恢复。

[0239] 从3只测试小鼠的左侧和右侧TA肌肉提取蛋白质,并进行蛋白质印迹分析。图38显示了经处理的TA肌肉中的体内肌营养不良蛋白恢复。在对照左侧TA肌肉中没有看到蛋白质表达,而所有三个右侧TA肌肉都展示不同水平的肌营养不良蛋白表达。来自小鼠1的右侧TA的蛋白质表达最强,而小鼠2和3有微弱但仍然存在的带。所披露的基于CRISPR/Cas9的基因编辑系统在体内起作用以在Δ52/mdx小鼠中一定程度上恢复肌营养不良蛋白表达。

[0240] 小鼠生理学测试。用含有SaCas9和gRNA的AAV8(n=10)或AAV9(n=10)重组病毒构建体处理经处理的小鼠(所有雄性hDMD-Δ52(het)/mdx(hemi)小鼠),并与未经处理的小鼠(hDMD-Δ52/mdx)(n=10)进行比较。使用载体PT366-179(SEQ ID NO:39)和PT366-183(SEQ ID NO:40)并使用本领域已知的方法创建AAV重组病毒构建体。将经处理的小鼠在6-8周龄之间尾静脉注射200μL病毒。小鼠在8周后进行测试。

[0241] 旷场距离测试。允许小鼠自由探索旷场竞技场30分钟。使用与计算机运行Fusion Activity软件接口的红外二极管来自动监测动物的活动和位置。连续收集数据并将其分档为5分钟的时间间隔。图39显示了与未经处理的16周龄小鼠相比,在8周龄时处理的16周龄小鼠移动总距离的所有时间点的平均值。图40显示了16周后总饲养姿势的所有时间点的平均值。统计资料为:单因素ANOVA,将每列均值与未经处理的均值进行比较,以及邓尼特事后

检验(Dunnett post hoc)(均值+/-SEM)。与未经处理的年龄匹配的小鼠相比,经AAV8和AAV9处理的小鼠显示出统计显著更大的行进距离(统计学显著)。与未经处理的年龄匹配的对照相比,所有经处理的小鼠显示出统计学显著增加的饲养姿势量。

[0242] 握力。测试16周的未经处理和经处理的小鼠的握力。为了测试握力,对小鼠进行了分别针对前后脚的3-5次试验。图41显示了平均值试验。握力以克力报告。如图41所示,与未经处理的年龄匹配的小鼠相比,经AAV9处理的小鼠显示出统计学显著增加的前爪握力。统计资料为:双因素ANOVA,图基事后检验(Tukey's test post hoc)。

[0243] cDNA PCR。使用RNEasy Plus通用迷你试剂盒(凯杰公司)处理来自小鼠心脏的组织。使用SuperScript VILO cDNA合成试剂盒将得到的RNA逆转录为eDNA。使用AccuPrime DNA聚合酶和外显子48(正向引物:gtttccagagctttacctgagaa(SEQ ID NO:89))和外显子54(反向引物:CTTTTATGAATGCTTCTCCAAG(SEQ ID NO:90))的引物对1μL cDNA进行PCR扩增。如果不存在缺失(即,外显子52仍然存在),则预期的带的大小为997nt,并且如果存在外显子52的缺失,则预期的带的大小为大约764nt。图42显示了面静脉注射AAV9("JA10(P2AAV9)")的36-50小时龄之间的P2小鼠,以及尾静脉注射3.3-7.7E12AAV8("JA11(TV AAV8)")或4.3-7.5E12AAV9("JA12(TV AAV9)")的成年小鼠。如图42所示,在经AAV9处理的P2小鼠(48-54小时龄的小鼠)和经AAV8和AAV9处理的成年小鼠中发生不同程度的编辑,这通过使用与外显子53结合的引物tttctgtgattttcttttggattg(SEQ ID NO:109)对缺失带进行测序而得到进一步证实。图43显示了代表性色谱图,显示来自Ja10小鼠1的缺失带的序列中外显子51和52的缺失。

[0244] 应理解以上具体实施方式和随附实例仅是说明性的,并且不应视为对本发明范围的限制,该范围仅由随附权利要求和它们的等效物限定。

[0245] 所披露的实施例的各种变化和修改对本领域技术人员将是显而易见的。可以在不脱离本发明的精神和范围的情况下进行此类变化和修改,包括但不限于与本发明的化学结构、取代基、衍生物、中间体、合成、组合物、配制品或使用方法有关的那些变化和修改。

[0246] 出于完整性的原因,本发明的各个方面将在以下编号的条款中阐述:

[0247] 条款1.一种包含靶向结构域的指导RNA(gRNA),该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:41、SEQ ID NO:42中列出的核苷酸序列或其互补序列。

[0248] 条款2.一种包含第一gRNA和第二gRNA的DNA靶向组合物,该第一gRNA分子和该第二gRNA分子包含靶向结构域,该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:41、SEQ ID NO:42中列出的核苷酸序列或其互补序列,其中该第一gRNA分子和该第二gRNA分子包含不同的靶向结构域。

[0249] 条款3.如条款2所述的DNA靶向组合物,其中该第一gRNA分子是SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID

NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15或SEQ ID NO:41,并且该第二gRNA分子是SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19或SEQ ID NO:42。

[0250] 条款4.如条款2或3所述的DNA靶向组合物,其中该第一gRNA分子选自下组,该组由以下组成:SEQ ID NO:1、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14和SEQ ID NO:15,并且该第二gRNA分子选自下组,该组由以下组成:SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18和SEQ ID NO:19。

[0251] 条款5.如条款2-4中任一项所述的DNA靶向组合物,其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:1中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:2中列出的核苷酸序列;(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列:(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(x)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列;(xi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;以及(xii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:41中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:42中列出的核苷酸序列。

[0252] 条款6.如条款2-5中任一项所述的DNA靶向组合物,进一步包含成簇规律间隔短回文重复序列相关(Cas)蛋白。

[0253] 条款7.如条款6所述的DNA靶向组合物,其中该Cas蛋白包含识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的Cas9分子。

[0254] 条款8.如条款6或7所述的DNA靶向组合物,其中该Cas蛋白包含具有SEQ ID NO:45的氨基酸序列的金黄色葡萄球菌Cas9分子。

[0255] 条款9.如条款2-8中任一项所述的DNA靶向组合物,其中该DNA靶向组合物包含SEQ ID NO:83的核苷酸序列、SEQ ID NO:84的核苷酸序列、SEQ ID NO:37的核苷酸序列和/或SEQ ID NO:38的核苷酸序列。

[0256] 条款10.一种分离的多核苷酸,包含如条款1所述的gRNA分子或如条款2-9中任一项所述的DNA靶向组合物。

[0257] 条款11.一种载体,包含如条款1所述的gRNA、如条款2-9中任一项所述的DNA靶向组合物或如条款10所述的分离的多核苷酸。

[0258] 条款12.一种载体,包含如条款6-9中任一项所述的DNA靶向组合物。

[0259] 条款13.一种载体,编码:(a)第一指导RNA(gRNA)分子,(b)第二gRNA分子和(c)至少一种识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的Cas9分子,其中该第一gRNA分子和该第二gRNA分子包含靶向结构域,该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19、SEQ ID NO:41、SEQ ID NO:42中列出的核苷酸序列或其互补序列,其中该第一gRNA分子和该第二gRNA分子包含不同的靶向结构域。

[0260] 条款14.如条款13所述的载体,其中该载体被配置为形成在人DMD基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂。

[0261] 条款15.如条款13或14所述的载体,其中该第一gRNA分子是SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:]5或SEQ ID NO:41,并且该第二gRNA分子是SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19或SEQ ID NO:42。

[0262] 条款16.如条款13-15中任一项所述的载体,其中该第一gRNA分子选自下组,该组由以下组成:SEQ ID NO:1、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14和SEQ ID NO:15,并且该第二gRNA分子选自下组,该组由以下组成:SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18和SEQ ID NO:19。

[0263] 条款17.如条款13-16中任一项所述的载体,该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:1中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:2中列出的核苷酸序列;(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序

列;(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(x)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列;以及(xi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列。

[0264] 条款18.如条款11-17中任一项所述的载体,其中该载体是病毒载体。

[0265] 条款19.如条款18所述的载体,其中该载体是腺相关病毒(AAV)载体。

[0266] 条款20.如条款19所述的载体,其中该AAV载体是AAV8载体或AAV9载体。

[0267] 条款21.如条款11-20中任一项所述的载体,其中该载体包含可操作地连接至编码该第一gRNA分子、该第二gRNA分子和/或该Cas9分子的核苷酸序列的组织特异性启动子。

[0268] 条款22.如条款21所述的载体,其中该组织特异性启动子是肌肉特异性启动子。

[0269] 条款23.一种细胞,包含如条款1所述的gRNA、如条款2-9中任一项所述的DNA靶向组合物、如条款10所述的分离的多核苷酸或如条款11-22中任一项所述的载体。

[0270] 条款24.一种试剂盒,包含如条款1所述的gRNA、如条款2-9中任一项所述的DNA靶向系统、如条款10所述的分离的多核苷酸、如条款11-22中任一项所述的载体或如条款23所述的细胞以及任选地使用说明书。

[0271] 条款25.一种修正细胞中突变型肌营养不良蛋白基因的方法,该方法包括向细胞给予如条款1所述的gRNA、如条款2-9中任一项所述的DNA靶向系统、如条款10所述的分离的多核苷酸或如条款11-22中任一项所述的载体。

[0272] 条款26.一种基因组编辑受试者中突变型肌营养不良蛋白基因的方法,该方法包括向该受试者给予包含如条款1所述的gRNA、如条款2-9中任一项所述的DNA靶向系统、如条款10所述的分离的多核苷酸、如条款11-22中任一项所述的载体或如条款23所述的细胞的基因组编辑组合物。

[0273] 条款27.如条款26所述的方法,其中将该基因组编辑组合物肌内、静脉内或其组合给予该受试者。

[0274] 条款28.如条款25-27中任一项所述的方法,其中修正该突变型肌营养不良蛋白基因包括核酸酶介导的非同源末端连接。

[0275] 条款29.一种治疗具有突变型肌营养不良蛋白基因的有需要的受试者的方法,该

方法包括向该受试者给予如条款1所述的gRNA、如条款2-9中任一项所述的DNA靶向系统、如条款10所述的分离的多核苷酸、如条款11-22中任一项所述的载体或如条款23所述的细胞。

[0276] 条款30.一种用于基因组编辑受试者中突变型肌营养不良蛋白基因的经修饰腺相关病毒载体,包含编码如条款1所述的gRNA的第一多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的Cas9分子的第二多核苷酸序列。

[0277] 条款31.如条款30所述的经修饰腺相关病毒载体,其中该经修饰腺相关病毒载体包含SEQ 1D NO:39或SEQ ID NO:40中列出的核苷酸序列。

[0278] 条款32.一种用于缺失包含外显子51的肌营养不良蛋白基因区段的组合物,该组合物包含:(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第一Cas9分子的多核苷酸序列,以及(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第二Cas9分子的多核苷酸序列,其中该第一和第二gRNA分子各自具有长度为19至24个核苷酸的靶向结构域,并且其中该第一载体和第二载体被配置为分别形成在人DMD基因的外显子51侧翼的第一内含子和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段。

[0279] 条款33.如条款32所述的组合物,其中该区段具有约50个碱基对至约2,000个碱基对的长度。

[0280] 条款34.如条款33所述的组合物,其中该区段具有约118个碱基对、约233个碱基对、约326个碱基对、约766个碱基对、约805个碱基对或约1611个碱基对的长度。

[0281] 条款35.如条款32-34中任一项所述的组合物,其中该第一Cas9分子和该第二Cas9分子是相同的。

[0282] 条款36.如条款35所述的组合物,其中该第一Cas9分子和该第二Cas9分子是金黄色葡萄球菌Cas9分子。

[0283] 条款37.如条款36所述的组合物,其中该第一Cas9分子和该第二Cas9分子是突变型金黄色葡萄球菌Cas9分子。

[0284] 条款38.如条款32-34中任一项所述的组合物,其中该第一Cas9分子和该第二Cas9分子是不同的。

[0285] 条款39.如条款38所述的组合物,其中该第一Cas9分子或该第二Cas9分子是金黄色葡萄球菌Cas9分子。

[0286] 条款40.如条款32-39中任一项所述的组合物,其中该第一Cas9分子和/或该第二Cas9分子包含具有SEQ ID NO:45的氨基酸序列的SaCas9分子。

[0287] 条款41.如条款32-40中任一项所述的组合物,其中该第一载体和/或该第二载体是病毒载体。

[0288] 条款42.如条款41所述的组合物,其中该第一载体和/或该第二载体是腺相关病毒(AAV)载体。

[0289] 条款43.如条款42所述的组合物,其中该AAV载体是AAV8载体或AAV9载体。

[0290] 条款44.如条款32-43中任一项所述的组合物,其中该肌营养不良蛋白基因是人肌

营养不良蛋白基因。

[0291] 条款45.如条款32-44中任一项所述的组合物,其中该第一gRNA分子和该第二gRNA分子包含靶向结构域,该靶向结构域包含SEQ ID NO:1、SEQ ID NO:2、SEQ ID NO:3、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14、SEQ ID NO:15、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18、SEQ ID NO:19中列出的核苷酸序列或其互补序列,其中该第一gRNA分子和该第二gRNA分子包含不同的靶向结构域。

[0292] 条款46.如条款32-45中任一项所述的组合物,其中该第一gRNA分子是SEQ ID NO:1、SEQ ID NO:3、SEQ ID NO:7、SEQ ID NO:8、SEQ ID NO:9、SEQ ID NO:10、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14或SEQ ID NO:15,并且该第二gRNA分子是SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:5、SEQ ID NO:6、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18或SEQ ID NO:19。

[0293] 条款47.如条款32-46中任一项所述的组合物,其中该第一gRNA分子选自下组,该组由以下组成:SEQ ID NO:1、SEQ ID NO:11、SEQ ID NO:12、SEQ ID NO:13、SEQ ID NO:14和SEQ ID NO:15,并且该第二gRNA分子选自下组,该组由以下组成:SEQ ID NO:2、SEQ ID NO:4、SEQ ID NO:16、SEQ ID NO:17、SEQ ID NO:18和SEQ ID NO:19。

[0294] 条款48.如条款32-47中任一项所述的组合物,其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:1中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:2中列出的核苷酸序列;(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(x)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID

NO:15中列出的核苷酸序列;以及(xi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列。

[0295] 条款49.如条款32-48中任一项所述的组合物,其中该第一载体包含SEQ ID NO:39中列出的核苷酸序列,并且该第二载体包含SEQ ID NO:40中列出的核苷酸序列。

[0296] 条款50.如条款32-49中任一项所述的组合物,用于在药物中使用。

[0297] 条款51.如条款32-50中任一项所述的组合物,用于在治疗杜氏肌营养不良中使用。

[0298] 条款52.一种细胞,包含如条款32-51中任一项所述的组合物。

[0299] 条款53.一种修正细胞中突变型肌营养不良蛋白基因的方法,包括向该细胞给予:(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第一Cas9分子的多核苷酸序列,以及(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第二Cas9分子的多核苷酸序列,其中该第一gRNA和第二gRNA分子各自具有长度为19至24个核苷酸的靶向结构域,并且其中该载体被配置为分别形成在人肌营养不良蛋白基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段。

[0300] 条款54.如条款53所述的方法,其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:1中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:2中列出的核苷酸序列;(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列:(x)包

含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列;以及(xi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列。

[0301] 条款55.如条款53或54所述的方法,其中该突变型肌营养不良蛋白基因包含提前终止密码子、被破坏的阅读框、异常剪接受体位点或异常剪接供体位点。

[0302] 条款56.如条款53-55中任一项所述的方法,其中该突变型肌营养不良蛋白基因包含导致提前终止密码子和截短的基因产物的移码突变。

[0303] 条款57.如条款53-55中任一项所述的方法,其中该突变型肌营养不良蛋白基因包含破坏阅读框的一个或多个外显子的缺失。

[0304] 条款58.如条款53-57中任一项所述的方法,其中该突变型肌营养不良蛋白基因的修正包括提前终止密码子的缺失、被破坏的阅读框的修正或通过破坏剪接受体位点或破坏剪接供体序列对剪接进行的调节。

[0305] 条款59.如条款53-58中任一项所述的方法,其中该突变型肌营养不良蛋白基因的修正包括外显子51的缺失。

[0306] 条款60.如条款53-59中任一项所述的方法,其中该突变型肌营养不良蛋白基因的修正包括同源定向修复。

[0307] 条款61.如条款60所述的方法,进一步包括向该细胞给予供体DNA。

[0308] 条款62.如条款53-61中任一项所述的方法,其中该突变型肌营养不良蛋白基因的修正包括核酸酶介导的非同源末端连接。

[0309] 条款63.如条款53-62中任一项所述的方法,其中该细胞是成肌细胞。

[0310] 条款64.如条款53-63中任一项所述的方法,其中该细胞来自患有杜氏肌营养不良的受试者。

[0311] 条款65.如条款53-64中任一项所述的方法,其中该细胞是来自患有杜氏肌营养不良的人类受试者的成肌细胞。

[0312] 条款66.如条款53-65中任一项所述的方法,其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:1中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:2中列出的核苷酸序列;(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;以及(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列。

[0313] 条款67.一种治疗具有突变型肌营养不良蛋白基因的有需要的受试者的方法,该方法包括向该受试者给予:(a)第一载体,其包含编码第一指导RNA(gRNA)分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻近基序(PAM)的第一Cas9分子的多核苷酸序列,以及(b)第二载体,其包含编码第二gRNA分子的多核苷酸序列和编码识别NNGRRT(SEQ ID NO:24)或NNGRRV(SEQ ID NO:25)的原型间隔子邻

近基序(PAM)的第二Cas9分子的多核苷酸序列,其中该第一gRNA和第二gRNA分子各自具有长度为19至24个核苷酸的靶向结构域,并且其中该载体被配置为分别形成在人肌营养不良蛋白基因的外显子51侧翼的第一和第二内含子中的第一和第二双链断裂,从而缺失包含外显子51的肌营养不良蛋白基因区段。

[0314] 条款68.如条款67所述的方法,其中该第一gRNA分子和该第二gRNA分子选自下组,该组由以下组成:(i)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:1中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:2中列出的核苷酸序列;(ii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(iii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(iv)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(v)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(vi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列;(vii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列;(viii)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:4中列出的核苷酸序列;(ix)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:19中列出的核苷酸序列:(x)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:14中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:15中列出的核苷酸序列;以及(xi)包含如下靶向结构域的第一gRNA分子,该靶向结构域包含SEQ ID NO:11中列出的核苷酸序列,以及包含如下靶向结构域的第二gRNA分子,该靶向结构域包含SEQ ID NO:18中列出的核苷酸序列。

[0315] 条款69.如条款68所述的方法,其中该受试者患有杜氏肌营养不良。

[0316] 条款70.如条款67-69中任一项所述的方法,其中将该第一载体和第二载体给予至该受试者的肌肉。

[0317] 条款71.如条款70所述的方法,其中该肌肉是骨骼肌或心肌。

[0318] 条款72.如条款71所述的方法,其中该骨骼肌是胫骨前肌。

[0319] 条款73.如条款67-72中任一项所述的方法,其中将该第一载体和第二载体肌内、静脉内或其组合给予该受试者。

[0320] 条款74.一种产生具有外显子52缺失(Δ52)的人肌营养不良蛋白基因(hDMD)的转基因啮齿动物胚胎的方法,该方法包括向啮齿动物胚胎给予如条款1所述的gRNA、如条款2-

9中任一项所述的DNA靶向系统、如条款10所述的分离的多核苷酸、如条款11-22中任一项所述的载体、如条款30或31所述的经修饰腺相关病毒载体或如条款32-51中任一项所述的组合物,从而缺失人肌营养不良蛋白基因的外显子52,并选择具有人肌营养不良蛋白基因外显子52缺失的转基因啮齿动物胚胎,其中该啮齿动物胚胎包含正常的人肌营养不良蛋白基因。

[0321] 条款75.如条款74所述的方法,其中该啮齿动物胚胎是小鼠胚胎。

[0322] 条款76.如条款74或75所述的方法,其中该转基因啮齿动物胚胎是杂合hDMD或杂合hDMD-Δ52。

[0323] 条款77.如条款74-76中任一项所述的方法,其中将包含含SEQ ID NO:41中列出的核苷酸序列的靶向结构域的第一gRNA分子和包含含SEQ ID NO:42中列出的核苷酸序列的靶向结构域的第二gRNA分子给予至该啮齿动物胚胎以缺失人肌营养不良蛋白基因的外显子52。

[0324] 条款78.如条款74-77中任一项所述的方法,进一步包括向该啮齿动物胚胎给予包含SEQ ID NO:27中列出的氨基酸序列的Cas蛋白。

[0325] 条款79.一种通过如条款74-78中任一项所述的方法产生的转基因啮齿动物胚胎。

[0326] 条款80.一种由如条款79所述的转基因啮齿动物胚胎产生的转基因啮齿动物。

附录

含JCR179(SEQ ID NO:37)的pDO240(gRNA为粗体)

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttttaaccaataggccgaaatcggca
aaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacg
tggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagtttttttg
gggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcga
acgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgc
gtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattcaggctgcgcaactgttgggaa
gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacg
ccagggttttcccagtcacgacgttgtaaaacgacggccagtgagcgcgcgtaatacgactcactatagggcgaattgggta
ccaagcttgcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgta
aacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatg
gactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc**AACAC**
ACAGCTGGGTTATCAGAGgttttagtactctggaaacagaatctactaaaacaaggcaaaatgccgtgttt
atctcgtcaacttgttggcgagatttttttgcggccgcccgcggtggagctccagcttttgttccctttagtgagggttaattgcg
cgcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagca
taaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaa
acctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctc
gctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccaca
gaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttg
ctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgaca
ggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacct
gtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctcc
aagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccgg
taagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagt
tcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaa
aaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgca
gaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattt
tggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagta
aacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgact
ccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctc
accggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctc
catccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacag
gcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccat
gttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatg
gcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgaga
atagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtg
ctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgt
gcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa
gggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatg
agcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccac

含JCR183(SEQ ID NO:38)的pDO240(gRNA为粗体)

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttttaaccaataggccgaaatcggca
aaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacg
tggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagtttttttg
gggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcga
acgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgc
gtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattcaggctgcgcaactgttgggaa
gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacg
ccagggttttcccagtcacgacgttgtaaaacgacggccagtgagcgcgcgtaatacgactcactatagggcgaattgggta
ccaagcttgcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgta
aacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatg
gactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc**AGAAC**
TGGTGGGAAATGGTCTAGgttttagtactctggaaacagaatctactaaaacaaggcaaaatgccgtgttt
atctcgtcaacttgttggcgagatttttttgcggccgcccgcggtggagctccagcttttgttccctttagtgagggttaattgcg
cgcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagca
taaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaa
acctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctc
gctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccaca
gaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttg
ctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgaca
ggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacct
gtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctcc
aagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccgg
taagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagt
tcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaa
aaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgca
gaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattt
tggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagta
aacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgact
ccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctc
accggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctc

catccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacag
gcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccat
gttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatg
gcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgaga
atagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtg
ctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgt
gcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaa
gggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatg
agcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccac

含JCR179PT366AAV 179(SEQ ID NO:39)的PT366-用于体内起作用的AAV质粒(gRNA为粗体;SaCas9为大写;NLS为小写、粗体且加下划线)

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgcccggcctcagtga
gcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcctctagactcgaggcgttgacatt
gattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggt
aaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaata
gggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagta
cgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggc
agtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactc
acggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcg
taacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaact
accggtgccaccatggccccaaagaagaagcggaaggtcggtatccacggagtcccagcagccAAGCGGAAC
TACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATC
GACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGA
GGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGA
AGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCT
GTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACC
CCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGA
GTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGT
GAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAG
ATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCA
GCTGGAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGA
TTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCA
GAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCT
GCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCT

TCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGC
ACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGA
CCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACG
AGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTG
TTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCT
CGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGC
CCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCC
GGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATC
CTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAATCT
GAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGG
GCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGG
ACGAGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTG
AAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAAAGAGATCCCCAC
CACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCAT
CCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCA
ACGACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAG
AAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGGA
TCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATC
GAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCT
GGAAGCCATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGT
GGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAACAA
GGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCA
TTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAA
GAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCA
AGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAG
AAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGG
CCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGA
AAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGA
AGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGC
CCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGA
CAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCC
GAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCAT
CACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACA
GCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTG
TACTCCACCCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCT
GAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACA
AGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAG

AAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTA
CAAGTACTACGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGG
ACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTGAAC
GCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGT
GAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGT
GTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAACT
ACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAG
ATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATC
AAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCT
GAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGG
AAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCC
AAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTA
TGAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGC**aaaaggccgg**
cggccacgaaaaaggccggccaggcaaaaaagaaaaagggatcctacccatacgatgttccagattacgcttacccat
acgatgttccagattacgcttacccatacgatgttccagattacgcttaagaattcctagagctcgctgatcagcctcgactgtg
ccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttccta
ataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggg
gaggattgggaagagaatagcaggcatgctggggaggtaccgagggcctatttcccatgattccttcatatttgcatatacga
tacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaat
aatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggc
tttatatatcttgtggaaaggacgaaacaccg**AACACACAGCTGGGTTATCAGAG**gttttagtactct
ggaaacagaatctactaaaacaaggcaaaatgccgtgtttatctcgtcaacttgttggcgagatttttgcggccgcaggaacc
cctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcc
cgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctgatgcggtattttctcctta
cgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcggg
tgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgcc
acgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgacccc
aaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgt
tctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatt
tcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatggtgca
ctctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacggg
cttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcac
cgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcag
gtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaat
aaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcgg
cattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggtta
catcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagtt

ctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgactt
ggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatg
agtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatggggg
atcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgt
agcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatgg
aggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgag
cgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagt
caggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaag
tttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatcctttttgataatctcatgaccaa
aatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatcctttttttctgcg
cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttcc
gaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaact
ctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggtt
ggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggag
cgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcgg
acaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatcttta
tagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacg
ccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgt

含JCR183(SEQ ID NO:40)的PT366-用于体内起作用(gRNA为粗体;SaCas9为大写;NLS为小写、粗体且加下划线)

cctgcaggcagctgcgcgctcgctcgctcactgaggccgcccgggcgtcgggcgacctttggtcgcccggcctcagtga
gcgagcgagcgcgcagagagggagtggccaactccatcactaggggttcctgcggcctctagactcgaggcgttgacatt
gattattgactagttattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacggt
aaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaata
gggactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagta
cgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttatgggactttcctacttggc
agtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcagtacatcaatgggcgtggatagcggtttgactc
acggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcg
taacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagagctctctggctaact
accggtgccaccatggccccaaagaagaagcggaaggtcggtatccacggagtcccagcagccAAGCGGAAC
TACATCCTGGGCCTGGACATCGGCATCACCAGCGTGGGCTACGGCATCATC
GACTACGAGACACGGGACGTGATCGATGCCGGCGTGCGGCTGTTCAAAGA
GGCCAACGTGGAAAACAACGAGGGCAGGCGGAGCAAGAGAGGCGCCAGA
AGGCTGAAGCGGCGGAGGCGGCATAGAATCCAGAGAGTGAAGAAGCTGCT
GTTCGACTACAACCTGCTGACCGACCACAGCGAGCTGAGCGGCATCAACC
CCTACGAGGCCAGAGTGAAGGGCCTGAGCCAGAAGCTGAGCGAGGAAGA

GTTCTCTGCCGCCCTGCTGCACCTGGCCAAGAGAAGAGGCGTGCACAACGT
GAACGAGGTGGAAGAGGACACCGGCAACGAGCTGTCCACCAAAGAGCAG
ATCAGCCGGAACAGCAAGGCCCTGGAAGAGAAATACGTGGCCGAACTGCA
GCTGGAACGGCTGAAGAAAGACGGCGAAGTGCGGGGCAGCATCAACAGA
TTCAAGACCAGCGACTACGTGAAAGAAGCCAAACAGCTGCTGAAGGTGCA
GAAGGCCTACCACCAGCTGGACCAGAGCTTCATCGACACCTACATCGACCT
GCTGGAAACCCGGCGGACCTACTATGAGGGACCTGGCGAGGGCAGCCCCT
TCGGCTGGAAGGACATCAAAGAATGGTACGAGATGCTGATGGGCCACTGC
ACCTACTTCCCCGAGGAACTGCGGAGCGTGAAGTACGCCTACAACGCCGA
CCTGTACAACGCCCTGAACGACCTGAACAATCTCGTGATCACCAGGGACG
AGAACGAGAAGCTGGAATATTACGAGAAGTTCCAGATCATCGAGAACGTG
TTCAAGCAGAAGAAGAAGCCCACCCTGAAGCAGATCGCCAAAGAAATCCT
CGTGAACGAAGAGGATATTAAGGGCTACAGAGTGACCAGCACCGGCAAGC
CCGAGTTCACCAACCTGAAGGTGTACCACGACATCAAGGACATTACCGCCC
GGAAAGAGATTATTGAGAACGCCGAGCTGCTGGATCAGATTGCCAAGATC
CTGACCATCTACCAGAGCAGCGAGGACATCCAGGAAGAACTGACCAATCT
GAACTCCGAGCTGACCCAGGAAGAGATCGAGCAGATCTCTAATCTGAAGG
GCTATACCGGCACCCACAACCTGAGCCTGAAGGCCATCAACCTGATCCTGG
ACGAGCTGTGGCACACCAACGACAACCAGATCGCTATCTTCAACCGGCTG
AAGCTGGTGCCCAAGAAGGTGGACCTGTCCCAGCAGAAAGAGATCCCCAC
CACCCTGGTGGACGACTTCATCCTGAGCCCCGTCGTGAAGAGAAGCTTCAT
CCAGAGCATCAAAGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCA
ACGACATCATTATCGAGCTGGCCCGCGAGAAGAACTCCAAGGACGCCCAG
AAAATGATCAACGAGATGCAGAAGCGGAACCGGCAGACCAACGAGCGGA
TCGAGGAAATCATCCGGACCACCGGCAAAGAGAACGCCAAGTACCTGATC
GAGAAGATCAAGCTGCACGACATGCAGGAAGGCAAGTGCCTGTACAGCCT
GGAAGCCATCCCTCTGGAAGATCTGCTGAACAACCCCTTCAACTATGAGGT
GGACCACATCATCCCCAGAAGCGTGTCCTTCGACAACAGCTTCAACAACAA
GGTGCTCGTGAAGCAGGAAGAAAACAGCAAGAAGGGCAACCGGACCCCA
TTCCAGTACCTGAGCAGCAGCGACAGCAAGATCAGCTACGAAACCTTCAA
GAAGCACATCCTGAATCTGGCCAAGGGCAAGGGCAGAATCAGCAAGACCA
AGAAAGAGTATCTGCTGGAAGAACGGGACATCAACAGGTTCTCCGTGCAG
AAAGACTTCATCAACCGGAACCTGGTGGATACCAGATACGCCACCAGAGG
CCTGATGAACCTGCTGCGGAGCTACTTCAGAGTGAACAACCTGGACGTGA
AAGTGAAGTCCATCAATGGCGGCTTCACCAGCTTTCTGCGGCGGAAGTGGA
AGTTTAAGAAAGAGCGGAACAAGGGGTACAAGCACCACGCCGAGGACGC
CCTGATCATTGCCAACGCCGATTTCATCTTCAAAGAGTGGAAGAAACTGGA

CAAGGCCAAAAAAGTGATGGAAAACCAGATGTTCGAGGAAAAGCAGGCC
GAGAGCATGCCCGAGATCGAAACCGAGCAGGAGTACAAAGAGATCTTCAT
CACCCCCCACCAGATCAAGCACATTAAGGACTTCAAGGACTACAAGTACA
GCCACCGGGTGGACAAGAAGCCTAATAGAGAGCTGATTAACGACACCCTG
TACTCCACCCGGAAGGACGACAAGGGCAACACCCTGATCGTGAACAATCT
GAACGGCCTGTACGACAAGGACAATGACAAGCTGAAAAAGCTGATCAACA
AGAGCCCCGAAAAGCTGCTGATGTACCACCACGACCCCCAGACCTACCAG
AAACTGAAGCTGATTATGGAACAGTACGGCGACGAGAAGAATCCCCTGTA
CAAGTACTACGAGGAAACCGGGAACTACCTGACCAAGTACTCCAAAAAGG
ACAACGGCCCCGTGATCAAGAAGATTAAGTATTACGGCAACAAACTGAAC
GCCCATCTGGACATCACCGACGACTACCCCAACAGCAGAAACAAGGTCGT
GAAGCTGTCCCTGAAGCCCTACAGATTCGACGTGTACCTGGACAATGGCGT
GTACAAGTTCGTGACCGTGAAGAATCTGGATGTGATCAAAAAAGAAAACT
ACTACGAAGTGAATAGCAAGTGCTATGAGGAAGCTAAGAAGCTGAAGAAG
ATCAGCAACCAGGCCGAGTTTATCGCCTCCTTCTACAACAACGATCTGATC
AAGATCAACGGCGAGCTGTATAGAGTGATCGGCGTGAACAACGACCTGCT
GAACCGGATCGAAGTGAACATGATCGACATCACCTACCGCGAGTACCTGG
AAAACATGAACGACAAGAGGCCCCCCAGGATCATTAAGACAATCGCCTCC
AAGACCCAGAGCATTAAGAAGTACAGCACAGACATTCTGGGCAACCTGTA
TGAAGTGAAATCTAAGAAGCACCCTCAGATCATCAAAAAGGGC**aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag**ggatcctacccatacgatgttccagattacgcttacccat
acgatgttccagattacgcttacccatacgatgttccagattacgcttaagaattcctagagctcgctgatcagcctcgactgtg
ccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttccta
ataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggg
gaggattgggaagagaatagcaggcatgctggggaggtaccgagggcctatttcccatgattccttcatatttgcatatacga
tacaaggctgttagagagataattggaattaatttgactgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaat
aatttcttgggtagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggc
tttatatatcttgtggaaaggacgaaacaccg**AGAACTGGTGGGAAATGGTCTAG**gttttagtactct
ggaaacagaatctactaaaacaaggcaaaatgccgtgtttatctcgtcaacttgttggcgagatttttgcggccgcaggaacc
cctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaaaggtcgcccgacgcc
cgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctgatgcggtattttctcctta
cgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcggcgcattaagcgcggcggg
tgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgctttcttcccttcctttctcgcc
acgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgatttagtgctttacggcacctcgacccc
aaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacggtttttcgccctttgacgttggagtccacgt
tctttaatagtggactcttgttccaaactggaacaacactcaaccctatctcgggctattcttttgatttataagggattttgccgatt
tcggcctattggttaaaaaatgagctgatttaacaaaaatttaacgcgaattttaacaaaatattaacgtttacaattttatggtgca

ctctcagtacaatctgctctgatgccgcatagttaagccagccccgacacccgccaacacccgctgacgcgccctgacggg
cttgtctgctcccggcatccgcttacagacaagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcac
cgaaacgcgcgagacgaaagggcctcgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcag
gtggcacttttcggggaaatgtgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaat
aaccctgataaatgcttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattcccttttttgcgg
cattttgccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggtta
catcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagtt
ctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatgactt
ggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccataaccatg
agtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgctttttgcacaacatggggg
atcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacaccacgatgcctgt
agcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaattaatagactggatgg
aggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataaatctggagccggtgag
cgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagttatctacacgacggggagt
caggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcattggtaactgtcagaccaag
tttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggtgaagatccttttgataatctcatgaccaa
aatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatccttttttctgcg
cgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagagctaccaactctttttcc
gaaggtaactggcttcagcagagcgcagataccaaatactgtccttctagtgtagccgtagttaggccaccacttcaagaact
ctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggtt
ggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacggggggttcgtgcacacagcccagcttggag
cgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcgg
acaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccagggggaaacgcctggtatcttta
tagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacg
ccagcaacgcggcctttttacggttcctggccttttgctggccttttgctcacatgt

pDO242(用于体外起作用的所有JCR89/91项目和JCR157/160项目的SaCas9；SaCas9为大写)(SEQ ID NO:83)

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttttaaccaataggccgaaatcggca
aaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacg
tggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagttttttg
gggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcga
acgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgc
gtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattcaggctgcgcaactgttgggaa
gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacg
ccagggttttcccagtcacgacgttgtaaaacgacggccagtgagcgcgcgtaatacgactcactatagggcgaattgggta
cCtttaattctagtactatgcaTgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagcc

catatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtc
aataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgccca
cttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgc
ccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggca
gtacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttg
gcaccaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtg
ggaggtctatataagcagagctctctggctaactaccggtgccaccATGAAAAGGAACTACATTCTGG
GGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTATGAA
ACAAGGGACGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGT
GGAAAACAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGAAA
CGACGGAGAAGGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTCGATTA
CAACCTGCTGACCGACCATTCTGAGCTGAGTGGAATTAATCCTTATGAAGC
CAGGGTGAAAGGCCTGAGTCAGAAGCTGTCAGAGGAAGAGTTTTCCGCAG
CTCTGCTGCACCTGGCTAAGCGCCGAGGAGTGCATAACGTCAATGAGGTG
GAAGAGGACACCGGCAACGAGCTGTCTACAAAGGAACAGATCTCACGCAA
TAGCAAAGCTCTGGAAGAGAAGTATGTCGCAGAGCTGCAGCTGGAACGGC
TGAAGAAAGATGGCGAGGTGAGAGGGTCAATTAATAGGTTCAAGACAAGC
GACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAGTGCAGAAGGCTTACCA
CCAGCTGGATCAGAGCTTCATCGATACTTATATCGACCTGCTGGAGACTCG
GAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCTTCGGATGGAAAG
ACATCAAGGAATGGTACGAGATGCTGATGGGACATTGCACCTATTTTCCAG
AAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTGTACAACGCC
CTGAATGACCTGAACAACCTGGTCATCACCAGGGATGAAAACGAGAAACT
GGAATACTATGAGAAGTTCCAGATCATCGAAAACGTGTTTAAGCAGAAGA
AAAAGCCTACACTGAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAG
GACATCAAGGGCTACCGGGTGACAAGCACTGGAAAACCAGAGTTCACCAA
TCTGAAAGTGTATCACGATATTAAGGACATCACAGCACGGAAAGAAATCA
TTGAGAACGCCGAACTGCTGGATCAGATTGCTAAGATCCTGACTATCTACC
AGAGCTCCGAGGACATCCAGGAAGAGCTGACTAACCTGAACAGCGAGCTG
ACCCAGGAAGAGATCGAACAGATTAGTAATCTGAAGGGGTACACCGGAAC
ACACAACCTGTCCCTGAAAGCTATCAATCTGATTCTGGATGAGCTGTGGCA
TACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAAGCTGGTCCCAAA
AAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAACCACACTGGTGGACG
ATTTCATTCTGTCACCCGTGGTCAAGCGGAGCTTCATCCAGAGCATCAAAG
TGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAATGATATCATTATCG
AGCTGGCTAGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGA
GATGCAGAAACGAAACCGGCAGACCAATGAACGCATTGAAGAGATTATCC

GAACTACCGGGAAAGAGAACGCAAAGTACCTGATTGAAAAAATCAAGCTG
CACGATATGCAGGAGGGAAAGTGTCTGTATTCTCTGGAGGCCATCCCCCTG
GAGGACCTGCTGAACAATCCATTCAACTACGAGGTCGATCATATTATCCCC
AGAAGCGTGTCCTTCGACAATTCCTTTAACAACAAGGTGCTGGTCAAGCAG
GAAGAGAACTCTAAAAAGGGCAATAGGACTCCTTTCCAGTACCTGTCTAGT
TCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCACATTCTGAATCTG
GCCAAAGGAAAGGGCCGCATCAGCAAGACCAAAAAGGAGTACCTGCTGG
AAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCGG
AATCTGGTGGACACAAGATACGCTACTCGCGGCCTGATGAATCTGCTGCGA
TCCTATTTCCGGGTGAACAATCTGGATGTGAAAGTCAAGTCCATCAACGGC
GGGTTCACATCTTTTCTGAGGCGCAAATGGAAGTTTAAAAAGGAGCGCAA
CAAAGGGTACAAGCACCATGCCGAAGATGCTCTGATTATCGCAAATGCCG
ACTTCATCTTTAAGGAGTGGAAAAAGCTGGACAAAGCCAAGAAAGTGATG
GAGAACCAGATGTTCGAAGAGAAGCAGGCCGAATCTATGCCCGAAATCGA
GACAGAACAGGAGTACAAGGAGATTTTCATCACTCCTCACCAGATCAAGC
ATATCAAGGATTTCAAGGACTACAAGTACTCTCACCGGGTGGATAAAAAG
CCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACAAGAAAAGACGA
TAAGGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAG
ATAATGACAAGCTGAAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTG
ATGTACCACCATGATCCTCAGACATATCAGAAACTGAAGCTGATTATGGAG
CAGTACGGCGACGAGAAGAACCCACTGTATAAGTACTATGAAGAGACTGG
GAACTACCTGACCAAGTATAGCAAAAAGGATAATGGCCCCGTGATCAAGA
AGATCAAGTACTATGGGAACAAGCTGAATGCCCATCTGGACATCACAGAC
GATTACCCTAACAGTCGCAACAAGGTGGTCAAGCTGTCACTGAAGCCATAC
AGATTCGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGACTGTCAAG
AATCTGGATGTCATCAAAAAGGAGAACTACTATGAAGTGAATAGCAAGTG
CTACGAAGAGGCTAAAAAGCTGAAAAAGATTAGCAACCAGGCAGAGTTCA
TCGCCTCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAACTGTATA
GGGTCATCGGGGTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATG
ATTGACATCACTTACCGAGAGTATCTGGAAAACATGAATGATAAGCGCCCC
CCTCGAATTATCAAAACAATTGCCTCTAAGACTCAGAGTATCAAAAAGTAC
TCAACCGACATTCTGGGAAACCTGTATGAGGTGAAGAGCAAAAAGCACCC
TCAGATTATCAAAAAGGGCagcggaggcaagcgtcctgctgctactaagaaagctggtcaagctaagaa
aaagaaaggatcctacccatacgatgttccagattacgcttaagaattcctagagctcgctgatcagcctcgactgtgccttct
agttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaa
tgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggagga
ttgggaagagaatagcaggcatgctggggaggtagcggccgcCCgcggtggagctccagcttttgttccctttagtgagg

gttaattgcgcgcttggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgag
ccggaagcataaagtgtaaagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttc
cagtcgggaaacctgtcgtgccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctct
tccgcttcctcgctcactgactcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatac
ggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaa
aggccgcgttgctggcgtttttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggc
gaaacccgacaggactataaagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgctt
accggatacctgtccgcctttctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtagg
tcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgag
tccaacccggtaagacacgacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcgg
tgctacagagttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagt
taccttcggaaaaagagttggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagca
gattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcac
gttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaag
tatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccata
gttgcctgactccccgtcgtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgag
acccacgctcaccggctccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaact
ttatccgcctccatccagtctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgc
cattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttac
atgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatca
ctcatggttatggcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaa
gtcattctgagaatagtgtatgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcaga
actttaaaagtgctcatcattggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgt
aacccactcgtgcacccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatg
ccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcaggg
ttattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtg
ccac

pJRH1(用于所有JCR179/183项目的SaCas9,SaCas9为大写;NLS为小写、粗体且加下划线)(SEQ ID NO:84)

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttttaaccaataggccgaaatcggca
aaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacg
tggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagtttttttg
gggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcga
acgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgc
gtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattcaggctgcgcaactgttgggaa
gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacg

ccagggttttcccagtcacgacgttgtaaaacgacggccagtgagcgcgcgtaatacgactcactatagggcgaattgggta
cctttaattctagtactatgcatgcgttgacattgattattgactagttattaatagtaatcaattacggggtcattagttcatagccc
atatatggagttccgcgttacataacttacggtaaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtca
ataatgacgtatgttcccatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacggtaaactgcccact
tggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcccgcctggcattatgcc
cagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctattaccatggtgatgcggttttggcag
tacatcaatgggcgtggatagcggtttgactcacggggatttccaagtctccaccccattgacgtcaatgggagtttgttttggc
accaaaatcaacgggactttccaaaatgtcgtaacaactccgccccattgacgcaaatgggcggtaggcgtgtacggtggg
aggtctatataagcagagctctctggctaactaccggtgccaccatggccccaaagaagaagcggaaggtcggtatccacg
gagtcccagcagccAAGCGGAACTACATCCTGGGCCTGGACATCGGCATCACCAG
CGTGGGCTACGGCATCATCGACTACGAGACACGGGACGTGATCGATGCCG
GCGTGCGGCTGTTCAAAGAGGCCAACGTGGAAAACAACGAGGGCAGGCGG
AGCAAGAGAGGCGCCAGAAGGCTGAAGCGGCGGAGGCGGCATAGAATCC
AGAGAGTGAAGAAGCTGCTGTTCGACTACAACCTGCTGACCGACCACAGC
GAGCTGAGCGGCATCAACCCCTACGAGGCCAGAGTGAAGGGCCTGAGCCA
GAAGCTGAGCGAGGAAGAGTTCTCTGCCGCCCTGCTGCACCTGGCCAAGA
GAAGAGGCGTGCACAACGTGAACGAGGTGGAAGAGGACACCGGCAACGA
GCTGTCCACCAAAGAGCAGATCAGCCGGAACAGCAAGGCCCTGGAAGAGA
AATACGTGGCCGAACTGCAGCTGGAACGGCTGAAGAAAGACGGCGAAGTG
CGGGGCAGCATCAACAGATTCAAGACCAGCGACTACGTGAAAGAAGCCAA
ACAGCTGCTGAAGGTGCAGAAGGCCTACCACCAGCTGGACCAGAGCTTCA
TCGACACCTACATCGACCTGCTGGAAACCCGGCGGACCTACTATGAGGGA
CCTGGCGAGGGCAGCCCCTTCGGCTGGAAGGACATCAAAGAATGGTACGA
GATGCTGATGGGCCACTGCACCTACTTCCCCGAGGAACTGCGGAGCGTGA
AGTACGCCTACAACGCCGACCTGTACAACGCCCTGAACGACCTGAACAAT
CTCGTGATCACCAGGGACGAGAACGAGAAGCTGGAATATTACGAGAAGTT
CCAGATCATCGAGAACGTGTTCAAGCAGAAGAAGAAGCCCACCCTGAAGC
AGATCGCCAAAGAAATCCTCGTGAACGAAGAGGATATTAAGGGCTACAGA
GTGACCAGCACCGGCAAGCCCGAGTTCACCAACCTGAAGGTGTACCACGA
CATCAAGGACATTACCGCCCGGAAAGAGATTATTGAGAACGCCGAGCTGC
TGGATCAGATTGCCAAGATCCTGACCATCTACCAGAGCAGCGAGGACATC
CAGGAAGAACTGACCAATCTGAACTCCGAGCTGACCCAGGAAGAGATCGA
GCAGATCTCTAATCTGAAGGGCTATACCGGCACCCACAACCTGAGCCTGAA
GGCCATCAACCTGATCCTGGACGAGCTGTGGCACACCAACGACAACCAGA
TCGCTATCTTCAACCGGCTGAAGCTGGTGCCCAAGAAGGTGGACCTGTCCC
AGCAGAAAGAGATCCCCACCACCCTGGTGGACGACTTCATCCTGAGCCCC
GTCGTGAAGAGAAGCTTCATCCAGAGCATCAAAGTGATCAACGCCATCAT

CAAGAAGTACGGCCTGCCCAACGACATCATTATCGAGCTGGCCCGCGAGA
AGAACTCCAAGGACGCCCAGAAAATGATCAACGAGATGCAGAAGCGGAA
CCGGCAGACCAACGAGCGGATCGAGGAAATCATCCGGACCACCGGCAAAG
AGAACGCCAAGTACCTGATCGAGAAGATCAAGCTGCACGACATGCAGGAA
GGCAAGTGCCTGTACAGCCTGGAAGCCATCCCTCTGGAAGATCTGCTGAAC
AACCCCTTCAACTATGAGGTGGACCACATCATCCCCAGAAGCGTGTCCTTC
GACAACAGCTTCAACAACAAGGTGCTCGTGAAGCAGGAAGAAAACAGCAA
GAAGGGCAACCGGACCCCATTCCAGTACCTGAGCAGCAGCGACAGCAAGA
TCAGCTACGAAACCTTCAAGAAGCACATCCTGAATCTGGCCAAGGGCAAG
GGCAGAATCAGCAAGACCAAGAAAGAGTATCTGCTGGAAGAACGGGACAT
CAACAGGTTCTCCGTGCAGAAAGACTTCATCAACCGGAACCTGGTGGATAC
CAGATACGCCACCAGAGGCCTGATGAACCTGCTGCGGAGCTACTTCAGAG
TGAACAACCTGGACGTGAAAGTGAAGTCCATCAATGGCGGCTTCACCAGC
TTTCTGCGGCGGAAGTGGAAGTTTAAGAAAGAGCGGAACAAGGGGTACAA
GCACCACGCCGAGGACGCCCTGATCATTGCCAACGCCGATTTCATCTTCAA
AGAGTGGAAGAAACTGGACAAGGCCAAAAAAGTGATGGAAAACCAGATG
TTCGAGGAAAAGCAGGCCGAGAGCATGCCCGAGATCGAAACCGAGCAGG
AGTACAAAGAGATCTTCATCACCCCCCACCAGATCAAGCACATTAAGGACT
TCAAGGACTACAAGTACAGCCACCGGGTGGACAAGAAGCCTAATAGAGAG
CTGATTAACGACACCCTGTACTCCACCCGGAAGGACGACAAGGGCAACAC
CCTGATCGTGAACAATCTGAACGGCCTGTACGACAAGGACAATGACAAGC
TGAAAAAGCTGATCAACAAGAGCCCCGAAAAGCTGCTGATGTACCACCAC
GACCCCCAGACCTACCAGAAACTGAAGCTGATTATGGAACAGTACGGCGA
CGAGAAGAATCCCCTGTACAAGTACTACGAGGAAACCGGGAACTACCTGA
CCAAGTACTCCAAAAAGGACAACGGCCCCGTGATCAAGAAGATTAAGTAT
TACGGCAACAAACTGAACGCCCATCTGGACATCACCGACGACTACCCCAA
CAGCAGAAACAAGGTCGTGAAGCTGTCCCTGAAGCCCTACAGATTCGACG
TGTACCTGGACAATGGCGTGTACAAGTTCGTGACCGTGAAGAATCTGGATG
TGATCAAAAAAGAAAACTACTACGAAGTGAATAGCAAGTGCTATGAGGAA
GCTAAGAAGCTGAAGAAGATCAGCAACCAGGCCGAGTTTATCGCCTCCTTC
TACAACAACGATCTGATCAAGATCAACGGCGAGCTGTATAGAGTGATCGG
CGTGAACAACGACCTGCTGAACCGGATCGAAGTGAACATGATCGACATCA
CCTACCGCGAGTACCTGGAAAACATGAACGACAAGAGGCCCCCCAGGATC
ATTAAGACAATCGCCTCCAAGACCCAGAGCATTAAGAAGTACAGCACAGA
CATTCTGGGCAACCTGTATGAAGTGAAATCTAAGAAGCACCCTCAGATCAT
CAAAAAGGGC**aaaaggccggcggccacgaaaaaggccggccaggcaaaaaagaaaaag**ggatcctacc
catacgatgttccagattacgcttaagaattcctagagctcgctgatcagcctcgactgtgccttctagttgccagccatctgttg

tttgcccctcccccgtgccttccttgaccctggaaggtgccactcccactgtcctttcctaataaaatgaggaaattgcatcgca
ttgtctgagtaggtgtcattctattctggggggtggggtggggcaggacagcaagggggaggattgggaagagaatagca
ggcatgctggggaggtagcggccgcccgcggtggagctccagcttttgttccctttagtgagggttaattgcgcgcttggcg
taatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgta
aagcctggggtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgt
gccagctgcattaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactga
ctcgctgcgctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcagg
ggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgttt
ttccataggctccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactata
aagataccaggcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctt
tctcccttcgggaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctggg
ctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacac
gacttatcgccactggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagt
ggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagtt
ggtagctcttgatccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaa
aggatctcaagaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatg
agattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggt
ctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtc
gtgtagataactacgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggct
ccagatttatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccag
tctattaattgttgccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgt
ggtgtcacgctcgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgc
aaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagca
ctgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgt
atgcggcgaccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatca
ttggaaaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccca
actgatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaata
agggcgacacggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggat
acatatttgaatgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccac

PT366中的NLS序列(SEQ ID NO:85)

AAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAG

pDO203-针对SpCas9的gRNA克隆进的通用主链;JCR94和JCR99置于粗体位置以在细胞中测试,然后从中制备mRNA以制备hDMD-Δ52/mdx小鼠(SEQ ID NO:86)

ctaaattgtaagcgttaatattttgttaaaattcgcgttaaatttttgttaaatcagctcatttttttaaccaataggccgaaatcggca
aaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggaacaagagtccactattaaagaacg
tggactccaacgtcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagtttttttg
gggtcgaggtgccgtaaagcactaaatcggaaccctaaagggagcccccgatttagagcttgacggggaaagccggcga
acgtggcgagaaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagtgtagcggtcacgctgcgc
gtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcccattcgccattcaggctgcgcaactgttgggaa
gggcgatcggtgcgggcctcttcgctattacgccagctggcgaaagggggatgtgctgcaaggcgattaagttgggtaacg
ccagggttttcccagtcacgacgttgtaaaacgacggccagtgagcgcgcgtaatacgactcactatagggcgaattgggta
ccgagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgactgtaa
acacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttttaaaattatgttttaaaatgg
actatcatatgcttatcgtaacttgaaagtatttcgatttcttggctttatatatcttgtggaaaggacgaaacacc**GGGTCT**
TCGAGAAGACCTgttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggc
accgagtcggtgctttttttccgcggtggagctccagcttttgttccctttagtgagggttaattgcgcgcttggcgtaatcatggt
catagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcataaagtgtaaagcctggg
gtgcctaatgagtgagctaactcacattaattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgc
attaatgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgactcgctgcg
ctcggtcgttcggctgcggcgagcggtatcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgc
aggaaagaacatgtgagcaaaaggccagcaaaaggccaggaaccgtaaaaaggccgcgttgctggcgtttttccataggc
tccgcccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccag
gcgtttccccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgg
gaagcgtggcgctttctcatagctcacgctgtaggtatctcagttcggtgtaggtcgttcgctccaagctgggctgtgtgcacg
aaccccccgttcagcccgaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgcca
ctggcagcagccactggtaacaggattagcagagcgaggtatgtaggcggtgctacagagttcttgaagtggtggcctaac
tacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttga
tccggcaaacaaaccaccgctggtagcggtggtttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaag
aagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaa
aaggatcttcacctagatccttttaaattaaaaatgaagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttac
caatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataact
acgatacgggagggcttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatca
gcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgtt
gccgggaagctagagtaagtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgct
cgtcgtttggtatggcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcgg
ttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattct
cttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcgac
cgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgt
tcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaactgatcttca
gcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgaca
cggaaatgttgaatactcatactcttcctttttcaatattattgaagcatttatcagggttattgtctcatgagcggatacatatttga
atgtatttagaaaaataaacaaataggggttccgcgcacatttccccgaaaagtgccac

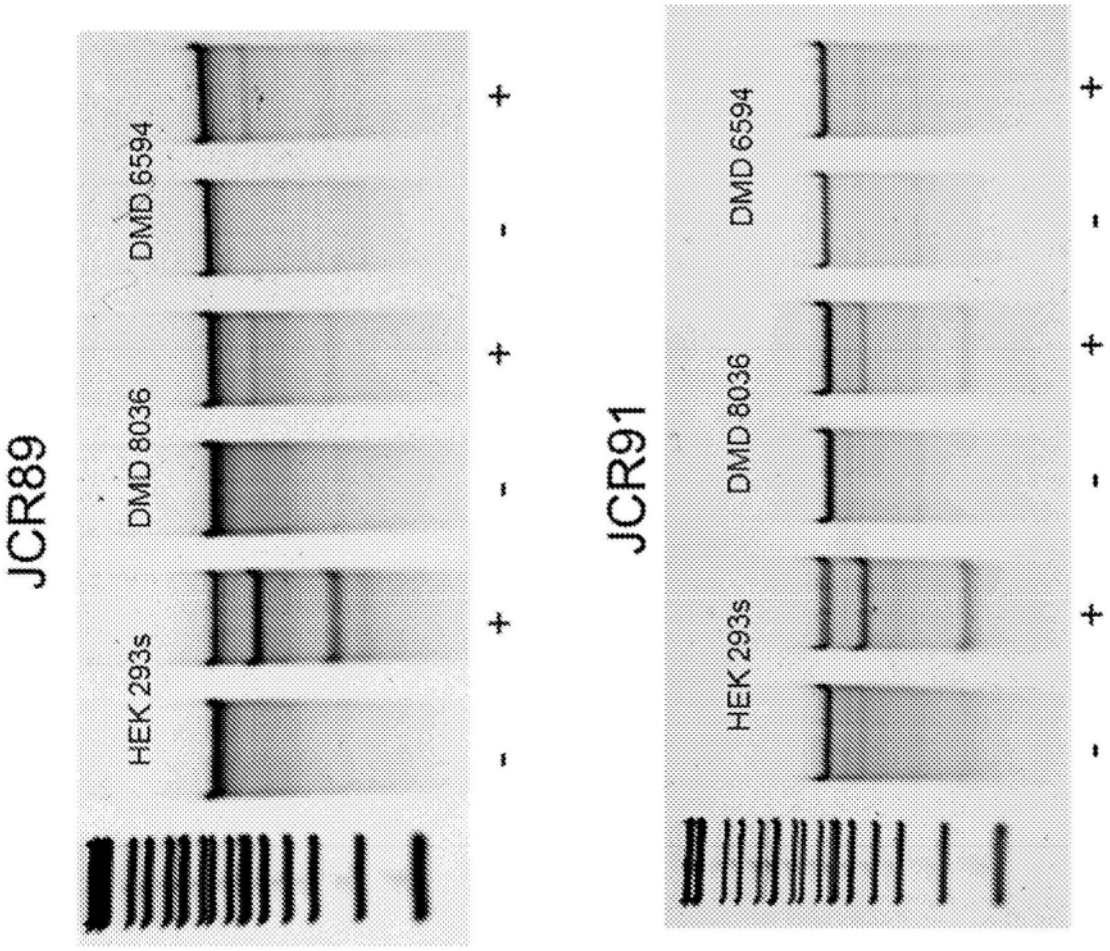

JCR89
HEK 293s
DMD 8036
DMD 6594
-
+
-
+
-
+
JCR91
HEK 293s
DMD 8036
DMD 6594
-
+
-
+
-
+

图1

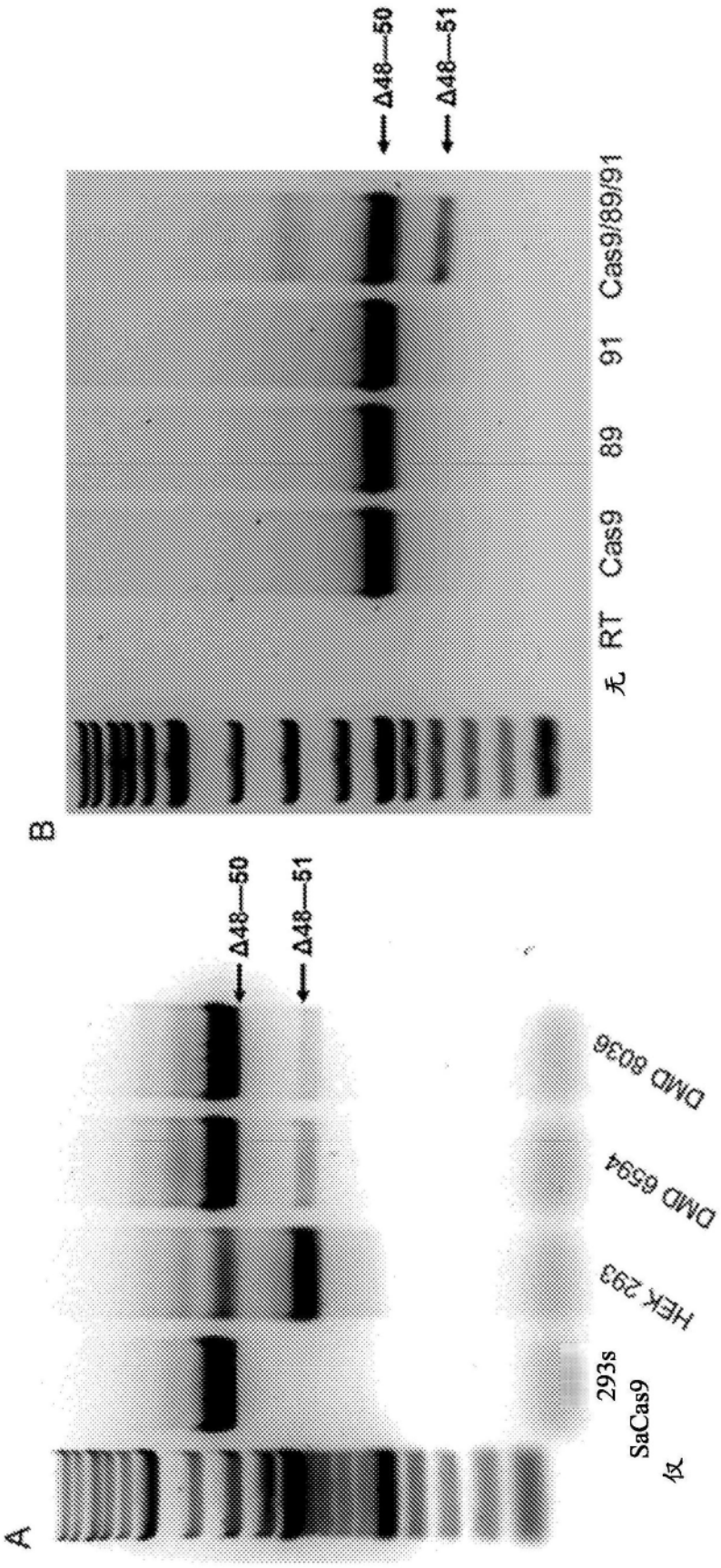
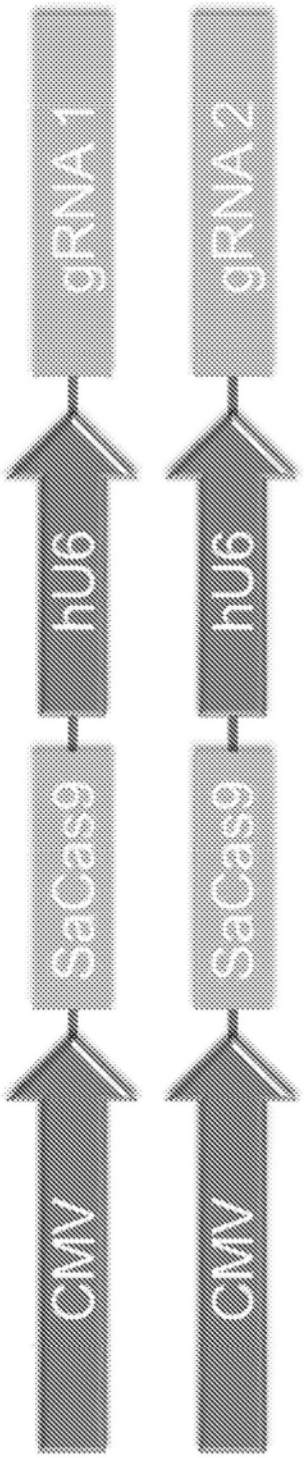

A
Δ48—50
Δ48—51
DMD 6036
DMD 6594
HEK 293
293s
SaCas9
仗
B
Δ48—50
Δ48—51
无 RT Cas9 89 91 Cas9/89/91

图2

递送设置



共同递送2种病毒w/JCR89, 91

图3

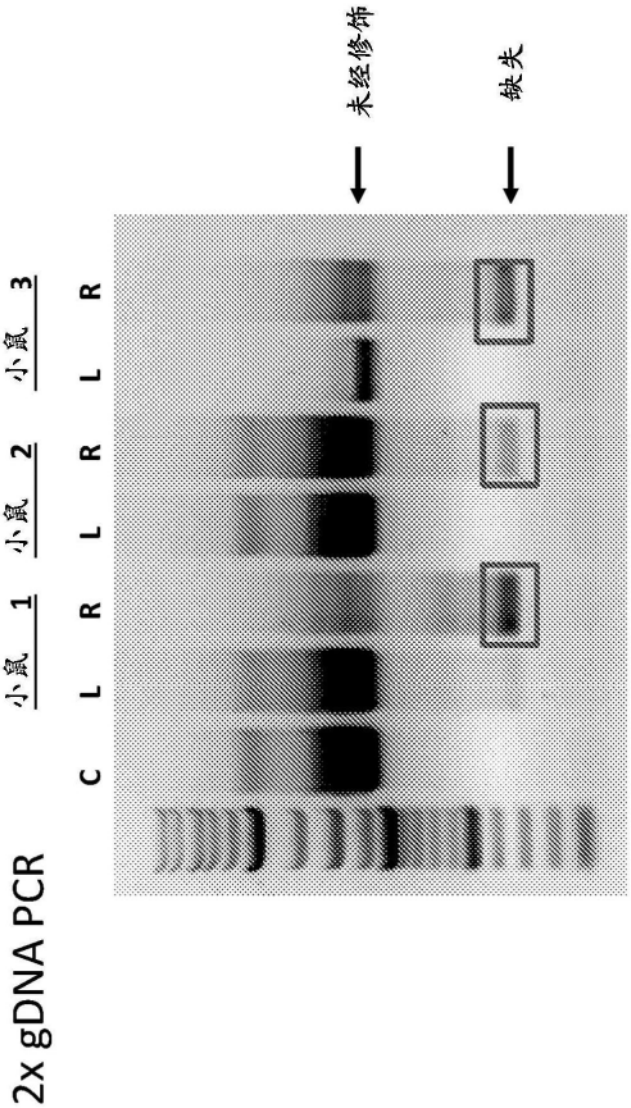
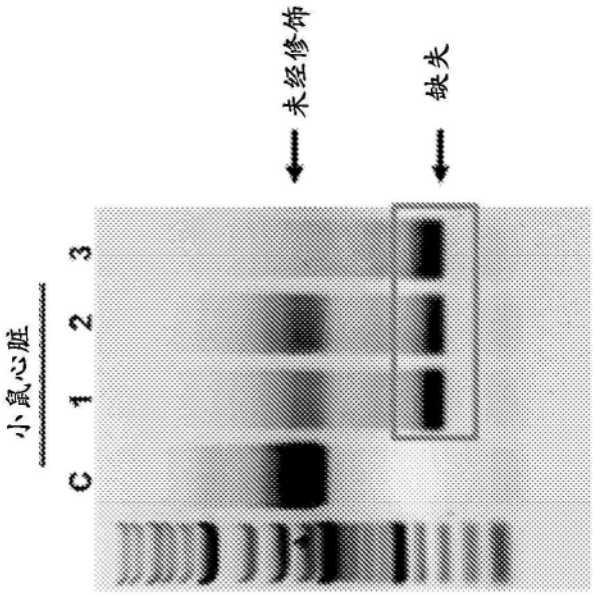


图4



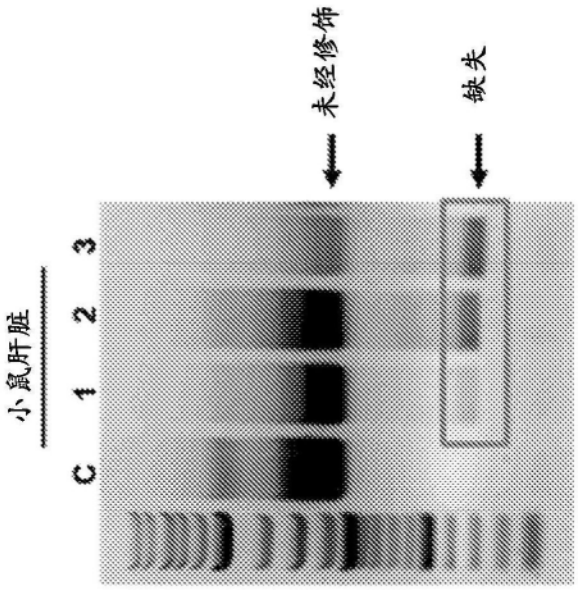


图5

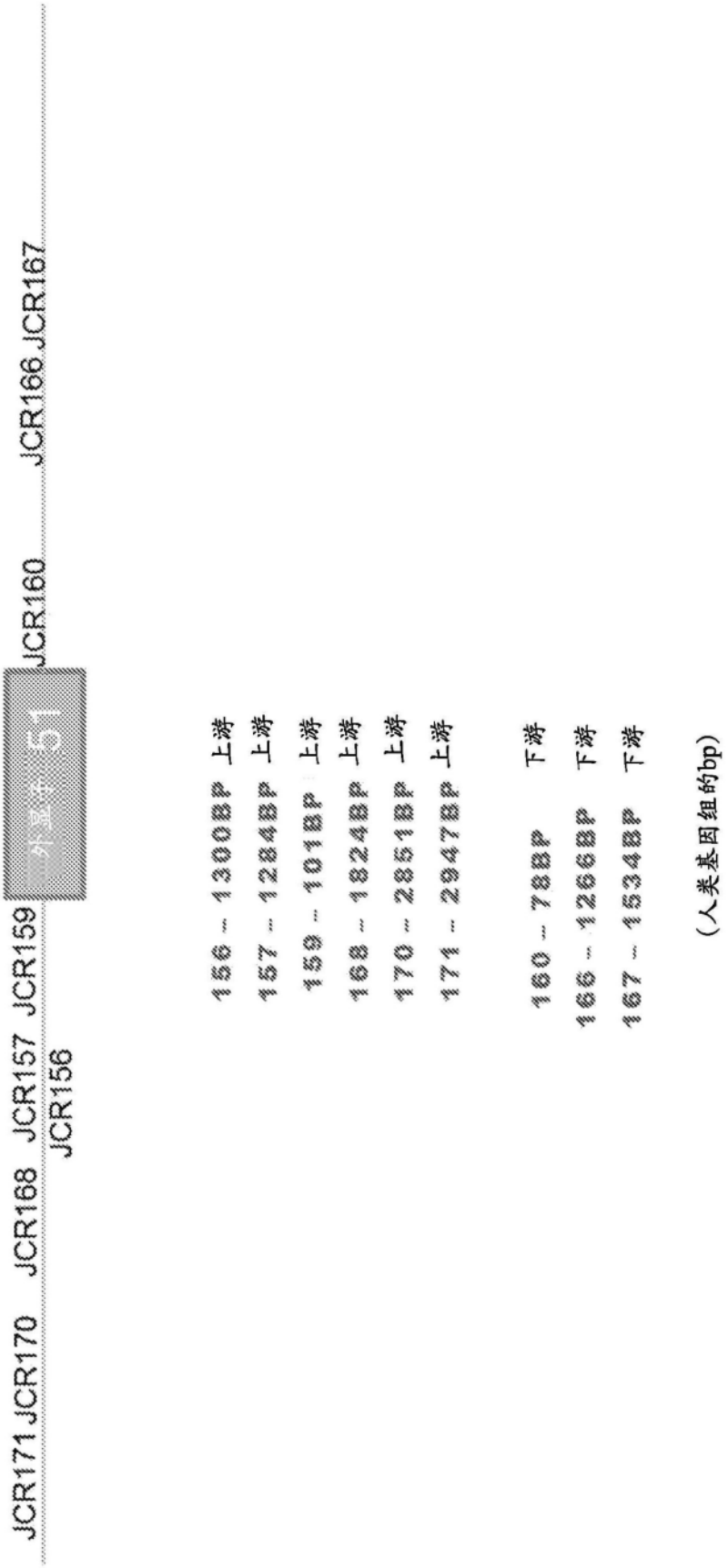

JCR171
JCR170
JCR168
JCR157
JCR159
JCR156
外显子 51
JCR160
JCR166
JCR167
156 – 1300BP 上游
157 – 1284BP 上游
159 – 101BP 上游
168 – 1824BP 上游
170 – 2851BP 上游
171 – 2947BP 上游
160 – 78BP 下游
166 – 1266BP 下游
167 – 1534BP 下游
(人类基因组的bp)

图6

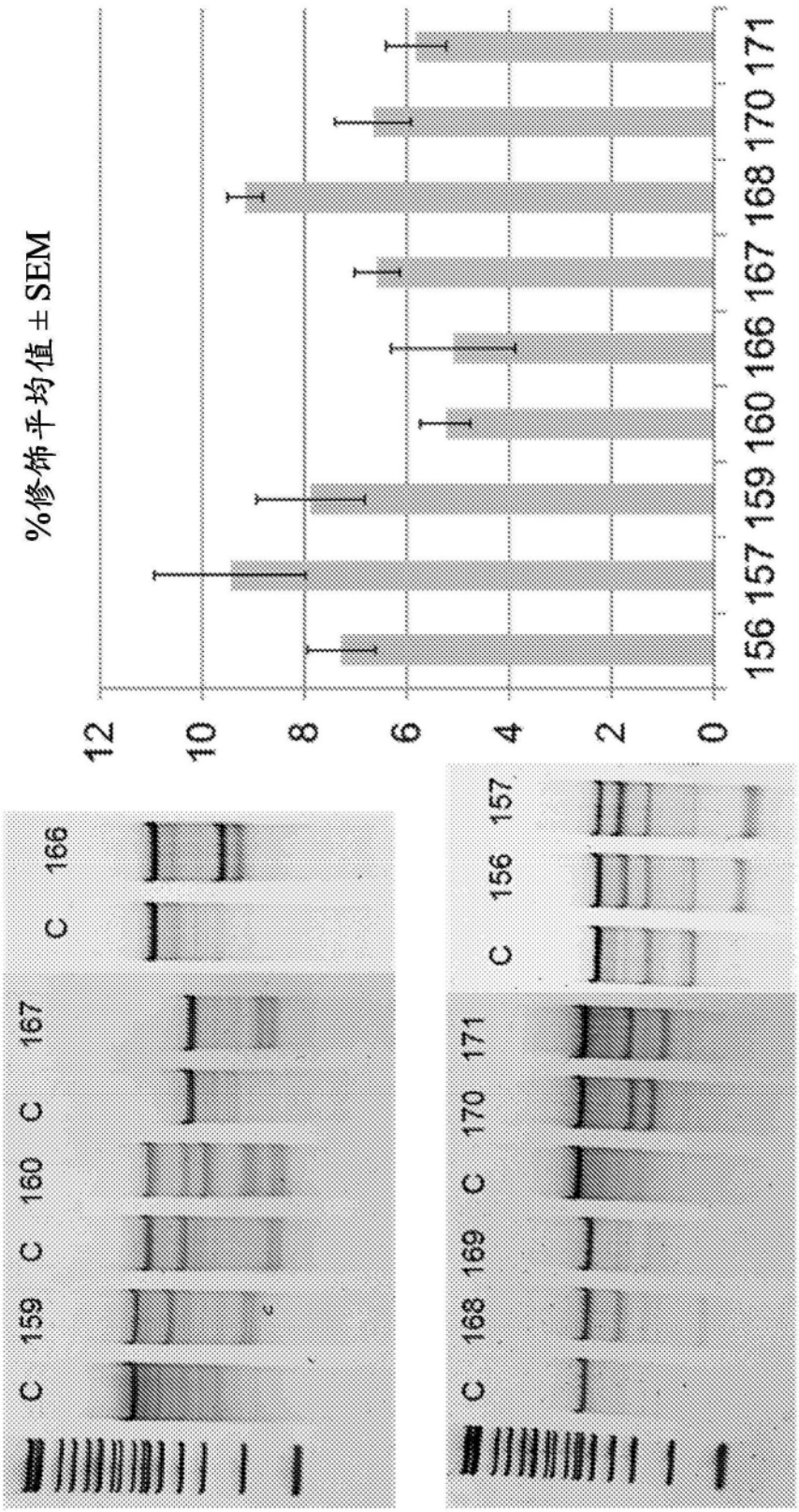

%修饰平均值 ± SEM
12
10
8
6
4
2
0
156 157 159 160 166 167 168 170 171
C 159 C 160 C 167
C 166
C 168 169 C 170 171
C 156 157

图7

gRNA	位置	距离 (nt)	%修饰 ± SEM	人脱靶（错配：基因组中的命中）	恒河猴脱靶
JCR156	上游	1300	7.28% ± 0.66	4 个错配 :23 3 个错配 3 2, 1 个错配 :0	4 个错配 :19 3 个错配 :7 2 个错配 :1 <2: 0
JCR157	上游	1284	9.45% ± 1.48	*在人类基因组中未找到原型间隔子* 4 个错配 :6 <4 个错配 :0	4 个错配 :5 <4 个错配 :0
JCR159	上游	101	7.88% ± 1.07	4 个错配 :16 <4 个错配 :0	4 个错配 :15 3 个错配 :1 <3 个错配 :0
JCR168	上游	1824	9.17% ± 0.35	4 个错配 :21 3 个错配 :1 <3 个错配 h: 0	4 个错配 :13 <4 个错配 :0
JCR170	上游	2851	6.66% ± 0.74	4 个错配 :1 3 个错配 :1 <3 个错配 :0	4 个错配 :3 <4 个错配 :0
JCR171	上游	2947	5.82% ± 0.59	4 个错配 :5 3 个错配 :1 <3 个错配 :0	4 个错配 :2 3 个错配 :1 2 个错配 1 <2 个错配 :0
JCR160	下游	78	5.25% ± 0.50	4 个错配 :2 <4 个错配 :0	4 个错配 :4 <4 个错配 :0
JCR166		1266	5.09% ± 1.21	4 个错配 :25 3 个错配 :5 <3: 0	4 个错配 :17 3 个错配 :5 <3 个错配 :0
JCR167	下游	1534	6.59% ± 0.44	4 个错配 :4 3 个错配 :2 <3 个错配 :0	4 个错配 10 <4 个错配 :0

图8

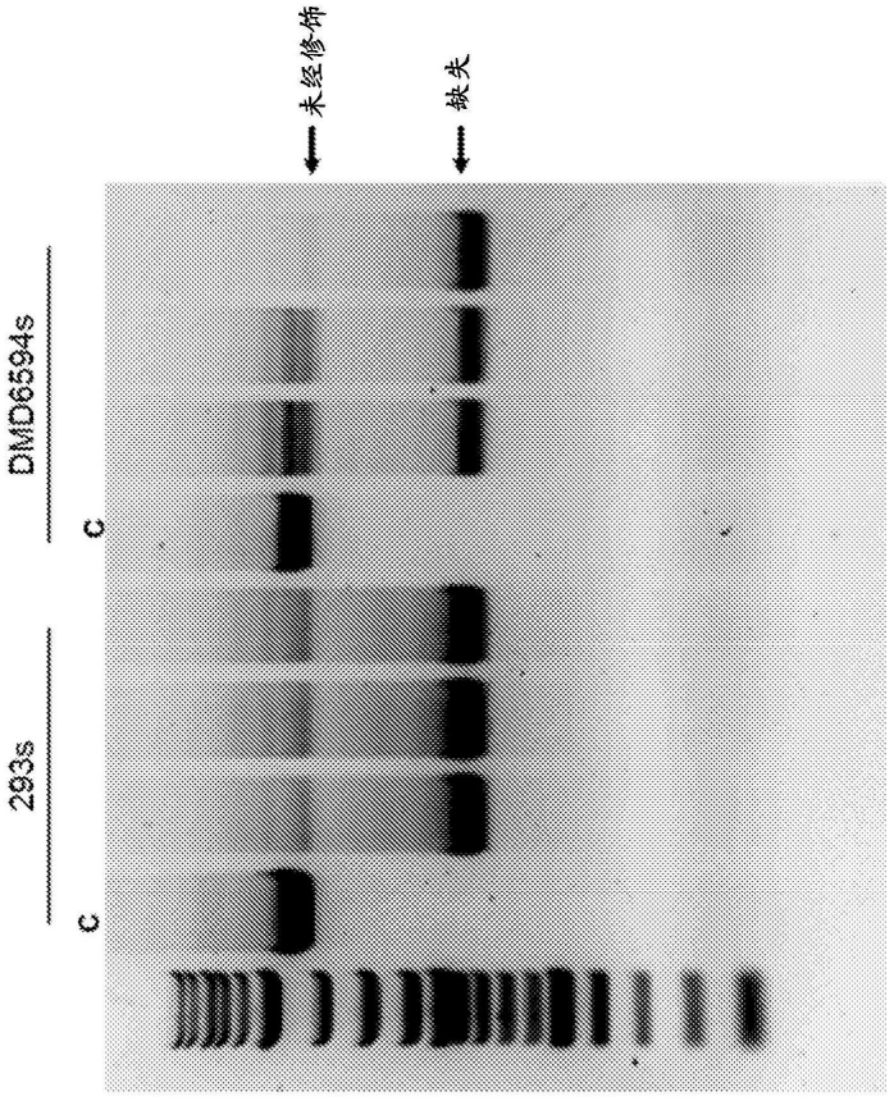


图9

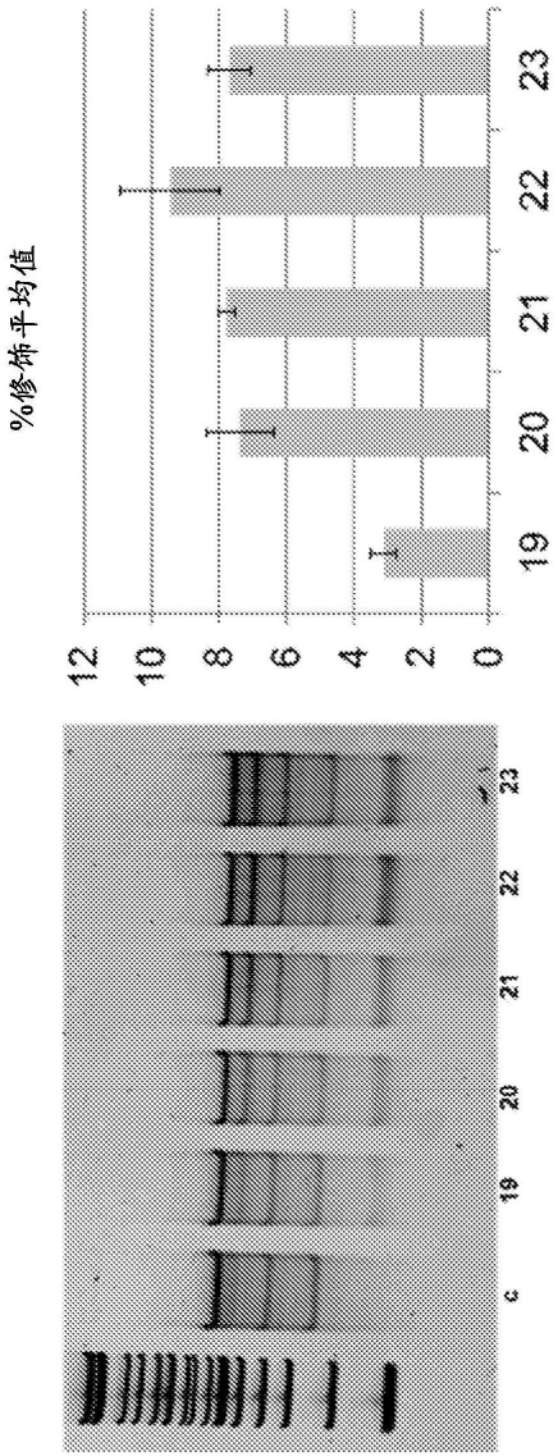


图10

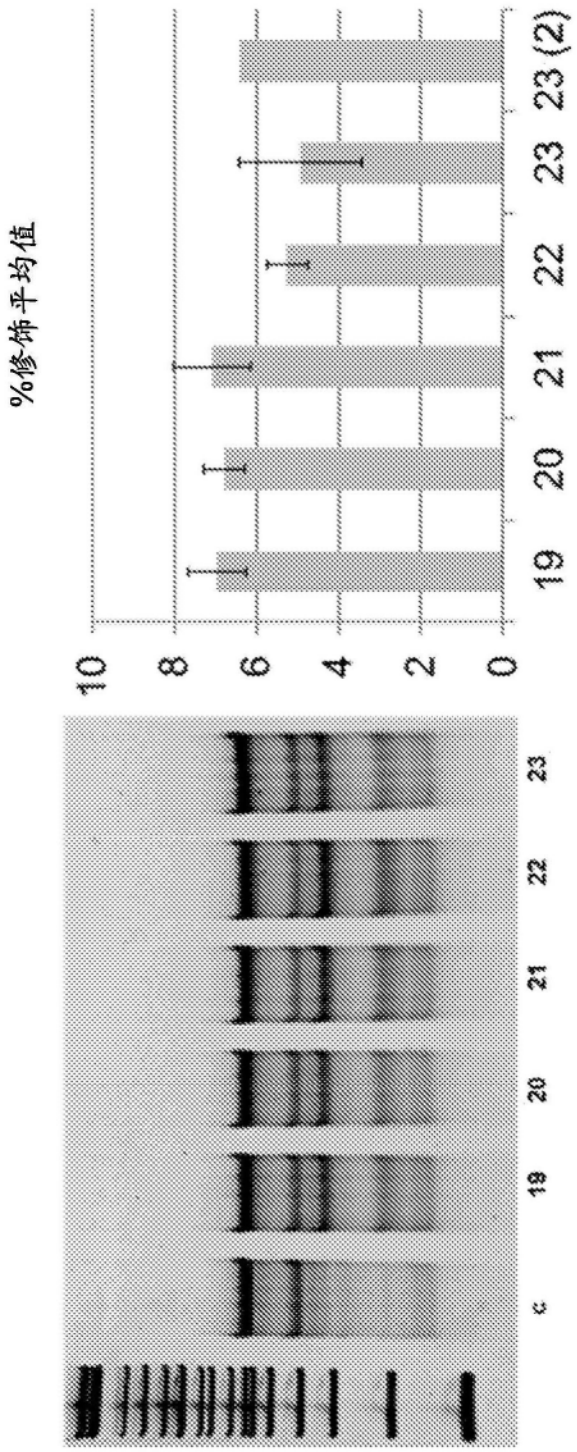


图11

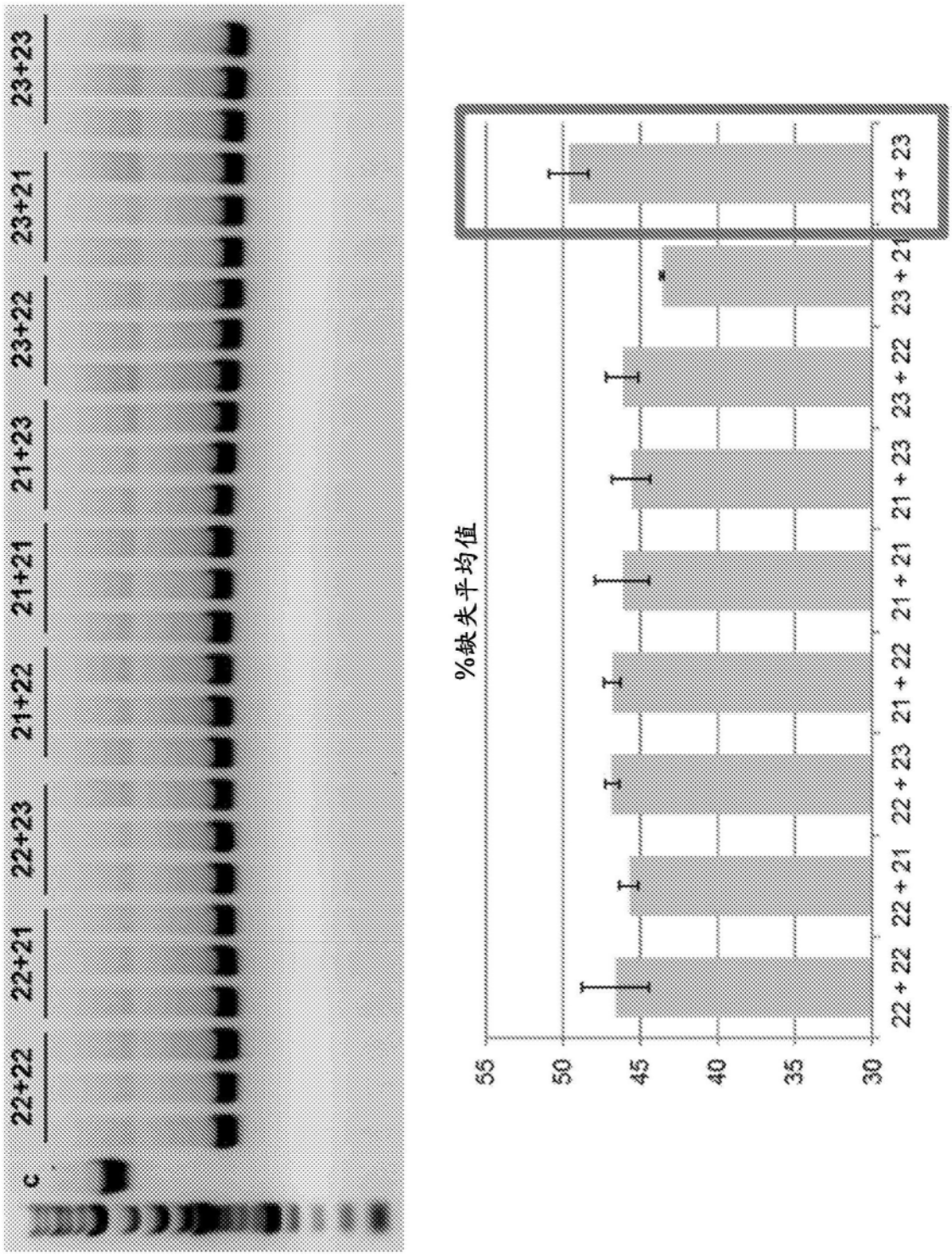

c
22+22
22+21
22+23
21+22
21+21
21+23
23+22
23+21
23+23
%缺失平均值
55
50
45
40
35
30
22 + 22
22 + 21
22 + 23
21 + 22
21 + 21
21 + 23
23 + 22
23 + 21
23 + 23

图12

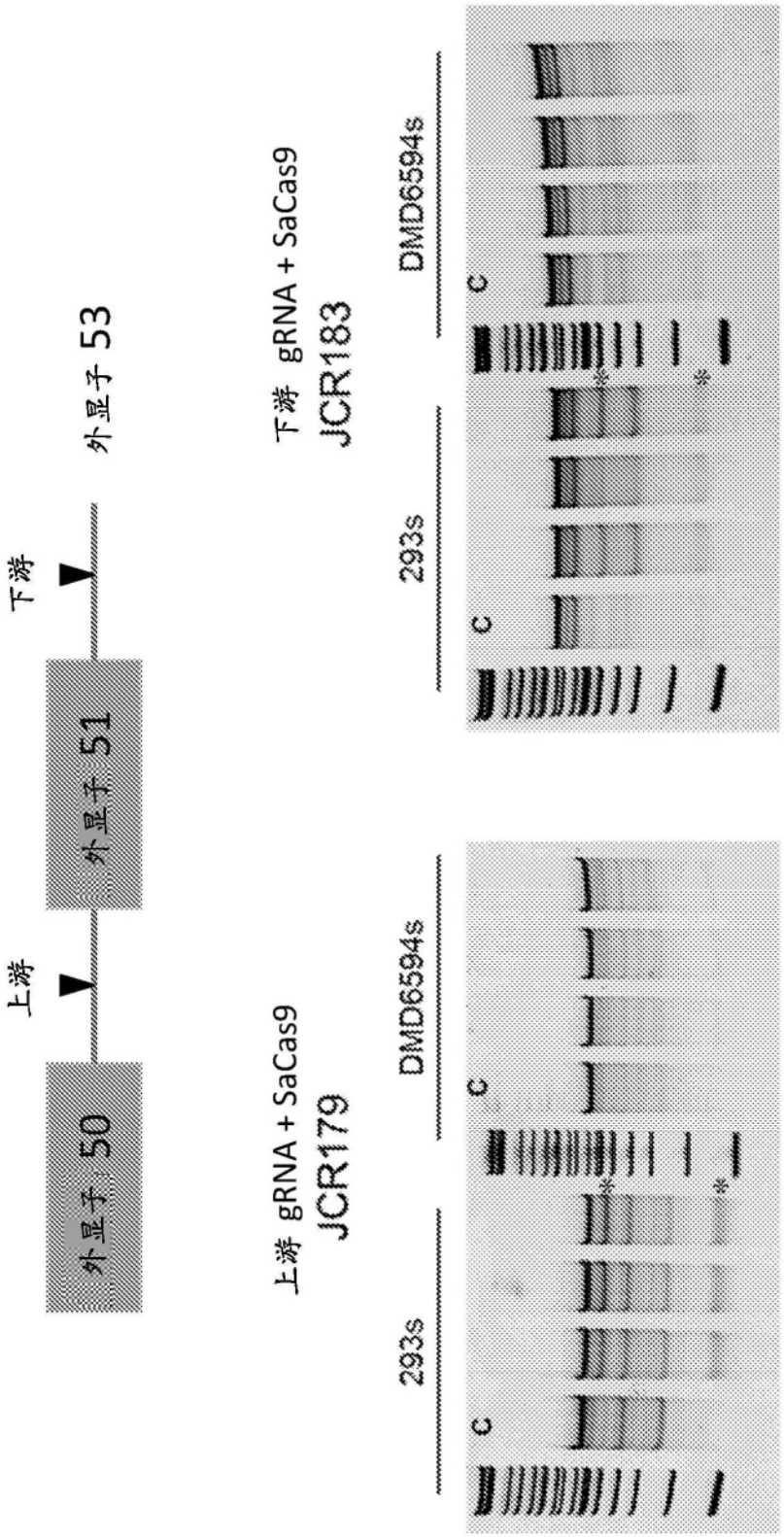

外显子 50
上游
外显子 51
下游
外显子 53
上游 gRNA + SaCas9
JCR179
293s
DMD6594s
c
c
下游 gRNA + SaCas9
JCR183
293s
DMD6594s
c
c

图13

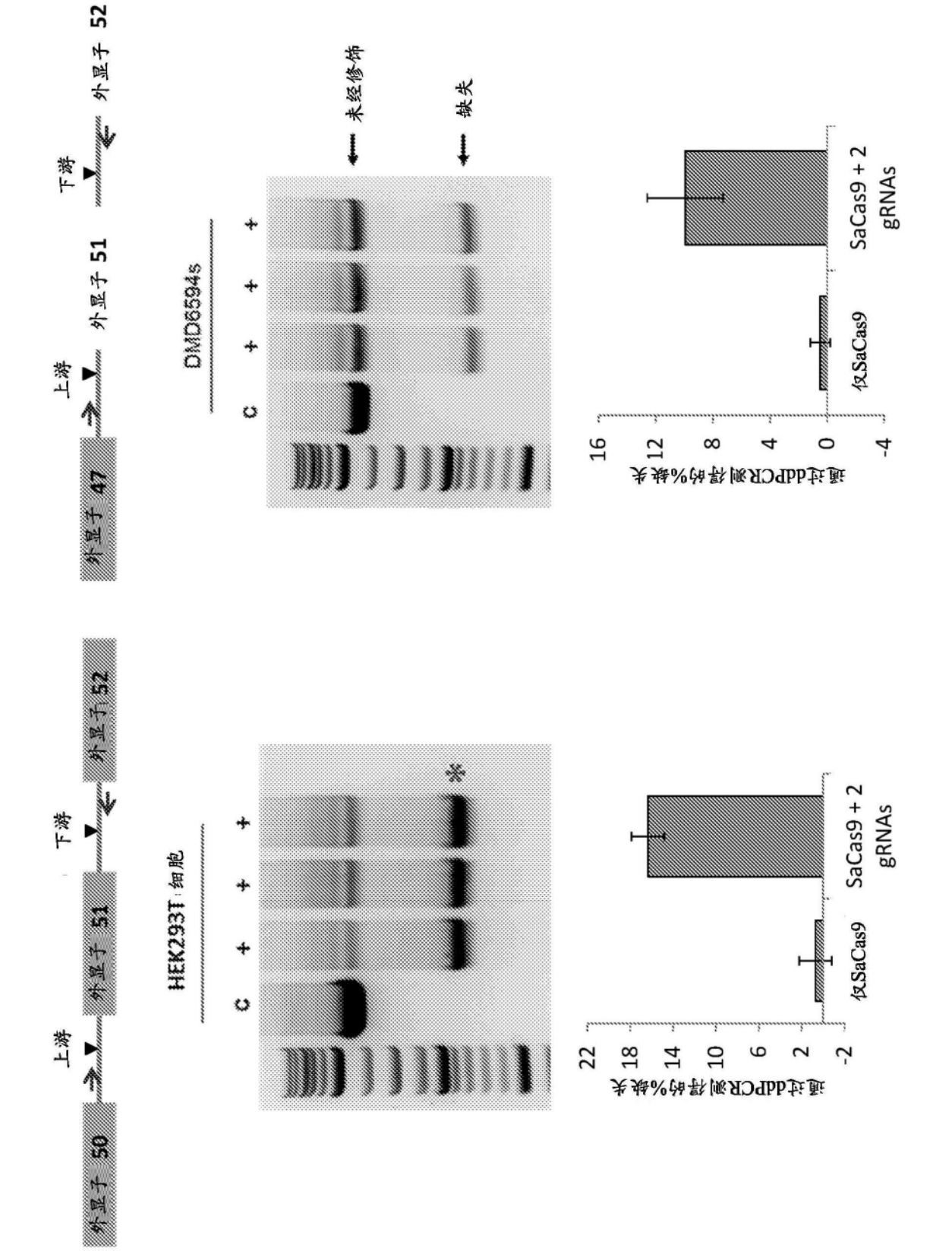

外显子 50
上游
外显子 51
下游
外显子 52
HEK293T 细胞
C
+
+
+
*
通过ddPCR测得的%缺失
22
18
14
10
6
2
-2
仅SaCas9
SaCas9 + 2 gRNAs
外显子 47
上游
外显子 51
下游
外显子 52
DMD6594s
C
+
+
+
未经修饰
缺失
通过ddPCR测得的%缺失
16
12
8
4
0
-4
仅SaCas9
SaCas9 + 2 gRNAs

图14

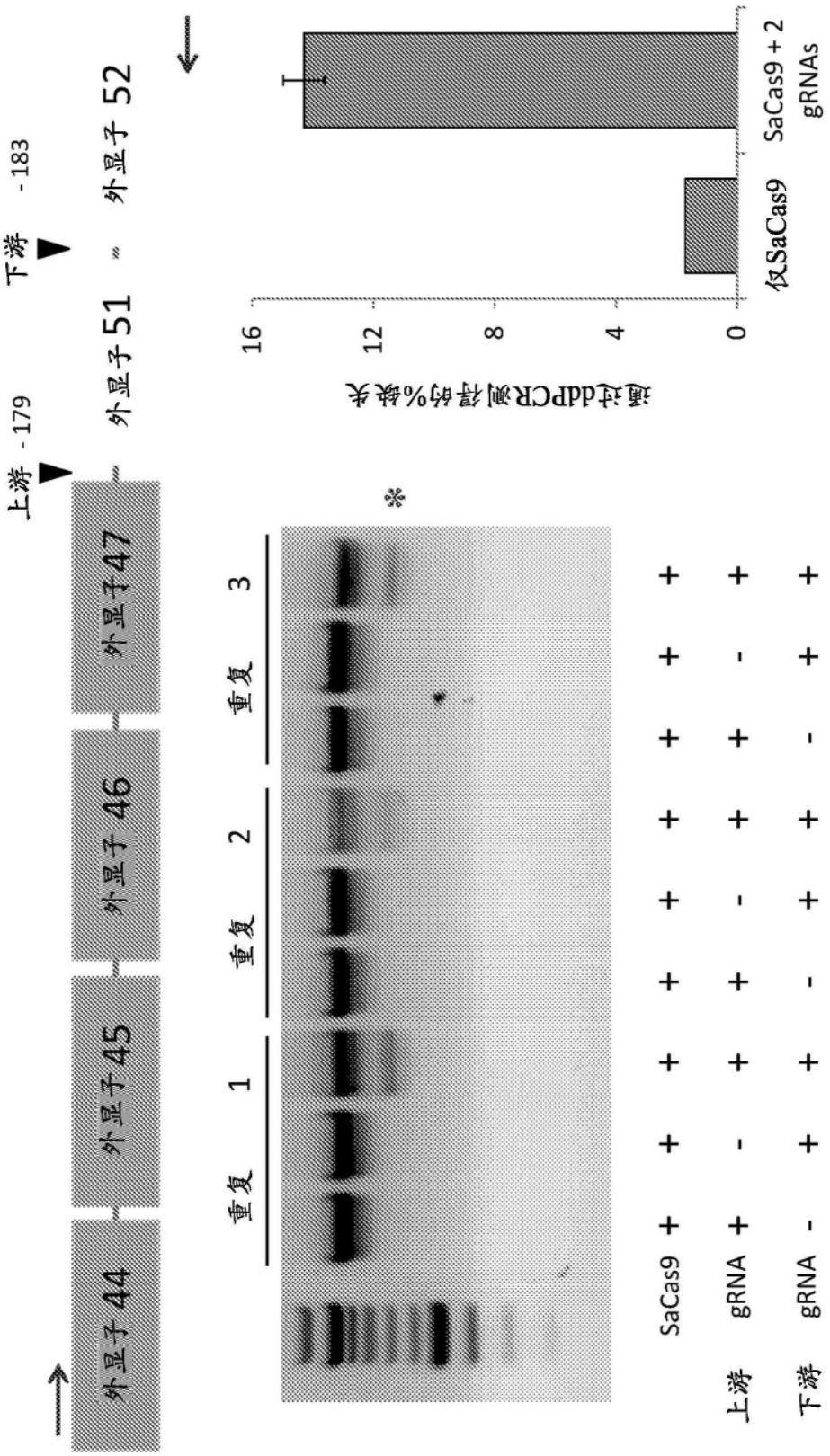

外显子44
外显子45
外显子46
外显子47
上游 - 179
外显子51
下游 - 183
外显子52
重复 1
重复 2
重复 3
SaCas9
上游 gRNA
下游 gRNA
通过ddPCR测得的%缺失
16
12
8
4
0
仅SaCas9
SaCas9 + 2 gRNAs

图15

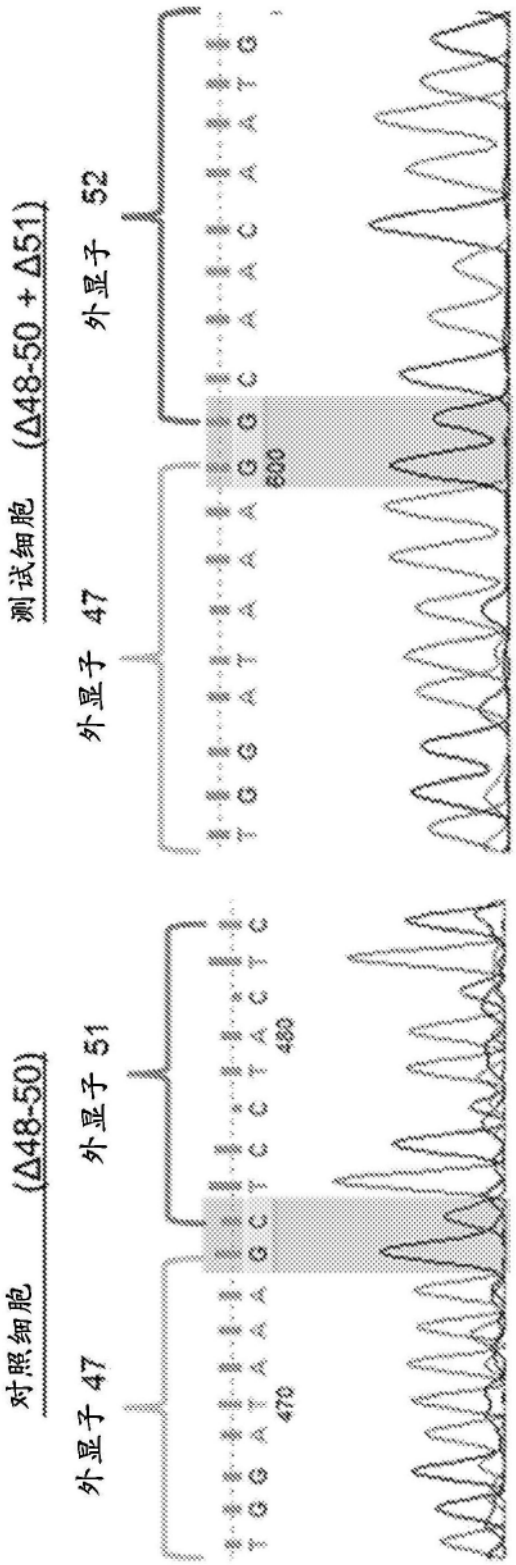

对照细胞 (Δ48-50)
外显子 47
外显子 51
470
480
测试细胞 (Δ48-50 + Δ51)
外显子 47
外显子 52
600

图16

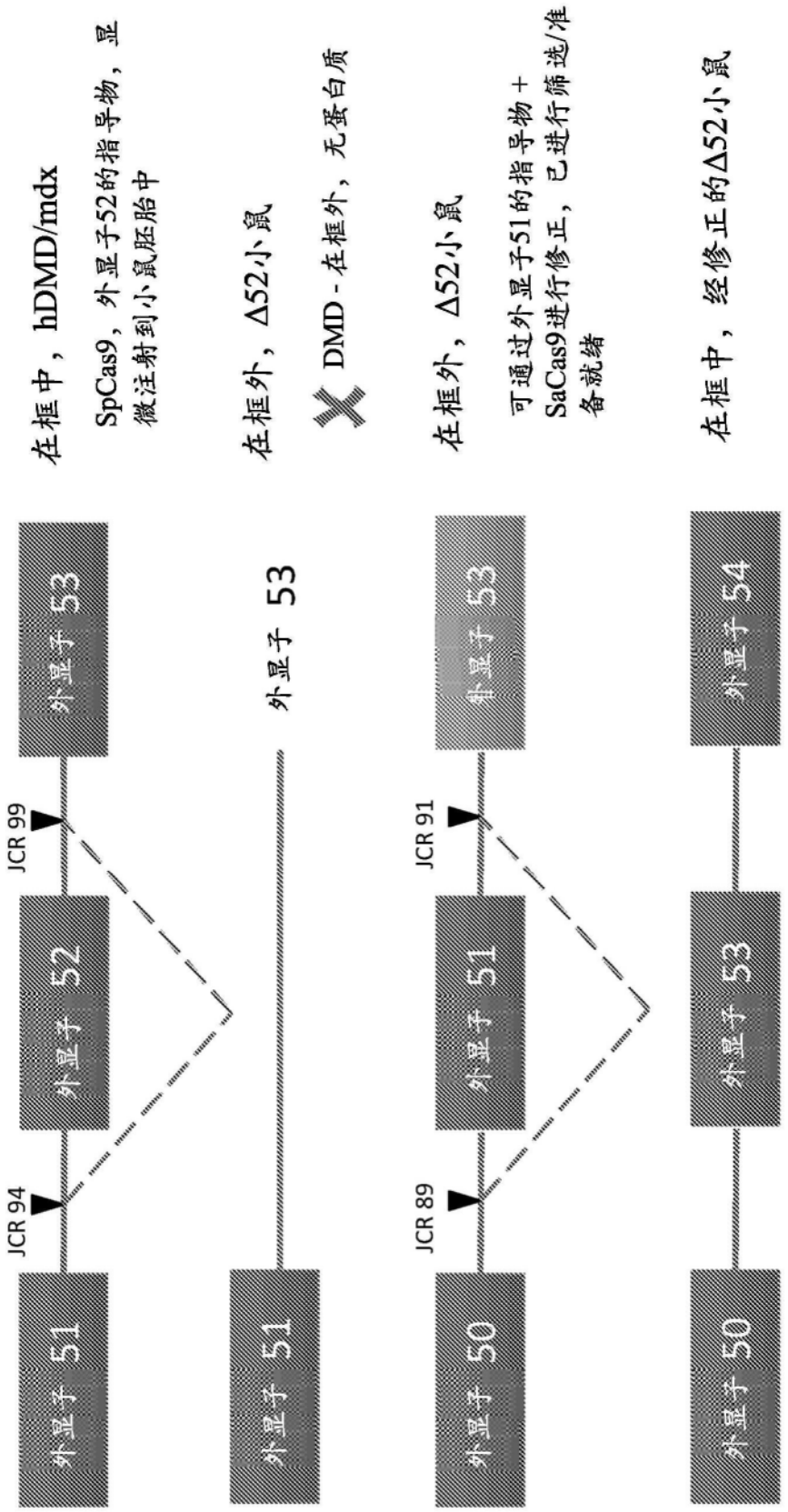

在框中，hDMD/mdx
SpCas9，外显子52的指导物，显微注射到小鼠胚胎中
在框外，Δ52小鼠
DMD - 在框外，无蛋白质
在框外，Δ52小鼠
可通过外显子51的指导物 + SaCas9进行修正，已进行筛选/准备就绪
在框中，经修正的Δ52小鼠
外显子 53
外显子 51
外显子 53
JCR 99
外显子 52
JCR 94
外显子 51
外显子 53
JCR 91
外显子 51
JCR 89
外显子 50
外显子 54
外显子 53
外显子 50

图17

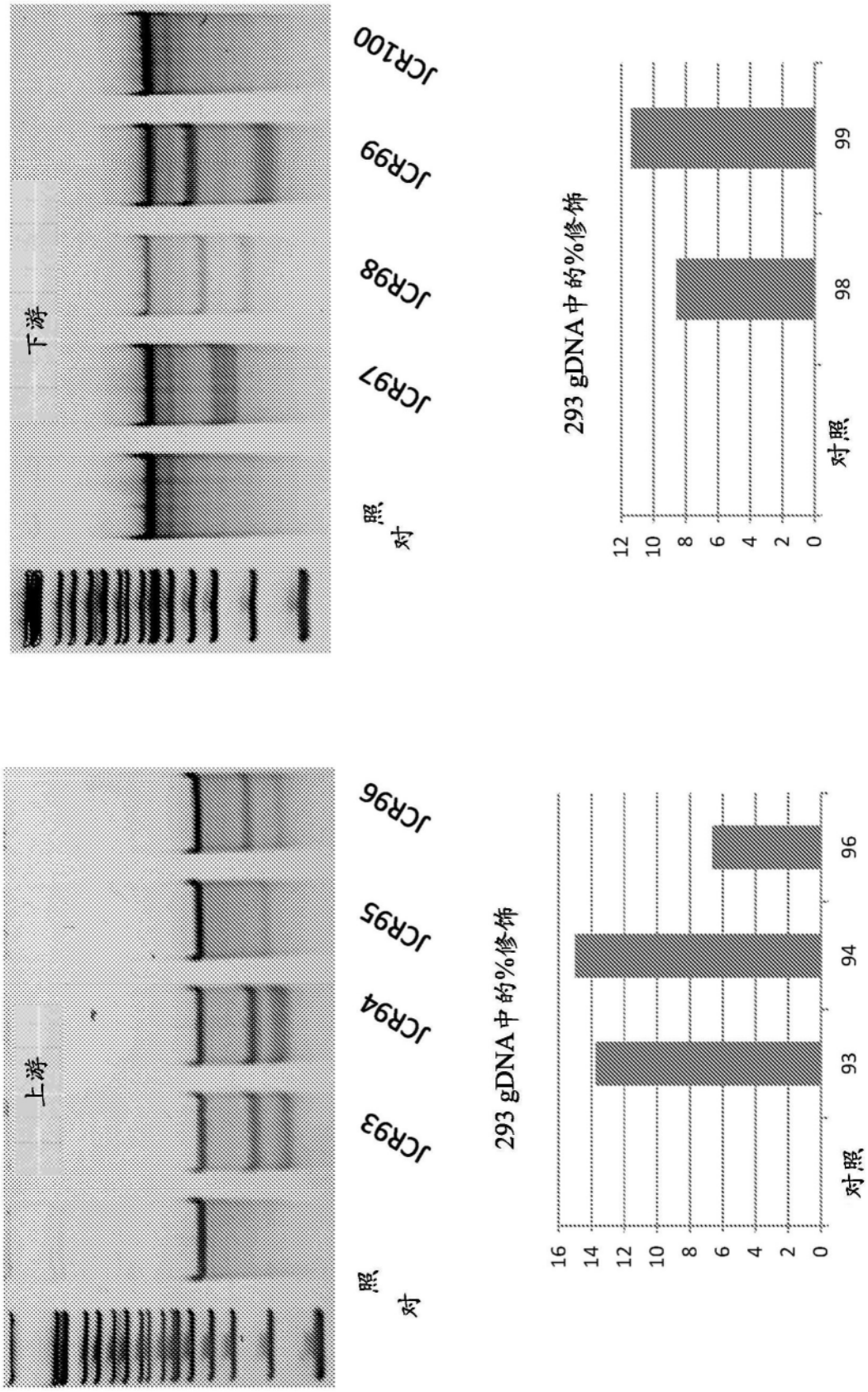

下游
照对
JCR97
JCR98
JCR99
JCR100
293 gDNA 中的%修饰
0
2
4
6
8
10
12
对照
98
99
上游
照对
JCR93
JCR94
JCR95
JCR96
293 gDNA 中的%修饰
0
2
4
6
8
10
12
14
16
对照
93
94
96

图18

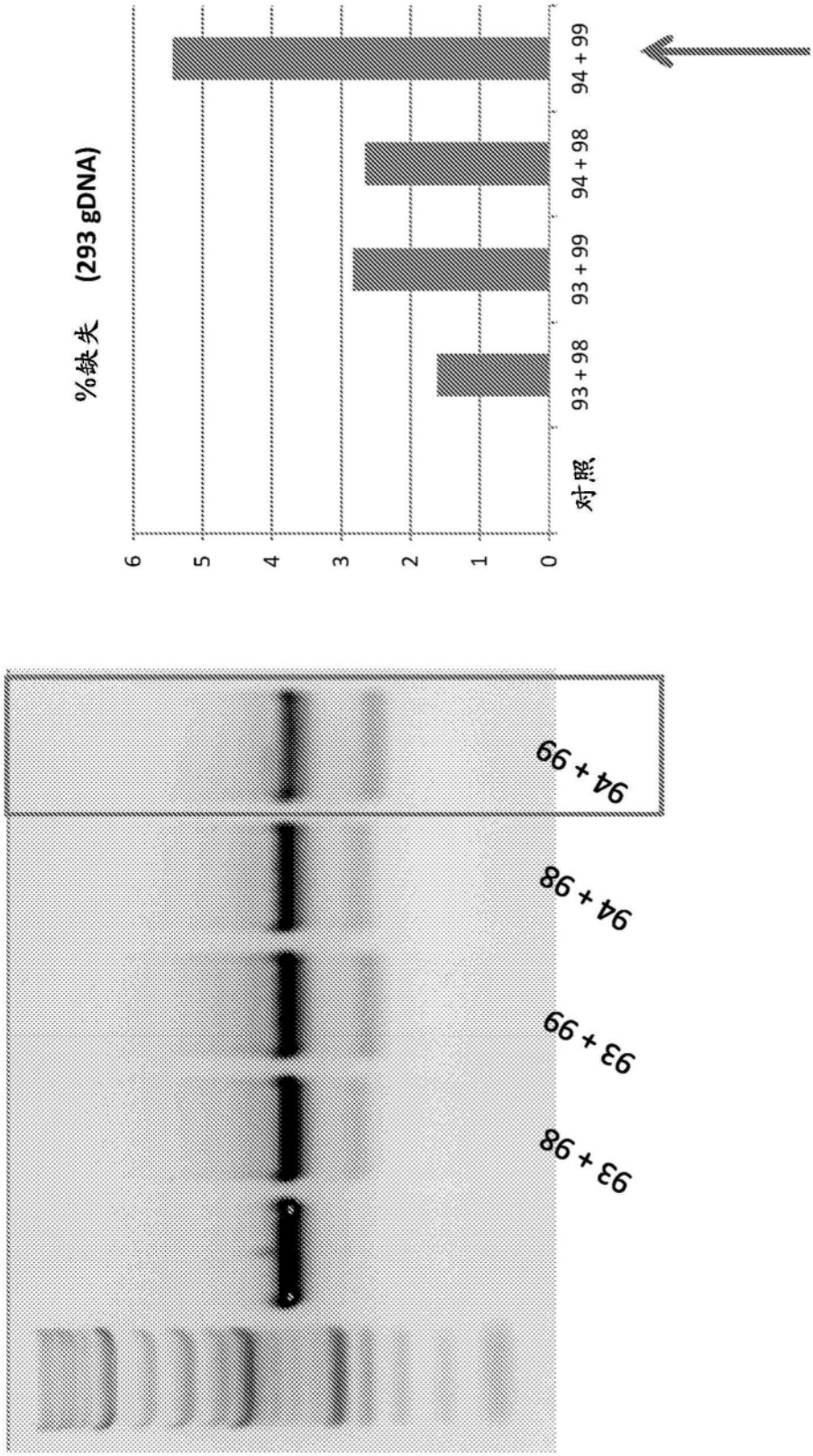

%缺失 (293 gDNA)
6
5
4
3
2
1
0
对照
93 + 98
93 + 99
94 + 98
94 + 99
93 + 98
93 + 99
94 + 98
94 + 99

图19

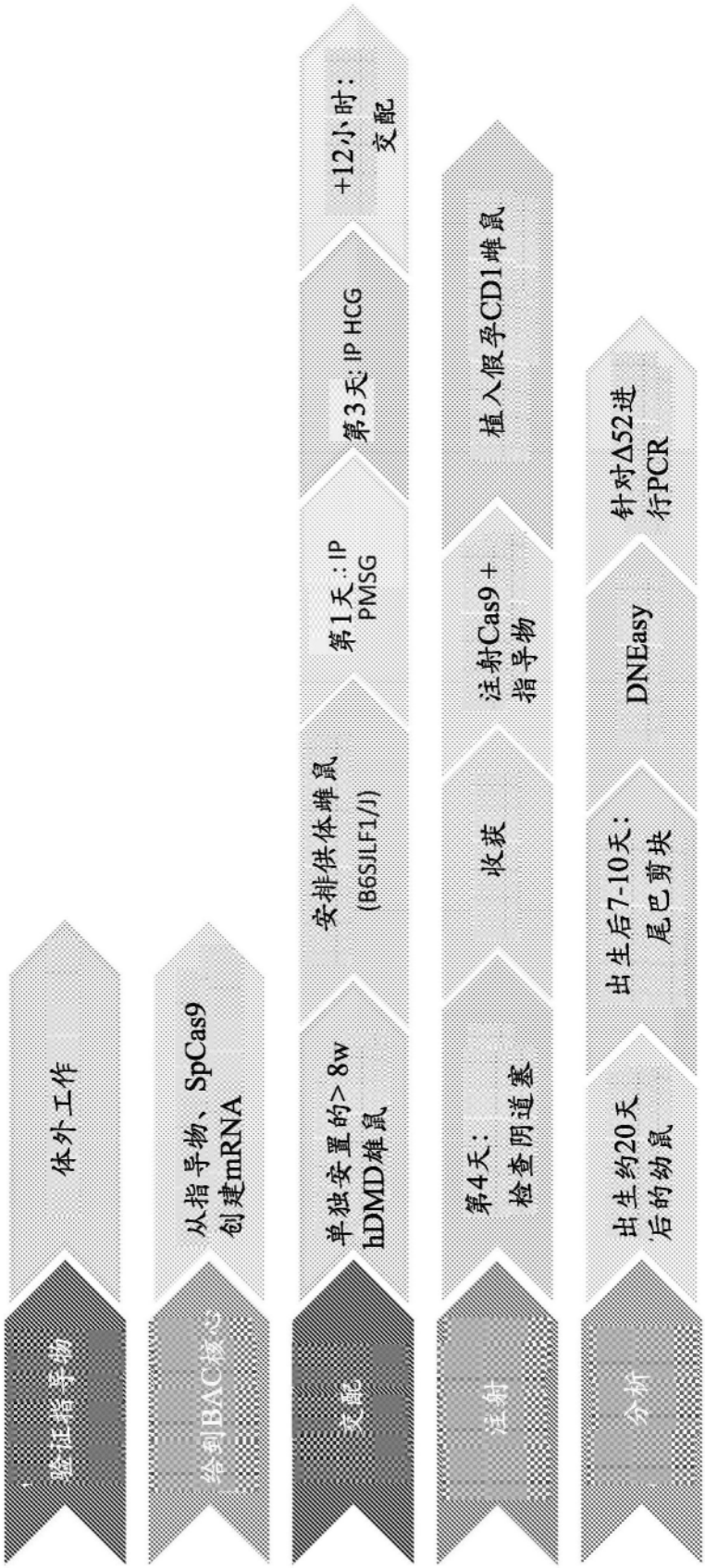

验证指导物
体外工作
给到BAC核心
从指导物、SpCas9
创建mRNA
交配
单独安置的> 8w
hDMD雄鼠
安排供体雌鼠
(B6SJLF1/J)
第1天：IP
PMSG
第3天: IP HCG
+12小时：
交配
注射
第4天：
检查阴道塞
收获
注射Cas9 +
指导物
植入假孕CD1雌鼠
分析
出生约20天
后的幼鼠
出生后7-10天：
尾巴剪块
DNEasy
针对Δ52进
行PCR

图20

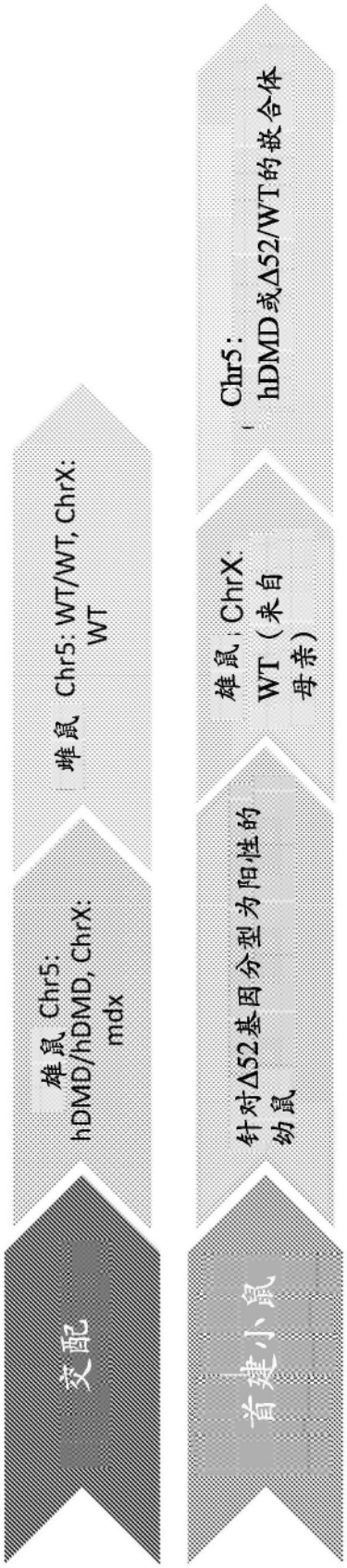

交配
雄鼠 Chr5: hDMD/hDMD, ChrX: mdx
雌鼠 Chr5: WT/WT, ChrX: WT
首建小鼠
针对Δ52基因分型为阳性的幼鼠
雄鼠 ChrX: WT（来自母亲）
Chr5: hDMD或Δ52/WT的嵌合体

图21

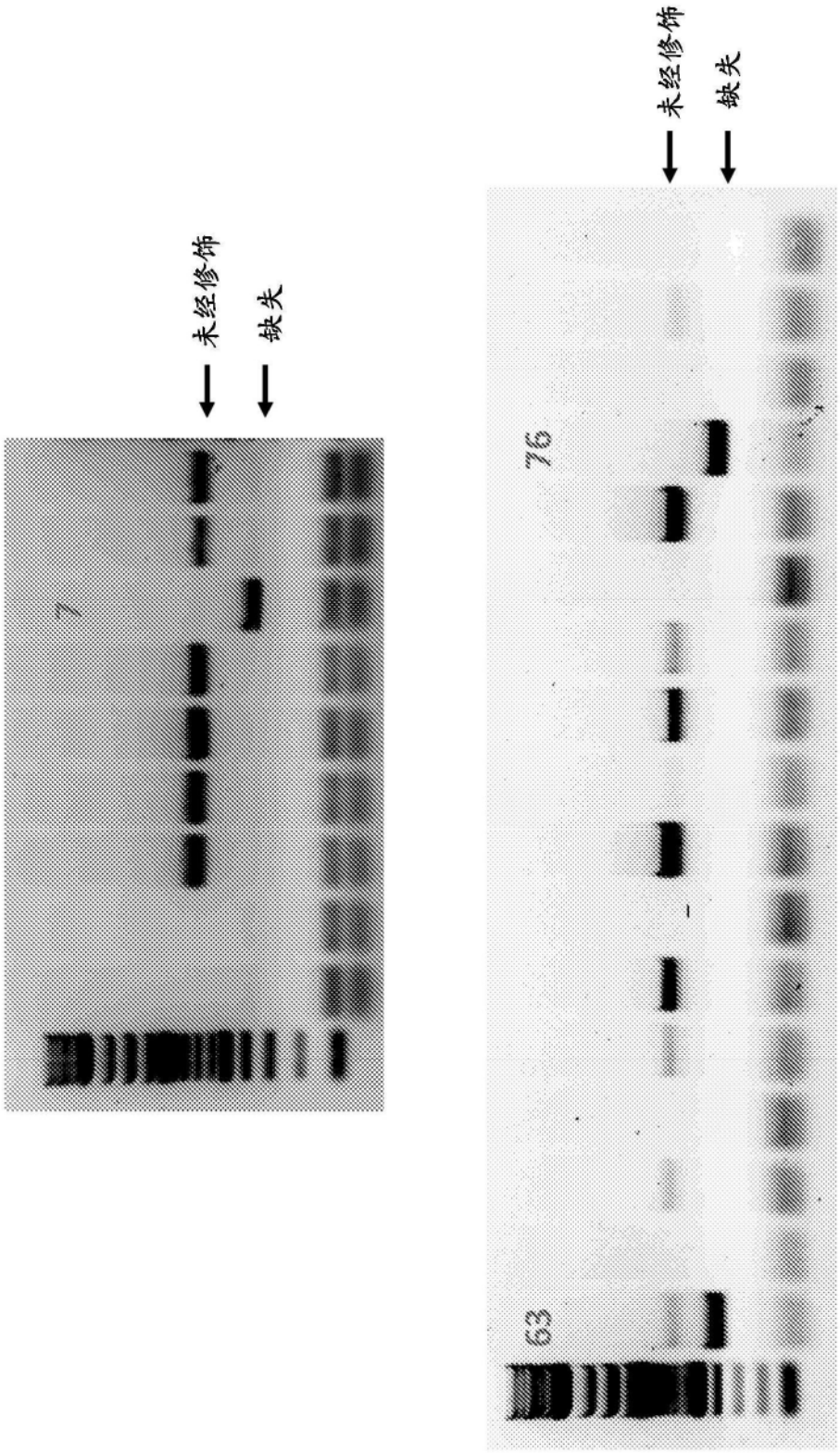


图22

紫色粗体： JCR94; 绿色加下划线： JCR99; 灰色斜体： 插入/缺失

Δ52 : AACGCTGAAGAACCCT**GAT**AATAGAAGAAATACATTTTTAAATCAATTCAGG (SEQ ID NO: 46)

首建鼠 76: AAGAACCCT**GAT**A –*TTATCTTAGTAGATTA* – ATAGAAGAAATACATTTTTAAA (SEQ ID NO: 47)

首建鼠 63: GCTGAAGAACCCT**GA**-------AAAAATACATTTTTTTATCAATTCAGG (SEQ ID NO: 48)

首建鼠 7: GAAT-----(19x)-----**GAT**-*TTTCTTG*TAGAAGAAATAACAATT----AAATC (SEQ ID NO: 49)

图23

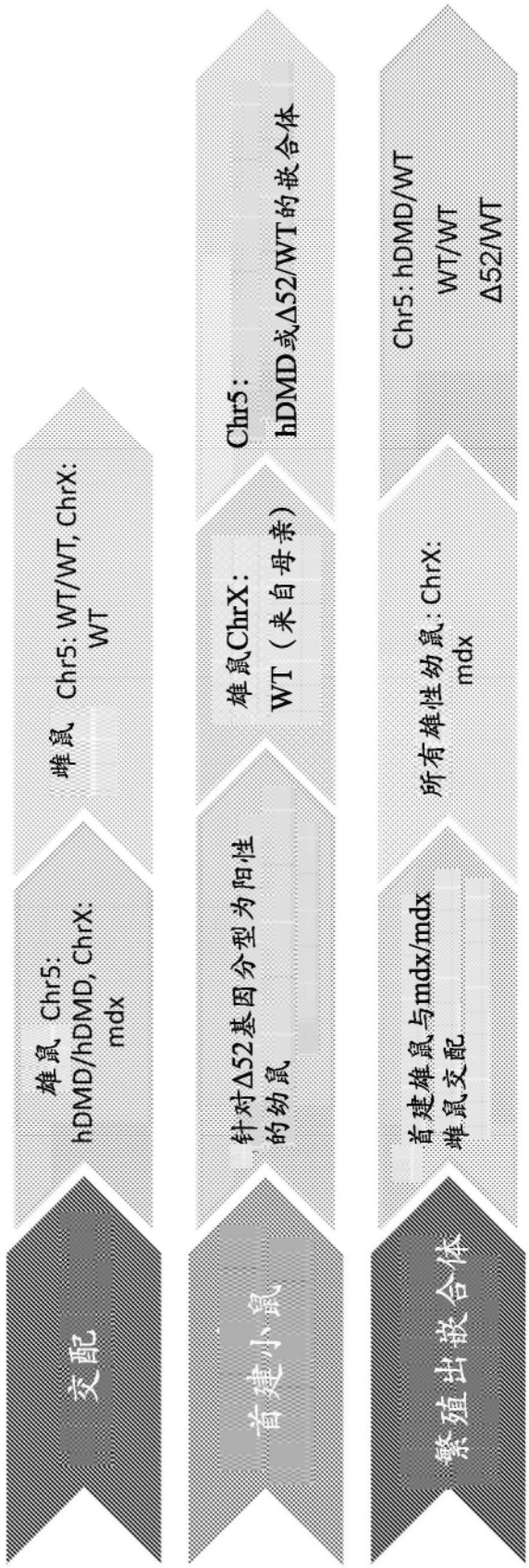

交配
雄鼠 Chr5: hDMD/hDMD, ChrX: mdx
雌鼠 Chr5: WT/WT, ChrX: WT
首建小鼠
针对Δ52基因分型为阳性的幼鼠
雄鼠ChrX: WT（来自母亲）
Chr5: hDMD或Δ52/WT的嵌合体
繁殖出嵌合体
首建雄鼠与mdx/mdx雌鼠交配
所有雄性幼鼠: ChrX: mdx
Chr5: hDMD/WT
WT/WT
Δ52/WT

图24

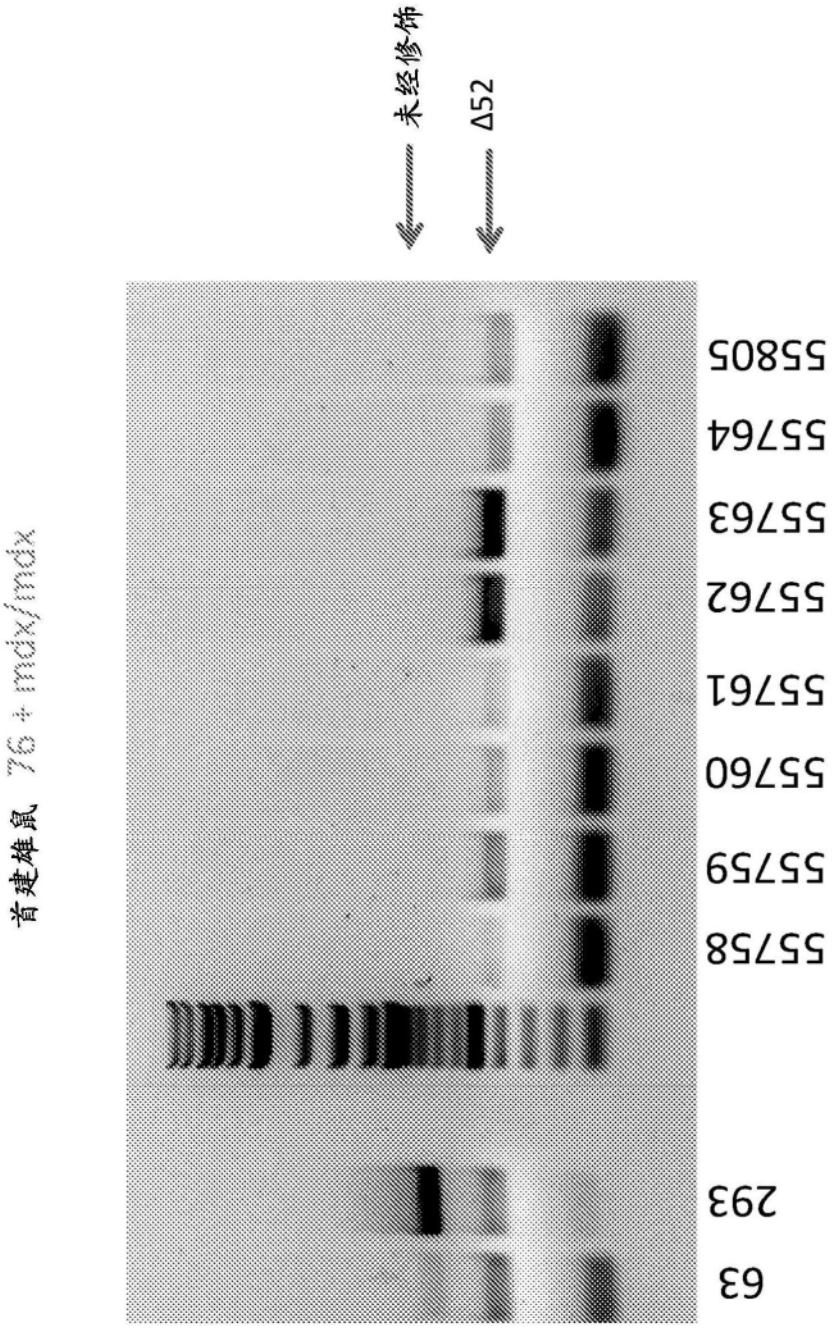

首建雄鼠 76 + mdx/mdx
未经修饰
Δ52
63
293
55758
55759
55760
55761
55762
55763
55764
55805

图25

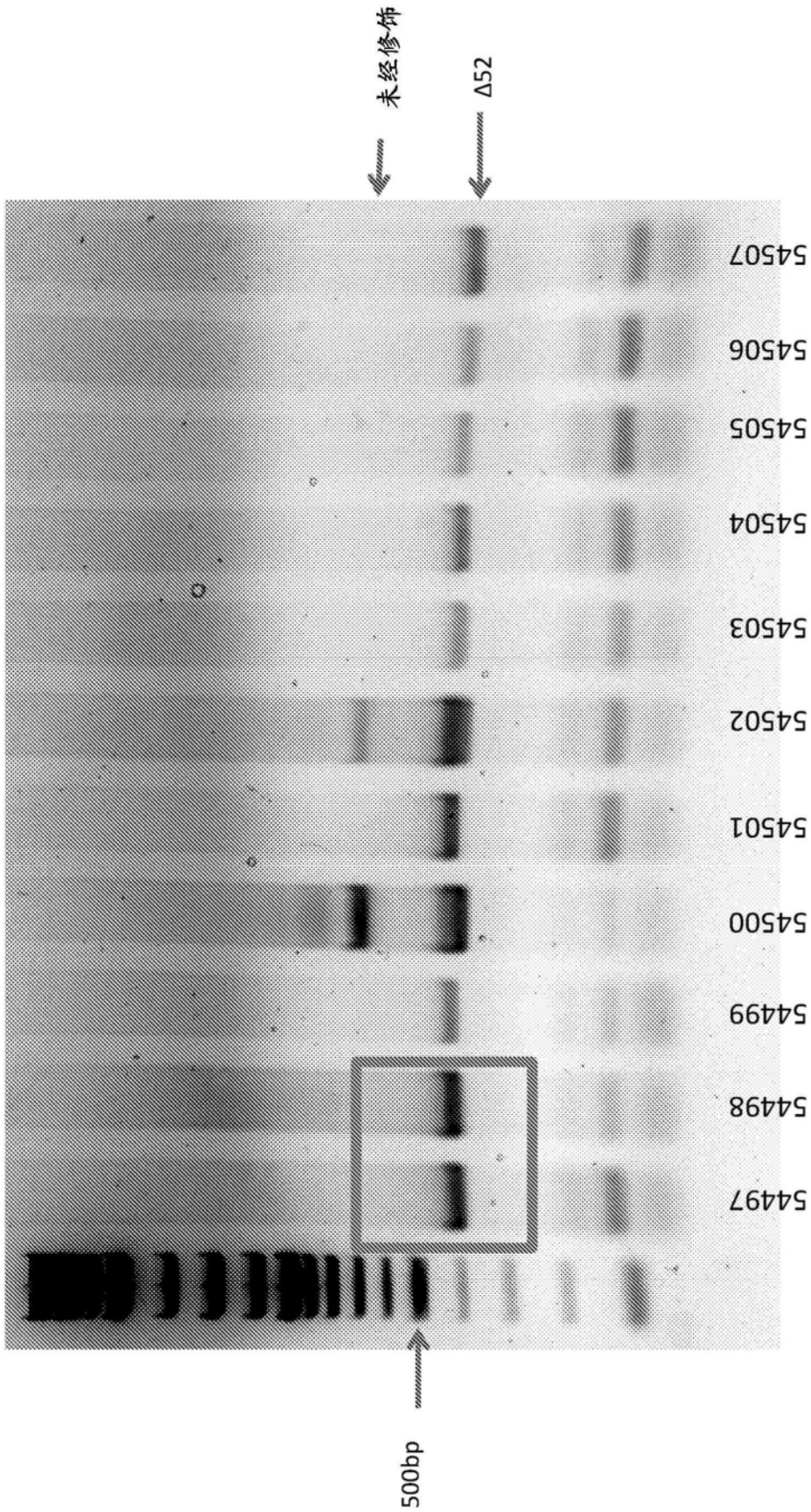

未经修饰
Δ52
500bp
54497
54498
54499
54500
54501
54502
54503
54504
54505
54506
54507

图26

紫色粗体： JCR94; 绿色加下划线： JCR99; 灰色斜体：插入/缺失

Δ52 : AACGCTGAAGAACCCT**GAT**AATAGAAGAAATACATTTTTAAATCAATTCAGG (SEQ ID NO: 50)

54497: AACGCTGAAGAACCCT**GAT***ATTATCTTAGTAGATTA*ATAGAAGAAATACATTTTTAAAT (SEQ ID NO: 51)

54498: AACGCTGAAGAACCCT**GAT***ATTATCTTAGTAGATTA*ATAGAAGAAATACATTTTTAAAT (SEQ ID NO: 52)

图27

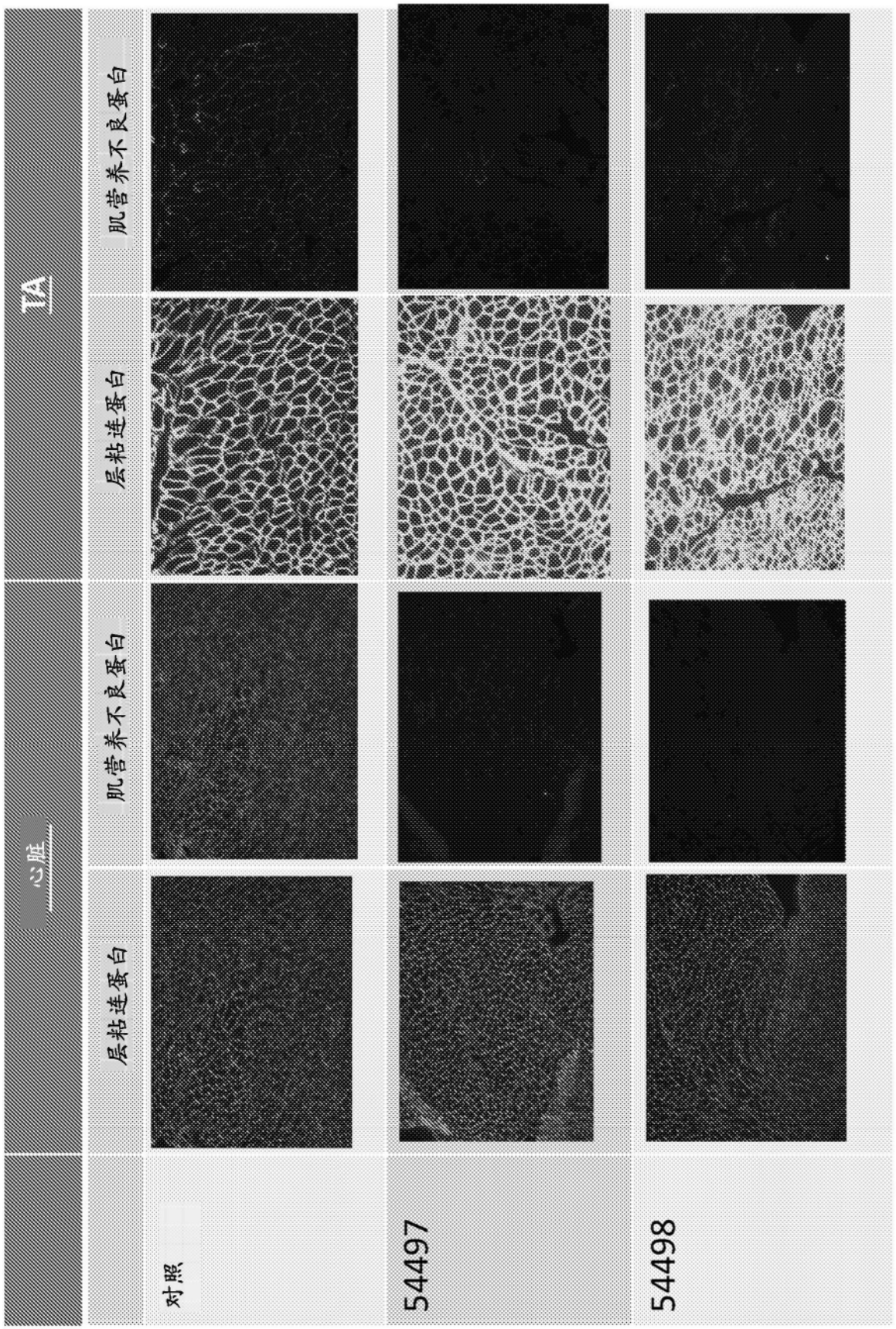


图28

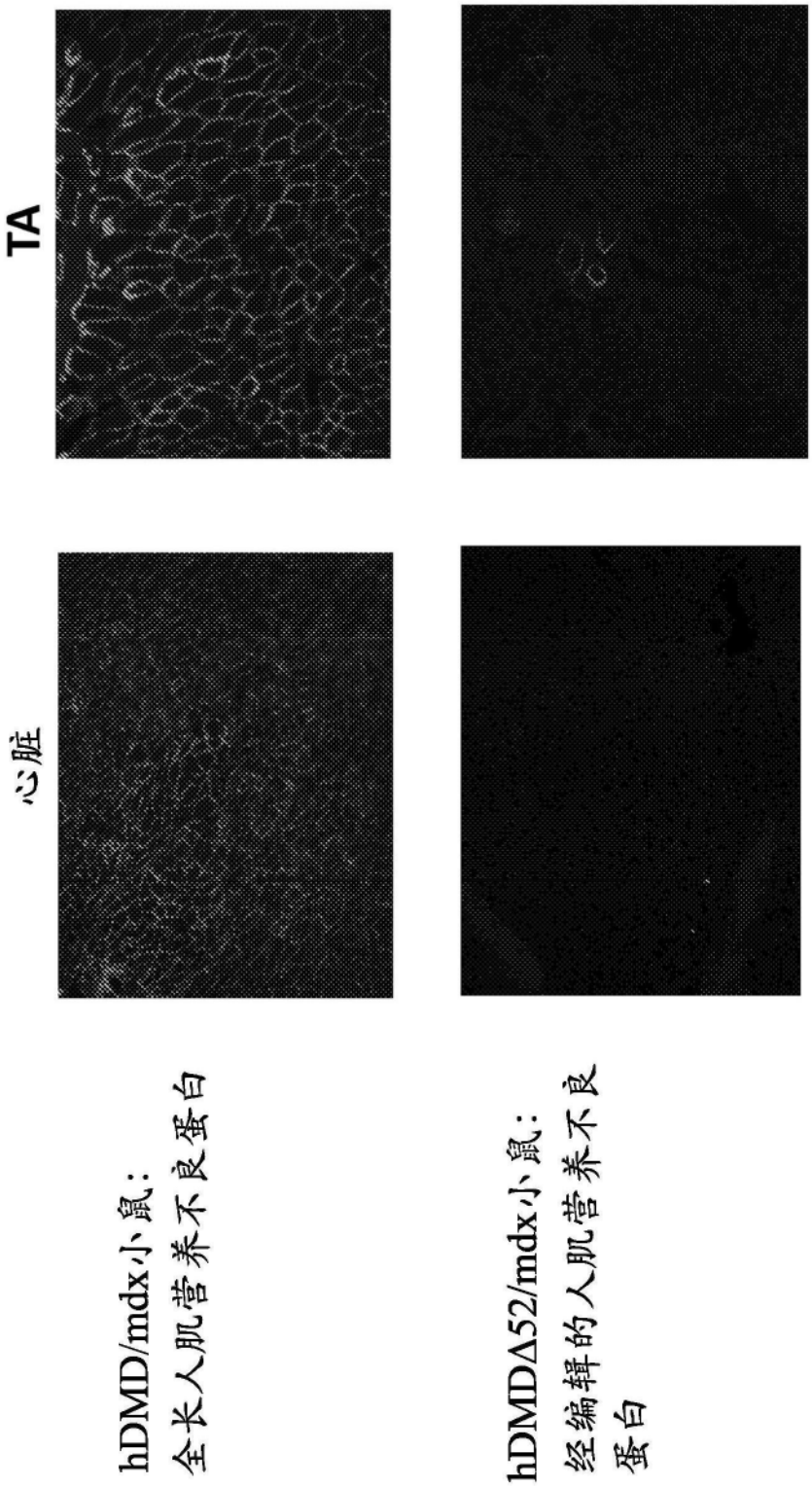


图29

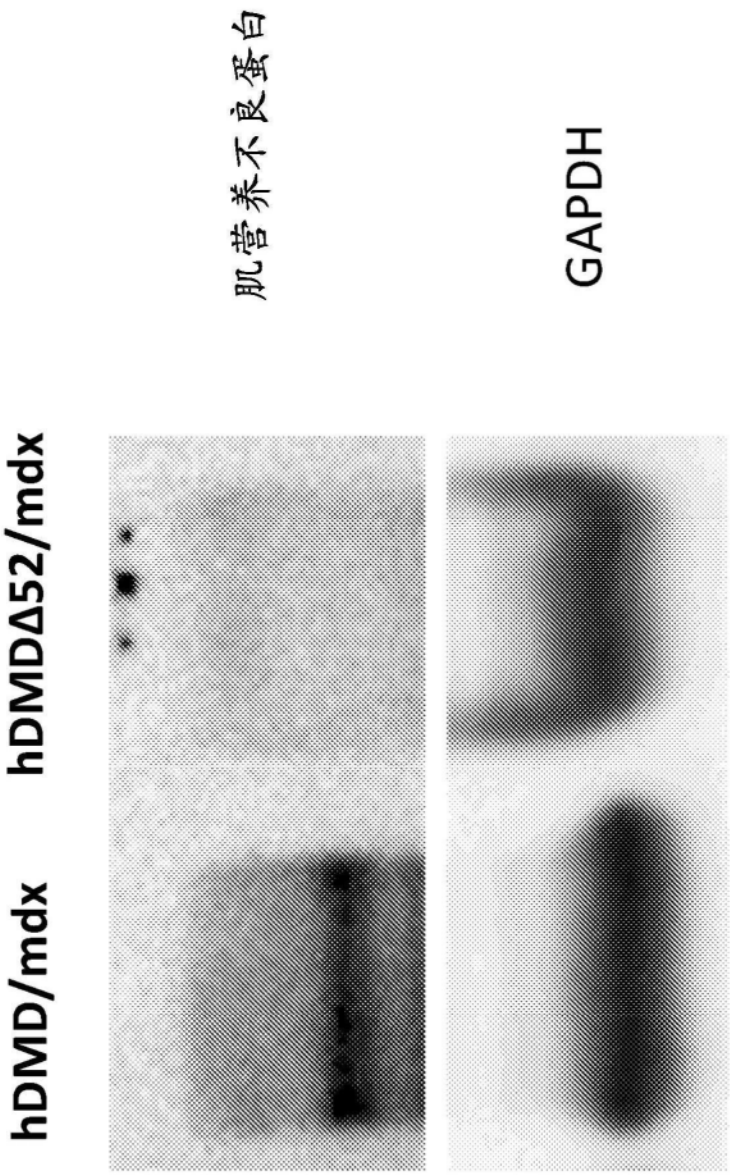
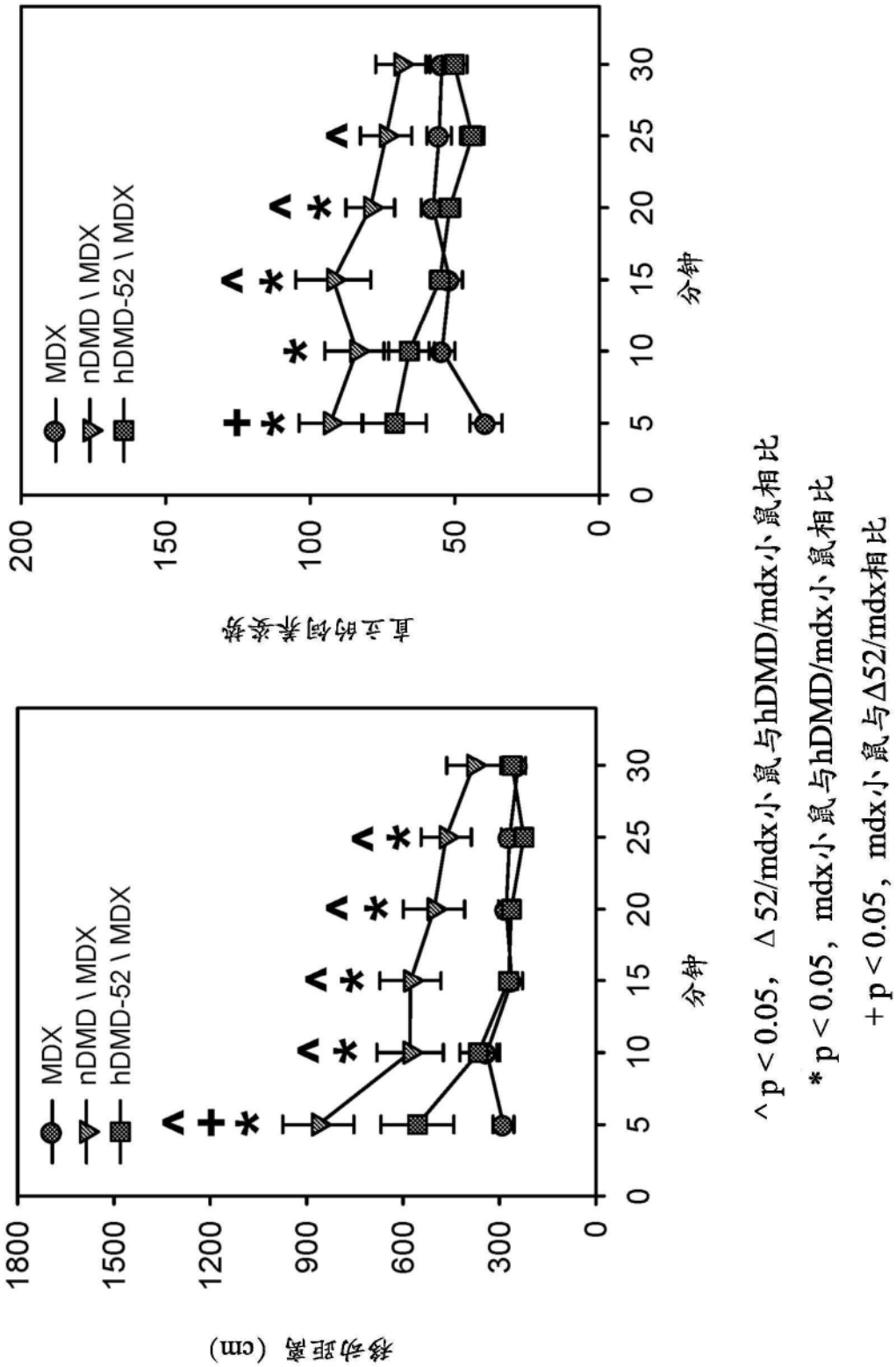


图30



^ $p < 0.05$，Δ 52/mdx小鼠与hDMD/mdx小鼠相比

* $p < 0.05$，mdx小鼠与hDMD/mdx小鼠相比

+ $p < 0.05$，mdx小鼠与Δ52/mdx相比

图31

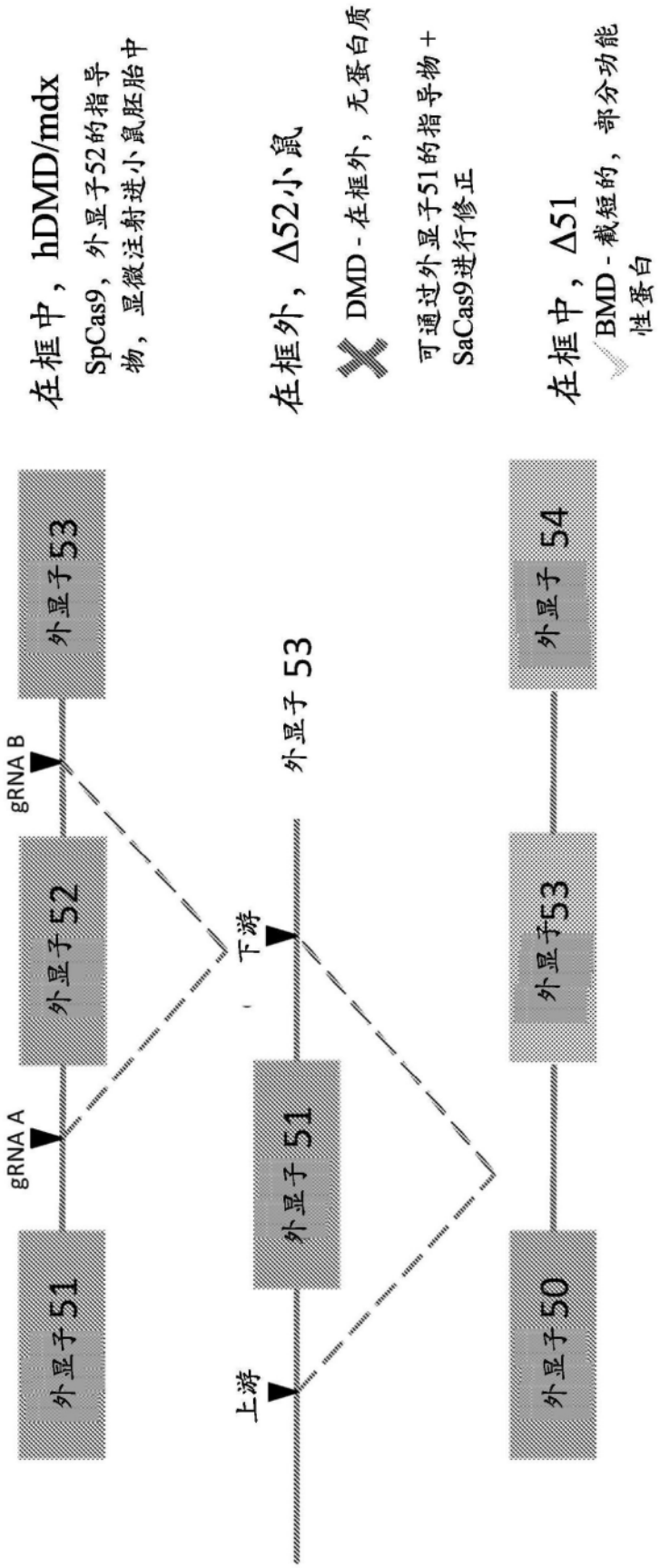

在框中，hDMD/mdx
SpCas9，外显子52的指导物，显微注射进小鼠胚胎中
在框外，Δ52小鼠
DMD - 在框外，无蛋白质
可通过外显子51的指导物 + SaCas9进行修正
在框中，Δ51
BMD - 截短的，部分功能性蛋白
外显子53
gRNA B
外显子52
gRNA A
外显子51
外显子 53
下游
外显子 51
上游
外显子 54
外显子53
外显子50

图32

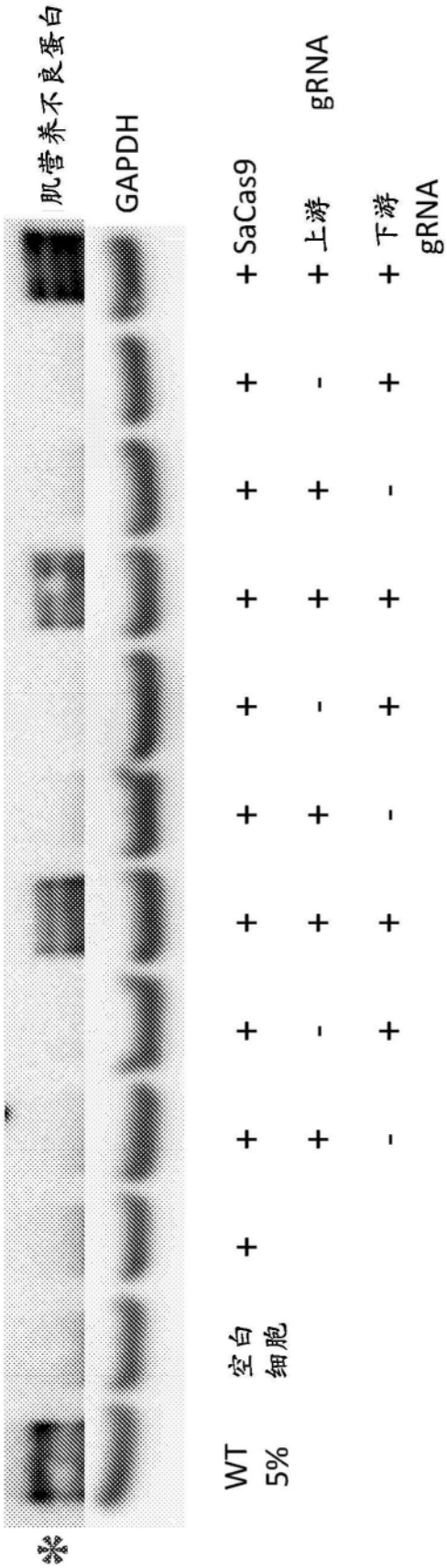


图33

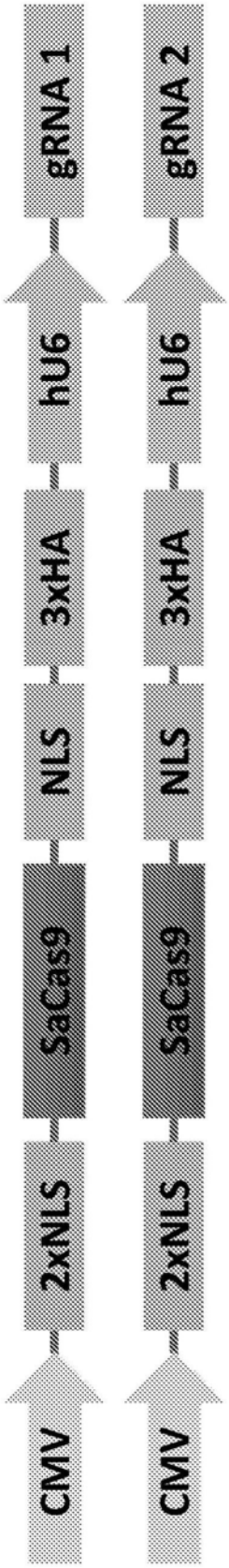

CMV
2xNLS
SaCas9
NLS
3xHA
hU6
gRNA 1
CMV
2xNLS
SaCas9
NLS
3xHA
hU6
gRNA 2

图34

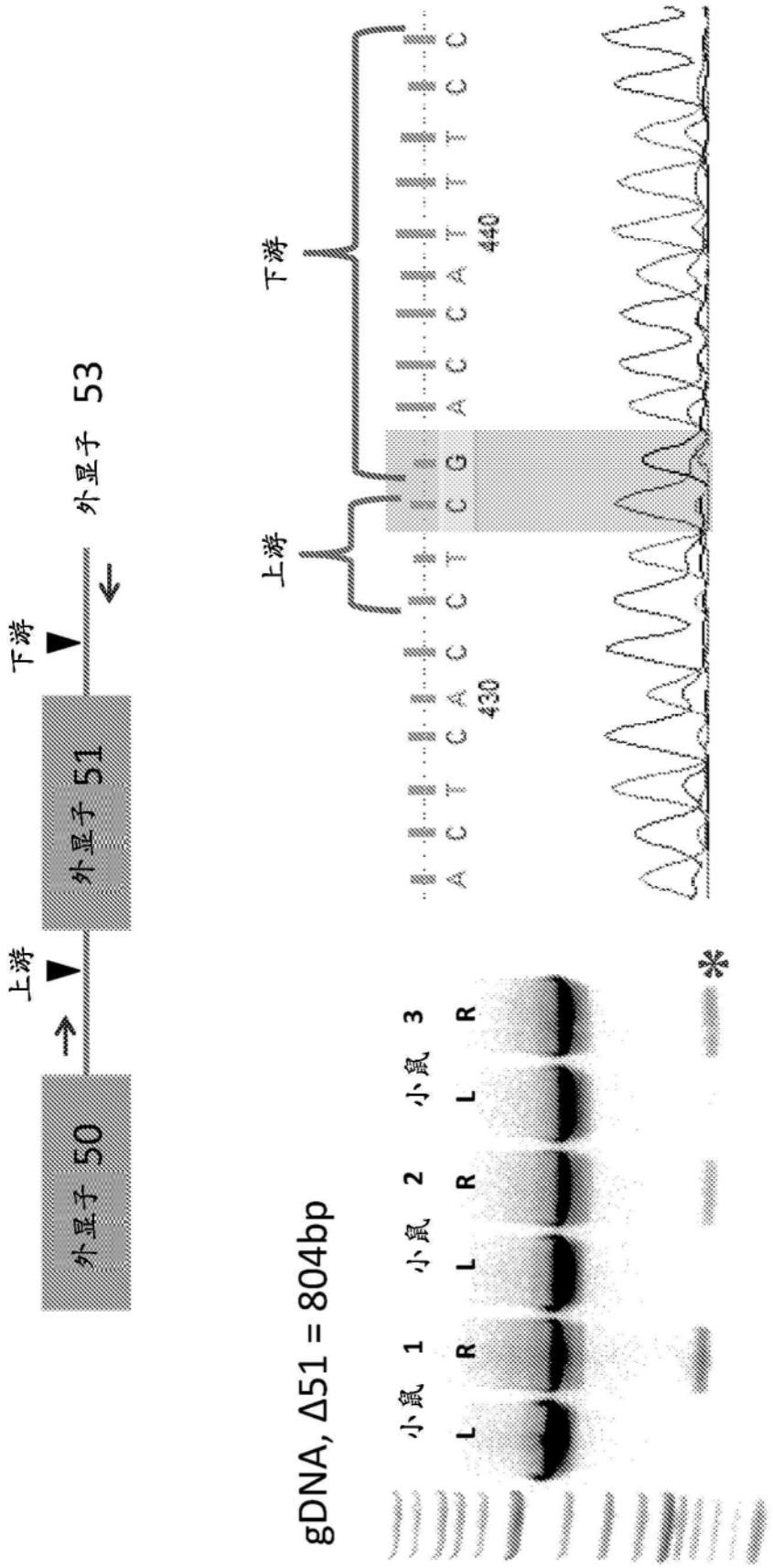

外显子 50
上游
外显子 51
下游
外显子 53
上游
下游
A C T C A C C T C G A C C A T T C C
430
440
gDNA, Δ51 = 804bp
小鼠 1
小鼠 2
小鼠 3
L R L R L R
*

图35

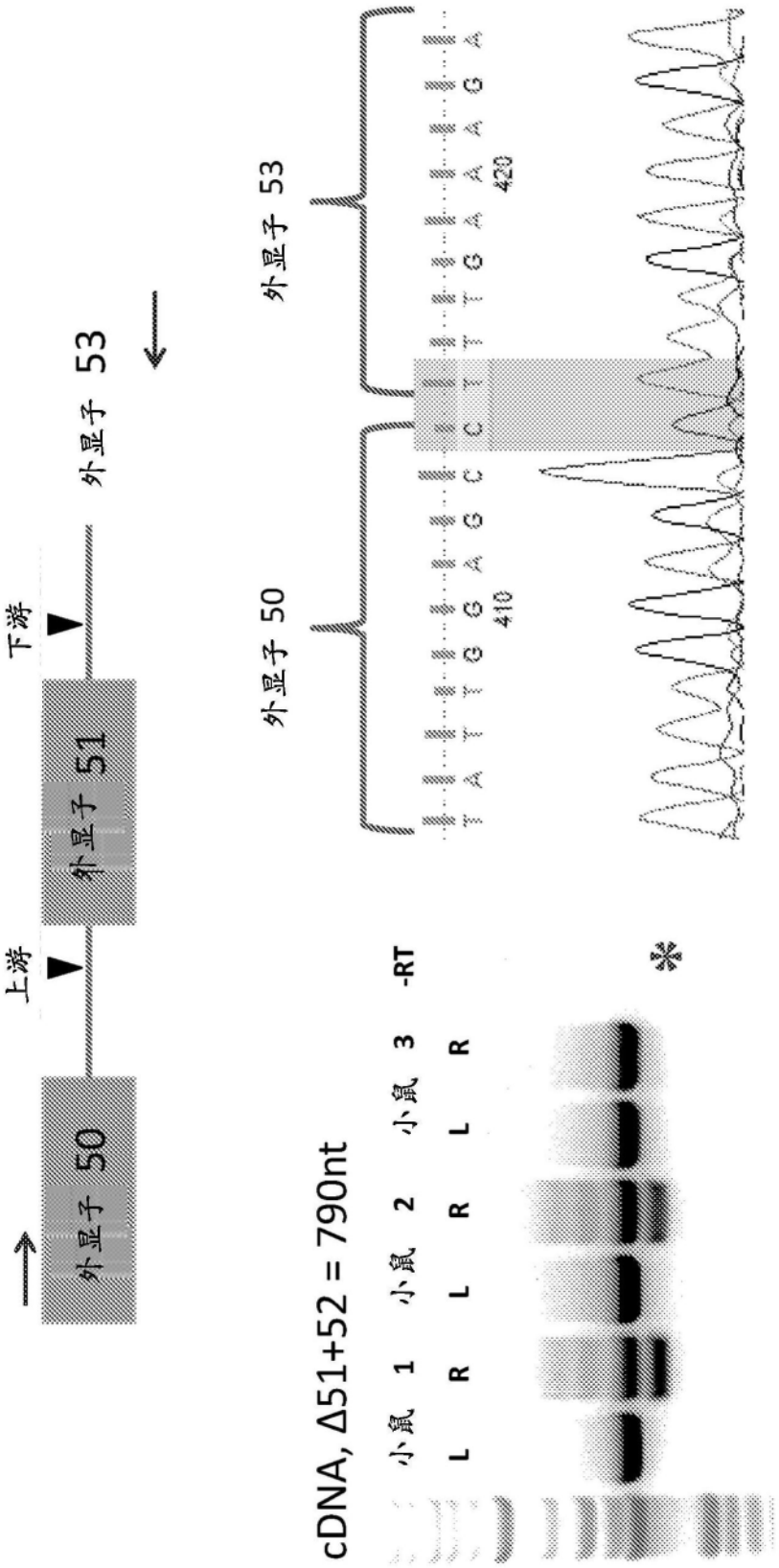

外显子 50
上游
外显子 51
下游
外显子 53
外显子 50
410
外显子 53
420
cDNA, Δ51+52 = 790nt
小鼠 1
小鼠 2
小鼠 3
-RT
L
R
L
R
L
R
*

图36

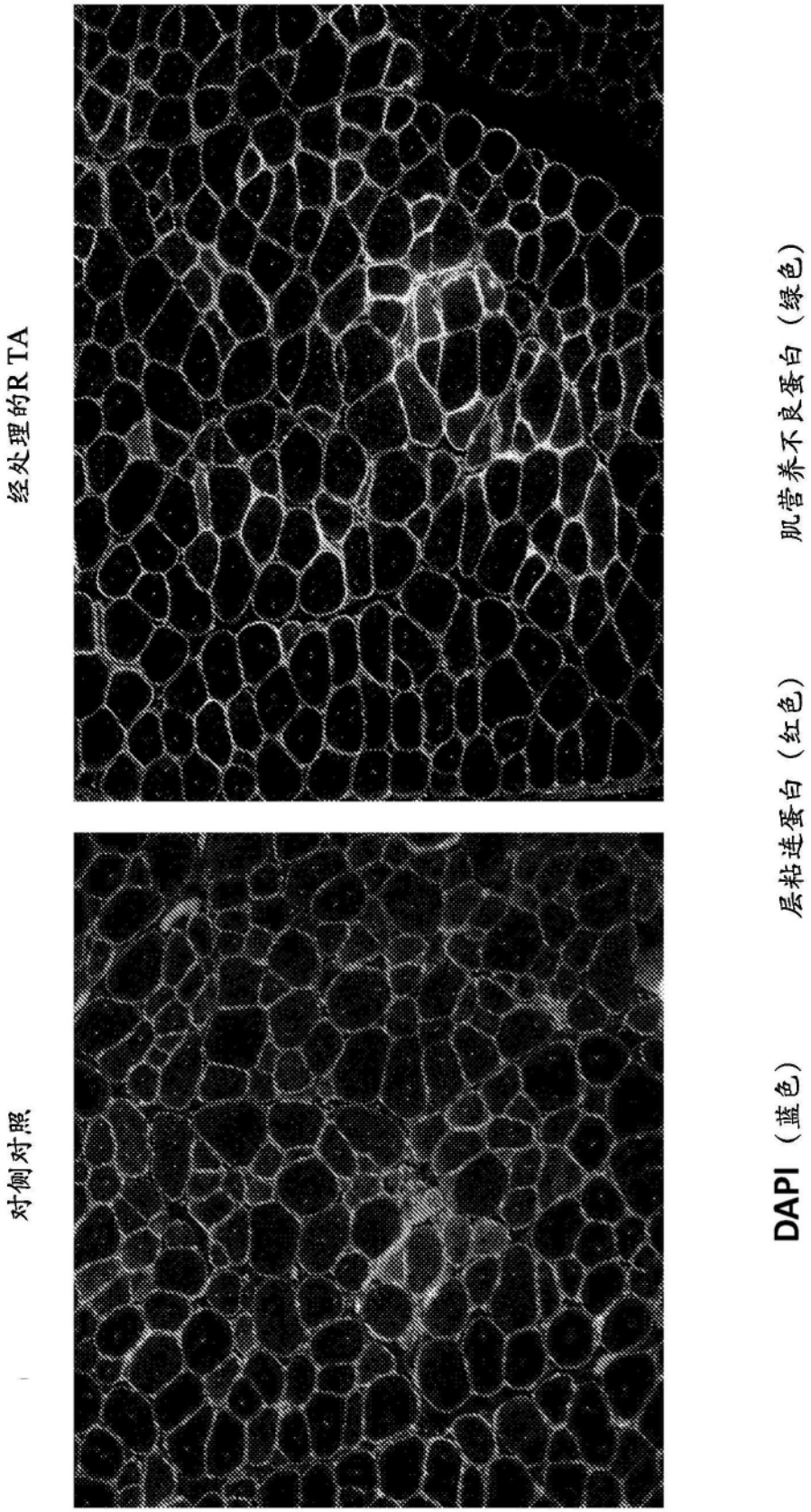


图37

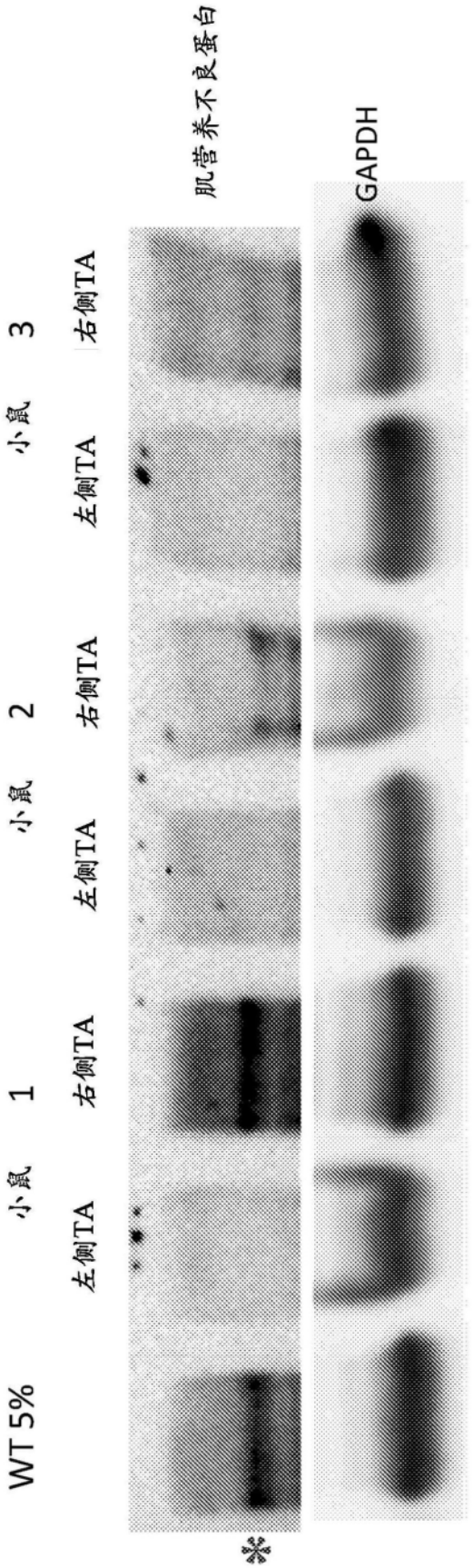


图38

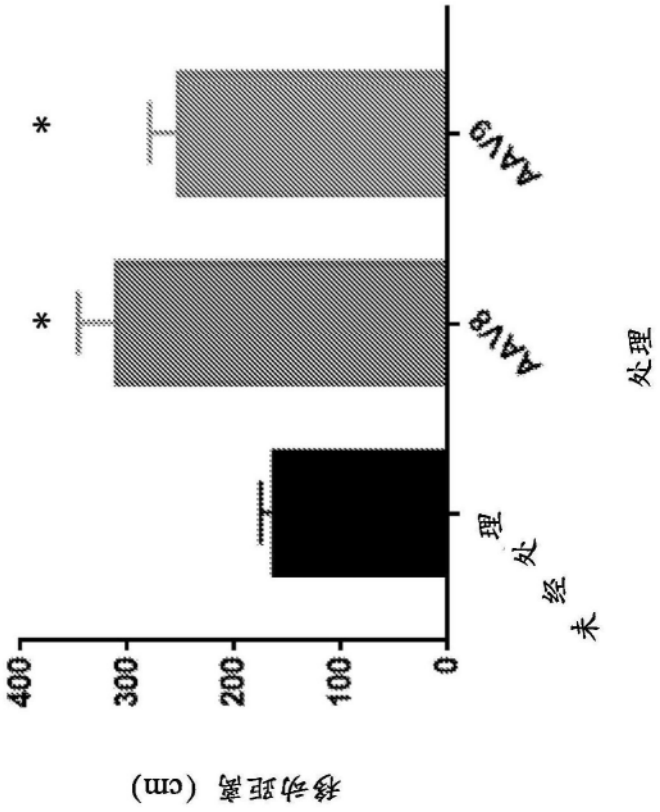

移动距离 (cm)
400
300
200
100
0
*
*
未经处理
AAV8
AAV9
处理

图39

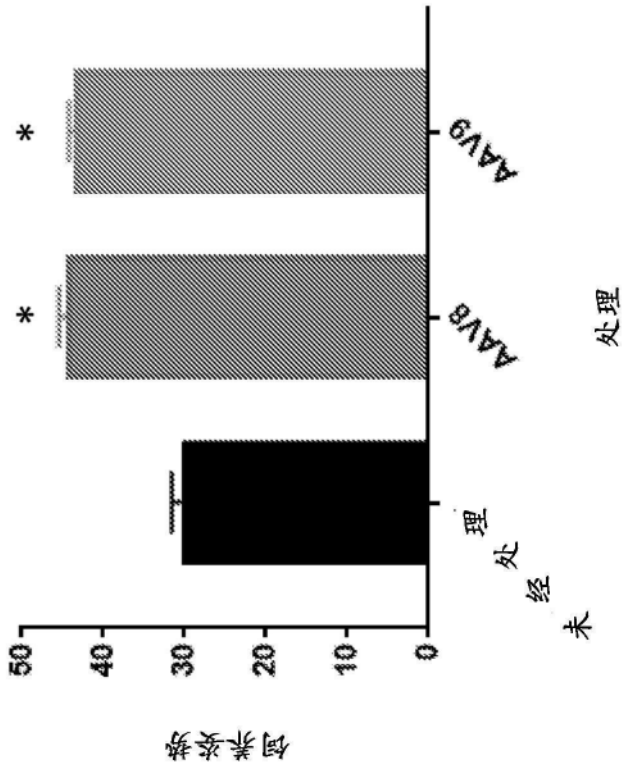

饲养姿势
50
40
30
20
10
0
*
*
未经处理
AAV8
AAV9
处理

图40

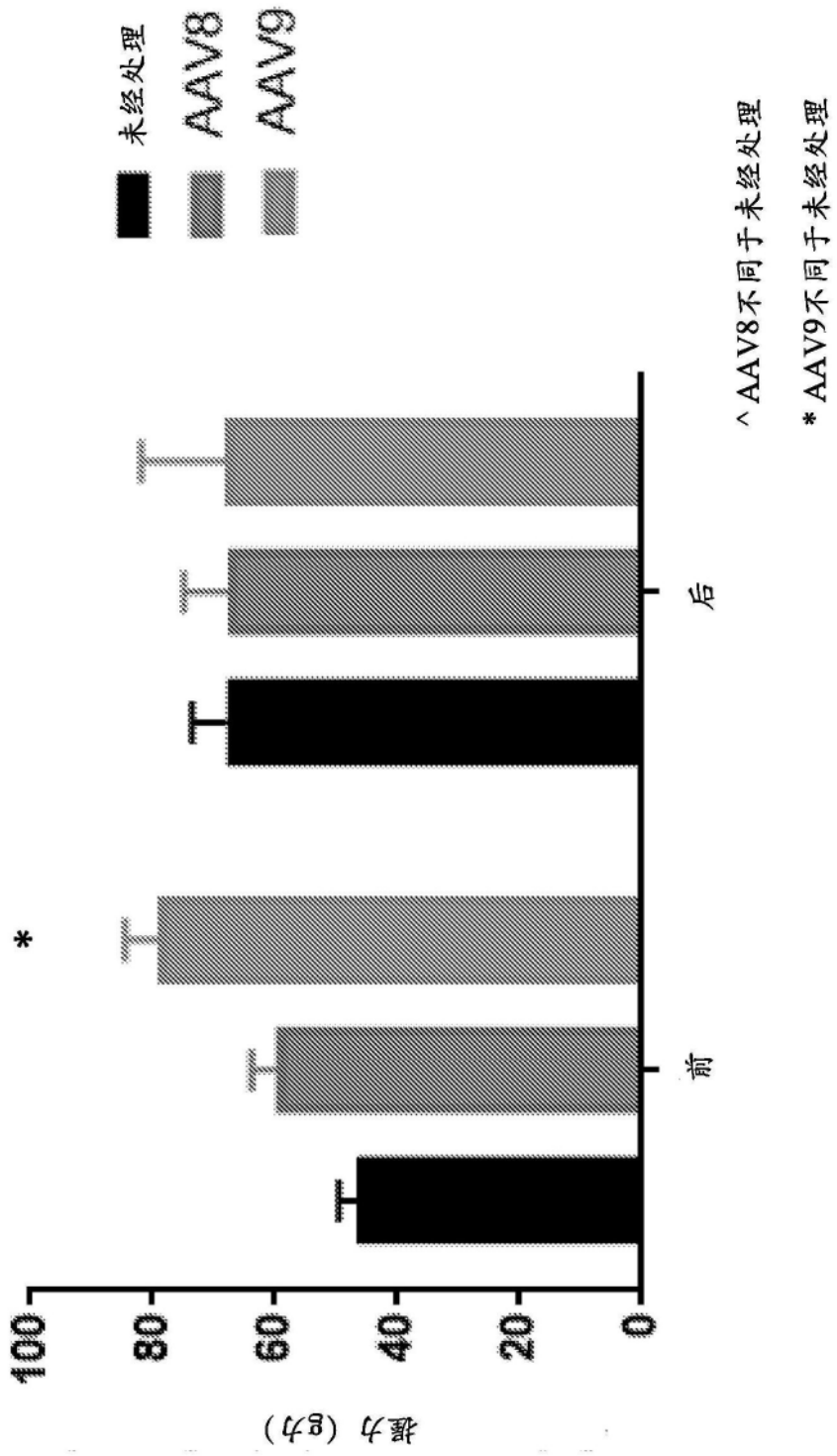

未经处理
AAV8
AAV9
^ AAV8不同于未经处理
* AAV9不同于未经处理
握力 (g力)
100
80
60
40
20
0
*
前
后

图41

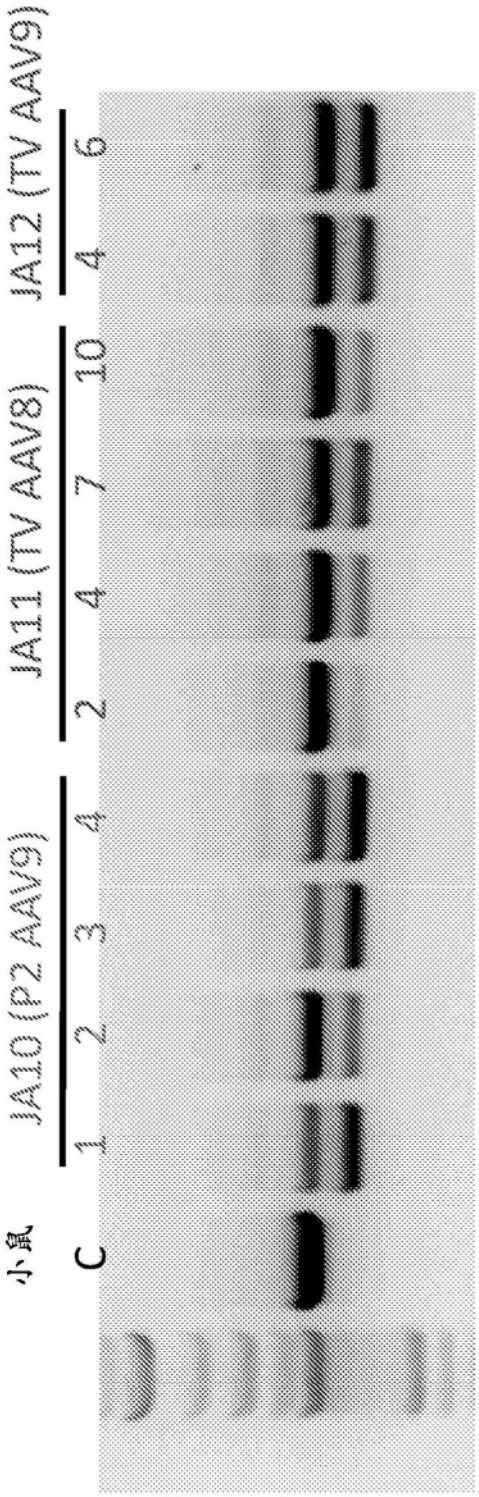


图42

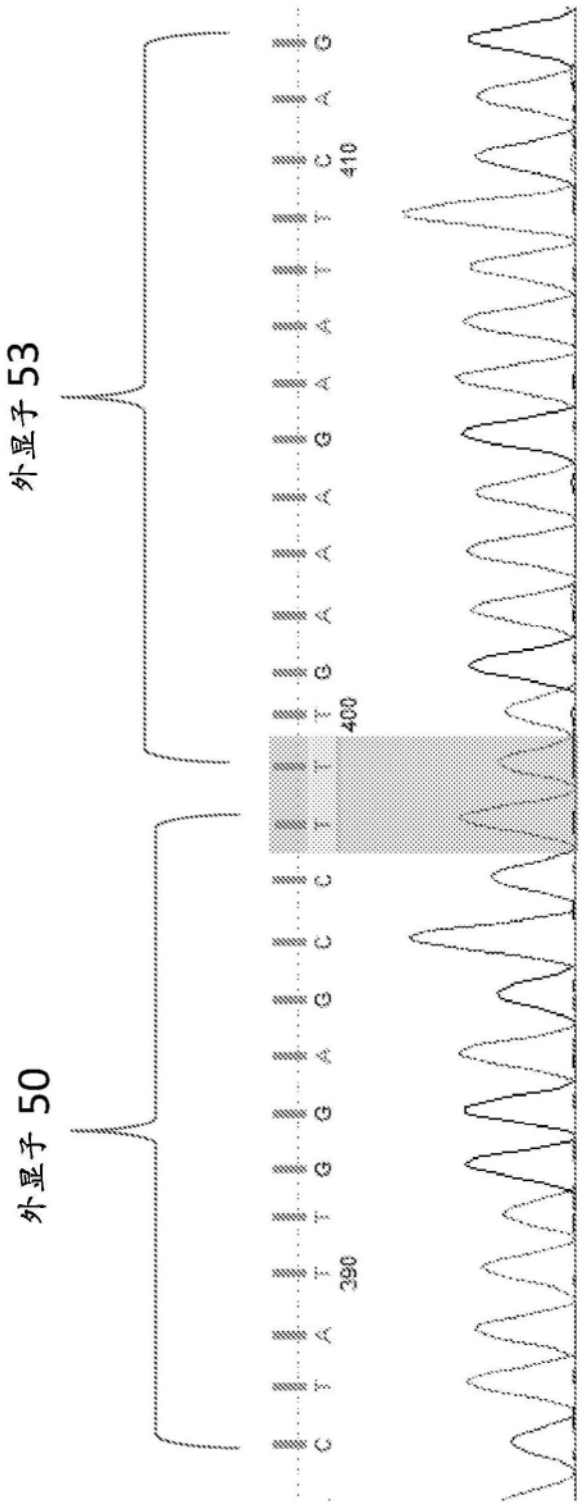


图43